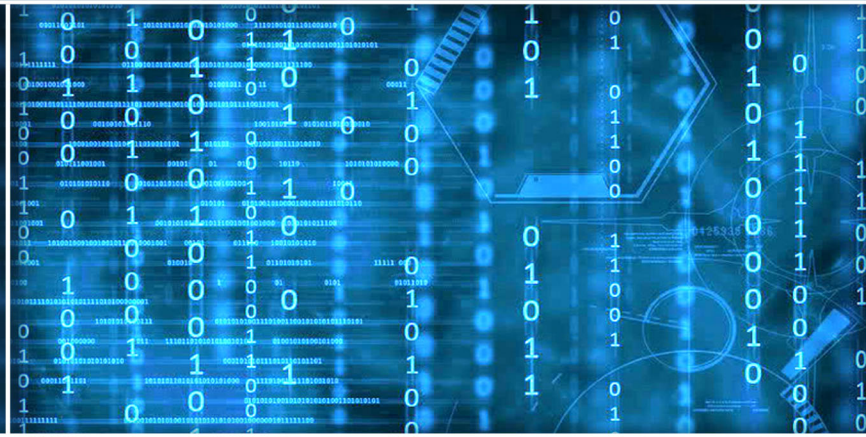


Volume 14 Issue 4

April 2023



ISSN 2156-5570(Online)

ISSN 2158-107X(Print)



Editorial Preface

From the Desk of Managing Editor...

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

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

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

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

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

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

Thank you for Sharing Wisdom!

Kohei Arai
Editor-in-Chief
IJACSA
Volume 14 Issue 4 April 2023
ISSN 2156-5570 (Online)
ISSN 2158-107X (Print)

Editorial Board

Editor-in-Chief

Dr. Kohei Arai - Saga University

Domains of Research: Technology Trends, Computer Vision, Decision Making, Information Retrieval, Networking, Simulation

Associate Editors

Alaa Sheta

Southern Connecticut State University

Domain of Research: Artificial Neural Networks, Computer Vision, Image Processing, Neural Networks, Neuro-Fuzzy Systems

Domenico Ciuonzo

University of Naples, Federico II, Italy

Domain of Research: Artificial Intelligence, Communication, Security, Big Data, Cloud Computing, Computer Networks, Internet of Things

Doroła Kaminska

Lodz University of Technology

Domain of Research: Artificial Intelligence, Virtual Reality

Elena Scutelnicu

"Dunarea de Jos" University of Galati

Domain of Research: e-Learning, e-Learning Tools, Simulation

In Soo Lee

Kyungpook National University

Domain of Research: Intelligent Systems, Artificial Neural Networks, Computational Intelligence, Neural Networks, Perception and Learning

Krassen Stefanov

Professor at Sofia University St. Kliment Ohridski

Domain of Research: e-Learning, Agents and Multi-agent Systems, Artificial Intelligence, e-Learning Tools, Educational Systems Design

Renato De Leone

Università di Camerino

Domain of Research: Mathematical Programming, Large-Scale Parallel Optimization, Transportation problems, Classification problems, Linear and Integer Programming

Xiao-Zhi Gao

University of Eastern Finland

Domain of Research: Artificial Intelligence, Genetic Algorithms

CONTENTS

Paper 1: An End-to-End Deep Learning System for Recommending Healthy Recipes Based on Food Images

Authors: Ledion Lico, Indrit Enesi, Sai Jawahar Reddy Meka

PAGE 1 – 7

Paper 2: An Automatic Framework for Number Plate Detection using OCR and Deep Learning Approach

Authors: Yash Shambharkar, Shailaja Salagrama, Kanhaiya Sharma, Om Mishra, Deepak Parashar

PAGE 8 – 14

Paper 3: Convolution Neural Networks for Phishing Detection

Authors: Arun D. Kulkarni

PAGE 15 – 19

Paper 4: A Radial Basis Network-based Early Warning Algorithm for Physical Injuries in Marathon Athletes

Authors: Ruisheng Jiao, Juan Luo

PAGE 20 – 27

Paper 5: Classification of Hand Movements Based on EMG Signals using Topological Features

Authors: Jianyang Li, Lei Yang, Yunan He, Osamu Fukuda

PAGE 28 – 36

Paper 6: Research on Automatic Intrusion Detection Method of Software-Defined Security Services in Cloud Environment

Authors: Xingjie Huang, Jing Li, Jinmeng Zhao, Beibei Su, Zixian Dong, Jing Zhang

PAGE 37 – 43

Paper 7: Experimental Analysis of WebHDFS API Throughput

Authors: Yordan Kalmukov, Milko Marinov

PAGE 44 – 50

Paper 8: Gradually Generative Adversarial Networks Method for Imbalanced Datasets

Authors: Muhammad Misdram, Muljono, Purwanto, Edi Noersasongko

PAGE 51 – 58

Paper 9: Human Fall Detection for Smart Home Caring using Yolo Networks

Authors: Bo LUO

PAGE 59 – 68

Paper 10: Investigation of You Only Look Once Networks for Vision-based Small Object Detection

Authors: Li YANG

PAGE 69 – 82

Paper 11: A Novel Data Aggregation Method for Underwater Wireless Sensor Networks using Ant Colony Optimization Algorithm

Authors: Lianchao Zhang, Jianwei Qi, Hao Wu

PAGE 83 – 93

Paper 12: A New Machine Learning-based Hybrid Intrusion Detection System and Intelligent Routing Algorithm for MPLS Network

Authors: Mohammad Azmi Ridwan, Nurul Asyikin Mohamed Radzi, Kaiyisah Hanis Mohd Azmi, Fairuz Abdullah, Wan Sifi Halimatul Munirah Wan Ahmad

PAGE 94 – 107

Paper 13: Plant Disease Classification and Adversarial Attack based CL-CondenseNetV2 and WT-MI-FGSM

Authors: Yong Li, Yufang Lu

PAGE 108 – 115

Paper 14: EMOCASH: An Intelligent Virtual-Agent Based Multiplayer Online Serious Game for Promoting Money and Emotion Recognition Skills in Egyptian Children with Autism

Authors: Hussein Karam Hussein Abd El-Sattar

PAGE 116 – 129

Paper 15: Adaptive Balance Optimizer: A New Adaptive Metaheuristic and its Application in Solving Optimization Problem in Finance

Authors: Purba Daru Kusuma, Ashri Dinimaharawati

PAGE 130 – 142

Paper 16: Improved Speaker Recognition for Degraded Human Voice using Modified-MFCC and LPC with CNN

Authors: Amit Moondra, Poonam Chahal

PAGE 143 – 151

Paper 17: Customer Segmentation of Personal Credit using Recency, Frequency, Monetary (RFM) and K-means on Financial Industry

Authors: Hafidh Rizkyanto, Ford Lumban Gaol

PAGE 152 – 162

Paper 18: Challenges of Digital Twin Technologies Integration in Modular Construction: A Case from a Manufacturer's Perspective

Authors: Laith Jamal Aldabbas

PAGE 163 – 167

Paper 19: SuffixAligner: A Python-based Aligner for Long Noisy Reads

Authors: Zeinab Rabea, Sara El-Metwally, Samir Elmougy, M. Z. Rashad

PAGE 168 – 172

Paper 20: Scrum: A Systematic Literature Review

Authors: Adrielle Cristina Sassa, Isabela Alves de Almeida, Tábata Nakagomi Fernandes Pereira, Milena Silva de Oliveira

PAGE 173 – 181

Paper 21: Artificial Intelligence Based Modelling for Predicting CO2 Emission for Climate Change Mitigation in Saudi Arabia

Authors: Sultan Alamri, Shahnawaz Khan

PAGE 182 – 189

Paper 22: Implementation of Revised Heuristic Knowledge in Average-based Interval for Fuzzy Time Series Forecasting of Tuberculosis Cases in Sabah

Authors: Suriana Lasaraiya, Suzelawati Zenian, Risman Mat Hasim, Azmirul Ashaari

PAGE 190 – 196

Paper 23: Event Feature Pre-training Model Based on Public Opinion Evolution

Authors: WANG Nan, TAN Shu-Ru, XIE Xiao-Lan, LI Hai-Rong, JIANG Jia-Hui

PAGE 197 – 206

Paper 24: A Novel Network Intrusion Detection System Based on Semi-Supervised Approach for IoT

Authors: Durga Bhavani A, Neha Mangla

PAGE 207 – 216

Paper 25: Iris Recognition Through Edge Detection Methods: Application in Flight Simulator User Identification

Authors: Sundas Naqeeb Khan, Samra Urooj Khan, Onyeka Josephine Nwobodo, Krzysztof Adam. Cyran

PAGE 217 – 229

Paper 26: Personalized Music Recommendation Based on Interest and Emotion: A Comparison of Multiple Algorithms

Authors: Xiuli Yan

PAGE 230 – 235

Paper 27: Research on Customer Retention Prediction Model of VOD Platform Based on Machine Learning

Authors: Quansheng Zhao, Zhijie Zhao, Liu Yang, Lan Hong, Wu Han

PAGE 236 – 244

Paper 28: Development of Computer Vision-enabled Augmented Reality Games to Increase Motivation for Sports

Authors: Bauyrzhan Doskarayev, Nurlan Omarov, Bakhytzhhan Omarov, Zhuldyz Ismagulova, Zhadra Kozhamkulova, Elmira Nurlybaeva, Galiya Kasimova

PAGE 245 – 252

Paper 29: Opposition Learning Based Improved Bee Colony Optimization (OLIBCO) Algorithm for Data Clustering

Authors: Srikanta Kumar Sahoo, Priyabrata Pattanaik, Mihir Narayan Mohanty, Dilip Kumar Mishra

PAGE 253 – 261

Paper 30: Fuzzy Rank-Based Ensemble Model for Accurate Diagnosis of Osteoporosis in Knee Radiographs

Authors: Saumya Kumar, Puneet Goswami, Shivani Batra

PAGE 262 – 270

Paper 31: A Novel Deep CNN-RNN Approach for Real-time Impulsive Sound Detection to Detect Dangerous Events

Authors: Nurzhigit Smailov, Zhandos Dosbayev, Nurzhan Omarov, Bibigul Sadykova, Maigul Zhekambayeva, Dusmat Zhamangarin, Assem Ayapbergenova

PAGE 271 – 280

Paper 32: Reversible De-identification of Specific Regions in Biomedical Images and Secured Storage by Randomized Joint Encryption

Authors: Prabhavathi K, Anandaraju M. B

PAGE 281 – 293

Paper 33: Texture Analytics for Accurate Person Recognition: A Multimodal Approach

Authors: Suchetha N V, Sharmila Kumari M

PAGE 294 – 299

Paper 34: Deep Learning Models for Crime Intention Detection Using Object Detection

Authors: Abdirahman Osman Hashi, Abdullahi Ahmed Abdirahman, Mohamed Abdirahman Elmi, Octavio Ernest Romo Rodriguez

PAGE 300 – 306

Paper 35: Hand Gesture Recognition Based on Various Deep Learning YOLO Models

Authors: Soukaina Chraa Mesbahi, Mohamed Adnane Mahraz, Jamal Riffi, Hamid Tairi

PAGE 307 – 319

Paper 36: Optimized Image Authentication Algorithm using Redundant Wavelet Transform Based Sift Descriptors and Complex Zernike Moments

Authors: Pooja Vijayakumaran Kallath, Kondaka Lakshmisudha

PAGE 320 – 326

Paper 37: Anchor-free Proposal Generation Network for Efficient Object Detection

Authors: Hoanh Nguyen

PAGE 327 – 335

Paper 38: Detecting Fraud Transaction using Ripper Algorithm Combines with Ensemble Learning Model

Authors: Vo Hoang Khang, Cao Tung Anh, Nguyen Dinh Thuan

PAGE 336 – 345

Paper 39: A Single-valued Pentagonal Neutrosophic Geometric Programming Approach to Optimize Decision Maker's Satisfaction Level

Authors: Satyabrata Nath, Purnendu Das, Pradip Debnath

PAGE 346 – 356

Paper 40: Multifaceted Sentiment Detection System (MSDS) to Avoid Dropout in Virtual Learning Environment using Multi-class Classifiers

Authors: Ananthi Claral Mary. T, Arul Leena Rose. P. J

PAGE 357 – 368

Paper 41: Implementation of CNN for Plant Identification using UAV Imagery

Authors: Mohd Anul Haq, Ahsan Ahmed, Jayadev Gyani

PAGE 369 – 378

Paper 42: An IoT-based Framework for Detecting Heart Conditions using Machine Learning

Authors: Mona Alnaggar, Mohamed Handosa, T. Medhat, M. Z. Rashad

PAGE 379 – 389

Paper 43: Solar Energy Forecasting Based on Complex Valued Auto-encoder and Recurrent Neural Network

Authors: Aymen Rhouma, Yahia Said

PAGE 390 – 395

Paper 44: Classification with K-Nearest Neighbors Algorithm: Comparative Analysis between the Manual and Automatic Methods for K-Selection

Authors: Tsvetelina Mladenova, Irena Valova

PAGE 396 – 404

Paper 45: The Impact of Design-level Class Decomposition on the Software Maintainability

Authors: Bayu Priyambadha, Tetsuro Katayama

PAGE 405 – 413

Paper 46: Listening to the Voice of People with Vision Impairment

Authors: Abeer Malkawi, Azrina Kamaruddin, Alfian Abdul Halin, Novia Admodisastro

PAGE 414 – 423

Paper 47: A Deep Learning based Approach for Recognition of Arabic Sign Language Letters

Authors: Boutaina Hdioud, Mohammed El Haj Tirari

PAGE 424 – 429

Paper 48: Improving QoS in Internet of Vehicles Integrating Swarm Intelligence Guided Topology Adaptive Routing and Service Differentiated Flow Control

Authors: Tanuja Kayarga, Ananda Kumar S

PAGE 430 – 437

Paper 49: A Comparative Performance Evaluation of Routing Protocols for Mobile Ad-hoc Networks

Authors: Baidaa Hamza Khudayer, Lial Raja Alzabin, Mohammed Anbar, Ragad M Tawafak, Tat-Chee Wan, Abir AISideiri, Sohail Iqbal Malik, Taief Alaa Al-Amiedy

PAGE 438 – 446

Paper 50: Deep Learning for Combined Water Quality Testing and Crop Recommendation

Authors: Tahani Alkhudaydi, Maram Qasem Albalawi, Jamelah Sanad Alanazi, Wejdan Al-Anazi, Rahaf Mansour Alfarshouti

PAGE 447 – 455

Paper 51: Context Aware Automatic Subjective and Objective Question Generation using Fast Text to Text Transfer Learning

Authors: Arpit Agrawal, Pragya Shukla

PAGE 456 – 463

Paper 52: Multistage End-to-End Driver Drowsiness Alerting System

Authors: Sowmyashree P, Sangeetha J

PAGE 464 – 473

Paper 53: Insights on Data Security Schemes and Authentication Adopted in Safeguarding Social Network

Authors: Nithya S, Rekha B

PAGE 474 – 483

Paper 54: A Novel Framework for Semi-supervised Multiple-label Image Classification using Multi-stage CNN and Visual Attention Mechanism

Authors: Joseph James S, Lakshmi C

PAGE 484 – 493

Paper 55: Comparison of Predictive Machine Learning Models to Predict the Level of Adaptability of Students in Online Education

Authors: Orlando Iparaguire-Villanueva, Carmen Torres-Ceclén, Andrés Epifanía-Huerta, Gloria Castro-Leon, Melquiades Melgarejo-Graciano, Joselyn Zapata-Paulini, Michael Cabanillas-Carbonell

PAGE 494 – 503

Paper 56: Analyzing WhisperGate and BlackCat Malware: Methodology and Threat Perspective

Authors: Mathew Nicho, Rajesh Yadav, Digvijay Singh

PAGE 504 – 519

Paper 57: A Cloud Native Framework for Real-time Pricing in e-Commerce

Authors: Archana Kumari, Mohan Kumar. S

PAGE 520 – 529

Paper 58: A Deep Learning Approach for Sentiment Classification of COVID-19 Vaccination Tweets

Authors: Haidi Said, BenBella S. Tawfik, Mohamed A. Makhlof

PAGE 530 – 538

Paper 59: Validate the Users' Comfortable Level in the Virtual Reality Walkthrough Environment for Minimizing Motion Sickness

Authors: Muhammad Danish Affan Anua, Ismahafezi Ismail, Nur Saadah Mohd Shapri, Wan Mohd Amir Fazamin Wan Hamzah, Maizan Mat Amin, Fazida Karim

PAGE 539 – 547

Paper 60: Fusion Privacy Protection of Graph Neural Network Points of Interest Recommendation

Authors: Yong Gan, ZhenYu Hu

PAGE 548 – 556

Paper 61: Identity Authentication Protocol of Smart Home IoT based on Chebyshev Chaotic Mapping

Authors: Jingjing Sun, Peng Zhang, Xiaohong Kong

PAGE 557 – 565

Paper 62: Hybrid Optimization with Recurrent Neural Network-based Medical Image Processing for Predicting Interstitial Lung Disease

Authors: K. Sundaramoorthy, R. Anitha, S. Kayalvili, Ayat Fawzy Ahmed Ghazala, Yousef A.Baker El-Ebiary, Sameh Al-Ashmawy

PAGE 566 – 574

Paper 63: Study on Tomato Disease Classification based on Leaf Image Recognition based on Deep Learning Technology

Authors: Ji Zheng, Minjie Du

PAGE 575 – 583

Paper 64: Research on Recommendation Model of College English MOOC based on Hybrid Recommendation Algorithm

Authors: Yifang Ding, Jingbo Hao

PAGE 584 – 593

Paper 65: Employee Information Security Awareness in the Power Generation Sector of PT ABC

Authors: Ridwan Fadlika, Yova Ruldeviyani, Zenfrison Tuah Butarbutar, Relaci Aprilia Istiqomah, Achmad Arzal Fariz

PAGE 594 – 603

Paper 66: ECAH: A New Energy-Aware Coverage Method for Wireless Sensor Networks using Artificial Bee Colony and Harmony Search

Authors: ZHOU Bing, ZHANG Zhigang

PAGE 604 – 616

Paper 67: Patient Health Monitoring System Development using ESP8266 and Arduino with IoT Platform

Authors: Jamil Abedalrahim Jamil Alsayaydeh, Mohd Faizal bin Yusof, Muhammad Zulhakim Bin Abdul Halim, Muhammad Noorazlan Shah Zainudin, Safarudin Gazali Herawan

PAGE 617 – 624

Paper 68: Integrated Methodology for Information Security Risk Management using ISO 27005:2018 and NIST SP 800-30 for Insurance Sector

Authors: Arief Prabawa Putra, Benfano Soewito

PAGE 625 – 633

Paper 69: An Effectivity Deep Learning Optimization Model to Traditional Batak Culture Ulos Classification

Authors: Rizki Muliono, Mayang Septania Iranita, Rahmad BY Syah

PAGE 634 – 638

Paper 70: Enhancing Customer Relationship Management Using Fuzzy Association Rules and the Evolutionary Genetic Algorithm

Authors: Ahmed Abu-Al Dahab, Riham M Haggag, Samir Abu-Al Fotouh

PAGE 639 – 649

Paper 71: Leader-follower Optimal Control Method for Vehicle Platoons to Improve Fuel Efficiency

Authors: Zhigang Li, Yushi Guo, Hua Wang, Jianyong Li, Yuye Xie, Jingyu Liu

PAGE 650 – 658

Paper 72: A Review of the Recent Progress on Crowd Anomaly Detection

Authors: Sarah Altowairqi, Suhuai Luo, Peter Greer

PAGE 659 – 669

Paper 73: Improving Brain Tumor Segmentation in MRI Images through Enhanced Convolutional Neural Networks

Authors: Kabirat Sulaiman Ayomide, Teh Noranis Mohd Aris, Maslina Zolkepli

PAGE 670 – 678

Paper 74: Rain Streaks Removal in Images using Extended Generative Adversarial-based Deraining Framework

Authors: Subbarao Gogulamudi, V. Mahalakshmi, Indraneel Sreeram

PAGE 679 – 690

Paper 75: A PSL-based Approach to Human Activity Recognition in Smart Home Environments

Authors: Yan Li

PAGE 691 – 698

Paper 76: Design of Virtual Experiment Teaching of Inorganic Chemistry in Colleges and Universities Based on Unity3D

Authors: Xia Hu

PAGE 699 – 708

Paper 77: The Effective 3D MRI Reconstruction Method Driven by the Fusion Strategy in NSST Domain

Authors: Jin Huang, Lei Wang, Muhammad Tahir, Tianqi Cheng, Xinping Guo, Yuwei Wang, ChunXiang Liu

PAGE 709 – 715

Paper 78: Integrating Dropout Regularization Technique at Different Layers to Improve the Performance of Neural Networks

Authors: B. H. Pansambal, A. B. Nandgaokar

PAGE 716 – 722

Paper 79: Comparison Review on Brain Tumor Classification and Segmentation using Convolutional Neural Network (CNN) and Capsule Network

Authors: Nurul Fatimah Binti Ali, Siti Salasiah Mokri, Syahirah Abd Halim, Noraishikin Zulkarnain, Ashrani Aizuddin Abd Rahni, Seri Mastura Mustaza

PAGE 723 – 731

Paper 80: Fuzzy Reasoning based Reliability Fault Prediction of CNC Machine Tools

Authors: Jie Yu, Tiebin Wang, Weidong Wang, Gege Zhao, Yue Yao

PAGE 732 – 737

Paper 81: Using Machine Learning Algorithm as a Method for Improving Stroke Prediction

Authors: Nojood Alageel, Rahaf Alharbi, Rehab Alharbi, Maryam Alsayil, Lubna A. Alharbi

PAGE 738 – 744

Paper 82: Deep Learning Localization Algorithm Integrating Attention Mechanism in Database Information Query

Authors: Yang Li, Xianghui Hui, Xiaolei Wang, Fei Yin

PAGE 745 – 752

Paper 83: Develop an Olive-based Grading Algorithm using Image Processing

Authors: Dongliang Jin

PAGE 753 – 760

Paper 84: Intelligent Abnormal Residents' Behavior Detection in Smart Homes for Risk Management using Fuzzy Logic Algorithm

Authors: Bo Feng, Lili Miao, HuiXiang Liu

PAGE 761 – 768

Paper 85: BREPubSub: A Secure Publish-Subscribe Model using Blockchain and Re-encryption for IoT Data Sharing Management

Authors: Hoang-Anh Pham

PAGE 769 – 776

Paper 86: A Review of Trending Crowdsourcing Topics in Software Engineering Highlighting Mobile Crowdsourcing and AI Utilization

Authors: Mohammed Alghasham, Mousa Alzakan, Mohammed Al-Hagery

PAGE 777 – 786

Paper 87: Evaluation of Wood Species Identification Using CNN-Based Networks at Different Magnification Levels

Authors: Khanh Nguyen-Trong

PAGE 787 – 796

Paper 88: A Review of Milgram and Kishino's Reality-Virtuality Continuum and a Mathematical Formalization for Combining Multiple Reality-Virtuality Continua

Authors: Cristian Pamparãu

PAGE 797 – 805

Paper 89: Evolutionary Design of a PSO-Tuned Multigene Symbolic Regression Genetic Programming Model for River Flow Forecasting

Authors: Alaa Sheta, Amal Abdel-Raouf, Khalid M. Fraihat, Abdelkarim Baareh

PAGE 806 – 814

Paper 90: Autonomous Motion Planning for a Differential Robot using Particle Swarm Optimization

Authors: Fredy Martinez, Angelica Rendon

PAGE 815 – 821

Paper 91: Software Effort Estimation using Machine Learning Technique

Authors: Mizanur Rahman, Partha Protim Roy, Mohammad Ali, Teresa Goncalves, Hasan Sarwar

PAGE 822 – 827

Paper 92: An Approach to Hyperparameter Tuning in Transfer Learning for Driver Drowsiness Detection Based on Bayesian Optimization and Random Search

Authors: Hoang-Tu Vo, Hoang Tran Ngoc, Luyl-Da Quach

PAGE 828 – 837

Paper 93: Discovering COVID-19 Death Patterns from Deceased Patients: A Case Study in Saudi Arabia

Authors: Abdulrahman Alomary, Tarik Alafif, Abdulmohsen Almalawi, Anas Hadi, Faris Alkhilawi, Yasser Alatawi

PAGE 838 – 850

Paper 94: Highly Accurate Deep Learning Model for Olive Leaf Disease Classification: A Study in Tacna-Perú

Authors: Erbert F. Osco-Mamani, Israel N. Chaparro-Cruz

PAGE 851 – 860

Paper 95: Enhanced MQTT Architecture for Smart Supply Chain

Authors: Raouya AKNIN, Youssef Bentaleb

PAGE 861 – 869

Paper 96: Hybrid Machine Learning-Based Approach for Anomaly Detection using Apache Spark

Authors: Hanane Chliah, Amal Battou, Maryem Ait el hadj, Adil Laoufi

PAGE 870 – 878

Paper 97: Egypt Monuments Dataset version 1: A Scalable Benchmark for Image Classification and Monument Recognition

Authors: Mennat Allah Hassan, Alaa Hamdy, Mona Nasr

PAGE 879 – 884

Paper 98: Exploring the Joint Potential of Blockchain and AI for Securing Internet of Things

Authors: Md. Tauseef, Manjunath R Kounte, Abdul Haq Nalband, Mohammed Riyaz Ahmed

PAGE 885 – 895

Paper 99: Opportunities and Challenges in Human-Swarm Interaction: Systematic Review and Research Implications

Authors: Alexandru-Ionuț, Ștefan Bogdan, Bogdan-Constantin Grădinaru, Ovidiu-Ionuț, Gherman, Mirela Danubianu, Laurențiu-Dan Milici

PAGE 896 – 902

Paper 100: IM2P-Medical: Towards Individual Management Privacy Preferences for the Medical Web Apps

Authors: Nguyen Ngoc Phien, Nguyen Thi Hoang Phuong, Khiem G. Huynh, Khanh H. Vo, Phuc T. Nguyen, Khoa D. Tran, Bao Q. Tran, Loc C. P. Van, Duy T. Q. Nguyen, Hieu M. Doan, Bang K. Le, Trong D. P. Nguyen, Ngan T. K. Nguyen, Huong H. Luong, Duong Hon Minh

PAGE 903 – 911

Paper 101: Anomaly Discover: A New Community-based Approach for Detecting Anomalies in Social Networks

Authors: Hedia Zardi, Hajar Alrajhi

PAGE 912 – 920

Paper 102: A Particle Swarm Optimization with Imbalance Initialization and Task Rescheduling for Task Offloading in Device-Edge-Cloud Computing

Authors: Hui Fu, Guangyuan Li, Fang Han, Bo Wang

PAGE 921 – 926

Paper 103: Prediction of Air Quality and Pollution using Statistical Methods and Machine Learning Techniques

Authors: V. Devasekhar, P. Natarajan

PAGE 927 – 937

Paper 104: Fraud Mitigation in Attendance Monitoring Systems using Dynamic QR Code, Geofencing and IMEI Technologies

Authors: Augustine Nwabuwe, Baljinder Sanghera, Temitope Alade, Funminiyi Olajide

PAGE 938 – 945

Paper 105: Reverse Supply Chain Management Through a Quantity Flexibility Contract: A Case of Stochastic Remanufacturing Capacity

Authors: Changhao Zhang

PAGE 946 – 951

Paper 106: Simulation Method of Port Petrochemical Industry Throughput Development under the Background of Integration of Port, Industry and City

Authors: Tingting Zhou, Chen Guo

PAGE 952 – 962

An End-to-End Deep Learning System for Recommending Healthy Recipes Based on Food Images

Ledion Lico¹, Indrit Enesi², Sai Jawahar Reddy Meka³

Goergen Institute for Data Science, University of Rochester, New York, USA^{1,3}

Electronic and Telecommunication Department, Polytechnic University of Tirana, Tirana, Albania²

Abstract—Healthy food leads to healthy living and it is a major issue in our days. Nutri-Score is a nutrition label that can be calculated from the nutritional values of a food and helps evaluating the healthiness of it. Nevertheless, we don't always have the nutritional values of the food, so it is not always easy identifying this label. In the same way, it is not easy finding the healthier option to a favorite food. In this paper an end-to end deep learning system is proposed to identify the Nutri-Score label and recommend similar but healthier recipes based on food images. A new dataset of images is extracted from the Recipe 1M and labeled with the Nutri-Score value calculated for each image. Pretrained models Resnet50, Resnet101, EfficientNetB2 and DensNet121 are tuned based on this dataset. The embeddings from the last convolutional layer of the input image are used to find its most similar neighbor based on KNN algorithm. The proposed system suggests recipes with the lowest Nutri-Score similar to the inputted image. Implementations show that the Resnet50 provides highest prediction accuracy.

Keywords—Deep learning; nutri-score; new dataset; healthy food; accuracy

I. INTRODUCTION

Nutritional food plays an important role in human health. There is an expression from nutritionists “We are what we eat”. Nutri-Score is a nutrition label that converts the nutritional value of products into a simple code of five colored letters (A-E), where A is the healthier food [1]. Each product is awarded a score based on a scientific algorithm. It takes into account the negative nutrients with a high energy value, large amount of sugar, saturated fats, salt and the positive ones with fibers, proteins, fruits, vegetables, nuts, rapeseed oil, walnut oil and olive oil. A set of 50718 food images is extracted from the Recipe 1M dataset [2]. For every image of this dataset the nutri-score value is calculated using Nutri-Score system and the dataset is labeled with labels A, B, C, D, E where A is the healthier food with lowest nutritional value [3].

A new end-to-end deep learning neural network (DNN) architecture is proposed as a combination of existing deep learning systems with k-nearest neighbors (KNN) algorithm. Its input is a food image and the output of the system is a set of recipes and their respective images classified with labels suggesting the user the closest foods regarding the input one. The DNNs considered are Resnet50 -Residual Network with 50 layers (ResNet50) [4], Resnet101-Residual Network with 101layers (ResNet101) [5], EfficientNetB2 [6] and Densely

Connected Convolutional Networks with 121 layers (DenseNet121) [7]. These systems are pre-trained using ImageNet dataset. The systems are then tuned using the new labeled dataset with 50718 labeled images. The embeddings of the last convolutional layer are inputted to the KNN classifier to train it. Whenever a user inputs a food image to the proposed system, the embeddings for this new image are obtained to find its nearest neighbors. The k-neighbors with the lowest Nutri-score and the corresponding recipes will be returned to the user. The accuracy for the four DNNs is calculated yielding that the combination of ResNet50 with KNN algorithm is the most accurate one regarding the new food dataset. Based on our internet searches this is the only algorithm that proposes to the user a healthy recipe only from food images.

A novel approach for recommending healthy recipes based on food images using a combination of deep learning and k-nearest neighbors' algorithm is proposed. The proposed method can help people make healthier food choices by providing personalized recommendations that take into account the nutritional value of different food products. Moreover, the experiments show that the combination of ResNet50 with KNN algorithm is the most accurate approach for this task. The paper introduces a new and innovative approach for recommending healthy recipes. Based on the experimental results it is shown that the proposed approach is effective and accurate.

The paper is organized as: Related works are treated in Section II. Section III describes the Nutri-Score evaluation system; Deep Neural Networks are described in Section IV; the new dataset is described in Section V; methodology is described in Section VI; experimental results in Section VII; conclusions and future work closes the paper with Section IX.

II. RELATED WORK

Several datasets and approaches have been developed in the recent years to address the topic. The [8] introduced a novel method to automatically recognize dishes pictures using Random Forests which allowed them to mine for parts simultaneously for all classes and to share knowledge among them. They also introduced the Food-101 visual classification dataset of 101 food categories, with 101k images. Im2Calories was introduced in [9] in which a classifier was trained on food images taken from 23 different restaurant and used to predict which food are present on the plate and the compute the corresponding calories. The system first performs a food

segmentation followed by a volume estimation. This is obtained by combining two Convolutional Neural Networks (CNNs) for the meal detection and food recognition, with a food image segmentation technique and Google's Places API used to recognize the restaurant. A bigger dataset was introduced in [10]. The dataset contained 720 639 training, 155 036 validation and 154 045 test recipes, containing a title, a list of ingredients, a list of cooking instructions and (optionally) an image. The [11] in collaboration with Facebook research introduced a recipe generation system that takes a food image as an input and outputs a sequence of cooking instructions, which are generated by means of an instruction decoder that takes as input two embeddings. The first one represents visual features extracted from an image, while the second one encodes the ingredients extracted from the image. A transformer-based instruction decoder is used to generate the cooking instructions. The [12] proposed RecipeNet, a system that returns to the user the recipe based on the inputted image. They use the embeddings obtained from the final convolutional layer of ResNet-50, ResNet-101, and DenseNet-121, then use K-NN to return the most similar images and recipes. Although, they don't take Nutri-Score into account when tuning the system and they cannot return the healthiest image. On the other hand they don't compare the models based on this classification task.

III. NUTRI-SCORE

Nutri-Score is a system for categorizing foods based on their nutritional value [1][13]. It guides consumers towards healthier food choices preventing the wide range of nutrition-chronic diseases. It was selected from French government from March 2017 to be displayed on products. It relies on the computation of nutrient profiling system from the United Kingdom Food Standard Agency [14]. Its output is: five classification letters corresponding to five colored labels from A to E, where A is the healthiest food and E is the detrimental one. Calculation of the score is based on the quantity of seven nutrient components found on 100g of food, which are high content of fruits and vegetables, fibers, and proteins (based on a rule of 2019 are added and the healthy oils) for a preferable score, while high content of sugar, saturated fatty acids, sodium yields a detrimental score. Special rules are added for "cheese", "added fats" and beverages. The system doesn't take in consideration the degree of food processing, vitamins, antioxidants, fiber type of additives. The calculation consists of three steps. In the first step the nutritional score of the food is assessed.

Ingredients which affect negatively the Nutri-Score compounding the Total_N_Score value are: high energy density per 100 (g of l), high sugar content, high content of saturated fatty acids and high salt content. Ingredients which affect positively the Nutri-Score results compounding the Total_P_Score value are content of fruits, vegetables, nuts and legumes; fiber content; protein content; content rapeseed, walnut and olive oil.

$$\text{Nutritional_Score} = \text{Total_N_Score} - \text{Total_P_Score} \quad (1)$$

In the second step, the Rayner's score is calculated for all the products in the same way, except cheese, vegetable, animal

fats, oils, and drinks. In the third step, the two scores are used to classify the food in one of five levels of the Nutri-Score system.

Based on the Nutri-Score evaluation system algorithm, 0 – 10 points are assigned for energy value and ingredients which should be limited in diet; 0 - 5 points for beneficial ingredients. To determine the value of a product and its corresponding letter, the sum of points for beneficial ingredients subtracts the sum of points for nonhealthy ingredients, the final score varies from -15 to +40 and with corresponding letters from A to E, where the lower the score the better nutritional food.

Nutri-Score is easily computable by industrial and public stakeholders; it encourages the food industry to improve the nutritional quality of the food supply.

IV. DEEP NEURAL NETWORKS

Deep Learning is a subset of machine learning that automates part of the feature extraction step of the process, thus eliminating some of the manual human intervention and also enables the usage of big data sets [15]. It consists of processing nodes arranged in layers. The system processes data between input and output layers to solve or predict specific task. The neural network needs to learn all the time solving specific tasks in a more qualified way or providing better results. When new data are inputted, the system learns how to act according to new situation [16]. Deep neural networks use layers of nodes to create high level functions from datasets. Different types of neural networks are distinguished between them from working principles, action of schemas and application areas.

Convolutional Neural Networks (CNN) implements convolution in its structure. It reduces the memory using weight sharing and the number of network parameters avoiding the over-fitting problem. Shared weight as well as space or time down sampling implemented in CNN, provide a certain amount of translation, scaling, and distortion invariances [17][18].

Deep Neural Networks are widely used and with great impact in the computer vision field. They solve different problems like image and face recognition, object detection, image classification etc.

A. ResNet-50 and ResNet-101

The template is designed so that author affiliations are not repeated each time for multiple authors of the same affiliation. Please keep your affiliations as succinct as possible (for example, do not differentiate among departments of the same organization). This template was designed for two affiliations.

ResNet50 is a Residual Neural Network of type Convolutional Neural Networks (CNN) with 50 layers. It has very good results in image classification. Number of layers is very important in Deep Learning. Additional layers can improve solution of complex problems or having better results. But it also increases the risk of saturation of accuracy levels which may slowly degrade after some point. The performance of the system may be decreased in training as well as in testing [19].

To increase the accuracy, ResNet uses residual blocks. It introduces the concepts of “skip connections”, which add outputs from previous layers to the outputs of stacked layers, enabling the model to learn the identity function. Residual blocks make possible facilitation of in learning the identity function, minimizing so the percentage error [20].

The first ResNet architecture was ResNet-34, using shortcut connections to turn a plain network into a residual one. Comparing with VGG neural networks, ResNets have fewer filters and less complexity. Layers have the same number of features for the same output feature map size. To preserve time complexity per layer, number of filters is doubled when the feature map size is halved. The VGG plain network is improved with “skip connections” or “shortcuts”. Identity shortcuts are used directly if input and output dimension are the same. Otherwise, extra zeros are padded to increase the dimensions of shortcuts or the projection shortcut is used to match the dimensions [20][21]. ResNet-50 is based on the ResNet-34 model, but the building blocks are modified in a bottleneck design using a stack of three layers. Each block of two layers in ResNet-34 is converted in a three layers bottleneck block yielding ResNet-50 architecture. It has much higher accuracy than Resnet-34 [22]. ResNet-101 or ResNet-152 residual networks use more three layers blocks than ResNet-50.

B. EfficientNet-B2

Scaling up the ConvNets is one of visible ways to achieve better accuracy. For example, ResNets scale from Resnet-50 to ResNet-152 by using more layers. It is possible to scale the models by scaling one or more of dimensions like width, depth, and resolution. EfficientNet models carefully balance network width, depth, and resolution to get better performance. Using new type of scaling method called compound coefficient, EfficientNet models uniformly scale across all the dimensions of width, depth, and resolution [23]. This provides added benefit over ResNet models in terms of accuracy and size of model. These models start with a baseline network designed using neural architecture search and use this new scaling method to obtain the family of models called EfficientNets. Unlike traditional methods which scale network width, depth, and resolution arbitrarily, EfficientNet family of models is formed by scaling these factors uniformly with a set of fixed scaling coefficients. Out of several versions available, we choose to use EfficientNet-B2 keeping in mind the memory and space constraints. This model is comparable in performance but more efficient than ResNet-50.

C. DenseNet

It is shown in practice that residual connections especially short connections between layers help convolutional networks be deeper, more accurate and efficient to train. DenseNets introduces connection between each layer to every other layer of the network in a feed-forward fashion. ConvNets with L layers have L connections one between each layer and its subsequent layer whereas DenseNets have $(L(L+1) / 2)$ direct connections. All previous layers' feature-maps are used as inputs into each layer, and its own feature-maps are used as inputs into all subsequent layers. DenseNets alleviate the vanishing-gradient problem, improve feature propagation, and

promote feature reuse. We use DenseNet121, the base model of DenseNet family keeping in mind space, computation and memory constraints [24].

D. K-Nearest Neighbors Classifier Algorithm

K Nearest Neighbors Algorithm is a supervised machine learning algorithm used to classify data. It a simple algorithm, easy to implement, its time execution increases significantly when data size increases. It finds the K smallest distances between the current data and the dataset, the output is the most frequent label in those K data. Usually, KNN algorithm is used for recommender systems, when the dataset size is not significantly high [25].

V. NEW DATASET

The images are extracted from the Recipe 1M dataset. There is a portion of the dataset that contains nutritional information. Also, it is populated with some data gotten from the web for the recipes without information. The Nutri-Score value is calculated for these recipes.

Each recipe is linked to several images. a new dataset of 50718 images with the respective nutri-score values is created.

In Fig. 1, the distribution of classes presented. It is seen that there is a little imbalance between classes, with class E being the minority class. The images are converted to 256x256 dimensions, normalized and augmented them. Next the dataset is split in a training and validation set with 90%-10% proportion in a stratified way to keep the ratios.

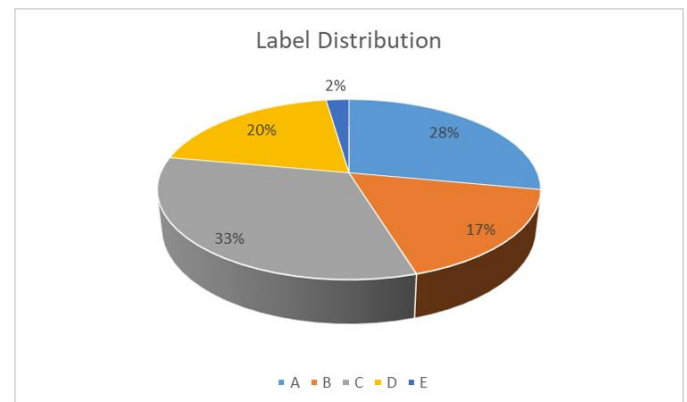


Fig. 1. Class distribution of the dataset.

VI. METHODOLOGY

Four deep learning neural networks, Resnet50, Resnet101, EfficientNetB2 and Densenet121, are considered for the healthy recipes system, which are pretrained using the ImageNet dataset.

The new dataset with 50718 labeled images from the Nutri-Score system is used to tune these pre-trained models. The deep learning pipeline is shown in Fig. 2.

The systems are evaluated and compared in terms of accuracy and loss. Cross-entropy loss and SGD optimizer are used for Resnet 50 and Resnet101 and Adam optimizer is used for EfficientNetB2 and DenseNet121. Servers equipped with GPU, are used for the training. Implementations are done using

Pytorch framework. Tensorboard was used for recording the models' loss and accuracies as well as for visualization.

Embeddings from the last convolutional layer are inputted to the KNN classifier to get the nearest neighbors with the lowest Nutri-Score value.

In our case, we need to return the nearest neighbors for the inputted image and it is returned based on the distance between the embeddings. KNN will evaluate the nearest neighbors based on the Minkowski distance, as it is a general form of Euclidean distance, as shown in equation 2:

$$d(x, y) = \left(\sum_{i=1}^n |x_i - y_i|^p \right)^{\frac{1}{p}} \quad (2)$$

KNN is a classifier that predicts the class of an element as the majority class among its K nearest neighbors.

The healthiest nearest neighbors for the input image are returned to the user together with the corresponding recipes.

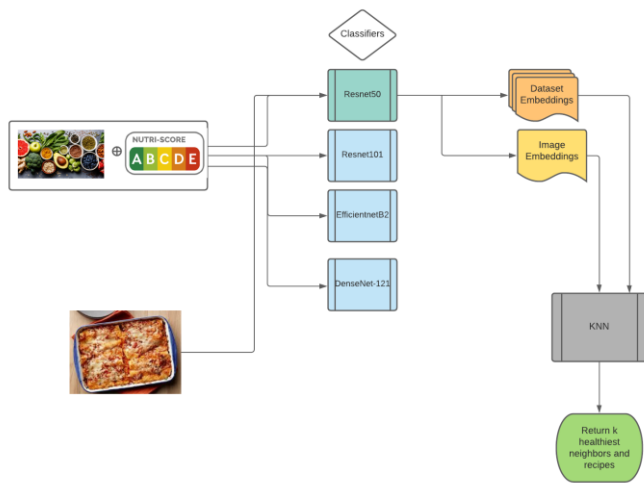


Fig. 2. Deep learning pipeline.

VII. EXPERIMENTAL RESULTS

A. Results from Deep Learning Neural Network Models

The new dataset was created and labeled. The training and test dataset loaders were created and were ready for the modelling part.

At first the pre-trained models on ImageNet are downloaded and the convolutional

Layers are frozen. An initial train is performed getting the best accuracy around 70%. Then the convolutional layers are unfrozen and full training of the networks is performed.

The learning rate scheduler decays the learning rate by a factor of 0.1 every 7 epochs. The SGD optimizer is used for the Resnet50 and Resnet101 models and Adam optimizer is used for Densenet121 and EfficientNetB2. Models are trained for 100 epochs and the best model is selected. Fig. 3 shows the comparison of classification models.

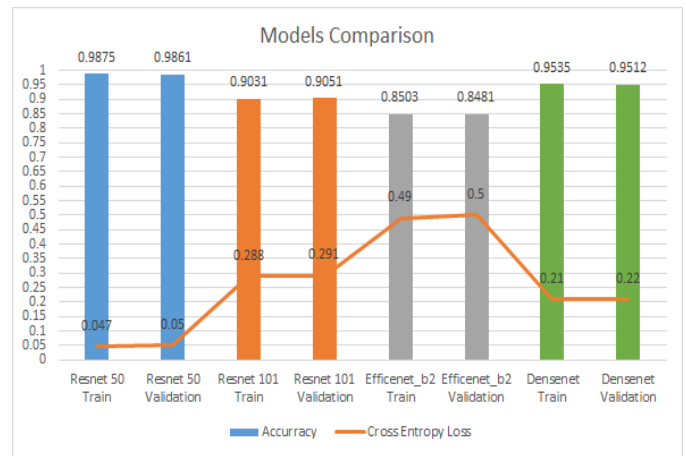


Fig. 3. Classification models comparison.

From the chart, it is shown that Resnet50 and Densenet121 are the best models giving respectively 98.61% and 95.12% of accuracy. Resnet 101 produced 90.51% of accuracy and performed poorer than Resnet50. This fact is explained that Resnet101 is a bigger model and our new dataset is relatively small. EfficientNetB2 is the worst performer with only 84.81%. Attempts are done for executions of EfficientNetB3 and EfficientNetB4 but we did not have the computational resources available to train these models so executions are limited ourselves to EfficientNetB2. To analyze the accuracy for every class, the confusion matrix for the two best models is printed, as shown in Fig. 4 and 5.

The results show that the model performs very well for minority classes also. Augmentation helped to deal with the imbalance. After defining best model, the next step is to get the embeddings and train a KNN model based on them.

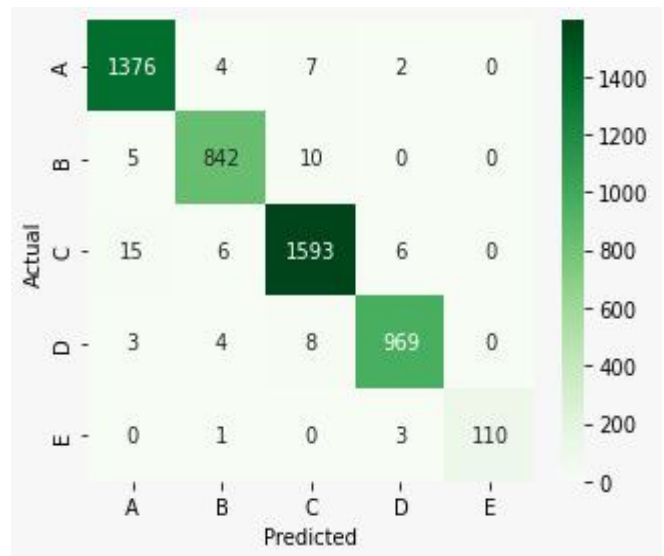


Fig. 4. Confusion matrix for resnet50.

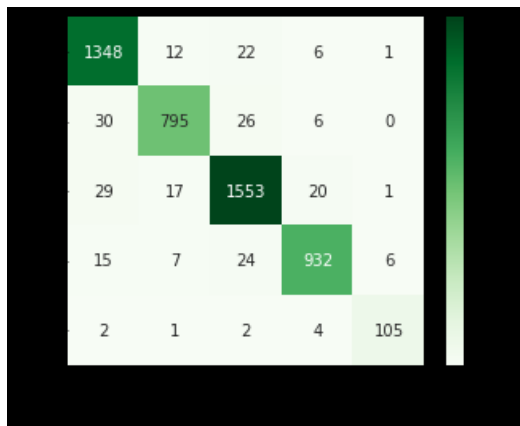


Fig. 5. Confusion matrix for densenet121.

B. Results from KNN Model

Since Resnet50 performed better than other models for the new labeled dataset, embeddings from the last convolutional layer of the Resnet 50 model are extracted.

The images based on the Euclidean distance between the embeddings in Tensorboard are plotted. The results are shown in Fig. 5.

As it can be seen from the plot images closer to the D label are also labelled with D, but some C-values in the surroundings exists. These images are the interested ones images they are similar but healthier. The number of neighbors to the KNN classifier is set to 24. The proposed system returns 24 similar food images according to the inputted one. This number is tunable and can be changed by the user.

Results of the experiment are shown in Fig. 6.



Fig. 6. Sample image input and its 24 outputs images recommended by the system.

As it is seen from the results, the images are similar to each other, and the class labels are very similar. Our main aim is to return the images and recipes with the lowest nutri-score value. The user might input the number of healthy recipes that he wants to get and the system will generate them. It is seen from the sample that the output is C-labelled has two B-labeled neighbors that are similar but healthier. These two images and the corresponding recipes will be returned to the user if he puts the recipe limit to two. According to the same logic, the first three neighbors will be returned for sample image and if the user inputs the recipe limit to three. As an example, we are generating the recipe for the second neighbor of the sample 2 image. The obtained recipe is shown in Fig. 7.

Recipe : Slow Cooker Fruit, Nuts, and Spice Oatmeal :

- Ingredient 1: (quantity:2 cup: 'oats')
- Ingredient 2: (quantity:2 cup: 'cranberries, dried, sweetened'),
- Ingredient 4: (quantity 2 tablespoons: 'salt, table'),
- Ingredient 5: (1/2 cup: 'nuts, almonds'),
- Ingredient 6: (1/2 cup: 'nuts, pecans'),
- Ingredient 7: (3 cup : 'water, bottled, generic'),
- Ingredient 8: (1 cup: 'milk, fluid, 1% fat, without added vitamin a and vitamin d'),
- Ingredient 9: (1 tablespoon: 'spices, cinnamon, ground'),
- Ingredient 10: (1 teaspoon: 'pumpkin, raw'),
- Ingredient 11: (2 teaspoon: 'butter, without salt'],

Instructions : 'Combine the oats, apple, cranberries, almonds, pecans, water, milk, cinnamon, pumpkin pie spice, and butter in a slow cooker. Cook on Low overnight or 8 hours.'

Fig. 7. Recipe for the second neighbor of sample 2 image , A-labeled.

It will be an A-class meal which is very healthy and it is similar to what is actually requested by sample 2 image. The proposed model is able to return the best and healthiest recipes for the input image. The user may decide the number of recipes that he wants to return.

User has also the possibility to set the maximum Nutri-score threshold that he can accept and the system will limit the result accordingly.

VIII. DISCUSSION OF RESULTS

In this paper, a deep learning-based approach for food image retrieval and recipe recommendation is presented. Results show that ResNet50 and DenseNet121 achieved the highest accuracies among the models evaluated. This is consistent with previous studies that have reported the superior performance of these models in image classification tasks [26][27]. Results also highlight the importance of transfer learning, as pre-trained models on ImageNet were used as a starting point for our models, which significantly improved their performance.

One important aspect of the proposed approach is the use of data augmentation to deal with the imbalanced dataset. This

allowed us to achieve high accuracies not only for the majority classes, but also for the minority classes. Data augmentation has been shown to be an effective technique to improve the performance of deep learning models, especially when dealing with limited datasets [28].

The use of KNN for food image retrieval and recipe recommendation is not a novel idea. However, the approach differs from previous studies in that we use the embeddings extracted from the last convolutional layer of the ResNet50 model as the feature vectors for KNN. This approach has been shown to be effective in several image retrieval tasks [29][30][31] and the obtained results confirm its effectiveness for food image retrieval as well.

One limitation of our study is the relatively small size of the dataset, which may affect the generalizability of our results. Moreover, our dataset contains only Western-style dishes, which may limit the applicability of our approach to other cuisines. Future studies should focus on collecting larger and more diverse datasets to improve the performance and generalizability of our approach.

The proposed approach provides a promising solution for food image retrieval and recipe recommendation. The high accuracies achieved by our models, especially for the minority classes, demonstrate the effectiveness of the approach. The use of KNN with embeddings extracted from deep learning models provides a fast and efficient way to retrieve similar food images and recommend healthy recipes. It has the potential to be useful for various applications, such as food logging, dietary analysis, and personalized meal planning.

IX. CONCLUSIONS AND FUTURE WORK

In this paper, an end-to-end deep learning neural network model is developed that recommends healthy meals to the users based on the imputed food images. A new dataset with 50718 records is created by extracting food images from Recipe 1M dataset and labeling them with the Nutri-Score value.

This dataset can be expanded even more in the future and can be used for healthy food classification tasks. Deep learning models ResNet50, ResNet101, EfficientNetB2 and DensNet121 are compared on the new dataset. Experimental results show that Resnet50 gives the maximum accuracy of 98.61 followed by Densnet121 with 95.12%. Embeddings from the last convolutional layer are inputted to the KNN model providing the user the healthiest recipes of food similar to the input image but healthier in nutritional values.

The results demonstrate the feasibility of using deep learning neural networks for recommending healthy meals based on food images. The new dataset created in this paper provides a valuable resource for future research in healthy food classification. The comparison of various deep learning models on this dataset highlights the importance of selecting an appropriate model architecture for the task at hand.

However, there are still some limitations that need to be addressed in future work.

Firstly, the dataset used in this study is relatively small and limited to Nutri-Score values. Expanding the dataset with

additional labels and nutritional information could further improve the performance of the model. Secondly, the proposed model is limited to recommending healthy meals based solely on food images, which may not be sufficient for some users. Future work could consider incorporating additional user information, such as dietary preferences and restrictions, to further personalize the recommendations.

REFERENCES

- [1] Hercberg, S., Touvier, M., Salas-Salvado, J., & on behalf of the Group of European scientists supporting the. (2021). The Nutri-Score nutrition label. *International Journal For Vitamin And Nutrition Research*. doi: 10.1024/0300-9831/a000722
- [2] Marin, J., Biswas, A., Ofli, F., Hynes, N., Salvador, A., & Aytar, Y. et al. (2021). Recipe1M+: A Dataset for Learning Cross-Modal Embeddings for Cooking Recipes and Food Images. *IEEE Transactions On Pattern Analysis And Machine Intelligence*, 43(1), 187-203. doi: 10.1109/tpami.2019.2927476
- [3] Moolya, D. et al. (2022) "Recipe generator using Deep Learning," *International Journal for Research in Applied Science and Engineering Technology*, 10(5), pp. 846-851. Available at: <https://doi.org/10.22214/ijraset.2022.42321>.
- [4] P. Yang, C. Dong, X. Zhao and X. Chen, "The Surface Damage Identifications of Wind Turbine Blades Based on ResNet50 Algorithm," 2020 39th Chinese Control Conference (CCC), Shenyang, China, 2020, pp. 6340-6344, doi: 10.23919/CCC50068.2020.9189408.
- [5] He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep Residual Learning for Image Recognition. 2016 IEEE Conference on Computer Vision and
- [6] S. Ethiraj and B. K. Bolla, "Classification of Astronomical Bodies by Efficient Layer Fine-Tuning of Deep Neural Networks," 2021 5th Conference on Information and Communication Technology (CICT), Kurnool, India, 2021, pp. 1-6, doi: 10.1109/CICT53865.2020.9672430.
- [7] Ezzat, D., Hassaniien, A.E. and Ella, H.A. (2021) "An optimized deep learning architecture for the diagnosis of COVID-19 disease based on gravitational search optimization," *Applied Soft Computing*, 98, p. 106742. Available at: <https://doi.org/10.1016/j.asoc.2020.106742>.
- [8] L. Bossard, M. Guillaumin, and L. Van Gool, "Food-101- mining discriminative components with random forests," in *European conference on computer vision*. Springer, 2014, pp. 446-461. 1
- [9] A. Meyers, N. Johnston, V. Rathod, A. Korattikara, A. Gorban, N. Silberman, S. Guadarrama, G. Papandreou, J. Huang, and K. P. Murphy, "Im2calories: towards an automated mobile vision food diary," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 1233-1241. 1
- [10] A. Salvador, M. Drozdal, X. Giro-i Nieto, and A. Romero, "Inverse cooking: Recipe generation from food images," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 10 453-10 462.
- [11] D. Raboy-McGowan and S. Lu, "Recipenet: Image to recipe/nutritional information generator," 2020.
- [12] Raboy-McGowan, D., & Lu, S. (2020). RecipeNet: Image to Recipe/Nutritional Information Generator.
- [13] Vermote, M. et al. (2020) "Nutritional content, labelling and marketing of breakfast cereals on the Belgian market and their reformulation in anticipation of the implementation of the nutri-score front-of-pack labelling system," *Nutrients*, 12(4), p. 884. Available at: <https://doi.org/10.3390/nu12040884>.
- [14] Jürkenbeck K, Mehlhose C, Zühlsdorf A (2022) The influence of the Nutri-Score on the perceived healthiness of foods labelled with a nutrition claim of sugar. *PLoS ONE* 17(8): e0272220. <https://doi.org/10.1371/journal.pone.0272220>
- [15] PirahanSiah, F. (2019) "Computer Vision, Deep Learning, Deep Reinforcement Learning." Available at: <https://doi.org/10.14293/s2199-1006.1.sor-uncat.clzwyuz.v1>.
- [16] Takagi, T., & Mizumoto, I. (2018). Parameter Optimization with Input/Output Data via DE for Adaptive Control System with Neural

- Network. *Journal Of Robotics, Networking And Artificial Life*, 5(1), 19. doi: 10.2991/jrnal.2018.5.1.5
- [17] Alzubaidi, L., Zhang, J., Humaidi, A., Al-Dujaili, A., Duan, Y., & Al-Shamma, O. et al. (2021). Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *Journal Of Big Data*, 8(1). doi: 10.1186/s40537-021-00444-8
- [18] Sirmorya, A., Chaudhari, M., & Balasinor, S. (2022). Review of Deep Learning: Architectures, Applications and Challenges. *International Journal Of Computer Applications*, 184(18), 1-13. doi: 10.5120/ijca2022922164
- [19] Alaeddine, H., & Jihene, M. (2021). Deep Residual Network in Network. *Computational Intelligence And Neuroscience*, 2021, 1-9. doi: 10.1155/2021/6659083
- [20] Patil, O.S. et al. (2022) "Deep residual neural network (resnet)-based Adaptive Control: A Lyapunov-based approach," 2022 IEEE 61st Conference on Decision and Control (CDC) [Preprint]. Available at: <https://doi.org/10.1109/cdc51059.2022.9992881>.
- [21] Image Recognition using Deep Learning Techniques. (2021). *Journal Of Xidian University*, 15(6). doi: 10.37896/jxu15.6/005
- [22] Zhang, L., & Schaeffer, H. (2019). Forward Stability of ResNet and Its Variants. *Journal Of Mathematical Imaging And Vision*, 62(3), 328-351. doi: 10.1007/s10851-019-00922-y
- [23] Tan, M., & Le, Q.V. (2019). EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. ArXiv, abs/1905.11946.
- [24] Li, G., Zhang, M., Li, J., Lv, F., & Tong, G. (2021). Efficient densely connected convolutional neural networks. *Pattern Recognition*, 109, 107610. doi: 10.1016/j.patcog.2020.107610
- [25] Zhang, X., Wang, S., Wu, Z., & Tan, X. (2022). Unsupervised image clustering algorithm based on contrastive learning and K-nearest neighbors. *International Journal Of Machine Learning And Cybernetics*. doi: 10.1007/s13042-022-01533-7
- [26] A. Salvador, N. Hynes, Y. Aytar, J. Marin, F. Ofli, I. Weber, and A. Torralba, "Learning cross-modal embeddings for cooking recipes and food images," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp.3020–3028. 1
- [27] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).
- [28] Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4700-4708).
- [29] Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data*, 6(1), 60.
- [30] Babenko, A., Slesarev, A., Chigorin, A., & Lempitsky, V. (2015). Neural codes for image retrieval. In *Proceedings of the European conference on computer vision* (pp. 584-599).
- [31] Radenović, F., Tolias, G., & Chum, O. (2016). CNN image retrieval learns from BoW: Unsupervised fine-tuning with hard examples. In *Proceedings of the European conference on computer vision* (pp. 3-20).

An Automatic Framework for Number Plate Detection using OCR and Deep Learning Approach

Yash Shambharkar¹, Shailaja Salagrama², Kanhaiya Sharma³, Om Mishra⁴, Deepak Parashar⁵
Symbiosis Institute of Technology Pune, Symbiosis International (Deemed) University, Pune, India^{1,3,4,5}
Computer Information System, University of the Cumberland's, Williamsburg, Kentucky, USA²

Abstract—The use of automatic number plate detection devices in safety, commercial, and security has increased over the past few years. Number plate detection using computer vision is used to provide fast and accurate detection and recognition. Lately, many computerized approaches have been developed for the identification of vehicle registration details based on license plate numbers using either Deep Learning (DL) methodologies. In the proposed framework, we used Optical Character Recognition (OCR) and a deep learning-based new approach for automatic number plate detection and recognition. A deep learning approach trains the model to recognize the vehicle. The vehicle registration plate area is cropped adequately from the image, and a Convolution Neural Network (CNN) uses OCR to identify numbers and letters. The Jetson TX2 NVIDIA target served as the model's training data source, and its performance has been tested on a public dataset from Kaggle database. We obtained the highest accuracy of 96.23%. The proposed system could recognize vehicle license plate numbers on real-world images. The system can be implemented at security checkpoint entrances in highly restricted areas such as military areas or areas surrounding high-level government agencies.

Keywords—Number plate detection; recognition; deep learning; OCR; image classification

I. INTRODUCTION

Automatic number plate Detection is also known as license plate detection or vehicle registration plate detection. It uses image processing technology to extract the registration plate from digital images or video [1]. Afterward, the information stored can be used to find various new models, some of which could be transaction gateways or traffic violation systems. Taking the context of a real-life problem into consideration, in practical applications, researchers have to deal with a variety of challenges, such as registration plate type, textual font, registration plate color and font, registration plate location, and environmental conditions such as lighting and weather. License plates become challenging to recognize. License plate formats vary from country to country: Different colors, languages, and fonts. Some plates have a different colored border than the background surrounding the plate, and some have a plain background, which indirectly adds challenges to capturing and recognizing car plates [2]. Variations in environmental conditions, such as lighting and image background, also affect the license plate recognition rate. Several studies have been proposed earlier.

YOLO is a unified model for object recognition [1]. This model can be built and trained directly on full frames. In contrast to classifier-based approaches, YOLO is introduced

with a loss function that directly corresponds to recognition performance, and the entire model is prepared together. Fast YOLO is the fastest general-purpose object detector in the literature, and YOLO advances state of art in real-time object detection. YOLO also generalizes to new domains, making it ideal for applications that rely on fast and robust object detection. Authors discuss the detection of a registration plate for a person who does not wear a helmet in [3]. The two-wheeler operator must have a helmet for his safety. In this study, they proposed a real-time detection of number plates using YOLO, for which they deployed the CNN layer. The accuracy of the model was 95.5%.

In [4], they DL method was proposed for vehicle detection using R-CNN and faster R-CNN, in which they tested and trained their model. The model has three steps: data acquisition, CNN design, and R-CNN. The mean average precision (map) values from faster R-CNN and R-CNN were approximately 0.73, 0.76, and 0.64, and 0.65, respectively. Another one had an overall accuracy of 94.98%. [5], proposed a deep learning model using an Image AI library for training. The model consists of four steps: video acquisition, detection of a vehicle, registration plate detection, and then registration plate character recognition. The model's accuracy was 96% for plate localization and 90% for character recognition was 90%. In [6], firstly, they made the image acquisition, after which they used pre-processing techniques like RGB to grayscale conversion, noise removal, and image banalization. For registration plate extraction, used Sobel's edge detection [7]. After that, the characters are segmented using the CNN layer for character recognition. The model had an overall accuracy of 95%. In [8] they proposed a highly accurate trained model for both registration plate detection and recognition [9]. The whole model was prepared by sharing a convolution layer with the above features. The proposed system had an accuracy of 94% for a jointly trained model. In [10], a framework for character detection using a differentiable binary process is presented. In [11], a framework requires text extraction from an input digital image in which the image has been processed and filtered thoroughly. They used the CV2 OpenCV library in Python for pre-processing and Tesseract for character recognition. Real-time plate detection was performed on a Bangla dataset [12]. For character recognition, the proposed model employed a CNN layer. The dataset used had a sample size of 5500, and the accuracy of successful recognition was 90.90% in a real-time study. In [13-14], YOLO is used for object detection. They trained and fine-tuned the CNN layer for each stage, then did the segmentation part. The SSIG dataset, which contained 2000 frames from 101 videos, was

used. The system had an accuracy of 93.53% with 47 frames per second. In [15], [18] a kernel that was partially restricted was focused on with the help of direct blind deconvolution and denoising with CNN on a real-time blur image that was acquired from a traffic surveillance system used a neural network that was trained on artificial data. They used deep learning mode [19]-[28], which includes image acquisition, plate detection, segmentation, and character recognition. The data set was the Turkish vehicle license plate dataset, which consists of real-world images obtained from a security surveillance camera. Unlike other countries, India, with a population of 1 billion, has a unique need for ANPRs. ANPR's primary uses are highway monitoring, parking lot management, and neighborhood law enforcement safety. Existing literature says that in India, within every four minutes, one person dies, and surprisingly, most of them are because of over-speeding. ANPR monitors the average speed of vehicles, and vehicle owner information can identify vehicles violating the traffic rules. In this case, e-challans can be automatically generated for the penalized plate owner. This helps maintain law and order, which in turn minimizes traffic fatalities. ANPR provides the best solution for giving parking management. Vehicles with registered license plates are automatically allowed into the parking area, but unregistered vehicles will be charged at check-in and check-out. According to the study, approximately 200,000 cars are stolen annually in India. This number can be reduced by taking appropriate measures and using this model to track vehicles. This allows law enforcement agencies to determine when, where, and through which route the stolen vehicle is stolen, if vehicle is stolen. This allows us to bring justice to such a powerhouse quickly. The main objectives of the proposed system are as follows. (1) To understand ANPR and apply it to a license plate recognition system. (2) To improve the performance of the number plate detection system using OCR and DL approach. (3) The proposed system efficiently used for detection and recognition of vehicle number plate.

The study is organized as follows: Section I provide the literature review on problem formulation and various existing methodologies; Section II describes the proposed procedures. Following that, in Section III, the results are analyzed and discussed. Section IV explains the conclusion and future scope.

II. PROPOSED METHOD

Fig. 1 depicts the proposed framework of number plate detection in four essential parts: a dataset, license plate detection, segmentation, and OCR. Modern rural, urban, and national highway networks have proliferated in recent decades. This created the need to monitor and manage traffic on the road efficiently. The primary goal of this study is to train a model to recognize and identify license plates from their images correctly. Different countries have different characteristics of license plates, such as their number system, colors, sign language, style (font), and size of a license plate, so further research is needed.

Deep learning (DL) is a branch of Machine Learning (ML), and ML is a subset of Artificial Intelligence (AI)

that mimics the way humans acquire certain kinds of knowledge. Deep learning is a critical component of data science, including statistics and predictive models. It benefits data scientists who must collect, analyze, and interpret large amounts of data. Deep Learning makes this process faster and easier. At its simplest, DL can be thought of as a way to automate predictive analytics. While traditional ML algorithms are linear, DL algorithms are built up in increasing complexity and abstraction layers. A shallow CNN quickly scans the test image and eliminates most sample windows. In the proposed approach, authors have used four layers to predict number plates. In the first layer, the number plate detection work is done; in the second and third layers, region extraction is calculated; and in the last layer, OCR is used. The automatic framework for number plate detection using OCR and deep learning architectures is shown in Fig. 2.

A. Dataset

The desired unit where deep learning problems fail is a need for a public dataset. As we know that, the wrong data sets can produce inaccurate or incorrect results. So, we need to ensure that there are no similarities between the content of a dataset other than its essential characteristics. Because India is a large country, hence need to consider all the contexts and ensure that the given dataset is well-balanced without favoring or disfavoring any entity, such as vehicle type or license plate layout. Firstly, the vehicle types present on the streets of a typical Indian city are cars, two-wheelers, buses, trucks, and pickups. So, every type of vehicle has its own format for displaying the registration plate. For example, most cars have one line of plates, but some people prefer a layout with two lines of plates. Registration plates belonging to Indians can be single-row or double-row and most two- and three-wheeler registration plates are two lines, but their layout differs. The font on both sides is also different, with some 3-wheeler auto rickshaw uses paint rather than printing the number plate. So here we found the basic first step in license plate recognition is to collect a database. In the proposed work dataset is downloaded from the Kaggle website, which includes license plates with significant variations in registration numbers collected by kaggle.com. Total of 4326 were trained on our model to identify the number plate and had used deep learning strategies as you can see in the Table I. It represents us the type of vehicle that were been trained as you can see the dataset is not biased as all type of vehicle are been taken into account.

B. Preprocessing

When images are acquired at this stage and passed to the algorithm for further testing and prediction, Image pre-processing is one of the crucial stages for any computer vision system. So, the main objective of this image pre-processing is to recognize the image and collect information regarding the image, which can be used in further processing [7]. Image pre-processing changes the operation performed on a non-figurative object. The primary goal is to improve image information to extract the necessary information from that image. Fig. 3 shows sample images from the custom dataset for experimental work.

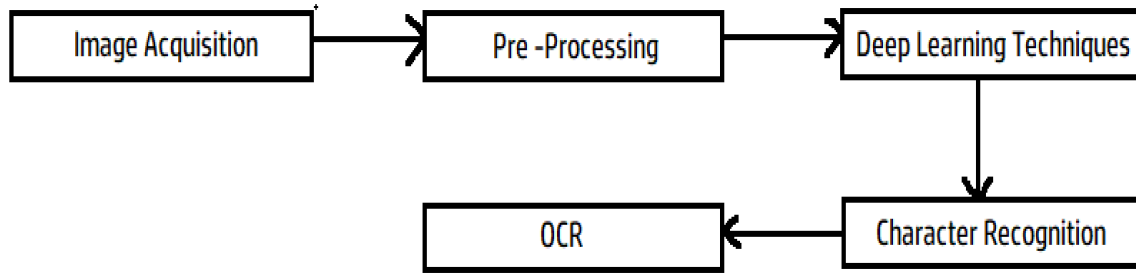


Fig. 1. The block diagram of proposed methodology.

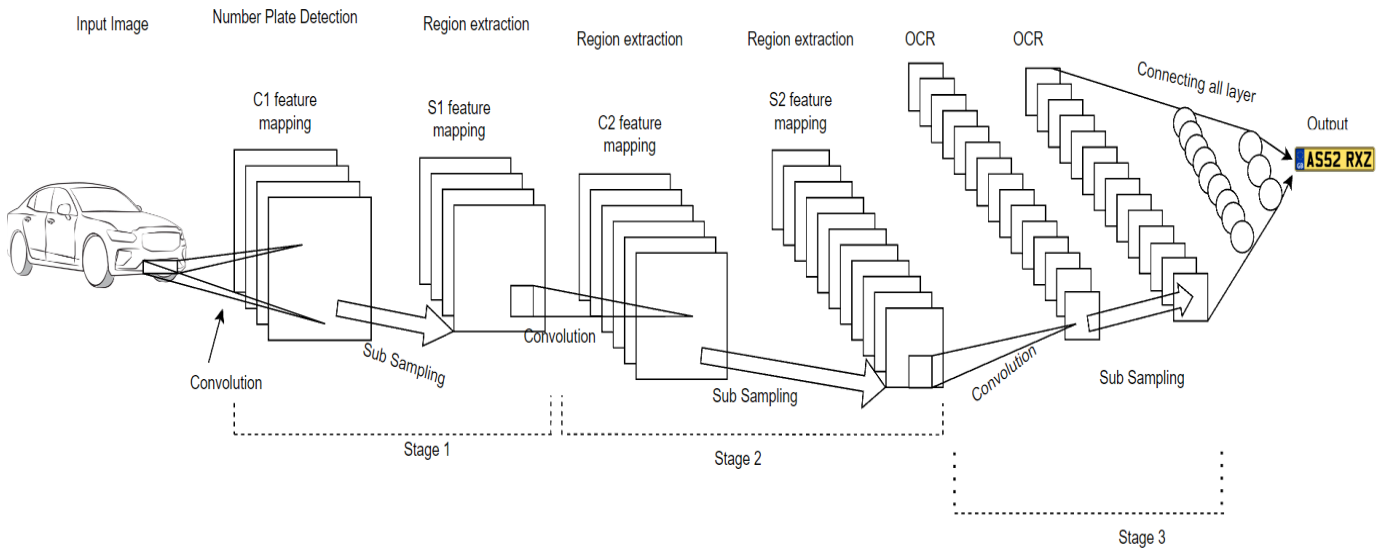


Fig. 2. An automatic framework for number plate detection using OCR and deep learning architecture.

C. Number Plate Detection

Most countries around the world have registration plates that are shaped like a rectangle. Moreover, the registration plate comes in different shapes and layouts in India. Sometimes it is seen that a two-wheeler uses a trapezoidal plate. In terms of size, the average size of the plate varies by vehicle, with the average size for cars being 500×120 mm and for motorcycles being 200×100 mm [11]. Three-wheeled vehicles such as an auto rickshaw have tiny plate painted mainly by a painter. Taking all the points into context, we put all the registration plates into a rectangular box. As shown in Fig. 2, the image of the car with the highlighted number plate is in the green-colored rectangular box. Fig. 4 shows the detection of registration plat.

The result uses a thresh holding algorithm on the number plate, and because of this thresh holding method, the plates are being segmented at a higher level. These thresholds help us recognize the character correctly.

D. Charactor Recognition

As discussed earlier, the main issue with character recognition is that there are many types and categories of the font used in the plate. So, the best way to train our desired model is to check for letters and digits to cover different typefaces, including those used in painted panels. Here, we used OCR for character recognition. The recognized and segmented character can be trained with our custom-built dataset for training and testing purposes.

A corresponding file was created for training and testing with a valid dataset of 4326 images. So, after all the model training, we figured out that our model is about 95% accurate at the detection range of the license plate and fails only when the license plate is a little smaller in size as compared to the standard plate size or when similar plates are present in the tested image.



Fig. 3. Sample images from the custom dataset for experimental work.

Now let's check how the OCR works. The attributes are extracted from the sample image. These features can be preferred as follows: (1) Distance to the wall (DTW): The calculation for such is based on active pixels. They are the highest intensity value possible, depending on the image format. For example, in an 8-bit grayscale format, the maximum level is 255; to calculate that, it is divided into 12 regions (4 rows x 3 columns). And the activated pixel is calculated concerning the horizontal and vertical areas within each region. (2) Cross-Time Feature (CTF): This feature includes the six widest white-to-black and black-to-white transitions across the character horizontally and vertically. The six widest shifts are computed for every row and column of character candidates. These intersections are sorted, and the first three rows and columns with the highest and lowest changes are saved along with their positions. (3) Active Region Ratio (ARR): The region is divided into four rows and three columns, respectively, and the area of active pixels is chosen as a critical feature for the block. (4) Height to Width Ratio (HWR): When talking about HWR, alphanumeric symbols for

any language have been bifurcated into three types, namely, those with more extensive and equivalent height and width. Table I depicted describes the count of vehicle.

The positive score represents the character, and the negative score represents the non-character symbol within the plate. Optical character recognition has been around for the last 80 years. However, large technology companies initially were the primary developers of products that recognized optical [18] characters. Advances in machine learning and deep learning have enabled individual researchers to develop algorithms and techniques that can more accurately identify handwritten manuscripts. We also see more and more researchers using - convolutional neural networks (CNNs) to recognize handwritten and machine-printed characters. This is since CNN-based architectures are well suited for recognition tasks where the input is an image. CNN's were initially used for object detection tasks in images. ImageNet Large Scale Visual Recognition Challenge (ILSVRC) is commonly used CNN-based architectures for visual recognition tasks.

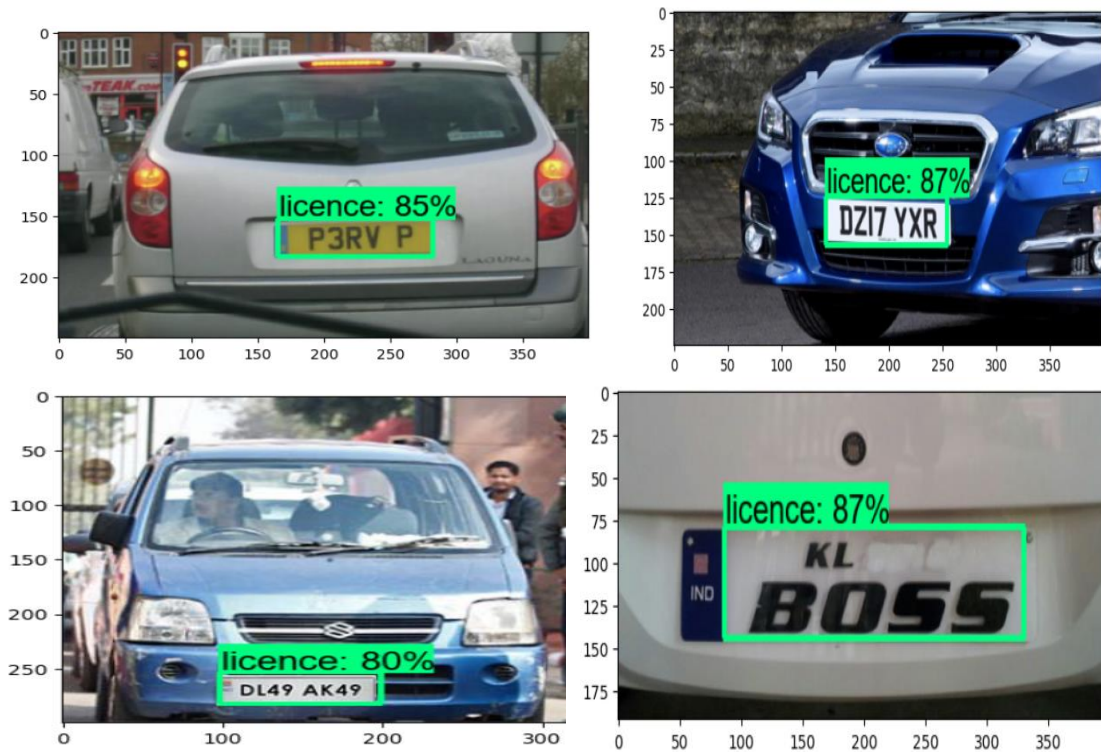


Fig. 4. Detection of registration plate in rectangular boxes.

TABLE I. DESCRIBES THE COUNT OF VEHICLE

Type of vehicle	Type of license plate	Font of license plate
Two Wheeler (300)	Printed (375)	Single line (273)
		Multiline (102)
Three Wheeler (637)	Painted (125)	Single line (89)
		Multiline (36)
Car (2963)	Printed (2957)	Single line (61)
		Multiline (96)
Truck & pickup (426)	Painted (480)	Single line (86)
		Multiline (394)
	Printed (6)	Single line (2873)
		Multiline (84)
	Painted (82)	Single line (2)
		Multiline (4)
	Painted (344)	Single line (80)
		Multiline (2)
		Multiline (284)

TABLE II. EXISTING METHOD FOR REGISTRATION PLATE DETECTION & RECOGNITION

Actual Registration plate	Predicted Registration Plate	Mismatched Recognition	Accuracy
P3RV P	P3RV P	0	900%
KL BOSS	KL 8OSS	1	84%
DL49 AK49	DL49 AK49	0	89%
DZI7 YXR	DZI7 YXR	0	85%
Mh 14 Gn 9239	Mh il Gn 9239	2	80%

III. RESULT AND DISCUSSION

The proposed system evaluates the car image from a given dataset. The most crucial factor in solving deep learning problems has an accurate dataset that is unbiased, because of which it produces inaccurate or skewed outcomes. It is necessary to ensure no similarities between the items in the dataset other than the characteristic of interest. After that, we came to the training part of the data set. We considered 4326 sample images and trained the model to detect the license plate. The results of OCR based approach are shown in Fig. 5. The captured registration plate is processed for letter segmentation. Separated characters are recognized using OCR (Optical Character Recognition). After which, the essences extracted from the registration plate are stored in an Excel spreadsheet. Table II shows the sample test cases on the real-time dataset.

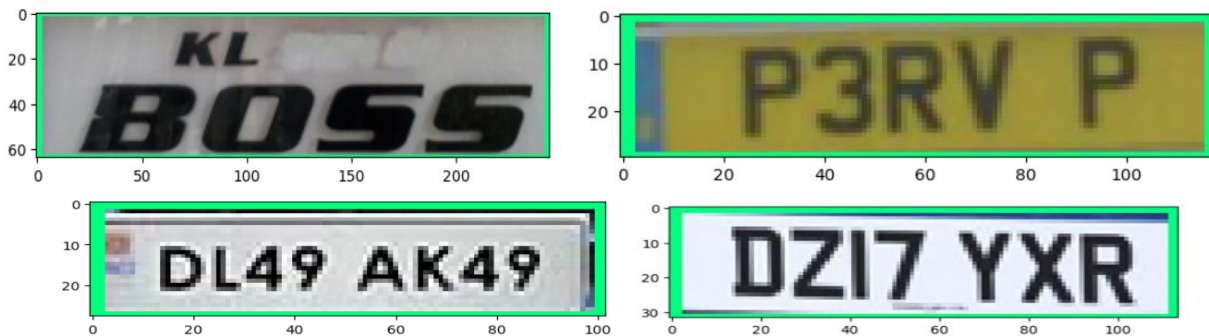


Fig. 5. Results of OCR approach to detect number plate.

TABLE III. COMPARISON WITH EXISTING METHDOS FOR NUMBER PLATE DETECTION AND RECOGNITION

Method/Ref.	Accuracy (%)	
	Detection	Recognition
YOLO, Response surface methodology (RSM) [5]	97 %	90 %
CNN, TensorFlow [16]	96.36%	93%
YOLO, CNN, Edge Detection [6]	94%	91%
RGB Scaling, CNN[9]	97%	94%
YOLO v3	91%	91%
Deep CNN [20]	98.1%	88.1%
Deep CNN, Fast RCNN [18]	99%	91%
RCNN, SVM, ANN [17]	94.4%	96%
RANSAC, SVM [16]	91.4%	91.4
RESNET, RCNN [3]	94.98%	92%
Proposed Method	100% (Vehicle Detection) 95% (License Plate Detection)	96.23%

The Table II clearly shows that a minimum of 80% accurate result can be obtained; in some cases, even 100% accurate prediction is achieved. Table III compares the proposed method’s performance and existing literature. It can be concluding from the Table III that the proposed method performs far better than the existing method.

IV. CONCLUSION

This study shows that in the presence of anomalies, the performance of deep learning techniques is relatively better and beneficial in the process of ANPR as per India’s conditions. Authors represented the entire end-to-end ANPR and noticed multiple lines of LP, non-uniform padding, plate shapes, fonts, and non-uniform font sizes. This presented a complete end-to-end pipeline for ANPR. For LPD, we proposed a model that works well in India and also introduced an alternate method for CR networks suitable for the Indian situation. Our LPD model achieved 96.23% accuracy with a detection threshold of 0.5, which only dropped because it has a 98% accuracy rate for small license plates on cars and buses and a total loss of less than 10% with a learning rate of 92%. The proposed network

showed a high average accuracy of 94.9%, including multiline panels, and predicted 93.7% of the characters with a higher confidence level than 90%. ANPR can identify vehicle owners, vehicle model identification, traffic voltage and control, and vehicle location tracking. This model can be further extended as a multilingual ANPR to automatically identify the language of the characters based on the training data.

REFERENCES

- [1] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016, pp. 779-788, doi: 10.1109/CVPR.2016.91.
- [2] S.S. Nadiminti, P.K. Gaur, and A. Bhardwaj, "Exploration of an End-to-End Automatic Number plate Recognition neural network for Indian datasets", *Computer Science, Environmental Science* 2022,doi: <https://doi.org/10.48550/arXiv.2207.06657>.
- [3] Abdullah Asım Yılmaz, Mehmet Serdar Guzel, İman Askerbeyli, Erkan Bostanci, "A Vehicle Detection Approach using Deep Learning Methodologies", *International Conference on Theoretical and Applied Computer Science and Engineering (ICTACSE)*, Ankara, 2018.
- [4] V Gnanaprakash, N kanthimathi, N Saranya, "Automatic number plate recognition using deep learning", *IOP Conference Series: Materials Science and Engineering, First International Conference on Circuits, Signals, Systems and Securities (ICSSSS 2020)*, December 2020, Tamil Nadu, India. 2021, doi: 10.1088/1757-899X/1084/1/012027.
- [5] S. Saunshi, V. Sahani, J. Patil, A. Yadav, S. Rathi, "License Plate Recognition Using Convolutional Neural Network", *IOSR Journal of Computer Engineering*, vol. 1, pp. 28-33, 2017.
- [6] P. Ravirathnam and A. Patawari, "Automatic License Plate Recognition for Indian Roads Using Faster-RCNN," *2019 11th International Conference on Advanced Computing (ICoAC)*, Chennai, India, 2019, pp. 275-281, doi: 10.1109/ICoAC48765.2019.246853.
- [7] H. Li, P. Wang and C. Shen, "Toward End-to-End Car License Plate Detection and Recognition With Deep Neural Networks," in *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 3, pp. 1126-1136, March 2019, doi: 10.1109/TITS.2018.2847291.
- [8] Z. Selmi, M. Ben Halima and A. M. Alimi, "Deep Learning System for Automatic License Plate Detection and Recognition," *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, Kyoto, Japan, 2017, pp. 1132-1138, doi: 10.1109/ICDAR.2017.187.
- [9] Minghui Liao, Zhaoyi Wan, Cong Yao, Kai Chen, Xiang Bai, "Real-time Scene Text Detection with Differentiable Binarization", *Proceedings of the AAAI Conference on Artificial Intelligence*, 2019, doi: 10.1609/aaai.v34i07.6812.
- [10] R. R. Palekar, S. U. Parab, D. P. Parikh and V. N. Kamble, "Real time license plate detection using openCV and tesseract," *2017 International Conference on Communication and Signal Processing (ICCSP)*, Chennai, India, 2017, pp. 2111-2115, doi: 10.1109/ICCSP.2017.8286778.

- [11] M. Atikuzzaman, M. Asaduzzaman and M. Z. Islam, "Vehicle Number Plate Detection and Categorization Using CNNs," *2019 International Conference on Sustainable Technologies for Industry 4.0 (STI)*, Dhaka, Bangladesh, 2019, pp. 1-5, doi: 10.1109/STI47673.2019.9068049..
- [12] Rayson Laroca, Evair Severo, Luiz A. Zanlorensi, Luiz S. Oliveira, Gabriel Resende Gonçalves, William Robson Schwartz and David Menotti, "A robust real-time automatic license plate recognition based on the yolo detector", *International Joint Conference on Neural Networks (IJCNN)* At: Rio de Janeiro, Brazil, 2018, <http://dx.doi.org/10.1109/IJCNN.2018.8489629>
- [13] Y. Yuan, W. Zou, Y. Zhao, X. Wang, X. Hu and N. Komodakis, "A Robust and Efficient Approach to License Plate Detection," in *IEEE Transactions on Image Processing*, vol. 26, no. 3, pp. 1102-1114, March 2017, doi: 10.1109/TIP.2016.2631901.
- [14] P. Svoboda, M. Hradiš, L. Maršik and P. Zemčík, "CNN for license plate motion deblurring," *2016 IEEE International Conference on Image Processing (ICIP)*, Phoenix, AZ, USA, 2016, pp. 3832-3836, doi: 10.1109/ICIP.2016.7533077.
- [15] R. Panahi and I. Gholampour, "Accurate Detection and Recognition of Dirty Vehicle Plate Numbers for High-Speed Applications," in *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 4, pp. 767-779, April 2017, doi: 10.1109/TITS.2016.2586520.
- [16] Muhammad Aasim Rafique, Witold Pedrycz, Moongu Jeon, "Vehicle license plate detection using region-based convolutional neural networks", *Soft Computing - A Fusion of Foundations, Methodologies and Applications*, vol. 22(19), pp 6429–644, 2018, doi: 10.1007/s00500-017-2696-2.
- [17] Chaowei Cheng, Liye Mei and Junhua Zhang, "License Plate Recognition via Deep Convolutional Neural Network", *2018 IOP Conf. Ser.: Earth Environ. Sci.* 189 062030, 2018, doi: 10.1088/1755-1315/189/6/062030.
- [18] I. Kilic and G. Aydin, "Turkish Vehicle License Plate Recognition Using Deep Learning," *2018 International Conference on Artificial Intelligence and Data Processing (IDAP)*, Malatya, Turkey, 2018, pp. 1-5, doi: 10.1109/IDAP.2018.8620744.
- [19] Li Zou, Meng Zhao, Zhengzhong Gao, Maoyong Cao, Huarong Jia and Mingtao Pei, "License Plate Detection with Shallow and Deep CNNs in Complex Environments", *Complexity*, vol. 2018, pp 1-6, 2018, doi: 10.1155/2018/7984653
- [20] A. Patel C. Patel, and D. Shah, "Automatic Number Plate Recognition System (ANPR): A Survey", *International Journal of Computer Applications*, vol. 69 (9), pp. 0975 – 8887, 2013, doi: 10.5120/11871-7665.
- [21] S. Mercy, I. Muthulakshmi, "Automatic number plate recognition using connected component analysis algorithm" *International*, *Journal For Technological Research In Engineering*, vol. 5 (7), 2018.
- [22] V. Jain et al., "Deep automatic license plate recognition system." *Proceedings of the Tenth Indian Conference on Computer Vision, Graphics and Image Processing*. ACM, pp, 1-6, 2016, doi: 10.1145/3009977.3010052.
- [23] Z. Selmi, M. Ben Halima and A. M. Alimi, "Deep Learning System for Automatic License Plate Detection and Recognition," *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, Kyoto, Japan, 2017, pp. 1132-1138, doi: 10.1109/ICDAR.2017.187.
- [24] D. Yao, W. Zhu, Y. Chen and L. Zhang, "Chinese license plate character recognition based on convolution neural network," *2017 Chinese Automation Congress (CAC)*, Jinan, China, 2017, pp. 1547-1552, doi: 10.1109/CAC.2017.8243013.
- [25] M. Z. Abedin, A. C. Nath, P. Dhar, K. Deb and M. S. Hossain, "License plate recognition system based on contour properties and deep learning model," *2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC)*, Dhaka, Bangladesh, 2017, pp. 590-593, doi: 10.1109/R10-HTC.2017.8289029.
- [26] L. Xie, T. Ahmad, L. Jin, Y. Liu and S. Zhang, "A New CNN-Based Method for Multi-Directional Car License Plate Detection," in *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 2, pp. 507-517, Feb. 2018, doi: 10.1109/TITS.2017.2784093..
- [27] W. Zhang, Y. Mao, Y. Han, "SLPNet: Towards End-to-End Car License Plate Detection and Recognition Using Lightweight CNN" *Pattern Recognition and Computer Vision. PRCV 2020. Lecture Notes in Computer Science*, vol. 12306, 2020, doi:10.1007/978-3-030-60639-8_25.
- [28] Y. Jamtsho, P. Riyamongkol, R. Waranusast, "Real-time Bhutanese license plate localization using YOLO", *ICT Express*, vol. 6, pp. 121-124, 2020, doi: <https://doi.org/10.1016/j.ict.2019.11.001>.

Convolution Neural Networks for Phishing Detection

Arun D. Kulkarni

Computer Science Department
The University of Texas at Tyler
Tyler, TX, 75799, USA

Abstract—Phishing is one of the significant threats in cyber security. Phishing is a form of social engineering that uses e-mails with malicious websites to solicitate personal information. Phishing e-mails are growing in alarming number. In this paper we propose a novel machine learning approach to classify phishing websites using Convolution Neural Networks (CNNs) that use URL based features. CNNs consist of a stack of convolution, pooling layers, and a fully connected layer. CNNs accept images as input and perform feature extraction and classification. Many CNN models are available today. To avoid vanishing gradient problem, recent CNNs use entropy loss function with Rectified Linear Units (ReLU). To use a CNN, we convert feature vectors into images. To evaluate our approach, we use a dataset consists of 1,353 real world URLs that were classified into three categories-legitimate, suspicious, and phishing. The images representing feature vectors are classified using a simple CNN. We developed MATLAB scripts to convert vectors into images and to implement a simple CNN model. The classification accuracy obtained was 86.5 percent.

Keywords—Classification; convolution neural networks; machine learning; phishing URLs

I. INTRODUCTION

Convolution Neural Networks (CNNs) present a tool that enables the computer to learn from image samples, extract internal representations, and classify images. Study into CNN has increased in recent years as the computing power is being available. CNNs have several advantages such as they do not require any feature extraction technique. CNNs extract features through convolution and pooling. Through unique layer designs CNNs can extract higher order statistics and non-linear correlations. Today, many CNN models are available in practice that can be executed efficiently with recent advantages in hardware like Graphical Processing Units (GPUs). CNNs need image data as an input. Conventional Machine Learning (ML) techniques require samples in the form of a feature vector for classification. The purpose of feature extraction to reduce data by measuring certain features or properties that distinguish input samples. Samples belonging to the same categories form clusters in the feature space. The classification problem essentially reduces to partitioning the feature space. When classes overlap in the feature space classifiers such Naïve Bayes classifier makes decisions based on posterior probabilities. Commonly used parametric and non-parametric classification techniques in ML include decision trees, neural networks, minimum distance classifier, Support Vector Machine (SVM), Naïve Bayes classifier, etc. For these conventional ML techniques, input is presented in the table form and each sample is represented by a feature vector, whereas for CNN models, data is presented in the form of an

image. To take advantage of CNNs one-dimensional feature vectors can be converted into two-dimensional images. A lot of data such as genomics, transcriptomic, methylation, mutation, text, spoken words, financial and banking are in non-image form and ML techniques used in these fields. Sharma et al. [1] have suggested a methodology to transform non-image data to image data. They have suggested a method to map a vector consisting of gene expression values to a feature matrix. In their method the location of a feature in the feature matrix depends upon similarity between feature values. In this paper, we suggest a new approach to map a feature vector to a feature matrix or the output image. In our approach, we divide the output image into regions and the gray value of each region is determined by a value in the feature vector. Each region in the output image represents a value in the feature vector. The number of regions in the output image is the same as the dimension of the feature vector. We have considered the problem of classification of Uniform Resource Locators (URLs) using a Deep Convolution Neural Network (DCNN). URL represents documents and other resources on the World Wide Web (WWW). Malicious web sites present a serious threat to cybersecurity. Malicious websites host unsolicited contents such as spam, phishing, viruses, etc. Many Machine Learning (ML) algorithms have been used to classify malicious URLs into classes such as legitimate, suspicious, or phishing. Common types of attacks using a malicious URL include Driven-by download, phishing, and spam. URLs consist of two components the protocol identifier and resource name. These two components are separated by a colon and two forward slashes. The common method to detect malicious URLs is the black-list method, which is database compiled over the period. ML approaches for URL classification use a set of URLs as training data and develop a model. To develop a model, one needs to extract features from URLs. In the present study, we have used a dataset consists of features 1,353 real world URLs that were classified into three categories-legitimate, suspicious, and phishing. The dataset contains ten attributes. The three classes: phishing, suspicious, and legitimate are denoted by -1, 0, and 1 respectively. We have converted the feature vectors into images that were stored in three folders. The folder names are the same as the class names. We developed a MATLAB script to map feature vectors into images, to implement a simple convolution neural network to classify URLs. The outline of the paper is as follows. The related work is provided in Section II. The proposed approach is described in Section III and implementation and results are provided in Section IV. Section V presents the conclusions and the future work.

II. RELATED WORK

This paper deals with phishing detection using convolution neural networks (CNN). CNN models are a part of artificial intelligence (AI). AI includes any technique that enables computers to mimic human behavior and reproduce or excel over human decision making to solve complex tasks independently or with minimum human intervention [2]. AI research deals with reasoning, knowledge representation, natural language processing. AI includes machine learning (ML) algorithms, Artificial Neural Networks (ANN), and deep learning networks. Machine learning evolved from pattern recognition and computational learning theory [3]. ANN models are biologically inspired. They learn from training samples and have used in pattern recognition since 1950s. Many ANN models with learning algorithms such as multilayer perceptron, backpropagation, Boltzmann machine, Hopfield net, neo cognition model etc. are available in practice [4,5,6,7]. Deep learning is a form of machine learning that enables computers to understand the world in terms of hierarchy of concepts. Convolution Neural Networks (CNNs) are special type of networks for processing data that have a known grid like structure [8]. DNNs discover in large datasets using the backpropagation algorithm [9]. CNNs are feedforward networks in that information flow takes place in one direction only, from their input to output [10]. CNN architectures in general consist of convolution and pooling layers that grouped in modules followed by fully connected layers. CNNs evolve into deep convolution neural networks (DCNN). DCNNs proven to be one of the best learning algorithms for understanding image contents and shown exemplary performance in image segmentation, detection, and retrieval tasks [11]. Recent developments in DCNN were possible because of availability of large data sets and graphical processing units (GPUs). With the ability of new programming framework, availability of data, and accessibility to GPUs many analytical models are developed [12]. DCNNs use gradient decent backpropagation algorithm. The use of Sigmoid activation functions leads to saturation resulting into slow convergence of gradient decent algorithm. The problem becomes sever as we go away from the output layers to hidden layers. The compound effect of saturation at multiple layers is known as vanishing gradient [13]. To avoid the vanishing gradient problem, DCNNs often use entropy loss functions with Rectified Linear Units (ReLU) in the output layer. Another issue with DCNNs is overfitting. Various regularization techniques such as dropout or bagging are used to overcome this problem [14].

Phishing URLs is one of significant threats in the world today. Commonly used technique for phishing URLs detection is blacklisting. Blacklists include sender blacklists and link blacklists. The effectiveness of using blacklists depends on update of databases that maintain blacklists. Phishing emails are sent from an Internet disguised as an email from a legitimate, trustworthy source. Many researchers have worked on phishing email detection. Gilehan and Taylor [15] used syntactic features for phishing detection. They presented the

comparison of sentence syntactic similarity and the difference in subjects and objects of target verbs between phishing emails and legitimate emails. Fang et al. [16] suggested a framework to detect phishing emails based on improved recurrent convolution neural networks (RCNN) with multilevel vectors and attention mechanism. In their approach to extract features they divide each email into multiple levels, the character and word level of the email header as well as the character and word level of the email body. Rashid et al. [17] propose an efficient machine learning based phishing detection technique. They first extract lexical, host and word vector features and using the principal component analysis to reduce the number of features and use the SVM model for classification. They use five principal components and obtain the efficiency of 95.66 percent. Machine learning techniques for phishing extract features that distinguish legitimates from phishing websites. Features are extracted from various sources such as URLs, page content, search engine, digital certificate, web traffic etc. Software based approaches are classified into machine learning based, blacklist based, and visual similarity based [18].

Zhang et al. [19] proposed a page content-based technique. Huang et al. [20] proposed an approach that is based on URL features. They have used 23 features from URL and used the SVM. The two classifier values are fed into the fusion model. Abdelhamid et al, [21] built a system for detecting phishing URLs based on associative classification. Hadi et al. [22] proposed an approach for detecting malicious URLs using only visible features from social networks. Kulkarni and Brown [23] have classified phishing URLs using machine learning techniques such as SVM, decision tree, Naïve based classifier, and ANN. Sahoo et al. [24] provide a comprehensive survey and structural understanding of malicious URL detection techniques using machine learning. Yang et al. [25] have proposed a spam filtering method based on multi-model fusion. During pre-processing they separate text and image data from an email. The text dataset to train Long-Short Term Memory (LSTM) and image datasets are used to train a CNN model. CNN architecture allows dealing with images effectively. CNN architecture employs a collection of neighborhood pixels as opposed to individual use of features by ML models [1]. Chiramdasu, et al. [26] explore the various ways of detecting malicious links from the host-based and lexical features of the URL to protect users from being subjected to identity theft attacks. We have used a CNN model for classification of URLs into three classes legitimate, suspicious and phishing. We used features that are extracted from URLs. To use CNN, we first converted feature vectors to images that were classified by the CNN model.

III. METHODOLOGY

In this paper, we propose a framework for classification of phishing URLs. In our approach we use the features that are extracted from URLs. Often phishing emails contain URLs of malicious websites. We use a simple DCNN to classify URLs from their feature vectors. The framework for the proposed approach is shown in Fig. 1. The second step in our approach converts the feature vectors into 2-D image matrices.

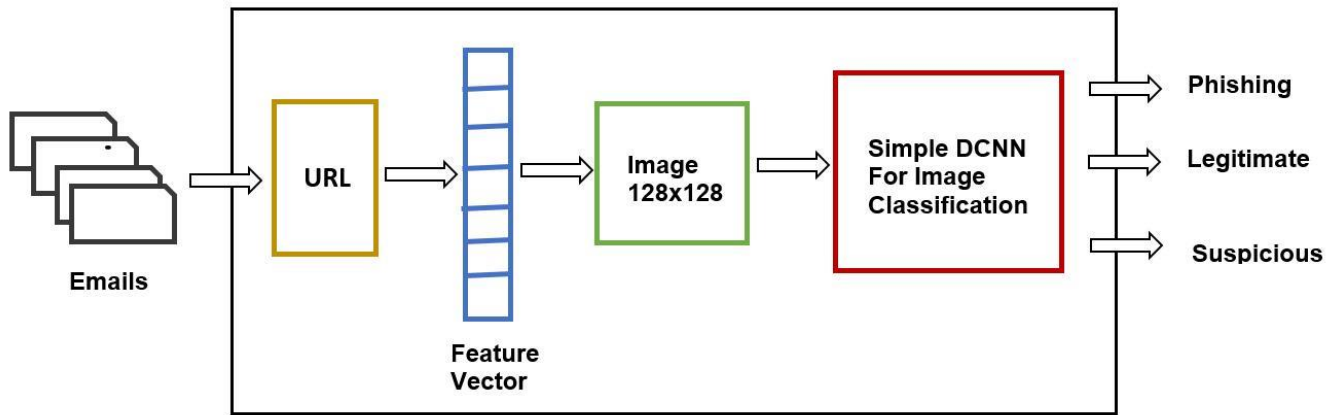


Fig. 1. Framework for phishing URL detection.

In our method we first normalize the values between 0 and 255 and create gray value regions based on the features in the feature vector, 0 corresponds to black and 255 corresponds to white. In between values are mapped to corresponding gray values. Fig. 2 shows the feature vector consists for four features and the output image. The gray values represent numeric values in the feature vector. There are four regions in the image, and each represents a feature value and the image represents the feature vectors. All feature vectors in the dataset are converted to the corresponding images. We created a datastore containing three folders one for each class. The labels of the folders are the same as the class labels. The images were of the size of 128 rows and 128 columns. The images were split into two datasets-training and testing datasets by randomly chosen images. The DCNN was trained with the training set images and was tested with images in the test dataset. Conventional neural networks with the backpropagation learning algorithms have been used for classifying feature vectors. Conventional neural networks during the learning phase use the mean squared error at the output layer and is propagated backwards to hidden layer to update the weights. That causes vanishing gradient problem. We use a simplified model of a DCNN as shown in Fig. 3.

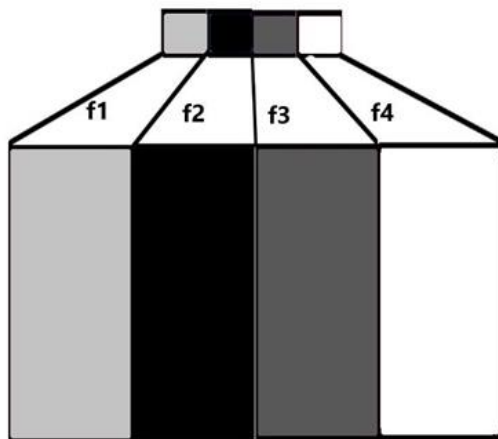


Fig. 2. Mapping a feature vector to an image.

The model consists of the input layer, convolution layer, batch normalization layer, ReLU layer, max pooling layer, fully connected layer, SoftMax layer, and classification layer. We can specify the input image size at the input layer. There are three convolution layers. The batch normalization layers normalize the activations and gradients propagating through the network, which makes the training an easier optimization problem. The batch normalization layers are followed by an ReLU layers. The max pooling layer is used to downsize the network and extract features. The fully connected, SoftMax and classifier layers map the feature vector to class labels. The output of the SoftMax layer consists of positive numbers that sum one that are used as class probabilities. To classify URLs, we have used three classes: phishing, legitimate, and suspicious.

IV. IMPLEMENTATION AND RESULTS

We used a dataset consists of 1,353 real word URLs that are classified into three categories a) phishing, b) legitimate, and c) suspicious. The dataset used in this paper is downloaded from University of California at Irvine (UCI) Machine Learning Repository [27]. The data set consists of ten features that are extracted from each URL. The features include URL of the anchor, Request URL, Server Form Handler (SFH), URL Length, Having “@” character, Prefix/Suffix, IP address, Sub Domain, Web Traffic, and Domain Age. These features are represented by numeric values such as -1, 0, and 1. We transformed each feature vector into a gray image by mapping numeric values in the feature vector to gray value regions. The images are classified using a DCNN model shown in Fig. 3. The image size for the input layer was set to 128x128. We used a 3x3 filter size and 8, 16, and 32 filters in the first, second, and third convolution layers, respectively. We used a 2x2 region size for the max pool layer. The number of units in the output layer was set to three as there three classes in the dataset.

We developed a MATLAB script to convert feature vectors into images. Three folders were created for three classes. The folder names are the same as the class names-phishing, legitimate, and suspicious. The output images were stored in the respective class folders. The images were classified by the simple DCNN. The total number of tuples in the data set is 1353 that represents 702 phishing, 548, legitimate, and 103

suspicious URLs. The samples were split into two datasets- the training and testing datasets. Randomly chosen seventy percent samples were used for training and thirty percent were used for validation. Fig. 4 shows randomly chose sixteen images from the training datasets. The classified images with class labels are shown in Fig. 5. The DCNN was trained using the training set data. Fig. 6 shows the accuracy and error curves with respect to epochs. The validation accuracy obtained was 85.47 percent in eight epochs.

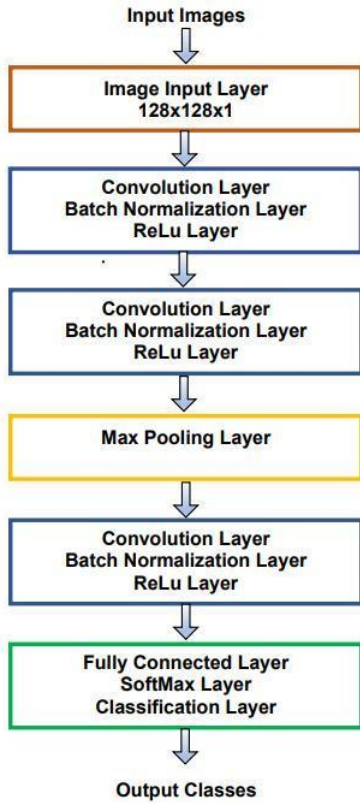


Fig. 3. Architecture for the simplified DCNN.

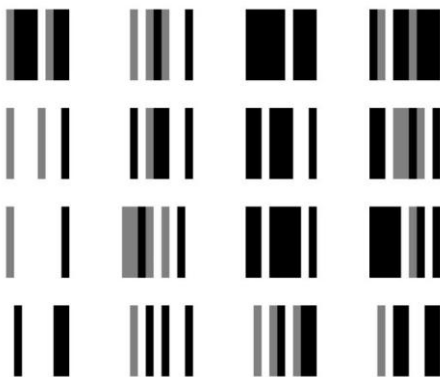


Fig. 4. Sample images from training set data.

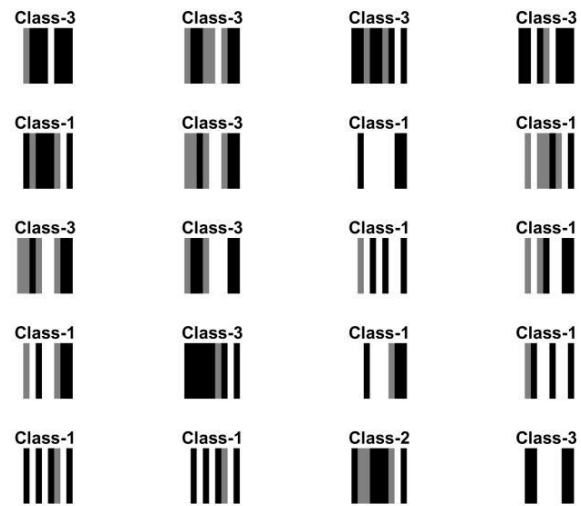


Fig. 5. Classified output images.

V. CONCLUSIONS

In this paper, we suggested a new approach to map non-image data to two-dimensional images so that data in the feature vectors form can be classified using CNN models. We mapped values in the feature vector to regions with different gray shades that are determined by feature values. We developed MATLAB script to convert feature vectors to images and classify them using a simple CNN model. The model was trained to classify real life URLs into three classes legitimate, suspicious, and phishing. We used randomly chosen seventy percent samples for training and thirty percent for testing. We obtained an accuracy of 85.56 percent. There are many ways to improve classification accuracy. In our method we used rectangular regions to map values in the feature vectors to corresponding gray regions in the output image. It is possible to use more complex shapes to define the regions. It is also possible to define shapes in the output image as a function of feature values. Furthermore, we can use DCNN models with a greater number of layers such as Alex Net, Res-Net, etc. to classify images obtained from the feature dataset. Our future work includes classifying data with DCNNs and testing the models with big datasets. In the present data set attributes consists of only three discrete values -1, 0, and 1. We plan to test the algorithm for features with multiple discrete values and explore complex shapes for mapping feature vectors to images and evaluate the suggested algorithm by comparing it with other machine learning algorithms.

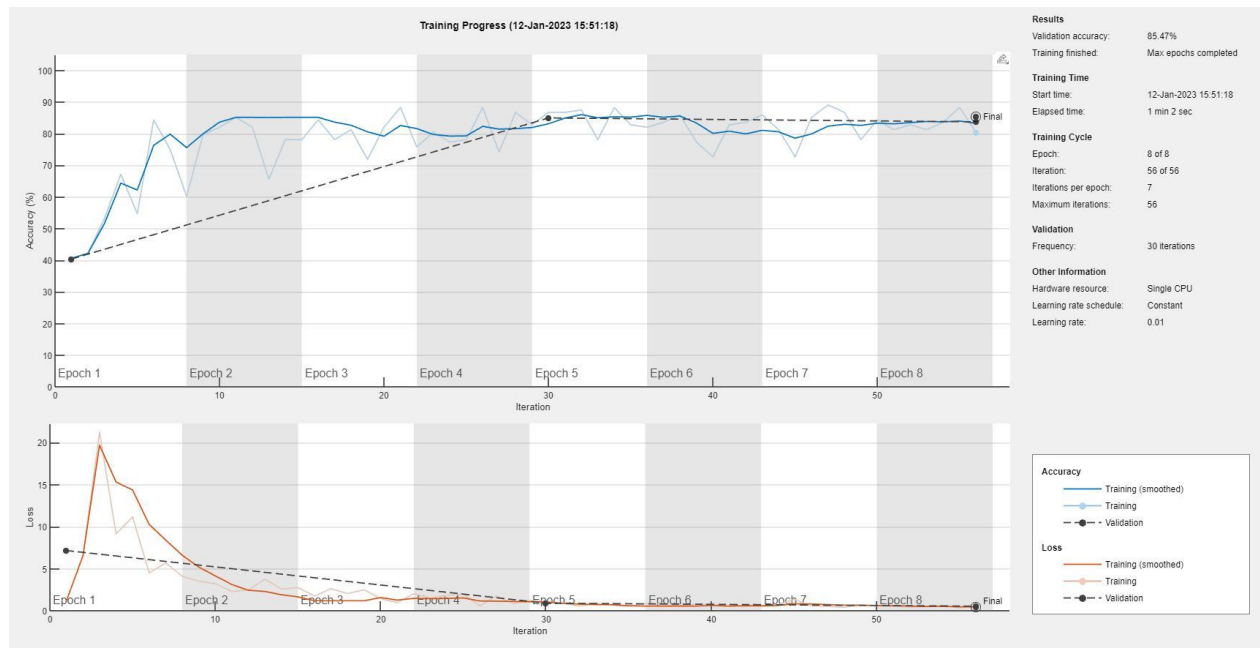


Fig. 6. Training progress plot.

REFERENCES

- [1] A. Sharma, E. Vans, D. Shigemizu, K. A. Boroevich, and T. Tsunoda, "Deep Insight: A methodology to transform a non-image data to an image for convolution neural network architecture". *Sci Rep* 9, 11399, 2019.
- [2] S. J. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, Global Edition. Pearson, Harlow, UK, 2022.
- [3] W. L. Hosch, *Machine Learning* [Online] Retrieved 2017-06-01 from <http://www.britannica.com/EBchecked/topic/1116194/machinelearning>.
- [4] J. J. Hopfield, "Neurons with graded response have collective computational properties like those of two state neurons," *Proceedings of the National Academy of Sciences*, 1984, vol. 81, pp. 3088-3092.
- [5] D. E. Rumelhart, J. L. McClelland, and the PDP Group, *Parallel Distributed Processing*, vol. I, MIT Press, Cambridge, MA, 1986.
- [6] R. P. Lippmann, "An introduction to computing with neural nets," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 32, pp. 4-22, 1987.
- [7] K. Fukushima, "Neural networks for visual pattern recognition," *Computer*, pp. 65-74, March 1988.
- [8] I. Goodfellow, Y. Bengin, and A. Courville, *Deep Learning*. The MIT Press, Cambridge, MA, USA, 2016.
- [9] Y. LeCun, Y. Bengin, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436-444, 2015.
- [10] W. Rawat, and Z. Wang, "Deep convolution neural networks for image classification: A comprehensive review", *Neural Computation*, vol. 29, pp. 2352-2449, 2017.
- [11] D. Cireşan, U. Meier, J. Schmidhuber, "Multi-column Deep Neural Networks for Image Classification," *Computer Vision and Pattern Recognition*, 2012, pp. 3642-3649.
- [12] E. Brynjolfsson, and A. McAfee, "The business of artificial intelligence", *Harvard Business Review*, pp. 1-20, 2017.
- [13] M. Tan and Q. V. Le, "Efficient Net: Rethinking Model Scaling for Convolutional Neural Networks," *Proceedings of the 36th International Conference on Machine Learning*, Long Beach, California, PMLR 97, 2019.
- [14] A. D. Kulkarni, "Deep Convolution Neural Networks for Image Classification", *International Journal of Advanced Computer Science and Applications*, Vol. 13, No. 6, pp 18-23, 2022.
- [15] P. Gilchan, and J. M. Taylor. "Using syntactic features for phishing detection." *arXiv preprint arXiv:1506.00037*, 2015.
- [16] Y. Fang, C. Zhang, C. Huang, L. Liu, and Y. Yang., "Phishing email detection using improved RCNN model with multilevel vectors and attention mechanism," *IEEE Access*, vol. 7, pp. 56329-56340, 2019.
- [17] J. Rashid, T. Mahmood, M. W. Nisar, and T. Nazir., "Phishing Detection Using Machine Learning Technique" 2020 First International Conference of Smart Systems and Emerging Technologies (SMARTTECH). doi: 10.1109/SMART-TECH49988.2020.00026
- [18] A. K. Jain and B. B. Gupta., "Comparative analysis of features based machine learning approaches for phishing detection", *IEEE International Conference on Computing for Global Development*, pp 2125-2129, 2016.
- [19] Y. Zhang, J. Hong, and L. Cranor, "CANTINA: a content-based approach to detecting phishing web sites" in *Proceedings of the 16th International Conference World Wide Web*, 2007.
- [20] H. Huang, L. Qian, Y. Wang, "A SVM-based technique to detect phishing URLs", *Inf. Technol. J.* vol. 11, no. 7, pp. 921-925, 2012.
- [21] N. Abdelhamid, A. Ayes, F. Thabtah, "Phishing Detection based Associative Classification", *Data Mining. Expert Systems with Applications (ESWA)*, vol. 41, pp 5948-5959, 2014.
- [22] W. Hadi, F. Aburrub, and S. Alhawari, "A new fast associative classification algorithm for detecting phishing websites", *Applied Soft Computing*, vol. 48, pp 729-734, 2016.
- [23] A. D. Kulkarni and L. Brown, "Phishing Websites Detection using Machine Learning". *Journal of Advanced Computer Science and Applications*, vol. 10, no. 7, pp. 8-13, 2019.
- [24] D. Sahoo, C. Liu, and C. H. Hoi, "Malicious URL detection using machine learning: A Survey", <https://arxiv.org/abs/1701.07179>, 2017.
- [25] H. Yang, Q. Liu, S. Zhou, and Y. Luo, "A spam filtering method based on multi-modal fusion, *Appl. Sci.* vol. 9, 2019. doi:10.3390/app9061152.
- [26] R. Chiramdasu, G. Srivastava, S. Bhattacharya, P. K.Reddy, T. R. Gadekallu, "A machine learning driven threat intelligence system for malicious URL detection", *Proceeding of 16th International Conference on Availability, Reliability, and Security*, August 2021, Arical 154, pp 1-7, doi.org/10.1145/3465481.3470029.
- [27] UCI Machine Learning Repository: Website Phishing Data Set (Online) <https://archive.ics.uci.edu/ml/datasets/Website+Phishing>.

A Radial Basis Network-based Early Warning Algorithm for Physical Injuries in Marathon Athletes

Ruisheng Jiao¹, Juan Luo^{2*}

College of Physical Education, Chizhou University, Chizhou, China¹

College of Foreign Languages, Chizhou University, Chizhou, China²

Abstract—For marathon runners, a single injury may affect their lifelong athletic career, so their injury management is very important. The current injury management for marathon runners has a certain lag, and the current injury warning is mainly based on manual teams, which is costly and poorly automated. To solve these problems, the study proposes a marathon athlete physical injury warning algorithm based on inertia weight adjustment optimized radial basis network. Particle swarm optimization technology has also been incorporated into early warning algorithms. Finally, an athlete injury and disease early warning model is constructed based on the algorithm. The results of performance tests show that the algorithm has a minimum fitness function value of 0.13, which is significantly lower than the current algorithm used for comparison. In the test with real data, the MAPE of the proposed algorithm was as low as 7.598% and the agreement of the hazard score results with the expert human assessment reached 100%. The results of the study indicate the practicality of the algorithm to assist work teams and perform early warning of physical injuries in athletes. However, the high number of iterations required is a limitation awaiting resolution.

Keywords—Radial basis neural network; exponentially decreasing inertia weights; early warning algorithm; sports injury; marathon; particle swarm; model building

I. INTRODUCTION

Marathon is one of the most popular athletic events. For marathoners, injury management and prevention are of the utmost importance. A single serious injury can end an athlete's career. However, there is a certain lag in the current injury management of marathon athletes. In addition, marathoner injury management is currently mainly through human analysis and decision making, which is costly and less automated [1]. To solve these problems, an early warning algorithm for physical injuries of marathon runners is proposed. The research gap of this study is focused on injury prevention for athletes. There are few studies focused on automation and algorithm applications, and there is still a large research gap in this field. Radial basis function neural networks have excellent approximation and global optimization capabilities, and are an effective tool for solving early warning problems [2]. Currently, there are few studies on the application of radial basis function networks in athlete injury prevention. This research can enrich these fields. To address the problem of balancing global and local search performance in radial basis networks, the Exponential Decreasing Inertia Weight (EDIW) strategy is introduced into radial basis networks. The physical injury warning algorithm for marathon runners is constructed

based on the radial basis network optimized by EDIW, and it is hoped that this study will bring meaningful results for injury management and warning of marathon runners.

This article has a total of V parts. The second part is related works, which reviews the research achievements in relevant fields in recent years, laying a foundation for this research. The third part is methods, which introduces the design idea of the algorithm and the construction process of the model. The fourth part is the experimental results, which show the performance of the proposed algorithm in the experiment. The fifth part is the conclusion, which summarizes the results of this study.

II. RELATED WORKS

Radial basis function neural network is a type of feedforward neural network that has superior performance and has been widely used in the prediction and warning fields. Zijie N led a team to build a mobile platform control system integrated with two Radial basis function neural networks, one for identifying the system's state and the other for predicting the mobile platform's deviation angle based on existing data [3]. The experimental results showed that the application of this algorithm during longitudinal driving reduced the correction time by 1.4 seconds and the overshoot angle by 7.4 degrees compared to traditional control algorithms. Additionally, Wang H and their research team proposed a robot fault-tolerant control model based on Radial basis function neural network prediction, which estimated external disturbances using an RBFNN and automatically handled hazardous factors employing trajectory tracking techniques [4]. The experimental results showed that this RBFNN model could predict external disturbances with an accuracy rate of over 70% and effectively mitigate their impact using tracking and vibration elimination techniques. Furthermore, Lian X and other researchers proposed a sliding mode controller based on an adaptive Radial basis function neural network that introduced a track modeling approach with 12 degrees of freedom [5]. Through performance testing and comparative analysis, the algorithm was found to perform high-precision satellite capture and release tasks. Current Radial basis function neural network research tends to focus on practical applications, with limited research investigating performance optimizations of RBFNN itself. This study addresses its own performance limitations to some extent by optimizing RBFNN prior to its use in practical applications.

Injury management and prevention in athletes has long been an important research topic in the field of sports. Ye and

Di studied injury and fatigue in a large number of winter Olympic athletes and continuously monitored their psychological status for injury prevention in winter athletes [6]. The results of the study showed that there was a significant correlation between the adequacy of athletes' preparation activities and the rationality of training programs and athletes' injuries. Wang and his research partner developed a mutual information sports injury warning model based on an attribute parsimony algorithm, which was designed for youth athletics [7]. Simulation experiments found that the model was able to warn youth track and field athletes of injuries with 80% correctness, but the model suffered from a local optimal solution. Bahr led his team to explore the characteristics of athletes' injuries and diseases based on epidemiological research methods, and they reached consensus on a set of recommendations for the latest sports injury and disease research and proposed an athlete epidemiological research report list extension [8]. The study provided a systematic understanding of the causes of injury and disease in athletes and developed protective measures accordingly. Li used data fusion techniques to analyze and assess potential injury factors in various sports and based on this, developed a dynamic chain model for early warning of risk factors for sports injuries [9]. The study provides a reference for athletes to avoid and reduce injury risk and guarantee normal training and competition, and the authors also applied the research results to tennis training and achieved scientific results. Chia and her research team studied injury prevention in athletes from a social marketing perspective and proposed a strategy to implement athlete injury prevention efforts using a social marketing mix [10]. The team analyzed in detail the useful features of the social marketing mix, including elements such as product, price, and location, and provided high-value recommendations on the corresponding injury prevention programs. According to the analysis results of the literature in the field of athlete injury management, it is found that there is less research on the balance between local and global search capabilities, and the application of radial basis function networks in the field of sports is also relatively lacking. This indicates that there is not much research focused on automation and algorithm applications in this field, and there is still large research space. Therefore, the improved radial basis network is applied to injury warning for marathon runners, hoping to bring practically meaningful research results to these fields.

III. RADIAL BASIS NETWORK-BASED PHYSICAL INJURY WARNING ALGORITHM CONSTRUCTION FOR MARATHON RUNNERS

A. Radial Basis Network Model for EDIW Optimization

Radial basis networks belong to feedforward neural networks and have excellent approximation and global optimization capabilities [11]. In addition, radial basis networks have a simpler structure compared to other feedforward networks and are therefore widely used in approximation, classification, and regression problems [12]. The marathon runner physical injury warning algorithm is based on a special radial basis network, and the topology of this network and its difference from the ordinary radial basis network are shown in Fig. 1.

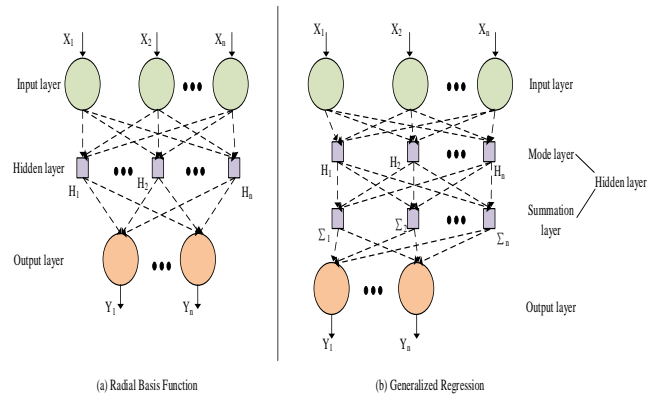


Fig. 1. Radial basis function network model.

Fig. 1(a) shows the structure of the ordinary radial basis network. Fig. 1(b) shows the adopted special radial basis network, which is based on nonlinear regression and is called generalized regression network. The difference from the ordinary radial basis network is that the hidden layer of the generalized regression network is a two-layer structure, i.e., the mode layer and the summation layer [13]. The mode layer is activated using a radial basis Gaussian function, while the summation layer performs direct and weighted summation of the output values of the mode layer, respectively [14]. This structure allows radial basis networks to optimize the warning effect by eliminating the need to adjust the connection weights and only changing the smooth factor to affect the activation function of the mode layer [15]. Current particle swarm optimized radial basis networks have received attention for their stronger global search capability and computational efficiency, and therefore damage warning algorithms also use particle swarm optimized generalized regression radial basis networks [16]. However, the global search and local exploration capabilities of such optimization networks are often not easily balanced, so how to adjust the inertia weights of the algorithm and achieve the best balance is a key issue for such networks [17]. In this study, an EDIW-based strategy is proposed, and the operational flow of the radial basis network model optimized by this strategy is shown in Fig. 2.

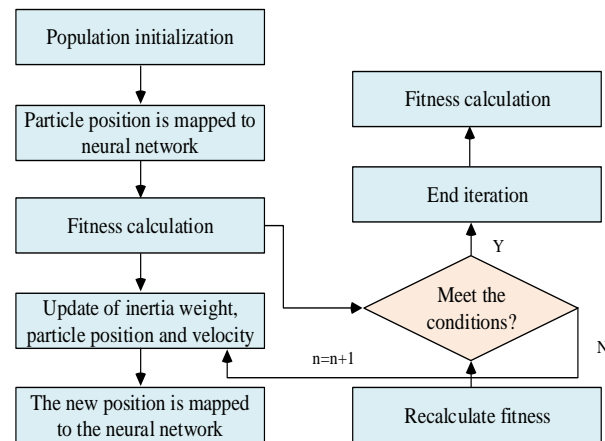


Fig. 2. Operation process of EDIW optimized RBF network model.

In this operational flow, the particle population is first initialized and the positions of the particles are mapped to the radial basis network. After that, the fitness is calculated and the inertia weights and particle positions and velocities are updated, and then the new values are mapped into the radial basis network. The mathematical expression of the EDIW strategy is shown in Eq. (1).

$$\omega(n) = \frac{n(\omega(\max) - \omega(\min))}{N} \quad (1)$$

In Eq. (1), n is the number of iterations of the neural network, and N is the maximum number of iterations. $\omega(\max)$ and $\omega(\min)$ are the maximum and minimum initial inertia weights, respectively. According to the mathematical properties of the EDIW expression, it is a linear function with the decline as shown in Eq. (2).

$$LD = \frac{\omega(\max) - \omega(\min)}{N} \quad (2)$$

The LD in Eq. (2) is the weight reduction. The inertia weights have two main roles for the motion damage warning algorithm, one is to adjust the influence of the historical velocity on the current velocity, and the other is to balance the global detection and local search ability. Therefore, at the beginning of the iteration, the inertia weights should decrease at a faster rate to ensure that the particle swarm can search the region where the feasible solution is located more quickly. At the later stage of the iteration, the inertia weight decreases at a significantly slower rate, thus limiting the step size of particle updates, which allows the particles to increase their ability to search for the optimal solution in the region of feasible solutions. Based on this theory, an EDIW strategy based on control parameters is proposed, as shown in Eq. (3).

$$\omega(n) = \exp\left(-\frac{Con * n}{N}\right) * (\omega(\max) - \omega(\min)) + \omega(\min) \quad (3)$$

In Eq. (3), Con is the control parameter whose value is always greater than 0. The strategy adds this parameter to the normal EDIW and uses it to change the drop in inertia weights. The decrease is affected by the variables Con , $\omega(\max)$ and $\omega(\min)$. Provided that the maximum and minimum initial inertia weights remain unchanged, the decrease in the weights according to the change in Con is shown in Eq. (4).

$$|\Delta\omega(n, Con)| = \frac{\exp\left(-\frac{Con * n}{N}\right) * (\omega(\max) - \omega(\min)) * Con}{N} \quad (4)$$

In Eq. (4), $\Delta\omega(n, Con)$ indicates how much the weight decreases with Con . The drop in weight becomes smaller as the number of iterations goes up. Assuming $n = N$, the expression of $|\Delta\omega(n, Con)|$ becomes Eq. (5).

$$|\Delta\omega(n, Con)| = Con * (\omega(\max) - \omega(\min)) * \exp(-Con) \quad (5)$$

The derivative function analysis of Eq. (5) is shown in Eq. (6).

$$\Delta' = \exp(-Con)(Con - 1) \begin{cases} > 0 & x < 1 \\ = 0 & x = 1 \\ < 0 & x > 1 \end{cases} \quad (6)$$

According to Eq. (6), when the value of the control parameter is greater than 1, the decline of the weights gradually converges to 0 with the increase of the control parameter. The maximum value of the decline is $\frac{1}{e}$, when the value of the control parameter is 1. So far the algorithm has achieved the design of the inertia weights with the number of iterations. The rate of decrease of the inertia weight decreases with the increase of the number of iterations. At the beginning of the iteration, the inertia weights are larger, which leads to the fact that the particles will retain more of the historical velocity. It is easy to see that the output results of the algorithm under this model are greatly influenced by the control parameters, so the values of the control parameters are important to ensure accurate results. Depending on the number of iterations, the selection of the control parameters also needs to satisfy different conditions. When $n = N$, the control parameters must be such that the inertia weights can reach or converge to $\omega(\min)$, and the inertia weights at this time are shown in Eq. (7).

$$\exp(-Con) = \frac{\omega(n) - \omega(\min)}{\omega(\max) - \omega(\min)} \quad (7)$$

In this case, the inertia weights are taken as shown in Table I.

TABLE I. INERTIA WEIGHTS

Con	1	2	3	4	5	6	7	8	9	10
$\omega(\max)$ =0.7	0.3	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
$\omega(\min)$ =0.2	84	65	33	09	04	01	01	00	00	00
$\omega(\max)$ =0.9	0.4	0.2	0.2	0.2	0.2	0.2	0.2	2.0	0.2	0.2
$\omega(\min)$ =0.2	58	95	34	11	05	02	01	1	00	00
$\omega(\max)$ =0.7	0.4	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3
$\omega(\min)$ =0.3	45	54	20	09	03	01	00	00	00	00
$\omega(\max)$ =0.9	0.5	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3
$\omega(\min)$ =0.3	21	79	29	13	05	02	01	00	00	00
$\omega(\max)$ =0.7	0.5	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4
$\omega(\min)$ =0.4	14	46	19	05	02	00	00	00	00	00

When $0 < n < N$, the control parameters need to be taken in such a way that the weights fall faster and then slower. In

this case, the values of the control parameters can lead to large differences in the drop curves of the inertia weights, and it is necessary to test the drop curves with different parameter values to determine the optimal parameter values.

B. Construction of Injury Influencing Factors Model and Injury Warning Algorithm for Marathon Runners

Most sports injury events in marathon runners are not triggered by a single factor, but by a plural number of influencing factors acting together [18]. Therefore a correct analysis of the influencing factors is the basis of the early warning algorithm. According to the definition of sports injuries, the risk factors that induce this type of injury can be classified as internal and external factors [19]. Internal factors refer to the physical condition of the athlete, including age, muscle strength, injury history, etc. External factors are environmental factors other than the athletes themselves, including terrain, weather, etc. [20]. However, the model of injury influencing factors from internal and external factors only lacks comprehensiveness, so stimulus-triggering factors were added as the third assessment dimension in the influencing factor model. Stimulus-predisposing factors are factors that amplify the likelihood of an athlete's injury while internal and external influences remain unchanged. When an athlete's internal or external influences are abnormal, the athlete is judged to be an injury prone individual [21]. Injury-prone individuals are significantly more likely to be injured in marathon sports than non-injury-prone individuals and require extra attention. When stimulus triggers are present, the likelihood of injury increases and the risk is higher for injury-prone athletes, at which point the algorithm needs to warn the athlete and their team in a timely manner. This study compiled injury-influencing factors for marathon athletes based on the opinions of a panel of professional track and field players and coaches, as well as a large number of actual situations in competition, as shown in Fig. 3.

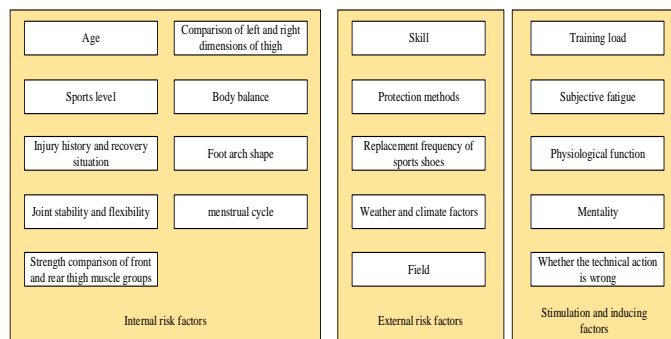


Fig. 3. Influencing factors of injury and disease of marathon athletes.

After importing the injury influencing factors into the radial basis network and training, the early warning algorithm is able to determine the injury risk of marathon runners by the input physical and other related information. The athlete's injury risk is classified as low, medium or high. The algorithm does not warn for low risk. For medium risk, the algorithm issues an alert and indicates the source of the risk to the athlete and his or her team. For high risk, the algorithm issues a warning and strongly advises the athlete not to participate in the competition or to take appropriate measures. The hidden layer activation

function of the radial basis network is a Gaussian function, whose mathematical expression is shown in Eq. (8).

$$f_j(x) = \exp\left(-\frac{\|x - c_j\|^2}{2\sigma_j^2}\right) \quad (8)$$

In Eq. (8), $f_j(x)$ represents the output value of the j hidden layer node. x It is the independent variable and the input to the network. c_j is the center vector of the kernel function of the j hidden layer node. Under this activation function, the output of the radial basis network is shown in Eq. (9).

$$y_j = \sum_{j=1} \omega_j f_j(x) \quad (9)$$

In Eq. (9), y_j is the output of the i th output layer node, and ω_{ij} is the connection weight between the i th and j th hidden nodes. Since the position vector of each particle in the radial basis network is composed of the function center, function width and connection weights, these three elements need to be defined [22]. First an error function needs to be defined, as shown in Eq. (10).

$$M = \frac{\sum_{q=1} E_q^2}{2} \quad (10)$$

In Eq. (10), M is the error function and E_q is the error of the input sample. The definition of the error is shown in Eq. (11).

$$E_q = d_q - \sum_{j=1}^3 \omega_{ij} \exp\left(-\frac{\|x - c_j\|^2}{2\sigma_j^2}\right) \quad (11)$$

Eq. (11) in d_q represents the value of the desired type of sample. The values vary according to the different athlete injury risk levels, 1 and 2 for low risk, 3 and 4 for medium risk, and 5 for high risk. According to the error function, the weights of the radial basis network output cells are shown in Eq. (12).

$$\omega_{ij}(m+1) = \omega_{ij}(m) - \eta \frac{\partial M(m)}{\partial \omega_{ij}} \quad (12)$$

In Eq. (12), $m+1$ is the value after iteration and m represents the current value of the individual variable. η represents the learning efficiency. The radial basis network implicit layer function centers are defined as shown in Eq. (13).

$$c_j(m+1) = c_j(m) - \eta \frac{\partial \eta(m)}{\partial c_j} \quad (13)$$

Eq. (13) in c_j is the center of the function. The width of the function is defined, as shown in Eq. (14).

$$\sigma_j(m+1) = \sigma_j(m) - \eta \frac{\partial M(m)}{\partial \sigma_j} \quad (14)$$

Finally, the fitness function of the radial basis network needs to be defined. The fitness function uses the relative error function between the actual output and the network output, and its mathematical expression is given in Eq. (15).

$$F = \frac{\sum_{r=1}^R \sum_{k=1}^K (y_{rk} - \dot{y}_{rk})}{R} \quad (15)$$

In Eq. (15), y_{rk} represents the first r output value of the k output neuron, and \dot{y}_{rk} is its actual value. R is the sample size and K is the number of output neurons. In the practical application of the physical injury warning algorithm for marathon runners, it is first necessary to initialize the particle population and map its position into the radial basis network. After training, the radial basis network calculates the global extremes of the particles. After updating the weights, the particle fitness is calculated again and iterated continuously. If the population fitness after iteration is better than the last one, the iterated extreme value is used as the new extreme value. Keep iterating until the global extreme value meets the fitness filtering condition to output the prediction result.

IV. PERFORMANCE AND APPLICATION EFFECT TEST OF PHYSICAL INJURY WARNING ALGORITHM FOR MARATHON RUNNERS

The experimental hardware environment is a computer system with I7 processor and 8G memory, and the programming environment is PYTHON, and the experimental procedure is to test the theoretical performance of the early warning algorithm by simulating the environment and test data set, and finally to conduct the practical application test by using the data of real marathon runners' bodies and other related factors. Since the performance of the proposed warning algorithm is greatly influenced by the control parameters, the optimal control parameters need to be confirmed first. The inertia weight variation curves under different control parameters are shown in Fig. 4.

The number of iteration steps of the radial basis network is set to 1000 in Fig. 4, and the maximum and minimum initial weights are 0.7 and 0.2, respectively. Fig. 4 illustrates that smaller control parameters ensure that the inertia weights change slowly at the beginning of the network iteration, allowing the algorithm more space to find the ideal region while larger control parameters enable the algorithm to arrive at the minimum initial weight when it is iterating to the maximum number of iterations. When the control weights are above eight, the inertia weights drop too fast and reach the lower bound at 400 iterations. And when the inertia weight is less than six, the inertia weight decreases too slowly, and the lower bound is still not reached by the maximum number of iterations. In general, when the control parameter ranges from 6 to 8, the inertia weight decreases more satisfactorily. This experiment takes the middle number 7 as the control parameter

of radial base network. After completing the control parameter setting, the simulation performance test of the warning algorithm was started. Firstly, experiments were conducted on the variation of the fitness function curve of the parameters with the number of iterations. In order to compare and determine the differences between the proposed algorithm and the existing algorithms, a normal radial basis network, a particle swarm optimized radial basis network and a linear decreasing inertia weight (LDIW) optimized radial basis network are used here as the comparison algorithms, and the results are shown in Fig. 5.

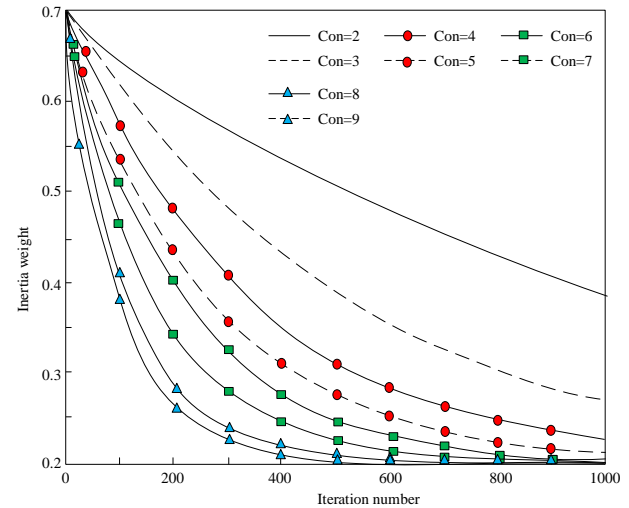


Fig. 4. Change of inertia weight under different control parameters.

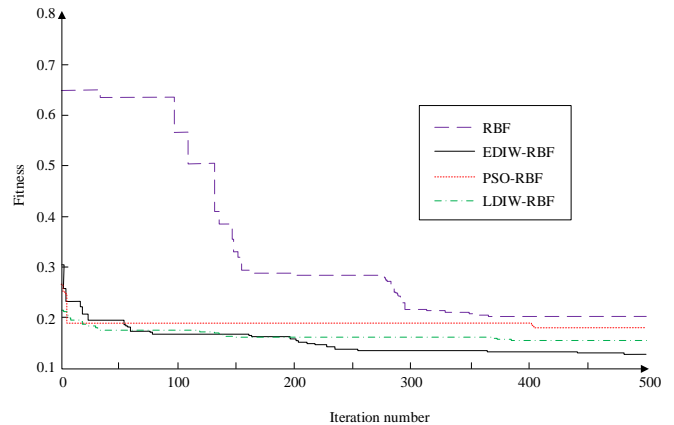


Fig. 5. Change of algorithm fitness function curve.

According to Fig. 5, it can be seen that the radial basis network without any optimization has the slowest convergence rate in terms of fitness, and the fitness that reaches stability after convergence is 0.22, which is greater than the other algorithms. The proposed EDIW optimized radial basis network shows a convergence speed very close to the other two optimized radial basis networks in the experiment and has the smallest fitness function value, which is 0.13. The four algorithms have the highest to lowest fitness function values in order of radial basis networks; particle swarm optimized radial basis networks, LDIW radial basis networks and the proposed

algorithm. The results illustrate that the proposed early warning algorithm has the smallest neural network training error and shows the highest accuracy with negligible differences between the convergence speed and similar optimization algorithms. The next experiment trained the algorithm using time-series data of injury influencing factors of marathon runners, and the output obtained is shown in Fig. 6.

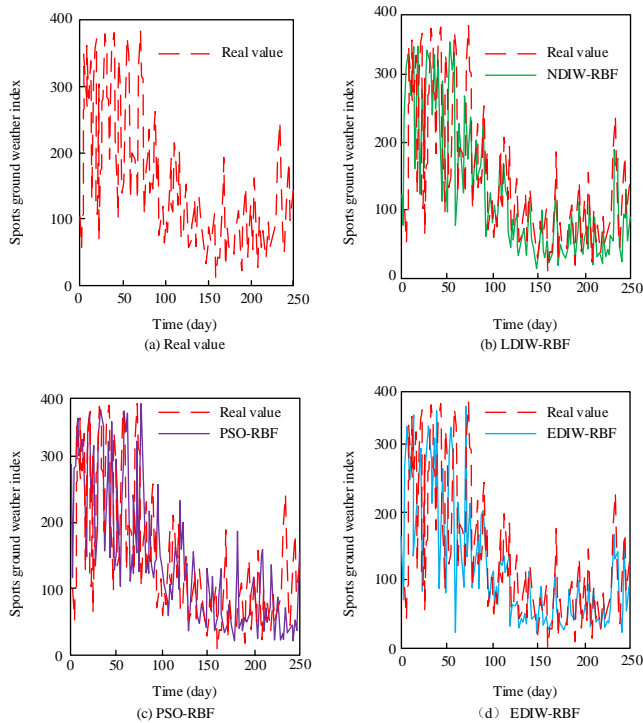


Fig. 6. Output curve of algorithm in training.

In Fig. 6, Fig. 6(a) shows the actual values of the training data; Fig. 6(b) shows the superimposed images of the output of the radial basis network optimized by LDIW and the actual values; Fig. 6(c) shows the superimposed images of the output of the radial basis network optimized by particle swarm and the actual values; Fig. 6(d) shows the superimposed images of the output of the proposed algorithm and the actual values. It can be seen that the weather index of the sports field in this dataset reaches a maximum of 390 and a minimum of 11, showing an overall trend of significant fluctuations followed by a decrease. All three algorithms used in this session are able to restore the trend of the real data, but some errors can still be observed. To further compare the errors in a quantitative and visual way, the error images of the three algorithms in this session were drawn, as shown in Fig. 7.

In Fig. 7, Fig. 7(a) shows the error image of the particle swarm optimized radial basis network; Fig. 7(b) shows the error image of the LDIW optimized radial basis network; and Fig. 7(c) shows the error image of the proposed algorithm. It is not difficult to see that the training errors under all three algorithms show a trend from large to small. Before the 80th iteration, the error values of all three algorithms fluctuate more drastically. Both the radial basis network with particle swarm optimization and the radial basis network with LDIW optimization show an error of more than 250. The maximum

error of the proposed algorithm, on the other hand, is 235, which shows an advantage in terms of the error maximum. After the 80th iteration, the errors of all three algorithms decreased significantly, indicating that the predictive stability of the algorithms for the test data increased with the number of iterations. The mean square error (MSE), root mean square error (RMSE) and mean absolute error (MAE) of the three algorithms in training were tested, and the three metrics of the proposed algorithms were found to be 2.521, 0.129, and 20684, which are the smallest among the three compared algorithms. This represents that the error of the proposed algorithm is the smallest of several algorithms. After completing the network training and related experiments, the actual accuracy of the early warning algorithm was further tested using real data sets, and the test results are shown in Table II.

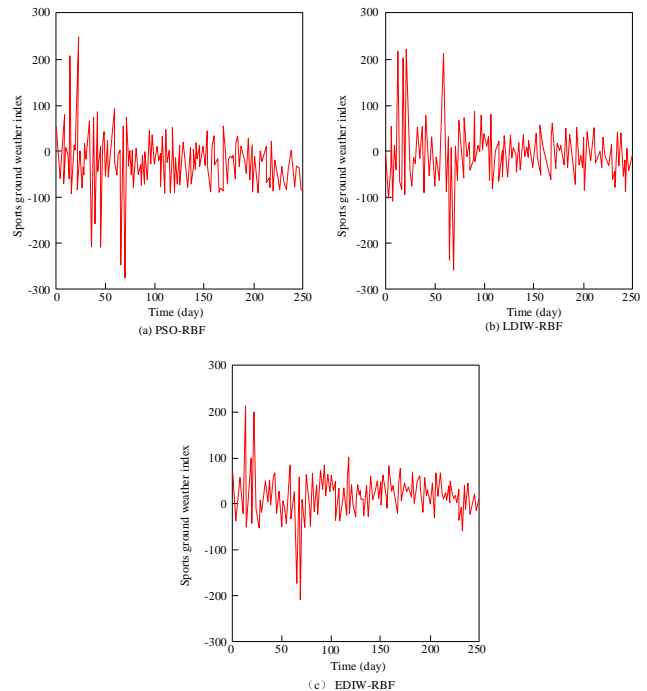


Fig. 7. Error comparison of different algorithms.

TABLE II. ALGORITHM ERROR TEST RESULTS UNDER REAL DATA

Case number	PSO-RBF	LDIW-RBF	EDIW-RBF	Actual value
1	224.34	223.14	234.75	231.50
2	247.48	245.87	236.61	223.00
3	241.01	250.63	241.11	281.00
4	247.00	248.47	241.75	215.00
5	249.63	246.33	239.07	247.00
Index	\	\	\	\
MAPE (%)	8.796	8.213	7.598*	\
MSE	6.846	5.268	4.996*	\
RMSE	0.011	0.010	0.008*	\

^a Note: "*" indicates that the error of this item is the lowest among the three algorithms.

Table II provides the error performance of the proposed algorithm and the two compared algorithms with the real data set and uses it to evaluate their actual accuracy. The proposed algorithm exhibits an RMSE of 0.008, MSE of 4.996, and

Mean Absolute Percentage Error (MAPE) of 7.598% in this session. The three errors metrics of the proposed algorithm are the lowest among the compared algorithms, which indicates that the actual accuracy of the proposed algorithm is better than the other two algorithms. This result further demonstrates the feasibility of applying the improved EDIW to the particle swarm radial basis function neural network, and its optimization effect is higher than other similar algorithms. Although the proposed algorithm has shown better performance than similar algorithms in predicting injury and illness factors in marathon runners, in reality most injury and illness risk factors management in marathon runners is still implemented by expert teams for human management. This management approach has proven to be effective, but consumes more human resources. Although the proposed algorithm can save labor cost, its consistency with expert team decision making still needs to be proven. Therefore, the experiments were conducted using the same set of athletes and race data, allowing the expert team and the algorithm to be evaluated separately, and the results are shown in Fig. 8.

Fig. 8(a) shows the results of the radial basis network with particle swarm optimization versus expert team decision making. Fig. 8(b) and Fig. 8(c) show the results of LDIW versus the output of the proposed algorithm versus the expert team decision, respectively. Although all three algorithms agree with the expert decision results in terms of the trend of the hazard level for different cases, only the hazard level scoring results of the proposed algorithm agree with the expert assessment by 100%. The agreement between the results of the particle swarm algorithm and the expert assessment is only 20%, while the agreement between the LDIW optimization algorithm and the expert assessment is 80%. This result indicates that the judgment of the radial basis network with particle swarm optimization has been very close to the judgment of experts who have worked in the industry for many years, which represents that compared to other algorithms, the learning effect of the radial basis network with particle swarm optimization is better, following the results of personal judgment by the close person.

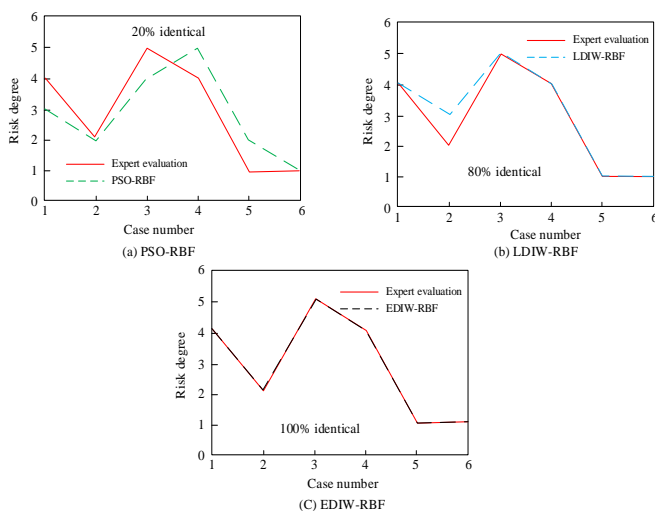


Fig. 8. Comparison between algorithm output and expert evaluation results.

V. CONCLUSION

To address the problems of lag and low automation in modern marathon tele-mobilization injury management, this study proposes a marathon athlete physical injury warning algorithm based on particle swarm radial basis neural network with EDIW optimization. The algorithm utilizes the EDIW strategy to optimize the search process of the neural network to ensure that the algorithm achieves an optimal balance between global search and local search. The minimum fitness function value of the proposed algorithm is 0.13, which is lower than the values of particle swarm radial basis network and LDIW radial basis network. In the training session of the neural network, the MSE, RMSE, and MAE of the proposed algorithm are 2.521, 0.129, and 20684, respectively, with lower errors than other comparative algorithms. In the test with real data, the MAPE of the proposed algorithm is as low as 7.598%, while the MAPEs of the LDIW radial basis network and the particle swarm radial basis network used as comparisons are both above 8.1%. In the comparison with the expert group's assessment, only the hazard score results of the proposed algorithm reached 100% agreement with the expert assessment. The experimental results demonstrate the practicality of this injury warning algorithm and its ability to enhance the automation of injury management for marathon runners. Although the construction of this algorithm has been successful, there are obvious limitations. Only when the number of iterations is high enough, the negative impact of parameter initialization of the proposed algorithm is small. However, this will lead to a large amount of computation, which leads to a high overhead of algorithm operation. In the case of a low budget, this algorithm may be difficult to apply to reality. How to reduce the computational complexity of the algorithm while ensuring performance is the focus of future research. In this regard, based on global optimization algorithms, using algorithmic search to find the most appropriate initial value, thereby reducing the amount of computation, is a promising direction.

FUNDING STATEMENT

This research was supported by 2022 Youth Project of Anhui Provincial Philosophy and Social Science Planning: Research on the governance innovation of public service of sports for the elderly in rural communities in Anhui Province (Project No.: AHSKQ2022D132).

DATA AVAILABILITY STATEMENT

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

REFERENCES

- [1] T. E. Andersen, M. W. Fagerland, and B. Clarsen, "Assessing the cumulative effect of long-term training load on the risk of injury in team sports," *BMJ Open Sport and Exercise Medicine*, vol. 8, no. 2, 2022, pp.1-20.
- [2] W. B. Lian, B. H. Liu, and X. Q. Liu, "The backstepping control of high-speed train speed tracking based on rbf neural network,"

- International Journal of Electrical Engineering: Transactions of the Chinese Institute of Engineers, Series E, vol. 28, no. 1, 2021, pp.35-42.
- [3] Z. Niu, P. Zhang, Y. Cui, and Z. Jun, "PID control of an omnidirectional mobile platform based on an RBF neural network controller," *Industrial Robot*, vol. 49, no. 1, 2022, pp.65-75.
- [4] H. Wang, X. Zhou, and Y. Tian, "Robust adaptive fault-tolerant control using RBF-based neural network for a rigid-flexible robotic system with unknown control direction," *International Journal of Robust and Nonlinear Control*, vol. 32, no. 3, 2022, pp.1272-1302.
- [5] X. Lian, J. Zhang, L. Chang, J. Song, and J. Sun, "Test mass capture for drag-free satellite based on RBF neural network adaptive sliding mode control," *Advances in Space Research: The Official Journal of the Committee on Space Research (COSPAR)*, vol. 69, no. 2, 2022, pp.1205-1219.
- [6] L. Ye and P. Di, "Optimizing the regulation and control of sports injury and fatigue of winter olympic ice and snow athletes based on injury prevention," *Revista Brasileira de Medicina do Esporte*, vol. 27, no. 2, 2021, pp.79-82.
- [7] X. Wang and Y. Guo, "Attribute reduction algorithm based early warning model of sports injury," *Informatica*, vol. 45, no. 5, 2021, pp.561-657.
- [8] R. Bahr, B. Clarsen, W. Derman, and K. Chamari, "International Olympic Committee consensus statement: methods for recording and reporting of epidemiological data on injury and illness in sport 2020 (including STROBE Extension for Sport Injury and Illness Surveillance (STROBE-SIIS))," *British Journal of Sports Medicine*, vol. 54, no. 7, 2020, pp.372-389.
- [9] Y. Li, "Construction of intelligent campus tennis players' body data monitoring and injury warning system based on data fusion," *Revista Brasileira de Medicina do Esporte*, vol. 27, no. spe2, 2021, pp.46-49.
- [10] L. Chia, C. W. Fuller, D. Taylor, and E. Pappas, "Mastering the topic, the message, and the delivery: Leveraging the social marketing mix to better implement sports injury prevention programs," *The Journal of Orthopaedic and Sports Physical Therapy*, vol. 52, no. 2, 2022, pp.55-59.
- [11] D. Zhang, W. Feng, L. Wei, and X. Hu, "RBF neural network PID space vector control of linear servo load simulator," *Mechatronic Systems and Control (formerly Control and Intelligent Systems)*, vol. 48, no. 4, 2020, pp.207-215.
- [12] J. Zhang, H. Sun, Y. Qi, and S. Deng, "Temperature control strategy of incubator based on RBF Neural Network PID," *World Scientific Research Journal*, vol. 6, no. 1, 2020, pp.7-15.
- [13] W. Huang and Y. Yang, "Water quality sensor model based on an optimization method of RBF Neural Network," *Computational Water, Energy, and Environmental Engineering*, vol. 09, no. 1, 2020, pp.1-11.
- [14] X. Peng, H. Yu, X. Zhu, and Y. Li, "Electro-hydraulic proportional position control using auto disturbance rejection based on RBF Neural Network," *Journal of Beijing Institute of Technology*, vol. 30, no. 1, 2021, pp.121-128.
- [15] H. Liu, B. He, P. Qin, X. Zhang, S. Guo, and X. Mu, "Sea level anomaly intelligent inversion model based on LSTM-RBF network," *Meteorology and Atmospheric Physics*, vol. 133, no. 2, 2021, pp.245-259.
- [16] Q. Li, Q. Xiong, S. Ji, Y. Yu, and H. Yi, "A method for mixed data classification base on RBF-ELM network," *Neurocomputing*, vol. 431, no. 1, 2021, pp.7-22.
- [17] J. Ostadieh and M. C. Amirani, "Introducing the hybrid 'K-means, RLS' learning for the RBF network in obstructive apnea disease detection using Dual-tree complex wavelet transform based features," *Journal of Electrical Bioimpedance*, vol. 11, no. 1, 2020, pp.4-11.
- [18] C. Avinash, D. P. Loretta, Y. Heather, and E. Angelo, "Modeling time loss from sports-related injuries using random effects models: an illustration using soccer-related injury observations," *Journal of Quantitative Analysis in Sports*, vol. 16, no. 3, 2020, pp.221-235.
- [19] M. Edama, H. Inaba, F. Hoshino, S. Natsui, and G. Omori, "The relationship between the female athlete triad and injury rates in collegiate female athletes," *PeerJ*, vol. 9, no. 4, 2021, pp.1-12.
- [20] C. J. Xie and L. Tian, "Application of Nano-Composites in the recovery of sports ligament injury," *OMICS International*, vol. 83, no. 4, 2021, pp.235-242.
- [21] J. Sabol, C. Kane, M. P. Wilhelm, J. C. Reneker, and M. B. Donaldson, "The comparative mental health responses between post-musculoskeletal injury and post-concussive injury among collegiate athletes: A systematic review," *International Journal of Sports Physical Therapy*, vol. 16, no. 1, 2021, pp.1-11.
- [22] J. Xi and H. Jiang, "Infrared multispectral radiation-temperature measurement based on RBF Network," *Infrared Technology*, vol. 42, no. 10, 2020, pp. 963-968.

Classification of Hand Movements Based on EMG Signals using Topological Features

Jiayang Li¹, Lei Yang², Yunan He³, Osamu Fukuda⁴

Mathematical Science Research Center, Chongqing University of Technology, Chongqing 400054, China^{1, 2, 3}
Graduate School of Science and Engineering, Saga University, Saga 840-8502, Japan⁴

Abstract—Hand movement classification based on Electromyography (EMG) signals has been extensively investigated in the past decades as a promising approach used for controlling upper prosthetics or robotics. Topological data analysis is a relatively new and increasingly popular tool in data science that uses mathematical techniques from topology to analyze and understand complex data sets. This paper proposes a method for classifying hand movements based on EMG signals using topological features crafted with the tools of TDA. The main findings of this work on hand movement EMG classification are as follows: (1) topological features are effective in classifying EMG signals and outperform other time domain features tested in the experiments; (2) the 0-th Betti numbers are more effective than the 1-st Betti numbers; (3) Betti amplitude is a more stable and powerful feature than other topological features discussed in this paper. Additionally, Betti curves were used to visualize topological patterns for hand movement EMG.

Keywords—EMG classification; persistent homology; topological features; betti curve

I. INTRODUCTION

Electromyography (EMG) is a bioelectrical signal generated by muscle cells, providing valuable information about muscle activity that can be used to recognize hand movements [1]. Over the past few decades, various techniques have been developed to discriminate hand movements from EMG signals [2], [3], [4], [5], and most follow a unified analysis pipeline that involves preprocessing, feature extraction, and classification. Feature extraction involves transforming raw data into a feature vector that is later fed into a classifier. The dimension of the feature vector is usually much smaller than the raw data, which helps to reduce redundant data and prevent learning from the curse of dimensionality [6], thus accelerating learning speed and generalization steps. Therefore, choosing the right features has a significant impact on classification performance.

Previous studies have evaluated the ability of various EMG features to recognize hand movements, which can be categorized into time and frequency domains. Time-domain features analyze EMG signals over time, and common ones include mean absolute value (MAV) and root mean square (RMS) [7]. Frequency-domain analysis involves measurements that describe specific aspects of the signal's frequency spectrum, with mean frequency and median frequency being two useful features [8]. In addition, there are also end-to-end learning models that try to learn a feature representation of raw signals instead of using hand-crafted features [9], [10]. These models map the initial input to the final outputs directly, making it difficult to classify features into a particular domain.

In this work, we introduce a new type of feature for hand movement classification called topological features. The primary method used in this study to extract topological features falls in the field of topological data analysis (TDA). TDA is an approach that extracts topological features and describes the geometric shapes of datasets using techniques from topology [11], [12], [13]. Topological invariants are flexible and independent of metrics and coordinates, which allows us to compare EMG data collected from various wearing manners of EMG sensors. Additionally, topological invariants characterize the overall features of datasets, making topology-based solutions less sensitive to noise and enabling the identification of EMG signal shapes despite countless deformations [14]. For a point set in a Euclidean space, we can obtain a filtration of simplicial complexes as the spatial scale changes. The filtration of simplicial complexes provides a multi-scale perspective to understand data. The most common constructions of the filtration of simplicial complexes are the Vietoris-Rips complex and the Čech complex [15], [16]. Persistent homology is proposed to measure the topological features of data that persist across all scales. Since topological features are detected over a wide range of spatial scales, it is more likely to find true patterns of shape behind the data rather than noise.

TDA has found applications across diverse fields, such as material science [17], [18], biomolecules [19], [20], oncology [21], sensor networks [22], and data science [23]. Numerous TDA-based methods have been proposed for time series analysis [24], [25], [26]. For example, Pereira and de Mello developed a time series clustering approach using topological features computed by persistent homology [27]. Khasawneh and Munch tracked the stability of stochastic dynamical systems with TDA [28]. Gidea and Katz detected market crashes with financial time series analysis using topological features [29]. Emrani et al. applied TDA for wheeze detection in breathing sounds [30]. TDA has also been employed in ECG signal analysis, with Ignacio et al. identifying Atrial Fibrillation using topological features [31] and Dlugas detecting arrhythmias with topological methods [32].

This study presents a novel and effective method for visualizing and recognizing hand movement EMG signals and provides guidance for selecting hyper-parameters. The main innovations of this work are as follows. First, our approach does not rely on a specific embedding dimension and delay to transform time series signals into point clouds for analysis. Instead, we explored various combinations of embedding dimensions and delays to determine the optimal values and understand the relationship between classification accuracy

and the embedding parameters. Secondly, a wide range of topological features (Wasserstein amplitude, Betti amplitude, landscape amplitude, and persistent entropy) were considered from the perspective of entropy and amplitude to classify EMG signals. The topological features derived from different dimensions of homology groups were considered separately and jointly. Finally, we pioneered the use of Betti curves to visualize the topological structure of EMG signals. Betti amplitude can be used as a metric to evaluate the difficulty in distinguishing two hand movements using EMG signals.

To the best of our knowledge, this is the first work that employs topological features for classifying hand movements based on EMG signals. Our experimental results show that the proposed topological features achieve higher accuracy than other time-domain features for EMG signal classification. The main contributions of this study are: (1) The information conveyed by the 0-th Betti numbers is found to be more effective than that conveyed by the 1-st Betti numbers in classifying EMG signals. Combining the 0-th and 1-st Betti numbers does not result in higher classification accuracy, but instead yields a lower accuracy than using only the 0-th Betti numbers. (2) Among the four topological features explored, Betti amplitude, which is the L_2 -metric between Betti curves, is stable and effective in classifying EMG signals, with its accuracy almost unaffected by the embedding dimension and embedding delay. (3) Inspired by the effectiveness of the 0-th Betti numbers in classifying EMG signals, we visualized EMG signals of 53 types of hand movements using 0-th Betti curves and clearly observed differences in the topological structure.

The rest of this paper is organized as follows: Section II introduces the general approach for classifying EMG signals using topological features, along with related concepts and tools. Section III delves into the experimental aspects, addressing topics such as hyperparameter selection, topological feature selection, and the effectiveness of topological features. Finally, Section IV summarizes the study and highlights potential directions for future research.

II. METHODS

This section will begin with a review of the fundamental concepts and methods of TDA that are relevant to this work. It will cover the concepts of simplex, simplicial complex, and persistent homology. Additionally, it will introduce the notion of a persistence diagram, which is one type of representation used in persistent homology. The latter part of this section will focus on the pipeline used to classify EMG signals. The pipeline involves transforming the time-series EMG signal into a metric space, followed by converting the metric space into a topological space. From this topological space, topological features are extracted and fed into classifiers.

A. Homology and Persistent Homology

The triangle is known to be one of the simplest geometric shapes in the plane, and we can combine triangles into a more complex shape. The simplex can be viewed as a generalization of the notion of the triangle in any dimension. Specifically, a p -simplex is a p -dimensional polytope which is the convex hull of its $p + 1$ geometrically independent vertices in the Euclidean space \mathbb{R}^n . For example, a 0-simplex is a point, a

1-simplex is a line segment, a 2-simplex is a triangle, a 3-simplex means a tetrahedron, etc. A simplicial complex K in the Euclidean space is a collection of simplices such that (1) each face of a simplex of K is a simplex; (2) the intersection of any two simplices of K is either empty or a common face of them. Simplicial complexes provide discrete representations for topological spaces.

Homology is one of the essential topological invariants for describing the intrinsic properties of spaces. In this work, we focus on the simplicial homology on simplicial complexes. Let K be a simplicial complex. Let \mathbb{F} be a field. Denote $C_p(K; \mathbb{F})$ the \mathbb{F} -linear space generated by the p -simplices of K . Then $C_*(K; \mathbb{F})$ is a chain complex with the boundary operator $\partial_p : C_p(K; \mathbb{F}) \rightarrow C_{p-1}(K; \mathbb{F})$ given by

$$\partial_p[v_0, v_1, \dots, v_p] = \sum_{i=0}^p (-1)^i [v_0, \dots, \hat{v}_i, \dots, v_p], \quad p \geq 1$$

for any simplex $[v_0, v_1, \dots, v_p]$ of K , where \hat{v}_i means omission of the term v_i . For $p = 0$, we denote $\partial_0 = 0$. Then the p -th homology group of K is defined by

$$H_p(K; \mathbb{F}) := \frac{\ker \partial_p}{\text{Im } \partial_{p+1}}, \quad p \geq 0.$$

The homology groups reflect the topological features of simplicial complexes. The 0-dimensional homology group detects the connected components of simplicial complexes, the 1-dimensional homology group detects the loops while higher dimensional homology groups detect higher dimensional voids or cavities. In essence, homology detects “holes” in a simplicial complex. Betti number, defined by $\beta_p = \dim H_p(K; \mathbb{F})$ is the usual topological invariant to describe the information of “holes”. The advantage of using Betti number to represent the topology pattern of data is that it is more intrinsic and more resistant to noise.

Persistent homology is the central method to detect the topological features and describe the geometric shapes of high-dimensional data in topological data analysis. The persistence is intended to focus on the multi-scale information of data sets. We build the persistent homology on data sets by a filtration of simplicial complexes. Given a collection of points in Euclidean space \mathbb{R}^n , the Vietoris-Rips complex \mathcal{R}_ϵ is the abstract simplicial complex whose p -simplices are the sets of $p + 1$ points which are pairwise within distance ϵ [15]. The Vietoris-Rips complex is the most frequently used filtration of complexes constructed from a point set. Let K be a simplicial complex equipped with a real-valued function $f : K \rightarrow \mathbb{R}$. Then we have a filtration of simplicial complex $\{K_\epsilon\}_\epsilon$ given by $K_\epsilon = \{\sigma \in K | f(\sigma) \leq \epsilon\}$. The (a, b) -persistent homology of K with respect to f is defined by

$$H_p^{a,b}(K; \mathbb{F}) := \text{Im}(H_p^a(K; \mathbb{F}) \rightarrow H_p^b(K; \mathbb{F})), \quad p \geq 0.$$

The (a, b) -persistent Betti number is given by $\beta_p^{a,b} = \dim H_p^{a,b}(K; \mathbb{F})$. There are two typical representations of persistent Betti numbers, the barcode, and the persistence diagram. The barcode and the persistence provide the visualization of the persistent homology. See Fig. 1 as an example. Consider a collection of points in a Euclidean plane \mathbb{R}^n . We obtain a filtration of simplicial complexes as the parameter ϵ grows. When $\epsilon = 0$, the Vietoris-Rips complex \mathcal{R}_0 is the

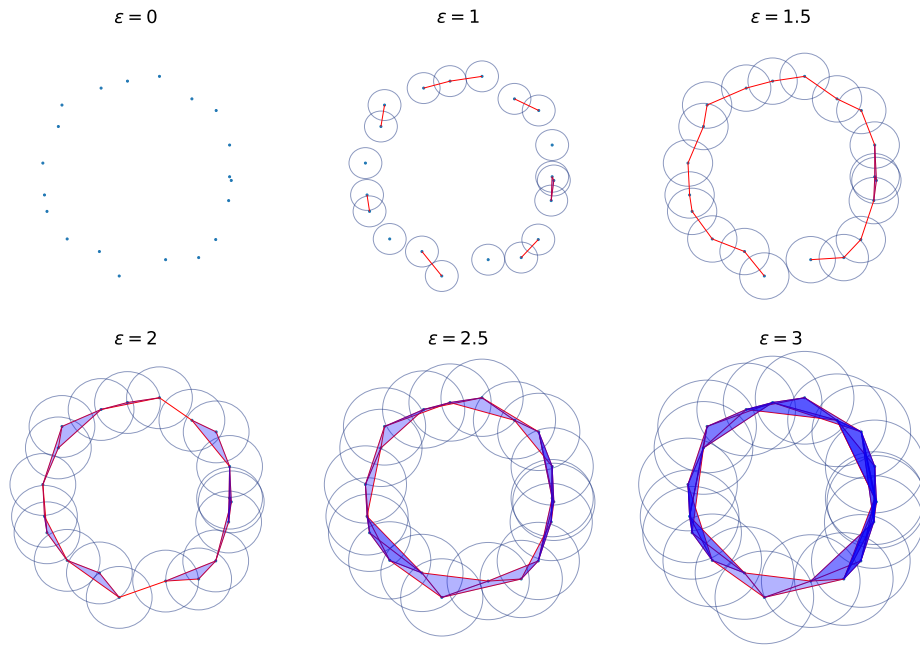


Fig. 1. A nested sequence of simplicial complexes as the radius increases.

simplicial complex of discrete points. By setting the sample points as the circle center and setting the scale parameter ϵ as the radius of a circle, the filtration of simplicial complexes can be constructed as follows: if two circles have an intersection, then add a line segment between the two points ($\epsilon = 1$). If three circles have intersections with each other, then add a triangle among the three points ($\epsilon = 2$). Thus simplices are added to the complex step by step, which generates a nested sequence of simplicial complexes as the radius increases. As the scale parameter grows, some generators come into existence at some parameters while some generators disappear at some parameters. The parameter that a generator α appear is called the birth time of α , and the parameter that α disappears is called the death time. The (birth, death) pairs are plotted in a diagram called persistence diagram. Persistence diagram provides a concise description of the topological changes over all scale parameters.

B. Topological Features

Although persistence diagrams are useful descriptors of data, they may not be suitable as input data for most machine learning models due to the absence of a natural linear structure. However, this limitation has been addressed through vectorizing persistence diagrams or using kernel methods. In this study, we extract four types of topological features by vectorizing persistence diagrams. These features have been successfully employed in other work for time series analysis. The four features can be categorized into two groups: entropy and amplitude. Entropy refers to persistent entropy, while amplitude includes Wasserstein amplitude, landscape amplitude, and Betti amplitude. The followings are the details of the topological features.

- (a) *Persistent entropy*: Persistent entropy is defined as the Shannon entropy of the persistence diagram. In general,

the lower the persistent entropy is, the simpler the shape of the data will be. Let D be a persistence diagram and $\alpha = (b_\alpha, d_\alpha)$ be a point in the diagram. Then, the persistent entropy of D is defined as:

$$E(D) := - \sum_{\alpha \in D} p_\alpha \log p_\alpha \quad (1)$$

where

$$p_\alpha = \frac{d_\alpha - b_\alpha}{\sum_{\alpha \in D} (d_\alpha - b_\alpha)} \quad (2)$$

- (b) *Amplitude*: The main idea behind amplitude is to partition a diagram into sub-diagrams based on homology dimension, and use a metric to measure the distance of each sub-diagram (or the derivative of each sub-diagram) with respect to the diagonal diagram. The diagonal diagram consists of only the diagonal line. Wasserstein amplitude computes the p -Wasserstein distance between sub-diagrams and the diagonal diagram, and it is defined as:

$$A_w = \frac{\sqrt{2}}{2} \sum_{\alpha \in D} ((d_\alpha - b_\alpha)^p)^{\frac{1}{p}} \quad (3)$$

Landscape amplitude computes the L_p distance between persistence landscapes derived from sub-diagrams and the persistence landscape derived from the diagonal diagram. Betti amplitude, on the other hand, computes the L_p distance between Betti curves derived from sub-diagrams and Betti curve derived from the diagonal diagram. Persistence landscapes and Betti curves are two other representations of topological signatures. In this study, we only consider the case of $p = 2$.

C. Time Delay Embedding

To compute persistent homology and extract topological features from the 1-dimensional time series EMG signals, they must first be represented in the form of a point cloud. This can be achieved using the method of time delay embedding, also known as Taken's embedding. Given a 1-dimensional time series $X(t)$, its time delay embedding can be described as a sequence of vectors in the following form:

$$X_i = [X(t_i), X(t_i + \tau), \dots, X(t_i + (d-1)\tau)] \in \mathbb{R}^d \quad (4)$$

where d is the embedding dimension and τ is the time delay. Each vector is treated as a point in a d -dimensional Euclidean space, and all vectors together constitute a point cloud. There are two main parameters that need to be set in time delay embedding: embedding dimension and time delay. The embedding dimension determines the dimension of the Euclidean space. A higher dimension means more information is embedded in one single point, but it is harder to find patterns from these points in a higher space. The embedding delay determines how long a term of memory is embedded in one single point. A longer term of memory means longer memory, but the resulting data points lose the short-term memory and local features. The selection of embedding dimension and delay determines the topology of the embedded point clouds and thus affects the classification results.

In most cases, EMG signals are collected from multiple electrodes, making them multivariate time series. To apply time delay embedding on multivariate signals, it is applied to each channel separately, and each channel of an EMG sequence corresponds to a point cloud. In our proposed method, we use the same embedding dimension and time delay for all channels. Considering a window of an n -channel EMG signal with the shape of $(m \times n)$, where m is the length of the window, we obtain n point clouds.

III. EXPERIMENTS

A. Dataset and Preprocessing Policy

The EMG dataset employed in our experiments is a publicly accessible dataset called NinaPro DB5 [33], which records muscle activity using two Thalmic Myo armbands. Each Myo armband is equipped with 8 electrodes, yielding a total of 16 channels of EMG signals collected. The dataset comprises 6 repetitions of 53 distinct movements (including rest) performed by 10 intact subjects. The Thalmic Myo already incorporates a 50 Hz notch filter, eliminating the need for additional filtering [34].

Before feeding the EMG signals into classification models, data preprocessing is necessary. To compare the results of our proposed method with other benchmarks, we must maintain consistency in data preprocessing. Thus, we followed the exact procedure outlined in [34]. This process involves dividing each detected repetition into 200-sample windows with an overlap of 100 samples. Subsequently, each window is labeled with its corresponding movement number. For training and testing dataset splitting, repetitions 1, 3, 4, and 6 were used for training, while repetitions 2 and 5 were designated for validation. Classification was performed on all 53 movements (including

rest). It is important to note that the rest movement's sample size is significantly larger than that of other movements, so it was reduced to avoid dataset imbalance. Python served as the programming language for conducting the experiment. We employed various Python packages, including Risper [35] for computing persistent homology and scikit-learn for building classifiers.

B. Selecting the Best Embedding Delay and Dimension

As outlined in Section II-C, we employed time delay embedding to transform time-series EMG signals into point clouds in Euclidean space. Time delay embedding involves two primary parameters: embedding dimension and delay. The selection of these parameters defines the topology of the embedded point clouds, which in turn influences the classification accuracy. To determine the optimal values for embedding dimension and delay in EMG classification, we generated point clouds for each channel by applying the same dimension and delay parameters, ranging from 2 to 10 and 1 to 9, respectively.

Subsequently, we calculated persistent homology on these point clouds and extracted persistence diagrams. The 0-th and 1-st Betti numbers were utilized to detect the number of connected components and the number of independent loops of simplicial complexes. This approach was chosen as computing higher-dimensional Betti numbers can be highly complex. Persistent entropy was then extracted as a feature descriptor from the persistence diagram. For a single channel, persistent entropy is a 2-dimensional vector in a plane, with coordinates corresponding to the 0 and 1-dimensional homology groups. As illustrated in Fig. 2, the two hypotheses being compared are labeled as H_0 and H_1 . The dataset used in this experiment contained recordings from 16 channels. When persistent entropy was computed for each channel, both H_0 and H_1 were produced, each 16-dimensional. Concatenating H_0 and H_1 resulted in a single 32-dimensional feature vector.

The EMG data was classified using persistent entropy of H_0 , H_1 , and their concatenation. Each feature was input into a random forest classifier separately. The classifiers' performance was then evaluated to identify which feature yielded the best results. Fig. 2 displays the classification outcomes. The heatmap reveals the following findings: (1) The persistent entropy of H_1 does not contribute any valuable information for EMG signal classification and exhibits poor performance. (2) Concatenating H_0 and H_1 does not significantly enhance classification accuracy. The highest classification accuracy achieved in our experiment was 71.86%, obtained with an embedding delay of 1 and a dimension of 6.

The optimal combination of embedding delay and dimension discovered in our previous experiment using persistent entropy and the random forest classifier may not be entirely convincing. Moreover, while the highest accuracy was attained at dimension 6, it is unclear whether this dimension is significantly better than dimensions 5 or 4, as the classification accuracies are only marginally different. To gain further insight into the effect of embedding delay and dimension on classification performance, we conducted a second experiment. Similar to the first experiment, we transformed the raw time-series data into point clouds for each channel using dimension

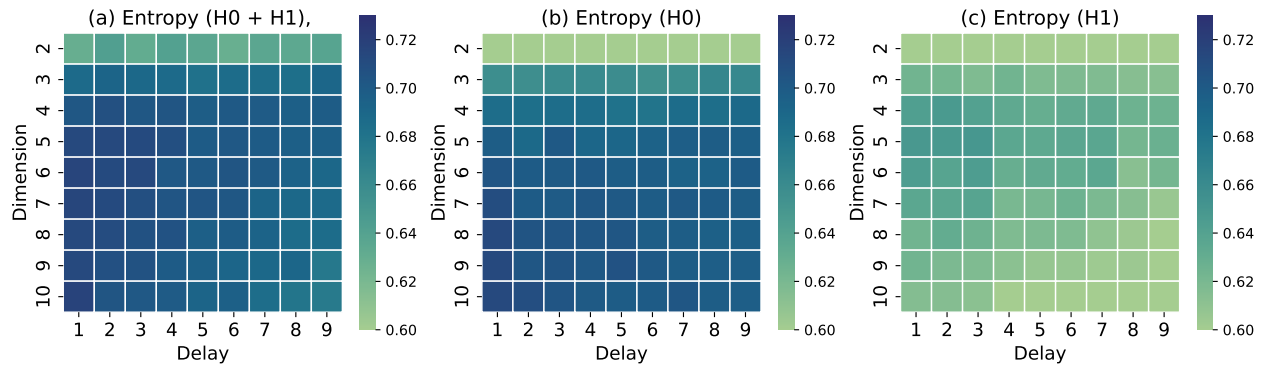


Fig. 2. Accuracy heatmap for various combinations of embedding delay and dimension.

and delay parameters selected from the range of 2 to 10 and 1 to 9, respectively, and computed persistent homology on these point clouds. We then extracted four types of topological features, namely persistent entropy, Wasserstein amplitude, landscape amplitude, and Betti amplitude. For each type of feature, we considered both H_0 , H_1 , and their concatenation. These features were then fed into both SVM and random forest classifiers for further analysis.

To explore the impact of embedding dimension on classification accuracy, we first calculated the mean accuracy for each embedding dimension across all delays and plotted the mean accuracy curve in Fig. 3. In the legend of the first subfigure, “entropy_h0_rf” denotes using persistent entropy of H_0 with a random forest classifier, while “entropy_rf” denotes using the concatenation of persistent entropy of H_0 and H_1 with a random forest classifier.

The figure reveals that: (1) The topological features of H_1 provide no meaningful information for EMG signal classification and perform poorly, consistent with the conclusion of the first experiment. The classification accuracy using the topological features of H_0 does not show a decreasing trend with any classifier, and even continues to increase beyond a dimension of 10 in the case of persistent entropy and Wasserstein amplitude. (2) Concatenating topological features of H_0 and H_1 generally does not result in a positive effect. The accuracy curve initially exhibits a rising trend, reaches a peak, and then declines. This occurs because H_1 carries less information and negatively impacts the concatenation of H_0 and H_1 . (3) Classification using Betti amplitude of H_0 proves effective and stable across all dimensions, while landscape amplitude is not a useful feature, as the landscape amplitude of H_0 , H_1 , and their concatenation did not yield acceptable accuracy.

To examine the effect of embedding delay on classification accuracy, we calculated the mean accuracy for each delay across all dimensions and plotted the results in Fig. 4. Our analysis reveals that: (1) For EMG signals, the classification accuracy using any topological feature (except landscape amplitude, which is not effective) of H_0 remains stable across all delays, indicating that embedding delay has only a minor influence on EMG signal classification. (2) When using the concatenation of topological features of H_0 and H_1 , accuracy decreases as delays increase. This can be attributed to the limited usefulness of the information carried by H_1 , which

can negatively impact H_0 . However, classification using Betti amplitude of H_0 remains stable across all delays.

Fig. 3 shows that the accuracy using topological features of H_0 has a trend of increasing even beyond a dimension of 10. To further investigate this trend, we examined dimensions ranging from 2 to 20, with an embedding delay of 1, as delay has only a minor influence on classification results. The classification results, shown in Fig. 5, reveal that: (1) Betti amplitude remains stable even at higher dimensions, without any significant decrease in accuracy. (2) For persistent entropy and Wasserstein amplitude, the increasing trend becomes less pronounced after dimension 10, and their accuracy converges to a maximum value around dimension 12. (3) The highest classification accuracy achieved is 73.93%, using Betti amplitude of H_0 at an embedding dimension of 3 and an embedding delay of 1 with SVM.

In summary, choosing the appropriate delay and dimension in time delay embedding and using Betti amplitude of H_0 as a topological feature can significantly enhance the accuracy of EMG signal classification. The main findings of the experiments are as follows:

- As the embedding dimension increases, the accuracy improves, eventually converging to a specific value. Delay has only minor effects on movement classification accuracy, so it can be simply set to 1. An exception is the case of Betti amplitude, where the classification model using Betti amplitude is robust to changes in dimension and delay. The optimal choices for delay and dimension for Betti amplitude are 1 and 3, respectively.
- H_0 contains more meaningful information than H_1 for EMG hand movement classification, regardless of the topological feature used. Combining the topological features of H_0 and H_1 has negative effects on H_0 and does not enhance accuracy.
- Betti amplitude of H_0 is the most effective and stable topological feature for classifying EMG signals, as it is robust to changes in dimension and delay and achieves the highest accuracy. Persistent entropy and Wasserstein amplitude are also good choices but require appropriate selection of embedding dimension and delay. Landscape amplitude is not an effective feature for EMG classification.

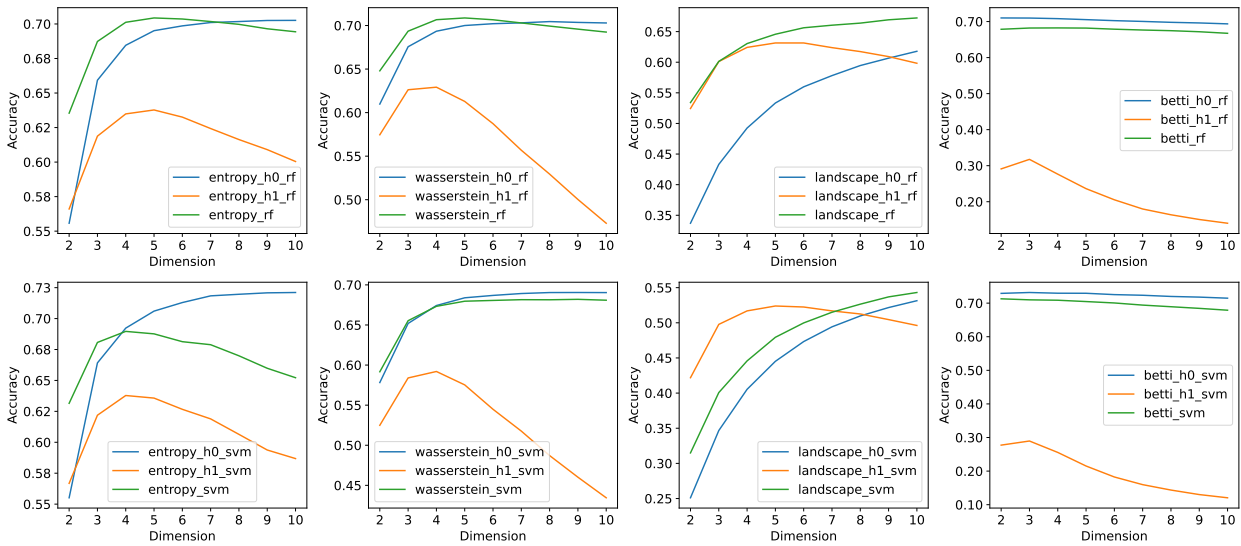


Fig. 3. The mean of the accuracy in each embedding dimension across all delays.

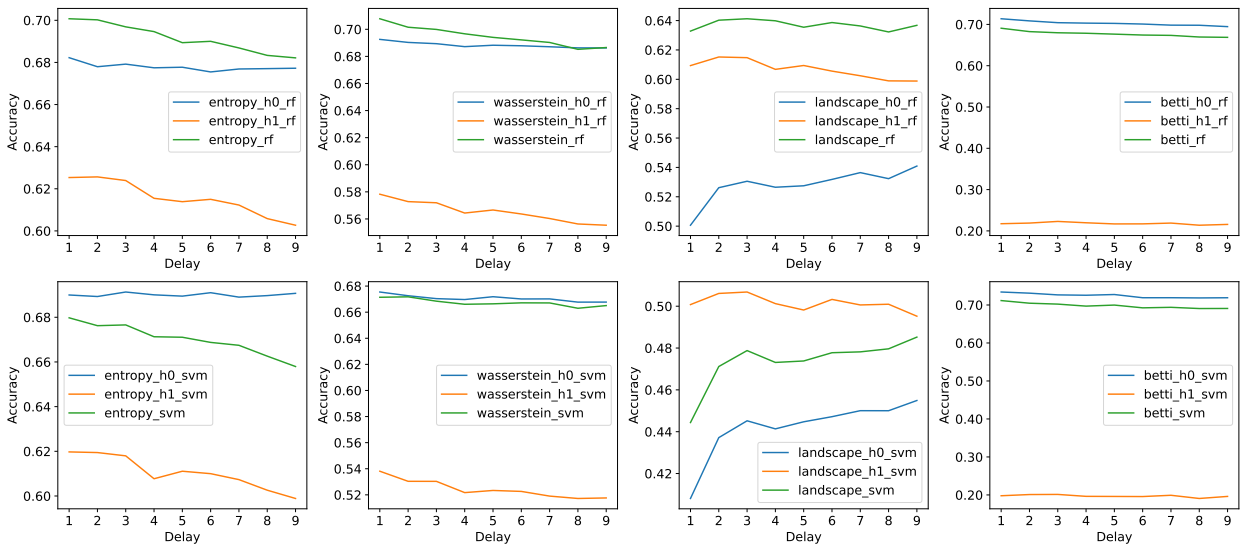


Fig. 4. The mean of the accuracy in each embedding delay across all dimensions.

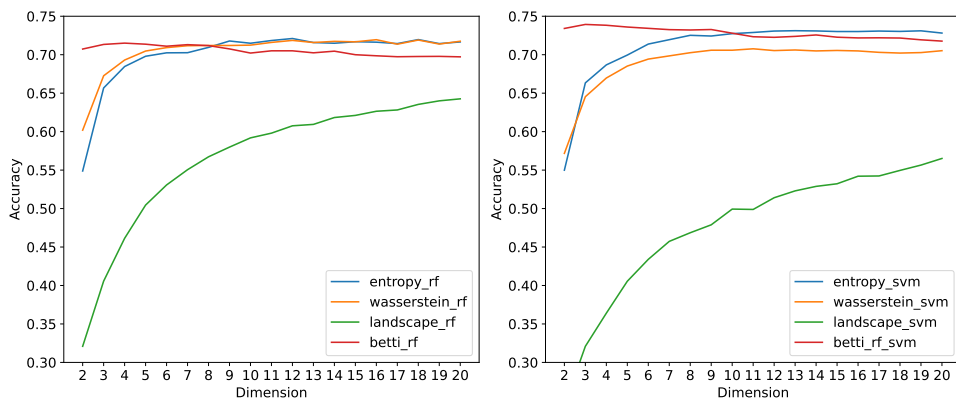


Fig. 5. The classification accuracy with embedding dimensions ranged from 2 to 20.

TABLE I. CLASSIFICATION ACCURACY USING TOPOLOGICAL FEATURES AND OTHER TIME DOMAIN FEATURES

Features		Classifier	Mean accuracy	Peak accuracy	
Topological features	Persistent entropy	SVM	73.02	73.12	
	Wasserstein amplitude		70.49	70.76	
	Landscape amplitude		56.32	56.51	
	Betti amplitude		73.50	73.93	
Time domain features	RMS (Root mean square)		70.36	70.63	
	TD (Time Domain Statistics)		68.19	68.48	
Topological features	Persistent entropy		Random forest	71.51	71.69
	Wasserstein amplitude			71.69	72.07
	Landscape amplitude	64.35		64.57	
	Betti amplitude	71.37		71.77	
Time domain features	RMS (Root mean square)	71.26		71.60	
	TD (Time Domain Statistics)	69.85		70.19	

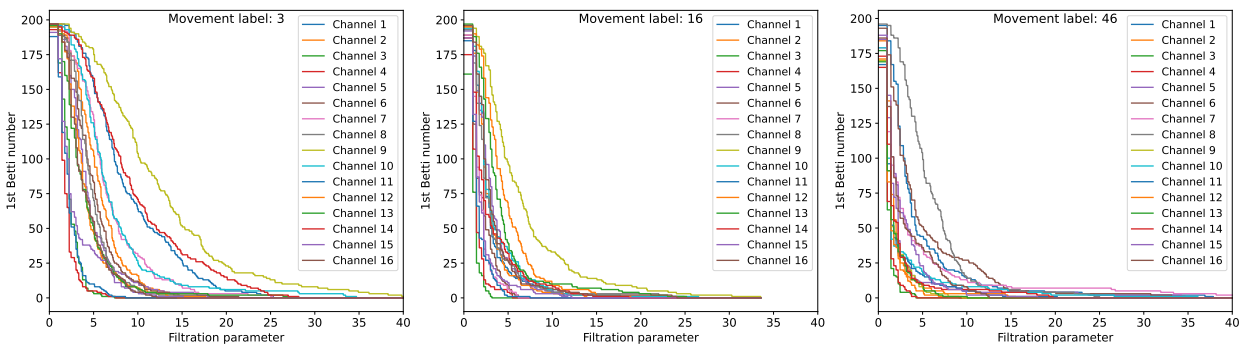


Fig. 6. Betti curves of one sample randomly selected from each of three types of hand movements, respectively.

C. Compare to Other Time-domain Features

To evaluate the performance of topological features relative to other feature types, we conducted an additional experiment using two time-domain features: Root Mean Square (RMS) and Time Domain Statistics (TS). RMS computes the root-mean-square value for each channel and constructs a 16-dimensional feature vector. In contrast, TS incorporates the mean absolute value, number of zeros, number of slope changes, and waveform length calculated from each channel [36]. We separately fed these two feature types, as well as the topological features with embedding delay and dimension determined based on the findings from previous experiments, into Support Vector Machines (SVM) and Random Forest for movement classification. We repeated the training process 10 times and recorded the classification accuracies. The mean and peak accuracy are displayed in Table I. Our results indicate that topological features, with the exception of landscape amplitude, outperform RMS and TS. Notably, the SVM model employing persistent entropy or Betti amplitude achieved a high accuracy of over 73%.

D. Visualize EMG Signals using Betti Curves

As previously mentioned, Betti amplitude is recognized for its robustness to changes in dimension and delay, and it has demonstrated the highest classification accuracy among other topological and time-domain features. This suggests that Betti amplitude may be better equipped to reveal the topology of EMG signals. However, Betti curve encodes even more topological information than Betti amplitude, as Betti ampli-

tude calculates only the L_2 distance between Betti curves. As a result, Betti curve can effectively differentiate topological patterns of distinct hand movements. This has inspired us to employ Betti curve for visualizing the topological patterns of various hand movements. For instance, we have plotted Betti curves of randomly selected samples from three types of hand movements in Fig. 6. It is important to note that only the 0-th Betti number, which counts the connected components in the topological space, is displayed in the figure, as we have demonstrated that only the information of H_0 is effective in classifying EMG signals. The figure clearly illustrates the distinct patterns of Betti curves in each channel for different hand movements.

To gain a broader understanding of the topological patterns for a specific type of hand movement, we calculated the average Betti curve by taking the arithmetic mean of Betti curves from all samples within that particular type of hand movement. This offers a more accurate representation of the general topological patterns for a given movement. By visualizing the average Betti curves of various hand movements in a single figure, we can easily observe the clear differences in topological patterns among different movements. This is demonstrated in Fig. 7, where the curve labeled “0” represents the “rest” movement, which is the fastest curve that drops to 0. This suggests that the topological space of “rest” has minimal topological complexity. Defining a distance metric on the average Betti curve could potentially aid in measuring the difficulty of discriminating a movement from EMG signals.

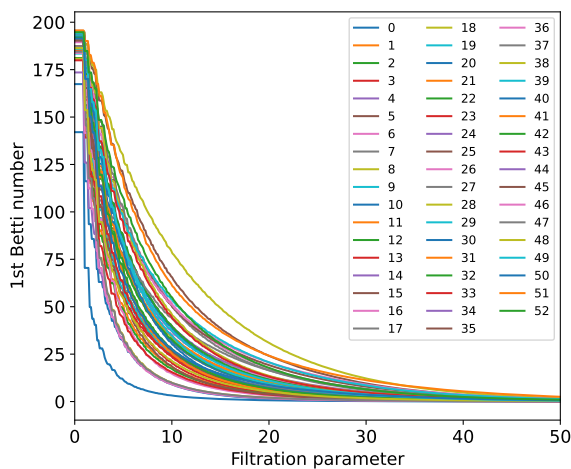


Fig. 7. The average Betti curves calculated for each of the 53 hand movements based on all training samples.

IV. CONCLUSIONS

The goal of this study was to determine whether topological features could enhance the accuracy of EMG signal classification for hand movements and contribute a methodology to the analysis of bio-electrical signals. The main findings of this study are as follows: (1) Topological features can effectively classify EMG signals, achieving the highest classification accuracy of 73.93%, outperforming the other tested time-domain features by nearly 2% in the experiment. (2) Topological features of H_0 prove more effective than those of H_1 . In general, higher embedding dimensions lead to increased accuracy, while embedding delay has a smaller impact on classification accuracy. (3) Among the four tested topological features, Betti amplitude is the most stable, and we have introduced Betti curves for visualizing the shape of hand movement EMG signals. Future research will explore whether these findings can be applied to other types of bio-electrical signals.

ACKNOWLEDGMENT

This research was funded by Scientific Research Foundation of Chongqing University of Technology and supported by the Science and Technology Research Program of Chongqing Municipal Education Commission (No. KJQN202101108).

REFERENCES

- [1] M. B. I. Reaz, M. S. Hussain, and F. Mohd-Yasin, "Techniques of emg signal analysis: detection, processing, classification and applications," *Biological procedures online*, vol. 8, pp. 11–35, 2006.
- [2] C. Sapsanis, G. Georgoulas, and A. Tzes, "Emg based classification of basic hand movements based on time-frequency features," in *21st Mediterranean conference on control and automation*. IEEE, 2013, pp. 716–722.
- [3] M. A. Aceves-Fernandez, J. Ramos-Arreguin, E. Gorrostieta-Hurtado, J. Pedraza-Ortega *et al.*, "Methodology proposal of emg hand movement classification based on cross recurrence plots," *Computational and mathematical methods in medicine*, vol. 2019, 2019.
- [4] O. Fukuda, T. Tsuji, M. Kaneko, and A. Otsuka, "A human-assisting manipulator teleoperated by emg signals and arm motions," *IEEE transactions on robotics and automation*, vol. 19, no. 2, pp. 210–222, 2003.

- [5] N. Rabin, M. Kahlon, S. Malayev, and A. Ratnovsky, "Classification of human hand movements based on emg signals using nonlinear dimensionality reduction and data fusion techniques," *Expert Systems with Applications*, vol. 149, p. 113281, 2020.
- [6] M. Köppen, "The curse of dimensionality," in *5th online world conference on soft computing in industrial applications (WSC5)*, vol. 1, 2000, pp. 4–8.
- [7] D. Tkach, H. Huang, and T. A. Kuiken, "Study of stability of time-domain features for electromyographic pattern recognition," *Journal of neuroengineering and rehabilitation*, vol. 7, no. 1, pp. 1–13, 2010.
- [8] A. Phinyomark, S. Thongpanja, H. Hu, P. Phukpattaranont, and C. Limsakul, "The usefulness of mean and median frequencies in electromyography analysis," *Computational intelligence in electromyography analysis-A perspective on current applications and future challenges*, vol. 81, p. 67, 2012.
- [9] U. Côté-Allard, C. L. Fall, A. Drouin, A. Campeau-Lecours, C. Gosselin, K. Glette, F. Laviolette, and B. Gosselin, "Deep learning for electromyographic hand gesture signal classification using transfer learning," *IEEE transactions on neural systems and rehabilitation engineering*, vol. 27, no. 4, pp. 760–771, 2019.
- [10] Y. He, O. Fukuda, N. Bu, H. Okumura, and N. Yamaguchi, "Surface emg pattern recognition using long short-term memory combined with multilayer perceptron," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2018, pp. 5636–5639.
- [11] G. Carlsson, A. Zomorodian, A. Collins, and L. J. Guibas, "Persistence barcodes for shapes," *International Journal of Shape Modeling*, vol. 11, no. 02, pp. 149–187, 2005.
- [12] A. Zomorodian and G. Carlsson, "Computing persistent homology," *Discrete & Computational Geometry*, vol. 33, no. 2, pp. 249–274, 2005.
- [13] G. Carlsson, "Topology and data," *Bulletin of the American Mathematical Society*, vol. 46, no. 2, pp. 255–308, 2009.
- [14] P. Y. Lum, G. Singh, A. Lehman, T. Ishkanov, M. Vejdemo-Johansson, M. Alagappan, J. Carlsson, and G. Carlsson, "Extracting insights from the shape of complex data using topology," *Scientific reports*, vol. 3, no. 1, pp. 1–8, 2013.
- [15] E. Carlsson, G. Carlsson, and V. De Silva, "An algebraic topological method for feature identification," *International Journal of Computational Geometry & Applications*, vol. 16, no. 04, pp. 291–314, 2006.
- [16] R. Ghrist, "Barcodes: the persistent topology of data," *Bulletin of the American Mathematical Society*, vol. 45, no. 1, pp. 61–75, 2008.
- [17] Y. Lee, S. D. Barthel, P. Dłotko, S. M. Moosavi, K. Hess, and B. Smit, "Quantifying similarity of pore-geometry in nanoporous materials," *Nature communications*, vol. 8, no. 1, pp. 1–8, 2017.
- [18] D. Chen, J. Liu, J. Wu, G.-W. Wei, F. Pan, and S.-T. Yau, "Path topology in molecular and materials sciences," *Journal of Physical Chemistry Letters*, vol. 14, no. 4, p. 954–964, 2023.
- [19] Z. Cang, L. Mu, and G.-W. Wei, "Representability of algebraic topology for biomolecules in machine learning based scoring and virtual screening," *PLoS computational biology*, vol. 14, no. 1, p. e1005929, 2018.
- [20] M. Wang, Z. Cang, and G.-W. Wei, "A topology-based network tree for the prediction of protein–protein binding affinity changes following mutation," *Nature Machine Intelligence*, vol. 2, no. 2, pp. 116–123, 2020.
- [21] A. Bukkuri, N. Andor, and I. K. Darcy, "Applications of topological data analysis in oncology," *Frontiers in artificial intelligence*, vol. 4, p. 659037, 2021.
- [22] V. De Silva, R. Ghrist *et al.*, "Homological sensor networks," *Notices of the American mathematical society*, vol. 54, no. 1, 2007.
- [23] V. Snášel, J. Nowaková, F. Xhafa, and L. Barolli, "Geometrical and topological approaches to big data," *Future Generation Computer Systems*, vol. 67, pp. 286–296, 2017.
- [24] N. Ravishanker and R. Chen, "Topological data analysis (tda) for time series," *arXiv preprint arXiv:1909.10604*, 2019.
- [25] S. Gholizadeh and W. Zadrozny, "A short survey of topological data analysis in time series and systems analysis," *arXiv preprint arXiv:1809.10745*, 2018.

- [26] Y. Umeda, J. Kaneko, and H. Kikuchi, "Topological data analysis and its application to time-series data analysis," *Fujitsu Scientific & Technical Journal*, vol. 55, no. 2, pp. 65–71, 2019.
- [27] C. M. Pereira and R. F. de Mello, "Persistent homology for time series and spatial data clustering," *Expert Systems with Applications*, vol. 42, no. 15-16, pp. 6026–6038, 2015.
- [28] F. A. Khasawneh and E. Munch, "Chatter detection in turning using persistent homology," *Mechanical Systems and Signal Processing*, vol. 70, pp. 527–541, 2016.
- [29] M. Gidea and Y. Katz, "Topological data analysis of financial time series: Landscapes of crashes," *Physica A: Statistical Mechanics and its Applications*, vol. 491, pp. 820–834, 2018.
- [30] S. Emrani, T. Gentimis, and H. Krim, "Persistent homology of delay embeddings and its application to wheeze detection," *IEEE Signal Processing Letters*, vol. 21, no. 4, pp. 459–463, 2014.
- [31] P. S. Ignacio, C. Dunstan, E. Escobar, L. Trujillo, and D. Uminsky, "Classification of single-lead electrocardiograms: Tda informed machine learning," in *2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA)*. IEEE, 2019, pp. 1241–1246.
- [32] H. Dlugas, "Electrocardiogram feature extraction and interval measurements using optimal representative cycles from persistent homology," *bioRxiv*, pp. 2022–02, 2022.
- [33] M. Atzori, A. Gijsberts, S. Heynen, A.-G. M. Hager, O. Deriaz, P. Van Der Smagt, C. Castellini, B. Caputo, and H. Müller, "Building the ninapro database: A resource for the biorobotics community," in *2012 4th IEEE RAS & EMBS International Conference on Biomedical Robotics and Biomechatronics (BioRob)*. IEEE, 2012, pp. 1258–1265.
- [34] S. Pizzolato, L. Tagliapietra, M. Cognolato, M. Reggiani, H. Müller, and M. Atzori, "Comparison of six electromyography acquisition setups on hand movement classification tasks," *PloS one*, vol. 12, no. 10, p. e0186132, 2017.
- [35] U. Bauer, "Ripser: efficient computation of vietoris–rips persistence barcodes," *Journal of Applied and Computational Topology*, vol. 5, no. 3, pp. 391–423, 2021.
- [36] B. Hudgins, P. Parker, and R. N. Scott, "A new strategy for multifunction myoelectric control," *IEEE transactions on biomedical engineering*, vol. 40, no. 1, pp. 82–94, 1993.

Research on Automatic Intrusion Detection Method of Software-Defined Security Services in Cloud Environment

Xingjie Huang*, Jing Li, Jinmeng Zhao, Beibei Su, Zixian Dong, Jing Zhang
State Grid Information and Telecommunication Branch, Beijing, 100053, China

Abstract—In a cloud environment, software defined security services are highly vulnerable to malicious virus attacks. In response to software security issues, this project plans to use machine learning technology to achieve automated detection of software security services in a cloud environment. Firstly, study the intrusion characteristics of software defined security services in cloud environments based on piecewise sample regression, and establish their statistical feature quantities. Then, using the method of decision statistical analysis, achieve its fixed identification. Finally, the intrusion characteristics of software defined security services in the cloud environment are studied and compared with the data in the cloud environment to obtain its power spectral density. On this basis, machine learning methods are used to extract features from software security services in the cloud environment, in order to achieve the goal of automatic extraction and optimization of software security services in the cloud environment. Through simulation experiments, the credibility of the proposed algorithm for software defined security services in the cloud environment was verified, and the attack characteristics of software defined security services in the cloud environment were effectively patched.

Keywords—Cloud environment; software; security services; invasion; detection; machine learning

I. INTRODUCTION

Software Defined Network (SDN) is a new network innovation architecture first proposed by the Clean-Slate research group of Stanford University in the United States. Its design idea is to separate the data layer from the control layer, in which the control layer is a programmable central controller, which can obtain global network information, status, etc., which is convenient for operators and researchers to manage and configure the network, and can deploy customized new protocols. The data layer includes routers, switches and other packet forwarding devices, which focus on providing simple data forwarding functions and can quickly process matching packets to meet the growing needs of users in the network. SDN has a wide range of application scenarios, such as big data analysis, cloud computing, Internet of Things, and car networking. However, like traditional networks, the new features brought by SDN architecture are also threatened by cyber-attacks [1].

Network intrusion detection technology refers to identifying and filtering abnormal traffic in the network through effective technical means, which is the basic method to ensure network security. Identifying abnormal traffic in time

and accurately can effectively reduce the impact of malicious attacks on the network and users [2]. At present, machine learning algorithm has attracted much attention in the field of network security, and intrusion detection technology based on machine learning has become a hot spot for researchers [3,4]. The study [5] proposes an intrusion detection system model by using different machine learning algorithms for different application scenarios. Autoencoder (AE) shows the advantages of the algorithm based on unsupervised learning in the task of feature extraction, feature reduction and classification, which promotes the progress of network attack traffic detection. The combination of automatic encoder and one-class support vector machine (OCSVM) algorithm can improve the anomaly detection rate and effectively reduce the training time of OCSVM algorithm, thus improving the performance of intrusion detection system. Therefore, it is of great theoretical and application value to study machine learning SDN intrusion detection technology based on automatic encoder and one-class support vector machine. Effectively detect the intrusion data of software-defined security service in cloud environment, and combine the statistical feature analysis and feature extraction methods of software-defined security service intrusion in cloud environment to fix the features and collect samples of software-defined security service intrusion in cloud environment. In research [6], an intrusion feature extraction method of software-defined security services in cloud environment based on blind balanced scheduling is proposed. Combined with the backtracking control method of intrusion nodes, the intrusion feature extraction of software-defined security services in cloud environment is realized, which improves the intrusion detection ability, but the optimization control ability of this method is not good, and the real-time performance of intrusion detection is not good. In study [7], an intrusion feature extraction method of software-defined security services in cloud environment based on spectral peak correlation search is proposed, which combines the autonomous positioning technology of intrusion nodes to realize intrusion feature extraction of software-defined security services in cloud environment, but the active optimization ability of this method is not strong. The most important feature of software-defined network is that it can program the network behavior, but the original software-defined network architecture is only limited to the programmability of the control plane, and the function of its forwarding plane is still limited by the fixed function hardware supported by the equipment provider. P4 can instantiate customized pipelines and stateful objects, support the implementation of complex

workflows, user-defined protocols and finite state machines, and use P4 switch to realize a reliable attack detection system to protect edge node resources from malicious network attacks, thus maximizing network utilization and effectively ensuring service quality [8].

Aiming at the above problems, this paper proposes a software-defined security service intrusion detection method in cloud environment based on machine learning algorithm. The statistical feature quantity of software-defined security service intrusion characteristics in cloud environment is constructed by using piecewise sample regression analysis method, and the intrusion characteristics of software-defined security service in cloud environment are fixed and identified by combining decision statistical analysis method. Combined with the feature fusion analysis method of classified intrusion samples, a big data distribution model of software-defined security service intrusion features in cloud environment is constructed, and the power spectral density feature quantity of software-defined security service intrusion features in cloud environment is extracted. The machine learning algorithm is used to adaptively optimize the extraction process of software-defined security service intrusion features in cloud environment, so as to realize automatic extraction and optimization of software-defined security service intrusion features in cloud environment. Finally, the simulation experiment analysis shows the superior performance of this method in improving the intrusion feature extraction ability of software-defined security services in cloud environment [9].

II. SOFTWARE-DEFINED SECURITY NETWORK INTRUSION NODE DISTRIBUTION MODEL AND INFORMATION SAMPLING

A. Cloud Software Defined Security Network Intrusion Distribution Structure Model

The data forwarding layer is located at the bottom of SDN architecture, and contains thousands of interconnected switches, which are responsible for forwarding data packets. If the switch is damaged, the message flowing through the switch will not be forwarded normally. In addition, the switch is the direct entrance for end users to access the network, and attackers can attack the switch by simply connecting to the switch port. Man-in-the-middle attack is a typical network intrusion method. Its main principle is to insert a proxy node between the source node and the target node, intercept the communication data, and tamper with the communication data without being discovered. The specific attack methods of man-in-the-middle attack include session hijacking, DNS spoofing and port mirroring. Man-in-the-middle attack between controller and switch is an ideal way to attack SDN network [10]. It can intercept and tamper with the forwarding rules of messages sent to switch, so as to control network forwarding. After that, attackers can carry out further attacks, such as black hole attacks. In addition, there may be no direct physical connection between the controller and the switch, that is, the data packet from the switch to the controller may pass through several other switches, so in the man-in-the-middle attack, all the switches and hosts directly connected to it on the communication path can easily be converted into proxy nodes. In order to realize the optimization of intrusion feature extraction of software-defined security service in cloud

environment, it is necessary to construct a distributed sensor model of software-defined security network in cloud environment under intrusion. Combined with the distributed design method of directed graph, blind forensics and intrusion node location are carried out for intrusion feature extraction of software-defined security service in cloud environment. Combined with feature distributed sampling technology [5], the intrusion node distribution model of software-defined security network nodes in cloud environment is constructed by using packet forwarding control technology, as shown in Fig. 1.

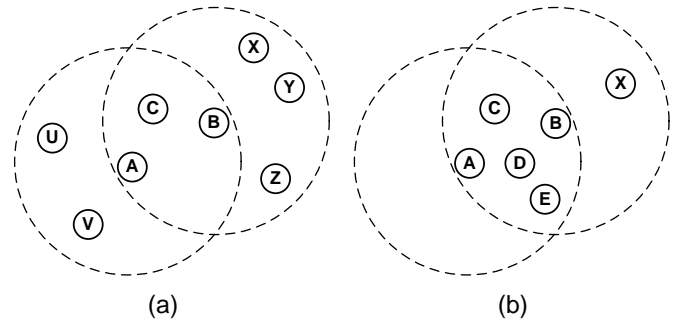


Fig. 1. Intrusion node distribution model of software-defined security network in cloud environment.

In the software-defined secure network node distribution model in the cloud environment shown in Fig. 1, when new data traffic in the network enters the OpenFlow switch for the first time, the OpenFlow switch will generate and process a Packet_In message to the control plane. By looking at the Packet_In message, the control plane will install the flow rules (dropping, forwarding and enqueueing) into the flow table and then send them to the OpenFlow switch. There is no authentication mechanism in this communication process if there is an incorrect traffic flow in the network [10].

Data plane refers to the interconnection infrastructure composed of hardware or software-based devices, which have basic packet forwarding functions, and these functions manage the received packets according to the flow rules set by the controller through the southbound interface. Each rule entry in the flow table consists of three fields: operation, counter and mode. The mode field defines the flow mode, and the flow mode is basically a collection of field values in the packet header. When a packet is received, the switch will search for a rule matching the field in its flow table. Once the switch matches such a rule, the counter of the rule will be incremented and the operation corresponding to the specific rule will be executed. Otherwise, the switch will inform the controller to ask for help or directly discard the packet.

The control plane is an independent and logically centralized server, also called the controller, which is used to handle the flow table installation of forwarding rules and monitor the status of data plane devices and links to integrate the global network view. Its main purpose is to manage the distributed forwarding devices of data plane in the network and provide operators with a simpler API, namely the northbound interface. SDN controller communicates with the switch through southbound API, such as OpenFlow protocol, and has a global view of the whole network topology. The originally

designed OpenFlow protocol promotes the decoupling of control plane and data plane. OpenFlow provides a southbound interface for the controller to establish a secure communication channel to manage forwarding device rules and receive status updates. These rules are the operations (such as forwarding, flooding and dropping) of the data packet determined according to the header information of the data packet (such as Ethernet /IP address and TCP/UDP port). The status includes the information discovered by the link layer, and the data packet and byte counters related to the flow table, which are used to support the global network view on the control plane [11].

The application plane is composed of a group of network applications, which provides users with quick response and various business requirements, such as network virtualization, topology discovery, traffic monitoring, security, load balancing and so on. The application communicates with the controller through the northbound API, such as RESTAPI, and the control layer provides the abstraction of physical network resources for the application layer, which enables the user to give the SDN controller other functions through the application and change the data flow rules without reconfiguring all the physical switches, so as to better manage and control the network behavior.

Under the network structure of SDN, equipment suppliers can focus on designing forwarding equipment for efficiently processing data packets, and network operators can easily test and deploy customized network management applications from a higher abstract level [12].

B. Software-defined Security Service Intrusion Information Sampling

Construct a node distribution model for extracting intrusion characteristics of software-defined security services in cloud environment, and use a binary directed graph to represent $G = (V, E)$, and the node's intrusion location is $\eta_n \in \Omega_\eta, n = 1, 2, \dots, N$. Construct a packet link forwarding protocol for network intrusion, and design a node directed graph model for network intrusion, and obtain a link forwarding control protocol for software-defined security services intrusion in cloud environment, which is expressed in the following form:

$$E(p) = [0, 0, \dots, 0, \underbrace{\frac{\gamma_{th} \sigma^2}{h_i [G - (n - k - 1) \gamma_{th}]}}_{i-1}, \dots, \frac{\gamma_{th} \sigma^2}{h_i [G - (n - k - 1) \gamma_{th}]}]^T \quad (1)$$

Set $f(p_i) = \frac{-L_i}{2} \log(1 + p_i)$, by calculating the cross-correlation coefficient between different categories of the hidden layer, according to the priority $E = E_1 \cup E_2 \cup E_3$ of connecting to the software-defined security service intrusion node in the cloud environment, and combining with the correlation detection method, obtain the statistical probability

distribution statistics of the software-defined security service intrusion neighbor node in the cloud environment as follows:

$$Vt(k) = \left\{ a_{s+t} \dots a_{t+1} a_t \dots a_1 \mid \overline{a_{s+t} \dots a_{t+1}} = k, a_i \in \{0, 1\}, 0 \leq k < 2^s \right\} \quad (2)$$

Assume $a_1, a_2 \in V$, $b_1, b_2 \in V'$, for Sink nodes $EHs(j)$ and $EHt(k)$ with software-defined security service intrusion feature distribution in the cloud environment, learn quickly and generalize the method, and get the associated feature quantity of intrusion feature detection as follows:

$$T_{l1} = \sqrt{F_{p1}^2 + F_{q1}^2} \quad (3)$$

According to the number of hidden layers of the structure, the feature validity of node intrusion is analyzed, and the reliability evaluation model of intrusion feature detection is

$F_{il} = \frac{1}{P_{il}}$, thus the feature link distribution model of software-defined security service intrusion nodes in cloud environment is constructed. The hidden layer and the output layer form a complete feature chain, which is expressed as $W(p) = G_T p^2 - Cp + \alpha T$, which $W(p)$ is a quadratic function of the software-defined security service intrusion feature link set SD in cloud environment. Combined with the decision statistical analysis method, the intrusion feature of software-defined security service in cloud environment is fixed and sample identified, which improves the effectiveness of feature extraction.

C. Information Sampling of Intrusion Characteristics

A big data distribution model of software-defined security service intrusion characteristics in the cloud environment is constructed by combining the classified intrusion sample feature fusion analysis method, and the power spectral density of software-defined security service intrusion characteristics in the cloud environment is extracted, and the statistics of blind forensics judgment of software-defined security service intrusion characteristics in the cloud environment are obtained as follows:

$$\mu(n) = \begin{cases} \beta_1 [1 - \exp(-\alpha_1 |e_{MCMA}(n)|^2)], & E[(|e(n)|^2)] > K \\ \beta_2 [1 - \exp(-\alpha_2 |e_{MCMA}(n)|^2)], & \dots \end{cases} \quad (4)$$

Wherein, K is the decision threshold for successful attack $\alpha_1 > 0$, $\alpha_2 > 0$, λ_{max} is the intrusion decision thresholds for software-defined security network in cloud

environment, $0 < \beta_1 < \frac{1}{\lambda_{max}}$, $0 < \beta_2 < \frac{1}{\lambda_{max}}$. When the root mean square error of one-step attack in each variable set is

satisfied as $MSE = E[(|e(n)|^2)] > K$, take smaller α_2

and β_2 to sample the characteristic information of intrusion characteristics. The fuzzy decision algorithm for sampling the characteristic information of intrusion characteristics is described as follows:

```

ROUTE_2      (Route       $u = u_{s+1} \dots u_{t+1} u_t \dots u_1 0$ 
 $v = v_{s+t} \dots v_{t+1} u_t \dots u_1 0$ )
{
 $x = u_{s+1} \dots u_{t+1}$ ,  $y = v_{s+t} \dots v_{t+1}$ ,  $I(x, y) = \emptyset$ ;
For each  $e_i$ , if ( $u_i \neq v_i$ )  $I(x, y) = I(x, y) + e_i$ ;
While ( $I(x, y) \neq \emptyset$ )  $e_i = \text{firstselect}(I(x, y))$ ; //
Set the success probability of feature extraction.

 $y = y_{n-1} y_{n-2} \dots y_0$  //

Forwarding control protocol for variable sets form  $x$  to
 $x + e_i$ ;  $x = x + e_i$ ;  $I(x, y) = I(x, y) - e_i$ ;
}

```

According to the above algorithm design process, an intrusion detection system ML-SDNIDS based on machine learning is designed under the SDN network architecture. The overall architecture is mainly divided into two parts: the control plane and the data plane. The control plane is mainly responsible for the feature processing of the data set, the algorithm training of the intrusion detection model, the creation of the flow table and the centralized management of the data plane. The data plane is composed of multiple P4 switches connected with each other, and is mainly responsible for malicious traffic detection and packet forwarding decision. The control plane uses the existing IDS data set in the network, inputs the data set into the feature extraction module, and then the feature extraction module processes the data packet to effectively select the original features. Then the feature mapping module selects the features with strong correlation, and maps different features to the automatic encoder network. The automatic encoder calculates the root mean square error (RMSE) between the new data and the fitting data, and then uses OCSVM classification algorithm to classify the RMSE to realize anomaly detection.

III. OPTIMIZATION OF INTRUSION DETECTION FOR SOFTWARE-DEFINED SECURITY SERVICES

A. Fixed Handling of Intrusion Characteristics of Software-defined Security Services

Intrusion detection system mainly includes three modules: information collection, data analysis and emergency response. The information collection module is the foundation of the whole system, which mainly collects host log information, network segment protocol datagram information and user behavior status. The data analysis module is the core module of

the whole system, which mainly uses some means such as statistics, pattern matching and anomaly detection algorithm to quickly analyze the data collected in the information collection stage and judge whether there is abnormal behavior in network activities. The emergency response module is to take corresponding protection measures in time to prevent further damage when intrusion is detected in the previous stage.

The control plane uses the existing IDS data set in the network, inputs the data set into the feature extraction module, then the feature extraction module processes the data packet, selects the effective features of the original features, then the feature mapping module selects the features with strong correlation, maps different features to the automatic encoder network, and the automatic encoder calculates the root mean square error (RMSE) between the new data and the fitting data. Then, RMSE is classified by OCSVM classification algorithm to realize anomaly detection, and the intrusion features of software-defined security services in cloud environment are fixed and optimized. This paper proposes an intrusion detection method of software-defined security services in cloud environment based on machine learning algorithm. The statistical analysis model is adopted to construct the feature spatial distribution structure model of software-defined security service intrusion characteristics in cloud environment, and the nonlinear time series of software-defined security service intrusion characteristics in cloud environment is obtained as follows:

$$x_i(t) = x_i^1(t) + x_i^2(t) + x_i^3(t) \quad (5)$$

Wherein,

$$x_i^1(t) = \sum_{k=1}^p \varphi_{k0} x_i(t-k) - \sum_{k=1}^q \theta_{k0} \varepsilon_i(t-k) + \varepsilon_i(t) \quad (6)$$

$$x_i^2(t) = \sum_{k=1}^p \sum_{l=1}^2 \varphi_{kl} [w_{i1}^l, \dots, w_{im}^l] [x_1(t-k), \dots, x_n(t-k)]^T \quad (7)$$

$$x_i^3(t) = - \sum_{k=1}^q \sum_{l=1}^2 \theta_{kl} [w_{i1}^l, \dots, w_{im}^l] [\varepsilon_1(t-k), \dots, \varepsilon_n(t-k)]^T \quad (8)$$

When the intruder attacks the target network, combining with the piecewise linear test method, the statistical characteristics of software-defined security service intrusion detection in cloud environment are obtained as follows:

$$DS = \{(x_0, t_0), (x_1, t_1), \dots, (x_i, t_i), \dots\} \quad (9)$$

Combining machine learning and genetic evolution methods, the feature set of software-defined security service intrusion features in cloud environment is automatically clustered, and the fuzzy fit degree of software-defined security service intrusion feature detection in cloud environment is obtained as follows:

$$\begin{aligned}
O(LOF_k(p)) &= O(\text{Ird}_k(p)) + O(N_{k-\text{dist}(p)}) \\
&\quad + O(\text{Ird}_k(o_i \in N_{k-\text{dist}(p)})) \\
&= O(m * n)
\end{aligned} \quad (10)$$

In the spatial distribution area of software-defined security service intrusion characteristics in cloud environment, the adaptive scheduling of software-defined security service intrusion characteristics in cloud environment is carried out by combining genetic evolution and statistical feature analysis methods.

B. Automatic Extraction and Optimization of Intrusion Characteristics of Software-defined Security Services in Cloud Environment

Combined with the decision statistical analysis method, the intrusion features of software-defined security services in cloud environment are fixed and identified, and the process of extracting intrusion features of software-defined security services in cloud environment is adaptively optimized by machine learning algorithm to realize automatic extraction and optimization of intrusion features of software-defined security services in cloud environment. The implementation steps are described as follows:

Step 1: Constructing a nonlinear feature sequence distribution set $k, k = 1, 2 \dots, p$ and an information weighting coefficient $w_{i1}^l, \dots, w_{in}^l$ of software-defined security service intrusion features in a cloud environment, and performing node initialization operation to meet the requirements.

Step 2: Calculate the spectral density of the cluster head node of the software-defined security service intrusion feature chain in the cloud environment, and calculate the formula of information sampling time delay in the storage space of the software-defined security service intrusion feature in the cloud

$$\tau^*(p) = \frac{\theta(p.d, q.d)}{1 + \alpha \cdot \delta(p.l, q.l)}$$

environment. construct the first route detection protocol under network intrusion.

Step 3: Machine learning algorithm is used for optimization control, and the coverage point set of intrusion feature distribution is obtained.

Step 4: Reconstruct the intrusion characteristics of software-defined security services in cloud environment in the dimension distributed feature space, and get the structural mapping output of the feature chain.

$$\Omega_i(t) = \frac{\gamma_{th} \sigma^2}{h_i [G - (N(l) - 1) \gamma_{th}]}$$

Step 5: Using the fuzzy optimization control method, the iterative formula of intrusion feature extraction is

$$p_i(l+1) = \min(p_{max}, \Omega_i(l+1))$$

. If the software defines the security service intrusion feature storage spectrum

component gain value $h_i \neq h_{min}(l)$ and $\Omega_i(l) > 0$ in the cloud environment, the partial derivative of intrusion feature location output is obtained:

$$\frac{\partial u_i}{\partial p_i} = \frac{G h_i}{\sum_{j \neq i} h_j p_j + \sigma^2} \left(\frac{1}{1 + \gamma_i} - \beta_{c_1} \right) \tag{11}$$

Step 6: According to the improved machine learning algorithm, the adaptive iteration is carried out until the convergence criterion is met, the coverage of intrusion feature distribution of software-defined security services in cloud environment is calculated, and the intrusion feature is extracted according to the coverage, and the end is over.

The implementation process of extracting and fixing intrusion features of software-defined security services in cloud environment is shown in Fig. 2.

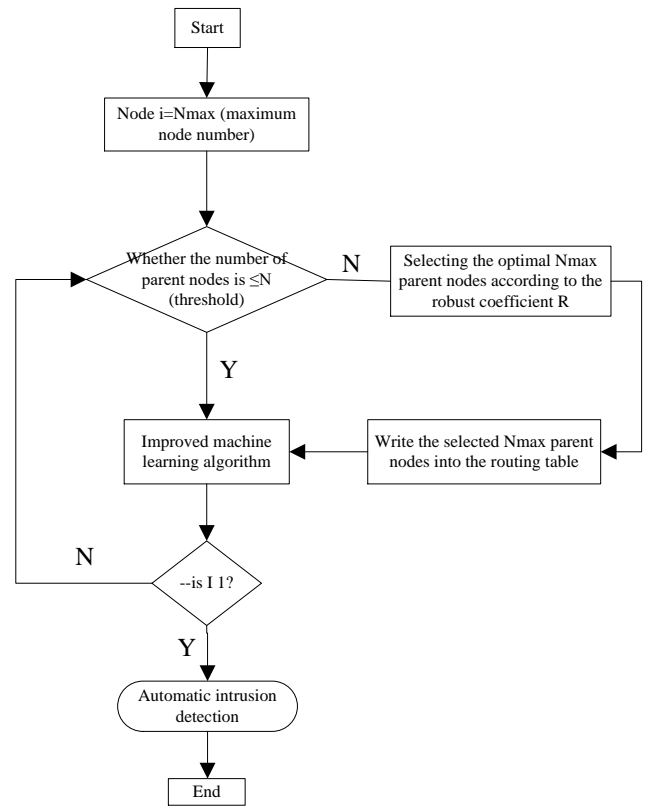


Fig. 2. Flow chart of network intrusion detection implementation.

IV. SIMULATION EXPERIMENT ANALYSIS

In Matlab 7 simulation environment, the simulation experiment of automatic extraction of intrusion characteristics of software-defined security services in cloud environment is carried out. Four P4 switches and four terminal hosts are adopted. The four switches are connected with each other to form a data forwarding plane, switch s1 is connected with hosts h1 and h2, and switch s2 is connected with hosts h3 and h4, and the hosts can communicate with each other. The configuration file describes in detail the network information configuration of each host (such as IP address, MAC address and default gateway), the configuration file path of P4 switch, and the connection path between the host and the switch. CIC-IDS2018 data set is dedicated to the analysis, testing and evaluation of network-based intrusion detection systems. The

data set is generated based on the creation of a user information file, which contains abstract expressions of events and user behaviors browsed by users in the network. The data set contains seven different attack scenarios: Brute-force, Heartbleed, Botnet, DenialofService(DoS), Distributed Denialofservice (DDOS), Webattacks and Infiltration of Modern WorkFrominside. The basic attack facility includes 50 computers as attack nodes, and the attack target includes 420 computers and 30 servers. The final data set includes the captured network traffic and system logs of each machine, and 80 features extracted from the captured traffic using CICFlowMeter-V3. The running environment of ML-SDNIDS is shown in Fig. 3.

```
p4@ubuntu:~/P4/tutorials/ML-SDNIDS/P4-NIDS$ make run
mkdir -p build pcap_logs
p4c-bm2-ss --p4v 16 -p4runtime-files build/basic_nids.p4.p4info.txt -o build/basic_nids.json basic_nids.p4
sudo python ../utils/run_exercise.py -t pod-topo/topology.json -j build/basic_nids.json -b simple_switch_grpc
Reading topology file.
Building mininet topology.
Configuring switch s3 using P4Runtime with file pod-topo/s3-runtime.json
- Using P4Info file build/basic_nids.p4.p4info.txt...
- Connecting to P4Runtime server on 127.0.0.1:50053 (bmw2)...
- Setting pipeline config (build/basic_nids.json)...
- Inserting 8 table entries...
- MyEgress.swid: (default action) => MyEgress.set_swid(swid=3)
- MyIngress.ipv4_lpm: (default action) => MyIngress.drop()
- MyIngress.ipv4_lpm: hdr.ipv4.dstAddr=['10.0.1.1', 32] => MyIngress.ipv4_forward(dstAddr=08:00:00:00:01:00, port=1)
- MyIngress.ipv4_lpm: hdr.ipv4.dstAddr=['10.0.2.2', 32] => MyIngress.ipv4_forward(dstAddr=08:00:00:00:01:00, port=1)
- MyIngress.ipv4_lpm: hdr.ipv4.dstAddr=['10.0.3.3', 32] => MyIngress.ipv4_forward(dstAddr=08:00:00:00:02:00, port=2)
- MyIngress.ipv4_lpm: hdr.ipv4.dstAddr=['10.0.4.4', 32] => MyIngress.ipv4_forward(dstAddr=08:00:00:00:02:00, port=2)
```

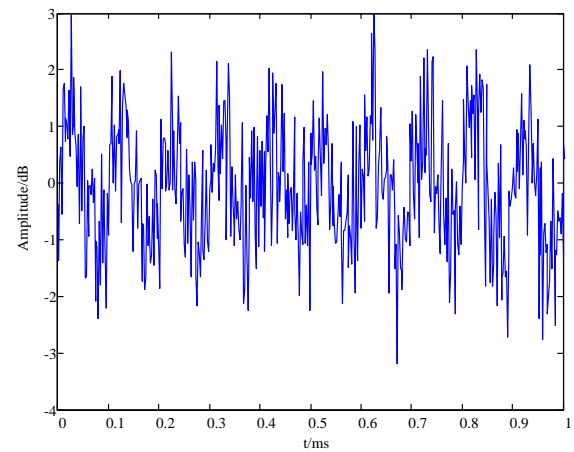
Fig. 3. ML-SDNIDS operating environment.

The fundamental frequency of software-defined security service intrusion feature collection in cloud environment is 20KHz, the coverage range of software-defined security service intrusion feature in cloud environment is 200*400, the spectral feature detection of software-defined security service intrusion feature in cloud environment is set to 12, the normalized termination frequency of information collection is 0.68Hz, and the modulation frequency of intrusion information detection varies between [120Hz and 1024Hz]. The intrusion characteristics of software-defined security services in cloud environment are extracted with signal-to-noise ratios of -5dB, 5dB and 20dB, respectively. CIC-IDS2018 data set is used. Due to the limited experimental environment, some data sets in CIC-IDS2018 total data set are selected for this test. The statistical information is as follows, including the total number of samples, the number of normal samples and the number of abnormal samples in each sub-data set. According to the above simulation environment and parameter settings, the intrusion characteristics of software-defined security services in cloud environment are extracted, and the original intrusion data collection is shown in Fig. 4.

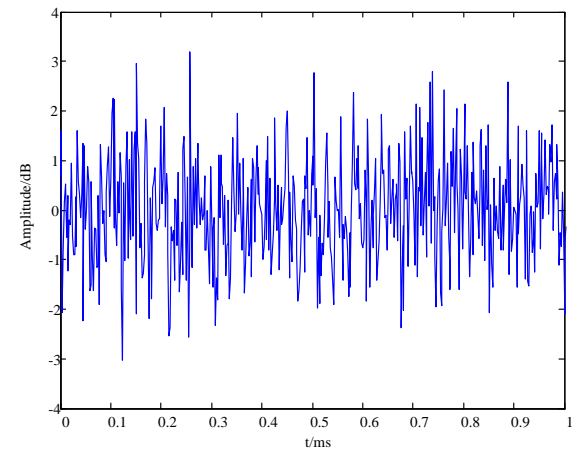
Taking the data collected in Fig. 4 as input, the intrusion feature extraction of software-defined security service in cloud environment is carried out, and the extraction result is shown in Fig. 5.

From the analysis of Fig. 5, it is known that the intrusion feature extraction of software-defined security services in cloud environment by this method has strong detection ability for the distribution of intrusion areas. The accuracy of different methods for intrusion feature extraction of software-defined

security services in cloud environment is tested, and the comparison results are shown in Table I. From the analysis of Table I, it is known that the intrusion feature extraction of software-defined security services in cloud environment by this method has high accuracy and good detection performance.



(a) Test sample.



(b) Training sample.

Fig. 4. Time domain waveform of software-defined security service intrusion characteristic sampling data.

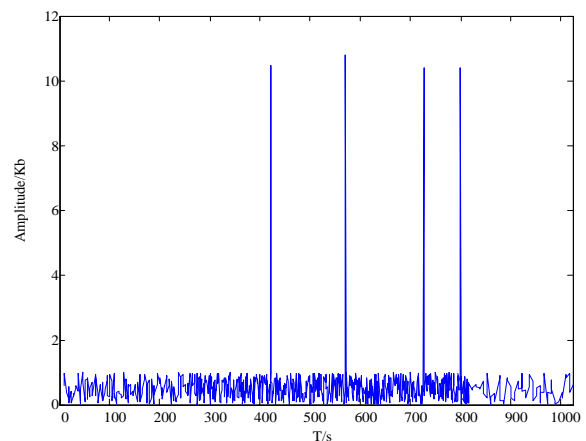


Fig. 5. Extraction results of intrusion characteristics of software-defined security services in cloud environment.

TABLE I. COMPARISON OF DETECTION PERFORMANCE

Iterations	The method in this paper	Neural network detection	Block matching detection
200	0.914	0.846	0.826
400	0.955	0.889	0.892
600	0.992	0.923	0.926
800	1	0.944	0.965

V. CONCLUSIONS

In this paper, the intrusion data of software-defined security service in cloud environment is effectively detected, combined with the statistical feature analysis and feature extraction method of software-defined security service intrusion in cloud environment, the feature fixation and sample collection of software-defined security service intrusion in cloud environment are carried out, and the software-defined security service intrusion forensics in cloud environment is realized according to the sample categories of software-defined security service intrusion in cloud environment. In this paper, an intrusion detection method of software-defined security service in cloud environment based on machine learning algorithm is proposed. The statistical feature quantity of software-defined security service intrusion characteristics in cloud environment is constructed by piecewise sample regression analysis method, and the intrusion characteristics of software-defined security service in cloud environment are fixed and sample identified by combining decision statistical analysis method. Combined with the feature fusion analysis method of classified intrusion samples, a big data distribution model of software-defined security service intrusion features in cloud environment is constructed, and the power spectral density feature quantity of software-defined security service intrusion features in cloud environment is extracted. The machine learning algorithm is used to adaptively optimize the extraction process of software-defined security service intrusion features in cloud environment, so as to realize automatic extraction and optimization of software-defined security service intrusion features in cloud environment. The research shows that this method has high accuracy in extracting intrusion features of software-defined security services in cloud environment, good reliability in network intrusion detection and strong ability in fixing intrusion features.

Finally, under the software-defined network structure, the control plane realized the intrusion detection model based on the combination of automatic encoder and support vector machine, and the data plane realized the intrusion detection system based on machine learning with P4 programming language. Feature extraction is carried out on the data plane, and the packet features are classified in the matching action

pipeline, and finally the decision classification of the packet is realized in the export pipeline. The final experimental results show that, in most cases, the accuracy of attack detection is higher than that of ninety-seven percent. Although the packet processing delay is increased by about five times, its efficiency is still millisecond.

ACKNOWLEDGMENT

The study was supported by Science and Technology Project of The State Grid Information and Telecommunication Branch "Research and design of key technologies for security intelligent detection and automation arrangement in a heterogeneous cloud environment" (NO. 529939220003).

REFERENCES

- [1] YANG Jianxi, ZHANG Yuanli, JIANG Hua, ZHU Xiaochen. Detection method of physical-layer impersonation attack based on deep Q-network in edge computing. *Journal of Computer Applications*, 2020, 40(11):3229-3235.
- [2] EBTEHAJ I, BONAKDARI H, ES-HAGHI M S. Design of a hybrid ANFIS-PSO model to estimate sediment transport in open channels. *Iranian Journal of Science and Technology-Transactions of Civil Engineering*, 2019, 43(4):851-857.
- [3] GHASEMI M, AKBARI E, RAHIMNEJAD A, et al. Phasor particle swarm optimization:a simple and efficient variant of PSO. *Soft Computing*, 2019, 23(19):9701-9718.
- [4] KHALILI A, SAMI A. SysDetect: a systematic approach to critical state determination for industrial intrusion detection systems using Apriori algorithm. *Journal of Process Control*, 2015, 2776:154-160.
- [5] Mernik M, Liu S H, Karaboga M D, et al. On clarifying misconceptions when comparing variants of the Artificial Bee Colony Algorithm by offering a new implementation. *Information Sciences*, 2015, 291(10):115-127.
- [6] MORADI M, KEYVANPOUR M R. An analytical review of XML association rules mining. *Artificial Intelligence Review*, 2015, 43(2):277-300.
- [7] BACH S H, BROECHELER M, HUANG B, et al. Hinge-loss Markov random fields and probabilistic soft logic. *Journal of Machine Learning Research*, 2017, 18:1-67.
- [8] FARNADI G, BACH S H, MOENS M F, et al. Soft quantification in statistical relational learning. *Machine Learning*, 2017, 106(12):1971-1991.
- [9] YUAN Chi. Identity-based dynamic clustering authentication algorithm for wireless sensor networks. *Journal of Computer Applications*, 2020, 40(11):3236-3241.
- [10] Yongmin LIU, Yujin YANG, Haoyi LUO, et al. Intrusion detection method for wireless sensor network based on bidirectional circulation generative adversarial network. *Journal of Computer Applications*, 2023, 43(1):160-168.
- [11] MOZAFFARI M, SAAD W, BENNIS M, et al. A tutorial on UAVs for wireless networks: applications, challenges, and open problems. *IEEE Communications Surveys & Tutorials*, 2019, 21(3):2334-2360.
- [12] YUAN X, HE P, ZHU Q, et al. Adversarial examples: attacks and defenses for deep learning. *IEEE Transactions on Neural Networks and Learning Systems*, 2019, 30(9):2805-2824.

Experimental Analysis of WebHDFS API Throughput

Yordan Kalmukov, Milko Marinov

Department of Computer Systems and Technologies, University of Ruse, Ruse, Bulgaria

Abstract—Data analysis is very important for the success of any business today. It helps to optimize business processes, analyze users' behavior, demands etc. There are powerful data analytics tools, such as the ones of the Hadoop ecosystem, but they require multiple high-performance servers to run and high-qualified experts to install, configure and support them. In most cases, small companies and start-ups could not afford such expenses. However, they can use them as web services, on demand, and pay much lower fees per request. To do that, companies should somehow share their data with an existing, already deployed, Hadoop cluster. The most common way of uploading their files to the Hadoop's Distributed File System (HDFS) is through the WebHDFS API (Application Programming Interface) that allows remote access to HDFS. For that reason, the API's throughput is very important for the efficient integration of a company's data to the Hadoop cluster. This paper performs a series of experimental analyses aiming to determine the WebHDFS API's throughput, if it is a bottleneck in integration of a company's data to existing Hadoop infrastructure and to detect all possible factors that influence the speed of data transmission between the clients' software and the Hadoop' file system.

Keywords—WebHDFS API; throughput analysis; data analytical tools; Hadoop Distributed File System (HDFS)

I. INTRODUCTION

Data analysis is a key point for success of any business organization. It allows companies to extract knowledge from the data they gather, to optimize their business processes and operations, to predict future failovers and to determine the right moment of maintenance. In general, users and clients are source of enormous amount of data. Companies can use these data to analyze users' behavior [1] and anticipate their demands. So, the analysis of any type of data can provide a significant competitive advantage of the company over rival businesses.

However, data analysis is a complex, time-consuming and computationally intensive task. Data analytics tools are usually either expensive or requires multiple high performance servers to run together with highly qualified IT experts to install and support them. This, of course, is not affordable for small and especially start-up companies. Fortunately, they can still use big data analysis tools, like the ones provided by the Hadoop's ecosystem, by hiring them as services "on demand". The "on demand" business model is a modern trend for hiring web-based (cloud-based) services and paying per request, rather than buying own expensive software and hardware. It provides maximum scalability and flexibility. According to it, computing resources are always available on the Internet and a company can use as many resources as it needs at the moment, while paying as much as it consumes.

Using remote data analytics services however, requires that the company share its data with the service provider [2]. In case big data analytics tools, being provided, are part of the Hadoop ecosystem, they will read the data from the "Hadoop Distributed File System" (HDFS). Although there are several ways to copy data to remote HDFS [3], the most preferable one is through the WebHDFS Application Programming Interface (API). It allows third-party applications to connect to remote HDFS file system and write/read files to/from it. As we are considering big data analysis, the amount of data being transferred is supposed to be large enough, so the WebHDFS API's throughput plays an important role in integration of a company's data to the Hadoop service provider. Other ways of data integration are reviewed in [4],[5],[6],[7],[8] and [9].

The aim of this work is to perform a series of experimental analyses to determine the WebHDFS API's throughput; if it is a bottleneck in integration of a company's data to existing Hadoop infrastructure; and to detect all possible external factors that influences the speed of data transmission between the clients' software and the WebHDFS API of existing Hadoop cluster.

The paper is structured as follows: Section II reviews some previous work done by other researchers. Section III describes the experimental system's architecture and the experimental setup in details. Section IV analyzes and discusses obtained experimental results. Finally, Section V ends the article with a conclusion, outlining and summarizing all key observations authors noticed during the experimental analysis.

II. RELATED WORK

HDFS allows management of large volumes of data using commodity items. This reinforces the need to provide robust data protection to facilitate file sharing in Hadoop, as well as having a trusted mechanism to verify the authenticity of shared files. This is the focus of [10], where the authors' attention is directed to improving the security of HDFS using a blockchain-enabled approach (hereafter referred to as BlockHDFS). User connects to the WebHDFS REST API, through which all data retrieval and modifications are implemented. In BlockHDFS, the blockchain is responsible for storing the file metadata. The costs incurred in storing HDFS file metadata on the blockchain are twofold. First, the WebHDFS API must read a file's metadata from HDFS as a hash value. Second, additional operations are required to store the metadata in the blockchain. However, since the metadata size is typically small, such overhead will neither introduce high latency for HDFS operations nor require a large amount of disk space for blockchain storage. The paper proposes a new approach to introduce blockchain (and more specifically, Hyperledger) to improve the security of the HDFS ecosystem.

Paper [11] discusses one of the most significant challenges of next-generation big data federation platforms, namely the access control in Hadoop systems. The paper critically analyzes and explores security limitations in Hadoop systems and presents a tool called “Big Data federation access broker” to address eight major Hadoop security limitations. To validate the performance of the broker, authors have conducted a set of experimental studies on a real Hadoop cluster. They made a comparison between read and write operations, performed through WebHDFS, in two cases - without using any security measures (pure WebHDFS access) and when using the authors' proposed broker model for accessing the big data. Performance analysis of operations executed over WebHDFS with files of various sizes was done as well.

Authors of [12] discuss the design of a data transfer service, called Stargate, to address the challenges of large data transfers over a WAN. Stargate implements a content-addressable protocol and multi-layered caching to cope with these challenges. It uses a novel approach that localizes computation, cache, and transfers to achieve efficient data access in cluster computing. Stargate is evaluated experimentally by comparing its performance with two widely used Hadoop data access methods - DistCP and WebHDFS. DistCP is a built-in Hadoop data delivery tool. DistCP preorders data, while WebHDFS provides data access on-demand. The elapsed times of three Hadoop benchmarks that have different I/O workloads were compared to evaluate efficiency. Experiments show that Stargate over WAN has comparable performance to HDFS running on a LAN. It also has lower overhead than WebHDFS, which is widely used for remote access to data from Hadoop clusters.

Apache Spark uses a cluster of compute-optimized servers on which the execution modules run, and a cluster of servers, optimized for performing storage operations and hosting the HDFS data. However, the network transfer from the data warehouse to the computing cluster becomes a serious obstacle for big data processing. Near-data processing (NDP) is a concept that aims to ease the network load in such cases by offloading some of the computing tasks to the storage cluster. Rachuri et al. present an architecture and basic principles of implementation of an NDP system for Spark [13]. HDFS can be configured to add redundancy by copying the same blocks of files across multiple data nodes to improve fault tolerance. It also provides an API - WebHDFS. In the proposed implementation, the authors take advantage of the replication factor to increase the number of data nodes that can perform operations related to offloading the computational tasks and intercept the WebHDFS communication between the client and the data node to perform NDP operations. Simulation results and experiments conducted on the developed prototype show that SparkNDP can help reduce the execution time of Spark queries compared to both - the default approach of not directing any tasks to the repository, and the direct NDP approach to offloading all tasks to the repository.

High Performance Computing (HPC) and Big Data are two trends that are starting to converge. In this process, aspects of hardware architectures, system support, and programming paradigms are revisited from both perspectives. The authors of [14] present their experience on this path of convergence. They

propose a framework through which some of the programming problems, arising from such integration, are solved. An integrated environment has been developed that integrates: (1) COMPS, a programming environment for developing and running parallel applications for distributed infrastructures; (2) Lemonade, a data mining and analysis tool; and (3) HDFS, the most widely used distributed file system for big data. In order to implement the integration between COMPS and HDFS, aspects of the available techniques for communication between external applications, in particular those written in Java and Python, and HDFS are considered. HDFS provides interfaces through a direct Java API, a command-line interface (CLI), a REST API (WebHDFS), and a C API (libhdfs). The proposed solution provides processing of large data transfers, with access to low-level functions.

WebHDFS allows users to connect to HDFS from outside the Hadoop cluster, which is especially useful when an external application needs to load data into or out of HDFS or work with the data stored in HDFS. WebHDFS also supports (for all HDFS users) operations such as reading files, writing to files, creating directories, changing access permissions, renaming, etc. The WebHDFS API is used for two functions in [15]: 1) after server-side processing is complete, this data is stored in HDFS via the WebHDFS API; and 2) when the created final data for visualization in raw text format is requested by clients, the data is passed to them via the WebHDFS API.

A system architecture combining the “IP multimedia subsystem (IMS)” platform and the Hadoop system used in the distributed storage of the IMS service resources is proposed in [16]. The result is a manageable Hadoop-based data center for telecommunication service providers. Interoperability between different systems is achieved through RESTful web services. The WebHDFS API is used to allow services to interact with HDFS, while the Oozie Web Services API is used for the compute service. The conducted tests prove the availability, scalability, and reliability of the proposed system. Experimental results show that system performance is improved, especially in terms of disk space utilization and system throughput.

Although HDFS works well with medium-sized and large files, its performance seriously degrades in case of multiple very small files. To overcome this shortcoming, the authors of [17] propose a system to improve the performance of HDFS using a distributed full-text search system. By indexing each file's metadata, such as name, size, date, and description, files can be quickly accessed through efficient metadata searches. Additionally, by consolidating many small files into one large file to be stored with better space and I/O efficiency, the negative performance impacts caused by directly storing each small file separately are avoided.

HDFS is a widely used open-source scalable and reliable file management system designed as a general-purpose distributed file storage solution. WebHDFS is a service for accessing data stored and maintained in HDFS. It runs on all nodes in the Hadoop cluster and provides a REST interface for data access. Unlike other file systems or data transfer tools, WebHDFS detects the layout of data blocks stored in HDFS. Using this block information, clients can directly access the

HDFS node (data node) on which the data is stored. This not only reduces data access latency, but also provides load balancing of data access requests. This motivated the authors of [18],[19],[20] to investigate the performance of HDFS in remote data access.

III. EXPERIMENTAL SETUP

To study the throughput of the WebHDFS API, a testing client must be developed to access the interface in both read and write modes. The client, for the current experiment, has been implemented in the php programming language. It uses the open source library PHP-Hadoop-HDFS [21], implemented and maintained by Aleksandr Kuzmenko. It is a wrapping library that does not do any specific data processing, but just composes the necessary HTTP requests to access the WebHDFS API. The access itself is done through the cURL (client URL library) library [22], distributed together with the php interpreter. The WebHDFS API could be accessed without PHP-Hadoop-HDFS library, but it facilitates the access, since the library frees the programmer from having to know the WebHDFS API itself. Instead, the programmer should only know the methods that the library implements and their input arguments. The architecture of the experimental system is presented on Fig. 1.

When performing the experiments, several key parameters should be monitored: total time; upload speed; download speed; bytes uploaded; bytes downloaded; and response HTTP code. Fortunately, all of these, together with many more parameters, are measured by the cURL library itself.

The throughput of the WebHDFS API is not the only limitation factor. The network speed is important as well, even more important. Even if the API itself allows the transfer of hundreds of megabytes per second, if the user's Internet speed is slow, then the API's throughput does not matter at all. Therefore, experiments should be performed from different type of computer networks:

1) *Internet*- This is the most important experiment, as this is the most realistic scenario for accessing the Hadoop cluster. Most likely, the greatest limitation factor will be the Internet connection speed.

2) *Corporate LAN of the service provider*: This experiment is important in case of a large company, having multiple offices, maintaining its own Hadoop infrastructure and MAN network between the different locations. Experiments could be done at multiple network speeds - 1 gbit/s and 100 mbit/s seem to be the most realistic speeds for a company's MAN.

Although writing to HDFS is more important than reading, both operations will be tested. Writing is more important since company's data should be saved to HDFS, before being analyzed by the Hadoop's data analytical tools. So, the data flow direction in general will be from the company to the Hadoop cluster. However, reading is also useful.

Experiments have been conducted with small, medium-sized and large files which are generated with the Windows' fsutil application. They contain only zeros (i.e. no meaningful information). Since we are making performance experiments, the content of the files does not matter at all, but their exact size does. It is important that their size can be precisely controlled.

The Hadoop cluster consists of 1 name node (2 x Intel Xeon Silver 4110, 32 total threads, 64 GB RAM) and 9 data nodes (Intel Xeon E-2124, 16 GB RAM). Servers are connected through 24 Port Gigabit switch HPE OfficeConnect 1820.

The experimental application, developed in php, runs on a laptop computer (Intel i7-7500U, 12 GB RAM) for all experiments in all types of networks. Using the same laptop for all experiments is intentionally done in order to ignore the influence of the client's hardware.

The access from the client to the WebHDFS API is done through the:

- 1) *Internet* – the access is done from a laptop computer, connected to a home router. According to the subscription, the guaranteed Internet speed is 80 mbit/s.
- 2) *University of Ruse's campus network* – the access is done from the same laptop, connected to any point of the university's campus-wide 100 mbit/s network.
- 3) *1 gbit/s cluster's switch* – the laptop is connected directly to the 1-gigabit switch of the Hadoop cluster.

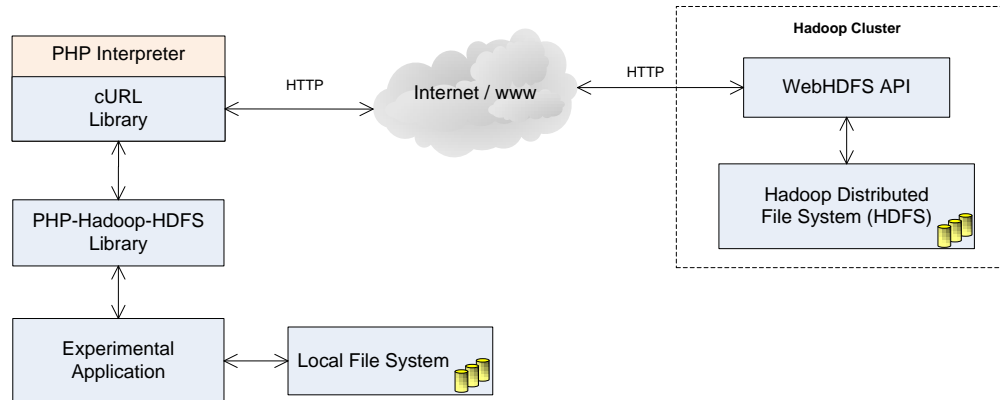


Fig. 1. Architecture of the system for experimental study of the WebHDFS API's throughput.

IV. EXPERIMENTAL RESULTS

A. Writing Data to HDFS

As previously mentioned, when integrating a company’s data to an existing Hadoop cluster, writing/saving files to WebHDFS is the most important operation. So, it is tested with priority.

Experiments started with small to medium-sized files from 10 to 100 MB, with a step of 10MB. Results are presented on Fig. 2. They show almost constant write speed when WebHDFS is accessed from the Internet and the university’s campus-wide 100 mb/s network. The transfer speed almost reaches the network’s capacity – with the 80 mbit/s Internet connection, the achieved speed is around 60 mbit/s, while within the 100 mbit/s campus-wide LAN, we achieve sustainable transfer of 85 mbit/s.

When the laptop is connected directly to the cluster’s gigabit switch, the writing speed is times higher and is increasing with the increase of the file size.

Since the API supports high-speed data transfer, we decided to go further and experiment with medium-sized files, from 100 to 300 MB with step of 50 MB, and large files from 500 to 1500 MB with step of 500 MB. Results are shown on Fig. 3 and 4 respectively. For larger files, upload speed becomes constant (about 800 mb/s) for the gigabit network as well. That proves the WebHDFS API supports very high writing speeds and could not be considered as bottleneck in the integration architecture. Most probably, the API just saves the incoming data to the HDFS file system without applying any complex processing on them.

As known from everyday usage of computer network and different types of file transfer, copying single large files is much more efficient than copying multiple smaller files. There are objective reasons for that, including metadata overheads. So, it is worth testing how much slower uploading multiple smaller files will be in respect to a single large file, having the same total size.

Three experiments have been performed with single large files of 100 MB, 200 MB and 300 MB, and 10 x 10 MB, 20 x 10 MB and 30 x 10 MB. The size of the single large file exactly matches the total sum of bytes of the respective many 10 MB files. Results are presented on Fig. 5. Expectedly, uploading a single large file is faster than uploading many smaller files, having the same total size as the large one.

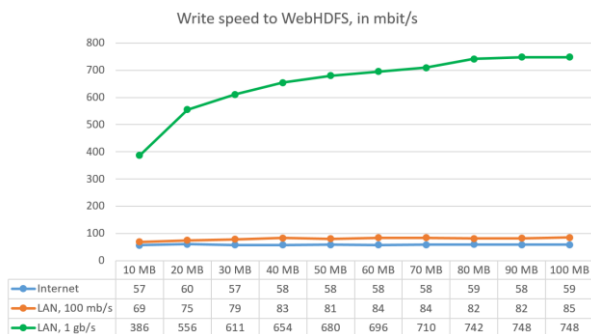


Fig. 2. Write speed to HDFS via the WebHDFS API with relatively small file sizes - from 10 MB to 100 MB.

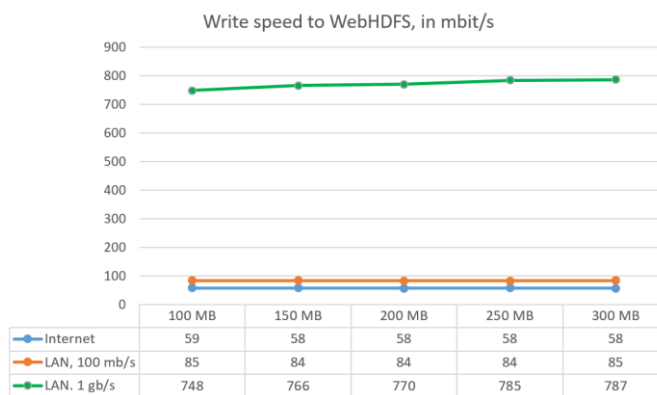


Fig. 3. Write speed to HDFS via the WebHDFS API. File sizes from 100 MB to 300 MB. Write speed is constant, independent on the file size.

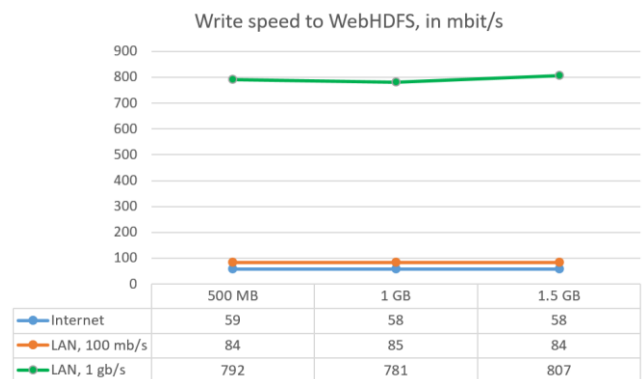


Fig. 4. Write speed to HDFS via the WebHDFS API. File sizes from 500 MB to 1500 MB. Write speed is still constant, although files have got very large.

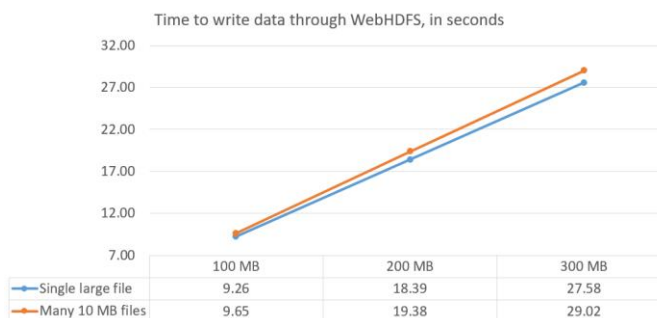


Fig. 5. Time to write data (single large file or multiple small files) to HDFS via the WebHDFS API.

B. Reading Data from HDFS

Although reading data from HDFS is less important operation when integrating a company’s data to existing Hadoop cluster, a series of experiments will be done by using the files, already uploaded to the HDFS through the WebHDFS API. The same client is used, as in the previous experiments for writing data, and runs on the same laptop computer as well. Results are shown on Fig. 6 (for files 10 to 100 MB), Fig. 7 (for files 100 to 300 MB) and Fig. 8 (for files 500 to 1500 MB).

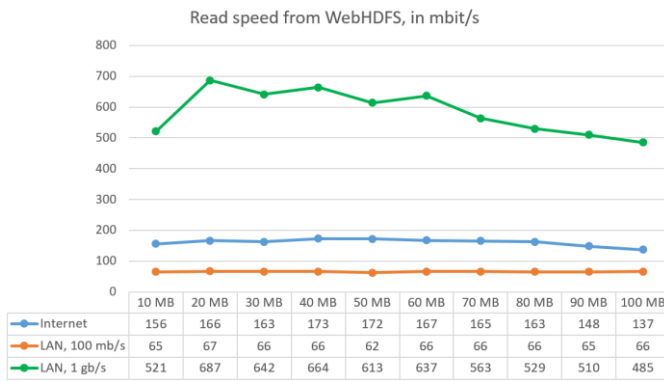


Fig. 6. Read speed from HDFS via the WebHDFS API with relatively small file sizes - from 10 MB to 100 MB.

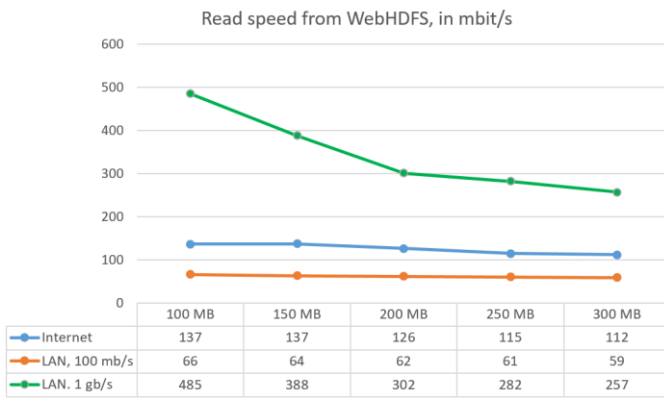


Fig. 7. Read speed from HDFS via the WebHDFS API. File sizes from 100 MB to 300 MB. On high-speed networks, read/download speed is decreasing with increasing the file size.

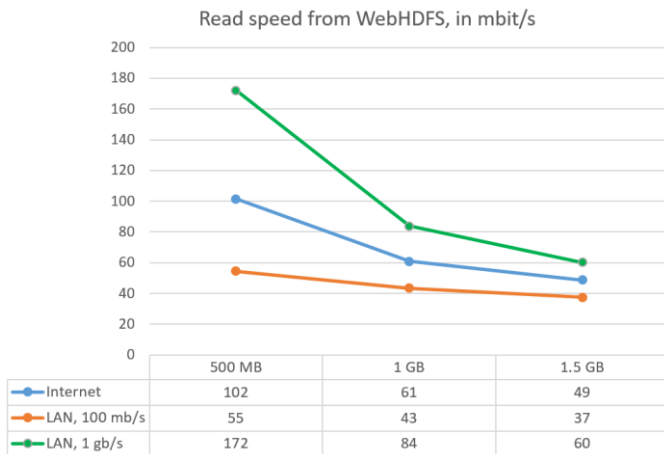


Fig. 8. Read speed from HDFS via the WebHDFS API. File sizes from 500 MB to 1500 MB. Read/download speed is decreasing with increasing the file size on all networks.

In contrast to the constant writing speed however, the reading (download) speed rapidly falls with increasing the file size – significantly noticeable for the high-speed 1 gbit/s connection. The initial suspicion/assumption was this decrease of the reading speed is related to the higher number of packets that the large files consist of. There is a sense in that – since larger files contains many more packets, more time may be

required to reconstruct the file from the higher number of packets. To test if the assumption is correct, the experiments have been repeated with measuring not just the total time, but transferring time and file saving time separately.

Results show that the transferring time, without the time needed to save the file within the local file system is commensurate with the total time. And time needed to reconstruct and save the file is actually very small and does not depend on the file size. So, the assumption is wrong. Since the communication between the client and the WebHDFS API is handled by the cURL library, distributed with the PHP interpreter, then the cause of the reading speed decrease could be either the cURL library itself or the API. To determine where the problem is, the WebHDFS API should be accessed in another way. It could be accessed directly through a browser, but it is not very convenient. Other tools like Postman or Rester are not very suitable as well, since they do not measure times. The cURL library however is not developed specially for PHP, but it is an open source project ported to almost any programming language. It is provided as a built-in application in Unix/Linux/macOS and could be downloaded as external stand-alone application for Windows.

We take the HTTP queries, generated by the PHP-Hadoop-HDFS library and run them from the cURL applications for MacOS and Windows. Results show that files are downloading (read) from HDFS with very high speed almost reaching the maximum throughput capacity of the relevant type of network, regardless of the file size. So, the causer of the read speed decrease in our experiments is determined to be the cURL library, distributed with the PHP interpreter. Further experiments will be done with different versions of PHP and its accompanying cURL library.

Another interesting and unexpected result occurred when reading files from WebHDFS over the Internet. The utilized home Internet subscription plan is guaranteeing speed of 80 mbit/s. However, reading was done at speeds up to 175 mbit/s. The connection between the home computer and the Internet Service Provider (ISP) is actually higher than the guaranteed 80mbit/s. Apparently, the ISP also has a high-speed connection to the university's network, where the Hadoop cluster is located. So, the access to the servers from the home computer is done in a kind of MAN network with speeds significantly higher than guaranteed Internet connection. Interestingly, this is not the case when writing data to WebHDFS. When writing (uploading) files, they are transferred at a speed no higher than the guaranteed Internet connection. This, however, is a specific case study related to the specific ISP and should not be considered as an essential part of the results of the experiments.

V. CONCLUSIONS

Since WebHDFS is the most preferable way for remote access of the distributed file system HDFS, it is important to know its capabilities, performance and throughput.

After performing dozens of experiments and additional analyses, the results could be summarized in the following conclusions:

1) WebHDFS API allows data exchange with the Hadoop Distributed File System (HDFS) at very high speeds, and in general it is not the limitation factor, but the speed of the network itself. In a 1 gbit/s network, we achieve data transfer rates of around 800 mbit/s. On slower networks, the network capacity is almost entirely compressed.

2) The speed of writing (uploading) files through the WebHDFS API does not depend on the size of the files, but remains constant and it is only limited by the network capacity.

3) In contrast to the constant write speed, the reading (download) speed through the WebHDFS API decreases rapidly as the file size increases.

4) However, the reason for the decreasing read speed is not in the WebHDFS API itself, but in the implementation of the cURL library, distributed together with the PHP interpreter. When accessing the API in alternative ways, the read speed remains constant and comparable to the write speed.

5) When reading files from the WebHDFS API by using PHP and cURL, it is mandatory that the PHP interpreter is configured to use a larger amount of RAM memory than the size of the files being read. This is expected since the data transfer happens in multiple small network packets, but in order to reconstruct the file from them, they must be stored and arranged in a common buffer (located within the RAM memory).

6) When writing files through the WebHDFS API, the amount of RAM memory that the PHP interpreter can work with is not of such importance because the data is read from the local file system in chunks, which are typically much smaller than the default memory limit of 128 MB. In this case, the buffer size is important on the receiving side – i.e. on the server where the WebHDFS API is running. This size is managed by Cloudera Manager and is large enough.

7) In relation to the above-mentioned, it has been observed that files larger than 2GB cannot be reliably written to HDFS via the WebHDFS API. For all of our attempts to upload a file larger than 2 GB, the server did not return any HTTP response, although in some cases the files were actually stored in HDFS. The fact that files of 1.99 GB were always reliably saved, but 2.0 GB were not, suggests that the reason might be a WebHDFS setting, maybe the server-side buffer in question or something else. Further analysis could help to determine the exact cause.

8) A large number of small files are transferred more slowly than a single large file of the same total size. This observation is absolutely expected given the fact that with many small files, many separate HTTP requests are made, each of which has time to resolve the domain to an IP address, time to connect, time to make and receive the request, etc.

The WebHDFS API allows data exchange with HDFS at very high speeds, so it could not be considered as a bottleneck in the integration of a company's data to an existing Hadoop cluster. However, by default, it does not perform any user access control by itself, so additional means should be designed and implemented, or integrated, to control user access

and guarantee data isolation (no third-party company should be able to access data of another company).

ACKNOWLEDGMENTS

This work is supported by the Science Fund of the “Angel Kanchev” University of Ruse and the European Regional Development Fund under grant BG05M2OP001-1.002-0002-C02 “Digitalization of Economy in Big Data Environment”.

REFERENCES

- [1] Ashayer, S. Yasrobi, S. Thomas and N. Tabrizi, "Performance Analysis of Hadoop Cluster for User Behavior Analysis", 2018 IEEE 20th International Conference on High Performance Computing and Communications; IEEE 16th International Conference on Smart City; IEEE 4th International Conference on Data Science and Systems (HPCC/SmartCity/DSS), Exeter, UK, 2018, pp. 805-809, doi 10.1109/HPCC/SmartCity/DSS.2018.00135.
- [2] Y. Kalmukov, M. Marinov. "Hadoop as a Service: Integration of a Company's Heterogeneous Data to a Remote Hadoop Infrastructure". International Journal of Advanced Computer Science and Applications, 13(4), pp. 49-55, 2022, DOI:10.14569/IJACSA.2022.0130406.
- [3] Y. Kalmukov, M. Marinov, T. Mladenova, I. Valova. "Analysis and Experimental Study of HDFS Performance". TEM Journal, 10(2), pp. 806-814, ISSN 2217-8309, DOI: 10.18421/TEM102-38, May 2021.
- [4] M. Sarnovsky, P. Bednar and M. Smatana, "Data integration in scalable data analytics platform for process industries", 2017 IEEE 21st International Conference on Intelligent Engineering Systems (INES), pp. 187-192, 2017.
- [5] J. Pokorný, "Integration of Relational and NoSQL Databases", Vietnam Journal of Computer Science, vol. 6, no. 4, pp. 389-405, 2019.
- [6] S. Ramzan, I.S. Bajwa, B. Ramzan, and W. Anwar, "Intelligent Data Engineering for Migration to NoSQL Based Secure Environments", IEEE Access, vol. 7, pp. 69042-69057, 2019.
- [7] Cholissodin, D. Seruni, J. Zulqornain, A. Hanafi, A. Ghofur, M. Alexander and M. Hasan, "Development of Big Data App for Classification based on Map Reduce of Naive Bayes with or without Web and Mobile Interface by RESTful API Using Hadoop and Spark", Journal of Information Technology and Computer Science, vol. 5(3), pp. 302–312, 2020.
- [8] Anilkumar and B. Shireesha, "A Study on Optimized Big Data Performance and its Industrial Development", Journal of Advanced Research in Technology and Management Sciences, vol. 1(1), pp. 1-11, 2019.
- [9] Erraissi, A. Belangour, A. Tragha, "A Comparative Study of Hadoop-based Big Data Architectures", International Journal of Web Applications, vol. 9(4), pp. 129-137, 2017.
- [10] Mothukuri, S. Cheerla, R. Parizi, Q. Zhang and K. Choo, "BlockHDFS: Blockchain-integrated Hadoop distributed file system for secure provenance traceability", Blockchain: Research and Applications, vol. 2(4), pp. 1-7, 2021.
- [11] F. Awaysheh, M. Alazab, M. Gupta, T. Pena and J. Cabaleiro, "Next-generation big data federation access control: A reference model", Future Generation Computer Systems, vol. 108, pp. 726-741, 2020.
- [12] Choi and J. Hartman, "Stargate: remote data access between Hadoop clusters", in Proc. of the 36th Annual ACM Symposium on Applied Computing (SAC '21), ACM, NY, USA, pp. 32–39, 2021.
- [13] S. Rachuri, A. Gantasala, P. Emanuel, A. Gandhi, R. Foley, P. Puhov, T. Gkoutouva and H. Lei, "Optimizing Near-Data Processing for Spark", 2022 IEEE 42nd International Conference on Distributed Computing Systems (ICDCS), pp. 636-646, 2022.
- [14] L. Ponce, W. Santos, W. Meira, D. Guedes, D. Lezzi and R. Badia, "Upgrading a high performance computing environment for massive data processing", Journal of Internet Services and Applications, Vol. 10:19, 2019.
- [15] Khan, Y. Li, A. Anwar, Y. Cheng, T. Hoang, N. Baracaldo and A. Butt, "A Distributed and Elastic Aggregation Service for Scalable Federated

- Learning Systems”, ArXiv, abs/2204.07767, 2022. <https://arxiv.org/abs/2204.07767>.
- [16] Y. Seraoui, M. Bellafkih and B. Raouyane, "A high-performance and scalable distributed storage and computing system for IMS services", 2016 2nd International Conference on Cloud Computing Technologies and Applications (CloudTech), Marrakech, Morocco, pp. 335-342, 2016.
- [17] W. Xu, X. Zhao, B. Lao, G. Nong, "Enhancing HDFS with a full-text search system for massive small files", The Journal of Supercomputing, vol. 77, pp. 7149-7170, 2021.
- [18] U. Özdil, and S. Ayvaz, "An experimental and comparative benchmark study examining resource utilization in managed Hadoop context", Cluster Computing, 2022. <https://doi.org/10.1007/s10586-022-03728-7>.
- [19] Raj, R. D'Souza, "A Review on Hadoop Eco System for Big Data", International Journal of Scientific Research in Computer Science, Engineering and Information Technology, vol, 5(1), pp. 343-348, 2019.
- [20] T. Ma, F. Tian and B. Dong, "Ordinal Optimization-Based Performance Model Estimation Method for HDFS", in IEEE Access, vol. 8, pp. 889-899, 2020.
- [21] PHP-Hadoop-HDFS Library, Pure PHP unified wrapper for WebHDFS and CLI, <https://github.com/adprofy/Php-Hadoop-Hdfs> (Accessed March 2023).
- [22] The cURL Project, <https://curl.se/> (Accessed March 2023).

Gradually Generative Adversarial Networks Method for Imbalanced Datasets

Muhammad Misdrum^{1*}, Muljono^{2*}, Purwanto³, Edi Noersasonko⁴

Faculty of Computer Science, Universitas Dian Nuswantoro, Semarang 51031 Indonesia^{1,2,3,4}

Faculty of Information Technology Merdeka University Pasuruan 67129 Indonesia¹

Abstract—Imbalanced dataset can cause obstacles to classification and result in a decrease in classification performance. There are several methods that can be used to deal the data imbalances, such as methods based on SMOTE and Generative Adversarial Networks (GAN). These methods are used for overcoming data oversampling so that the amount of minority data can increase and it can reach a balance with the majority data. In this research, the selected dataset is classified as a small imbalanced dataset of less than 200 records. The proposed method is the Gradually Generative Adversarial Network (GradGAN) model which aims to handle data imbalances gradually. The stages of the GradGAN model are adding the original minority dataset gradually so that it will create new minority datasets until a balance of data is created. Based on the algorithm flow described, the minority data is multiplied by the value of the variable that has been determined repeatedly to produce new balanced minority data. The test results on the classification of datasets from the GradGAN model produce an accuracy value of 8.3% when compare to that without GradGAN.

Keywords—Classification; imbalance; GAN model; GradGAN model; significant oversampling

I. INTRODUCTION

The data mining solving on classification is a research topic that still needs contribution [1]. Because every use the dataset often result in data imbalances which can reduce the performance both of images and classification accuracy [2], [3]. In research on data imbalance, many methods have been offered but the method's robustness has not been satisfactory. This phenomenon is shown by existence of several data classification results that are less than [4], [5]. There are several cases of imbalanced data, for example, in the medical field, regarding disease prediction [6] and hepatitis diseases detection [7]. Imbalanced data classification is a vital problem [8], and the key is to create a flexible and correct method of classifying minority and majority data. This raises an urgent need for a better solution to the data imbalance problem so that the classification process can be more optimal.

The sampling method can be carried out at the preprocessing stage, for example, in the case of missing values, which has less significant impact on machine learning outcomes [9]. In the majority class, the use of under-sampling technique is relatively easier to do to balance the data [10]. Whereas in the minority class, the over-sampling method that commonly uses Synthetic Informative Minority Over-Sampling (SMOTE) algorithm to find synthetic data, is currently effective [11], [12]. Another over-sampling method

for dealing with synthetic data is Borderline-SMOTE [13]. There is a drawback of the SMOTE algorithm, namely there is no consideration of neighboring information from minority class samples, resulting in over generation. The Modified Fuzzy-Neighbor Weighted Algorithm has also been proposed, the classification results are also better [14]. Although existing classification methods have been able to overcome data imbalances, a new approach is still needed to overcome the challenges of the problem of heavy imbalanced data flows [15]. The online ensemble learning algorithm has also been used for unbalanced data streams [16]. Recently, a relatively new method has been developed to overcome over-sampling techniques for imbalanced data, namely the Generative Adversarial Network (GAN) method; the goal is to overcome the difficulties of minority data samples to be more balanced with the majority data [4], [17]. Initially, the GAN model was implemented in deep learning machines, namely about human faces, adopting images so as to produce better images [18]. Another implementation of GAN, is that it models complex real-world image data and normalizes data imbalances [19]. There are several developments of GAN models, including Triple Generative Adversarial Nets (TripleGANs) [20], and Senti Generative Adversarial Networks (SentiGAN) [21]. The development of GANs is not only for deep learning but also addressing imbalances of machine learning in data mining. For example, to classify public data of Bank Marketing, Credit, and Lending Club from the UCI repository, the GAN model and the random forest method produce several sub-classes of sample data and these sub-classes will be combined with the original dataset to form a new minority dataset [4]. Generative Adversarial Networks (GAN) model is also used to create new text from the scarcity of the Dialectal Arabic dataset which generates annotations (notes or comments) with automatic generation [21].

Learning methods that are commonly used and have good performance are Naïve Bayes (NB), k-Nearest Neighbour (k-NN), Support Vector Machine (SVM), and Decision Tree (D-Tree). NB has the advantage that it is simple and works well in the real world. The weakness is that the dataset to be tested must be in numerical form [22], [23]. K-NN has advantages, which are effective and robust against noisy training data [24]. SVM has the advantage of being able to generalize and is known to produce high accuracy values but is less than optimal when applied to imbalanced data [25]. Meanwhile, the drawback is that it is difficult to implement in large cases and is developed for two-class problems [26]. The D-Tree method has the advantage of a simpler and more straightforward, specific, and flexible way of making decisions, while the

weaknesses are frequent overlap and difficulty in designing a more optimal one [27]. The random forest ensemble applied to predict heart disease produces a better accuracy value than the other methods [28]; besides that, it is applied to the classification of Sarcastic Tweet, resulting in the highest accuracy [22].

Based on the above explanation, to overcome the heavy flow of imbalanced data, this research proposes a Gradually Generative Adversarial Network (GradGAN) model. The GradGAN model works to handle oversampling data with a resampling technique, namely adding minority data gradually. This technique is expected to be optimally applied to all dataset sizes. In future research, the GradGAN model can be developed and relied upon to deal with data imbalance problems. To test the quality of the data resulting from the

GradGAN model, data can be classified by using the NB, SVM, k-NN, D-Tree and RF methods because they are common methods and have good classification performance. The dataset used in this research is a small imbalanced dataset that is taken from the UCI repository.

II. RELATED WORK

So far, the obstacle faced by researchers in handling datasets in machine learning and deep learning applications is data imbalance. The reason is because the data imbalance makes the performance of the machine not optimal. In fact, the results of research on classification show that classification performance is less significant. There are several methods in research to handle the classification and imbalance of datasets.

TABLE I. LIST OF RESEARCH SUPPORTING PAPER REFERENCES

Papers	Classification method					Imbalanced Method									
	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
[2]														✓	
[3]						✓				✓					
[4]					✓				✓						
[5]	✓	✓		✓				✓							
[6]	✓		✓			✓									
[7]				✓							✓				
[8]	✓	✓	✓	✓	✓	✓									
[9]	✓		✓			✓									
[10]					✓				✓						
[11]									✓						
[12]									✓						
[13]		✓											✓		
[14]									✓			✓			
[15]	✓	✓	✓												
[16]	✓					✓									
[17]				✓		✓									
[18]		✓	✓	✓											
[19]		✓	✓	✓											
[20]			✓		✓										
[21]					✓										
[22]		✓		✓	✓										
[23]									✓						
[24]	✓		✓		✓	✓			✓						
[25]						✓	✓								
[26]	✓		✓	✓			✓								
[27]															
[28]															✓
Proposed	✓	✓	✓	✓	✓	GradGAN									

Abbreviations: a)NB; b)SVM ; c)D-Tree; d)k-NN; e)RF; f)SMOTE; g)Borderline-SMOTE; h)Radial-Based Undersampling; i)GAN; j)PWIDB; k)Ada- NN; l)SentiGAN; m)TripleGANs; n)BAGAN;o) LSTM

The development model in this research refers to the literature research in Table I. The GAN model used in this paper is developed into a Balancing GAN (BAGAN). The BAGAN model requires a large amount of data to function properly [2]. In order to achieve automatic balancing, Piece-Wise Incremental Data Re-Balancing (PWIDB) is combined with the Racing Algorithm (RA) technique and incremental iterative balancing techniques. According to [3], PWIDB outperforms other balancing techniques. The GAN model serves to overcome data imbalance and handle minority data samples. Several data samples based on a random forest model were broken down into sub-sections and then combined with the original data to form new minority data. The most ideal selected dataset is large size. The advantage in classification has produced good AUC and F1 [4]. The Radial-Based Under sampling method is proposed to handle imbalanced data to get good results, by combining the NB, SVM, and k-NN classification [5]. The Synthetic Informative Minority Oversampling (SMOTE) Technique method is used to balance training data by creating artificial minority data. The test results of the Naive Bayes, SVM-RBF, C4.5, and RIPPER classification methods produce the best accuracy with a value of 89.5%, 90% precision, 89.4% recall, 89.5% F-score and 83.5% Kappa [6]. In this paper, the problem to be solved is to classify the hepatitis dataset that is not balanced.

To increase the accuracy of the k-NN method due to the influence of data imbalance, an Adaptive-Condensed NN (Ada-CNN) method has been developed and then the accuracy value can reach a maximum [10]. In this paper, it is explained that to improve accuracy, it must balance the data, the method used is SMOTE. To find higher accuracy, data has been tested by the common classification methods, namely SVM, Random Forest, KNN, Naive Bayes, and Decision Tree. Then the general classification methods are compared with the proposed method using a voting multiclassifier; the results are better than the general method [11]. To overcome the lack of ROC values due to imbalance data, this research used the SMOTE method. In order to know the ROC value, the NB and D-Tree methods are used for data classification [12]. To overcome the decline in classification performance due to imbalanced data, this paper applies a Generative Adversarial Network (GAN) model. The workflow begins with creating artificial data so that minority sub-data are formed. Furthermore, from some of the minority sub-data, they are formed together to form a new minority data. Classification algorithm used is the Random Forest (RF) method [17]. Generative Adversarial Network (GAN) model was first proposed by Goodfellow et al. [18]. Invoked as a model to bridge between supervised and unsupervised learning in 2014, it has been hailed as the most exciting ideas in machine learning in the last ten years. To overcome imbalances level in classification, object detection and pixels in segmentation, this paper applies the Generative Adversarial Neural Networks (GANs) model [19]. The GAN model is promising in its application in handling image forms, but there are weaknesses between the generator and discriminator which are not optimal and cannot control the resulting sample. To overcome this problem, research is applied using a generator, discriminator and classifier called the triple Generatif Adversarial Net (Triple-GAN), which has produced a good classification [20]. The GAN model

collaborated with sentimental to form sentimental GAN (SentiGAN) has been succeeded in text creation domain. The modified SentiGAN utility will amplify a small data set and produce a variety of high-quality sentences in different Arabic dialects obtained from the MADAR data set. And it can produce a higher number of sentences than the original data, and this method can reduce vocabulary and only use common words. Even though there are slight deficiencies in the resulting text, the key features detected can help the classification remain strong. So the Generator process is not only effective and consistent but can also improve classification when used from the original dataset and the resulting dataset from the model process [21]. To find the accuracy value of the 14 datasets, in this paper, the classification methods used include Naive Bayes (NB), SVM, D-Tree, and k-NN. And from the test results, the NB method has produced the best accuracy value [22]. To overcome class imbalance, in this paper, the algorithm chosen is SMOTE. Meanwhile, the NB method is used for classification. The test results have resulted in an accuracy value of 88.5%, more significant than the R algorithm of 87.5% [23]. To overcome the imbalance problem, the popular SMOTE algorithm is used while the k-NN method is chosen for classification. Imbalanced data test results produce lower accuracy, while data that is balanced has greater accuracy [24]. The case of credit assessment can not be just any method that can be applied because it can hinder the assessment work itself. A good method to use is the SVM method in collaboration with Least Squares SVM (LS-SVM), It is shown from the test results with eight data set that it produces better performance when compared to other methods such as D-Tree and k-NN. But, in this research, SVM is also supported by meta-heuristics to be better [25]. Another research on the treatment of breast cancer uses several classification methods, the best results is the SVM method with an accuracy of 80.4% [26]. In implementing the algorithm, it is very vulnerable to the existence of data, and sometimes it is found that the data is imbalanced. To overcome that case, in this research has utilized decline tree with the D-Tree and Random Forest (FR) methods for the data balance process. The test results show that all classifications are strongly influenced by the balance of the data to achieve maximum results [27]. Detecting heart disease with feature selection and Random Forest Ensemble methods, the resulting accuracy is better by 99% [28]. Comparison of SVM, k-NN, Maximum Entropy and Random Forest methods for classification, the best accuracy value is Random Forest [22]. The GAN model can be applied in several cases to handle dataset imbalances. For example, in the field of financial anti-fraud with smaller data samples, the GAN model can be collaborated with the Long Short-Term Memory (LSTM) network algorithm so that the problem of data imbalance can be taken seriously [28]. Both of these models will share roles to process data in a time sequence completed by the Long Short-Term Memory network, while the GAN is to distribute selected real data which will produce data that is similar to the original data [23]. In this research, the SMOTE algorithm was used to overcome data imbalance, while the best method was chosen to classify, namely NB, D-Tree and RF, for example in the case of breast disease prediction [24]. The class with a small number of observations is called the minority class, while the class with the largest number of observations is called the

majority class. In this paper, to handle data imbalance, the SMOTE and Borderline-SMOTE algorithms were selected, while to measure the accuracy value, they relied on the Safe Level Graph [13]. In the end, the GAN model will produce data similar to the original [23]. In this paper, to overcome data imbalances, the SMOTE algorithm is used, while for classification, the method chosen is D-Tree [26]. The preprocessing process is a step for preparing the dataset to be tested. The new original dataset is usually still not perfect, there are still deficiencies, for example an empty record is found, it is necessary to handle imputation to perfect the dataset. The goal of imputation is to keep the number of datasets ideal. There are several imputation methods to choose from, including Zero, Mean/Median, k-NN, Multivariate Imputation with Chained Equations (MICE), Deep Learning (Datawig) [26].

In the previous research, the GAN model algorithm can be described in Eq. (1), and the algorithm flow [4], [18] as follows:

- Collect m noise samples $\{z(1).....z(m)\}$ from the noise chamber $P_g(z)$.
- Collect m data samples $\{x(1).....x(m)\}$ from the $P_{data}(x)$ data set.
- Update the discriminator by promoting a random gradient. The specific formula is as follows:

$$\Delta\theta \frac{1}{m} = \sum_{i=1}^m \left[\log D^{(i)} + \log \left(1 - D \left(G \left(Z^{(i)} \right) \right) \right) \right] \quad (1)$$

The next step:

- Group m noise samples $\{z(1)....z(m)\}$ from the noise space $P_g(z)$.
- Refine the generator by reducing the random gradient. The specific formula is as Eq. (2):

$$\Delta\theta \frac{1}{m} = \sum_{i=1}^m \log \left(1 - D \left(G \left(Z^{(i)} \right) \right) \right) \quad (2)$$

The GAN flow diagram can be described as Fig. 1.

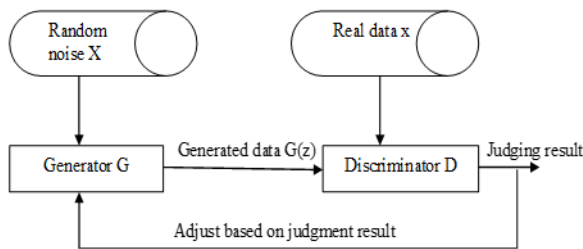


Fig. 1. GAN algorithm flowchart.

The first GAN produces two sub-models namely Generator G (generative model) and Discriminator D (discriminative model). In the initial algorithm, the generative model G based on available noise generates some data. Then Discriminator D will assess whether the data obtained is in the form of real data or data generated by the generative model G. The purpose of the generative model G is to make the resulting data as close as possible to the actual data and cannot be easily identified by

the discriminatory model D. The purpose of the discriminator the discriminative model D is to distinguish between real data and data generated by the generative model G [4].

III. PROPOSED METHODOLOGY

The steps of the research methodology process are carried out, starting from the dataset collection stage to the process stage of producing the expected output from the research. In this research, there are several steps as follows.

A. Data Collection

The selected dataset are Hepatitis, Immunotherapy, and Echocardiogram medical data, all of are taken from the UCI repository, where all of them are datasets with a relatively small number of less than 200 records and are not balanced [11].

B. Pre-Processing

The preprocessing process is a step for preparing the dataset to be tested. The new original dataset is usually still not perfect, and there are still deficiencies; for example, if an empty record is found, it is necessary to handle imputation to perfect the dataset. The goal of imputation is to keep the sum of the datasets perfect. There are several imputation methods to choose from; and in this research the imputation method used was the k-NN method. [26].

C. Proposed Model GradGAN

The GradGAN method is a development and modification of the GAN model. The working pattern of the Gradually Generative Adversarial Network (GradGAN) model is to overcome imbalanced data gradually until it gets balance. The process in GradGAN will involve a generator function to generate random datasets which will create majority and minority data from random samples of the original random dataset. And the discriminator that functions knows that the majority data are original data from the generator results. In this research, the main concern is minority data. Minority data will be processed with the GradGAN model to produce a new minority data sample. The next step is that the minority data will be multiplied by the variable value gradually so that new minority values will be created until a balance value is formed with the original majority data. Then classification is carried out where the new minority data serves as test data and the original majority data as training data. This GradGAN model is a new discovery model and has never been used by other researchers in dealing with imbalanced data. Following is a Fig. 2 of a flowchart and its explanation.

The process in GradGAN will involve a generator function to generate random datasets, which will create majority and minority data from random samples of the original random dataset. Moreover, the discriminator that functions knows that the majority and minority data are original data from the generator results. In this research, the main concern is minority data. Minority data will be processed with the GradGAN model to produce minority data samples. The next step is that the minority data will be multiplied by a variable value gradually so that new minority values will create until a balance value forms with the original or original majority data. Then the two data, namely the new minority data, serve as test

data and the original majority data as training data for classification. This GradGAN model is a new discovery model and has never been used by other researchers in dealing with imbalanced data.

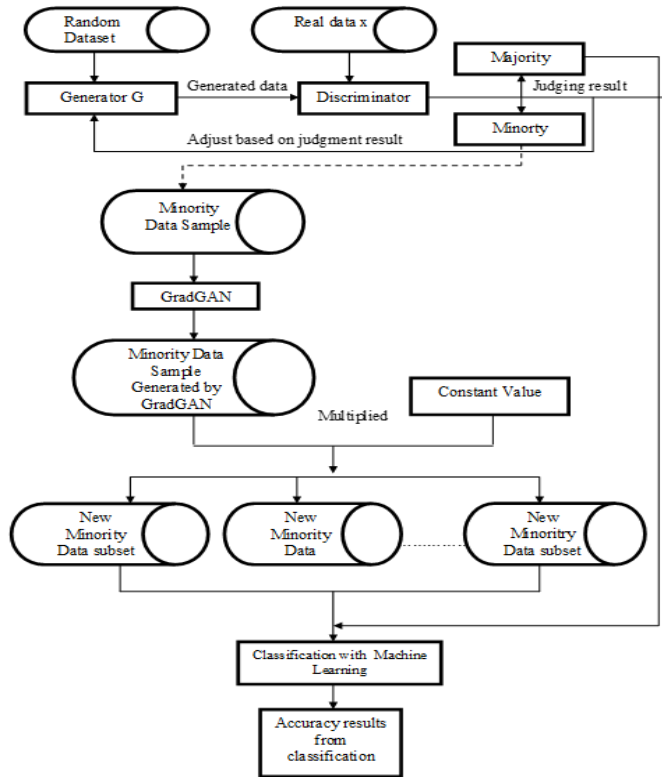


Fig. 2. Flowchart of the GradGAN algorithm.

Based on the above explanation, the GadGRAN model can be explained by calculating the equations in detail as follows:

- A sample set of x random datasets ($d^{(1)} \dots d^{(x)}$) with a total dataset of the $Dx^{(d)}$ variable will be generated to produce a total of majority and minority data.
- Determine the sample set x majority data class ($a^{(1)} \dots a^{(x)}$) with the number of the $Mx^{(a)}$ variable.
- Determine the sample set x minority data class ($i^{(1)} \dots i^{(x)}$) with the number of the $Mx^{(i)}$ variable.

In this research, the primary concern of data oversampling is to resample minority data with the $Mx(i)$ variable multiplied by a variable value to add minority data called new minority data. The goal is to create new minority data so that there is a balance with the original majority data. A mathematical formula can calculate with the following Eq. (3).

- Determine the sample x new minority data that is close to the balance ($i2^{(1)} \dots i2^{(x)}$) with the amount of data the $Mx(i2)$ variable.
- Determine the sample x variable value ($k^{(1)} \dots k^{(x)}$) with the value of the $Kx(k)$ variable.

- Calculate the number of new minority data using Eq. (3) or it can be written with Eq. (4).

$$Mx(i2) = Mx(i) (\sum_{k=0}^x Kx(k)) \quad (3)$$

$$Mx(i2) = \{(Mx(i)) (Kx(k))\} \quad (4)$$

The $Mx(i2)$ variable describes the number of new minority data, which results from the calculation of the constant the $Kx(k)$ variable multiplied by the number of original minority data in the $Mx(i)$ variable. This calculation is carried out in stages until there is a balance between the number of the $Mx(i2)$ and $Mx(a)$ variables' original majority data.

The GradGAN model, which has been described mathematically, can be described in the form of a Pseudo-code algorithm as follows:

1. Start
2. Input minority dataset $Mx(i)$;
3. Input Dataset majority $Mx(a)$;
4. Initialize $Mx(i2) = 0$;
5. Initialize $Kx(k) = 1$;
6. Compute the new minority value $Mx(i2) = (Mx(i)) * Kx(k)$
7. Is the value of $Mx(a) \geq Mx(i2)$ then
8. $Kx(k) = Kx(k) + 1$;
9. Goto 5
10. Else
11. Print $Mx(i2)$
12. Classification process
13. End

In the process of the algorithm for measuring balance, the $Kx(k)$ variable is multiplied by the minority data $Mx(i)$ on gradually, where $Kx(k)$ is variable value (1, 2, ..., k). $Mx(i2)$ or new minority data is the product of multiplying the original minority $Mx(i)$ with $Kx(k)$ variable, where k variable represents the number of variable values. In the next process, the minority data will be multiplied by a variable value gradually to create another new minority data until it finds data that is close to class balance.

This algorithm can be applied to all imbalanced data of all types of data sizes. In this research has been tested on three imbalanced datasets, which have different attributes and number of records. All data can be implemented well in this algorithm and produce a significant average accuracy value. The limitations of this algorithm cannot be applied to imbalanced datasets with missing values or with empty attributes.

IV. RESULTS AND DISCUSSION

The dataset is described and shown in Table II. In the description of the Hepatitis dataset, there are two classes representing death, "Die" and life, "live" In the Echocardiogram dataset, there are two classes representing death, "Dead" and life, "Alive". In the Immunotherapy dataset, there are two classes represented "No" and "Yes"; the purpose is to assist treatment. Treatments can be stopped if a positive "Yes" or a negative "No" is treated.

TABLE II. ORIGINAL DATASET DESCRIPTION

Dataset	Number of records	Number of features	Imbalanced ratio	Minority class	Majority class	Number of classes
Hepatitis	155	20	80 : 20	“Die”	“Live”	2
Echocardiogram	131	13	70 : 30	“Alive”	“Dead”	2
Immunotherapy	90	8	80 : 20	“No”	“Yes”	2

Fig. 3 compares the majority class and the minority class between the original dataset and the dataset that has been processed with the GradGAN model in stages with the final value. In the original dataset, the imbalance can obtain from the training data, which describes the majority class, and the test data, which describes the minority class. Furthermore, the GradGAN model functions to gradually increase the number of minority class data to become new minority data that is close to balance. By generating the original data, the calculation results can be seen in Fig. 3.

The three datasets in the GradGAN model to achieve balance can be seen in the comparison of the majority and the minority as follows: hepatitis dataset 124: 124, echocardiogram 107: 96, immunotherapy 71: 57.

In Fig. 3, the data balance stage process has been carried out. It starts from balancing the original data to form majority and minority data. Then the next process is applying the Gradually Generative Adversarial Network (GradGAN) model, balancing the original data gradually, forming a comparison of the original majority data with the new minority data gradually. Each comparison of the datasets formed by majority and minority data will be carried out with a classification experiment. The majority of the experimental results resulted in increase in the accuracy value of each dataset in the five classification methods. The result is shown in Table III.

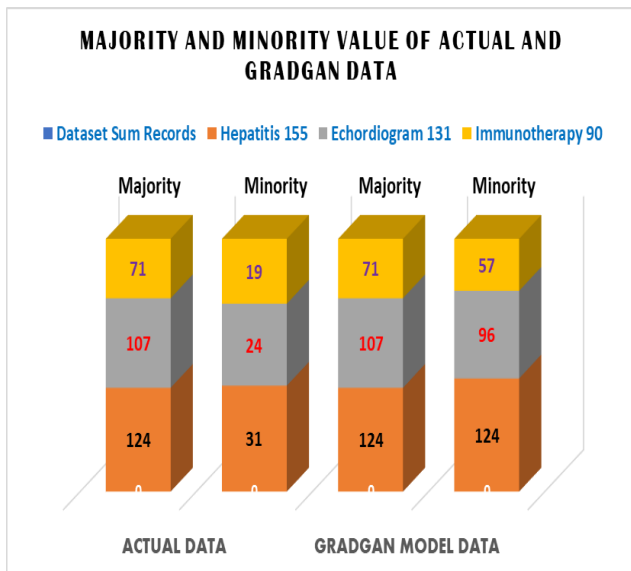


Fig. 3. Comparison chart of majority class and minority class original data and GradGan model.

In Table III, the results of the experimental accuracy of the two datasets are the original dataset and the dataset that already has the influence of the Gradually Generative Adversarial Network (GradGAN) model. First, the lowest classification

value for the original dataset is 66%, obtained from the hepatitis dataset, while the highest is 83% from the echocardiogram dataset. Second, is the classification of the dataset that has been balanced with the GradGAN model. Classification produces an accuracy value of at least 68% and a maximum of 94%. Of the three original datasets and datasets that already have influence of the GradGAN model that It has been tested on five classification methods, and there is an average increase in accuracy of approximately 8,3%.

TABLE III. THE RESULTS OF CALCULATING THE ACCURACY VALUE OF THE DATASET

Dataset Name	Methods	Accuracy Values (%)		
		(a)	(b)	(c)
Original Datasets	NB	73	75	73
	SVM	81	79	81
	D-Tree	66	82	81
	k-NN	77	65	75
	RF	79	83	82
GradGAN Model Influence Dataset	NB	80	85	69
	SVM	84	84	70
	D-Tree	84	87	94
	k-NN	71	71	75
	RF	88	89	93

Abbreviations:(a)Hepatitis;(b)Echocardiogram;(c)Immunotherapy

A description of the dataset is a way to find out the character of the data. From the description of the selected datasets in this research, it can be seen that there is an imbalance in data. The evidence of data imbalance is an unequal comparison between the majority class and the minority class. An imbalance will affect the resulting accuracy value in the classification. To overcome the imbalance of the dataset in this research, a new innovation was developed from the GAN model, namely the GradGAN model. The flow of the GradGAN algorithm is shown in Fig. 2. The generation process starts from a random dataset and a discriminator to find out the original dataset so as to form an imbalanced majority and minority data. The goal of the GradGAN model is to form some new, larger minority data gradually. In this research, to increase the number of minority data gradually, by multiplying the number of variables (1, 2, ... k) that has been determined. So that it will produce several data sets which eventually form new minority data that are balanced or close to balance, as shown in the Fig. 3. The dataset that will be tested is the original imbalanced dataset and the balanced dataset using the five selected classification methods. The results of the classification accuracy can be seen in Table III. Experiments from three balanced datasets have been carried out to prove that the GRadGAN model has good classification performance by producing a better increase in accuracy values. Table IV shows a collection of classification results from the original dataset and the GradGAN model influence dataset. The results show that a balanced dataset has a significant increase in accuracy values, namely hepatitis by 88%, echocardiogram 89%, and immunotherapy 94%. The increased accuracy of the

original dataset and the GradGAN model for each dataset, namely hepatitis 7%, echocardiogram 6% and immunotherapy 12% with an average increase in accuracy of 8.3%. Hypothesis analysis can be described in that, the influence of the GradGAN model can increase the accuracy value significantly and in the future it can be applied to data other than the data that has been tested.

TABLE IV. THE INCREASE IN THE RESULTS OF THE ACCURACY VALUE ON THE ORIGINAL DATASET AND THE GRADGAN MODEL

Datasets	(a)	(b)	(c)
Hepatitis	81%	88%	7%
Echocardiogram	83%	89%	6%
Immunotherapy	82%	94%	12%
Average Accuracy Improvement	82%	90,3%	8,3%

Abbreviations:(a)Originaldatasets;(b)GradGANmodel;(c)Accuracy improvement

V. CONCLUSION

Based on the results of the research discussion above, it can be concluded that the dataset is getting closer to being balanced and the results are getting better. In this research, accuracy values were generated from the classification of three datasets, both of original and influence datasets from the GradGAN model. Experimental results involving five classification methods and three datasets, the application of the GradGAN model resulted in an increase in the accuracy value of 8,3%, compared to the original dataset. Thus the hypothesis can be concluded that the GradGAN model is very influential in the process of handling imbalanced datasets. Evidence of increased accuracy from imbalanced datasets and balanced datasets resulting from the resampling process of the GradGAN model is shown in Table IV. This means that by applying the GradGAN method the accuracy results are superior to those without GradGAN. In this research, the model was only tested on small imbalanced datasets, but other trials need to be carried out on large imbalanced datasets.

VI. CONFLICTS OF INTEREST

The author has verified that they do not have any competing interests, as mentioned in their affirmation.

VII. AUTHORS' CONTRIBUTION

Conceptualization of paper topics, M. Misdrum and Muljono; research methodology, M. Misdrum and Muljono; validation of research results, E. Noersasongko, Purwanto, Muljono; the formal analysis and the research investigation, M. Misdrum and Muljono; the resources, M. Misdrum, E. Noersasongko, Purwanto and Muljono; accuration spatial datasets, M. Misdrum, and Fandi Yulian Pamuji; writing—original draft preparation, M. Misdrum; writing—review and editing, E. Noersasongko, Purwanto and Muljono; visualization data and the research results, M. Misdrum; supervision, E. Noersasongko, Purwanto, and Muljono; spatial and attribute data collector, M. Misdrum, and Fandi Yulian Pamuji.

ACKNOWLEDGMENTS

These research findings are part of a thesis at Universitas Dian Nuswantoro, Indonesia. Development of research results is supported by Universitas Merdeka Pasuruan, Indonesia.

REFERENCES

- [1] M. Bramer, Principles of Data Mining, vol. 53, no. 9. 2019.
- [2] G. Huang and A. H. Jafari, "Enhanced balancing GAN: minority-class image generation," *Neural Comput. Appl.*, vol. 8, 2021, doi: 10.1007/s00521-021-06163-8.
- [3] R. A. Mohammed, K. W. Wong, M. F. Shiratuddin, and X. Wang, "Pwldb: A framework for learning to classify imbalanced data streams with incremental data re-balancing technique," *Procedia Comput. Sci.*, vol. 176, pp. 818–827, 2020, doi: 10.1016/j.procs.2020.09.077.
- [4] Q. Shu, T. Hu, and S. Liu, "Random Forest Algorithm Based on GAN for Imbalanced Data Classification," *J. Phys. Conf. Ser.*, vol. 1544, no. 1, 2020, doi: 10.1088/1742-6596/1544/1/012014.
- [5] M. Koziarski, "Radial-Based Undersampling for imbalanced data classification," *Pattern Recognit.*, vol. 102, 2020, doi: 10.1016/j.patcog.2020.107262.
- [6] N. Nnamoko and I. Korkontzelos, "Efficient treatment of outliers and class imbalance for diabetes prediction," *Artif. Intell. Med.*, vol. 104, no. January, p. 101815, 2020, doi: 10.1016/j.artmed.2020.101815.
- [7] N. G. Siddappa and T. Kampalappa, "Adaptive condensed nearest neighbor for imbalance data classification," *Int. J. Intell. Eng. Syst.*, vol. 12, no. 2, pp. 104–113, 2019, doi: 10.22266/IJIES2019.0430.11.
- [8] G. A. Pradipta, R. Wardoyo, A. Musdholifah, and I. N. H. Sanjaya, "Improving classification performance of fetal umbilical cord using combination of SMOTE method and multiclassifier voting in imbalanced data and small dataset," *Int. J. Intell. Eng. Syst.*, vol. 13, no. 5, pp. 441–454, 2020, doi: 10.22266/ijies2020.1031.39.
- [9] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, no. February 2017, pp. 321–357, 2002, doi: 10.1613/jair.953.
- [10] T. Zhou, W. Liu, C. Zhou, and L. Chen, "GAN-based semi-supervised for imbalanced data classification," 2018 4th Int. Conf. Inf. Manag. ICIM 2018, pp. 17–21, 2018, doi: 10.1109/INFOMAN.2018.8392662.
- [11] I. Goodfellow et al., "Generative adversarial networks," *Commun. ACM*, vol. 63, no. 11, pp. 139–144, 2020, doi: 10.1145/3422622.
- [12] V. Sampath, I. Mautua, J. J. Aguilar Martín, and A. Gutierrez, A survey on generative adversarial networks for imbalance problems in computer vision tasks, vol. 8, no. 1. Springer International Publishing, 2021.
- [13] C. Li, K. Xu, J. Zhu, and B. Zhang, "Triple generative adversarial nets," *Adv. Neural Inf. Process. Syst.*, vol. 2017-Decem, pp. 4089–4099, 2017.
- [14] X. A. Carrasco, A. Elnagar, and M. Lataifeh, "A Generative Adversarial Network for Data Augmentation: The Case of Arabic Regional Dialects," *Procedia CIRP*, vol. 189, pp. 92–99, 2021, doi: 10.1016/j.procs.2021.05.072.
- [15] S. Taheri and M. Mammadov, "Learning the naive bayes classifier with optimization models," *Int. J. Appl. Math. Comput. Sci.*, vol. 23, no. 4, pp. 787–795, 2013, doi: 10.2478/amcs-2013-0059.
- [16] T. Sajana and M. R. Narasingarao, "Classification of imbalanced malaria disease using naïve bayesian algorithm," *Int. J. Eng. Technol.*, vol. 7, pp. 786–790, 2018, doi: 10.14419/ijet.v7i2.7.10978.
- [17] A. M. De Carvalho and R. C. Prati, "Improving kNN classification under Unbalanced Data. A New Geometric Oversampling Approach," *Proc. Int. Jt. Conf. Neural Networks*, vol. 2018-July, no. Cmcc, pp. 1–6, 2018, doi: 10.1109/IJCNN.2018.8489411.
- [18] R. Y. Goh and L. S. Lee, "Credit Scoring: A Review on Support Vector Machines and Metaheuristic Approaches," *Adv. Oper. Res.*, vol. 2019, 2019, doi: 10.1155/2019/1974794.
- [19] T. A. Khan, K. A. Kadir, S. Nasim, M. Alam, Z. Shahid, and M. S. Mazliham, "Proficiency Assessment of Machine Learning Classifiers: An Implementation for the Prognosis of Breast Tumor and Heart Disease classification," *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 11, pp. 560–569, 2020, doi: 10.14569/IJACSA.2020.0111170.
- [20] C. O. Truiçã and C. A. Leordeanu, "Classification of an imbalanced data set using decision tree algorithms," *UPB Sci. Bull. Ser. C Electr. Eng. Comput. Sci.*, vol. 79, no. 4, pp. 69–84, 2017.
- [21] D. C. Yadav and S. Pal, "Prediction of heart disease using feature selection and random forest ensemble method," *Int. J. Pharm. Res.*, vol. 12, no. 4, pp. 56–66, 2020, doi: 10.31838/ijpr/2020.12.04.013.

- [22] R. Kumar and J. Kaur, Random forest-based sarcastic tweet classification using multiple feature collection, vol. 163. Springer Singapore, 2020.
- [23] Z. Zhang et al., "A generative adversarial network-based method for generating negative financial samples," *Int. J. Distrib. Sens. Networks*, vol. 16, no. 2, 2020, doi: 10.1177/1550147720907053.
- [24] 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.*, vol. 11, no. 8, pp. 54–63, 2020, doi: 10.14569/IJACSA.2020.01110808.
- [25] C. Bunkhumpornpat and S. Subpaiboonkit, "Safe level graph for synthetic minority over-sampling techniques," 13th Int. Symp. Commun. Inf. Technol. Commun. Inf. Technol. New Life Style Beyond Cloud, Isc. 2013, no. September 2013, pp. 570–575, 2013, doi: 10.1109/ISCIT.2013.6645923.
- [26] A. M. Sowjanya and O. Mrudula, "Effective treatment of imbalanced datasets in health care using modified SMOTE coupled with stacked deep learning algorithms," *Appl. Nanosci.*, no. 0123456789, 2022, doi: 10.1007/s13204-021-02063-4.
- [27] W. Badr, "6 Different Ways to Compensate for Missing Values In a Dataset (Data Imputation with examples)," *Towar. Data Sci.*, 2019, [Online]. Available: <https://towardsdatascience.com/6-different-ways-to-compensate-for-missing-values-data-imputation-with-examples-6022d9ca0779>.
- [28] A. Muslim, A. B. Mutiara, R. Refianti, C. M. Karyati, and G. Setiawan, "Comparison of accuracy between long short-term memory-deep learning and multinomial logistic regression-machine learning in sentiment analysis on twitter," *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 2, pp. 747–754, 2020, doi: 10.14569/ijacsa.2020.0110294.

Human Fall Detection for Smart Home Caring using Yolo Networks

Bo LUO¹

Engineering Training Center, Chongqing Kechuang Vocational College
Yongchuan 402160, Chongqing, China

Abstract—In order to help the elderly and limit the incidence of falls that result in injuries, effective fall detection in smart home applications is a challenging topic. Many techniques have been created employing both vision and non-vision-based technologies. Many researchers have been drawn to the vision-based technique amongst them because of its viability and application. However, there is still room for improvement in the effectiveness of fall detection given the poor accuracy rate and high computational cost issues with current vision-based techniques. This study introduces a new dataset for posture and fall detection, whose photo images were gathered from Internet resources and data augmentation. It employs YOLO networks for fall detection purpose. Furthermore, different YOLO networks are implemented on our dataset to address the most accurate and effective model. Based on assessment parameters including accuracy, F1 score, recall, and mAP, the performance of the various YOLOv5n, s and YOLOv6s versions are compared. As experimental results showed, the YOLOv5s performed better than other.

Keywords—YOLO; computer vision; fall detection; smart home; caring

I. INTRODUCTION

In recent years, improvements in information and communication technologies (ICTs) and the innovations that followed them have significantly altered people's lives and given rise to elegant surroundings, cities, and societies. By observing our environments and making decisions to produce desired results, technologies like artificial intelligence (AI) and the Internet of Things (IoT) improve our quality of life. As the foundation of cities and societies, houses play a crucial role in developing smart living. They are anticipated to be major enablers of smart cities and societies [1]. Life expectancy is rising, and fertility is significantly declining in the modern world due to several socioeconomic causes [2]. An automated home-based solution to lessen the pressure on staffed health services and give frequent insight into fall risk would be an alluring alternative strategy [3] as more senior people struggle to preserve their independence and live in their own homes. At the same time, healthcare difficulties are becoming more and more significant due to the rise in the number of senior people worldwide. Elderly individuals who live alone must use human motion capture technology to address these problems. Also, by observing an elder's posture, it is possible to track how healthy they are, and if high-risk postures, such as falling over, are noticed, a warning may be sent. In addition to increasing the effectiveness of posture recognition, these systems will lighten the workload on human resources [4]. Due to considerations

including shifting viewing angles, body occlusion, and significant variations in human posture, figuring out how to recognize human posture is extremely difficult [5].

The three basic categories are sensor-based, vision-based, and radio-based HAR [5]. Sensor-based types rely on information gathered by sensors to identify human activity. For instance, a light sensor's activation can signal movement in a certain region or activity. Vision-based types rely on picture and video data formats to identify human activity. As an illustration, motions in smart home recordings imply doing specific tasks like cooking or strolling. Radio-based technology relies on the information and characteristics of signals to detect human activity. Wearable motion sensors that track the movement of body parts might detect particular behaviors like sitting or walking [6].

Systems for recognizing gestures and actions based on video analysis have been thoroughly investigated. Since vision-based data are less expensive and simpler to gather than sensor-based data, the most recent research has been on vision-based HAR. Thus, this study only partially covers the vision-based HAR investigations; it only includes a limited and representative sample. HAR is used in many applications, including surveillance systems, behavior analysis, gesture recognition, patient monitoring systems, and ambient assisted living (AAL). Recent research has focused on fall detection rather than fall risk prediction [7].

Falls are a common occurrence in the course of human growth. Falls happen when kids learn to stand, walk, climb, run, and engage in other activities. In a similar vein, falls also grow more common as we age. Though most falls are minor mishaps, as people age, falls become far more common and serious. It frequently underestimates falls' effect on people and society since they are just a natural part of life [8] [9]. Worldwide, falls are a significant health and financial problem. While falls inflict a high cost on healthcare systems in terms of in-patient and long-term care, they can also have indirect psychological impacts, such as reducing or avoiding physical activity out of fear of falling, in addition to the immediate direct medical repercussions. From an economic standpoint, it is obvious that any additional demand on the health system has a direct financial cost to society; nevertheless, lost productivity also has a hidden cost that is frequently disregarded. It should be no surprise that fall detection research and innovation have been sparked by the frequency and effects of falls, the risk of mortality, healthcare expenditures, and lost social and economic output [8].

The HAR problem has long been addressed using machine learning (ML) techniques, including random forest (RF), Bayesian networks, Markov models, and support vector machines (SVM). Traditional ML algorithms have demonstrated impressive performance in tightly regulated settings with little input data. However, they could be more efficient and take a lot of time to produce since they need several pre-processing stages and appropriate hand-crafted characteristics. External feature use also results in subpar incremental learning or unsupervised learning results. Due to its remarkable performance in several study domains, including object detection and identification, picture classification, and natural language processing (NLP), deep learning has attracted much community attention in recent years. Deep learning significantly decreases the effort required to select the best features compared to typical machine learning algorithms. The deep learning framework has also been effective with unsupervised learning and reinforcement learning. As a result, more and more newly released HAR frameworks are through deep learning [7].

Deep learning, particularly the convolutional neural network (CNN), which is modeled after the hierarchical processing of the human visual cortex, has had great success in the last few years when it comes to classifying images. CNN is a useful feature extraction and classification technique because it can automatically learn discriminative features from training data [10]. One or more of CNN's exciting application areas include speech recognition, object detection, video processing, object classification and segmentation, and natural language processing [11].

The two primary kinds of target detection algorithms based on CNN are two-stage detection algorithms, which divide target detection into two steps: finding and recognizing. The traditional technique, known as the Region-Convolutional Neural Network (R-CNN), performs poorly and cannot keep up with real-time demands. Fast regions with CNN (Fast R-CNN) and faster regions with CNN (Faster R-CNN) are offered because of further advancements based on R-CNN, although they still need to satisfy people's demands for real-time performance. The other is a one-stage detection technique, which combines target placement and recognition into a single phase for optimization. Single-shot multi-box detector (SSD) and YOLO series models are classic examples of this method [12].

In this paper, different versions of YOLO networks are investigated for fall detection which is utilized in smart home application for elderly caring. This investigation intends to address the most of less memory space usage and highest accuracy detection rate in fall detection.

The format of this essay is as follows: The background is offered in Section II, the technique is described in Section III, along with the training and testing procedure, the discussion and analysis of the findings are found in Section IV, and the conclusion and future work are found in Section V.

II. RELATED WORKS

Several recent deep learning-based approaches have been reported to improve the accuracy of fall detection recognition.

In this section, we will have a brief overview of several methods of this type.

The suggested approach establishes a framework for fall protection research. In contrast to providing raw data to the cloud for real-time prediction of fall occurrences, [13] offered a framework for fall detection through LSTM networks that utilized edge devices like a laptop for computation. The suggested system used the open-source Apache Flink streaming engine, a low-cost MetaMotionR sensor from MblentLab, and three-axis accelerometer raw data. The architecture has been trained and tested using a portion of the public MobiAct dataset. The created system found that the waist was the ideal location for placing sensors. With a 95.8% accuracy rate, the suggested framework can forecast autumn occurrences from current fall data. Performance can be improved by using many sensors and data streams.

The study [14] has presented a fall detection system design with health monitoring features. Utilizing low-power enabled low-power wide-area network (LPWAN) technology, the system combined Edge computing, Fog computing, and a compression method to transport data. This reduced the latency of the system. LSTM and RNN networks have been developed on the edge computer to identify falls from the incoming data. Raw data is delivered to the cloud via these edge gateways for online analysis, along with real-time notifications and alarms. The suggested architecture extends battery life and allows operation in regions with weak network access. Using the suggested approach and the MobiAct dataset, fall event predictions have been made with an average accuracy of 95% and a precision of 90%. The system performance may be increased by making a few adjustments and combining various techniques.

To more reliably and speedily identify fall behavior, [15] provides a fall detection approach based on a video in a complicated environment. The following is the paper's primary contribution: First, a YOLOv3 network model for the detection method is suggested. Second, the Pascal VOC data set format creates the human fall detection data set. Next, a self-built data set is used to train and test the network model on a GPU server. The experimental data demonstrates that the algorithm's mAP is 0.83 and its AP of down is 0.97, which are better than other conventional algorithms and have a strong resilience and detection impact.

In [16], a noise-tolerant FDS is shown operating with missing values in the data. This study seeks to use DL techniques for wearable sensor-based fall detection when faced with difficulty identifying missing values in data. On the SisFall and UP-Fall datasets, two public accessible datasets, the suggested method applies RNN with an underlying stack of BiLSTM blocks. When a value in the sensor data disappears, BiLSTMs can quickly obtain the long-range context thanks to their innate capacity to store long-term dependencies from both the past and the future. This explains why BiLSTM is a good fit for our noise-tolerant fall detection system that uses sequential sensor data. On the benchmark datasets for SisFall and UP-Fall Detection, the system outperforms the current state of the art with the accuracy, sensitivity, and specificity values of 97.21%, 97.41%, 96.38%, and 91.45%, respectively.

Due to its ability to maintain long-term dependencies from the past and future, the results show that BiLSTM is an appropriate model option to manage missing variables for wearable fall detection systems.

Based on the Fast Pose Estimation approach, [17] suggests a revolutionary human fall detection solution. The method employs 1Dimensional Convolutional Neural Network (1D-CNN) and Time-Distributed Convolutional Long Short-Term Memory (TD-CNN-LSTM) models to categorize retrieved data from picture frames accurately. As a result, the suggested approach effectively contributes to accurate human fall detection by employing the Fast Pose Estimation technique. The original URFD films were subjected to rotation, brightness, horizontal flip, and gamma correction augmentation procedures for dataset preparation. The result was an improved version of the URFD dataset. This resulted in a total of 560 movies, comprising 240 videos of falls and 320 videos of daily activities, which were then used to evaluate the produced models.

III. METHODOLOGY

A. YOLO based Networks

Basically, YOLO is a pre-trained object detector that is trained to recognize everyday objects like tables, chairs, cars, phones [18]. To create a model that can detect human falls, we used different versions of yolo, among which the yolov5s model obtained better results.

1) *YOLO v5 network*: The YOLO v5 target detection model, which avoids the recomputation of candidate areas in the two-stage series and has high identification precision and quick inference, is the most representative target detection model in the one-stage series. The four primary model structures of the YOLO v5 architecture are YOLO v5l, YOLO v5x, YOLO v5m, and YOLO v5s [19], with decreasing order of network complexity. The YOLO v5n model, which has just 1.9 MB parameters and the same model depth as the YOLO v5s model but a network width half that of the YOLO v5s, was later developed to adapt the solution to mobile devices.

The backbone, neck, and head are the three fundamental parts of the YOLO v5 basic architecture. Fig. 1(a) to 1(d) show the module composition for the basic design. A CNN, which creates visual characteristics by combining many fine-grained pictures, is one of the Backbone structures. Convolution operations such as 2D convolution, 2D regularization, and SiLU activation are carried out by the conv module, which serves as the architecture's fundamental convolution unit [20].

2) *YOLO v6 Network*: The second autonomous detector used in this investigation is the single-stage object detection framework for industrial applications, YOLOv6, the most recent member of the YOLO family. Although this model does not belong to the official YOLO series, it reportedly exceeds YOLOv5 in terms of detection precision and inference speed, making it more effective for industrial applications. YOLOv6 includes many improvements in the Backbone, Neck, Head blocks and also training strategies. For instance, the Neck and Backbone in YOLOv6 have been redesigned by using Rep-

PAN and EfficientRep structures, respectively, based on the idea of hardware-aware neural network. EfficientRep Backbone can make use of hardware computing power, such as GPU, and it also has strong representation capabilities. Rep-Pan Neck, is more accurate and faster than PANet and SPP. YOLOv6 Head is decoupled by adding a layer between the network and the final Head, which in turn improves the performance during the training process [21]. The architecture of YOLOv6 is shown in Fig. 2.

B. Dataset

A bespoke fall detection dataset containing two directory pictures and labels was made using photographs gathered from various sources. The study's classifications refer to sitting, walking, and falling. Two subdirectories—Training (333 photos) and Val (111 images)—are included in the images directory and are utilized for different purposes. Here, we have text files with labels for that specific image in this directory. The Labels directory has two subdirectories, train and Val. Our dataset is shown in various cases in Fig. 3.

The 333 photos from the dataset were increased to 1092 images utilizing Roboflow to enhance our model. A maximum of three enhanced versions of each image were created by randomly applying blur (up to 1 px), noise (up to 5% px), brightness (between -40% and +40%), exposure (between -35% and +35%), rotation hue (between -50° and +50°), and exposure (between -35% and +35%).

The dataset related to validation consists of three sets, including one main dataset and two other sets obtained by applying preprocessing and augmentations. In one of our preprocessing suites: resizing (stretched to 416x640), automatic contrast adjustment (using adaptive equalization), grayscale (applied) and increments: brightness (between -40% and +40%), exposure (between -21% and +21%) and in another set of pre-processing: resizing (stretched to 640x480), automatic contrast adjustment (using adaptive equalization), and gain: brightness (between -40% and +40%), exposure (between -21% and +21%), 90° rotate (clockwise, counterclockwise) we used. Examples of augmented images are shown in Fig. 3. Lastly, the labeled pictures are divided into a training (70%) and validation set (30%).

C. Google Colab

We used Google Colab, which provides free access to powerful GPUs. All training and testing tasks are performed using a 12GB NVIDIA Tesla T4 GPU. Our model was trained for 20 epochs with batch size of 16, image size of 640 and with YOLOv5 default adjustment for other hyper parameters. Fig. 4 shows the Google Colab's details for our model implementation.

D. Training and Testing

It is usually a good idea to start with a model that has already been trained on very big datasets and then use its weights to train an object detector. Even if the trained weights do not have the items needed for this experiment, this is OK. Transfer learning is the name for this process. In order to speed up network learning, beginning weights from a pre-trained model are utilized, which comprises weights from the COCO

dataset. Additionally advantageous is the fact that fewer data will be needed. [18].

Our total dataset consists of 1425 images, 75% of which are used for training, 25% for validation. 75% of training includes 1092 images, 25% of validation includes 333 images.

IV. EXPERIMENTAL RESULTS AND PERFORMANCE ANALYSIS

In this section, we introduce the experimental results and model performance analysis, and show the training outcomes with using pretraining weights and compare the three models of YOLO.

A. Experimental Results

We trained our model with YOLOv5n, YOLOv5s, YOLOv6 which are different versions of YOLO model, YOLOv5s performed better. Some examples of model predictions for new and unseen images are shown in Fig. 5, and label 0 for falling, label 1 for walking, and label 2 for sitting.

B. Model Evaluation

The gathered experimental data were compared in this work using various assessment criteria, including accuracy, recall, and mean average precision (mAP). The true positive rate, or TPR, is a statistic used to assess the likelihood that items from the real world would be correctly identified. When a model produces no false negatives, which indicates that there are no bounding boxes that are not recognized but ought to be detected, it has a high recall. Eq. (1) provides the mathematical form for the recall.

$$R = \frac{TP}{TP+FN} = \frac{TP}{\text{Total Ground Truths}} \quad (1)$$

The real positive and false negative are denoted in the equation above by the letters TP and TN, respectively. Eq. (2) defines *precision* as the percentage of correctly anticipated positives, often known as the positive predictive value. The accurate model creates no false positives (FP) and only detects important things.

$$P = \frac{TP}{TP+FP} = \frac{TP}{\text{Total Predictions}} \quad (2)$$

According to Eq. (3), AP is the area under the PR curve, and mAP is the average of all AP values across various classes and categories (3).

$$mAP = \frac{1}{n} \sum_{i=1}^n AP_i \quad (3)$$

where n is the number of classes [22].

We YOLOv5n, YOLOv5s, and we taught YOLOv6s. In Fig. 6 and 7, we have shown the confusion matrix, F1 confidence and Precision-Recall curves, respectively. In addition, we have included the parameters of each model for better evaluation.

Table I show the performance results for different Yolo based models. These models are YOLOv5n, YOLOv5s and YOLOv6s. As shown in the Table, YOLOv5s presents better results compared to other models.

YOLOv5s model obtained better results.

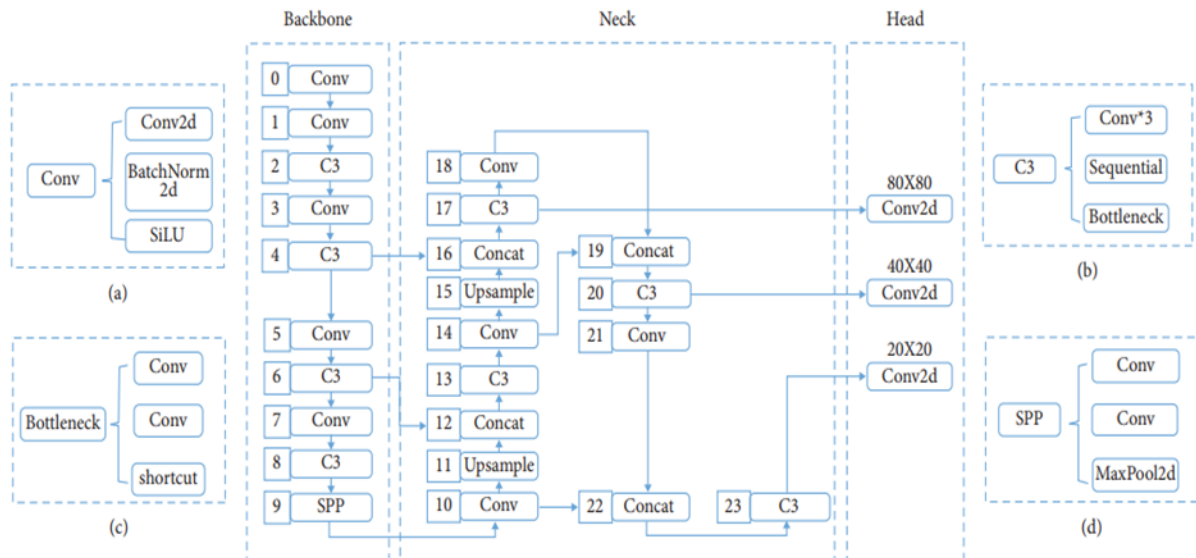


Fig. 1. Yolov5 architecture [20].

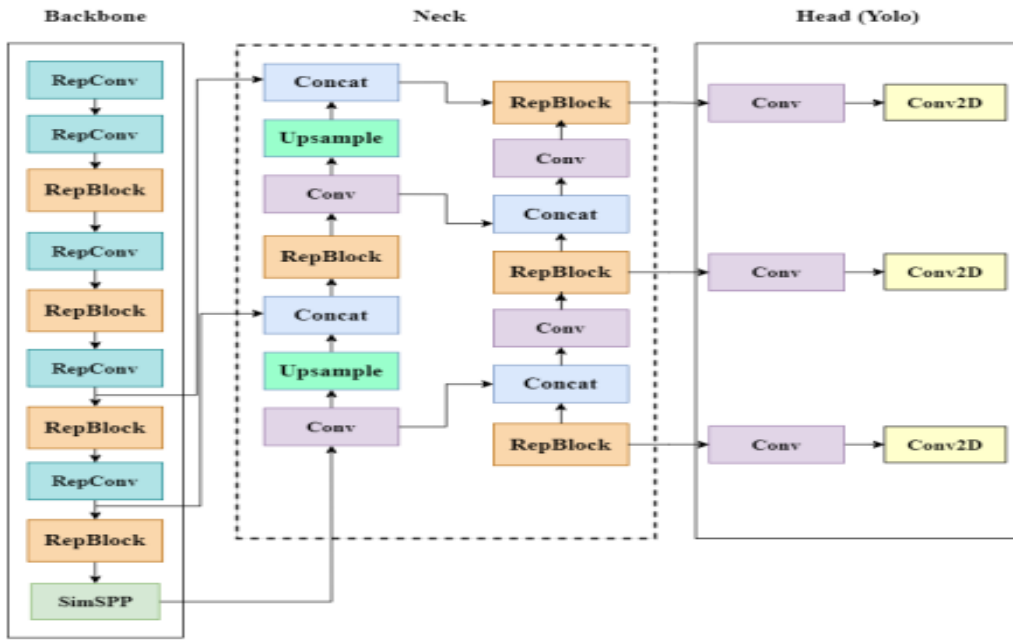


Fig. 2. Network structure of YOLOv6 [21].



Fig. 3. Sample images from dataset.

```

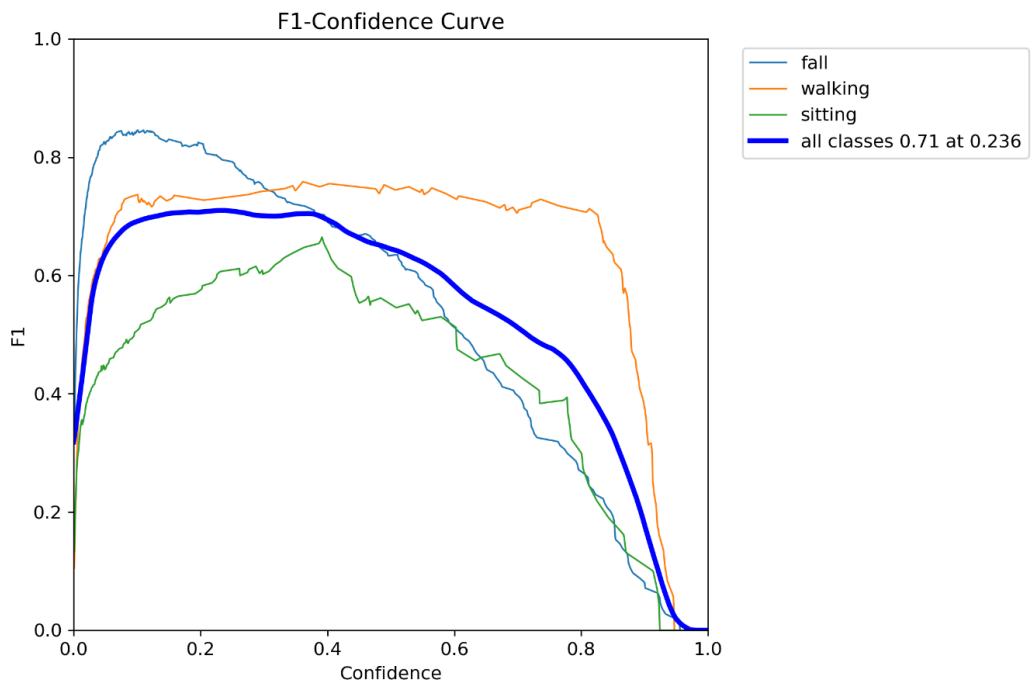
+-----+
| NVIDIA-SMI 460.32.03   Driver Version: 460.32.03   CUDA Version: 11.2   |
+-----+
| GPU  Name          Persistence-M| Bus-Id        Disp.A | Volatile Uncorr. ECC |
| Fan  Temp   Perf    Pwr:Usage/Cap|      Memory-Usage | GPU-Util  Compute M. |
|-----+-----+-----+-----+-----+-----+
|  0  Tesla T4             Off   | 00000000:00:04:0 Off |             0         |
| N/A  63C    P8      11W / 70W   |  0MiB / 15109MiB |           0%      Default |
|-----+-----+-----+-----+-----+

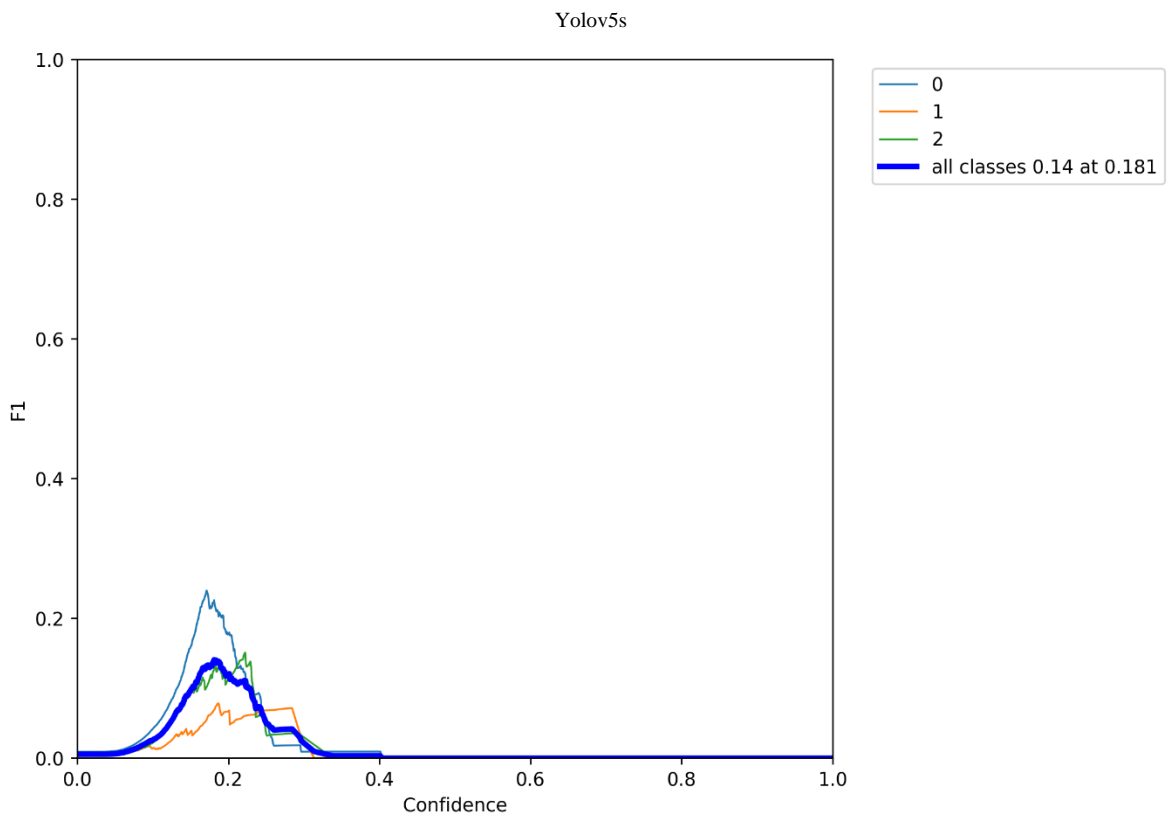
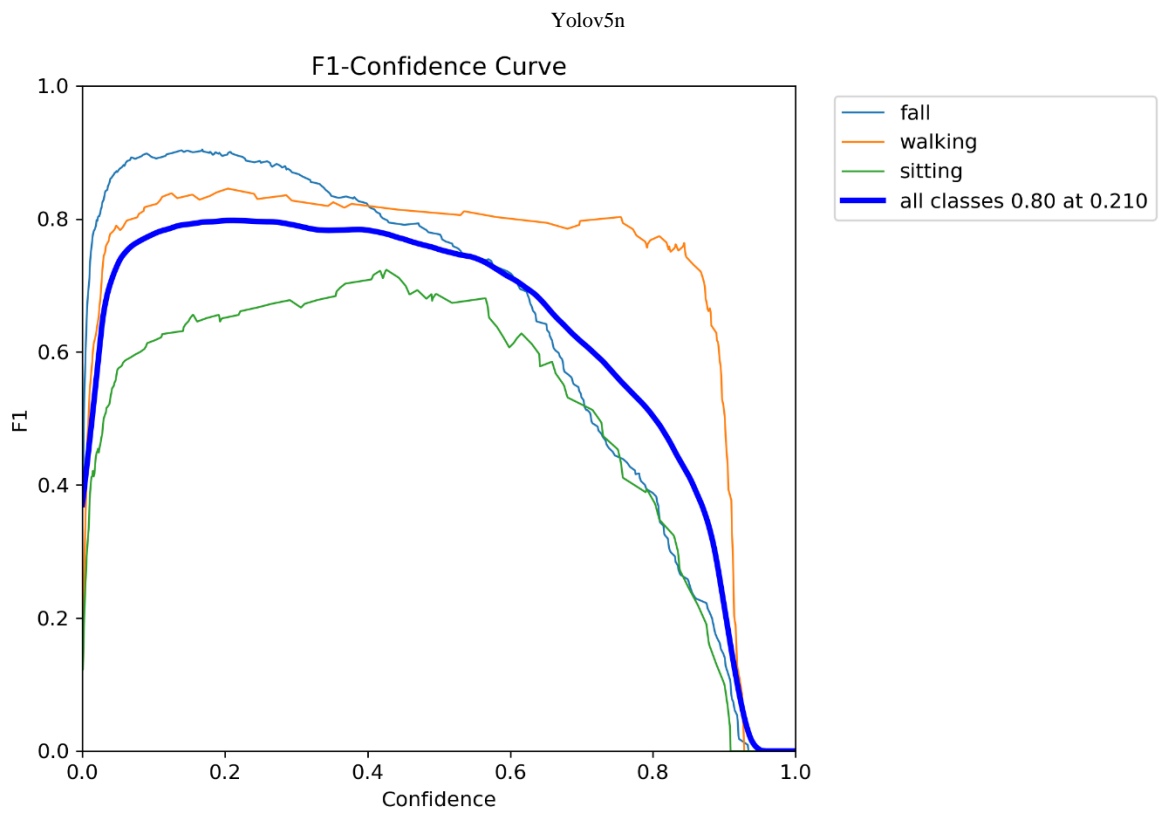
```

Fig. 4. Details of google colab' GPU.



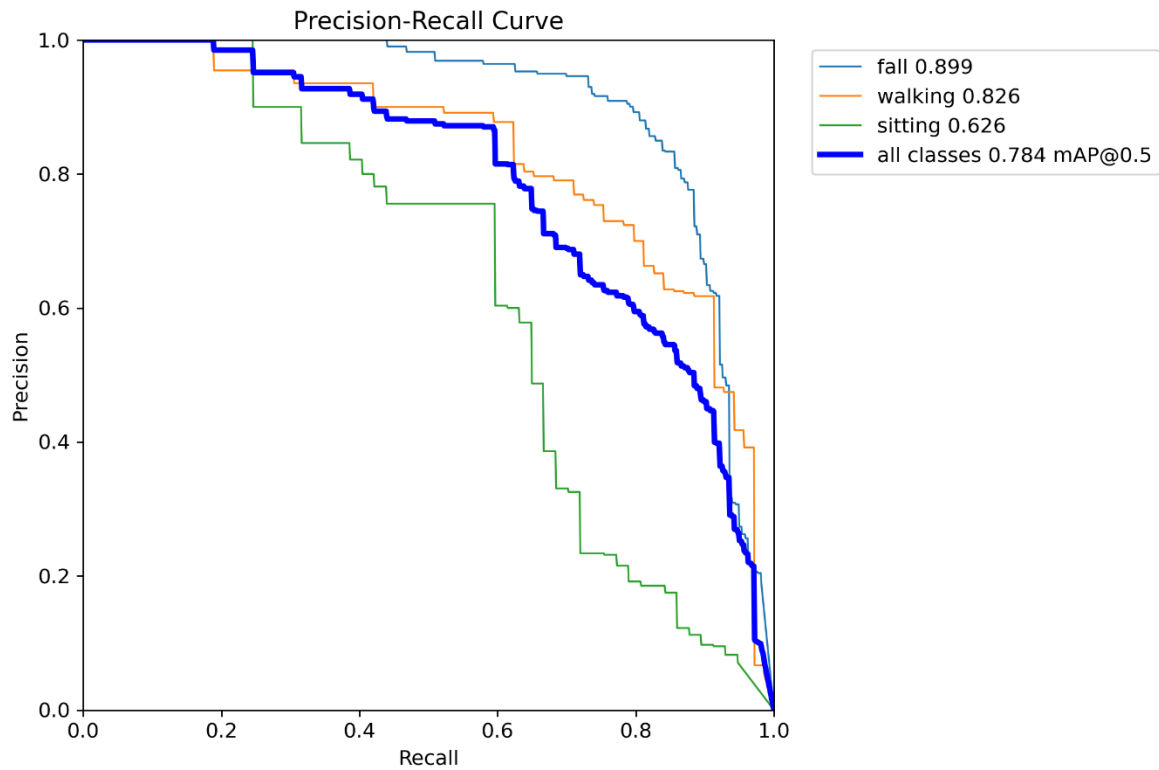
Fig. 5. Experimental results.



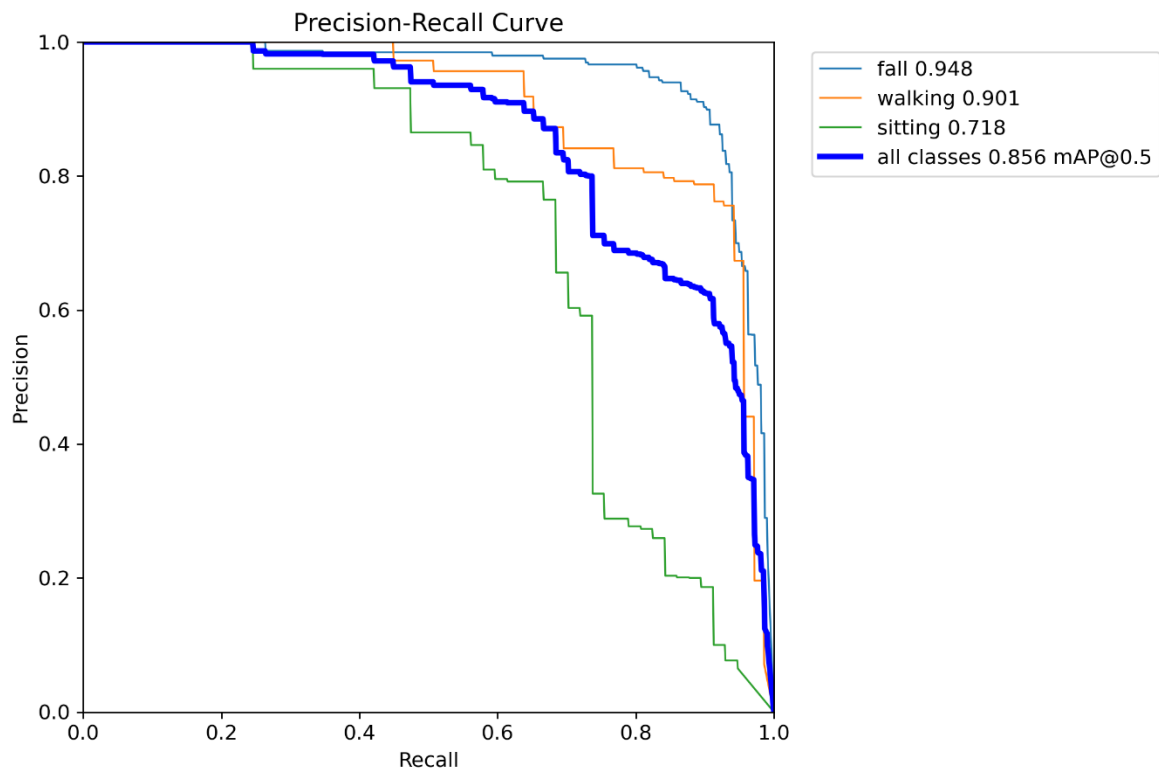


Yolov6s

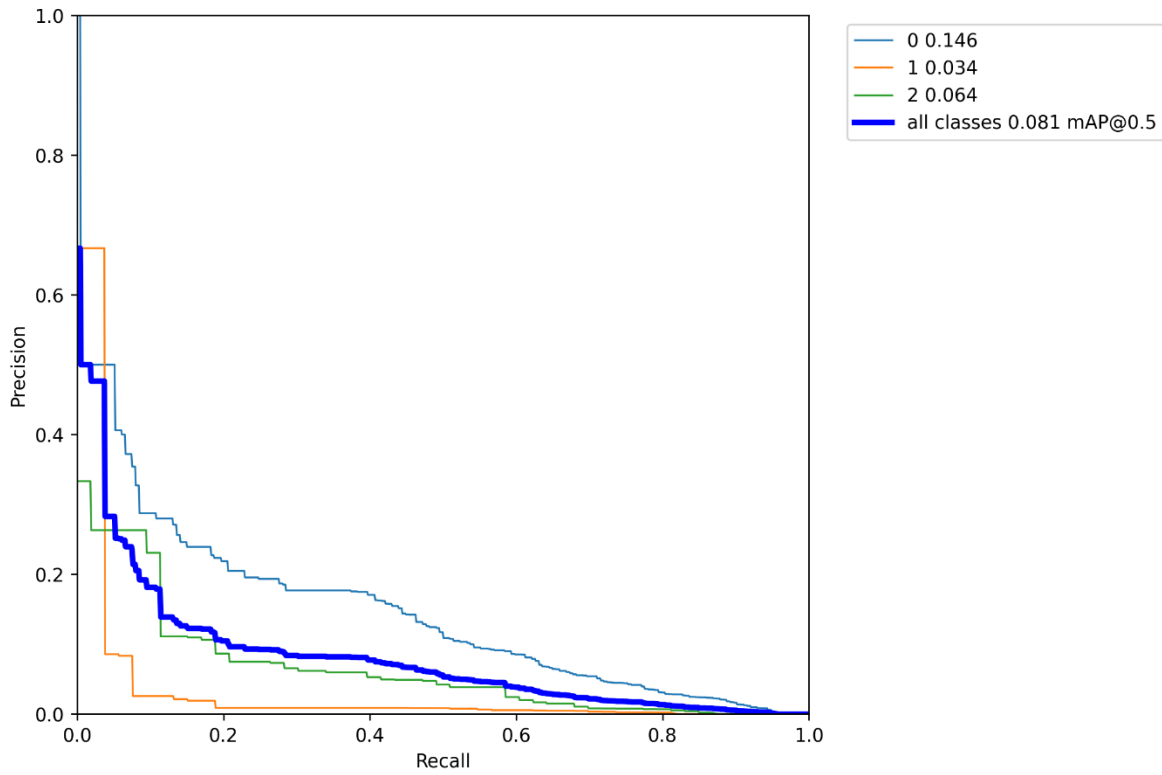
Fig. 6. F1-Confidence curves.



Yolov5n



Yolov5s



Yolov6

Fig. 7. Precision-recall curves.

TABLE I. RESULT OF MAP, RECALL, PRECISION

Model	MAP@0.5	MAP@0.5: 0.95	Precision	Recall
YOLOv5n	0.845	0.45	0.844	0.7
YOLOv5s	0.948	0.588	0.938	0.84
YOLOv6s	0.741	0.215	0.95	0.884

V. PERFORMANCE COMPARISON

This section presents performance comparison of different human fall detection methods. These methods involve OpenPose + LSTM [23], OpenPose + CNN [24], OpenPose + three thresholds [25] and the proposed method. In order to conduct fair comparison, these methods are experimented in current dataset and evaluated the performance using the procedure stated in [26]. Table II shows the performance comparison for existing methods.

TABLE II. PERFORMANCE COMPARISON BETWEEN METHODS

Model	OpenPose + LSTM	OpenPose + CNN	OpenPose + three thresholds	our proposed method
Accuracy	0.936	0.917	0.924	0.948

As shown in Table II, our proposed method presents better results compared to other existing methods in terms of the accuracy rate.

VI. CONCLUSION

This study investigated the advantages of data augmentation for the human posture dataset, and a dataset consisting of 1092 images for training and 374 images for testing was created from three classes, i.e. falling, walking and sitting. The performance of different YOLOv5 and YOLOv6 kinds was compared in the second stage based on accuracy, recall, and mAP. Based on the results, YOLOv5s outperformed other fall detection algorithms and had the greatest mAP @ 0.5, 0.948 and mAP @ 0.5: 0.95, which was 0.588. Finally, the proposed method is compared to other existing methods to demonstrate the outperformed method. The proposed method in this study is sensitive to the camera field of view for elderly fall detection that can be effectively influenced on accuracy rate in real applications. For future works, may employ other sensors combine with vision-based sensors to improve accuracy rate detection.

REFERENCES

- [1] Alqahtani, E., et al., Smart homes and families to enable sustainable societies: A data-driven approach for multi-perspective parameter discovery using bert modelling. *Sustainability*, 2022. 14(20), pp. 13534.
- [2] Zolfaghari, S., E. Khodabandehloo, and D. Riboni, TraMiner: Vision-based analysis of locomotion traces for cognitive assessment in smart-homes. *Cognitive Computation*, 2022. 14(5), pp. 1549-1570.
- [3] Forbes, G., S. Massie, and S. Craw, Fall prediction using behavioural modelling from sensor data in smart homes. *Artificial Intelligence Review*, 2020. 53(2), pp. 1071-1091.
- [4] Quan, W., et al., Human posture recognition for estimation of human body condition. *Journal of Advanced Computational Intelligence and Intelligent Informatics*, 2019. 23(3), pp. 519-527.
- [5] Yu, N. and J. Lv, Human body posture recognition algorithm for still images. *The Journal of Engineering*, 2020. 2020(13), pp. 322-325.
- [6] Mohamed, M., A. El-Kilany, and N. El-Tazi, Future Activities Prediction Framework in Smart Homes Environment. *IEEE Access*, 2022. 10: p. 85154-85169.
- [7] Dang, L.M., et al., Sensor-based and vision-based human activity recognition: A comprehensive survey. *Pattern Recognition*, 2020. 108: p. 107561.
- [8] Santos, G.L., et al., Accelerometer-based human fall detection using convolutional neural networks. *Sensors*, 2019. 19(7), pp. 1644.
- [9] Jokanović, B. and M. Amin, Fall detection using deep learning in range-Doppler radars. *IEEE Transactions on Aerospace and Electronic Systems*, 2017. 54(1), pp. 180-189.
- [10] Kamel, A., et al., Deep convolutional neural networks for human action recognition using depth maps and postures. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2018. 49(9), pp. 1806-1819.
- [11] Khan, A., et al., A survey of the recent architectures of deep convolutional neural networks. *Artificial intelligence review*, 2020. 53(8), pp. 5455-5516.
- [12] Chen, T., Z. Ding, and B. Li, Elderly Fall Detection Based on Improved YOLOv5s Network. *IEEE Access*, 2022. 10, pp. 91273-91282.
- [13] Ajerla, D., S. Mahfuz, and F. Zulkernine, A real-time patient monitoring framework for fall detection. *Wireless Communications and Mobile Computing*, 2019. 2019.
- [14] Queralta, J.P., et al. Edge-AI in LoRa-based health monitoring: Fall detection system with fog computing and LSTM recurrent neural networks. in 2019 42nd international conference on telecommunications and signal processing (TSP). 2019. IEEE.
- [15] Wang, X. and K. Jia. Human fall detection algorithm based on YOLOv3. in 2020 IEEE 5th International Conference on Image, Vision and Computing (ICIVC). 2020. IEEE.
- [16] Waheed, M., H. Afzal, and K. Mehmood, NT-FDS—A Noise Tolerant Fall Detection System Using Deep Learning on Wearable Devices. *Sensors*, 2021. 21(6), pp. 2006.
- [17] Salimi, M., J.J. Machado, and J.M.R. Tavares, Using deep neural networks for human fall detection based on pose estimation. *Sensors*, 2022. 22(12): p. 4544.
- [18] Hatab, M., H. Malekmohamadi, and A. Amira. Surface defect detection using YOLO network. in Proceedings of SAI Intelligent Systems Conference. 2020. Springer.
- [19] Zhou, F., H. Zhao, and Z. Nie. Safety helmet detection based on YOLOv5. in 2021 IEEE International Conference on Power Electronics, Computer Applications (ICPECA). 2021. IEEE.
- [20] Dai, G., L. Hu, and J. Fan, DA-ActNN-YOLOV5: Hybrid YOLO v5 Model with Data Augmentation and Activation of Compression Mechanism for Potato Disease Identification. *Computational Intelligence and Neuroscience*, 2022. 2022.
- [21] Aburaed, N., et al. A Study on the Autonomous Detection of Impact Craters. in IAPR Workshop on Artificial Neural Networks in Pattern Recognition. 2023. Springer.
- [22] Ali, L., et al., Development of YOLOv5-Based Real-Time Smart Monitoring System for Increasing Lab Safety Awareness in Educational Institutions. *Sensors*, 2022. 22(22), pp. 8820.
- [23] Jeong, S., Kang, S. and Chun, I., 2019, June. Human-skeleton based fall-detection method using LSTM for manufacturing industries. In 2019 34th International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC) (pp. 1-4). IEEE.
- [24] Xu, Q., Huang, G., Yu, M. and Guo, Y., 2020. Fall prediction based on key points of human bones. *Physica A: Statistical Mechanics and its Applications*, 540, p.123205.
- [25] Chen, W., Jiang, Z., Guo, H. and Ni, X., 2020. Fall detection based on key points of human-skeleton using openpose. *Symmetry*, 12(5), p.744.
- [26] Ali, M.A., Hussain, A.J. and Sadiq, A.T., 2022. Human Fall Down Recognition Using Coordinates Key Points Skeleton. *International journal of online and biomedical engineering*, 18(2), pp.88-104.

Investigation of You Only Look Once Networks for Vision-based Small Object Detection

Li YANG¹

Department of Electronic Information Engineering, Leshan Vocational and Technical College
Leshan 614099, Sichuan, China

Abstract—Small object detection is a challenging issue in computer vision-based algorithms. Although various methods have been investigated for common objects including person, car and others, small object are not addressed in this issue. Therefore, it is necessary to conduct more researches on them. This paper is focused on small object detection especially jewellery as current object detection methods suffer from low accuracy in this domain. This paper introduces a new dataset whose images were taken by a web camera from a jewellery store and data augmentation procedure. It comprises three classes, namely, ring, earrings, and pendant. In view of the small target of jewellery and the real-time detection, this study adopted the You Only Look Once (Yolo) algorithms. Different Yolo based model including eight versions are implemented and train them using our dataset to address most effective one. Evaluation criteria, including accuracy, F1 score, recall, and mAP, are used to evaluate the performance of the various YOLOv5, YOLOv6, and YOLOv7 versions. According to the experimental findings, utilizing YOLOv6 is significantly superior to YOLOv7 and marginally superior to YOLOv5.

Keywords—YOLOv7; YOLOv6; YOLOv5; computer vision; jewellery detection; small object detection; real-time detection

I. INTRODUCTION

Target detection techniques based on deep learning have recently received much attention because of their strong generalizability, which has coincided with the growth of deep-learning theory and improved computer performance [1]. Convolutional neural networks dominate deep learning (CNN). Using the input picture as training data, the convolution network may successfully learn the key characteristics of the recognised item. Repeated training steadily boosts the training model's performance to provide accurate target detection outcomes [2].

Whether or not candidate areas are formed, the current popular object detection methods may be classified into two-stage algorithms and one-stage detection techniques. The RCNN [3], SPP (Space Pyramid Pooling)-Net [4], Fast-RCNN [5], Faster-RCNN [6], Mask-RCNN [7], and other algorithms are examples of the former. The latter primarily consists of the YOLO algorithm (YOLOv1 [8], YOLOv2 [9], YOLOv3 [10], YOLOv4 [11], YOLOv5 [12], etc.) and the SSD [13] algorithm. While two-stage algorithms may achieve high accuracy, the lengthy detection time makes it challenging to meet the real-time requirement in common object identification applications. The one-step detection method has taken centre stage in the study of object detection due to its benefits of high precision and quick speed. Early object identification systems,

such as YOLOv1, YOLOv2, etc., often had a network topology that consisted of a fully-connected layer on top of many convolution layers. The model's capacity to recognize several scales was significantly constrained by only calculating the feature map of a set size. In this case, the algorithm's detection accuracy for tiny objects could be more optimal [14].

Tiny/small object recognition is a difficult topic in computer vision, and several solutions have been put out to deal with it. Tiny/small object identification techniques have a long history that dates back to the early 2000s, a time when classic feature-based techniques were widely used. Later, with the development of deep learning, researchers started looking at object detection techniques based on deep neural networks. Faster R-CNN, YOLO, SSD, and RetinaNet, among other well-known object detection frameworks, were first developed for identifying large objects, but they have now been upgraded to handle small/tiny objects. With techniques like Yolo, EfficientDet, and Sparse R-CNN reaching cutting-edge performance on tiny/small object identification benchmarks, there has been tremendous research advancement in this field recently. By altering network topologies and incorporating and enhancing additional datasets, researchers attempt to enhance the outcomes for tiny object identification. Another clear answer to this problem is to improve the input image resolution, although it increases processing time [15].

In this paper, we introduce a new dataset whose images were taken by a web camera. The main objective of this study is to address an effective method to deal with small object detection challenge with generated dataset. In order to generate the dataset, data augmentation and YOLO are used to detect jewellery object as one of popular small objects. The reason to use Yolo based object detection in this study, because of their high efficiency in terms of accuracy rate and low computation complexity in object detection. In view of the small target of jewellery and the real-time detection, we chose YOLOv7 [16], YOLOv6 [17] and YOLOv5 models and compare their results.

In summary, this paper has the following contributions:

- 1) Generating a new dataset for tiny objects: We gathered about 2,500 photos from a webcam which includes three categories of jewellery (earrings, ring and pendant) and then augmented them to about 6,500 images.
- 2) Generating new Yolo-based model on different versions of YOLOv7, YOLOv6 and YOLOv5 algorithms.

3) Conducting extensive evaluations and analysis for the generated models and presented comprehensive performance comparison on the methods.

The rest of this paper consists as follows; Section II presents the background of the study. Section III discusses the research methodology. The results and analysis are presented in Section IV. Finally, his paper concludes in Section V.

II. BACKGROUND

In 2015, researcher Joseph Redmon and colleagues introduced YOLO algorithm. The YOLO series outperforms preceding versions in terms of speed and the ability to find tiny things [18]. It just takes one iteration of the algorithm to spread over a picture for the YOLO to detect things in real time [19]. In this study, various YOLO model iterations were used for both training and testing.

A. YOLOv5

A number of object recognition architectures that have already been trained using the MS COCO dataset are available in YOLOv5, which was released in 2020. Because of its quick speed and great accuracy, it is one of the most well-known detection algorithms. The photos are divided into a grid system by YOLOv5, and each grid cell is in charge of identifying items inside its own area. When several objects are present, this method offers a particular benefit.

The YOLOv5 is a deep learning-based platform available in five distinct iterations, ranging in size from the tiny YOLOv5 nano version, designed for mobile and embedded devices, to the gigantic YOLOv5x large version [18, 20, 22].

The author did not publish a detailed paper, but only launched a repository on Github and updated improvements there. Fig. 1 illustrates the backbone, neck, and head of the YOLOv5 architecture, which may be understood by analyzing its structural code [18, 21, 23]: Cross Stage Partial Networks (CSP) and the focal structure make up the backbone. The focus structure down samples the input data dimension while the

original data is kept. The model's capacity for learning is enhanced, and its memory use is decreased, thanks to the CSP Network's ability to extract important information.

The neck part combines the acquired characteristics and transmits them to the prediction layer using Feature Pyramid Networks (FPN) and the Path Aggregation Network (PAN). The FPN up samples the high-level feature data via top-to-bottom communication and prediction fusion. The underlying pyramid, PAN, communicates important positioning properties top-to-bottom, aiding in the distinguishing of similar objects of various scales and sizes.

The final output vectors, consisting of bounding boxes, class probabilities, and object scores, are provided by the output layer's head by applying anchor boxes to the features. Including the focus and CSP layers is the primary change in YOLOv5. The focusing layer decreases layers, parameters, FLOPS, and CUDA memory to increase forward and reverse speeds. The backbone layer's CSP layer tries to extract specific data and carry out more extensive activities. In YOLOv5, the meshing ideas from the original YOLO algorithm have been retained [18].

B. YOLOv6

The single-stage object detection framework for industrial applications, MT-YOLOv6 (created by the Meituan firm, hence the prefix "MT"), has a hardware-friendly, effective design and excellent performance [24].

Numerous upgrades to the Backbone, Neck, and Head blocks as well as training methods are included in YOLOv6. For instance, utilizing Rep-PAN and EfficientRep structures, respectively, the Neck and Backbone in YOLOv6 have been redesigned in accordance with the notion of a hardware-aware neural network. Strong representational capabilities are combined with hardware computational capability, such as GPU, in the EfficientRep Backbone. Rep-Pan Neck outperforms PANet and SPP in terms of accuracy and speed.

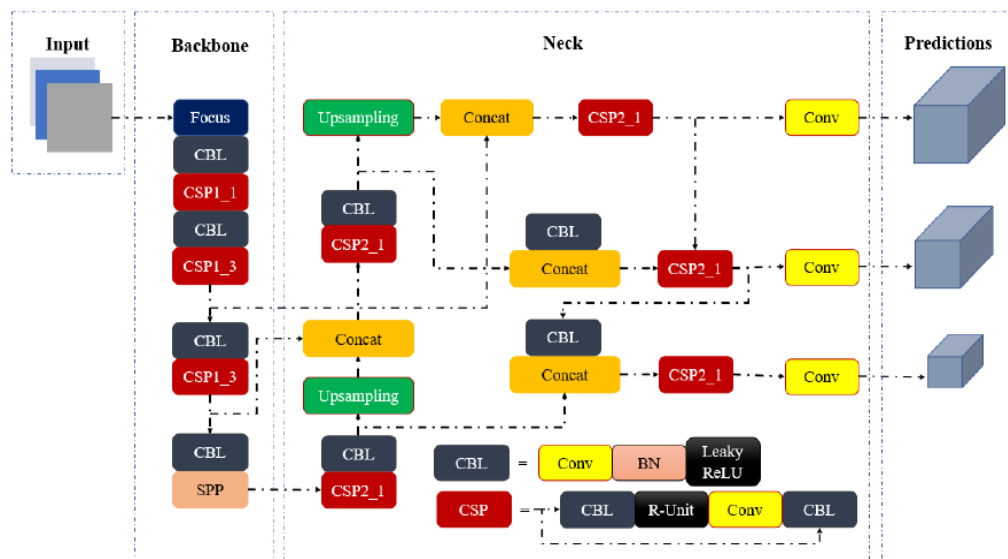


Fig. 1. Network structure of YOLOv5 [18].

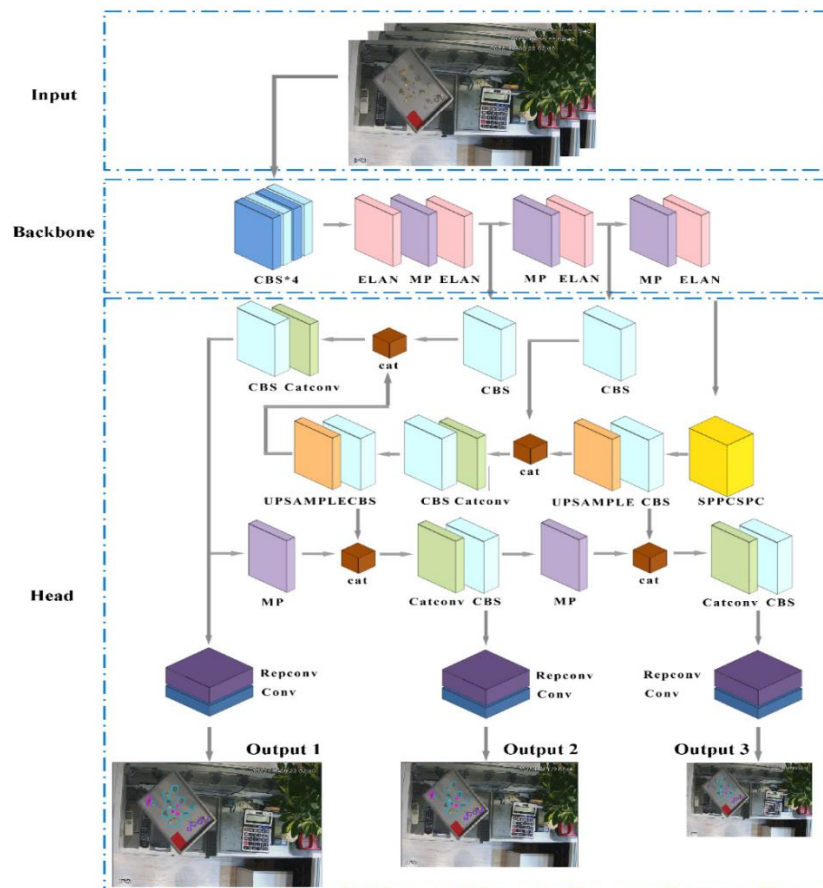


Fig. 3. Network structure of YOLOv7 [30].

In [32] YOLOv5 network modifications were made for aerial small target detection. By utilizing the first effective channel attention module, they altered the backbone, and the channel attention pyramid approach was suggested. Consequently, the module for identifying large items was removed in order to improve the identification of tiny things, and a detect layer was introduced to look for smaller objects. Finally, transposed convolution was used to produce upsampling rather than the already used closest neighbor interpolation. With the suggested technique, the mAP for the VEDAI dataset was 6.9%, for the xView dataset it was 6.7%, for the DOTA dataset it was 2.7%, and for the Arirang dataset it was around 2.4% for the small car class.

To address the problem of detecting small fish [33] presented a YOLOv5-based model. It tries to address the issues of inaccurate location and insufficient information for detecting underwater targets. First, they proposed combining the attention mechanisms CA and C3 structure to increase the network's ability to retrieve crucial information. Next, they suggested expanding the YOLOv5's three detection layers to four in order to address the issue of numerous, intensive detection tasks. Finally, in order to improve convergence time and lessen erroneous regression findings, GIOU loss was used in place of EIOU loss. The experimental findings demonstrated that the revised algorithm affected various indicators differently; mAP@0.50 achieved 94.9%, which was more accurate. The detection effect was 24.6 percent greater, and

there were 248 picture detections overall—49 more than with YOLOv5. Performance for target detection was enhanced. Poor underwater fish swarm detection, tiny target location, few pixels, and low accuracy issues are all resolved.

In [31] some generic work for detecting tiny objects was done. They showed that one of the reasons for the poor average accuracy for small objects is the need for more representation of small things in training data. This is especially true for today's most advanced object detectors, which require a large enough training set of objects to guarantee that the predicted anchors match. They suggested two ways to improve the initial MS COCO database in order to resolve the problem. They first demonstrated how oversampling photos with small items during training may easily enhance performance on small objects. Second, they recommended an improved method based on pasting microscopic items. Compared to the state-of-the-art achieved by Mask R-CNN on MS COCO, their trials showed a 9.7% relative improvement, for instance, segmentation and a 7.1% relative improvement for object detection for tiny objects. The collection of augmentation techniques that have been suggested provides a trade-off between the accuracy of predictions for small and big objects.

In [34], their technique produced counting dense flocks of hemp ducks using positive detection findings. To improve the network structure of the YOLOv7 algorithm, three CBAM modules were added to the backbone network. SE-YOLOv7 and ECA-YOLOv7 were introduced for comparison studies

along with an updated YOLOv7 algorithm that includes an attention mechanism. As a consequence of the experimental findings, CBAM-YOLOv7 was shown to have greater accuracy, as well as somewhat enhanced recall, mAP@0.5, and mAP@0.5:0.95 values. There was no change in the computational demand, and the FLOPS only slightly increased by 0.02 G. They also provided two labelling methods: whole-body labelling and head-only tagging, taking into consideration the overlap problem with hemp duck labelling frames. The full-body frame labelling approach showed a greater detection effect, whereas the head-only labelling method resulted in the loss of a significant amount of feature information.

2) *Jewellery detection*: Although there are not many articles available for jewellery recognition work, especially jewellery in store, some similar works can be mentioned.

Images of jewellery have been categorized using a machine-learning method [35]. They employed various methods. The first method takes advantage of the characteristics of AlexNet's support vector machine and support vector machine extracted from the input pictures. The Inception-v3 model is used in the second technique to carry out the same task. The results of the experiments showed that both techniques worked well, although Inception-v3 had a 0.9% higher success rate. In order to improve consistency, the Inception-v3 was then used to train the dataset from scratch. SVM has a 98.30% accuracy rating compared to 99.19% for Inception-v3.

In [36], the author focuses on picture recognition methods for automatically classifying stone-grinding flaws. After that, the stone quality is classified using the computed pertinent image attributes. A binary decision tree-based technique that uses decision thresholds modified from a training dataset does classification. In the end, the accuracy and time complexity of the proposed method are compared with more than twenty cutting-edge machine learning algorithms, and the results are competitive: on the D1 dataset, the best accuracy was reached among all tested algorithms; on the D2 dataset, 7% poorer prediction than the best algorithm was achieved. In addition, the algorithm consistently outperformed all others in all tests.

A novel technique for counting pearls was presented in [37]. The model is composed of an approach to counting and object detection. After a thorough investigation, they examine the key performance metrics of nine object detection algorithms. The results demonstrate that pearl identification can be accomplished using Faster R-CNN with ResNet152, which was pretrained on the pearl dataset, and only needs 15.8 ms of inference time with a counter following the initial loading of the model. Additionally, performance for precise counting and pearl recognition of natural or artificial light is promising. Additionally, the network obtained 100% accuracy in counting pearls and a recall/AR@100 (medium) of 95% for pearl identification.

In [38], they generated a model for automated coin identification and recognition. In contrast to image processing techniques, which focus on the extraction of color, shape, and edge information, the majority of coin identification systems

now in use are based on the physical characteristics of the coins. They have suggested a deep learning strategy for the recognition and detection of Indian coinage. AlexNet, a convolutional neural network that has already been trained, is trained using features including textures, colors, and forms. More than 1600 photos were used to train the model. They used a pre-assembled collection of photos to train AlexNet in their trials, which proved that there was sufficient training data. Results obtained demonstrated that the suggested technique outperformed more established methods.

III. METHODOLOGY

YOLO was developed as a pre-trained object detector that can identify common items, including tables, chairs, automobiles, phones, and more [39]. We suggest a detection approach based on YOLO techniques to create a model that could recognize jewellery. Also useful in real-time applications are our models.

A. Dataset

Our dataset consists of three various categories of jewellery (earrings, ring and pendant). The dataset includes images taken from our webcam. We collected our photos from a jewelry store webcam that was fixed in place. Fig. 4 shows some examples of our dataset. The dataset contains different image sizes including small target objects, which is more challenging to detect.

We selected photos of different types of shapes, sizes, resolutions, angles and different numbers of samples in each image. Using Roboflow, the 2456 photos from the dataset were increased to roughly 6500 images in order to improve our model. The following random processing steps were applied to each image: horizontal flip, rotation (between -15° and $+15^\circ$), saturation (between -20% and $+20\%$), exposure (between -10% and $+10\%$), and brightness (between -20% and $+20\%$). A maximum of three enhanced versions were produced for each image. In Fig. 5, examples of enhanced pictures are displayed. A training set (93%), a validation set (6%), and a test set (1%), each comprised of the labelled images, are then created.

B. Google Colab

Utilizing Google Colab, which offers free usage of powerful GPUs, was helpful. A 12GB NVIDIA Tesla T4 GPU, shown in greater detail in Fig. 6, is used for all training and testing workloads. Our whole model was trained with an image size of 640 pixels and 50 training iterations. Other hyperparameters were adjusted using the YOLO default settings.

C. Transfer Learning

It is always a good idea to begin the training process for an object detector with a model that has already been built using weights from extremely sizable datasets. Even if the training weights do not have the test items, this is OK. Transfer learning refers to this procedure. In order to help the network learn more quickly, a pre-trained model that uses weights from the COCO dataset as its initial weights is employed. Additionally advantageous is the fact that fewer data will be needed [39].

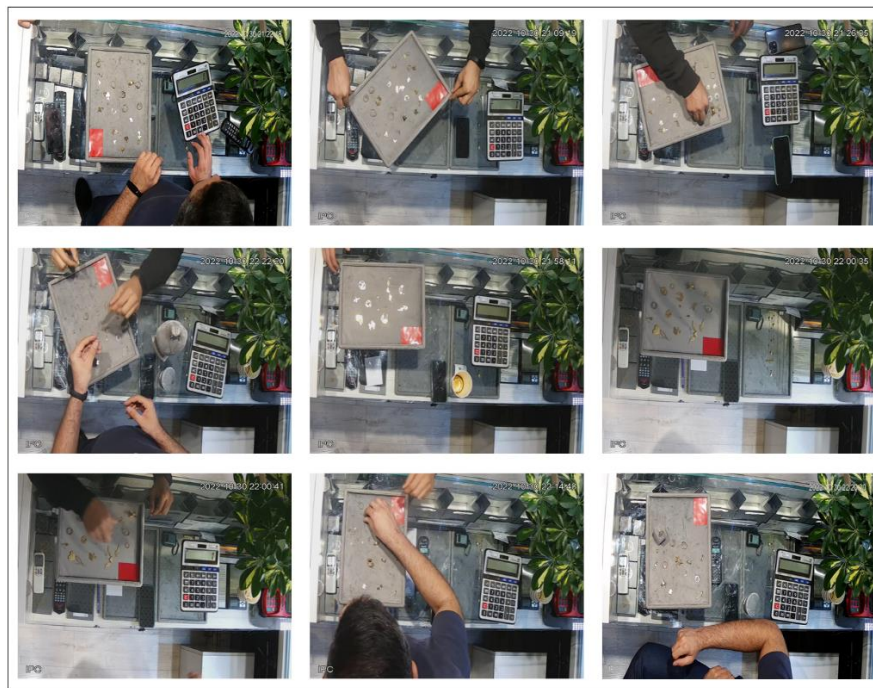


Fig. 4. Sample images from dataset.



Fig. 5. Sample images of augmented images.

```

+-----+-----+-----+
| NVIDIA-SMI 460.32.03   Driver Version: 460.32.03   CUDA Version: 11.2   |
+-----+-----+-----+
| GPU Name      Persistence-M| Bus-Id        Disp.A | Volatile Uncorr. ECC | |
| Fan  Temp  Perf  Pwr:Usage/Cap|      Memory-Usage | GPU-Util  Compute M. |
|               |                  |              |   MIG M.   |
+-----+-----+-----+
|   0   Tesla T4             Off   | 00000000:00:04:0  Off   |    0%        0      | |
| N/A   63C   P8             11W / 70W |  0MiB / 15109MiB |             Default  |
|               |                  |              |   N/A   |
+-----+-----+-----+

```

Fig. 6. Details of Google Colab GPU.

IV. RESULTS AND ANALYSIS

In this section, we introduce the experiment's details, and then we show the training results using pretraining weights.

A. Model Evaluation

In the proposed study, the gathered experimental findings were compared using a variety of assessment metrics, including accuracy, recall, F1-score, confusion matrix, and mean average precision (mAP). The true positive rate, or TPR, indicates how probable it is that things from the real world will be correctly identified. When a model produces no false negatives, which indicates that there are no bounding boxes that are not recognized but ought to be detected, it has a high recall. Eq. (1) provides the mathematical model for the recall [18]:

$$R = \frac{TP}{TP+FN} = \frac{TP}{\text{Total Ground Truths}} \quad (1)$$

The real positive and false negative are denoted in the equation above by the letters TP and FN, respectively. Eq. (2) defines precision as the percentage of correctly anticipated positives, often known as the positive predictive value. The exact model creates no false positives and only detects important things (FP) [18].

$$P = \frac{TP}{TP+FP} = \frac{TP}{\text{Total Predictions}} \quad (2)$$

The F-1 score, as stated in Eq., is the harmonic mean of the accuracy and recall scores (3).

$$F - 1 = 2 * \frac{P * R}{P + R} \quad (3)$$

Eq. (4) defines AP as the area under the PR curve and defines mAP as the average of all AP values across all classes/categories.

$$mAP = \frac{1}{n} \sum_{i=1}^n AP_i \quad (4)$$

where n is the number of classes [18].

We trained YOLOv5n, YOLOv5s, YOLOv5m, YOLOv6t, YOLOv6n, YOLOv6s, YOLOv7tiny and YOLOv7. In Fig. 7 and 8, we have shown F1-confidence and precision-recall curves respectfully. Correspondingly, the results of validation's mAP, recall and precision are shown in Table 1. In addition, we put the training time and parameter values of each model for better evaluation.

B. Model Losses

The results of loss functions of models are shown in Fig. 9 to 11:

C. Analysis

We trained our model with YOLOv5n, YOLOv5s, YOLOv5m, YOLOv6n, YOLOv6t, YOLOv6s, YOLOv7tiny and YOLOv7. Fig. 12 to 14 provide examples of model predictions for the fresh and undiscovered images. Although YOLOv7 is the newest version and was launched in July 2022, it is allegedly better at object recognition than YOLOv5 and YOLOv6. All models that were trained on our jewellery dataset showed promising performance. However, we explored first YOLOv6 and next YOLOv5 are performing better than YOLOv7, and both YOLOv7tiny and YOLOv7 training accuracies were far worse on our dataset.

Due to their recent publication, the poor accuracies of the current YOLO versions may have resulted from insufficient experimentation, tweaking, and correction [40].

As shown by the results, the smallest model YOLOv5n achieved mAP@0.5 of 0.865 and the largest model, YOLOv7 achieved mAP@0.5 of 0.60. YOLOv6s model almost outperformed all other models and YOLOv6n is in second place. The worst result is for YOLOv7tiny except for training time which is the lowest with 2.393 hours. In the next place, the YOLOv7 has the worst performance. Due to the number of their parameters, YOLOv7 took the longest, clocking in at 4.518 hours. The detection speed of the model decreases as the number of parameters increases.

Another problem that should be investigated is that in all YOLOv5 models, the number of undetected objects is higher. It is feasible that expanding the dataset will boost the model's precision.

D. Additional Analysis

We analyzed two other cases. One is to change the Augmentation strategy and the second is to train with more epochs. We augmented our images as described in Section III A with RoboFlow. We also know that there are techniques to augment data in the structure of YOLO algorithms. For further investigation, we once removed all data augmentation techniques in the YOLOv5s and YOLOv6s algorithms. And next time we only removed the mosaic and mixup techniques. The difference in accuracy and speed between the original models with the modified models is shown in Table II. Interestingly, we noticed that when we remove the three data augmentation, results in YOLOv5s increases slightly and in YOLOv6s is almost equal. But in both models the training time decreases by a large margin.

Finally, to evaluate training more epochs, we trained the YOLOv6s algorithm with 100 epochs, the results compared to 50 epochs are in shown Table III. It shows that training with more epochs is slightly better. Fig. 12to 14 show the experimental results for Yolov5, 6 and 7.

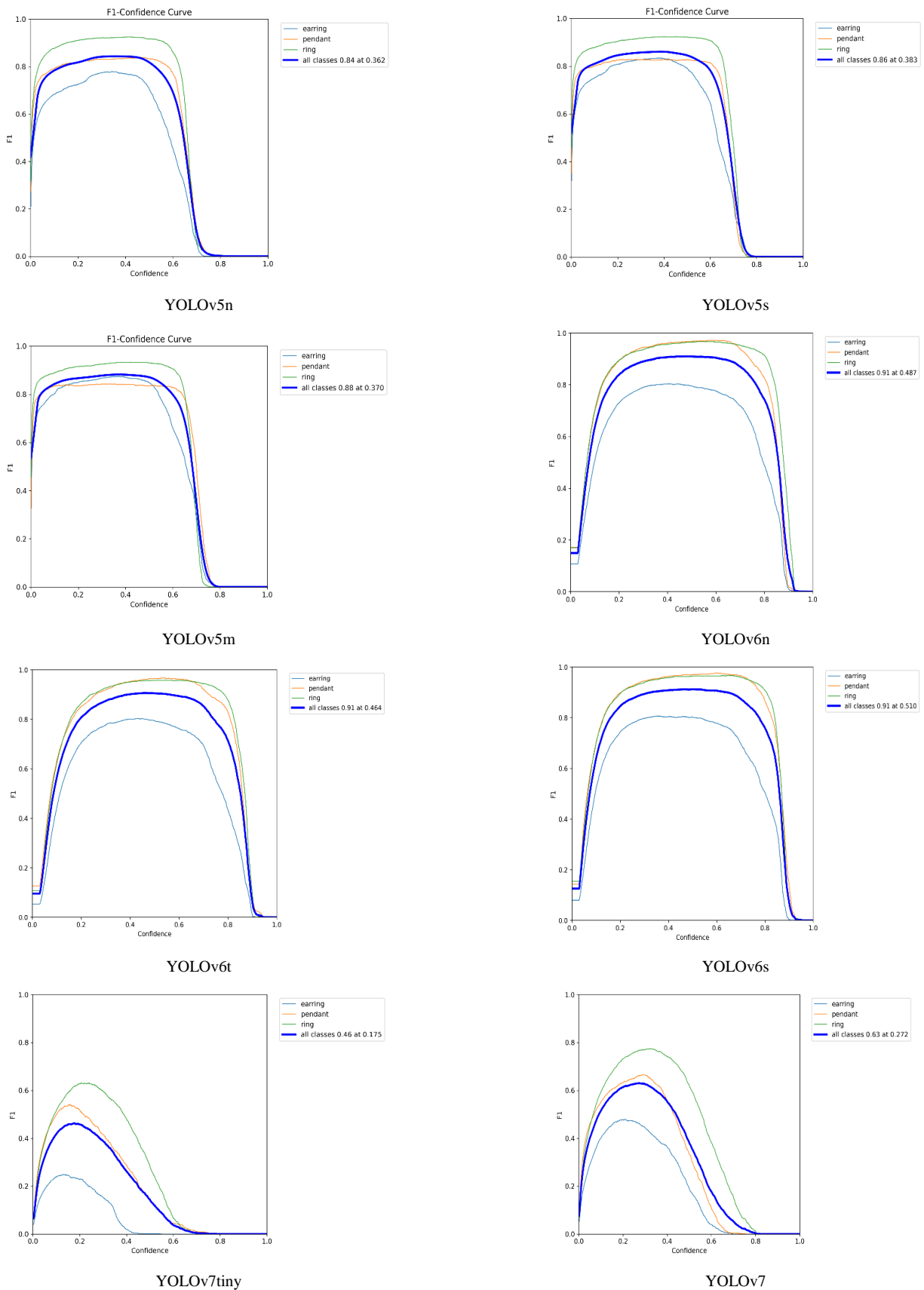


Fig. 7. F1-Confidence curves.

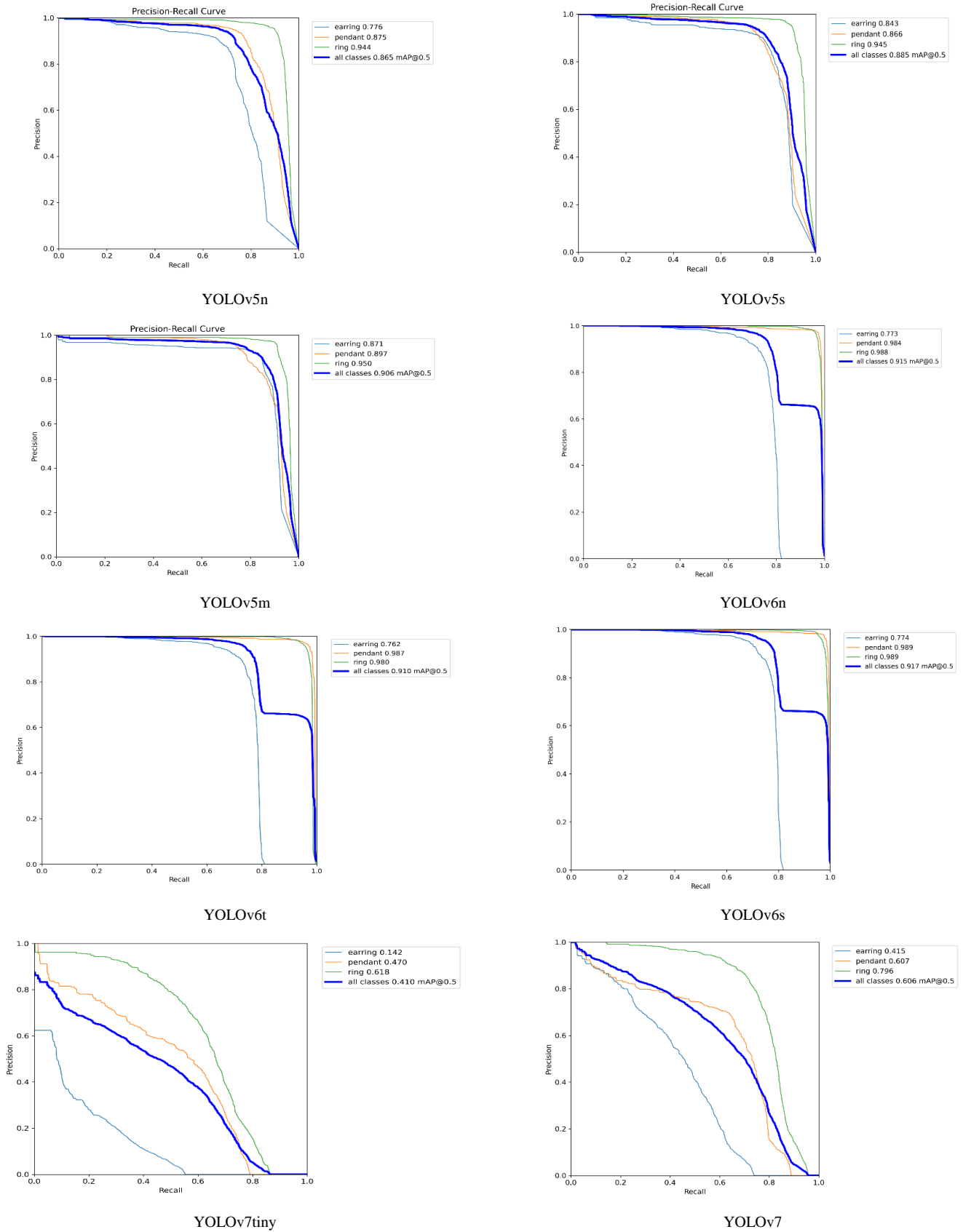


Fig. 8. Precision-Recall curves.

TABLE I. RESULT OF MAP, RECALL, PRECISION AND TRAINING TIME

	<i>MAP@0.5</i>	<i>MAP@0.5:0.95</i>	<i>Precision</i>	<i>Recall</i>	<i>Params(M)</i>	<i>Trainig Time(hour)</i>
YOLOv5n	0.865	0.547	0.902	0.794	1.76	4.206
YOLOv5s	0.885	0.585	0.913	0.817	7.01	4.431
YOLOv5m	0.906	0.615	0.932	0.84	20.86	4.404
YOLOv6n	0.915	0.598	0.946	0.883	4.30	3.670
YOLOv6t	0.91	0.595	0.936	0.884	9.67	3.823
YOLOv6s	0.917	0.61	0.955	0.881	17.19	3.726
YOLOv7tiny	0.41	0.14	0.48	0.45	6.01	2.393
YOLOv7	0.60	0.24	0.694	0.59	37.2	4.518

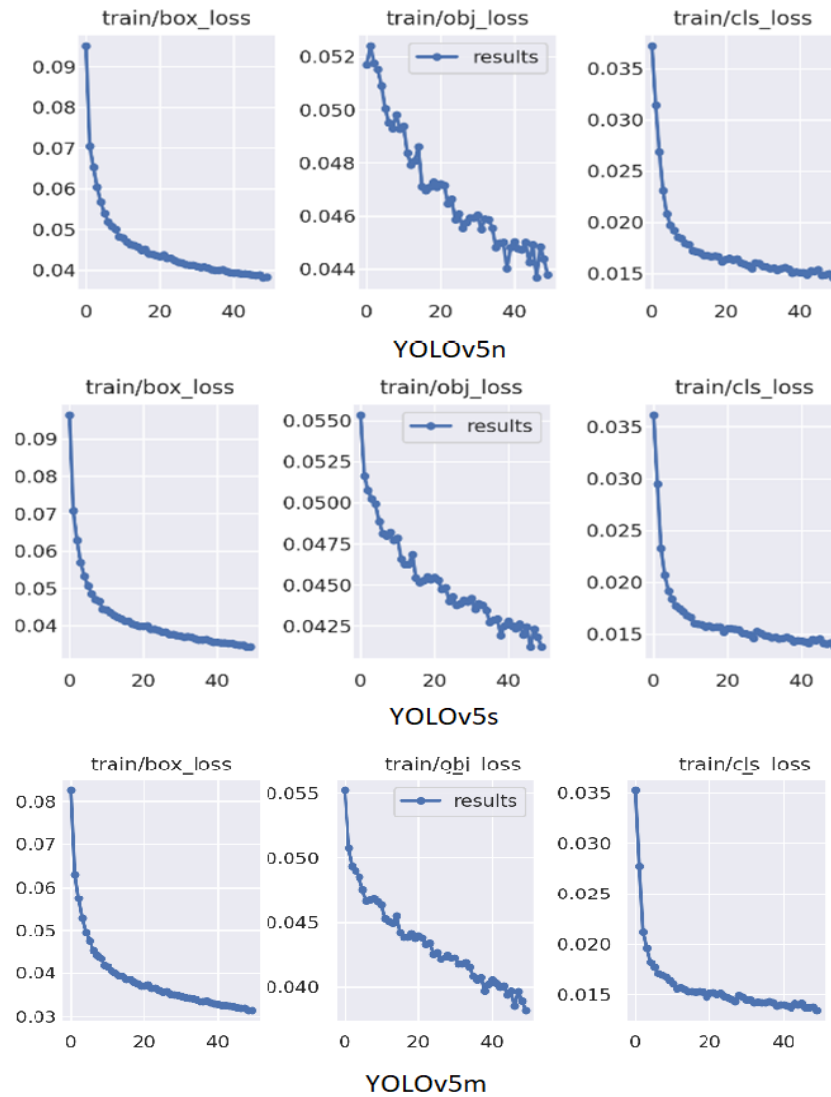


Fig. 9. Model Losses of YOLOv5.

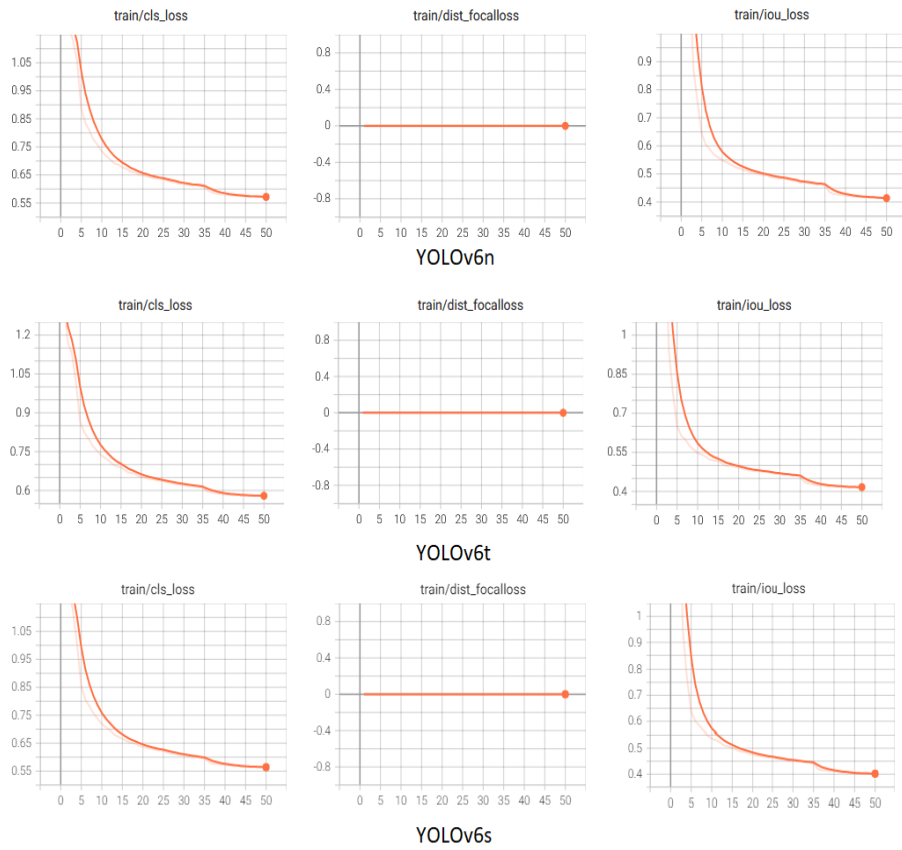


Fig. 10. Model Losses of YOLOv6.

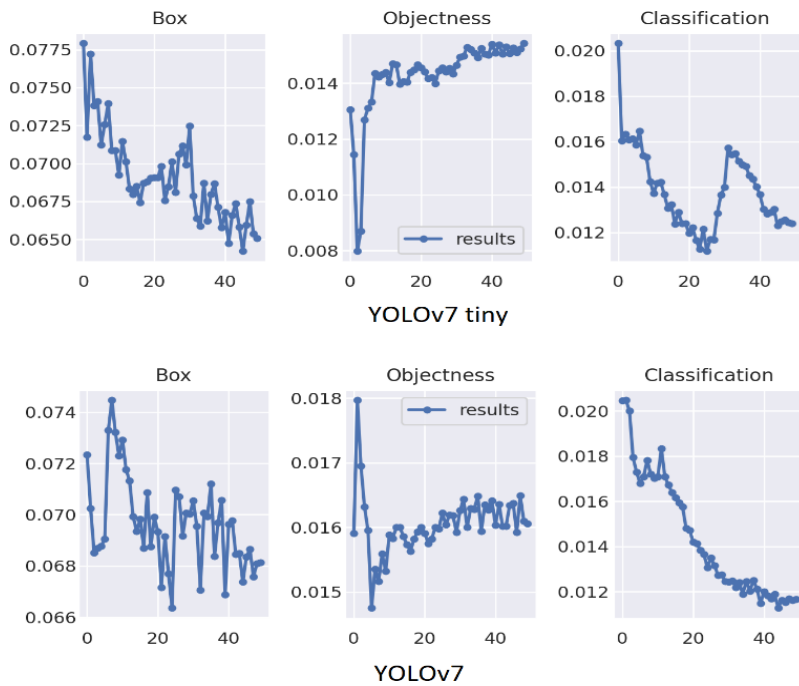


Fig. 11. Model Losses of YOLOv7.



Fig. 12. Samples for detection with YOLOv5.



Fig. 13. Samples for detection with YOLOv6.

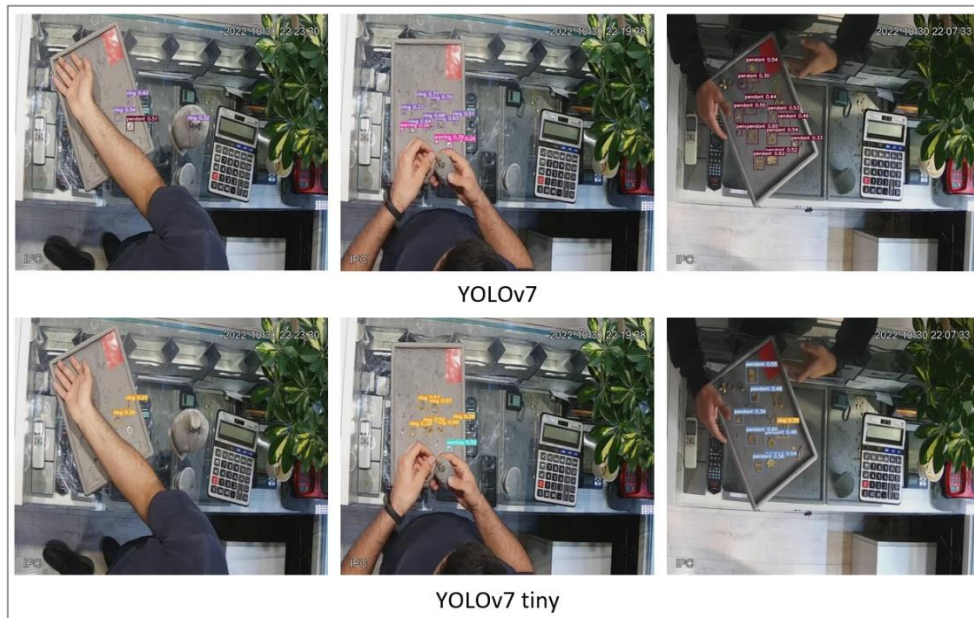


Fig. 14. Samples for detection with YOLOv7.

TABLE II. RESULTS OF AUGMENTATION ABLATION. THE “ORIGINAL” MODEL IS THE SAME AS THE PREVIOUS SECTIONS WITHOUT CHANGES AND WITH ALL THE AUGMENTATIONS. “W/O ALL” DENOTES WITHOUT ALL YOLO AUGMENTATION TECHNIQUES. AND “W/O 3” DENOTES WITHOUT JUST MOSAIC, MIXUP AND COPY_PASTE YOLO AUGMENTATION TECHNIQUES

	<i>MAP@0.5</i>	<i>MAP@0.5:0.95</i>	<i>Precision</i>	<i>Recall</i>	<i>Trainig Time(hour)</i>
YOLOv5s (Original)	0.885	0.585	0.913	0.817	4.431
YOLOv5s (W/O All)	0.885	0.535	0.901	0.838	1.734
YOLOv5s (W/O 3)	0.905	0.62	0.915	0.839	2.117
YOLOv6s (Original)	0.917	0.61	0.955	0.881	3.726
YOLOv6s (W/O All)	0.893	0.505	0.935	0.869	2.129
YOLOv6s (W/O 3)	0.916	0.612	0.944	0.893	2.407

TABLE III. RESULTS OF INCREASING EPOCHS

	<i>MAP@0.5</i>	<i>MAP@0.5:0.95</i>	<i>Precision</i>	<i>Recall</i>	<i>Trainig Time(hour)</i>
YOLOv6s (50 epochs)	0.917	0.61	0.955	0.881	3.726
YOLOv6s (100 epochs)	0.934	0.645	0.5	0.906	7.893

V. CONCLUSION

In this work firstly a dataset consisting of 6k images of three classes, i.e., earrings, ring and pendant was created. Our photos were taken from a webcam fixed in a jewellery store. It has been tried that the photos be in different lighting conditions and angles as well as with different qualities. We have also used the benefits of data augmentation for our dataset. The performance of several YOLOv5, YOLOv6, and YOLOv7 variations was then compared based on the algorithm's accuracy, recall, mAP, and training time. YOLOv6s outperformed other algorithms for identifying jewellery, achieving the greatest mAP@0.5, of 0.917, and mAP@0.5:0.95, of 0.6, according to the results. The fastest algorithm was YOLOv7tiny with 2.393 hours of training time for 50 epochs. Also, we analyzed the model's augmentation techniques and training with more epochs. The proposed method in this study has advantages compared to other existing

methods because of high accuracy rate and low computation complexity. For future study, the results of the work show that we have reached good accuracy, but there is more work to be done, especially for the YOLOv7 algorithms, which can be compensated for by further optimization and investigation and the execution of more epochs. Another direction for future study on tiny/small object detection is to explore the use of attention mechanisms in deep neural networks. The use of attention mechanisms and generative models could help improve the performance of tiny/small object detection models and lead to better results on real-world applications.

REFERENCES

- [1] Zheng, J., et al., Insulator-Defect Detection Algorithm Based on Improved YOLOv7. Sensors, 2022. 22(22), pp. 8801.
- [2] Xianbao, C., et al., An improved small object detection method based on Yolo V3. Pattern Analysis and Applications, 2021. 24(3), pp. 1347-1355.

- [3] Girshick, R., et al. Rich feature hierarchies for accurate object detection and semantic segmentation. in Proceedings of the IEEE conference on computer vision and pattern recognition. 2014.
- [4] He, K., et al., Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE transactions on pattern analysis and machine intelligence, 2015. 37(9), pp. 1904-1916.
- [5] Girshick, R. Fast r-cnn. in Proceedings of the IEEE international conference on computer vision. 2015.
- [6] Ren, S., et al., Faster r-cnn: Towards real-time object detection with region proposal networks. Advances in neural information processing systems, 2015. 28.
- [7] He, K., et al. Mask r-cnn. in Proceedings of the IEEE international conference on computer vision. 2017.
- [8] Redmon, J., et al. You only look once: Unified, real-time object detection. in Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.
- [9] Redmon, J. and A. Farhadi. YOLO9000: better, faster, stronger. in Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.
- [10] Redmon, J. and A. Farhadi, Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767, 2018.
- [11] Bochkovskiy, A., C.-Y. Wang, and H.-Y.M. Liao, Yolov4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934, 2020.
- [12] J, G. Ultralytics/yolov5 Available online: <https://github.com/ultralytics/yolov5/releases/tag/v6.1>. 2022.
- [13] Liu, W., et al. Ssd: Single shot multibox detector. in European conference on computer vision. 2016. Springer.
- [14] Liu, H., et al., SF-YOLOv5: A Lightweight Small Object Detection Algorithm Based on Improved Feature Fusion Mode. Sensors, 2022. 22(15), pp. 5817.
- [15] Benjumea, A., et al., YOLO-Z: Improving small object detection in YOLOv5 for autonomous vehicles. arXiv preprint arXiv:2112.11798, 2021.
- [16] Wang, C.-Y., A. Bochkovskiy, and H.-Y.M. Liao, YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. arXiv preprint arXiv:2207.02696, 2022.
- [17] Li, C., et al., YOLOv6: a single-stage object detection framework for industrial applications. arXiv preprint arXiv:2209.02976, 2022.
- [18] Ali, L., et al., Development of YOLOv5-Based Real-Time Smart Monitoring System for Increasing Lab Safety Awareness in Educational Institutions. Sensors, 2022. 22(22), pp. 8820.
- [19] Conley, G., et al., Using a deep learning model to quantify trash accumulation for cleaner urban stormwater. Computers, Environment and Urban Systems, 2022. 93, pp. 101752.
- [20] Ahmad, T., et al., Detecting Human Actions in Drone Images Using YoloV5 and Stochastic Gradient Boosting. Sensors, 2022. 22(18), pp. 7020.
- [21] Thuan, D., Evolution of Yolo algorithm and Yolov5: The State-of-the-Art object detection algorithm. 2021.
- [22] Yang, S.J., et al., Assessing microscope image focus quality with deep learning. BMC bioinformatics, 2018. 19(1), pp. 1-9.
- [23] Guo, Y., et al., Improved YOLOV4-CSP Algorithm for Detection of Bamboo Surface Sliver Defects With Extreme Aspect Ratio. IEEE Access, 2022. 10, pp. 29810-29820.
- [24] Horvat, M. and G. Gledec, A comparative study of YOLOv5 models performance for image localization and classification.
- [25] Aburaed, N., et al. A Study on the Autonomous Detection of Impact Craters. in IAPR Workshop on Artificial Neural Networks in Pattern Recognition. 2023. Springer.
- [26] Yun, J.-S., S.-H. Park, and S.B. Yoo, Infusion-Net: Inter-and Intra-Weighted Cross-Fusion Network for Multispectral Object Detection. Mathematics, 2022. 10(21), pp. 3966.
- [27] He, K., et al. Deep residual learning for image recognition. in Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.
- [28] Huang, G., et al. Densely connected convolutional networks. in Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.
- [29] Ding, X., et al. Repvgg: Making vgg-style convnets great again. in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021.
- [30] Wu, D., et al., Detection of Camellia oleifera Fruit in Complex Scenes by Using YOLOv7 and Data Augmentation. Applied Sciences, 2022. 12(22), pp. 11318.
- [31] Kisantal, M., et al., Augmentation for small object detection. arXiv preprint arXiv:1902.07296, 2019.
- [32] Kim, M., J. Jeong, and S. Kim, ECAP-YOLO: Efficient Channel Attention Pyramid YOLO for Small Object Detection in Aerial Image. Remote Sensing, 2021. 13(23), pp. 4851.
- [33] Li, J., et al., CME-YOLOv5: An Efficient Object Detection Network for Densely Spaced Fish and Small Targets. Water, 2022. 14(15), pp. 2412.
- [34] Jiang, K., et al., An Attention Mechanism-Improved YOLOv7 Object Detection Algorithm for Hemp Duck Count Estimation. Agriculture, 2022. 12(10), pp. 1659.
- [35] Singh, V. and P. Kaewprapha, A comparative experiment in classifying jewelry images using convolutional neural networks. Science & Technology Asia, 2018, pp. 7-17.
- [36] Hurtik, P., M. Burda, and I. Perfilieva. An image recognition approach to classification of jewelry stone defects. in 2013 Joint IFSA World Congress and NAFIPS Annual Meeting (IFSA/NAFIPS). 2013. IEEE.
- [37] Hou, M., et al., PDC: Pearl Detection with a Counter Based on Deep Learning. Sensors, 2022. 22(18), pp. 7026.
- [38] Tajane, A., et al. Deep learning based indian currency coin recognition. in 2018 International Conference On Advances in Communication and Computing Technology (ICACCT). 2018. IEEE.
- [39] Hatab, M., H. Malekmohamadi, and A. Amira. Surface defect detection using YOLO network. in Proceedings of SAI Intelligent Systems Conference. 2020. Springer.
- [40] Chen, E., et al., Real-time detection of acute lymphoblastic leukemia cells using deep learning. bioRxiv, 2022.

A Novel Data Aggregation Method for Underwater Wireless Sensor Networks using Ant Colony Optimization Algorithm

Lianchao Zhang¹, Jianwei Qi^{2*}, Hao Wu³

Tangshan Maritime Institute Tangshan
Hebei, 063000, China^{1, 2}

ShiJiaZhuang University of Applied Technology, Shijiazhuang
Hebei, 050081, China³

Abstract—Underwater Wireless Sensor Networks (UWSNs) have a wide range of applications for monitoring the ocean and exploring the offshore environment. Sensor nodes are typically dispersed throughout the area of interest at different depths in these networks. Sensor nodes on the seabed must use a routing protocol in order to communicate with surface-level nodes. The suitability assessment considers network resources, application requirements, and environmental factors. By combining these factors, a platform for resource-aware routing strategies can be created that meet the needs of different applications in dynamic environments. Numerous challenges and problems are associated with UWSNs, including the lack of battery power, instability of topologies, a limited bandwidth, long propagation times, and interference from the ocean. These problems can be addressed through the design of routing protocols. The routing protocol facilitates the transfer of data between source and destination nodes. Data aggregation and UWSN protocols are widely used to achieve better outcomes. This paper describes an energy-aware algorithm for data aggregation in UWSNs that uses the improved ACO (Ant Colony Optimization) algorithm to maximize the packet delivery ratio, improve the network lifetime, decrease end-to-end delay, and use less energy.

Keywords—UWSNs; routing; data aggregation; energy efficiency; ant colony optimization algorithm

I. INTRODUCTION

During the past few years, wireless and emerging technologies have experienced significant advancements, especially the Internet of Things (IoT) [1, 2], Wireless Sensor Networks (WSNs) [3], artificial intelligence [4, 5], machine learning [6-8], smart grids [9], Blockchain [10], 5G connectivity [11, 12], and cloud computing [13], all of which have proven to be beneficial to society in a number of ways. Underwater WSNs (UWSNs) allow devices to receive, process, and communicate embedded in water for monitoring and exploration at different depths [14]. These devices are equipped with sensors that send information to a surface station after being received from the underwater environment [15]. The data are then processed based on the requirements of the application. The creation of UWSNs has been inspired by several factors, including the study of geological processes on the ocean's surface, the identification of mines, predicting and visualizing climate change, the analysis of the human impact on marine ecosystems, the discovery of areas containing

underwater oil, the prevention of accidents, the tracking of mammals, fish, and other microorganisms, as well as the protection of water borders against invaders [16-18].

Sensor networks used for underwater communications differ from those used for wired communications. Underwater networks must be able to withstand extreme pressure and temperature changes. They also require specialized communication protocols to account for the longer transmission times and greater signal loss associated with underwater transmission. Additionally, acoustic waves are used for underwater communications, as radio waves cannot propagate through water. As a first point, energy consumption varies concerning the sort of application. Underwater networks must also be designed to be energy-efficient to ensure long-term operation. The second characteristic of such networks is that they typically work towards shared objectives rather than representing specific individuals. Maximizing throughput instead of ensuring relative fairness between nodes is the goal. Thirdly, the number of hops, link distance, and reliability are correlated in these networks. Several short hops rather than a single long hop are used in underwater networks; hence multi-hop data delivery uses less energy than single-hop data delivery. The end-to-end reliability of packet routing is, however, affected by many hops, especially when the environment is harsh underwater. As a final factor, individual companies using inexpensive equipment typically install these networks, so strict interoperability is unnecessary [19]. The subsurface environment may complicate UWSN design. Host conditions pose significant challenges in terms of node movement and 3D topology. Some underwater applications, such as detection and rescue missions, require ad hoc deployment, which is often unplanned and needs networks to be deployed in short timeframes [20].

UWSNs have attracted researchers' attention thanks to their wide range of applications in several sectors. Monitoring environmental conditions, oil and gas extraction under the sea, surveillance of military operations, smart farming, and communication are some of the many applications of UWSN [21]. UWSNs suffer from several significant issues, including highly energy-consuming, insufficient processing power of nodes, and short-term lifespans in routing protocols [22]. Therefore, it is a research challenge to reduce energy

*Corresponding Author.

consumption and processing in UWSNs to extend their lifetimes. There are two main types of data aggregation in UWSNs: unstructured and structured [23]. The amorphous category has no predefined structure for data gathering, so local information is utilized to aggregate data. Structured data aggregation networks have a defined structure. Structured data aggregation is classified into two categories: flat and hierarchical. A flat structure has objects that contribute to sensing and perform similar functions. Hierarchical approaches typically organize sensor nodes logically and enable them to communicate securely. There are three primary methods of implementing hierarchical structures: centralized, clustered, and tree-based [24].

The ACO algorithm for aggregating data from UWSNs is improved in this work. The first step is a description of network and optimization models. Second, an enhanced ACO algorithm is presented based on improvements to heuristic information, pheromone update procedures, and evaporation parameters. To improve the heuristic data of the conventional ACO algorithm, we provide heuristic information that considers distance and remaining energy. Also, the suggested adaptive scheme for the evaporation parameter enables the algorithm to perform a global search and reach a higher convergence rate. The article is organized as follows. In the second part, previous techniques are discussed. The proposed technique is thoroughly explained in Section III. Section IV presents the simulation results. The paper's fifth and final section offers some ideas for future research.

II. RELATED WORK

As scientists construct UWSNs, they are challenged by significant mobility, propagation delays, limited bandwidth, and restricted battery and memory capacities. Tran and Oh [25] proposed a clustering approach that overcomes the UWSN's limitations. It is divided into four stages: the initiation stage, cluster head selection stage, clustering stage, and data aggregation stage. Efforts are being made to reduce the network's energy consumption, increase throughput, minimize data redundancy, and guarantee data accuracy.

UWSN sensor nodes continually lose power while performing underwater monitoring tasks. With limited replacement options for underwater sensor nodes, energy saving is key to increasing their lifespans. Data aggregation and clustering are potentially energy-saving techniques. An improved data aggregation method for cluster-based UWSNs is presented by Goyal, et al. [26], which employs a TDMA transmission schedule to eliminate collisions between clusters and an efficient sleep-wake-up algorithm to aggregate the sensed data, and by combining data aggregation, scheduling, and fusion, well-known existing protocols are improved to minimize energy consumption. Compared to existing protocols, the proposed scheme performs better concerning energy consumption, delay, and packet delivery rate than existing approaches.

A similarity function-based data aggregation with a Semaphore process is used by Ruby and Jeyachidra [27] to reduce energy consumption in UWSNs. A Date Palm Tree approach is used to cluster sensor nodes. Data aggregation nodes use Minkowski distance to check if readings collected from cluster members are similar. Data aggregators and cluster heads implement the Semaphore concept to ensure optimal network performance and reduce energy consumption. Packets can be moved from the data aggregators to the cluster heads through the message queue. Similarity measures in the proposed algorithm would lead to an improvement in link quality, redundancy, data delay, and energy consumption.

Krishnaswamy and Manvi [28] propose a palm tree-based approach for data aggregation and routing in UWASN. Petioles, rachis, leaflets, and spines make up the structure of palm trees. The sink node is connected to the junction of petioles. The proposed scheme creates fronds and leaflets by connecting them through spines. Several factors are taken into account when selecting master center nodes at the petiole junction, such as interconnection, petiole angle, Euclidean distance, and residual energy. The third step involves the identification of local centers at either end of the leaflet and linking them to the master centers via mobile agents. As for the fourth step, local aggregation considers leaflet nodes at local centers and carries them to a connected master center. Simulations are performed under various UWASN situations to assess the effectiveness of the scheme.

Wan, et al. [29] present an energy-efficient adaptive clustering routing protocol for UWSNs. Using a hierarchical network structure, the algorithm determines the size of the competition radius based on the distance between cluster heads and the base station. By avoiding early death, cluster heads can avoid excess competition radius and excessive energy burden. Based on the nodes' residual energy and transmission paths' energy loss, the algorithm can select the cluster head with the largest residual energy. This will optimize network energy consumption. To balance energy, routing rules are determined by residual energy levels to select between a single-hop routing node that is more energy-efficient and a multi-hop routing node that is more energy-efficient. According to simulation experiments, the proposed algorithm results in significant energy savings to the AFP protocol and DEBCR algorithm.

IoT-enabled depth base routing (IDBR) was proposed by Farooq, et al. [30] in order to maximize energy efficiency. MATLAB simulation was used to compare the IDBR with the standard DBR protocol. Both methods (IDBR and DBR) are analyzed based on delay, base station utilization, network lifetime, and energy consumption, alive nodes. Based on simulations, it is found that IDBR is 27.7% more energy efficient than DBR and improves network stability. IDBR also utilizes surface sinks more than DBR since sinks are used as relays, forwarding data to the base station directly, which gives field nodes more power. It increases the accessibility and security of the sensed data while improving the network's lifetime.

III. PROPOSED METHOD

The proposed method for data aggregation problems in UWSNs usually has a high delay, or they do not consider this problem. This section proposes a technique through the ACO algorithm to solve this problem, reduce energy consumption, and extend the underwater network's lifespan. The ACO algorithm always chooses the routes with the least number of hops to route data packets between sources and sink nodes.

A. Network Model

According to the suggested network model, all sensor nodes are dispersed over a three-dimensional space with fixed sensor placements, and the base station knows their locations. Since they are close, all sensors can send and receive data to and from the base station. The proposed model represents the

network as a graph $G=(V, E)$, where V is the set of nodes and E is the links between nodes, as seen in Fig. 1. In the model, sensors receive data from other sensors, integrate it with their data, create a packet without considering the amount of data received, and then send the packet. The problem is planned by a routing plan to deliver integrated packets from the sensors to the base station to reduce the energy consumption of the sensors, increase the network lifetime, and reduce the end-to-end delay. One solution is to use multi-hop communication by integrating the associated data. The problem is modeled as a graph with sensors representing the graph nodes. A maximum sensing range is considered for each node, which adjusts its neighborhood group and routing table. Each source node senses the sensing area regularly and sends data to the next node to receive from the base station.

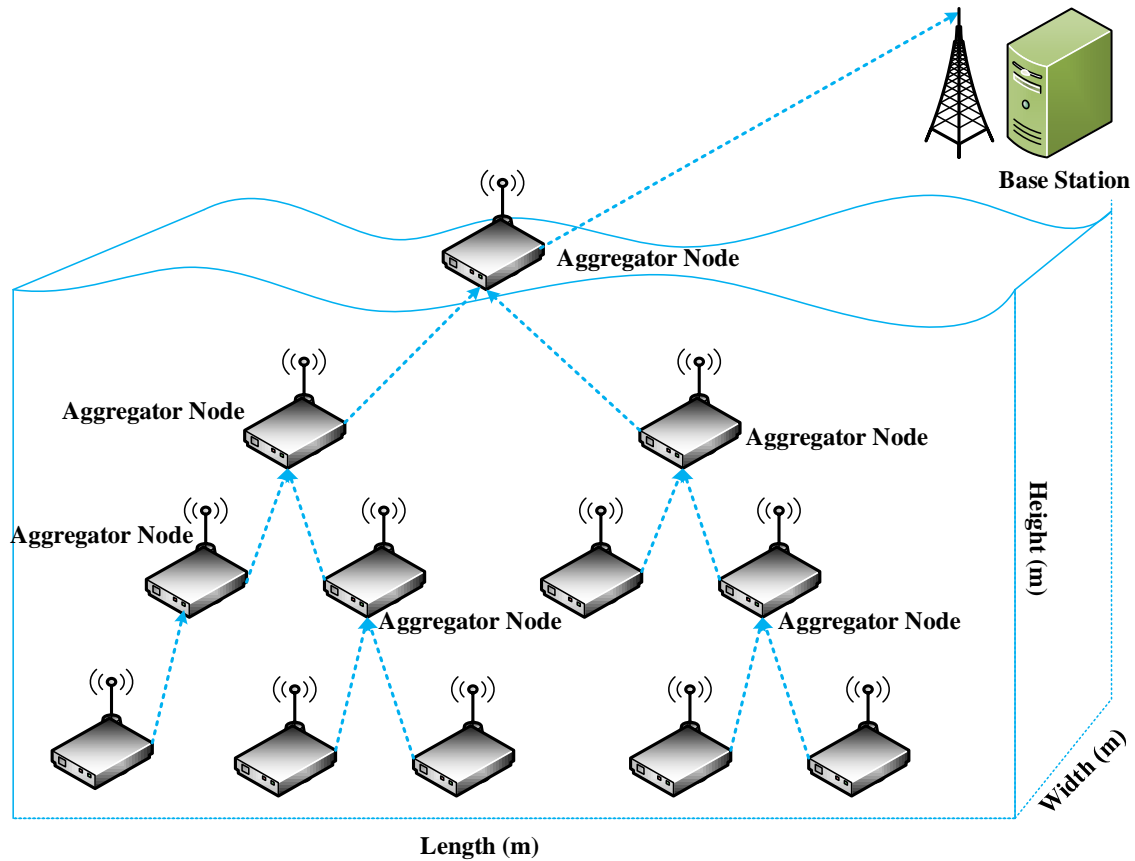


Fig. 1. The proposed network model.

B. Optimization Models

1) *Attenuation model*: Attenuation results from the conversion of sound energy to heat, and the energy absorbed by the water is accounted for according to the signal frequency. The Throb model [31] has the simplest equations for calculating attenuation based on signal frequency. The Throb equation is Eq. (1), where f is the signal frequency in kHz.

$$\alpha = 0.11 \times \frac{f^2}{1+f^2} + 44 \times \frac{f^2}{4100+f^2} + 2.75 \times 10^{-4} \times f^2 + 0.003 \quad (1)$$

2) *Transmission loss model*: The loss of transmission (no connection between the transmitter and the receiver) occurs because of the reduction in the sound intensity on the route (communication link) between the transmitter and the receiver. It depends on the range of the nodes and the signal attenuation. The transmission loss is defined and calculated by Eq. (2) and Eq. (3), where SS is the spherical distribution factor in the three-dimensional environment calculated as follows:

$$TL = SS + \alpha \times 10^{-3} \quad (2)$$

$$SS = 20 \times \log r \quad (3)$$

α is the attenuation coefficient, calculated by Eq. 1, and r is the communication area of the nodes in meters.

3) *Signal-to-noise ratio (SNR) model*: Generally, SNR is a criterion that indicates the proportion of the desired signal versus noise in the network. SNR is defined such that a higher ratio means that the amount of signal is more significant than the noise. In wireless sensor networks underwater, the SNR of the signal transmitted by a node is defined as the sum of the source level, transmission loss (TL), noise level (NL), and directional index (DI) of the signal transmission. This ratio is defined as follows:

$$SNR = SL + TL + NL + DI \quad (4)$$

SL depends on the Transmission Power (P_t) and Transmission Power Intensity (I_t), defined as follows:

$$SL = 10 \times \log\left(\frac{I_t}{0.067 \times 10^{-18}}\right) \quad (5)$$

The I_t of an underwater signal is calculated as follows according to the P_t :

$$I_t = \left(\frac{P_t}{2 \times \pi \times 1m \times d}\right) \quad (6)$$

Where I_t unit is m^2 because this ratio is estimated for a signal within a one-meter distance of the source node for shallow water, and d is the distance in meters between the nodes. TL is estimated using Eq. (2), and the considered DL is zero because of assuming that the receiver and transmitter of the signals (hydrophones) are distributed in an appropriate direction. NL in an underwater wireless sensor network is the sum of disturbance noises, the ships transportation noise, sound noise, and thermal noise, calculated by the following equation:

$$N(F) = N_t(f) + N_s(f) + N_w(f) + N_{th}(f) \quad (7)$$

In which the values of $N_t(f)$, $N_s(f)$, $N_w(f)$, and $N_{th}(f)$ are estimated as follows:

$$10 \log N_t(f) = 17 - 30 \log(f) \quad (8)$$

$$10 \log N_s(f) = 40 + 20(s - 0.5) + 26 \log(f) \quad (9)$$

$$10 \log N_w(f) = 40 + 7.5\sqrt{w} + 20 \log(f) - 40 \log(f + 0.4) \quad (10)$$

$$10 \log N_{th}(f) = -15 + 20 \log(f) \quad (11)$$

S in Eq. (9) is the transportation factor, in Eq. (10), wind speed ranges from 0 to 1, and f in all of the above equations is the frequency in kHz.

4) *Delay model*: The suggested technique uses a propagation delay model to calculate delay using Eq. (12). Underwater propagation delay depends on the underwater sound speed and distance between the nodes.

$$t_p = \frac{d}{c} \quad (12)$$

Eq. (12)'s terms d and c represent the distance between two nodes and the sound speed (m/s), respectively. Sound speed for underwater acoustic communications is calculated using Eq. (13). A sound wave can be mechanical energy transferred by the source node. This wave can be propagated from one

particle to another one through the ocean proportional to the sound speed.

$$C = 1449 + 4.6T + 0.055T^2 - 5.304 \times 10^{-2}T^2 + 2.374 \times 10^{-4}T^3 + \quad (13)$$

In Eq. (13), T stands for temperature, D for depth, and S is the saltness of water (parts per 1000). Since the sea depth varies between 0 and 1500 meters, the water temperature and saltness reduce with sound speed.

5) *Energy consumption model for UWSN*: Usage of energy in an underwater channel for data transmission between two nodes with These equations are used to calculate distance d :

$$E(d) = E_t(d) + E_r(d) \quad (14)$$

$$E_t(d) = l(E_{elec} + E_{amp}) + P_t \times \frac{l}{h \times B(d)} \quad (15)$$

$$E_r(d) = l(E_{elec} + E_{DA}) + P_r \times \frac{l}{h \times B(d)} \quad (16)$$

In the above equations, P_t and P_r are transmission and reception powers to transfer and receive energy E_t and E_r , l is the size of the data packet, $B(d)$ is the existing bandwidth, and h are bandwidth efficiency (bps/Hz), calculated by applying Eq. (17):

$$h = \log_2(1 + SNR) \quad (17)$$

E_{elec} is the required energy to process one bit of the packet (data package), E_{amp} is the amount of energy consumed. These are the values they follow: $E_{elec} = 50nJ/bit$ and $E_{amp} = 10pJ \times bit^{-1} \times m^{-2}$. Moreover, E_{DA} in Eq. (16) is the required energy for data aggregation. In the proposed method, data aggregation is performed by each parent node. The energy consumption of this process is estimated using Eq. (18), where l is the length of the transmitted packet, EDa is the consumed energy for data aggregation, and n is the count of the aggregator node's children.

$$E_{DA} = l \times E_{Da} \times n \quad (18)$$

6) *Lifetime model*: The network lifetime is calculated by applying Eq. (19), where $e_{initial}$ The initial energy of the system. e_{total} The sum of the required energy for data transmission and reception is estimated using Eq. (14).

$$Lifetime = \frac{e_{initial}}{e_{total}} \quad (19)$$

7) *The objective function*: According to the models discussed in the previous sections, all parameters should be optimized to achieve the desired result. Therefore, the objective function used in this paper is a multi-objective optimization function given by Eq. (20), where E is the consumed energy, L is the lifetime, D is the end-to-end delay, α is attenuation, and TL is the transmission loss rate. W_1 , W_2 , W_3 , W_4 , and W_5 are the mentioned function coefficients used for the normalization of the results. These coefficients are between 0 and 1, and $W_1 + W_2 + W_3 + W_4 + W_5 = 1$.

$$Fitness = W_1 \times E + W_2 \times L + W_3 \times D + W_4 \times \alpha + W_5 \times TL \quad (20)$$

C. The Proposed Method

The proposed method uses the ACO algorithm to create the aggregation tree, and one ant is considered for each node. First, each creates a route through the heuristic function and pheromone set. The route is stored as a solution. The proposed procedural steps are as follows.

1) *Solution generation by ants*: This step generates a primary population, initialises parameters, and the ants present a solution.

- Step 1-1: Generating primary population

The number of nodes in a network and the primary ant population are related; additionally, each ant has an array with the IDs of the nodes on the route to the destination. Fig. 2(a) shows three arrays of the response. The last index shows the base station, and the first index of each array shows the source node. Fig. 2(b) displays the generated tree for the solutions.

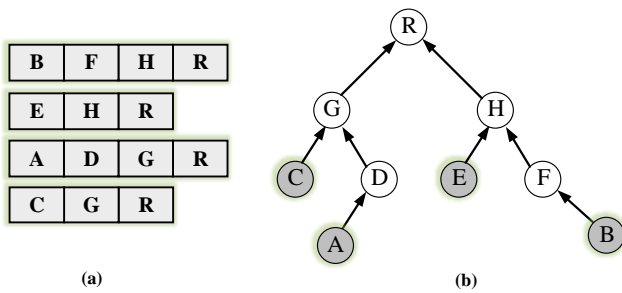


Fig. 2. Ant solutions.

- Step 1-2: The parameters initialization

In this step, the parameters should be set based on the problem. Thus, the correct values for the primary pheromone of the route and the heuristic function should be defined.

- The heuristic function definition: various functions can be used for the heuristic function in this problem. This function is proportional to the cost of the links. In the proposed method, the inverse Euclidean, a node's distance from the following node, is considered the heuristic function. Hence, an $n \times n$ matrix is generated to save the distance between the nodes. The heuristic function is shown in η that η_{ij} is the reverse distance between the i^{th} and j^{th} nodes.
- Pheromone definition: the amount of the pheromone between the i^{th} and j^{th} nodes shows the amount of the route goodness based on the ants' experience in the previous round. When the ants reach their destination, the pheromone is updated. In this paper, the primary pheromone for each route is the reverse Euclidian, the distance from the base station to the next node. An $n \times n$ matrix is created to save the primary pheromone in each route.

- Step 1-3: Solution generation by the ants

The ants are the answer generator routines that generate the answer by random movement on the graph $G = (V, E)$. The ants should consider the policies and the limitations of the problem when moving on the graph. They search for a route to the base station to generate a solution. The probability of selecting the next step is computed by Eq. (21). The definitions related to Eq. (21) are presented in Table I. One of the nodes is selected randomly after calculating the probability of each node selection as the next step. The roulette wheel sampling method (a random selection method of a discrete distribution) is used for this aim. The ants continue their step-by-step movement until the ant reaches the base station. Each ant's solution includes a generated route from the source node to the base station that is saved in an array with the length of the route.

$$P_{i \rightarrow j}^k = P_{ij}^k \begin{cases} \frac{(\tau_{ij})^\alpha \times (\eta_{ij})^\beta}{\sum_{u \in N_i^k} (\tau_{iu})^\alpha \times (\eta_{iu})^\beta} & j \in N_i^k \\ 0 & j \notin N_i^k \end{cases} \quad (21)$$

- Step 1-4: The obtained solution evaluation

The generated answer by each ant is evaluated in this step. At first, to determine the distance between the nodes, Eq. (22) is used.

$$d = d + \text{distance}(n(i) + (n(i+1))) \quad (22)$$

In Eq. (22), $n(i)$ is the source node, $n(i+1)$ is the next node, and d shows the total distance between the source node and the base station. The entire route distance is calculated step-by-step and saved in d . Then the energy consumption, delay, lifetime, attenuation, and SNR of each route are computed based on the equations in the mentioned models.

2) *Pheromone update*: This step includes adding the pheromone and its evaporation. The goal of adding a pheromone is to increase the amount of the pheromone related to the optimal solutions performed after each repetition of the algorithm. The pheromones are evaporated in each round. Pheromone evaporation helps the ants to forget previously learned unacceptable solutions.

- Step 2-1: Adding pheromone

Shedding pheromone is performed in this step based on the obtained solutions from the ant colony optimization algorithm. Some pheromone is added to each edge from the i^{th} to j^{th} node passed by the k^{th} ant proportional to the cost of the k^{th} ant solution using Eq. (23). Q is a constant considered 1. Less cost of the key leads to adding more pheromones to the route.

$$\tau_{ij} = \tau_{ij} + Q / \text{ant}(k). \text{cost} \quad (23)$$

- Step 2-2: Pheromone evaporation

The route's pheromone is evaporated proportional to Eq. (24), in which ρ is the evaporation rate with the value of 0.05.

$$\tau_{ij} = (1 - \rho) \times \tau_{ij} \quad (24)$$

IV. EXPERIMENTAL RESULTS

MATLAB is used for the implementation and data analysis of the proposed method because this simulator is used in many types of research because of its high ability for mathematical calculations, high capability to show the results, high efficiency, and high data volume. Simulation parameters and the used variables of the planned method are presented in Table II.

Three scenarios show the proposed method's efficiency rather than the previous ones. End-to-end delay and the energy consumption of the proposed technique are compared with the previously proposed techniques in the first and second scenarios. The idea of using multiple sinks is identified in the third scenario, and its performance is compared with the methods that use only one sink. The first scenario has been performed based on the data of [32], and the parameters of this paper are shown in Table III. In this scenario, as shown in Fig. 3 to 6, the proposed method delay outperforms the Clustered-based Multipath Shortest-distance Energy efficient Routing protocol (CMSE2R), and the genetic algorithm regarding network lifetime, energy consumption, end-to-end delay, and packet delivery rate.

The second scenario has been performed based on the data of [33], and the parameters of this paper are shown in Table IV. In this scenario, as displayed in Fig. 7, the proposed method performs well regarding end-to-end delay compared to Depth Base Routing (DBR). The third scenario has been performed based on the data of [34], and the parameters of this paper are shown in Table V. Fig. 8 and 9 show that the planned method performs better in terms of energy consumption and delay than Firefly mating optimization inspired Routing Protocol (FFRP) and Particle Swarm Optimization (PSO) algorithm. The simulation findings demonstrate that the suggested method needs low energy compared with the previous methods, and its end-to-end delay

is less than the previously proposed methods. A new idea is proposed and implemented to improve this method in which multiple sinks are used for data gathering instead of one sink. Also, the network lifetime improves in addition to the energy and delay reduction.

Considerations for the subsurface environment should be taken into account when evaluating UWSNs. The host conditions pose major challenges in terms of continuous node movement and 3D topology. Additionally, some underwater applications, including detection or rescue missions, are often ad hoc in nature, requiring both rapid deployment of networks and no advance planning. The routing protocol should be able to determine the location of the nodes in such circumstances without requiring any prior knowledge of the network. Furthermore, the network should be able to reconfigure itself under dynamic conditions so that the communication environment can be optimized. A significant consideration in the selection of a system is the relationship between the communication range and data rate and the specific conditions. Even when configured for higher data rates, a deep-water system may not be suitable for shallow water.

In practice, manufacturers' specifications of maximum data rates are mostly useful for establishing the upper performance bounds, which are not always achievable under specific conditions. Well-funded users have purchased multiple systems and tested them in specific environments in order to determine whether they are suitable for their needs. There is a need for an international effort to standardize the tests for acoustic communications, but this is not as simple as it sounds since private organizations or even government organizations that conduct comprehensive tests do not generally publish the results of their studies. Additionally, there is a lack of global standards for acoustic communication systems, which makes it difficult to compare the performance of different systems. Without a unified set of standards, it is difficult to determine which system is best suited for a particular application.

TABLE I. DEFINITIONS OF VARIABLES IN EQ. (21)

Parameters	Definition
P_{ij}^k	Probability of choosing node j as the next step by ant k
(τ_{ij})	The pheromone flow rate from node i to node j
(η_{ij})	The reverse of the distance between nod i to node j
N_i	The set of nodes
α	The controlling parameter for the relative effect of the amount of pheromone
β	The controlling parameter for the relative effect of the amount of heuristic function

TABLE II. SIMULATION PARAMETERS

Parameters	Description	Value
A(m×n×z)	Sensing area	100*100 – 500*500
N	Number of nodes	100-500
MaxIT	Max iteration	100-1000
L	Packet length	2-10 byte
$E_{initial}$	Initial energy	0.25 – 5 j
α	Ant colony constant	1
ρ	Pheromone evaporation rate	0.006
τ	Initial pheromones	1
Q	Ant colony constant	1
β	Ant colony constant	1

TABLE III. PARAMETERS FOR THE FIRST SCENARIO

Parameters	Values
No. of Nodes	300
Network Size	1500m x 1500m
Initial Energy	70 J
Simulation time	1000 sec
Data Packet size	64 bytes
Transmission range	100 m to 150 m
MAC Protocol (Shin & Kim, 2008)	802.11-DYNAV
The surface sink distance difference	100 m
Energy consumption for receiving	0.75w
Energy consumption for idle listening	8mw
Energy consumption for transmitting	2w

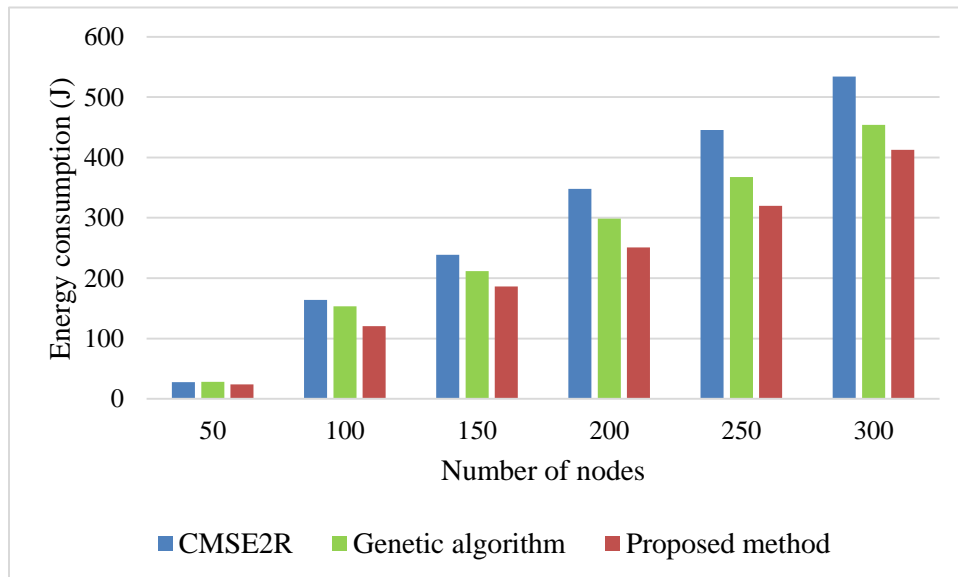


Fig. 3. Energy consumption comparison.

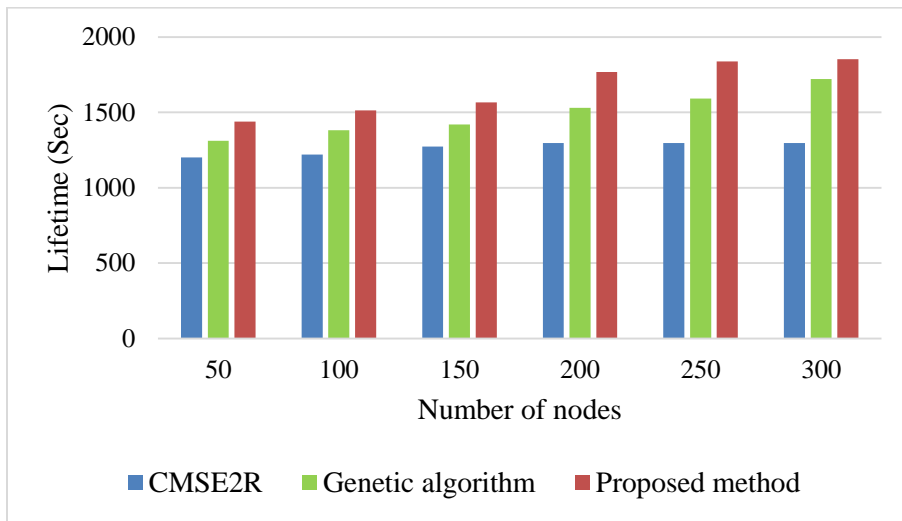


Fig. 4. Network lifetime comparison.

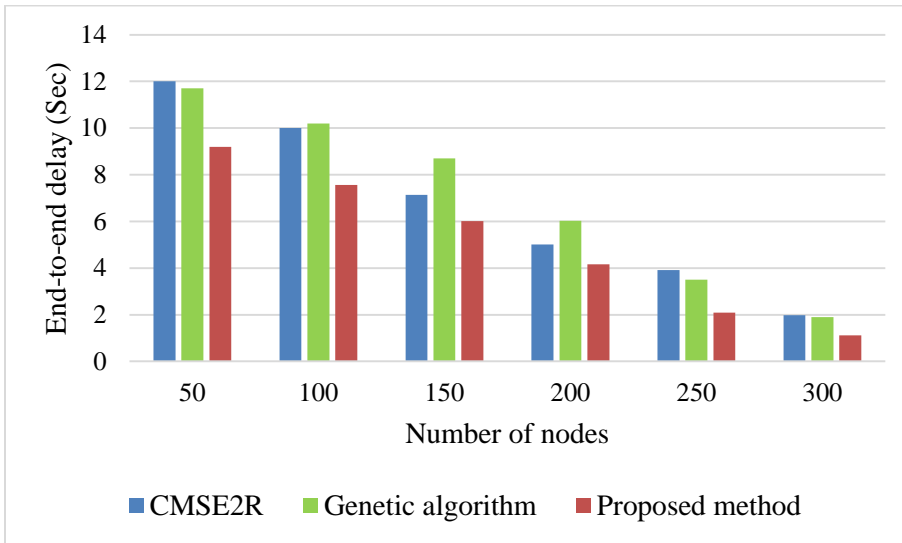


Fig. 5. End-to-end delay comparison.

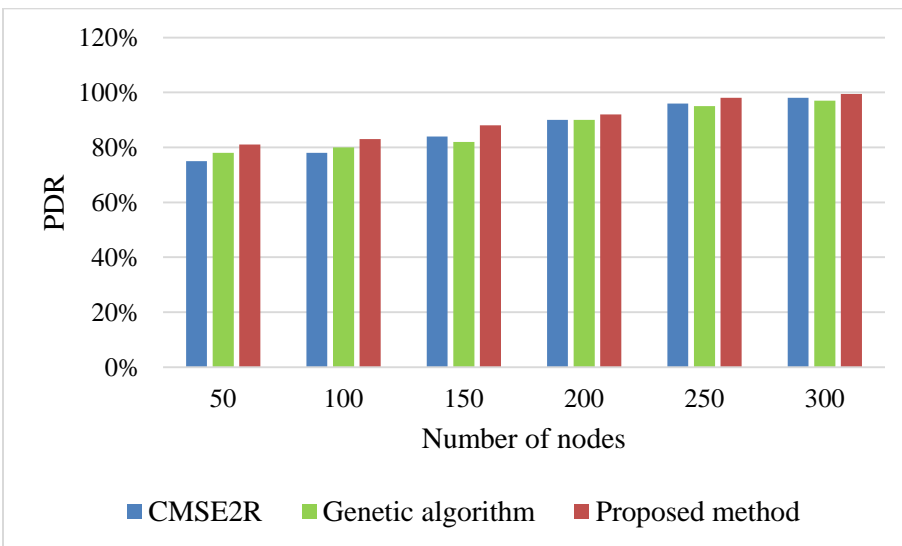


Fig. 6. Packet delivery rate comparison.

TABLE IV. PARAMETERS FOR THE SECOND SCENARIO

Parameters	Values
Data Packet size	512 bytes
Network Size	3000m x 3000m
No. of Nodes	30-150
Range (Transmitter/Receiver) of nodes	100m
Simulation time	70 sec

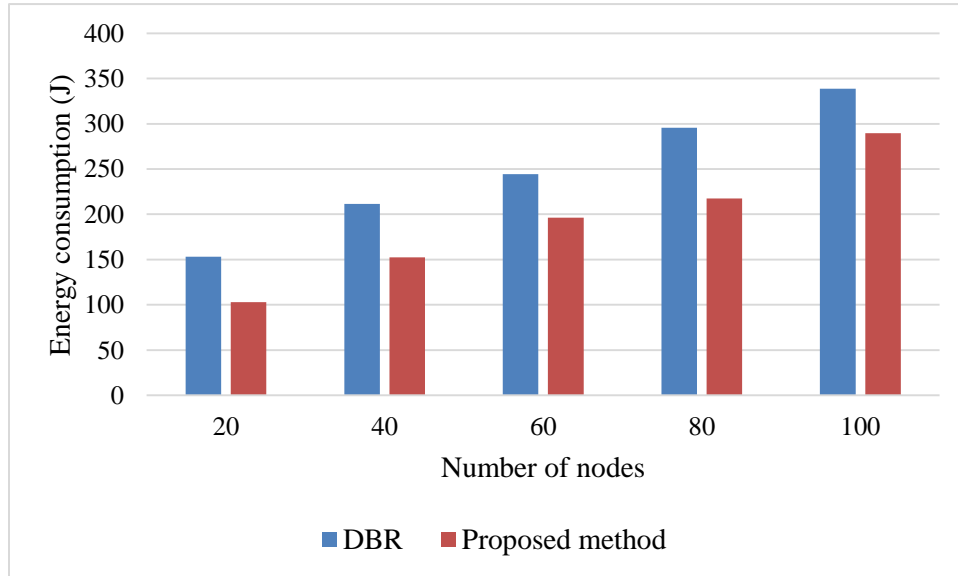


Fig. 7. Energy consumption comparison.

TABLE V. PARAMETERS FOR THE THIRD SCENARIO

Parameters	Values
Initial Energy	100 j
Network Size	1000m x 1000m
No. of Nodes	300
Data Packet size	30 bytes
Range (Transmitter/Receiver) of nodes	150m
Simulation time	130 sec

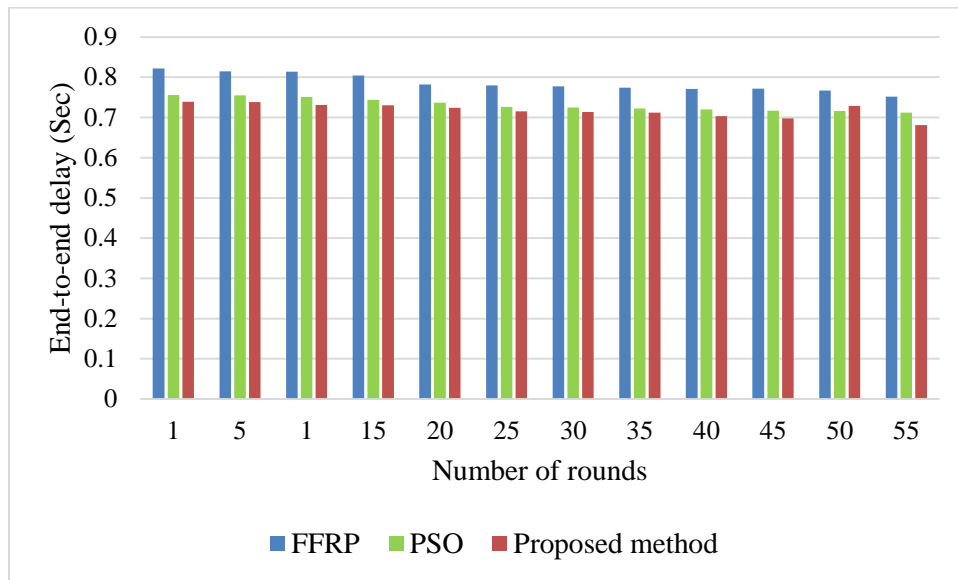


Fig. 8. End-to-end delay comparison.

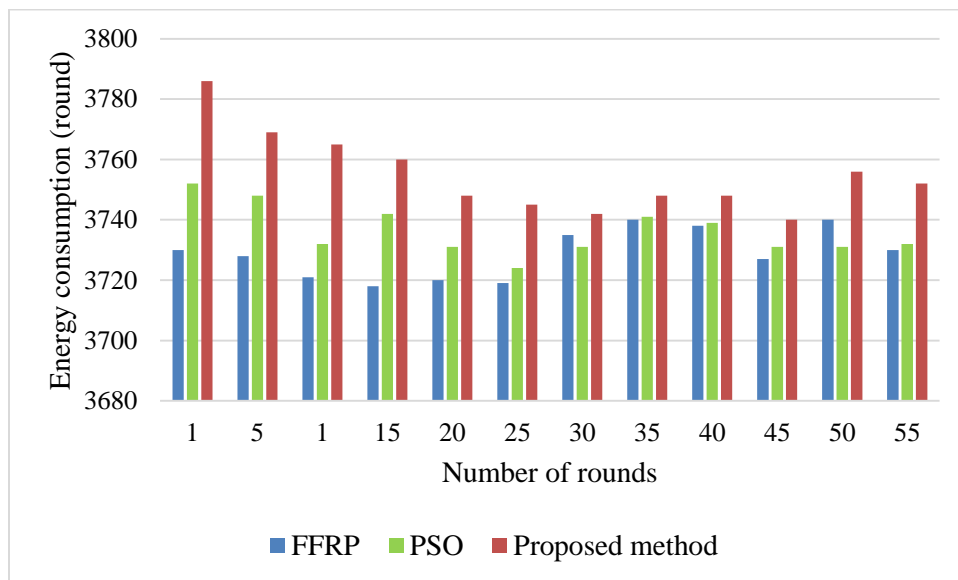


Fig. 9. Energy consumption comparison.

V. CONCLUSION

In order to mitigate the issue of excessive energy consumption in UWSNs, this work developed an energy-efficient data aggregation technique with the modified ACO algorithm. The paper made the following contributions. The heuristic information was enhanced using the distance factor and the residual energy of nodes. This research also included an improved adaptive technique for updating the evaporation parameter for the pheromone update mechanism, which can increase the algorithm's global search capacity and convergence rate. As a third point, this paper proposes ant searches. Simulation findings regarding packet delivery rate, end-to-end latency, network lifetime, and energy consumption demonstrate that the suggested technique surpasses existing ones. To achieve effective data aggregation, the aggregator must wait until the data is collected from various sensors and

transmit them to the sink without any collisions or delays. In order to accomplish this, effective scheduling techniques are required. Therefore, in the future work we will develop a scheduling technique that will allow for the effective aggregation of data from the sensors. As the aggregated data is transmitted to the sink without delay or loss of quality, it considers both spatial and temporal co-relationships among the data. For energy balanced networks, the mobility of sensor nodes should also be taken into consideration in an intra- and inter-cluster data aggregation process.

REFERENCES

- [1] F. Kamalov, B. Pourghebleh, M. Gheisari, Y. Liu, and S. Moussa, "Internet of Medical Things Privacy and Security: Challenges, Solutions, and Future Trends from a New Perspective," *Sustainability*, vol. 15, no. 4, p. 3317, 2023.

- [2] M. Mohseni, F. Amirghafouri, and B. Pourghebleh, "CEDAR: A cluster-based energy-aware data aggregation routing protocol in the internet of things using capuchin search algorithm and fuzzy logic," *Peer-to-Peer Networking and Applications*, pp. 1-21, 2022.
- [3] A. Kumar et al., "Optimal cluster head selection for energy efficient wireless sensor network using hybrid competitive swarm optimization and harmony search algorithm," *Sustainable Energy Technologies and Assessments*, vol. 52, p. 102243, 2022.
- [4] S. A. Saeidi, F. Fallah, S. Barmaki, and H. Farbeh, "A novel neuromorphic processors realization of spiking deep reinforcement learning for portfolio management," in *2022 Design, Automation & Test in Europe Conference & Exhibition (DATE)*, 2022: IEEE, pp. 68-71.
- [5] F. Vahedifard, S. Hassani, A. Afrasiabi, and A. M. Esfe, "Artificial intelligence for radiomics; diagnostic biomarkers for neuro-oncology," *World Journal of Advanced Research and Reviews*, vol. 14, no. 3, pp. 304-310, 2022.
- [6] J. Akhavan and S. Manoochehri, "Sensory data fusion using machine learning methods for in-situ defect registration in additive manufacturing: a review," in *2022 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS)*, 2022: IEEE, pp. 1-10.
- [7] R. N. Jacob, "Non-performing Asset Analysis Using Machine Learning," in *ICT Systems and Sustainability: Proceedings of ICT4SD 2020*, Volume 1, 2021: Springer, pp. 11-18.
- [8] C. Han and X. Fu, "Challenge and Opportunity: Deep Learning-Based Stock Price Prediction by Using Bi-Directional LSTM Model," *Frontiers in Business, Economics and Management*, vol. 8, no. 2, pp. 51-54, 2023.
- [9] S. H. Haghshenas, M. A. Hasnat, and M. Naeni, "A Temporal Graph Neural Network for Cyber Attack Detection and Localization in Smart Grids," *arXiv preprint arXiv:2212.03390*, 2022.
- [10] S. Meisami, M. Beheshti-Atashgah, and M. R. Aref, "Using Blockchain to Achieve Decentralized Privacy In IoT Healthcare," *arXiv preprint arXiv:2109.14812*, 2021.
- [11] R. Singh et al., "Analysis of Network Slicing for Management of 5G Networks Using Machine Learning Techniques," *Wireless Communications and Mobile Computing*, vol. 2022, 2022.
- [12] A. Mehbodniya et al., "Energy-Aware Routing Protocol with Fuzzy Logic in Industrial Internet of Things with Blockchain Technology," *Wireless Communications and Mobile Computing*, vol. 2022, 2022.
- [13] B. Pourghebleh, A. A. Anvigh, A. R. Ramtin, and B. Mohammadi, "The importance of nature-inspired meta-heuristic algorithms for solving virtual machine consolidation problem in cloud environments," *Cluster Computing*, pp. 1-24, 2021.
- [14] X. Wei, H. Guo, X. Wang, X. Wang, and M. Qiu, "Reliable data collection techniques in underwater wireless sensor networks: A survey," *IEEE Communications Surveys & Tutorials*, vol. 24, no. 1, pp. 404-431, 2021.
- [15] J. Luo, Y. Chen, M. Wu, and Y. Yang, "A survey of routing protocols for underwater wireless sensor networks," *IEEE Communications Surveys & Tutorials*, vol. 23, no. 1, pp. 137-160, 2021.
- [16] A. Ismail, X. Wang, A. Hawbani, S. Alsamhi, and S. Abdel Aziz, "Routing protocols classification for underwater wireless sensor networks based on localization and mobility," *Wireless Networks*, pp. 1-30, 2022.
- [17] N. Subramani, P. Mohan, Y. Alotaibi, S. Alghamdi, and O. I. Khalaf, "An efficient metaheuristic-based clustering with routing protocol for underwater wireless sensor networks," *Sensors*, vol. 22, no. 2, p. 415, 2022.
- [18] K. Saeed, W. Khalil, S. Ahmed, I. Ahmad, and M. N. K. Khattak, "SEECR: Secure energy efficient and cooperative routing protocol for underwater wireless sensor networks," *IEEE Access*, vol. 8, pp. 107419-107433, 2020.
- [19] H. Khan, S. A. Hassan, and H. Jung, "On underwater wireless sensor networks routing protocols: A review," *IEEE Sensors Journal*, vol. 20, no. 18, pp. 10371-10386, 2020.
- [20] A. Prasanth, "Certain investigations on energy-efficient fault detection and recovery management in underwater wireless sensor networks," *Journal of Circuits, Systems and Computers*, vol. 30, no. 08, p. 2150137, 2021.
- [21] N. Goyal, M. Dave, and A. K. Verma, "Protocol stack of underwater wireless sensor network: classical approaches and new trends," *Wireless Personal Communications*, vol. 104, no. 3, pp. 995-1022, 2019.
- [22] D. Anuradha, N. Subramani, O. I. Khalaf, Y. Alotaibi, S. Alghamdi, and M. Rajagopal, "Chaotic search-and-rescue-optimization-based multi-hop data transmission protocol for underwater wireless sensor networks," *Sensors*, vol. 22, no. 8, p. 2867, 2022.
- [23] S. Fattah, A. Gani, I. Ahmedy, M. Y. I. Idris, and I. A. Targio Hashem, "A survey on underwater wireless sensor networks: Requirements, taxonomy, recent advances, and open research challenges," *Sensors*, vol. 20, no. 18, p. 5393, 2020.
- [24] P. Joshi and A. S. Raghuvanshi, "Hybrid approaches to address various challenges in wireless sensor network for IoT applications: opportunities and open problems," *International Journal of Computer Networks and Applications*, vol. 8, no. 3, pp. 151-187, 2021.
- [25] K. T.-M. Tran and S.-H. Oh, "A data aggregation based efficient clustering scheme in underwater wireless sensor networks," in *Ubiquitous Information Technologies and Applications: Springer*, 2014, pp. 541-548.
- [26] N. Goyal, M. Dave, and A. K. Verma, "Improved data aggregation for cluster based underwater wireless sensor networks," *Proceedings of the National Academy of Sciences, India Section A: Physical Sciences*, vol. 87, no. 2, pp. 235-245, 2017.
- [27] D. Ruby and J. Jeyachidra, "Semaphore based data aggregation and similarity findings for underwater wireless sensor networks," *International Journal of Grid and High Performance Computing (IJGHPC)*, vol. 11, no. 3, pp. 59-76, 2019.
- [28] V. Krishnaswamy and S. S. Manvi, "Palm tree structure based data aggregation and routing in underwater wireless acoustic sensor networks: Agent oriented approach," *Journal of King Saud University-Computer and Information Sciences*, 2019.
- [29] Z. Wan, S. Liu, W. Ni, and Z. Xu, "An energy-efficient multi-level adaptive clustering routing algorithm for underwater wireless sensor networks," *Cluster Computing*, vol. 22, no. 6, pp. 14651-14660, 2019.
- [30] U. Farooq et al., "IDBR: Iot Enabled Depth Base Routing Method for Underwater Wireless Sensor Network," *Journal of Sensors*, vol. 2021, 2021.
- [31] A. Sehgal, I. Tumar, and J. Schonwalder, "Variability of available capacity due to the effects of depth and temperature in the underwater acoustic communication channel," in *OCEANS 2009-EUROPE*, 2009: IEEE, pp. 1-6.
- [32] M. Ahmed, M. A. Soomro, S. Parveen, J. Akhtar, and N. Naem, "CMSE2R: clustered-based multipath shortest-distance energy efficient routing protocol for underwater wireless sensor network," *Indian J. Sci. Technol.*, vol. 12, no. 8, 2019.
- [33] S. A. B. Andrabi and M. Kumar, "Energy Efficient Routing Technique for Underwater Wireless Sensor Networks," *International Journal of Advanced Science and Technology*, vol. 29, no. 11s, pp. 859-868, 2020.
- [34] M. Faheem et al., "FFRP: dynamic firefly mating optimization inspired energy efficient routing protocol for internet of underwater wireless sensor networks," *IEEE Access*, vol. 8, pp. 39587-39604, 2020.

A New Machine Learning-based Hybrid Intrusion Detection System and Intelligent Routing Algorithm for MPLS Network

Mohammad Azmi Ridwan¹, Nurul Asyikin Mohamed Radzi²,

Kaiyisah Hanis Mohd Azmi³, Fairuz Abdullah⁴, Wan Siti Halimatul Munirah Wan Ahmad⁵

Institute of Power Engineering-Department of Electrical and Electronic Engineering-College of Engineering, Universiti Tenaga Nasional (UNITEN), 43000 Kajang, Selangor, Malaysia^{1,2,3,4}

Faculty of Engineering, Multimedia University, 63100 Cyberjaya, Malaysia⁵

Abstract—Machine Learning (ML) is seen as a promising application that offers autonomous learning and provides optimized solutions to complex problems. The current Multiprotocol Label Switching (MPLS)-based communication system is packed with exponentially increasing applications and different Quality-of-Services (QoS) requirements. As the network is getting complex and congested, it will become challenging to satisfy the QoS requirements in the MPLS network. This study proposes a hybrid ML-based intrusion detection system (ML-IDS) and ML-based intelligent routing algorithm (ML-RA) for MPLS network. The research is divided into three parts, which are (1) dataset development, (2) algorithm development, and (3) algorithm performance evaluation. The dataset development for both algorithms is carried out via simulations in Graphical Network Simulator 3 (GNS3). The datasets are then fed into MATLAB to train ML classifiers and regression models to classify the incoming traffic as normal or attack and predict traffic delays for all available routes, respectively. Only the normal traffic predicted by the ML-IDS algorithm will be allowed to enter the network domain, and the route with the fastest delay predicted by the ML-RA is assigned for routing. The ML-based routing algorithm is compared to the conventional routing algorithm, Routing Information Protocol version 2 (RIPv2). From the performance evaluations, the ML-RA shows 100 percent accuracy in predicting the fastest route in the network. During network congestion, the proposed ML outperforms the RIPv2 in terms of delay and throughput on average by 57.61 percent and 46.57 percent, respectively.

Keywords—Machine learning; intrusion detection system; routing algorithm; quality of service; communication system

I. INTRODUCTION

Multi-protocol label switching (MPLS) routing technique for telecommunication networks was invented in the late 1990s as a more efficient alternative to the traditional Internet Protocol (IP) routing [1]. In contrast to traditional network protocols, which route data packets according to the source-to-destination (S2D) addresses, MPLS routes traffic from one node to another according to predefined labels in the packet header. These labels may contain information related to quality of service (QoS) such as traffic latency, jitter, packet loss and downtime, which allows network traffic to be prioritized according to its importance. One of the most noteworthy advantages of MPLS is its independence from any

protocol or transport medium. It supports IP-based, Ethernet-based, asynchronous transfer mode (ATM), and frame relay transmission [1]. Other benefits of MPLS include [1]: 1) providing good QoS performance for latency-insensitive applications such as video and mission-critical data; 2) allowing data and voice applications to coexist on the same network; 3) allows the pre-programming of different types of data with distinct priorities and service classes; and 4) offers network scalability to users.

However, as the number of applications and users grows exponentially, conventional MPLS networks are likely to become more complicated and require stringent QoS regulations to ensure network reliability, delay tolerance, and throughput. Routing assignment (RA) algorithms used in conventional networks are typically based on the shortest path with fixed rules. They may not give the optimal QoS, particularly in a complicated network architecture. While standards and algorithms have been developed to increase the efficiency of the existing networks, it is anticipated that conventional approaches would be unable to meet growing demand while maintaining QoS. Additionally, the networks are facing threats from cyber attackers that take advantage of network vulnerabilities, resulting in extensive network disruption and significant damage to an organization's reputation. These challenges emphasize the critical importance of intelligent routing techniques and network security protection, such as that provided by network intrusion detection systems (IDSs).

An IDS monitors network traffic for signals of hostile activity by building a predictive model that can discriminate between attack and normal network flows. However, despite decades of advancements, existing IDSs continue to face detection accuracy challenges by reducing the false alarm rates and the identification of unknown threats [2]. Additionally, the fixed rules of IDS systems are vulnerable to threats including Denial-of-Service (DoS) and brute force [3]. To safeguard the network from such vulnerabilities, researchers all around the world have created cutting-edge IDS with the integration of machine learning (ML) algorithms. ML is capable of rapidly identifying patterns in a variety of data and solving complex, multi-dimensional problems with little to no human intervention. Since intrusion detection is a

The research and publication of this article was funded by UNITEN BOLD grant J510050002/2021061 and 202210020YCU.

classification problem, ML can be one of the promising candidates for IDS in the network. A Learning-based system is used in ML-based IDS to identify possible attack classes based on the behavior of an incoming packet. These ML-based IDSs offer various advantages over conventional systems, including lower computational loads and greater flexibility, as well as the ability to detect novel attacks and capture the complex features of attack behavior [4].

This research addresses several issues in MPLS networks, including improving network QoS and enhancing network security using ML algorithms. This work proposes a hybrid supervised ML-based IDS and ML-based RA algorithm (herewith will be referred to as ML-IDS and ML-RA, respectively) that is trained using an ML-IDS dataset generated using a simple data extraction method. The proposed ML-IDS is a security intrusion classification algorithm that analyses the incoming traffic patterns and network conditions and then classifies the traffic as legitimate or potentially intrusive. Afterwards, the ML-RA intelligently computes a route that is predicted to provide the best QoS requirements under any network condition. In short, the main contributions of this research work are as follows: 1) we proposed an ML-IDS algorithm that uses a simple data extraction method from the network to train the classifier without compromising the accuracy; 2) we developed an ML-RA algorithm that predicts the QoS parameters and performs path computing for the incoming traffic with various priorities in different traffic conditions; and 3) we introduce the first hybrid ML-IDS and routing algorithm (RA) that enhances network security and QoS.

This paper is organized as follows. Section II provides the literature review. The formulated ML-IDS and ML-RA methodology is presented in Section III. The discussion on the findings in the evaluation of the proposed strategy is presented in Section IV. Finally, Section V refocuses on the purpose of the research and draws conclusions for this study.

II. LITERATURE REVIEW

A. Routing Strategies in the Literature

One of the networking fundamentals responsible for selecting a path for packet transmission is network traffic routing. With proper network routing management, it is possible to achieve a QoS-compliant and cost-effective route, especially through the implementation of ML in network routing. However, ML-based traffic routing is often challenging because of various constraints including complex and dynamic topologies, diverse traffic, and unique QoS requirements. In routing optimization problems, traffic and route matrices can be used to describe the input and output of ML algorithms [5]. To predict or select a path for incoming traffic, ML algorithms must learn the correlation between traffic inputs and link conditions. The recent applications of ML in routing can be divided into five routing objectives, which are discussed as follows:

1) *Routing by predicting network parameters:* In today's network operations and administration, it is critical to predict network parameters such as path or connection quality, delay, throughput, optical signal to noise ratio (OSNR), and

incoming traffic. ML aims to improve overall network performance by learning from past data or the environment. For example, Alvizu et al. [6] trained an artificial neural network (ANN) using a public dataset from Milan to forecast the traffic load and variation and calculate the best resource allocation via dynamic optical routing in software defined networks (SDNs), thereby reducing energy consumption. In contrast, Choudhury et al. [7] introduced a hybrid machine learning (ML) model based on the Gaussian process (GP) method to forecast traffic volume for each traffic engineering tunnel over time, followed by forecasting the optical performance of new wavelengths in a multi-vendor environment.

2) *Routing for QoS improvement:* By controlling the network's delay, jitter, bandwidth, and packet loss ratio, a good QoS can be attained. However, with the explosion of traffic volume in the network, it can be challenging to fulfil the QoS specifications for each incoming traffic. Due to the network's complexity, conventional algorithms to improve the QoS parameters may be impractical. To meet the QoS requirements, researchers are continuously developing and refining unique solutions, such as ML-based algorithms, to maximize throughput while minimizing latency.

Nakayama et al. [8] proposed a routing scheme using the Markov Chain Monte Carlo algorithm to reduce the worst-case end-to-end delay of all the front-haul flows of the centralized radio access network (C-RAN) and ensure that all flows meet the latency requirements. The proposed solution successfully reduces all flow's latency, demonstrating that the ML-based approach can address the CRAN's queuing delay problem. Additionally, Stampa [9] proposed a deep RL agent that can optimize routing in accordance with a predefined target metric, such as the delay requirement in SDN. The deep RL model automatically adapts to the current traffic conditions and proposes a customized configuration that minimizes network delay. The suggested deep RL agents were able to reliably calculate total traffic intensities, and the average delay is less than the benchmark of 100,000 randomly created routes.

3) *Low computation routing scheme:* Incorporating ML in the network may lead to high computational load, especially when dealing with high dimensional input features or when using deep learning (DL)-based algorithms. With that, several works proposed a low computational routing scheme. For example, Hendriks et al. [10] proposed Q2-RA, which is hybrid of Q-routing and Multi-Agent Reinforcement Learning (MARL) RAs. In the algorithm, ad-hoc wireless nodes decide on a route by selecting the neighbor with the best Q-value as the next-hop destination. Although this algorithm has an additional modified reward function to meet the QoS criterion, it is comparable to Q-routing. Only training traffic is provided throughout the learning process to obtain the Q-values on the available path until it converged inside a predetermined threshold. The transmission of data traffic then starts once the rate of sending learning traffic has drastically lowered. The

suggested Q2-RA performs better than the ad hoc on-demand distance vector method with QoS awareness and is more flexible to network condition changes.

Martin et al. [11] proposed a classifier that was trained using labelled Risk Weighted Assets (RWA) configurations and solved using inductive logic programming (ILP). The classifier can offer online network setup for newly arriving traffic matrices once it has been trained. In response to rapidly changing traffic patterns, it can dynamically adapt and reconfigure the network because of the quick computation of RWA configurations. Instead of calculating ILP for every incoming traffic, the network will remember prior ILP solutions and allocate a path in accordance with the historical data. As compared to the ILP method, this approach reduces computational time by up to 93%.

4) *Congestion-control routing*: Congestion is one of the main concerns for network providers as it can degrade the overall network performance. Through congestion control, network stability, fair resource allocation, and a reasonable packet loss ratio are all made possible [12]. In conventional routing protocols, previous network abnormalities, such as network congestion, are not learned. As network traffic keeps growing, the network is put under a lot of strain, which creates problems with resource management and allocation that affect traffic QoS. Given that the majority of networks are still using outdated routing systems, this congestion problem is getting more critical [13]. Additionally, routing systems were created for fixed networks that determine the shortest paths using distance vectors or link costs. Eventually, the network will experience excessive traffic load, which will severely degrade network performance. The conventional RAs frequently commit the same routing mistake when this condition recurs, leading to an unmanageable rise in delay and packet error rate. This is where the predictive ML models come in to attempt to overcome the congestion issue.

Due to the inflexibility of route selection in circuit-switched networks, their total routing performance is usually constrained. This issue is highlighted by the least loaded (LL) routing protocol, one of the network's routing protocols. Because of the excessive capacity consumption under conditions of heavy load, this routing protocol may result in subpar performance and overall inefficiency [5]. To assist the LL routing performance, a novel online-based supervised Naive Bayes (NB) classifier is proposed in [14]. The classifier forecasts the likelihood of future circuit blocking between node pairs. When a service is provided or denied, the network snapshot is stored as historical information to determine the best route for new service connections. The proposed solution outperforms the least-load and short-path conventional routing protocols in terms of the minimum number of extra hops, lowest blocking probability, and least amount of network capacity overconsumption.

Another example of congestion control is demonstrated by Tang et al. [13]. The authors proposed a real-time DL-based intelligent network traffic control method based on deep convolutional NN (deep CNN) with uniquely characterized

input and output to represent the wireless mesh network backbone. The performance of the proposed scheme is compared with Open Shortest Path First (OSPF), Intermediate System to Intermediate System (IS-IS), and Routing Information Protocol (RIP). The simulation results showed that the DL-based routing scheme is superior to other routing protocols, as 98.7% of congestion cases are avoided.

5) *Load-balancing routing*: The bursty nature of SDN packet traffic creates a network load imbalance. Yao et al. [15] proposed a pair of ML-aided load balance routing schemes that take queue utilization (QU) into account to address this issue. The aim is to improve load-balance routing by reducing the packet loss ratio and improving the worst throughput. To deal with network congestion caused by a sudden traffic burst, ANN algorithms predict the QU for the next time slot. The predicted value is used to guide intelligent routing decisions. When compared to the shortest path approach, the proposed scheme improves packet loss ratio and throughput while increasing delay by 20%.

The next-generation wireless network (NGWN) is a network service and operation interface that can support multiple standards such as 5G, Wi-Fi, and cognitive radio networks. However, the volume of traffic in the current communication infrastructure is expanding rapidly that the router's speed may not be sufficient to keep pace. Additionally, the NGWN's real-time load balance request cannot be satisfied and served by using conventional routing schemes that are solely based on standard rules and have limited computing capacity [16]. To anticipate the network queue state, which is one of the measures for making wise routing decisions, Yao et al. [16] suggested a load balancing routing based on NN. The proposed algorithm is compared to shortest path-based algorithms such as Bellman-Ford (BF) and Queue-Utilization BF (QUBF), in terms of throughput and delay. According to the results, the proposed technique reached the highest throughput while incurring a 20% delay over the BF approach. The proposed algorithm also can predict the next-hop path with the smallest buffer and thus improve load balancing.

B. Challenges in Routing

From the literature review, recent related works on ML-based RA have proven to route the traffic effectively. Almost all related works from literature successfully overcome the limitations of conventional routing protocols. Despite ML's superiority in routing in communication networks, there are still some challenges to consider, discussed as follows.

1) *Trade-off between accuracy and computational load*: The trade-off between accuracy and computational load in ML-based RA using classical ML and DL is similar to the issue discussed for ML-based IDS. This trade-off must be considered because ML routing decisions need to be swift and in real-time to avoid processing delay. Unlike the proposed ML-RA in the literature which used DL-based classifiers, our work aims to use ML-based regressions including DT-regressions and LR by Gibbs Sampling (LRgs) to predict the delay in all available routes. DT and LRgs are well-known for

their simplicity and interpretability, but they are prone to overfitting. To address this issue, we will train, test, and evaluate our proposed ML-RA under various network and traffic conditions.

2) *Lack of congestion and link failure scenario for routing*: The majority of associated studies in the literature test the effect of their proposed ML-based routing scheme under congested network condition, as evident from the studies in [13], [17]–[22]. Only a few research, such as [23], develop ML-based RA that considered both congestion and link or node failures. Therefore, our proposed ML-RA considers both congestion and link failure scenarios.

3) *Traffic modelling for performance evaluation*: As the network becomes more complex and congested with traffic of varying priorities, it is critical that the ML-based routing mechanism can deliver traffic while meeting QoS requirements. However, most related works only consider single traffic type for routing, which may not be feasible to resemble real-world traffic with different priorities. There has been little research into ML-based routing for traffic with varying QoS requirements. For example, in [24], the Transmission Control Protocol (TCP) and User Datagram Protocol (UDP) traffic are considered, which correspond to voice over Internet Protocol (VoIP) and video traffic. While

the authors [10] used three traffic priorities; high, medium and low priority, they did not explicitly specify the traffic types. Our proposed ML-RA will consider three traffic types which are expedited forwarding (EF), assured forwarding (AF), and best-effort (BE) traffic, which correspond to VoIP, close-circuit television (CCTV), and data transfer, respectively. It is expected that our proposed ML-RA can successfully route all traffic within their QoS requirements.

4) *Quality datasets for training*: When generating traffic in the network simulator, it is crucial that the traffic pattern is not random or static as demonstrated in [9], [11], [14], [17], [18], [22], to preserve the quality of the dataset used to train the ML algorithms. To improve the quality of the dataset, the traffic packets can be modified to resemble legitimate traffic properties, such as the standardized data rates and size. In this work, the ML-RA dataset is constructed using EF, AF, and BE traffic following the standard of VoIP, CCTV, and file transfers. More details on the traffic properties will be elaborated in Section III. In addition, the traffic is also modelled using the typical EF, AF, and BE traffic mixture ratios of 20:40:40 for [25].

Table I summarizes the discussed ML works in routing with their advantages and shortcomings.

TABLE I. SUMMARY OF RECENT ML-BASED RAS WITH THEIR ADVANTAGES AND SHORTCOMINGS

Authors	Routing Objectives	Description	Issues of Conventional Routing Protocol	ML Method	Advantages	Shortcomings
Yao et al. [15]	Load balancing	Proposed a pair of ML-assisted load-balancing RAs that consider QU, to improve load-balance routing by packet loss ratio reduction and improving the worst throughput	High computational complexity for QoS RA	DL	Improved global realignment and more efficient network optimization	<ul style="list-style-type: none">• High computational load• Only considers two traffic patterns (Steady and congest)
Yao et al. [16]	Load balancing	Proposed NN-based load-balancing RA to predict network queue status to make intelligent routing decisions	Traditional RAs cannot always serve the NGWN effectively	DL	Enhance the bit error rate, throughput, and delay	<ul style="list-style-type: none">• High computational load• Do not consider link failure• Do not consider different traffic priorities
Fadulullah et al. [23]	Predicting network parameters	Proposed a value iteration architecture-based deep RL routing approach, which includes the network node's adjacency matrix as learning parameters. The method can forecast the next node until the destination is reached	High computational cost to address RWA problems	DL	Ensures more stable network performance in the event of network topology changes	<ul style="list-style-type: none">• High computational load• Do not consider different traffic priorities
Murudkar et al. [18]	Predicting network parameters	Proposed a User Specific-Optimal Capacity Shortest Path RL routing in 5G networks is to establish the resource-based optimum-capacity shortest route for a user between S2D pairs	Challenging spectrum resource optimization	RL	Quickly determine the shortest route with the highest capacity	<ul style="list-style-type: none">• Only constant bit rate traffic in the simulations• Do not consider different traffic priorities• Do not consider node failures condition
Salani et al. [17]	Predicting network parameters	Proposed an integration of RF-based estimation for routing and spectrum assignment for quality of things	Inaccessible of perfect transmission knowledge to train the ML model	RF	Saves up to 30% on spectrum occupation.	<ul style="list-style-type: none">• Traffics in the simulations are generated randomly• Do not consider different traffic priorities• Do not consider link failure

Vashishth et al. [22]	Low computation routing scheme	Proposed cascade learning, an ensemble-based ML that combines LR and NN classifiers. Using the ML-based Probabilistic Routing Protocol and the History of Encounters and Transitivity (MLProph) as input, the logistic algorithm will generate two probabilities: delivered or not delivered	Context-free routing protocols suffer from high network overhead ratio and congestion	Ensemble LR and DL	Enhance message delivery probability, network overhead ratio, average hop count, and message drop rate.	<ul style="list-style-type: none"> Data generated is based on a flooding-based routing protocol to train the ML algorithms Do not consider different traffic priorities Do not consider node failures condition
Vashishth et al. [21]	Low computation routing scheme	Proposed a routing approach based on the Gaussian mixture (GM) model classifier in the Opportunistic Internet of Things (OppIoT) that increases message delivery probability	Fixed S2D path is non-existent, making routing very challenging	GM	Increase message delivery probability, the average hop count, the number of dropped packets, and the network overhead ratio	<ul style="list-style-type: none"> Delay is not one of the performance parameters Do not consider node failures
Hendriks et al. [10]	Low computation routing scheme	Proposed a hybrid of Q-routing and MARL RA. Ad-hoc wireless nodes use the algorithm to make routing decisions by selecting the neighbor with the best Q-value as the next hop	Traditional RAs have a high overhead load to be used in an ad hoc environment	RL	Outperforms well-known ad-hoc RAs in dynamic environments with QoS constraints	<ul style="list-style-type: none"> Do not consider node failures or congestion
Li et al. [14]	Congestion-control routing	Proposed an online-based supervised NB classifier for performance improvement. The classifier predicts the likelihood of future circuit blocking and uses the data to select routes for future service connections	Fixed route oriented significantly limits the routing performance and flexibility	NB	Saves ~ 90% of the time for the learning process, significantly speeding up simulation studies.	<ul style="list-style-type: none"> Traffic in the simulations is generated randomly Do not consider different traffic priorities
Tang et al. [13]	Congestion-control routing	Proposed a real-time DL-based intelligent network traffic control based on DCNN for a wireless mesh backbone	OSPF for training the ML-RA, which lacks the necessary intelligence to handle newly occurring situations	DL	Avoid 98.7% of congestion cases	<ul style="list-style-type: none"> High computational load Do not consider different traffic priorities Do not consider link failure scenario
Pasca et al. [20]	QoS improvement	Proposed an application-aware multipath flow routing framework integrating ML in SDN for traffic classification. The algorithm assigns paths based on QoS requirements of available parameters e.g., bandwidth and delay	Traditional static routing is slow to respond to network changes and slow to converge	NB, DT, Bayesian Network and SVM	Provide better routing configuration effectively reduces the network delay	<ul style="list-style-type: none"> Do not consider link failure Do not improve on delay
Nakayama et al. [8]	QoS improvement	Proposed a routing scheme based on the Markov Chain Monte Carlo algorithm to decrease the worst-case end-to-end delay of all CRAN front-haul flows and ensure that all flows fulfill the latency requirements	The current QoS-aware routing scheme ignores frame-level queuing delay.	Markov Chain Monte Carlo	All flows have a delay that is less than the threshold and meets the requirements.	<ul style="list-style-type: none"> Do not consider node failures or congestion Different class traffic is considered but did not mention explicitly
Stampa et al. [9]	QoS improvement	The authors designed and tested a deep RL agent in SDN that could significantly improve routing based on delay requirements	Limited routing optimization capabilities	Deep RL	Achieves improved delay when compared to the non-DRL agent routing scheme	<ul style="list-style-type: none"> Data samples used are 100,000 gravity generated traffic matrix Do not consider different traffic priorities Do not consider link failure and congestion
Mao et al. [19]	QoS improvement	Proposed Tensor-based Deep Belief Architectures (TDBA), in which traffic patterns from the edge router are fed into TDBA to build a path to all edge routers	Conventional routing cannot cope with the complex environment	DL	Achieves zero packet loss rate	<ul style="list-style-type: none"> High computational load Do not consider different traffic priorities Do not consider link failure condition

Alvizu et al. [6]	Predicting network parameters	Trained an ANN for forecasting the traffic load and variation and calculate the best resource allocation in SDN	Over-provisioning yields inefficiency and high operational costs	DL	The proposed scheme yields an optimality gap above 3%	<ul style="list-style-type: none"> • Training using a public dataset • Do not consider link failure
-------------------	-------------------------------	---	--	----	---	---

III. DEVELOPMENT OF THE HYBRID ML-IDS AND ML-RA

The ML-IDS dataset development is carried out via Graphical Network Simulator 3 (GNS3) by varying the network inputs. The normal incoming traffic is generated via the OSTINATO traffic generator, while the attack traffic is generated via Low Orbit Ion Cannon (LOIC). The output of the simulations in GNS3 for the ML-IDS algorithm will be the actual normal or attack label, while the output for ML-RA is the actual delay for all available routes between a S2D pair. All inputs and outputs are extracted via Wireshark and tabulated in a CSV file to build the ML-IDS and ML-RA datasets.

For ML-IDS, since it is a classification-based algorithm, the dataset is fed into MATLAB to train several ML classifiers. In contrast, ML-RA is a regression-based algorithm, and the ML-RA dataset is fed to train ML regression models. Both datasets are split into 70% training dataset and 30% testing dataset for performance evaluation during the algorithm development phase.

To further improve the performance of the ML models, the hyperparameters are optimized iteratively in MATLAB until the performance, i.e., error rate, converges to a constant value. Then, all the ML models are further tested using new data. The new data consists of new input features but without the actual output label. It is up to the ML models to provide predictions on the new data. The model which provides the most promising performance such as accuracy, precision, and F-measure are chosen for the proposed ML-IDS and ML-RA. Finally, both ML-IDS and ML-RA are cascaded together to build a new hybrid algorithm to enhance network security and improve network delay and throughput.

The simulation setup, the proposed algorithm, system parameters, and simulation scenarios are discussed in detail as follows.

A. The Simulation Setup

The network environment for the hybrid ML-IDS and ML-RA in the MPLS network system is depicted in Fig. 1. The network consists of eight edge routers, R1, R2, R5, R6, R7, R8, R9, and R10, which can be used as ingress or egress label edge routers (LER). Concurrently, R3 and R4 are normal label switch routers (LSRs) in the MPLS domain. All edge routers are linked to various types of traffic, such as VoIP, CCTV, and file transfers. The network is built in four ring topologies: 1) R1, R2, R3, R4, R5, and R6 form the main ring; 2) R1, R4, R5, R7, and R8 form the second ring and serve as node protection for R4; 3) R2, R3, and R9 form the third ring; and 4) the final ring is made up of R3, R6, and R10. The third and fourth rings protect the links between R2-R3 and R3-R6, respectively. In this network environment, all links are active, and the conventional RA and proposed ML-RA must compute the route for all traffic.

GNS3 is used to build the network for simulation, data collection, and performance analysis. OSTINATO and LOIC are used to generate normal and attack traffic, respectively. All edge routers are connected to either a Virtual PC (VPC), Windows virtual machine (VM), or OSTINATO traffic generator. The VPC is placed in the edge routers for traffic monitoring while the VM is used to mimic an actual Window's PC in the network for file transfers, generate DoS traffic and for traffic monitoring. Note that generated traffic can be sent using several streams simultaneously using different protocols at different rates.

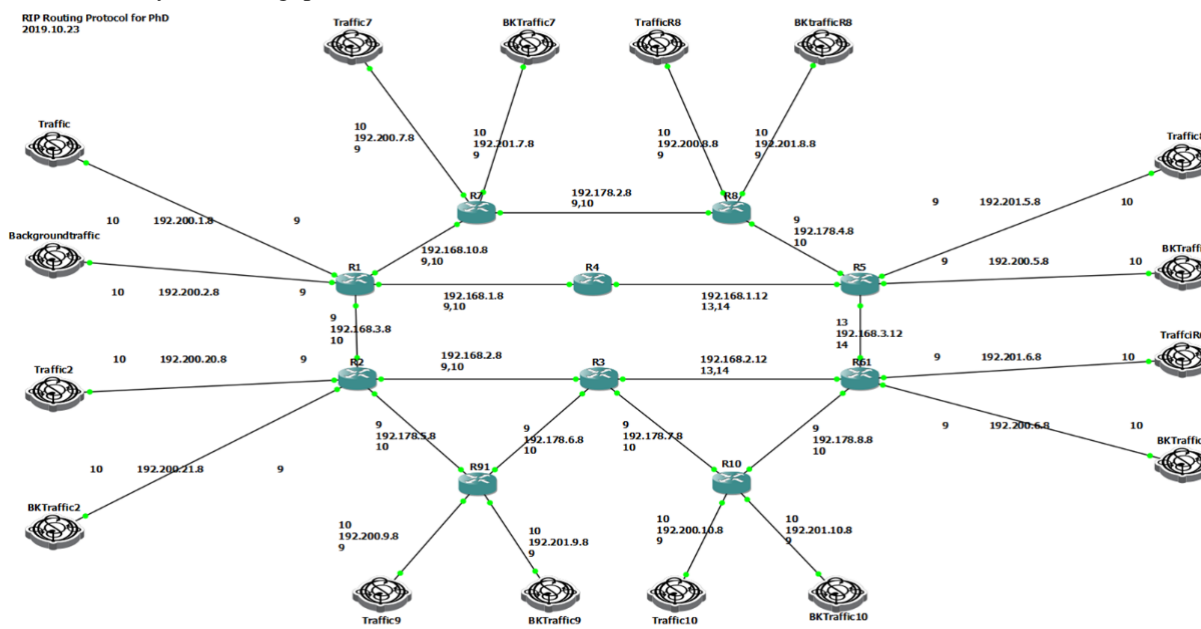


Fig. 1. Network environment in GNS3.

To launch a DoS attack, the LOIC tool is first installed in Window's VM in GNS3. After configuring the virtual Ethernet port of the VM, the IP address of the client or server is entered in the network as the DoS attack target, followed by the DoS method and packet flooding speed. The LOIC will flood the route leading to the targeted client. Once the DoS attack began, all network Virtual Private Clouds (VPCs) lost connectivity to the targeted client, while all clients connected to the route that links to the target client are also affected by the DoS attack. This scenario illustrated the severe damage of a DoS attack in the network domain as the attack affects not only the victim but also the devices linked to it.

B. Hybrid Supervised ML-IDS and ML-RA

This section discusses how ML-IDS and ML-RA are cascaded together to form complete ML-based security and QoS enhancement algorithms. Fig. 2 shows the framework of the proposed hybrid ML-IDS and ML-RA. The ML-IDS is incorporated at every ingress router that is R1, R2, R5, R6, R7, R8, R9, and R10, while the rest of the routers only focus on forwarding the traffic as computed by the ML-RA. ML-IDS is a classifier-based supervised ML that will predict the incoming traffic as normal or attack. In contrast, ML-RA, is a regression-based supervised ML that predicts the delay of all possible routes between the S2D pair.

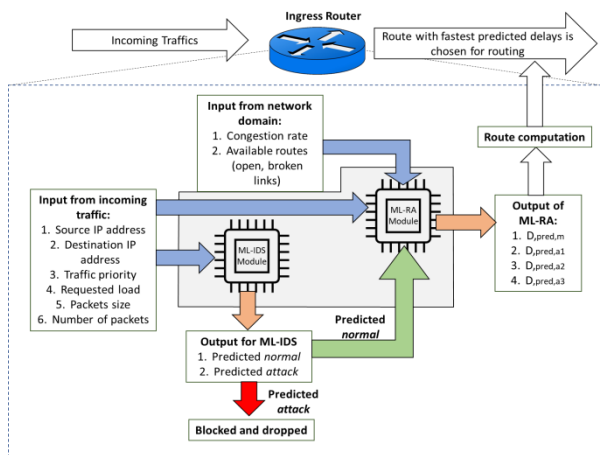


Fig. 2. Framework for the hybrid ML-IDS and ML-RA.

As incoming packets enter the ingress router, network information such as S2D IP address, traffic priority, requested load, packet size, number of packets, and data rates are fed into the ML-IDS. Additional network domain information, such as congestion rate and available routes, is required as inputs for the ML-RA. The output for the ML-IDS will classify the incoming traffic. If it is predicted as attack traffic, the ML-IDS will block and drop the traffic from entering the MPLS network. However, if it is predicted as normal by ML-IDS, the traffic will feed into the ML-RA module for route computation. To build the ML-IDS, the dataset constructed from the data generated using LOIC and OSTINATO is

tabulated in a CSV file format. The ML-IDS dataset is later fed into the Classification Learner App (CLA) in MATLAB, which trains model to classify data using supervised ML. This application allows users to import datasets, select features, specify validation schemes, train models, and assess results. Automated training of the ML algorithms allows users to select the models with the best classification model types. These automated training features ease the model development and evaluation process by eliminating the trial-and-error process to choose the best ML classifiers.

After importing the dataset into the CLA, the data is ready to be trained by a series of ML algorithms, and the best performing algorithm will be chosen for the proposed ML-IDS. The overall process to develop the proposed ML-IDS is as shown in Fig. 3. The dataset is split into a training dataset and test dataset with a split ratio of 70:30 [26]. However, the split ratio must be carefully adjusted to avoid over-fitting or under-fitting. A trial-and-error basis is generally adopted until the accuracy saturates.

The dataset of ML-RA is built in GNS3, covering the incoming traffic and network information from the MPLS domain. Simulations are conducted in GNS3 using different traffic parameters in the OSTINATO traffic generator for EF, AF, and BE traffic. For congestion, another OSTINATO traffic generator is run in GNS3 by bursting continuous packets. Broken links are simulated by simply closing the link in GNS3. The available routes in the network are performed by using RIPv2 routing protocol. RIPv2 was chosen due to ease of configurations in the Cisco emulator in GNS3 and because RIPv2 utilizes shortest-path routing scheme regardless of network conditions.

At first, a random S2D pair is chosen. Then, by using trace IP route in the Cisco Command Line Interface (CLI), GNS3 will show the main route computed by the RIPv2. Then, the main route is purposely closed to allow the RIPv2 to recompute to other available routes. The process is repeated until there are no more alternatives for S2D pair. For this research, the alternative routes are already trained by the ML-RA using the RIPv2 routing protocol. The advantage of this method is that ML-RA algorithm no longer needs to manually compute alternative routes in the network when a sudden surge of traffic in the network occurs. The ML-RA algorithm is trained so that it will predict the QoS parameters on all available routes and quickly assign a path for the traffic in any incoming packets and network conditions. For each S2D pair, different EF, AF, and BE traffic configurations, and network conditions are run in the simulations for several iterations to improve the ML-RA training rate. Since the delay and throughput of each iteration are not always precisely constant due to processing delay and limitation of the simulation platform, training it with several iterations will allow the ML-RA to foresee the patterns and trends in the network.

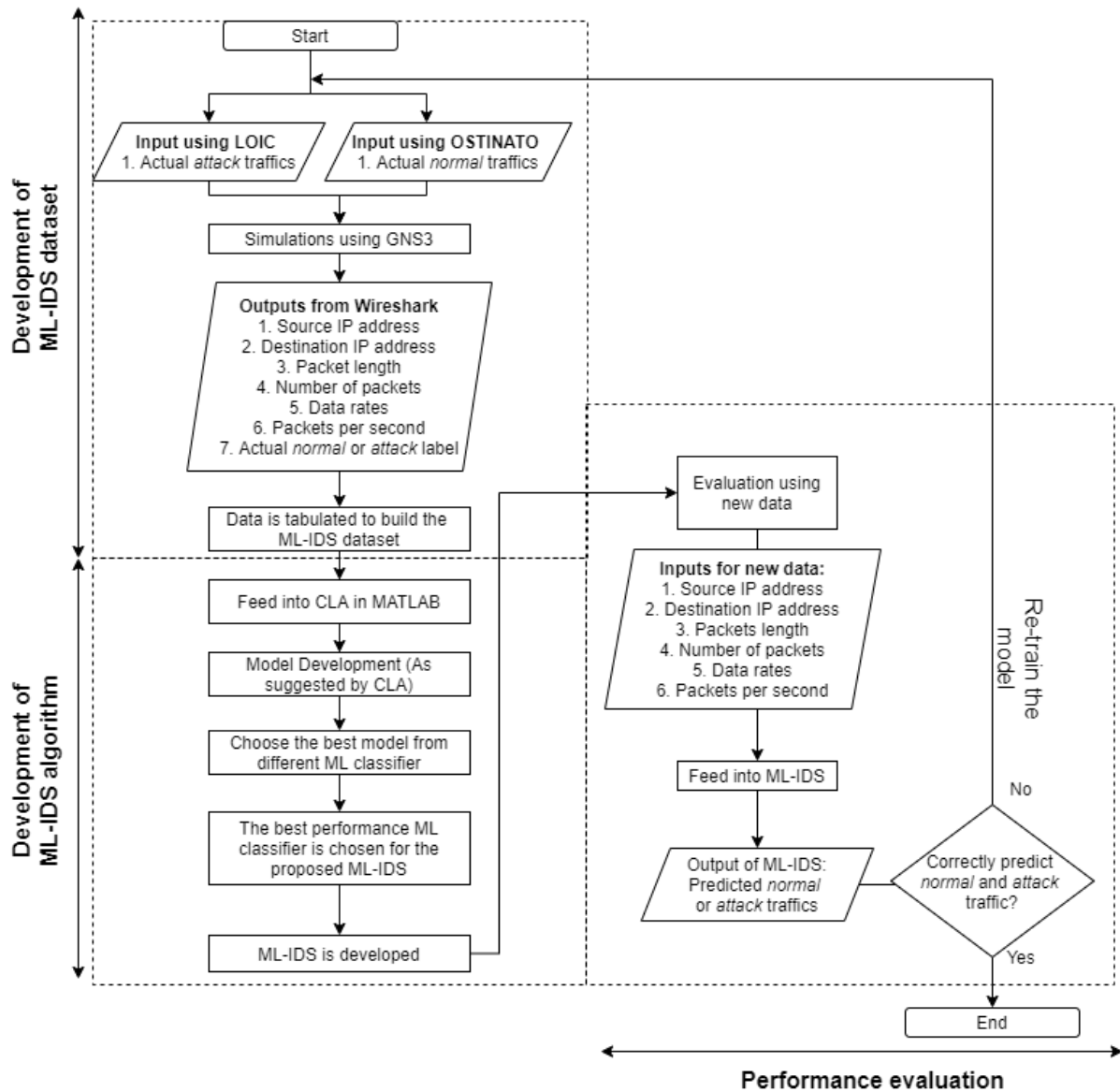


Fig. 3. Flowchart for the development of ML-IDS.

There are a total of 21 features for the ML-RA dataset. The first five input features are the information of the incoming traffic while the rest are on network conditions. To simplify the dataset, network congestion and broken links share the same column, where the value “0” denotes zero congestions, value “20” denotes congestions, while “100” denotes a broken or closed link. The data extracted from the simulations is up to half-a-million of iterations for all possible S2D pairs with different traffic and network conditions.

The ML-RA dataset is then fed into MATLAB's Regression Learner App (RLA) to create a regression model that predicts the delay for each route in the network. The RLA eases the ML development work by suggesting several regression models that fit the ML-RA dataset. For this case, the linear and tree-based regressions algorithms, including medium tree, course tree, and fine tree are suggested by the RLA. All the regression's algorithm is then compared with

their prediction speed, training time and RMSE. The best one among all will be chosen as the ML model for the ML-RA. Delay in this context is defined as the total time taken from the moment the first packets enter the MPLS domain to the time the last packets are received at the receiver end. The predicted delay by the ML-RA for the main route, Alternative Route 1 (ALT1), Alternative Route 2 (ALT2), and Alternative Route 3 (ALT3) are denoted as $D_{pred,m}$, $D_{pred,a1}$, $D_{pred,a2}$ and $D_{pred,a3}$, respectively. The predicted delays by the ML-RA will be used for route computation. The flowchart for the route computation is as shown in Fig. 4. The route computation for each traffic will be based on four network conditions, namely, all routes available, one or two routes down, three routes down, and all routes down based on the predicted delay. If the route is not available either due to broken links for node failures, the ML-RA will predict a value of “∞” to the route, indicating that the predicted delay is too high and should be avoided.

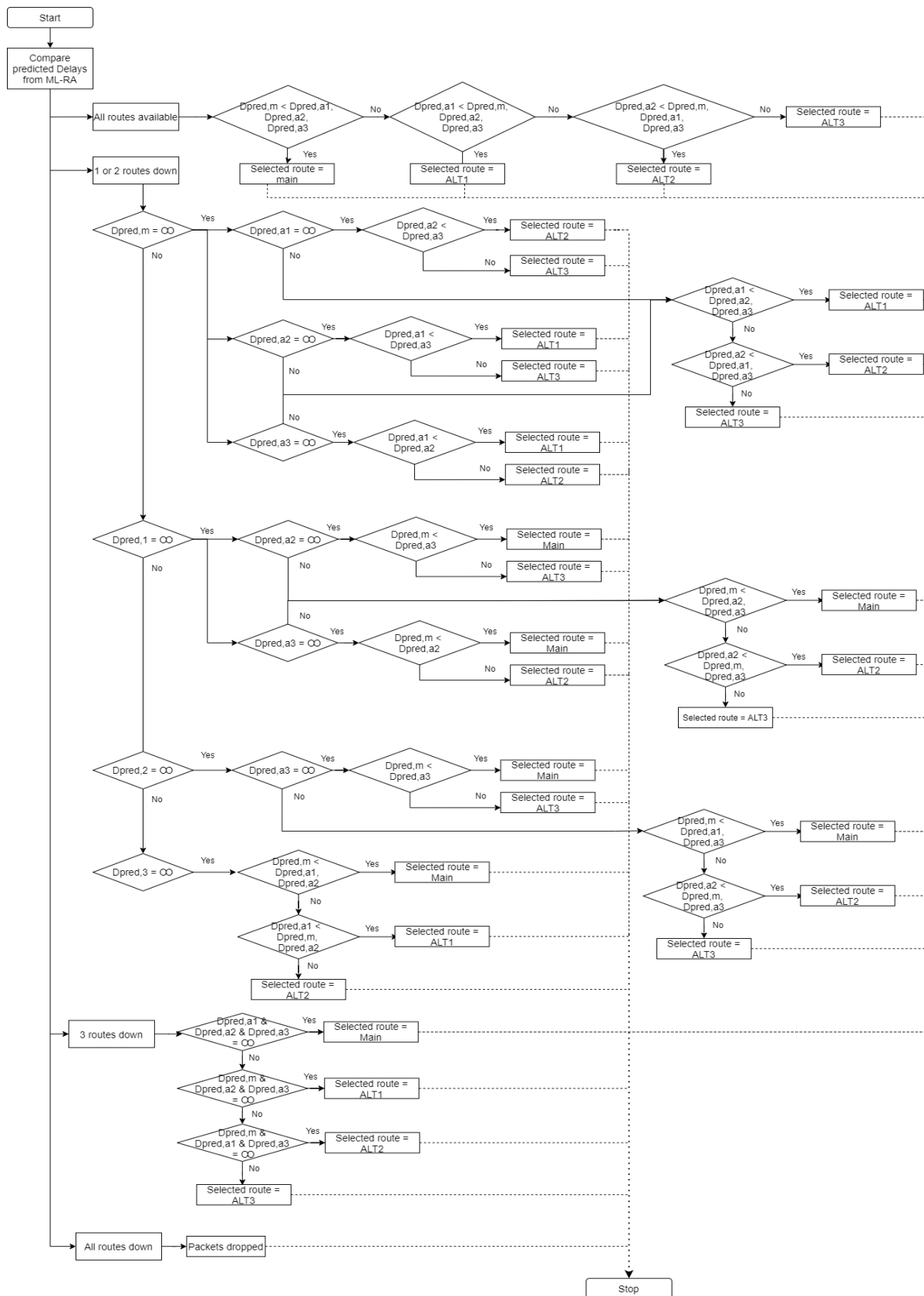


Fig. 4. Route computation flowchart for the proposed ML-based hybrid IDS and intelligent routing algorithm.

As shown in Fig. 4, when all routes are available, the ML-RA module will rank each of the predicted delays from lowest to highest. Then the route with the lowest predicted delays will be chosen for the traffic, starting with EF, followed by AF and BE traffic. In the event where all routes are down, ML-RA will drop all the packets until a route is back to available. One of the advantages of ML-RA is that from the predicted delay alone is already sufficient to understand the network conditions. For instance, when the link is congested, the predicted delay will be high. While, when the link is down or node failure, the predicted delay is “∞”. This eases the route computation works to choose the best route for the traffic without processing too much network information.

C. System Parameters

The design parameters for ML-IDS are shown in Table II. Note that the traffic generated by both OSTINATO and LOIC traffic generators in GNS3 follows the IEEE 802.3 standard, where the packet length is between 64 to 1520 [27]. The OSTINATO traffic generator is limited to only 20 Mbps for normal traffic, while LOIC generates attack traffic up to 40 Mbps, with up to 40,000 packets per second. However, due to limited computing storage, the number of packets per session is capped at 20,000 packets per session.

The design parameters for the ML-RA are as shown in Table III. The traffic in the network comprises of EF, AF, BE, and congested traffic. The requested load begins at 20% with an increment of 10% up to 100% according to the traffic mixture ratio of 20:40:40 for EF, AF, and BE, respectively. Our finding shows that the maximum network capacity in GNS3 is at 20 Mbps. With that, at 100% requested load, the throughput for EF, AF and BE is 4 Mbps, 8 Mbps, and 8 Mbps, respectively. The packet length for EF traffic is fixed at 160 bytes, following the Cisco bandwidth calculator [28]. Due to the limited computational storage of the GNS3 VM and ease of data extraction, the packet size for AF and BE traffic is fixed at 500 bytes. However, for congestion traffic, the length of the packet is set at random between 60 to 1550 bytes with the number of packets up to 20,000 packets per session using the “random” feature in OSTINATO.

D. Simulation Scenarios for Performance Study

Four S2D pairs are chosen for ML-RA validations: R6 to R2, R2 to R5, R10 to R7, and R5 to R3. These S2D pairs were selected because there are one main route and three alternative routes; namely ALT1, ALT2, and ALT3 available and it involves almost the entire link in the network, which is suitable for performance study. It is worth mentioning that there are also other S2D pairs available. However, because the four pairs with all the available routes already involve the entire links and LSRs in the network, it is deemed sufficient.

For ease of explanation on the routing, each of the routes in the network is given a unique LSP ID, as shown in Fig. 5. For instance, the route between R1 to R4 is LSP 1. For routes with more than one hop, for example, the route between R1 to R6 is denoted as LSP 1.2.3, which represents a total of three links. The main and alternative routes computed by the RIPv2 routing protocol for R6 to R2, R2 to R5, R10 to R7, and R5 to R3 are tabulated in Table IV. When the network is not

congested, the shortest path offers the best QoS parameters. RIPv2 will always compute the shortest path, which in this case, is the main route between an S2D pair regardless of network congestion. The downside is, when there is a sudden change in the network, RIPv2 may not be able to perform at its peak. When the main route is broken, the RIPv2 will recompute the next best route, which is the ALT1 and so forth.

To force the traffic to choose another alternative route, traffic engineering is configured via CLI in the Cisco router’s Internetwork Operating System. A sample case study is demonstrated in this paper, in which the main route for an S2D pair is congested, with one network route is down. The purpose of this sample study is to investigate how the proposed ML-RA can compute the path for different network conditions. The accuracy of the ML-RA is considered good when it can avoid congested routes and broken links. As a result, it is expected that the ML-RA will offer better QoS performance compared to RIPv2 that only considers the shortest path. The objective of ML-RA is to predict and compute the best routes in the network, regardless of the network conditions. The simulation details in GNS3 for ML-RA performance study are summarized in Table V for R6 to R2, R2 to R5, R10 to R7, and R5 to R3. The performance study focuses on the accuracy of the ML-RA to choose the route with the predicted best QoS. Then, the delay and throughput for the route are computed by the ML-RA and compared with RIPv2.

TABLE II. DESIGN PARAMETERS FOR THE PROPOSED ML-IDS

Design Parameter	Description
Packet length	60 bytes < Packet length < 1520 bytes
Number of packets	Up to 20,000 packets per session
Data rates	Up to 40 Mbps
Packets per second	Up to 40,000 packets per second for DoS attack
ML-based classifier	GBT, RF, DT, DL, FLM, LR, and GLM
Dataset	Proposed ML-IDS dataset and CICIDS-2018
Actual traffic type	The actual traffic from the dataset’s label (normal or attack)

TABLE III. DESIGN PARAMETERS FOR THE PROPOSED ML-RA

Design Parameter	Description
Traffic Priority	EF, AF and BE
Requested Load	20%, $\Delta \pm 10\%$ up to 100%
Packet Length	Background traffic (60 bytes to 1550 bytes), EF (160 bytes), and AF and BE (500 bytes)
Number of Packets	Up to 20,000 packets per session
Data Rates	Up to 20 Mbps for 100% load
Congestion Rates	Fixed at 20 Mbps
ML-based Regressions	Linear Model Regressions, Medium Tree Regressions, Fine Tree Regressions, and Course Tree Regressions
Dataset	Proposed ML-RA dataset
Routes	Main route, ALT1, ALT2, ALT3
Delay	Actual delay of the traffic from simulation during the training phase

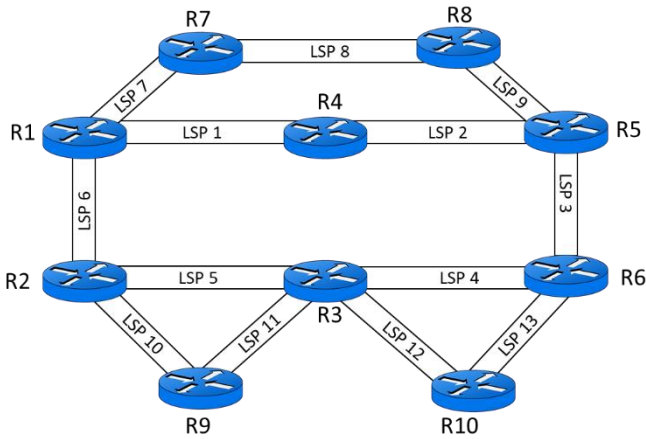


Fig. 5. Network environment with LSP ID.

TABLE IV. MAIN AND ALTERNATIVE ROUTES BY RIPV2 ROUTING PROTOCOL

Source router	Destination router	Main route (LSP)	ALT1 (LSP)	ALT2 (LSP)	ALT3 (LSP)
R6	R2	4.5	13.12.11.10	3.2.1.6	3.9.8.7.6
R2	R5	6.1.2	6.7.8.9	5.4.3	10.11.12.13.3
R10	R7	12.5.6.7	12.11.10.6.7	13.3.9.8	13.3.2.1.7
R5	R3	3.4	3.13.12	2.1.6.5	9.8.7.6.5

IV. RESULTS AND DISCUSSION

This section will discuss the results of the proposed system, along with some of its limitations and the future direction of the hybrid ML-IDS and ML-RA system.

A. Simulation Results

Delay is an essential parameter in evaluating the performance of ML-RA, especially for EF and AF traffic, as they require stringent delay requirements. There is no strict timing delay requirement for BE traffic. However, for comparison purposes, the delay for BE traffic is compared in different network conditions and different routing protocols and measured via the analysis tool in GNS3's Wireshark. Provided that it is within the delay standards, the lower the delay, the better the performance of ML-RA.

The next performance matrix evaluated is throughput, which is defined as the amount of data transferred between a S2D router to show the performance between ML-RA and RIPv2 RA [29]. AF and BE traffic require high amount of throughput compared to the EF traffic. It is also expected that the higher the congestion rate and the number of hop count for selected LSP will lower the throughput performance. Similar to delay, throughput is also measured via the analysis tool in GNS3's Wireshark.

Delay and throughput performance parameters are chosen for benchmarking purposes. This is because one of the problem statements is that the shortest path in the network will not provide the best QoS especially when the network resources are limited. The delay comparison between ML-RA and RIPv2 will show which RA performs the best with

different traffic and network conditions. The delay performance parameter is also helpful to evaluate the accuracy of the ML-RA to predict the fastest route in the network.

TABLE V. SIMULATION DETAILS FOR ML-RA PERFORMANCE STUDY

Source router	Destination router	Main route (LSP)
R6	R2	LSP 4.5 is congested and LSP 3 is closed
R2	R5	LSP 6.1.2 is congested and LSP 3 is closed
R10	R7	LSP 12.5.6.7 is congested and LSP 3 is closed
R5	R3	LSP 3.4 is congested and LSP 6 is closed

When the traffic reaches a receiver beyond the standard delay requirements, it is considered as packet loss. This shows that the delay parameter alone is sufficient to describe how the traffic is being forwarded in the network. Conversely, the throughput performance parameter is chosen to verify the capacity rate of the traffic to reach the destination. Having higher throughput proves that the route computed by the RA offers better quality for the incoming traffic. For the performance simulation, a total of four S2D pairs are chosen, which are R6 to R2, R2 to R5, R10 to R7, and R5 to R3 as summarized in Table V.

The sample case studies the performance of ML-RA when the main route is congested, and one of the routes to reach the destination router is down. With that, for all the S2D pairs, there are only two routes available, which are the congested main route and ALT1. Based on the traffic and network conditions, the routes computed by ML-RA for R6 to R2, R2 to R5, R10 to R7, and R5 to R3 are as shown in Table VI. The computed route for this case for all S2D pairs is the ALT1. The result is as expected where ML-RA accurately rerouted the traffic to the normal ALT1 as the main route is congested. As has been explained, since RIPv2 always chooses the shortest path regardless of network congestion, RIPv2 still computed the main route for routing. In terms of delay, it is expected that as the load increases, the delay also increases due to a higher number of packets being forwarded to the destination. The delay comparisons between ML-RA and RIPv2 for all S2D pairs are shown in Fig. 6. The lower the delay of the ML-RA compared to RIPv2, the better the performance of the algorithm.

In Fig. 6, the solid line represents the delay for the route computed by ML-RA, while the dotted line represents the delay for RIPv2. The delay for both routes computed by ML-RA and RIPv2 increase as the offered load for all traffic increases from 20% to 100%. However, the delay for ML-RA is lower compared to RIPv2 for all traffic types. This is because ML-RA rerouted the traffic immediately via the ALT1. The results prove that, although the ALT1 comprises of a higher number of hops, the delays are lower compared to the congested main route with a lesser number of hops. For instance, the delay for R6 to R2, as shown in Fig. 6(a), even though the main route comprises of only one hop, but since the main route is congested, the delay for ALT1 which comprises of three hops contributes to a lower delay.

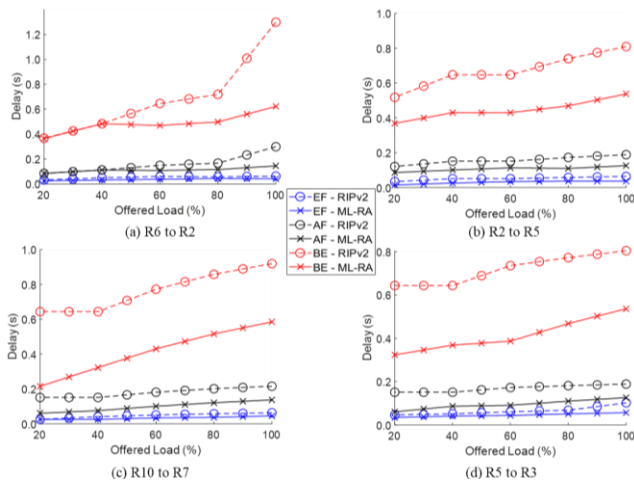


Fig. 6. Delay for the sample case.

TABLE VI. SIMULATION DETAILS FOR ML-RA PERFORMANCE STUDY

Source router	Destination router	Predicted fastest route by ML-RA
R6	R2	LSP 13.12.11.10
R2	R5	LSP 6.7.8.9
R10	R7	LSP 12.11.10.6.7
R5	R3	LSP 3.13.12

However, from 20% to 40% offered load for R6 to R2 S2D pair, the delay for BE traffic for both ML-RA and RIPv2 are almost in-line with each other as shown in Fig. 6(a). This is because the ratio of the number of hops between ALT1 and main route is almost triple, and the traffic from 20% to 40% offered load is too small to notice any delay difference. Nonetheless, beyond 40% offered load, the delay difference is more significant. At 100% offered load, the delay for BE traffic reduces from 1.2974 s to 0.6228 s for the route computed by ML-RA. While in Fig. 6(b) to 6(d), even though the difference in terms of number of hops between the main route and ALT1 is only by one additional hop, there is also a significant delay difference between the route computed by ML-RA and RIPv2. Since RIPv2 only computes the shortest available path in the network, the main route is chosen even though it will result in a higher delay. The delay comparison between ML-RA and RIPv2 for EF traffic for all pairs is not as significant as AF and BE traffic. This is because the data size and data rate for EF traffic are smaller compared to AF and BE.

The throughput is shown in Fig. 7. As opposed to delay, the higher the throughput offered by the ML-RA compared to RIPv2, the better the performance of the algorithm. Since the design parameters for the traffic is set at 20:40:40 traffic mixture ratio for EF, AF, and BE traffic, respectively, both AF and BE should theoretically have the same amount of throughput. However, due to the simulators' computation load, slight throughput fluctuation is expected. The throughput for all four S2D pairs shows a similar trend, which increases with increasing offered load from 20% to 100% for all traffic. However, since ML-RA successfully rerouted the traffic to the ALT1 for all four pairs, the throughput offered by the routes computed by ML-RA is higher compared to RIPv2.

For R6 to R2 S2D, the throughput for AF and BE traffic is almost in-line to each other from 20% to 60% offered load, as shown in Fig. 7. This is because the delay differences for both ML-RA and RIPv2 are not significant enough to produce a higher throughput difference. While for the rest of the S2D pairs where the number of hops difference between the congested main route and ALT1 is just addition by one, the delay difference between ML-RA and RIPv2 is even more significant which correlates to higher throughput difference.

The throughput and delay improvements for all S2D pairs are as shown in Table VII. In terms of throughput, the improvements for EF, AF, and BE traffic are up to 75.0%, 57.1%, and 57.1%, respectively. While in terms of delay, the improvements are up to 50.0%, 44.4%, and 44.4%, respectively. The results prove the superiority of ML-RA to predict and compute the fastest route in the network. For all S2D pairs, ML-RA learned from historical data that for this network condition, the fastest route would be ALT1. As a result, ML-RA intelligently computes ALT1 as the preferred route in the future when given the similar network condition.

From the results, the ML-RA is proven to be able to predict the route with the lowest delay and highest throughput, outperforming the RIPv2. ML-RA avoided the closed and congested routes and accurately computed the other faster alternative routes. Even when all routes are congested, ML-RA accurately predicts the fastest route in the network. Even though the route computed by ML-RA is the same as RIPv2, the goal of predicting the fastest path is met and maintained. Although only four S2D pairs are considered, the main and alternative routes cover the entire network, and almost all the LSRs in the network are involved. Thus, the ML-RA is expected to perform similarly for other S2D pairs.

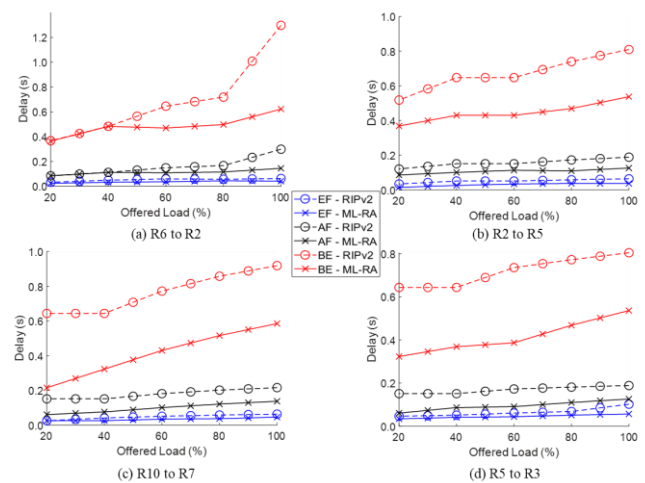


Fig. 7. Throughput for the sample case.

TABLE VII. THROUGHPUT AND DELAY IMPROVEMENT BY ML-RA

Source	Destination	Delay improvement (%)			Throughput improvement (%)		
		EF	AF	BE	EF	AF	BE
R6	R2	50.0	44.4	44.4	33.3	51.3	52.0
R2	R5	42.8	33.3	33.6	75.0	50.0	50.0
R10	R7	27.7	36.3	36.3	38.4	57.1	57.1
R5	R3	44.4	33.3	33.3	22.2	50.0	50.0

B. Limitations and Future Work of the Hybrid ML-IDS and ML-RA System

From the performance evaluations of the proposed ML-IDS and ML-RA, the system shows 100% accuracy in predicting the fastest route, as well as shows better performance against the RIPv2 in terms of delay and throughput. Despite the great performance of the proposed system, there may be several limitations that can be further improved in future research. For example, the proposed system relies on simulation data generated in GNS3 for dataset development. Note that the dataset was constructed using EF, AF, and BE traffic following the standard of VoIP, CCTV, and file transfers. However, this dataset may not accurately reflect the traffic patterns and network behavior of real-world environments. Thus, it may be advantageous to validate the performance of the system with real-world traffic data to confirm its practicality and effectiveness. Additional research on the influence of the parameters for the dataset on the performance of the system may also be investigated. The system's performance could also be assessed against different datasets in addition to evaluating its performance and scalability on larger and more complicated networks. Moreover, other ML techniques such as deep learning or reinforcement learning could be integrated to improve the accuracy and resilience of the proposed system. It is also advantageous to validate with real-world traffic data to confirm its practicality and effectiveness. Finally, the proposed system's implementation in a real-world environment could be performed to evaluate its applicability and feasibility in practical scenarios.

V. CONCLUSION

This study develops a hybrid ML-IDS and ML-RA algorithm to enhance the MPLS network's resiliency and improve traffic QoS. The proposed ML-IDS is a classification-based ML algorithm that learns the traffic pattern at the ingress router. Based on historical data, ML-IDS predicts and classifies the incoming traffic as normal or attack. The predicted attack traffic will be denied access to the network domain and discarded. While the predicted normal traffic will be queued according to their priority. This study prioritizes EF, AF, and BE traffic, which correspond to VoIP, CCTV, and data transfers, all having unique delay and bandwidth requirements. The GNS3 network setup is used for simulation and data collection. In particular, the LOIC traffic generator is used in the setup to simulate a DoS attack, while the OSTINATO traffic generator is used to generate the EF, AF, and BE traffic. Another OSTINATO traffic generator is used to burst background traffic in the network to simulate network congestion. The data in the network is collected as datasets. The output label for the ML-IDS and ML-RA datasets is the actual traffic type (normal or attack) and the actual delay for all available routes for all S2D pairs. The datasets are fed into MATLAB, which is then used to train classifiers for the ML-IDS algorithm and regressions for the ML-RA algorithm. For the performance evaluation, the ML-RA algorithm is compared to RIPv2. From the performance evaluations, the ML-RA shows 100% accuracy in predicting the fastest route in the network. During network congestion, the proposed ML outperforms the RIPv2

in terms of delay and throughput on average by 57.61% and 46.57%, respectively.

REFERENCES

- [1] S. Premkumar and V. Saminadan, "Performance evaluation of smart grid communication network using MPLS," Proc. 2017 IEEE Int. Conf. Commun. Signal Process. ICCSP 2017, vol. 2018-Janua, pp. 2116–2120, Feb. 2018, doi: 10.1109/ICCSP.2017.8286779.
- [2] G. Andresini et al., "Multi-Channel Deep Feature Learning for Intrusion Detection," IEEE Access, vol. 8, pp. 53346–53359, 2020, doi: 10.1109/ACCESS.2020.2980937.
- [3] M. A. Ridwan et al., "Applications of Machine Learning in Networking: A Survey of Current Issues and Future Challenges," IEEE Access, vol. 9, pp. 52523–52556, 2021, doi: 10.1109/ACCESS.2021.3069210.
- [4] P. Mishra et al., "A detailed investigation and analysis of using machine learning techniques for intrusion detection," IEEE Commun. Surv. Tutorials, vol. 21, no. 1, pp. 686–728, Jan. 2019, doi: 10.1109/COMST.2018.2847722.
- [5] K. Yu et al., "Machine learning driven network routing," 2019 6th Int. Conf. Syst. Informatics, ICSAI 2019, pp. 705–712, Nov. 2019, doi: 10.1109/ICSAI48974.2019.9010507.
- [6] R. Alvizu et al., "Matheuristic With Machine-Learning-Based Prediction for Software-Defined Mobile Metro-Core Networks," J. Opt. Commun. Networking, Vol. 9, Issue 9, pp. D19–D30, vol. 9, no. 9, pp. D19–D30, Sep. 2017, doi: 10.1364/JOCN.9.000D19.
- [7] G. Choudhury et al., "Two use cases of machine learning for SDN-enabled IP/optical networks: Traffic matrix prediction and optical path performance prediction [Invited]," J. Opt. Commun. Netw., vol. 10, no. 10, pp. D52–D62, Oct. 2018, doi: 10.1364/JOCN.10.000D52.
- [8] Y. Nakayama et al., "Low-latency routing for fronthaul network: A Monte Carlo machine learning approach," IEEE Int. Conf. Commun., Jul. 2017, doi: 10.1109/ICC.2017.7996439.
- [9] G. Stampa et al., "A Deep-Reinforcement Learning Approach for Software-Defined Networking Routing Optimization," Sep. 2017.
- [10] T. Hendriks et al., "Q 2 -Routing: A Qos-aware Q-Routing algorithm for Wireless Ad Hoc Networks," Int. Conf. Wirel. Mob. Comput. Netw. Commun., vol. 2018-October, pp. 108–115, Dec. 2018, doi: 10.1109/WIMOB.2018.8589161.
- [11] A. Martin et al., "Network resource allocation system for QoE-aware delivery of media services in 5G networks," IEEE Trans. Broadcast., vol. 64, no. 2, pp. 561–574, Jun. 2018, doi: 10.1109/TBC.2018.2828608.
- [12] R. Boutaba et al., "A comprehensive survey on machine learning for networking: evolution, applications and research opportunities," J. Internet Serv. Appl., vol. 9, no. 1, pp. 1–99, Jun. 2018, doi: 10.1186/s13174-018-0087-2.
- [13] F. Tang et al., "On Removing Routing Protocol from Future Wireless Networks: A Real-time Deep Learning Approach for Intelligent Traffic Control," IEEE Wirel. Commun., vol. 25, no. 1, pp. 154–160, Feb. 2018, doi: 10.1109/MWC.2017.1700244.
- [14] L. Li et al., "Naïve Bayes classifier-assisted least loaded routing for circuit-switched networks," IEEE Access, vol. 7, pp. 11854–11867, 2019, doi: 10.1109/ACCESS.2019.2892063.
- [15] H. Yao et al., "Machine learning aided load balance routing scheme considering queue utilization," IEEE Trans. Veh. Technol., vol. 68, no. 8, pp. 7987–7999, Aug. 2019, doi: 10.1109/TVT.2019.2921792.
- [16] H. Yao et al., "A machine learning approach of load balance routing to support next-generation wireless networks," 2019 15th Int. Wirel. Commun. Mob. Comput. Conf. IWCMC 2019, pp. 1317–1322, Jun. 2019, doi: 10.1109/IWCMC.2019.8766546.
- [17] M. Salani et al., "Routing and Spectrum Assignment Integrating Machine-Learning-Based QoT Estimation in Elastic Optical Networks," Proc. - IEEE INFOCOM, vol. 2019-April, pp. 1738–1746, Apr. 2019, doi: 10.1109/INFOCOM.2019.8737413.
- [18] C. V. Murudkar and R. D. Gitlin, "Optimal-Capacity, Shortest Path Routing in Self-Organizing 5G Networks using Machine Learning," 2019 IEEE 20th Wirel. Microw. Technol. Conf. WAMICON 2019, Apr. 2019, doi: 10.1109/WAMICON.2019.8765434.

- [19] B. Mao et al., "A Tensor Based Deep Learning Technique for Intelligent Packet Routing," 2017 IEEE Glob. Commun. Conf. GLOBECOM 2017 - Proc., vol. 2018-Janua, pp. 1–6, Jul. 2017, doi: 10.1109/GLOCOM.2017.8254036.
- [20] S. T. V. Pasca et al., "AMPS: Application aware multipath flow routing using machine learning in SDN," 2017 23rd Natl. Conf. Commun. NCC 2017, Oct. 2017, doi: 10.1109/NCC.2017.8077095.
- [21] V. Vashishth et al., "GMMR: A Gaussian mixture model based unsupervised machine learning approach for optimal routing in opportunistic IoT networks," Comput. Commun., vol. 134, pp. 138–148, Jan. 2019, doi: 10.1016/j.comcom.2018.12.001.
- [22] V. Vashishth et al., "A Machine Learning Approach Using Classifier Cascades for Optimal Routing in Opportunistic Internet of Things Networks," Annu. IEEE Commun. Soc. Conf. Sensor, Mesh Ad Hoc Commun. Networks Work., vol. 2019-June, Jun. 2019, doi: 10.1109/SAHCN.2019.8824952.
- [23] Z. M. Fadlullah et al., "Value Iteration Architecture Based Deep Learning for Intelligent Routing Exploiting Heterogeneous Computing Platforms," IEEE Trans. Comput., vol. 68, no. 6, pp. 939–950, Jun. 2019, doi: 10.1109/TC.2018.2874483.
- [24] S. Troia et al., "Machine-Learning-Assisted Routing in SDN-Based Optical Networks," Eur. Conf. Opt. Commun. ECOC, vol. 2018-Septe, Nov. 2018, doi: 10.1109/ECOC.2018.8535437.
- [25] N. A. Mohamed Radzi, "Recent Dynamic Bandwidth Allocation Algorithm Methods in Ethernet Passive Optical Network," Int. J. New Comput. Archit. their Appl., vol. 4, no. 4, pp. 167–176, 2014, doi: 10.17781/P0016.
- [26] H. Brink et al., Real-World Machine Learning. Manning, 2016.
- [27] LAN/MAN Standards Committee of the IEEE Computer Society, "IEEE Standard for Ethernet," IEEE Stand. Ethernet, 2018.
- [28] "Cisco Community," CISCO, 2010. <https://community.cisco.com/t5/voice-over-ip/bandwidth-calculation/td-p/1649134> (accessed Jun. 28, 2022).
- [29] S. U. Masruroh et al., "Performance evaluation of routing protocol RIPv2, OSPF, EIGRP with BGP," in 2017 International Conference on Innovative and Creative Information Technology (ICITech), 2017, pp. 1–7, doi: 10.1109/INNOCIT.2017.8319134.

Plant Disease Classification and Adversarial Attack based CL-CondenseNetV2 and WT-MI-FGSM

Yong Li¹, Yufang Lu^{2*}

School of Information Science and Engineering, Guilin University of Technology
Guilin, China

Abstract—In recent years, deep learning has been increasingly used to the detection of pests and diseases. Unfortunately, deep neural networks are particularly vulnerable when attacked by adversarial examples. Hence it is vital to explore the creation of intensely aggressive adversarial examples to increase neural network robustness. This paper proposes a wavelet transform and histogram equalization-based adversarial attack algorithm: WT-MI-FGSM. In order to verify the performance of the WT-MI-FGSM, we propose a plant pests and diseases identification method based on the coordinate attention mechanism and CondenseNetV2: CL-CondenseNetV2. The accuracy of CL-CondenseNetV2 on the PlantVillage dataset is 99.45%, which indicates that the improved CondenseNetV2 model has a more significant classification performance. In adversarial sample experiments using WT-MI-FGSM and CL-CondenseNetV2, experimental results show that when CL-CondenseNetV2 is attacked by the adversarial algorithm WT-MI-FGSM, the error rate reaches 89.8%, with a higher attack success rate than existing adversarial attack algorithms. In addition, the accuracy of CL-CondenseNetV2 is improved to 99.71% by adding the adversarial samples generated by WT-MI-FGSM to the training set and performing adversarial training. The experiments demonstrate that the adversarial examples caused by WT-MI-FGSM can improve the model's performance.

Keywords—Adversarial examples; FGSM; plants diseases and pests; attention mechanism; CondenseNetV2

I. INTRODUCTION

A country's agricultural sector is vital to its economic growth, with the potential to both stimulate and directly impact the national economy's development or stagnation. The stability of agriculture is intrinsically linked to social stability and national self-sufficiency; therefore, it is crucial to achieve the steady and sustainable development of agriculture. To achieve sustainable development in the agricultural field, we must first perform well in the prevention and control of crop diseases and insect pests, guarantee that the prevention and control measures are scientific and safe, successfully control diseases and insect pests, and promote eco-logical development in the agricultural field in a benevolent and sustainable direction [1].

Since the advent of deep learning, deep learning-based picture identification has been a popular topic in the image recognition community, finding widespread application in areas such as facial recognition, transportation, and healthcare. The most traditional network models are GoogLeNet, VggNet, and ResNet [2]-[4]. In recent years, the agricultural sector has also made extensive use of deep learning. Progress has been

made in the identification of plant diseases and pests as a result of the extensive research undertaken by scientists. Based on the AlexNet network, LV and others use batch normalization, PRelu activation function, etc. to improve network convergence and avoid over-fitting. They also combine extended convolution and multi-scale convolution to improve network feature extraction ability, demonstrating that the algorithm of feature enhancement can effectively improve the network's feature extraction ability and recognition accuracy [5]. Pandian et al. utilized image enhancement technology based on image processing and deep learning to improve the crop disease data set. Additionally, they expanded and improved the data set with the antagonistic generating network and neural pattern transfer using migration learning technology. Using this method, the experimental findings show that the improved data set can reach higher precision [6]. Durmus used tomato photos from the PlantVillage dataset to train numerous deep neural networks, and the accuracy of networks such as SqueezeNet significantly improved due to this [7-8]. Using plant images in the visible spectrum, Lily proposed a straightforward and reliable method for diagnosing plant diseases [9]. In his research work, Kaur proposed the DAG-ResNet model and utilized it to discover a number of tomato illnesses [10]. The accuracy was 98.8%. ALVARE et al. integrate FasterR-CNN, SSD, RFCN, VggNet, ResNet, and other feature extractors to obtain a notable recognition and classification effect [11]. Wenliang Tang used conditional convolution, channel attention module, and knowledge distillation to improve the model [12]. The accuracy was 97.6%. Agriculture diseases and pests have a high degree of similarity, a more dispersed and intricate distribution, a greater difficulty in classification, and a greater need for classification networks. Based on the classic CondenseNetV2 model, this work introduces CL-CondenseNetV2, a method for identifying agriculture diseases and pests that combines the coordinate attention mechanism and CondenseNetV2. The model incorporates a flexible and lightweight coordinate attention mechanism and embeds the position information into the channel attention in order to detect and identify the target area with greater precision. As a result of these enhancements, the CL-CondenseNetV2 model now has an identification accuracy of 99.45%, allowing for highly precise detection of agricultural diseases and pests.

Although deep neural networks perform well in most classification tasks, they are vulnerable when faced with adversarial samples. Adversarial samples are a class of samples formed by intentionally adding subtle perturbations to a dataset, which can induce the network model to misclassify

*Corresponding Author

and threaten the model's safety. However, on the other hand, for model designers, adversarial samples can be used as an effective tool to evaluate the security and robustness of the model. They can effectively improve the correctness and security of the model classification through adversarial training. Adversarial attack algorithms can be classified into two categories according to the mainstream classification methods: black-box attack and white-box attack, and white-box attack algorithms include FGSM, DeepFool, C&W, etc. [13]-[15]. Black-box attack algorithms currently have single-pixel and local search attack algorithms, etc. [16].

However, there are fewer examples of improving the classification success of the model by adding adversarial examples for adversarial training. This paper proposes a new DNN called CL-CondenseNetV2 that adds a coordinate attention module to CondenseNetV2. A significant number of comparative studies have shown that our network model performs well. In addition, this paper introduces wavelet variation and histogram equalization in the image domain based on the MI-FGSM algorithm to propose a new adversarial attack algorithm WT-MI-FGSM. The adversarial examples generated by this algorithm can be used to train CL-CondenseNetV2. Training with adversarial examples will help us to improve the accuracy of the classification of plant diseases and the security and robustness of the model.

II. ADVERSARIAL ATTACK ALGORITHM

FGSM, a white-box attack technique built on the production of adversarial example gradients, is the most widely used adversarial assault algorithm today. By iteratively computing the gradient, Alexey et al introduced I-FGSM, which significantly enhances the fit of the adversarial sample to the model [17]. After introducing the momentum factor based on I-FGSM, Dong et al suggested the MI-FGSM approach, which greatly increased the success rate of black-box assaults and successfully enhanced the migrability of the generated adversarial samples [18]. The MI-FGSM algorithm function is shown in (1) and (2).

$$g_{t+1} = \mu \cdot g_t + \frac{\nabla_{x_t^{adv}} J(x_t^{adv}, y)}{\|\nabla_{x_t^{adv}} J(x_t^{adv}, y)\|_1} \quad (1)$$

$$x_{t+1}^{adv} = Clip_{x_t}^{\varepsilon} x_t^{adv} + \alpha \cdot sign(g_{t+1}) \quad (2)$$

In the formula, $sign()$ is a symbolic function, and $\nabla_x J(x, y)$ represents a gradient. ε is the size of the neighborhood. x_{t+1}^{adv} is the adversarial example generated by iterating $t+1$ times. t represents the number of iterations. The step length can be obtained by $\alpha = \varepsilon/T$. It ensures that the adversarial examples generated are in the neighborhood of x . Where μ is the decay factor of the momentum term, and g_{t+1} denotes the cumulative gradient iteration $t+1$ times. The role of the $Clip$ function is to constrain the adversarial examples within the ε -neighborhood of the original image x to satisfy the Infinite norm constraint.

In this study, wavelet transform and histogram equalization are successfully integrated with MI-FGSM to create the WT-MI-FGSM, a more effective adversarial attack algorithm. The overall flow of WT-MI-FGSM is shown in Fig. 1 below. By means of an adversarial attack algorithm, the original example is utilized to produce an adversarial example. First, the wavelet transform and histogram equalization are performed on the origin example to obtain a $224 \times 224 \times 3$ image. The loss function is then computed using the acquired images as input into the model. Iteratively updating the example along the gradient of the loss function is followed by the addition of perturbations. If the requirements are unmet, the iteration will continue until it succeeds. Finally, output confrontational examples. The WT-MI-FGSM algorithm function is shown in (3). Where D is the image enhancement function. It includes wavelet transform and histogram equalization.

$$g_{t+1} = \mu \cdot g_t + \frac{\nabla_{x_t^{adv}} J(D(x_t^{adv}), y)}{\|\nabla_{x_t^{adv}} J(D(x_t^{adv}), y)\|_1} \quad (3)$$

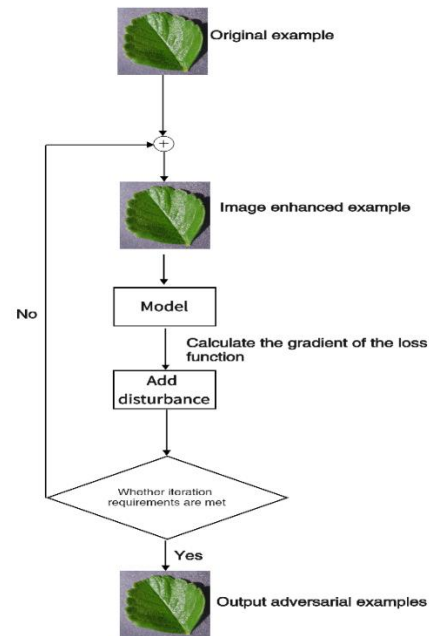


Fig. 1. WT-MI-FGSM attack concrete steps.

III. CL-CONDENSENETV2 MODEL DESIGN

A. CondenseNetV2

Huang Gao's team proposed DenseNet and CondenseNet in 2017, and DenseNet establishes a dense connection mechanism that allows each layer in the network to be directly connected to its preceding layer in the same block to achieve feature reuse, which enables DenseNet to reduce the total number of parameters and improve efficiency significantly [19]. CondenseNet introduces a full-dense connection and pruning mechanism based on DenseNet [20]. The full-dense connection enables the network to establish a dense connection between different blocks while combining with average pooling to achieve stitching between feature maps of various sizes, thus gaining more robust feature reuse. CondenseNet can achieve

almost the same accuracy as DenseNet with just 1/10 the training time thanks to the pruning mechanism, which enables the network to prune the irrelevant weights during the training phase and reduce network redundancy.

Nevertheless, features in DenseNet and CondenseNet will remain the same once they are formed, drastically ignoring the potential value of some features. CondenseNetV2, a powerful yet lightweight neural network based on CondenseNet, was suggested by Gao Huang in 2021 [21]. CondenseNetV2 introduces a sparse feature reactivation mechanism that enables the network to learn to choose a few potentially redundant features. The efficiency of the deep network's feature reuse is increased by concurrently cropping and updating these redundant features to make them better suited to feature learning. CondenseNetV2 achieves better performance than DenseNet and CondenseNet at a low computational cost, and achieves excellent performance on image classification and detection tasks. Table I shows the CondenseNetV2 network structure.

TABLE I. CONDENSENETV2 NETWORK STRUCTURE

Layers	Input	ConenseNetV2
Convolution	224×224	3×3 Conv , stride 2
Dense Block (1)	112×112	$\begin{bmatrix} 1 \times 1L - Conv \\ 3 \times 3G - Conv \\ SFR \text{ module} \end{bmatrix} \times 4$ (k=8)
Transition Layer (1)	112×112	2×2 average pool, stride2
Dense Block (2)	56×56	$\begin{bmatrix} 1 \times 1L - Conv \\ 3 \times 3G - Conv \\ SFR \text{ module} \end{bmatrix} \times 6$ (k=16)
Transition Layer (2)	56×56	2×2 average pool, stride2
Dense Block (3)	28×28	$\begin{bmatrix} 1 \times 1L - Conv \\ 3 \times 3G - Conv \\ SFR \text{ module} \end{bmatrix} \times 8$ (k=32)
Transition Layer (3)	28×28	2×2 average pool, stride2
Dense Block (4)	14×14	$\begin{bmatrix} 1 \times 1L - Conv \\ 3 \times 3G - Conv \\ SFR \text{ module} \end{bmatrix} \times 10$ (k=64)
Transition Layer (4)	14×14	2×2 average pool, stride2
Dense Block (5)	7×7	$\begin{bmatrix} 1 \times 1L - Conv \\ 3 \times 3G - Conv \\ SFR \text{ module} \end{bmatrix} \times 8$ (k=128)
Classification Layer	1×1	7×7 global average pool 1000D fully-connected, SoftMax

B. Introduction of Coordinate Attention Mechanism

Attention Mechanism is a unique structure within a machine learning model that is used to automatically calculate and learn the contribution of input data to output data. Common modules for attention mechanisms include SE, CBAM, etc. SE is only concerned with the weighting of channels [22]. Despite the fact that CBAM simultaneously considers the weight allocation of channels and spaces, redundant convolution pooling operations result in the loss of

some useful information [23]. Coordinate Attention (CA) mechanism is an attention mechanism that can embed position information into channel attention introduced in this paper [24]. In comparison to SE, CBAM, and other attention mechanisms, this attention mechanism is not only capable of capturing cross-channel information, but also direction perception and position perception information, in addition to being lightweight and flexible. Coordinate attention operations consist of coordinate information embedding and coordinate attention generation. Coordinate attention module is shown in Fig. 2.

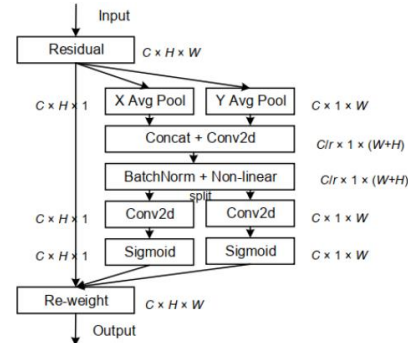


Fig. 2. Coordinate attention module.

1) *Coordinate information embedding*: Channel attention frequently employs two-dimensional global pooling to encode global spatial information, but this operation typically makes it challenging to save target location data. In order to prevent this, the Coordinate attention mechanism decomposes the two-dimensional pooling of channel attention into two one-dimensional feature coding processes and collects features along two spatial directions. Equation (4) and formula (5) represent the horizontal and vertical coordinate coding operations of input x .

$$z_c^h(h) = \frac{1}{W} \sum_{0 \leq i \leq W} x_c(h, i) \quad (4)$$

$$z_c^w(w) = \frac{1}{H} \sum_{0 \leq j \leq H} x_c(j, w) \quad (5)$$

Where x_c denotes the c th channel component of the input data, h and w respectively reflect the data's height and width. These two coding operations allow the attention module to capture the long-term dependency along one spatial direction and to store precise location information along the other spatial direction, thereby assisting the network in more precisely identifying the target information of pests and diseases on crops.

2) *Coordinate attention generation*: In order to make full use of the global receptive field obtained through the embedding operation of coordinate information and encode the accurate position information, first embed the coordinate information into the obtained features for concatenate operation, and then send them into the convolution module with the shared convolution core of 1×1, reduce its dimension to the original C/r , and then send the feature map F_1 after

batch normalization into the nonlinear activation function to obtain the feature map f in the form of $1 \times (W + H) \times C/r$. Equation (6) is shown below.

$$f = \delta \left(F_1 \left(\left[z^h, z^w \right] \right) \right) \quad (6)$$

Where $[\cdot, \cdot]$ is the splicing operation along the spatial dimension, δ is the nonlinear activation function, F_1 is the convolution change function, and f is the intermediate feature map that encodes the spatial information in the horizontal and vertical directions.

After the above operations, f is decomposed into f^h and f^w along the spatial dimension, and 1×1 convolution and nonlinear activation operations are performed on them to obtain the attention weights g^h and g^w of the feature map in the horizontal and vertical coordinate directions respectively. Equation (7) and (8) are as follows.

$$g^h = \sigma \left(F_h \left(f^h \right) \right) \quad (7)$$

$$g^w = \sigma \left(F_w \left(f^w \right) \right) \quad (8)$$

Where σ is the sigmoid activation function, F^h and F^w represent the convolution change function of the characteristic components f^h and f^w , respectively.

Finally, expand g^h and g^w , calculate the coordinate attention mask by matrix multiplication, and act on the input to get the output Y of the attention module:

$$y_c(i, j) = x_c(i, j) \times g_c^h(i) \times g_c^w(j) \quad (9)$$

C. CL-CondenseNetV2

The above coordinate attention is added to the CondenseNetV2 network to obtain the basic structure of CL-CondenseNetV2 as shown in Fig. 3. In each layer of the proposed network, LGC is first used to select important features for feature learning, and after obtaining new features, the SFR module is used to reactivate the previous features. On this basis, we added the coordinate attention module. The coordinate attention module improves the recognition accuracy of the model by making the network model lightweight while enabling the model to locate and identify the target region more accurately. Since plant disease features are distributed in different positions on the front of leaves, the classification network needs to accurately pay attention to the spatial location of disease features. Therefore, coordinate attention can significantly improve the recognition accuracy of plant diseases

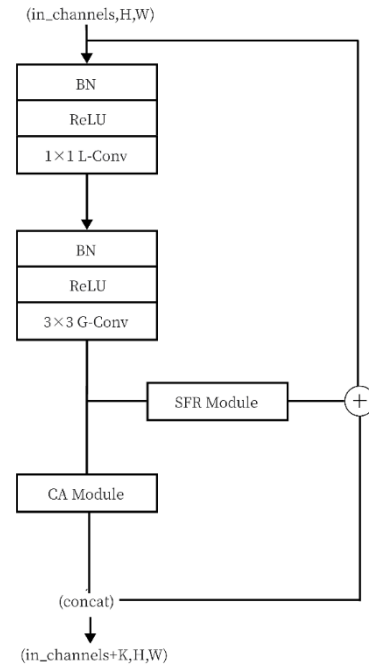


Fig. 3. CL-CondenseNetV2 structure.

IV. PREPARATION FOR THE EXPERIMENT

A. Experimental Environment

The experiment's environment setting is as follows: The graphics card is an NVIDIA GeForce RTX 3060, and the Windows 64-bit system CPU is an 8-core, 16-thread AMD Rayon R7-5800H processor. The memory is DDR4 16G, and the hard drive storage is 512G SSD. The software environment consists of Anaconda 4.10.3 and CUDA 11.6. The model is built and trained using the Python programming language and PyTorch framework.

B. Parameter Setting

In the model training parameter settings, the training batch size is set to 16, the test batch size to 8, and the number of iterations to 50 rounds (Epochs). The employed optimizer is SGD (Stochastic Gradient Descent) [25]. The rate of learning is set at 0.001. Adopting the learning rate exponential decay approach. Gamma has been adjusted to 0.9. Loss is represented by the SoftMax cross-entropy loss function. The definition of the function is:

$$L = - \sum_{k=1}^n \sum_{i=1}^C t_{ki} \lg y_{ki} \quad (10)$$

Where, n is the pixel of the picture; t_{ki} is the probability that pixel k belongs to the category i ; y_{ki} is the probability of predicting the pixel k as the category i for the classification network.

C. Data Set

The data set utilized in this experiment is PlantVillage, which consists of tens of thousands of photos of healthy and diseased plants annotated by plant pathologists and is available for free download at www.PlantVillage.org. All photographs in the PlantVillage database were captured at experimental research stations affiliated with American institutions (Pennsylvania, Florida, Cornell, etc.). The data set consists of 54303 health and disease images split into 38 categories, and it is still growing [26]. The image and caption of several plants pest data sets are depicted in Fig. 4.

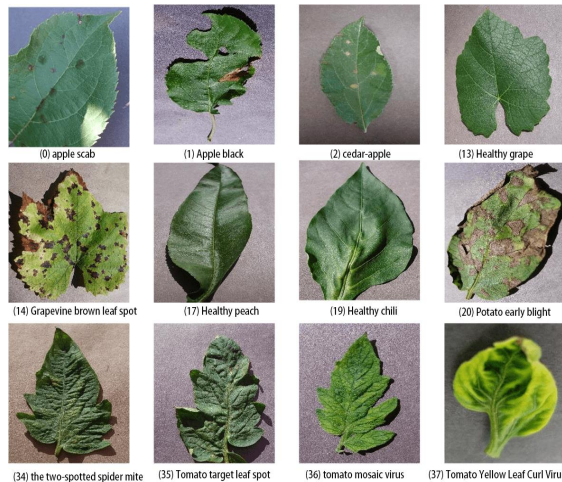


Fig. 4. Examples of images and labels of some crop pest data sets.

D. Data Preprocessing

By using a data enhancement approach, this work also increases the dataset. First, the data image is separated into a training set and a test set with an 8:2 ratio, and then the training set's data image is expanded to 81454 images to serve as the expanded training set. The resolution of the data image is finally rebuilt to 256×256 resolution, and the image is then standardized using the mean and standard deviation. Not only can data preparation imitate the actual agricultural setting and increase the diversity of training samples, but it can also improve the model's robustness and prevent overfitting.

E. Transfer Learning

In the realm of artificial intelligence, there exists a method known as "transfer learning". First, acquire knowledge in the source domain, and then apply it to the target domain, so that the target domain can achieve superior learning outcomes [27]. We can use this technique to reduce the number of training samples required by the model, eliminate the time-consuming and inefficient "ab initio" training process, accelerate network model training, and improve their overall learning efficiency. It has found widespread application in the field of image classification.

The improved model employs the model fine-tuned transfer learning method, employs the CondenseNetV2 pre-training model trained on the ImageNet large open dataset, and combines the transfer learning fine-tuning method to apply its parameters to the CL-CondenseNetV2 model, and uses it to identify plants diseases and pests.

F. Performance Metrics

For each of the experiments examined in this study, the evaluation metrics Accuracy, Precision, Recall, and Specificity are used to assess how well the network performed in identifying the test pictures. The evaluation metrics are calculated using the following equation.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (11)$$

$$Precision = \frac{TP}{TP + FP} \quad (12)$$

$$Recall = \frac{TP}{TP + FN} \quad (13)$$

$$Specificity = \frac{TN}{TN + FP} \quad (14)$$

Where TP is the true case, FP is the false positive case, FN is the false negative case, and TN is the true negative case. Accuracy is the percentage of samples that the model properly recognizes and categorizes to the total samples, which is often used to evaluate the overall accuracy of the model. Precision is the ratio of true cases to the number of positive cases classified by the model; Recall is the ratio of true cases to all positive cases; Specificity is the ratio of true negative cases to all negative cases.

V. PREPARATION FOR THE EXPERIMENT OF CLASSIFICATION

A. Experiment of Classification

1) Comparison of experimental effects between CL-CondenseNetV2 and CondenseNetV2: The original CondenseNetV2 and CL-CondenseNetV2 are trained on the extended dataset to evaluate how well the improved approach of this model works. Fig. 5 and 6 depict a comparison of the accuracy curve and loss value curve of the original network and the modified network, while Table II depicts a comparison of the accuracy and loss value results.

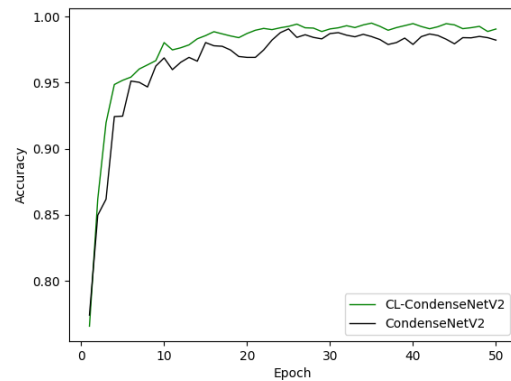


Fig. 5. Comparison of accuracy curves between CL- CondenseNetV2 and CondenseNetV2.

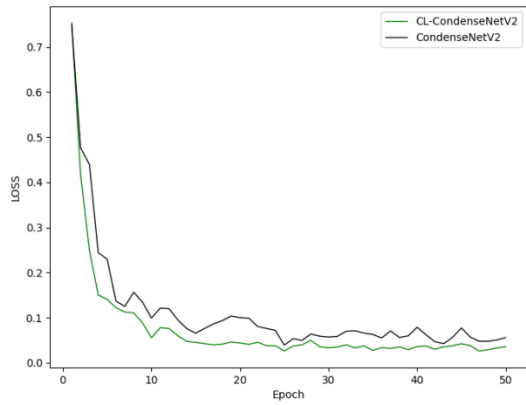


Fig. 6. Comparison of loss curve between CondenseNetV2 and CL-CondenseNetV2.

TABLE II. ACCURACY AND LOSS OF CL- CONDENSENETV2 AND CONDENSENETV2

Algorithms	Acc (%)	Pre (%)	Recall (%)	Spe (%)	Params (M)	Loss
CL-CondenseNetV2	99.45	98.85	98.17	99.89	6.1	0.0259
CondenseNetV2	99.07	96.69	97.03	99.22	6.1	0.0389

As can be shown in Fig. 5 and 6, the CL-CondenseNetV2 suggested in this study is superior than the CondenseNetV2 network model when it comes to the classification and recognition of pictures, as well as the recognition of crop diseases and insect pests. The accuracy and speed of CL-CondenseNetV2's accuracy convergence are faster than those of the original network, and the loss value curve is more consistently steady. According to Table I, the recognition rate of the original network is 99.07 percent, whereas the upgraded network model enhances the recognition rate of crop diseases and pests by 0.38 percent, demonstrating the viability of the improved model CL-CondenseNetV2.

2) *Comparison of experimental effects between CL-CondenseNetV2 and other models:* To further validate the benefits of the CL-CondenseNetV2 network model, employ three traditional networks, Vgg16, ResNet18, and ResNet50, to train on the improved data set and conduct comparative experiments with CL-CondenseNetV2 in the same experimental environment, using the same training parameters and training timeframes. Fig. 7 and 8 depict a comparison between the accuracy curve and the loss value curve.

Fig. 7 demonstrates that the CL-CondenseNetV2 network model maintains certain advantages. CL-CondenseNetV2's accuracy curve converges more rapidly during training, is more stable, and can essentially maintain its accuracy advantage over Vgg16, ResNet18, and ResNet50.

Fig. 8 shows that the CondenseNetV2 loss value curve has fallen greatly and rapidly, and that the loss value curve is more steady than those of Vgg16, ResNet18, and ResNet50, indicating that the network is more robust. This demonstrates that the CL-CondenseNetV2 network model is preferable. Table III displays a comparison of the experimental outcomes of each model.

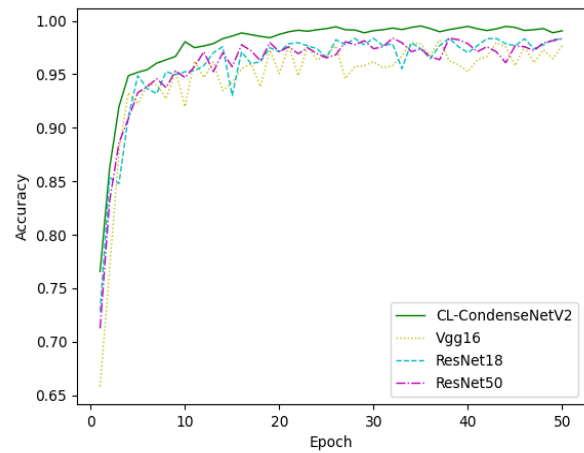


Fig. 7. Comparison of accuracy curves of each model.

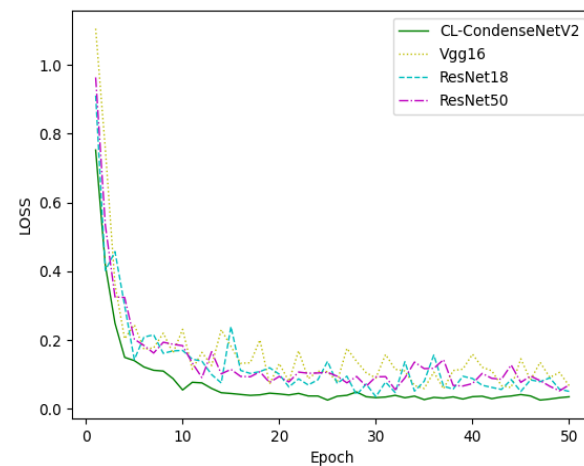


Fig. 8. Comparison of loss curve of each model.

TABLE III. ACCURACY AND LOSS OF EACH MODEL

Algorithms	Acc (%)	Pre (%)	Recall (%)	Spe (%)	Params (M)	Loss
CL-CondenseNetV2	99.45	98.85	98.17	99.89	6.1	0.0259
ResNet18	98.35	96.48	93.57	99.38	11.2	0.0573
ResNet50	98.37	95.37	95.52	99.85	23.5	0.0527
Vgg16	98.21	96.18	96.65	99.34	134.4	0.0596

Table III shows that under the same experimental conditions, CL-CondenseNetV2 achieved the highest accuracy of 99.45, the lowest loss value of 0.0259, and the least parameters of 6.1M, which were superior to Vgg16, ResNet18, and ResNet50. The accuracy of CL-CondenseNetV2 is 1.08 percentage points higher than the highest of the remaining deep learning models, ResNet50, while the number of parameters is 17.4M less than that of ResNet50. When compared to Vgg16, whose parameters is as high as 134.4M, the reductions are even more significant. CL-CondenseNetV2 outperforms other methods when taking into consideration both the network model's parameters and the recognition accuracy, indicating that the enhanced model may be put to better use in the detection of plant diseases and pests.

B. Experiment of Adversarial Example

1) *Preparation for the experiment:* To fairly validate the performance of the WT-MI-FGSM proposed in this paper, the experimental preparation for the adversarial attack experiments is approximately the same as when a model such as CL-CondenseNetV2 is trained. The experiments use the PlantVillage dataset. This experiment utilized smaller perturbations to make them more difficult to identify rather than setting the hyperparameters as the norm in the momentum method. Using small perturbations can increase the attack algorithm's success rate when compared to other adversarial attack methods. The maximum perturbation $\epsilon = 0.3$; the number of iterations $T = 10$; the step size $\alpha = 0.03$; and the fading factor $\mu = 1.0$. Controlled studies employing the white-box attack techniques of I-FGSM, MI-FGSM, and FGSM, respectively, were also carried out to further confirm the efficacy of the performance of WT-MI-FGSM.

2) *Experimental results:* This experiment contrasts four white-box adversarial attack algorithms—WT-MI-FGSM, I-FGSM, MI-FGSM, and FGSM—attack Vgg16, ResNet50, ResNet18, CondenseNetV2, and CL-CondenseNetV2, and generates both adversarial and original cases, as illustrated in Fig. 9. According to the experimental findings, all five models are susceptible to adversarial assaults, and all four of these attacks have a high success rate. The attacked models' recognition accuracy was lowered by roughly 91%. It is important to note that the differences between the original instances and the adversarial examples produced by the WT-MI-FGSM given in this research are negligible and challenging for the human eye to detect. The adversarial examples and original examples generated by the experiment are shown in Fig. 9.

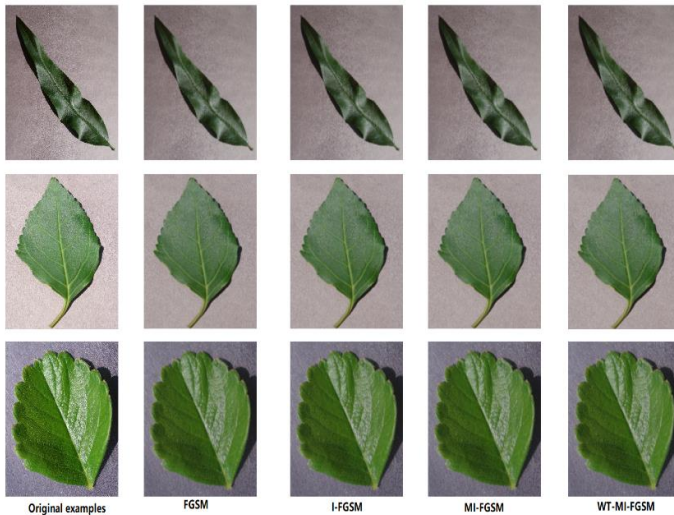


Fig. 9. Adversarial examples generated by various algorithms.

The recognition rate of each model reduces dramatically when attacked by the adversarial attack algorithm, as seen in Table IV. Vgg16 has the fastest drop when attacked by the WT-MI-FGSM algorithm proposed in this research, with a recognition rate of 0.6%, followed by ResNet18 and ResNet50,

with a recognition rate of 2.8% and 3.1%. And closely followed by CondenseNetV2, with a recognition rate of 5.3%. CL-CondenseNetV2 still has the greatest recognition rate after receiving the attack, with 10.2%. In addition, Table IV compares the performance of WT-MI-FGSM with other traditional adversarial attack algorithms. The success percentages of FGSM, I-FGSM, and MI-FGSM after CL-CondenseNetV2 received assaults from each algorithm are 85.9%, 86.5%, and 87.3%, respectively. They are all less than 89.8% of WT-MI-FGSM. WT-MI-FGSM outperforms MI-FGSM, which has the highest attack success rate among classic attack algorithms, by 1.9%. It is observed that the proposed adversarial attack algorithm WT-MI-FGSM has the best performance in this paper.

TABLE IV. ACCURACY AND LOSS OF EACH MODEL

Algorithms	ResNet 18	ResNet 50	Vgg1 6	CondenseNet V2	CL-CondenseNet V2
FGSM	88.5%	88.2%	92.1%	86.1%	85.9%
I-FGSM	89.8%	89.3%	92.4%	87.4%	86.5%
MI-FGSM	91.2%	90.1%	93.2%	89.1%	87.3%
WT-MI-FGSM	97.2%	96.9%	99.4%	94.7%	89.8%

Adversarial examples can increase model accuracy and robustness. The adversarial examples generated by WT-MI-FGSM are added to the training set, and the individual models are adversarial trained. Table V shows the accuracy of each model after training. CL-CondenseNetV2 has an accuracy of 99.71%, which is 0.26% greater than without adversarial training. Other models' accuracy has also increased. The experimental results show that adversarial examples generated by WT-MI-FGSM can improve model performance.

TABLE V. THE EXPERIMENT OF ADVERSARIAL TRAINING

Models	mAP (%)	mAP(%) (Slow)	mAP(%) (Medium)	mAP(%) (Fast)
CL-CondenseNetV2	99.71	99.07	99.97	99.57
CondenseNetV2	99.38	97.29	99.88	97.93
ResNet18	98.69	93.76	99.67	96.79
ResNet50	98.73	95.75	99.85	95.78
Vgg16	98.63	94.56	99.57	96.85

C. Discussion

To further verify that the CL-CondenseNetV2 model which is added to adversarial examples has a higher recognition rate of plant diseases and pests, it is compared with the DAG-ResNet model in literature [10] and the CondConvSENet detection model in literature [12]. The experimental results are shown in the Table VI.

TABLE VI. EXPERIMENTAL RESULTS

Models	Acc (%)
CL-CondenseNetV2 (Add adversarial examples)	99.71
CL-CondenseNetV2	99.45
DAG-ResNet	98.80
CondConvSENet	97.60

As can be seen from the table, the recognition rate of the CL-CondenseNetV2 model is higher than that of other models, while the classification recognition rate of the CL-CondenseNetV2 model after adding counter samples is far higher than that of other models. The feasibility and necessity of adding adversarial samples in model training are illustrated.

VI. CONCLUSIONS

In this paper, the proposed CL-CondenseNetV2 based on CA attention and CondenseNetV2 effectively improves the network's attention to feature space and enhances the accuracy of identifying agricultural diseases and pests. CL-CondenseNetV2 obtains 99.45% recognition accuracy in comparative studies with many models, outperforming classic CondenseNetV2, ResNet18, ResNet50, and Vgg16. This paper proposes a new adversarial attack algorithm WT-MI-FGSM based on MI-FGSM with the introduction of wavelet transform and histogram equalization. The comparison experiments use different adversarial attack algorithms against various models. The experimental results reveal that WT-MI-FGSM has a greater attack success rate than FGSM, I-FGSM, and MI-FGSM when compared to conventional adversarial attack methods, and the perturbations are too small to be recognized by human eyes. Furthermore, the adversarial samples generated by WT-MI-FGSM are added to the training set. After adversarial training, the recognition rate of CL-CondenseNetV2 may reach 99.71%, which is 0.26% higher than the accuracy rate without adversarial training, effectively increasing the model recognition's accuracy and robustness. Adversarial training is an efficient method for increasing the model's robustness. However, it has drawbacks such as sluggish training speed and overfitting when trained on tiny data sets. As a result, enhancing the performance of adversarial training will be the main focus of future study.

REFERENCES

[1] Di, T. Analysis on integrated control of plant diseases and pests in landscaping. World Tropical Agriculture Information, 2022 ,05: 63-64.
[2] Szegedy, C. et al. (2014) Going deeper with convolutions, arXiv.org. Available at: <https://arxiv.org/abs/1409.4842>.
[3] SIMONYAN,Karen;ZISSERMAN,Andrew.Very deep convolutional networks for large-scale image recognition.arXiv preprint arXiv:1409.1556, 2014.
[4] HE,Kaiming,et al. Deep residual learning for image recognition. In:Proceedings of the IEEE conference on computer vision and pattern recognition. 2016. p. 770-778.
[5] Lv M, Zhou G, He M, et al. Maize leaf disease identification based on feature enhancement and dms-robust alexnet[J]. IEEE Access, 2020, 8: 57952-57966.
[6] Pandian J A, Geetharamani G, Annette B. Data augmentation on plant leaf disease image dataset using image manipulation and deep learning

techniques[C]. 2019 IEEE 9th International Conference on Advanced Computing (IACC). IEEE, 2019: 199-204.
[7] Durmus, H.; Gunes, E.O.; Kirci, M. A hybrid approach for noise reduction-based optimal classifier using genetic algorithm: A case study in plant disease prediction. In Proceedings of the 2017 6th International Conference on Agro-Geoinformatics, Fairfax V A, USA, 7–10 August 2017; pp. 1–5.
[8] Iandola, F.N.; Moskewicz, M.W.; Ashraf, K.; Han, S.; Dally, W.J.; Keutzer, K. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <1MB model size. arXiv 2016, arXiv:1602.07360.
[9] Guadarrama, L.; Paredes, C.; Mercado, O. Plant Disease Diagnosis in the Visible Spectrum. Appl. Sci. 2022, 12, 1023–1049.
[10] Kaur, M.; Bhatia, R. Development of an improved tomato leaf disease detection and classification method. In Proceedings of the 2019 IEEE Conference on Information and Communication Technology (CICT), Jeju, Korea, 6–18 October 2019; pp. 1–5.
[11] Fuentes A, Yoon S, Kim S C, et al. A robust deep-learning-based detector for real-time tomato plant diseases and pests recognition[J]. Sensors, 2017, 17(9): 2022.
[12] Tang, W.; Huang, Z. Lightweight model of tomato leaf diseases identification based on knowledge distillation. Jiangsu J. Agric.Sci. 2021, 37, 9.
[13] Goodfellow, I.J.; Shlens, J.; Szegedy, C. Explaining and Harnessing Adversarial Examples. arXiv 2019, arXiv:1412.6572.
[14] Moosavi-Dezfooli, S.; Fawzi, A.; Frossard, P. DeepFool: A simple and accurate method to fool deep neural networks.In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 2574–2582
[15] Carlini, N.; Wagner, D.A. Towards evaluating the robustness of neural networks. In Proceedings of the 2017 IEEE Symposium on Security and Privacy (SP), San Jose, CA, USA, 22–26 May 2017; pp. 39–57.
[16] Huang, S.; Papernot, N.; Goodfellow, I.; Duan, Y.; Abbeel, P. Adversarial attacks on neural network policies. arXiv 2017, arXiv:1702.02284.
[17] Kurakin, A., Goodfellow, I. J., & Bengio, S. (2018). Adversarial examples in the physical world. In Artificial intelligence safety and security (pp. 99-112). Chapman and Hall/CRC.
[18] Dong, Y., Liao, F., Pang, T., Su, H., Zhu, J., Hu, X., & Li, J. (2018). Boosting adversarial attacks with momentum. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 9185-9193).
[19] HUANG, Gao, et al. Densely connected convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition.2017. p. 4700-4708.
[20] Huang, G., Liu, S., Van der Maaten, L., & Weinberger, K. Q. (2018). Condensenet: An efficient densenet using learned group convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2752-2761).
[21] Yang, L., Jiang, H., Cai, R., Wang, Y., Song, S., Huang, G., & Tian, Q. (2021). Condensenet v2: Sparse feature reactivation for deep networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 3569-3578).
[22] Hu, J. et al. (2019) Squeeze-and-excitation networks, arXiv.org. Available at: <https://arxiv.org/abs/1709.01507>.
[23] Woo, S. et al. (2018) CBAM: Convolutional Block Attention Module, arXiv.org. Available at: <https://arxiv.org/abs/1807.06521>.
[24] Hou, Q., Zhou, D. and Feng, J. (2021) Coordinate attention for efficient mobile network design, arXiv.org. Available at: <https://arxiv.org/abs/2103.02907>.
[25] Robbins, H.E. A Stochastic Approximation Method. Ann. Math. Stat. 2007,22, 400–407. [SGD]
[26] Hughes D, Salathé M. An open access repository of images on plant health to enable the development of mobile disease diagnostics[J]. arXiv preprint arXiv: 1511.08060, 2015.[17]
[27] Bousmalis, K. et al. (2016) Domain separation networks, arXiv.org. Available at: <https://arxiv.org/abs/1608.06019>.

EMOCASH: An Intelligent Virtual-Agent Based Multiplayer Online Serious Game for Promoting Money and Emotion Recognition Skills in Egyptian Children with Autism

Hussein Karam Hussein Abd El-Sattar

Department of Mathematics & Computer Science, Ain Shams University, Faculty of Science,
Abbassia 11566, Cairo, Egypt

Abstract—Autism, often known as "autism spectrum disorders (ASD)," is one of the most common developmental disabilities that affect how people learn, behave, communicate, and interact with others. Two crucial everyday tasks that people with ASD typically struggle with are managing finances and recognizing emotions. As the online gaming sector grows and develops, the question of why this type of media can't be used as a useful educational tool for those with ASD arises. This paper discusses this issue via a novel virtual agent-based multiplayer online serious game referred to as "EMOCASH," which aims to improve these important tasks for Egyptian children with ASD and achieve transfer of acquired knowledge to real-world situations via a 3D virtual shop scenario that was designed using the Autism ASPECTSS™ Design Index. EMOCASH served as an instrument for investigating the following research question: What role does technology play in the education of those with ASD? Numerous sub-questions that were related to the primary question were also addressed. A variety of usability metrics were used to assess effectiveness, efficiency and satisfaction aspects.

Keywords—Autism; virtual agents; serious games; digital technology; AI; online gaming; usability and accessibility

I. INTRODUCTION

Individuals with ASD frequently experience difficulties in daily living skills, including conceptual, social, or practical skills. The importance of daily living skills for people with ASD is especially important since they have a direct impact on the growth of abilities like self-determination and autonomy [1]. Numerous studies have used technological interventions to enhance the daily living (DL) skills for individuals with learning disabilities, including housekeeping [2], first aid training [3], money management skills [4]–[7], learning math subject [8], raise awareness about the effects of littering in [9] and diabetes in [10]. Scholars discovered that people with ASD commonly suffer from mindreading, or theory of mind (ToM) skills [11]–[14], and anthropomorphism, which is the extension of ToM to non-human agents [15]–[16]. ToM refers to our ability to observe the mental states (intentions, emotions, knowledge, and beliefs) of other people. Previous work [17]–[21] reported that autistic people's ToM skills are less developed when they mind-read human beings than when they mind-read non-human agents having anthropomorphic traits, such as animals and cartoons. For instance, the authors of [21]

demonstrate how anthropomorphizing faces—that is, placing them on objects like vehicles—can help people with ASD better recognize emotions. We carried out a study by expanding this idea to assist Egyptian children with ASD in practicing handling money, recognizing emotions, and teaching them how to behave properly in a shopping center using a collaborative authentic learning environment (CALE) instead of the conventional on-screen display apps. The app shows autistic children the potential everyday activities that could happen when a child goes shopping. Contrary to prior approaches, the learning environment was designed using the Autism ASPECTSS™ Design Index [22], [23] to adapt to the needs of all types of users. Moreover, the scenario supports multiplayer, which is more engaging and motivating than single-player training games. We expand on these foundations. The paper is organized as follows: the problem statement, the research questions that were raised, and the purpose of this paper are presented in Section II. Section III discusses previous studies connected to both money management and emotion recognition for people with ASD. Some theoretical background is summarized in Section IV. Section V introduces the methodology used. The development processes of the proposed system, including system architecture, design, and implementation used to develop the system, are presented in Section VI. The clinical characteristics and practices of the participants are listed in Section VII. Section VIII describes the different usability metrics used to assess the effectiveness, efficiency, and user satisfaction of the produced application, along with a comparison with existing techniques. Section IX concludes the paper and highlights future work.

II. OBJECTIVES AND RESEARCH QUESTIONS

While transitioning from adolescence to maturity, everyone must take a huge and difficult step. Because of their impairment symptoms, people with ASD typically have significant problems carrying out basic daily tasks on their own. Traditional intervention techniques, including the Picture Exchange Communication System (PECS) [24], social skills classes, narrative, role-playing, and so on, are costly, time-consuming, staff intensive, boring for patients due to exercise repetition, and usually have long waiting lists, rendering them ineffective. Internet-driven rapid advancements in ICT, such as serious games (SGs), VR, AR, virtual agents, AI, and robot,

among others [25]-[30] have opened creative and promising scenarios for therapists to improve DL skills for individuals with ASD. However, studies on how using authentic learning environments and engaging in online games affect learning outcomes for those with ASD are noticeably lacking. Taking into account the importance of user-built environment interaction for autistic persons, this paper addressed these issues using an innovative virtual agent-based multiplayer online SG called "EMOCASH," which served as an instrument for investigating the following research questions: (1) Is it possible to employ agent technology to create intelligent agents that can be included in online games for persons with ASD? (2) What advantages do online games provide for people with ASD? (3) How can we improve the appeal and motivation of education for those with ASD? (4) Which learning environments are most effective for this purpose? (5) What aspects of game elements are taken into consideration when using online gaming to educate autistic people? Concerning the first and second questions, the usability of virtual agent technology via online gaming for people with ASD is investigated in this paper using CALE. Online gaming enhances user experience, encourages immersion, social interaction and competitiveness, reduces autism symptoms, and promotes the satisfaction of working independently. When playing an online game, there is no eye contact or facial expression required, which can help reduce anxiety and any other unusual behaviors. Online games give people with ASD the opportunity to interact with their peers in a safe and less stressful learning environment that does not require regular face-to-face interaction; as a result, they can build communication skills that promote friendships and socializing (For instance, please refers to [31]-[34]). Regarding the third and fourth questions, we employed the Autism ASPECTSS™ Design Index [22] to accommodate specific autistic needs as a design development tool to support environments conducive to learning for those with ASD. The concepts of authentic learning [35]-[36] and authentic design thinking [37]-[38] were also used. In the methodology section, numerous key game elements will be examined together with relevant information on each one's usefulness for our developed game in response to the fifth question.

III. RELATED WORK

The use of serious games (SGs) to teach DL skills to persons with ASD has been proposed in a number of studies. These studies were categorized into four categories: conceptual (e.g., money, science, etc.), practical (e.g., use of money, travel and transportation, etc.), social (e.g., emotions, communication, etc.), and general skills. According to [30], the percentage degree of studies targeted at all of those learning categories is given as follows: conceptual skills (25.53%), practical skills (8.51%), social skills (36.17%) and general skills (29.79%). Furthermore, 1.06 percent of research papers were devoted to discussing money management techniques for those with ASD, demonstrating the pressing need for more studies in this field. Different DL skills such as money management skills [4]-[7], emotion recognition [39]-[45], among others have been successfully improved using SGs in special-needs education. However, those methods would not fit with the Egyptian lifestyle, currency and language. As an illustration, Caria et al.

[5] designed a single-player Web-based game that assists people with ASD in acquiring skills to help them understand the notion of Euro currency and its usage in real life in comparison to Bangladeshi Taka currency [7]. The authors in [45] conducted an intervention program using animated vehicles with real emotional faces to improve emotion recognition for Chinese children with ASD compared with Western children in [21]. Additionally, the vast majority of earlier studies concentrated on a single player in a two-dimensional gaming environment, whereas there is a notable lack of research on the development of collaboration skills through multiplayer online games in three dimensions, which support emotions and social interaction among participants and are more engaging and motivating than a single player. One more significant issue is that designated learning environments need to take sensory issues into consideration in order to meet learners' demands. This paper considers all these issues by applying the Autism ASPECTSS™ Design Index [22] as a design development tool for the EMOCASH game's environment. ASPECTSS™ is an abbreviation for: Acoustics, Spatial Sequencing, Escape Space, Compartmentalization, Transitions, Sensory Zoning, and Safety.

IV. BACKGROUND AND FUNDAMENTAL CONCEPTS

A. Serious Games

Serious games (SGs) are a form of education and entertainment that is designed to educate as well as amuse. Unlike games that are just played for fun, games intended for learning are labeled "serious" [46]-[47]. The key concept of serious gaming is the implementation of gaming elements, attributes, mechanisms and flow states to engage learners toward achieving real-life goals.

B. Activity Theory

Activity Theory (AT) is a theoretical framework for examining how people behave in a certain setting. According to AT, the essence of an activity is the interaction between a subject (a human doer) and an object (the item being done). Activity systems [48]-[49], a model of which is shown as a triangle (see Fig. 1). Simply, activity theory is all about 'who is doing what, why, and how' [49].

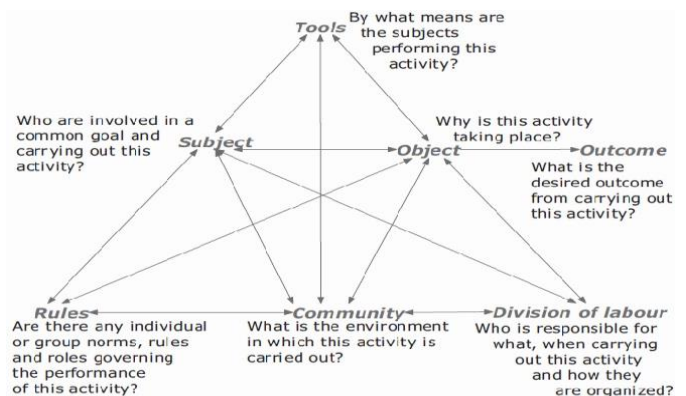


Fig. 1. Structure of activity theory.

C. Virtual Learning Environment and Virtual Agents

The term "Virtual Learning Environment" (VLE) refers to a design environment for teaching and learning. There are several types of VLE, such as single-user, multi-user, and collaborative VLE (CVLE) [50]. Each one of them has a unique avatar that the user may use to engage freely in the digital world. Avatars might take the form of a cartoon character, an entirely abstract shape, or a human-like image. The avatar in our system is a 3D representation of a rabbit with real faces grafted onto it. A software system that is embedded in a particular environment and has the capacity to behave autonomously there in order to achieve its design goals is known as a virtual agent, also known as an intelligent virtual agent (IVA) [51]. IVAs should be autonomous, proactive, reactive, and socially capable. There are several AI agents, so an environment for the IVA design work has to be created, which is called the PEAS (Performance Measure, Environment, Actuator, and Sensor) system [52]. The PEAS System is used to classify similar agents together. An agent's success is assessed using performance metrics; an actuator is a part of the agent that releases the action's output into the environment; and sensors are the receptive parts of an agent that takes in data. The qualities of the environment influence an agent's behavior [53]. Different types of intelligent agent programs based on their degree of perceived intelligence and capability are used in AI, namely: reflex agents, model-based agents, goal-based agents, utility-based agents, and learning agents [52]. Each kind of agent program combines particular components in particular ways to generate actions. The EMOCASH game uses reflex agents where the agent function is based on condition-action rules as demonstrated in Fig. 2.

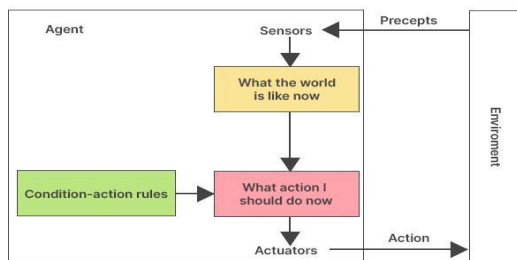


Fig. 2. Schematic diagram of reflex agents program.

V. METHODOLOGY

A. Methodology Workflow

According to contemporary conceptions of effective learning, learning is most successful when it is proactive, reactive, social, experiential, problem-based, and offers quick feedback [48]-[49]. Learner achievements and engagement, as well as active learning, are the two major elements of effective learning. One of the most significant elements influencing learner achievements, success and outcomes is the learning environment [53]. According to the sensory definition of autism [54] the problem of the sensory environment and its relationship to autistic behaviour appears to be the key to designing for autism. This description serves as the foundation for the Autism ASPECTSS™ Design Index [22], [23] which we employed in our methodology as a design development tool for the EMOCASH game's environment. For active learning,

learning environment should be authentic with authentic tasks. Learning in the real-world is authentic learning. To support learners' engagement towards the 21st century skills, the concepts of authentic learning and authentic design thinking [35]-[38], should be employed in educational contexts. In order to engage learners in 21st-century skills, service providers must consider the following five dimensions of 21st-century skills, which were in-depth examined in [37]: cognitive, metacognitive, social-cultural, productivity, and technological. Design thinking (DT) shows a variety of characteristics when used in educational contexts, according to [37] in chapter 9. These characteristics include: 1) DT comes out of social processes where ideas are being formed, clarified, and refined. Collaboration is an important foundation for DT in educational settings; 2) DT involves knowledge creation and is more closely characterized by an iterative and non-linear process that contains five phases (Empathize, Define, Ideate, Prototype, and test), as illustrated in Fig. 3; and 3) DT doesn't have well-defined and well-ordered design stages. Instead, ideas go forth and backward through social interactions until they are wholly embraced by the team. Training individuals with ASD necessitates repeating tasks, which might reduce participants' motivation and interest. To maximize the training outcomes while maintaining the participants' motivation, our methodology takes into consideration the idea of authentic design thinking, game elements' diversity, game attributes, and task activities and introduces an activity system as a framework to embed intelligent agents in SGs' development, as shown in Fig. 4.

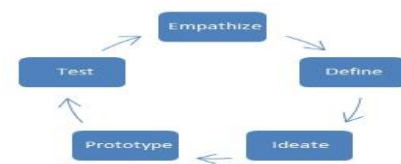


Fig. 3. Design thinking (DT) cycle.

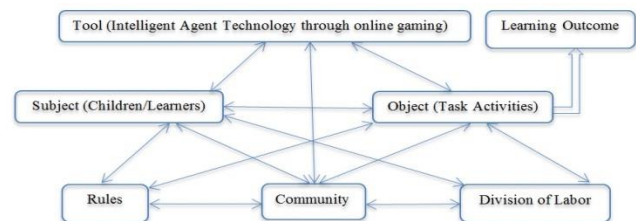


Fig. 4. Learning as an activity system.

As demonstrated in Fig. 4, this system effectively combines learning and technology while also capturing real-world learning environments, which is essential for 21st-century skills [37]. Task activities are the activities and interactions designed to keep the learner engaged and learning in the game's environment. In our approach, the task activities are used as the building blocks of the SG design, and the learning objectives are defined according to the kind of tasks, their number, and difficulty. We now respond to the research question listed in section II: How can we improve the appeal and motivation of education for those with ASD? Authenticity and sensory design aspects [22], [54] should be addressed throughout the design phase of the specified learning

environment. This produces the answer to this question. When the word "authentic" is added to the term design thinking, it transforms into a creative process that aids service providers in coming up with meaningful approaches to their issues by giving them the chance to experiment, develop and prototype models, gather feedback, and redesign instructional materials to improve their learners' learning and performance. Therefore, the finest resource is your expert judgment and relevant stakeholders. As shown in Fig. 5, the methodology workflow is iteratively developed and built from the study of the core design components of an authentic learning environment [35]-[36] and authentic design thinking [37]-[38], as well as the following rules and recommendations that are directly gathered from relevant stakeholders in the area of special needs education using a multi-stage process known as participatory design or a similar user-centered design (UCD) [55]. UCD is defined as a philosophy that places the learners at the center of the design process, taking into account their characteristics, needs, skills, desires, behavior, etc.

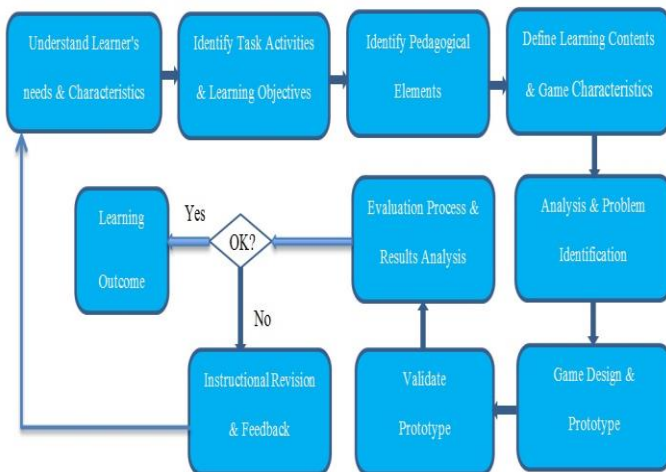


Fig. 5. Methodology workflow.

- Learner-centered: The interests and desires of your learners need to always be your starting point. What, for example, are their interests and needs? What engages them?
- Based on your learner needs, determine a possible goal or outcome. Link the outcome to anything in real life to highlight relevance and authenticity (e.g., shopping in the supermarket, in our case).
- Divide the skills your learners will learn in order to accomplish the outcome.
- Assess your learners both before and after teaching: How do they now perform in these skills? What do they now know? What new skills do they now have? This issue was addressed in our case through usability testing in terms of learnability rate.
- Community-driven design: Individuals struggle to address the entire system; a multidisciplinary group of people with a range of expertise is required to find a solution. We employed the idea of participatory design

[55] to satisfy the needs and preferences of the learners in order to accomplish this goal.

- Give the AMT model some attention: Provide learners with the opportunity to concentrate on the three learning goals of acquisition, meaning-making, and transfer (AMT), as shown in Fig. 6.



Fig. 6. The AMT model.

B. Pedagogical Agents

To facilitate and improve the learning process, virtual characters, or artificial agents, are commonly used in computer games via a VLE. These individuals, also known as "pedagogical agents," may serve in a range of roles, including that of teachers, subject-matter experts, mentors, motivators, or companions [56]. Pedagogical agents are computer agents designed and built to support education by enhancing learners' capacity for spontaneous recall and information retention. It has often an animated persona that responds to the action of the learner [57]. Numerous pedagogical agents have been created and assessed by scholars from different perspectives [58]-[60]. We used anthropomorphic agents—3D models of rabbits with real faces superimposed on them—to construct the pedagogical agents for the EMOCASH game. Additionally, we made an effort to make the rabbit's form larger in order to grab the child's attention. Similar findings have been made by [17]-[19], who claim that autistic individuals' ToM skills are less developed when they read the minds of humans than when they read the minds of non-human agents with anthropomorphic traits like animals. The EMOCASH game may be played by many users using a CALE, and users can interact with one another via their pedagogical agents (see Fig. 7).

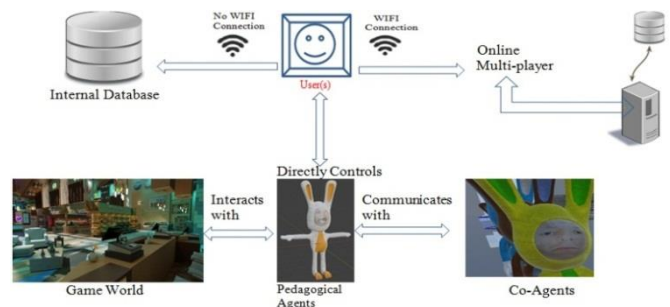


Fig. 7. An illustration of a CALE with pedagogical agents.

C. EMOCASH Gameplay

Gameplay defines how the learner and the game interact with each other; it simply means playing the game. EMOCASH gameplay is briefly discussed below:

- The psychologist describes the therapy process. To start the game, each participant has to log in.
- Game Acts: It is defined as the highest-level element in the EMOCASH game which is divided into many

parts, including the registration page, login screen, play mode, evaluation mechanism to allow tracking of players' progress, etc. Six distinct types of 3D mini-game engines, each with a different level of difficulty, are included in the play mode. Participants must register the first time they play the game. Each participant's results and learning progress are tracked in the assessment file folder.

- **Scenes:** The gameplay takes place in acts, where each act is divided into scenes. Scenes take place in one or more scenarios with different levels of difficulty.
- **Therapeutic game challenge:** The essence of the gameplay is the challenges-actions relationship (what challenges the game has and what actions the player can take to achieve the goals).
- **Currency recognition mini-game:** This mini-game attempts to help autistic children recognize the several primary Egyptian cash denominations and their corresponding values as a tutorial via two engines. The first engine that will be shown is an animated book (GameBook) that has all the major Egyptian cash denominations. The participants' task is to open each page, see and freely move the provided cash notes in all directions in accordance with the six DOFs as depicted in Fig. 8(a). The second engine's job is to put together an image of the currency note (e.g., a one, five, ten, twenty, etc.) from a variable number of pieces that are spread out across a table surface in random placements. The puzzle's final appearance after being properly solved is depicted by a target image in the upper portion of the screen. The task requires the participants to piece together the required image using the materials they have collected, as illustrated in Fig. 8(b). By the time the session is over, the participants will have two levels of easy quizzes that gradually get harder. For instance, we ask the participant to choose one of four currency notes that are displayed on the screen.

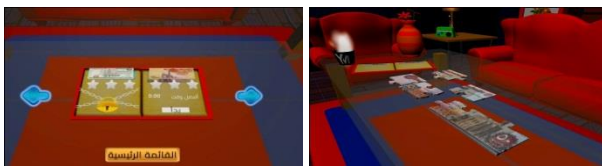


Fig. 8. (a) An animated game book containing all Egyptian currency denominations (b) Puzzle game.

- **Emotion Recognition mini-game:** With the use of numerous settings for purchasing (such as being a customer) and selling (such as being a cashier) inside a 3D virtual shop environment, this game enables autistic children to reinforce some facial expressions used when shopping. For instance, if you need to buy something but don't have enough money, you'll feel "sad"; alternatively, you'll feel "happy." When asking for assistance when shopping, the same feelings will manifest as depicted in Fig. 9. To engage participants and confirm that they were able to understand the required activities, the game start-up user interface

featured a tutorial for an animated human cartoon face with distinct emotions incorporating visual and audio stimuli (see Fig. 10).



(a) A snapshot of some players asking for some assistance



(b) The message indicates that the player received a favorable response (a happy feeling)



(c) A snapshot of some players asking for some assistance



(d) The message indicates that the player received an unfavorable response (a sad feeling)

Fig. 9. Snapshots of recognizing some human facial emotions in various contexts while shopping.

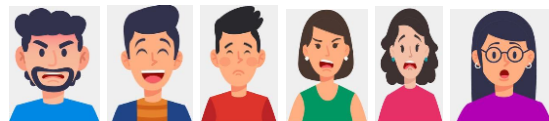


Fig. 10. A Screenshot for an animated designed human cartoon faces with different emotions.

- **3D virtual shop scenario:** This mini-game aims to teach the process of shopping to autistic children by achieving the transfer of acquired knowledge to a real-world scenario, such as shopping in the supermarket. The scenario employs a (CALE) with a simplified navigation and interaction interface that encourages emotions and interactions among participants, who can play alone or with their instructors. For agent movement in a virtual world, the A* (A-Star) path-finding search algorithm is used [61] (see Fig. 11). It addresses the problem of finding a shortest path from any coordinate in the game world to another.



Fig. 11. The A* (A-star) algorithm in action.

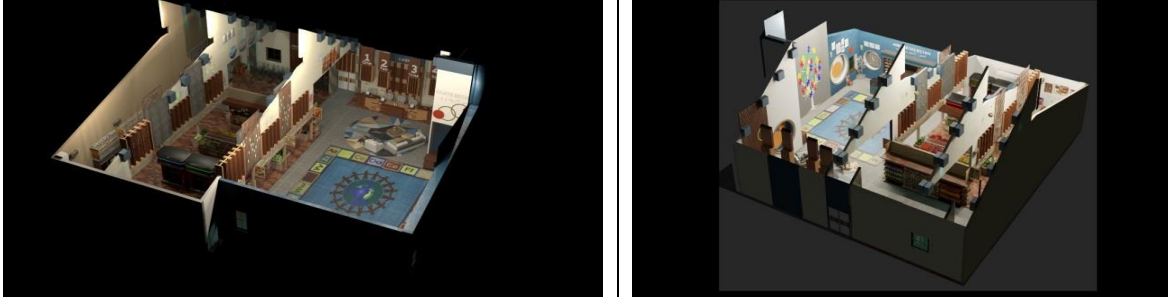
- By the end, feedback sessions are planned and the results are analyzed.




D. Game Elements and Attributes

The parts that make up a game and give learners an engaging experience are known as game elements. Every game element must have attributes that may be utilized to affect how

they are used when developing SGs. Scholars [5], [62]-[64] have reported that incorporating game elements into teaching methods for people with ASD can improve their learning abilities. For instance, the authors of [62] consider game elements as intrinsic and extrinsic factors to increase a learner's motivation. The majority of previous research did not explicitly identify the game elements or specifically state or offer information on the games' elements they employed while creating their solutions. Even when they did, they didn't provide adequate information on the efficiency of the particular game elements. We now respond to the research question listed in section II: What aspects of game elements are taken into consideration when using online gaming to educate autistic people? The EMOCASH game offers a variety of useful game elements with relevant information on each one's effectiveness, as indicated in Table I.

TABLE I. EMOCASH'S GAME DESIGN ELEMENTS FOR LEARNING

Game Elements	Description
Game World/Environment	<p>It is a virtual environment where players go while playing the game. As illustrated below, the gaming environment in the EMOCASH game is created as a 3D virtual shop scenario based on the principle of authentic learning with authentic tasks, like grocery shopping, for teaching the shopping process to autistic children. It was designed in accordance with the Autism ASPECTSS™ Design Index [22], [23], [54] to take into account the specific needs of autistic people and facilitate the learning process for those populations.</p> 
Clear Goals	When people are aware of their responsibilities, the goals are crystal clear. Giving meaningful challenges and establishing clear norms are closely tied to setting clear goals. Clear objectives improve attention. To hold players' attention and make sure they will be able to understand the necessary tasks, the primary objectives and rules of each mini-game are given at the beginning of the introduction to the entire EMOCASH game.
Challenges	Challenges are game tasks or exercises that require effort to perform. In the EMOCASH game, there are a lot of tasks for each mini-game to be completed. Once achieved, some rewards are provided.
Rewards	A reward is a component of the game that gives the players satisfaction and inspires them to work harder. Extrinsic rewards, such as points, badges, and leader boards, are available through EMOCASH, as well as intrinsic rewards, which are given for successfully completing each mini-game throughout the entire game in exchange for resources (such as earning virtual currency with a numeric value) and reputation (such as a certain number of stars).
Best Moments	It is a fun feature that motivates the participants during the game. The virtual gained items from the prize gallery are stored in an inventory which can be reached by clicking on the "prizes customization" button on the main screen to enable the participants to experience a personalized training by meeting their curiosity and virtually using the items (e.g., wearing a glass) through a variety of customization options.
Status	The status is a categorization of participants based on their scores. In EMOCASH, the motivation is based on the aim of participants to reach a high status by earning more virtual currency and gaining more stars. The rating number of stars shows great achievements. The higher the star's number rating is, the better its status.
Leader boards	A visual representation of the participant's progression with respect to others. The leader boards exploit social emotions such as feelings of "fame" and "status" while allowing comparison between participants and enhancing competition among them.
Competition	Competition modes are: single-player competition against the game environment; competition between two or more players; cooperation of two or more players against the game environment; and team-based competition. The EMOCASH game supports the first two modes.
Points	Game points are like grades in the educational system. It is a numerical representation of the player's performance in achieving goals. Points are assigned to each activity as a function of the number of goals reached and the time needed to be finished. Each mini-game mode assigns different points to participants. Points can be divided into levels, and levels can be presented with badges.
Badges	Badges are virtual goods that have a visual representation. They are awarded to participants after completing certain challenges or reaching certain achievements. In the EMOCASH game, we used it to represent participants' status or reputation.
Levels	Levels can have different meanings in games. Levels can refer to the rating of the participant based on his/her score or can be related to the difficulty of the game. The EMOCASH game supports both. In terms of difficulty, each mini-game has two levels of difficulty, namely basic and advanced, where the level of complexity increases very gradually in order to keep the participants motivated and to provide them with a continuous challenge (see figures below).

	  
Flow state	To feel the fun, users have to be in the channel of flow state. The flow theory is widely accepted to be one of the fundamental models for improving the game experience. When an individual experiences flow, they are said to be in a "flow state."
Feedback	For generating the flow it is important that activities provide immediate and clear feedback which can be provided with the help of visual and audio elements. As a demonstration, the figure below shows the feedback provided for money comparison and counting mini-game scenarios.
Progress bar	A progress bar was utilized to graphically represent the proportion of completed tasks in order to display the participants' progress for each activity. For each activity, a green color will emerge if the exercise is successfully solved, signifying that the participant has given the correct response; otherwise, a red one will be displayed.
Progression	Player growth and development.
Visual Aesthetics	Include visual elements such as the overall look and feel of the game.
Game Mechanics	The procedures and rules of the game.
Story	The series of events that occur as players play the game.
Technology	It is the medium through which the tale will be told, the mechanics will occur, and the visual aesthetics will take place. We used intelligent agent technology through online gaming for the EMOCASH game. Moreover, the learning environment was designed using the Autism ASPECTSS™ Design Index to adapt to the needs of all types of users
Game Fantasy	Refers to the environmental contexts that provide virtual world imagery
Sensation	Multimedia presentation of the virtual world.
Emotions	Games are good for creating emotions among participants. Those emotions can be created through gameplay, storytelling, or socialization. The EMOCASH game is a multi-player SG which supports both emotions and socialization among participants.
Avatars	A visual representation of a player character. The EMOCASH game uses anthropomorphic agents in the form of 3D models of rabbits grafted with real faces.

VI. SYSTEM DEVELOPMENT AND IMPLEMENTATION

A. System Architectural Design

Architectural design defines the interconnection and resource interfaces between system subsystems components and modules in ways suitable for their detailed design and overall configuration management. As training requires repetitive tasks, having a large number of mini-games is quite important to avoid boredom in the long term. Fig. 12 shows a sample architectural design overview of our EMOCASH game architecture, which is composed of six 3D mini-games, each one with two difficulty levels (basic and advanced). The complete navigational system for the EMOCASH game is shown in Fig. 13.

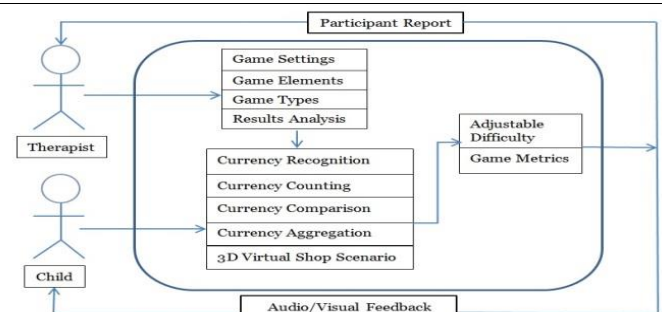


Fig. 12. EMOCASH's system architecture.

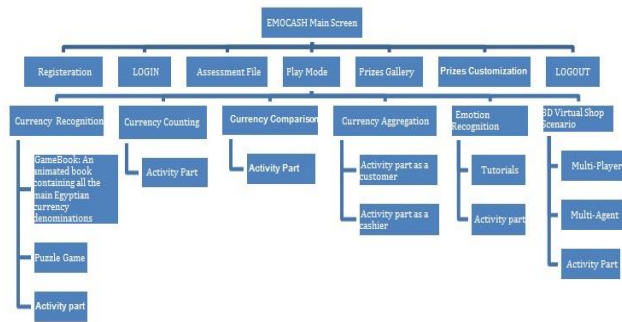


Fig. 13. EMOCASH's system navigation scheme.

B. Learning Mechanics-Game Mechanics (LM-GM) Model for EMOCASH Gameplay

This subsection analyzes the relevance of learning mechanics and game mechanics for the EMOCASH gameplay

through the Learning Mechanics-Game Mechanics mapping model (LM-GM) [65]. Several authors have offered definitions of the idea of game mechanics. However, there are no universally agreed-upon definitions of game mechanics, as evidenced by the wealth of literature [66]-[68]. For instance, game mechanics are described as "something that connects the player's actions with the purpose of the game and its main challenges" in a definition offered by [68]. On the other hand, activities or design patterns of mechanics that explicitly support learning but are not playable mechanics are referred to as learning mechanics. Table II displays the analysis of both the learning and game mechanics for the EMOCASH gameplay through the LM-GM model [65], [66], as well as how participants use them and how they were actually implemented.

TABLE II. THE LM-GM MODEL ANALYSIS FOR THE EMOCASH GAMEPLAY

Game Mechanics	Learning Mechanics	Implementation	Usage/Description
Cascading information/ Tutorials	Guidance/ Tutorial	The main goals and regulations of each mini-game are explained at the start of the introduction to the entire EMOCASH game with the help of a brief animation and voice.	Main goals are clearly presented in the game's introduction to keep participants' attention and ensure that they will be able to comprehend the necessary tasks.
Cooperation/ Collaboration	Emotion/ Competition/ Participation	The gaming environment in the EMOCASH game is designed as a 3D virtual shop scenario using the Autism ASPECTSS™ Design Index and is based on the idea of authentic learning with authentic tasks, such as shopping at the supermarket.	Games are good for creating emotions among participants. Those emotions can be created through gameplay, or socialization. Using a CVLE, the EMOCASH game is a multi-player SG which supports both emotions and socialization among participants.
Time pressure	Motivation Action/Task	Count down clocks/Response time	Extrinsic motivation.
Meta-Game Mechanic (Rewards that can be earned during the actual gameplay)/ Collecting (Elements of virtual rewards can be represented by virtual objects, which can be collected by the player)	Motivation/ Rewards/ Incentive	All six types of 3D mini-game engines include this feature.	The player earns virtual currency with a numeric value and some stars for each correct answer and for each mini-game successfully completed in the entire game.
Questions & Answers	Assessment	Once the player finishes practicing the currency recognition mini-game, he or she will have to play two simple and fun quiz games called puzzle and matching games before proceeding to the next task.	All five types of 3D mini-game engines include an activity part with two levels of difficulty, namely, basic and advanced, where the level of complexity increases very gradually in order to keep the participants motivated and provide them with a continuous challenge.
Rewards	Motivation/ Instructional/ Feedback	Extrinsic rewards, such as points, badges, and leader boards, are available through EMOCASH, as well as intrinsic rewards, which are given for successfully completing each mini-game throughout the entire game in exchange for resources (such as earning virtual currency with a numeric value) and reputation (such as a certain number of stars).	Extrinsic and intrinsic rewards are used to keep participants engaged in SGs and keep them happy and satisfied.
Behavioral momentum	Incentive/ Motivation/ Participation	The virtual gained items from the prize gallery are stored in an inventory which can be reached by clicking on the "prizes customization" button on the main screen to enable the participants to experience a personalized training by meeting their curiosity and virtually using the items through a variety of customization options.	It is a fun feature that motivates the participants during the game.
Feedback	Motivation/ Feedback	Points/levels/response time/badges/ emotional sounds/ audio-visual feedback/Progress bar.	Extrinsic rewards that facilitate both motivation and flow state.
Capture (How many points or counters a player has won or captured)	Action/Task Participation	Stars earned and the score of each mini-game completed.	Use information to solve problems. The accomplishment of goals during gaming is represented by stars, which are gathered with each new goal attained. The rating number of

defines how strong they are)			stars shows great achievements.
Challenge	Questions & Answers/ Action/Task /Participation/ Feedback/ Assessment	<ul style="list-style-type: none"> ▪ Challenges in games must match the player's skill level ▪ Games should provide different levels of challenge for different players ▪ The level of challenge should increase as the player progresses through the game 	Challenges are game tasks or exercises that require effort to perform. In the EMOCASH game, there are a lot of tasks for each mini-game to be completed. Once achieved, some rewards are provided. For instance, in the case of a puzzle game, the main challenge is the participant's ability to piece together the required image using the materials they have collected from a variable number of pieces that are spread out across a table surface in random placements.

C. Implementation

Our system is organized in multiple levels and implemented using Unity. Blender software was used to create the learning environment as well as the 3D models of various objects (e.g., rabbits) and import them into Unity. Since different currencies have a range of different coin values, or denominations, we employed the greedy change algorithm [69] in our implementation of the change problem.

VII. METHOD

A. Participants

In this study, we recruited a sample of fourteen distinct autistic children between the ages of seven and fifteen to play the EMOCASH game for at least one and a half hours each week, for a total of seven weeks of open trial. These groups, referred to as "Group #1," received an autism diagnosis from a psychiatrist and a clinical psychologist. Ten children (72%) were boys, while four (28%) were girls. We also invited a group of seven typically developing volunteer children, referred to as "Group#2," who tested the game in order to measure the effectiveness of using the developed application. Ethical approval was granted by the Egyptian General Ethics Standards.

B. Procedure and Materials

The intervention protocol used for participants consists of three different phases: pre-intervention, intervention, and post-intervention phases. During pre-intervention sessions, children completed IQ testing representing verbal and performance IQ taken from the 2nd Wechsler Abbreviated Scale of Intelligence (WASI- II), and parents filled out the Social Responsiveness Scale, 2nd edition (SRS-2) and the Vineland Adaptive Behavior Scales, 2nd edition (VABS-II) to confirm the children's diagnostic status. At the end of the pre-intervention session, participants and their parents were introduced to the EMOCASH game. A seven-week, three-times-per-week EMOCASH training program was part of the participant's intervention procedure. After a 7-week intervention period, participants and their parents participated in the post-intervention assessment. In this phase, a questionnaire is created to assess the expectations and satisfaction of both parents and children playing EMOCASH game.

VIII. EVALUATION, USABILITY TESTING AND RESULTS

A. Usability Testing with Experts

Usability testing is a technique used to assess the effectiveness and satisfaction of a developed application by its users and experts. Four subject-matter experts in different fields were asked to participate; two of them were experts in psychology and psychiatry, while the other two were experts in

game design and teaching, respectively. Furthermore, to identify the usability issues of the developed application in terms of efficiency and satisfaction, it was evaluated not only from the experts' perspective but also with input from people who are not disabled (Group#2). The purpose of these tests serves the function of gathering system feedback. A list of the six mini-games engine's interface and the tasks to be carried out in the EMOCASH game were given to the experts for usability analysis to be sure that it is suitable for use by children with ASD without any issues. Five themes were used to group the perspectives of experts on the EMOCASH gameplay: Effectiveness, usefulness, enjoyment, ease of use, and attitudes toward future usage. Most experts agreed that the proposed game was effective because of the following factors: (1) goals were clearly presented in the game's introduction to the entire EMOCASH game with the help of a brief animation and voice, which improve attention and help with their learning; (2) the presentation of multimedia resources (audio, visual, etc.) is done simultaneously; (3) immediate and clear feedback was given with the help of multimedia resources; (4) clear goals and feedback support concentration, which is one of the learning outcomes; (5) learner activities that use authentic learning closely resemble the real-world situations that professionals encounter while practicing; and (6) promoting practical life skills like money management can enhance a person's autonomy and self-determination. According to experts in the psychology and psychiatric disciplines, the game is useful since it is affordable, promotes self-confidence, may aid with motor skill development, and saves time for the teachers. Experts in game design and teaching fields found that the game is enjoyable because of the following reasons: (1) it attempts to increase both participant concentration and the sense of curiosity in individuals with ASD; (2) it contains audio-visual feedback and both extrinsic (e.g., points, badges, leader boards, etc.) and intrinsic rewards that facilitate motivation and flow state and, as a result, promote players' engagement, learning, and motivational outcomes; and (3) playing online games, which improve the user's experience, immersion, foster social engagement and competition, and reduce autistic symptoms, increases the enjoyment of working independently. In terms of usability, the majority of experts agreed that the game is useful, easy to use, and user-friendly, and that the challenge-skill balance matches the player's skill level. In addition, they reported that the game's mechanics and interface are straightforward, simple to understand, and non-intrusive, allowing for simple access to the game. In terms of attitude, all experts agreed that the game could be objectively used as a teaching tool for treating children with ASD in a cost-effective manner. They also found that the game is enjoyable, fascinating, and beneficial.

B. Evaluation Instrument for User Satisfaction

To evaluate the functionality of the system we have distributed a questionnaire evaluation form to assess expectations and satisfaction among parents and their children. Parents rated their expectations and satisfaction by filling out several questionnaires using a 10-point Likert scale to during pre-intervention and post-intervention measurement. Sample questionnaires are given in Table III.

TABLE III. A SAMPLE ASSESSMENT QUESTIONNAIRE FOR PARENT EXPECTATIONS AND SATISFACTION

Q#1	Parent expectations sample questionnaire	Parent satisfaction sample questionnaire
1	Do you think that this EMOCASH game is educational for your child?	How motivated was your child to play the EMOCASH game?
2	How much do you expect your child's skills will advance after playing the EMOCASH game?	Do you think the EMOCASH game had an effect on your child's performance on the different daily living (DL) tasks/skills?

After finishing the EMOCASH game, participants were also asked to provide their opinions, including what they enjoyed and disliked about it, as well as their favorite game. For this purpose, we utilized the Fun Toolkit [70], to assess participants' satisfaction. Moreover, all experts and ordinary volunteer children answered the System Usability Scale (SUS) questionnaire [71] after playing the game. The SUS questionnaire contains 10 questions about ease of learning, efficiency, ease of memorization, occurrence of execution errors, and level of satisfaction. Each question has a five-point scale varies from one (totally disagree) to five (totally agree). By combining the contributions of the scores of each item, the SUS questionnaire's results were analyzed. The results were positive between experts and ordinary volunteer children regarding their acceptance.

C. Game Data and Usability Metrics

The EMOCASH game's usability was assessed using the game data of each participant. The data for the game is stored in internal structures and contains the participant's score for each game successfully completed. The system creates reports in the assessment file folder when the participant has finished the games, providing the therapist with feedback. In order to test the application with users and to measure the efficiency and effectiveness of the application's use in education, we considered the following usability metrics [72] (see Fig. 14) where,

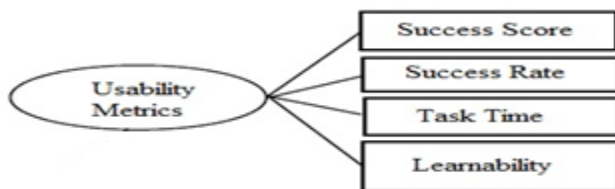


Fig. 14. Usability metrics.

1) *Success Score*: It refers to whether the user accomplished or didn't finish the required task (Number of completed tasks/Total number of attempts).

2) *Success Rate*. This includes several levels of success: a) Complete success: the user completed the task with no errors and exactly as specified; b) Minor issue success: the user completed the task but encountered a minor issue; c) Major issue success: the user completed the task but encountered a major issue; and d) Failure: the user was unable to complete or finish the required task.

3) *Task Time*: The amount of time it takes the user to complete the task.

4) *Learnability Rate*: it takes into account both how simple a task is for users to perform the first time they use the interface and how many tries it takes for them to get it right. Fig. 15 depicts the learnability rate for the participants' responses using conventional and non-conventional ways after practice with our EMOCASH game. It is clear that after practicing with our game, their performance improved significantly.

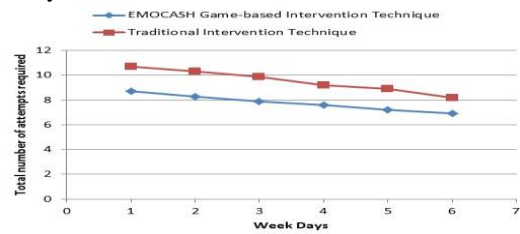


Fig. 15. Learnability rate.

D. Results and Discussion

This section describes the findings of the responses collected from experts, parents, and their children. This was done to help children recognize the EMOCASH game as a learning tool and determine which features of the game are more attractive. The results of the collected data show that the game is very useful and the results are given as follows:

- Most of them agreed that the system is helpful, straightforward to use with a user-friendly learning environment, and has simple content with good topic flow.
- Regarding game motivation, most of the participants found the game to be fun to play, especially when applying newly learned information through the 3D virtual store setting, which fosters feelings and sociability among players who can play alone or with their educators.
- Concerning game usability, most of the participants were able to play the game because their preferences were achieved via factors like simple GUI, audio-visual feedback, and both extrinsic and intrinsic rewards. Most experts rated the usability of the EMOCASH game as "excellent"
- The majority of the experts, typical volunteer children, as well as parents and their children with ASD, were satisfied with the EMOCASH game. Experts indicate that the game appears to have outstanding potential and can be objectively used as a teaching aid for assisting individuals with ASD. Parents also reported that seven

weeks of EMOCASH's use significantly improved their children's performance on the different money and emotion recognition tasks.

- The gathered data also explores the beneficial effects of involvement in online gaming as an effective teaching strategy for those with ASD. People with ASD can communicate more easily and participate in the gaming community without having to disclose their handicap by playing online games.

E. Comparisons

According to the taxonomy discussed in [46], [47], Table IV compares our proposed framework (EMOCASH) with other existing frameworks previously discussed during this research. As demonstrated in Table IV, all of the prior research only took into account a single player in a two dimensional gaming world, as opposed to multi-player collaborative games in three dimensions, which support both emotions and social interaction among participants and facilitate collaborative learning, which are more motivating and engaging than a single one. In addition, this research is one of the few that uses authentic learning activities like going

shopping while adhering to the autism ASPECTSS™ Design Index, a set of architectural design guidelines for individuals with autism. The EMOCASH game possesses the following merits that set it apart from previous works: (1) the scenario uses a CALE, which supports multiplayer, which is more engaging and motivating than single-player training games; (2) the sensory environment issues and their relationship to autistic behavior were taken into consideration by applying the Autism ASPECTSS™ Design Index as a design development tool for the EMOCASH game's environment to adapt to the needs of all types of users; (3) it explores the beneficial effects of involvement in online gaming as an effective teaching strategy for those with ASD; (4) our strategy takes into account the wide range of game elements and attributes to enhance training outcomes while keeping participant motivation; (5) the learning mechanics-game mechanics (LM-GM) model idea was used to assess and examine the efficacy of the proposed game in order to keep participants' interest and motivation; and (6) the game was developed using a number of design principles, such as guidelines from the literature, guidance from learning experts and instructors, and feedback from a wide range of stakeholders, including those who are not impaired.

TABLE IV. COMPARISON BETWEEN OUR PROPOSED FRAMEWORK AND THE OTHER EXISTING FRAMEWORKS

Criteria	Sources/References										
	EMOCASH	[5]	[7]	[6]	[4]	[42]	[41]	[40]	[43]	[44]	[39]
Learning Topics	Conceptual and social skills	Conceptual skills	Conceptual skills	Conceptual skills	Conceptual skills	Social skills	Social skills	Social skills	Social skills	Social skills	Social skills
Learning Objectives	Learning money & emotion recognition	Learning how to manage money	Learning money concept	Learning how to manage money	Learning how to manage money	Learning Emotion Recognition (LER)	LER	LER	LER	LER	LER
Game Interface	3D	2D	2D	2D	2D	2D	2D	3D	3D	2D	2D
Target Audience	Autistic disorder	High-Functioning ASD (High-F-ASD)	Autistic children	Intellectual disability	Cognitive impairments	Autistic children (AC)	Autistic children (AC)	AC	High-F-ASD	Autistic children (AC)	Autistic children (AC)
Interaction Style	Traditional IO (T-IO)	T-IO	T-IO	T-IO	T-IO	T-IO	T-IO	T-IO	T-IO	T-IO	T-IO
Feedback	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Progress monitoring	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of players	Single/Multi-player	Single player	Single player	Single player	Single player	Single player	Single player	Single player	Single player	Single player	Single player
Collaborative learning	Yes	No	No	No	No	No	No	No	No	No	No
Social Presence	Yes	No	No	No	No	No	No	No	No	No	No
Sensory Environment Design Issues	Yes	No	No	No	No	No	No	No	No	No	No
Game portability	Home/Hospital	Home	Home	Home	Home	Home/Hospital	Home	Home/Hospital	Hospital	Home/Hospital	Home/Hospital
Usability Testing	Yes	Yes	No	Yes	Yes	No	No	Yes	Yes	Yes	Yes
Knowledge Transfer to	Yes, via 3-D Virtual	Yes, through an	No	No	Yes, via a Simplified	No	No	No	No	No	No

Real-world Scenarios	Shop Scenario	Interaction with a Vending Machine (using coins only).			Cash Accounting.						
----------------------	---------------	--	--	--	------------------	--	--	--	--	--	--

IX. CONCLUSION AND FUTURE WORK

The number of people with ASD continues to grow worldwide. As a result of the incapacity symptoms, they often experience serious difficulties in performing daily living activities independently. Money management and recognizing emotions are two essential daily living activities that people with ASD generally struggle with. When designing a learning environment for them, care should be taken to take into account any sensory challenges they may have. The problem of the sensory environment and its connection to autistic behavior seems to be the key to designing for autism. There are few studies that look at sensory design issues while developing children's learning environments, and there is a noticeable lack of studies on how using virtual environments and playing online games affects social and relational interactions for persons with ASD. This paper addresses these issues and makes an effort to close this gap using a novel virtual agent-based multiplayer online serious game called "EMOCASH," which aims to enhance these crucial tasks for Egyptian children with ASD and achieve transfer of learned skills to real-world contexts via a 3D virtual shop scenario created using the Autism ASPECTSS™ Design Index. EMOCASH served as an instrumental tool for investigating numerous relevant research questions. A variety of usability metrics, were used to assess effectiveness, efficiency, and satisfaction aspects. This study is one of the few implementing activities related to real-world scenarios, such as shopping in the supermarket for people with autism using intelligent agent technology through online gaming. Our findings indicate that, when compared to non-agent methods, virtual agent technology through online gaming, reinforced by sensory design aspects, may be used as an objectively successful learning technique for individuals with ASD. Our future work will focus on three directions: 1) expanding the number of activities developed and making the application available on various platforms; 2) analyzing the impact on larger populations of autistic children with various categories; and 3) employing Brain-Computer Interface technology as a physiological measuring instrument that retrieves and uses information about an individual's mental state, which is closely related to the ToM and anthropomorphism.

ACKNOWLEDGMENT

The author appreciates the help and insightful conversation he received from Eng. Islam Abd El-Sattar, Mr. Mohamed Abdullah, and Mr. Yousef Salah.

REFERENCES

[1] Nikou, S. A. and Economides, A. (2021). A Framework for Mobile-Assisted Formative Assessment to Promote Students' Self-Determination. *Future Internet*. 13. 116. 10.3390/fi13050116.

[2] Yeni, S., Cagiltay, K. and Karasu, N., (2020). Usability investigation of an educational mobile application for individuals with intellectual disabilities. *Universal Access in the Information Society*, 19:619–632, <https://doi.org/10.1007/s10209-019-00655-0>

[3] Urturi, Z.S., Zorrilla, A.M., and Zapirain, B.G. (2012). A Serious Game for Android Devices to Help Educate Individuals with Autism on Basic First Aid. In: Omatu, S., Depaz Santana, J., Gonzalez, S., Molina, J., Bernardos, A., Rodriguez, J. (eds) *Distributed Computing and AI. Advances in Intelligent and Soft Computing*, Vol. 151, Springer, Berlin, Heidelberg. DOI: 10.1007/978-3-642-28765-7_74

[4] Abdel Wahed, Shiruk et. al. (2020). QUALY: A Money Management Application for People with Cognitive Impairments. 14th International Conference on Game Based Learning ECGBL 2020, 10.34190/GBL.20.109

[5] Caria, S., Paternò, F., Santoro, C. and Semucci, V. (2018). The Design of Web games for helping young High-Functioning Autistics in learning how to manage money. *Mob. Netw.* Vol. 23, pp. 1735–1748

[6] Lopez-Basterretxea, A., Mendez-Zorrilla, A. and Garcia-Zapirain, B. (2014). Telemonitoring Tool based on Serious Games Addressing Money Management Skills for People with Intellectual Disability. *Int. J. Environ. Res. Public Health*, Vol. 11, pp. 2361-2380, doi:10.3390/ijerph11030236

[7] Arshia Z. H., Bushra, T. Z., Fatema T. Z., Johra M., Tasmih Md. Mustafizur R., Hasan S. F., Syed , I. A. (2011). Developing the concept of money by interactive computer games for autistic children. In *Conf. Rec. 2011 IEEE Int. Symposium on Multimedia*, pp. 559–564

[8] Nur Syaheera, B. S., Hamzah Asyrani, B. S., Nor Saradatul Akmar, B. Z., & Tuty Asmawanty, B. A. (2023). MMZ: A study on the implementation of mathematical game-based learning tool. *International Journal of Advanced Computer Science and Applications*, Vol. 14 (1).

[9] Ayman A., Hind B., Mayda, A., Hind, A., & Eman, S., (2021). DoItRight: An Arabic gamified mobile application to raise awareness about the effect of littering among children. *International Journal of Advanced Computer Science and Applications*, Vol. 12, No. 12.

[10] Moosa, A. M. et. al. (2020). Designing a mobile serious game for raising awareness of diabetic children. *IEEE Access*, DOI: 0.1109/ACCESS.2020.3043840

[11] Karolina, D. Kamil, D., Lucci, R. B., and Anita, B. (2020). Therapeutic programs aimed at developing the theory of mind in patients with autism spectrum disorders – available methods and their effectiveness. *Psychiatria Pol.* Vol. 54, No. 3, pp. 591–602, doi: <https://doi.org/10.12740/PP/108493>.

[12] Rosello, B., Berenguer, C., Baixauli, I., García, R., and Miranda, A. (2020). Theory of Mind Profiles in Children with Autism Spectrum Disorder: Adaptive/Social Skills and Pragmatic Competence. *Frontiers in Psychology*, Vol. 11, doi:10.3389/fpsyg.2020.567401.

[13] Roberto, M., Claudio, M., Rodolfo, V., Angeles, Q., and Victor Hugo, C. D. (2019). Developing a software that supports the improvement of the Theory of Mind in Children with ASD. *IEEE Access*, doi: 10.1109/ACCESS.2018.2890220.

[14] Baron-Cohen, S. (2014). Theory of mind and autism: A review. *International Review of Research in Mental Retardation*, vol. 23

[15] Ruud, H., Michaela, K., Kohinoor, M. D., Laura, J., Kami, K., Richard, R. and Emily, S. C. (2021). Exploring the relationship between anthropomorphism and Theory-of-Mind in brain and behavior. *Human Brain Mapping*, vol. 42, pp. 4224–4241, doi: 10.1002/hbm.25542

[16] Atherton, G. and Cross, L. (2018). Seeing more than human: Autism and anthropomorphic theory of mind. *Frontiers in Psychology*. vol. 9, no. 528, pp. 1–18, doi: 10.3389/fpsyg.2018.00528

[17] Liam, C., Myles, F. and Gray, A.(2019). The Animal in Me: Enhancing Emotion Recognition in Adolescents with Autism Using Animal Filters. *Journal of Autism and Dev. Disorders*, vol. 49, pp. 4482–4487, <https://doi.org/10.1007/s10803-019-04179-7>

[18] Atherton, G. and Cross, L. (2019). Animal faux pas: Two legs good four legs bad for theory of mind, but not in the broad autism spectrum. *The Journal of Genetic Psychology*, 180(2–3), pp. 81–95, <https://doi.org/10.1080/00221325.2019.1593100> .

- [19] Whyte, E. M., Behrmann, M., Minshew, N. J., Garcia, N. V., and Scherf, K. S. (2016). Animal, but not human, faces engage the distributed face network in adolescents with autism. *Developmental Science*, vol. 19, no. 2, pp. 306–317
- [20] O’Haire ME., McKenzie, SJ. Beck, AM. and Slaughter, V. (2013). Social Behaviors Increase in Children with Autism in the Presence of Animals Compared to Toys. *PLoS ONE* 8(2): e57010, doi:10.1371/journal.pone.0057010
- [21] Golan, O., Ashwin, E., Granader, Y., McClintock, S., Kate, D., Victoria, L. and Baron-Cohen, S. (2010). Enhancing emotion recognition in children with Autism Spectrum Conditions: an intervention using animated vehicles with real emotional faces. *Journal of Autism and Developmental Disorders*, vol. 40, pp. 269-279
- [22] Mostafa, M., (2015). Architecture for Autism: Built Environment Performance in Accordance to the Autism ASPECTSS™ Design Index. *Design Principles and Practices: an International Journal Annual Review*, 8 (1): 55- 71. doi:10.18848/1833-1874/CGP/v08/38300
- [23] Hussein, K., (2023). Towards smart authentic learning environment using IT through online gaming for promoting money management skills in Egyptian children with autism. Submitted for publication.
- [24] Almurashi, H., Bouaziz, H., Alharthi, R., Al-Sarem, W., Hadwan, M., and Kammoun, S. (2022). Augmented Reality, Serious Games and Picture Exchange Communication System for People with ASD: Systematic Literature Review and Future Directions. *Sensors*, vol. 22, https://doi.org/10.3390/s22031250
- [25] Petersen, G., Petkakis, G. and Makransky, G. (2022). A study of how immersion and interactivity drive VR learning. *Computers & Education*, Vol. 179, https://doi.org/10.1016/j.compedu.2021.104429.
- [26] Desideri, L., Pérez-Fuster, P. and Herrera, G. (2021). Information and Communication Technologies to Support Early Screening of Autism Spectrum Disorder: A Systematic Review. *Children*, Vol. 8, No. 93, https://doi.org/10.3390/children8020093.
- [27] Zhang, K. and Aslan, A. B. (2021). AI technologies for education: Recent research & future directions. *Computers and Education: Artificial Intelligence*, Vol. 2, https://doi.org/10.1016/j.caeai.2021.100025.
- [28] Abu-Amara, F., Bensefia, A., Mohammad, H. and Tamimi, H. (2021). Robot and virtual reality-based intervention in autism: a comprehensive review. *Int. J. Inf. Technol.*, Vol. 13, No. 5, pp. 1879–1891 https://doi.org/10.1007/s41870-021-00740-9.
- [29] Pérez-Fuster, P., Sevilla, J. and Herrera, G. (2019). Enhancing daily living skills in four adults with autism spectrum disorder through an embodied digital technology-mediated intervention. *Res. Autism Spectrum Disorder*, Vol. 58, pp. 54–67.
- [30] Valencia, K., Rusu, C., Daniela, Q. and Erick, J. (2019). The Impact of Technology on People with Autism Spectrum Disorder: A Systematic Literature Review. *Sensors*, doi: 10.3390/s19204485
- [31] Raith L, Bignill J, Stavropoulos V, Millear P, Allen A, Stallman HM, Mason J, De Regt T, Wood A and Kannis-Dymand L. (2022). Massively Multiplayer Online Games and Well-Being: A Systematic Literature Review. *Front. Psychol.* 12:698799, (2022), doi: 10.3389/fpsyg.2021.698799
- [32] Mikhailova, O. B. (2019). High school students involved and not involved in MMORPG: creativity and innovativeness. *International Journal of Cognitive Research in Science, Engineering and Education (IJCRSEE)*, 7(2), 29-39.
- [33] Gallup, J., Serianni, B., Duff, C., & Gallup, A. (2016). An exploration of friendships and socialization for adolescents with autism engaged in massive multiplayer online role-playing games (MMORPG). *Education and Training in Autism and Developmental Disabilities*, 51(3), 223-237
- [34] Zhang, Y., Song, H., Liu, X., Tang, D., & Chen, Y. (2017). Language learning enhanced by massive multiple online role-playing games (MMORPGs) and the underlying behavioral and neural mechanisms. *Frontiers in Human Neuroscience*. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5332359
- [35] Lombardi, M. (2007). Authentic learning for the 21st century: An overview. Ed. By Oblinger, D. ELI Paper 1: EDUCAUSE Learning Initiative. https://net.educause.edu/ir/library/pdf/ELI3009.pdf.
- [36] Nicaise, M, Gibney, T, & Crane, M. (2000). Toward an understanding of authentic learning: Student Perceptions of an Authentic Classroom. *Journal of Science Education and Technology*, vol. 9, no. 1, pp. 79-94, https://doi.org/10.1023/A:1009477008671
- [37] Koh, J.H.L., Chai, C.S., Wong, B., Hong, H.Y. (2015). Design Thinking and 21st Century Skills. In: Design thinking for education, Chapter#3. Springer, Singapore, https://doi.org/10.1007/978-981287-444-3_3
- [38] Noel, K.H.C. & Saranya, E. (2015). Authentic Design Thinking for Special Education Teachers: Two Case Studies with a Special Focus on Autism. *Journal of humanities and social science (IOSR-JHSS)*, Vol. 20 (3). Doi:10.6084/M9.FIGSHARE.1353821.V1
- [39] Leandro, M. A. et al. (2019). ALTRIRAS: A Computer Game for Training Children with Autism Spectrum Disorder in the Recognition of Basic Emotions. *International Journal of Computer Games Technology*, Volume 2019, https://doi.org/10.1155/2019/4384896.
- [40] Grossard, C. et al. (2019). Teaching Facial Expression Production in Autism: The Serious Game JEMImE. *Creative Education*, Vol. 10.
- [41] Min, F., Alissa, N. A., Jianyu, F., Philippe, A., and Sheng, J. (2018). EmoStory: A Game-based System Supporting Children’s Emotional Development. Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems, 2018, https://doi.org/10.1145/3170427.3188594.
- [42] Fridenson-Hayo, S. et al. (2017). Emotiplay: a serious game for learning about emotions in children with autism: results of a cross-cultural evaluation. *European Child and Adolescent Psychiatry*, Vol. 26, No. 8, pp. 979-992, doi: 10.1007/s00787-017-0968-0.
- [43] Serret, S., et al. (2014). Facing the challenge of teaching emotions to individuals with low- and high-functioning autism using a new serious game: a pilot study. *Molecular Autism*, Vol. 5, No. 37.
- [44] Alves, S., Marques, A., Queirós, C. and Orvalho, V., (2013). LIFEisGAME prototype: A serious game about emotions for children with autism spectrum disorders. *PsychNology Journal*, 11(3), 191–21.
- [45] Yan, Y. Liu, C., Ye, L. and Liu, Y. (2018). Using animated vehicles with real emotional faces to improve emotion recognition in Chinese children with autism spectrum disorder. *PLoS ONE*, Vol. 13, No. 7.
- [46] Fedwa, L., Mohamad, E., and El Sadik A., (2014). An Overview of Serious Games. *Int. Journal of Computer Games Technology*, vol. 2014, pp.1-15.
- [47] Plass, J., Homer, D., and Kinze, K., (2015). Foundations of Game-Based Learning. *Educational Psychologist*, Vol. 50, No. 4, pp. 258–283.
- [48] Nikou, S.A., (2023). Student motivation and engagement in maker activities under the lens of the activity theory: A case study in a primary school. *J. Comput. Educ.*, https://doi.org/10.1007/s40692-023-00258-y00258-y
- [49] Hasan, H. & A. Kazlauskas, A., (2014). Activity Theory: who is doing what, why and how?," In H. Hasan (Eds.), *Being practical with theory: A Window into Business Research*, Wollongong, Australia: THEORI, 2014.
- [50] Rajendran, G. (2013). Virtual environments and autism: a developmental psychopathological approach, *Journal of Computer Assisted Learning*, 29, 334–347.
- [51] Fox, J., Beveridge, B. M. and Glasspool, D. W. (2003). Understanding intelligent agents: Analysis and Synthesis. *AI Communications*, vol. 16, no. 3, pp 139–152.
- [52] Russell S. and Norvig, P. (2010). *Artificial Intelligence: A Modern Approach*. Prentice-Hall.
- [53] Cayubit, R.F.O. (2022). Why learning environment matters? An analysis on how the learning environment influences the academic motivation, learning strategies and engagement of college students. *Learning Environments Research* 25, 581-599, https://doi.org/10.1007/s10984021-09382-x.
- [54] Ghazali, R., Md. Sakip, S. R., & Samsuddin, I. (2018). The Effects of Sensory Design on Autistic Children. *Asian Journal of Behavioural Studies*, 3(14), 68–83. https://doi.org/10.21834/ajbes.v3i14.165
- [55] Dehkordi, S. R., Ismail, M., & Diah, N. M. (2022). Game-Based Learning Application for Children with Autism Spectrum Disorder using Participatory Design. *International Journal of Academic Research in Progressive Education and Development*, 11(1), 1-13

- [56] Kim, Y. and Baylor, A. L. (2016). Research-based design of pedagogical agent roles: A review, progress, and recommendations. *Int. J. Artif. Intell. Educ.* Vol. 26, 160–169, doi: 10.1007/s40593-015-0055-y.
- [57] Schroeder, N. L., Romine, W. L. & Craig, S. D. (2017). Measuring pedagogical agent persona and the influence of agent persona on learning. *Comput. Educ.*, vol. 109, pp. 176–186, doi:10.1016/j.compedu.2017.02.015
- [58] Neelu Jyothi, A. et al. (2022). Investigative Study on the Effects of Pedagogical Agents on Intrinsic, Extraneous and Germane Cognitive Load: Experimental Findings With Dyscalculia and Non-Dyscalculia Learners, IEEE Access, DOI: 0.1109/ACCESS.2021.3115409.
- [59] Armando, M., Ochs, M. & Régner, I. (2022). The Impact of Pedagogical Agents' Gender on Academic Learning: A Systematic Review. *Front. Artif. Intell.* 5:862997, doi: 10.3389/frai.2022.862997
- [60] Johnson, W., & Lester, J. C. (2015). Face-to-Face Interaction with Pedagogical Agents, Twenty Years Later. *International Journal of Artificial Intelligence in Education*, vol. 2
- [61] Lutami, P. S., Athallah, F. R., Nur Huda, Y. & Zain, F. D. (2022). A Review of Pathfinding in Game Development. *Journal of Computer Engineering: Progress, Application and Technology*, vol 1, pp. 47-56.
- [62] Janet T. et al. (2022). Games and Rewards: A Scientometric Study of rewards in Educational and Serious Games, IEEE Access.
- [63] Chen, J., Wang, G., Zhang, K., Wang, G. and Liu, L. (2019). A pilot study on evaluating children with autism spectrum disorder using computer games. *Comput. Hum. Behav.*, pp. 204–214.
- [64] Hamari, J., Koivisto, J. and Sarsa, H. (2014). Does Gamification Work? A Literature Review of Empirical Studies on Gamification. In 2014 47th Hawaii Int. Conf. on System Sciences (pp. 3025–3034). IEEE. doi:10.1109/HICSS.2014.377.
- [65] Lim, T., Louchart, S., Suttie, N. et al., (2013). Strategies for effective digital games development and implementation. In Y. Baek & N. Whitton (Eds.), *Cases on digital game-based learning: methods, models, and strategies* (pp. 168–198). Hershey, PA, USA: IGI Global.
- [66] Arnab, S. Lim, T., Carvalho, M. B. et al. (2015). Mapping learning and game mechanics for serious games analysis: Mapping learning and game mechanics. *British Journal of Educational Technology*, Vol. 46, No. 2, pp. 391–411. <http://doi.org/10.1111/bjet.12113>
- [67] Salen K. and Zimmerman, E. (2010). *Rules of play: Game design fundamentals*. Cambridge, MA: The MIT Press, 2010
- [68] Sicart, M. (2008). Designing game mechanics. *International Journal of Computer Game Research*, Vol. 8, No. 2
- [69] Phillip E. C. Compeau and Pavel A. Pevzner, (2018). *Bioinformatics Algorithms: An Active Learning Approach*. 3rd edition, ISBN:978-0-9903746-3-3, Active Learning Publishers
- [70] Gavin, S. & Matthew H., (2012). Investigating Children's Opinions of Games: Fun Toolkit vs. This or That. *Proc. of the 11th international Conf. on Interaction Design and Children*, pp. 70-77. <https://doi.org/10.1145/2307096.2307105>
- [71] McLellan, S., Muddimer, A. & Camille, P. S. (2011). The effect of experience on System Usability Scale Ratings. *Journal of Usability Studies*, vol. 7, no. 2, pp.56-67, 2011
- [72] Albert W., & Tullis, T., (2013). *Measuring the User Experience: Collecting, Analyzing, and Presenting Usability Metrics (Interactive Technologies)*. San Mateo, CA, USA: Morgan Kaufmann.

Adaptive Balance Optimizer: A New Adaptive Metaheuristic and its Application in Solving Optimization Problem in Finance

Purba Daru Kusuma, Ashri Dinimaharawati

Computer Engineering, Telkom University, Bandung, Indonesia

Abstract—Adaptability becomes important in developing metaheuristic algorithms, especially in tackling stagnation. Unfortunately, almost all metaheuristics are not equipped with an adaptive approach that makes them change their strategy when stagnation happens during iteration. Based on this consideration, a new metaheuristic, called an adaptive balance optimizer (ABO), is proposed in this paper. ABO's unique strategy focuses on exploitation when improvement happens and switching to exploration during stagnation. ABO also uses a balanced strategy between exploration and exploitation by performing two sequential searches, whatever circumstance it faces. These sequential searches consist of one guided search and one random search. Moreover, ABO also deploys both a strict acceptance approach and a non-strict acceptance approach. In this work, ABO is challenged to solve a set of 23 classic functions as a theoretical optimization problem and a portfolio optimization problem as the use case for the practical optimization problem. In portfolio optimization, ABO should optimize the quantity of ten stocks in the energy and mining sector listed in the IDX30 index. In this evaluation, ABO is competed with five other metaheuristics: marine predator algorithm (MPA), golden search optimizer (GSO), slime mold algorithm (SMA), northern goshawk optimizer (NGO), and zebra optimization algorithm (ZOA). The simulation result shows that ABO is better than MPA, GSO, SMA, NGO, and ZOA in solving 21, 18, 16, 11, and 8, respectively, in solving 23 functions. Meanwhile, ABO becomes the third-best performer in solving the portfolio optimization problem.

Keywords—Optimization; metaheuristic; adaptability; portfolio optimization; IDX30

I. INTRODUCTION

Metaheuristics is a popular method used in various optimization problems. In the cloud system, the genetic algorithm (GA) was modified in service caching and task offloading to improve resource efficiency and user satisfaction [1]. A deep convolutional neural network is enriched with a gorilla troop optimizer (GTO) to improve its capability in diagnosing skin cancer [2]. The whale optimization algorithm (WOA) was used to solve portfolio optimization based on the FTSE100 index [3]. An improved sparrow search algorithm (ISSA) was developed to improve the high-intensity focused ultrasound (HIFU) technology that is used for tumor treatment [4]. A combination of tabu search (TS) and simulated annealing (SA) is used to solve the coupled task scheduling of the heterogeneous multiple automated guided vehicles (AGV) in a manufacturing system [5]. Its popularity comes from two

reasons. The first reason is that there are a huge number of metaheuristics already existing today. The second reason is that metaheuristic uses a stochastic approach to be efficient enough in solving large-scale optimization problems with limited computational resources [6]. Moreover, metaheuristic is also flexible enough to solve various kinds of problems by abstracting the problem. It focuses on the objectives and constraints of these problems. Then, it uses a trial-and-error mechanism to improve the solution through iteration. Meanwhile, this stochastic approach means that all metaheuristics do not guarantee finding the optimal global solution but only the high-quality or quasi-optimal one [7]. Besides, metaheuristic is also challenged with the optimal local issue.

One important consideration in metaheuristics is adaptability. Adaptability is important, especially in facing the circumstance of whether the current search produces a better solution or not. Each metaheuristic was developed based on a strategy for finding a better solution. This new solution is then used for the next iteration for several purposes. In some metaheuristics, new solutions are used to choose the reference for the guided search. Some metaheuristics rank the solutions and then split them into several groups where each group performs its strategy. Some other metaheuristics rank the solutions to eliminate the worst solution or some worst solutions for the next iteration.

Unfortunately, almost all metaheuristics are not adaptive enough. Many metaheuristics do not care about the quality of the new solution relative to the previous solution. The new solution still replaces the current solution, although this new solution is not better than the existing solution. This mechanism can be found in many metaheuristics, such as grey wolf optimizer (GWO) [8], MPA [9], GSO [10], SMA [11], darts game optimizer (DGO) [12], butterfly optimization algorithm (BOA) [13], chameleon swarm algorithm (CSA) [14], tunicate swarm algorithm (TSA) [15], squirrel search optimizer (SSO) [16], coronavirus optimization algorithm (COVIDOA) [17], white shark optimizer (WSO) [18], stochastic paint optimizer (SPO) [19], and so on. In some metaheuristics, a strict-acceptance approach is applied. Through this approach, a new solution is accepted to replace the current solution only if it is better than the current one. This approach can be found in many latest metaheuristics, such as the pelican optimization algorithm (POA) [20], guided pelican algorithm (GPA) [21], total interaction algorithm (TIA) [22], three-on-three optimizer (TOTO) [23], average and

subtraction-based optimizer (ASBO) [24], northern goshawk optimizer (NGO) [25], zebra optimization algorithm (ZOA) [26], coati optimization algorithm (COA) [27], fennec fox optimization (FFO) [28], chef-based optimization algorithm (CBOA) [29], modified honey badger algorithm (MHBA) [30], flower pollination algorithm (FPA) [31], football game based optimizer (FBGO) [32], red fox optimization algorithm (RFO) [33], and so on.

On the other hand, few metaheuristics are adaptive enough when it fails to improve. In KMA [34], the population size increases if stagnation occurs and decreases when improvement occurs. There is a static number of the increasing or decreasing population. The population size can increase until the maximum population size, while the population size can decrease until the minimum population size. In an artificial bee colony (ABC) [35], the bee performs a full random search after it fails to improve for certain periods.

The other consideration is the use case used to evaluate the metaheuristic when it was first introduced. In general, all metaheuristics were tested by using theoretical optimization problems. This theoretical problem consists of a set of mathematical functions. The set of 23 classic functions has been widely used in the first introduction of many metaheuristics, such as in the first introduction of KMA [34]. Other functions are CEC 2015, CEC 2017, and so on. In some studies, the new metaheuristics were also challenged to solve practical problems. Some optimization problems in mechanical engineering are commonly used, such as pressure vessel design problems, speed reducer design problems, welded beam design problems, and tension-compression spring design problems. The power flow optimization problem in the energy sector is also a common use case. Unfortunately, a study that uses optimization problems in the financial sector is rare.

Based on this consideration, especially on the adaptability and use case issues, this work is aimed to develop a new metaheuristic that is adaptive enough to tackle the stagnation problem. This stagnation can be detected, especially when the optimization process fails to improve the quality of the current solution during the iteration.

The main scientific contributions presented in this work are described below:

- 1) A new metaphor-free metaheuristic called as adaptive balance optimizer (ABO) is presented.
- 2) This work presents the adaptive strategy used in ABO, especially in switching between exploration and exploitation.
- 3) The performance of ABO is evaluated by implementing it to solve both theoretical optimization problem (a set of 23 classic functions) and practical optimization problem (portfolio optimization problem).
- 4) The performance of ABO is also competed with five other metaheuristics: MPA, SMA, GSO, NGO, and ZOA.
- 5) The hyper parameter evaluation is performed to evaluate the performance of ABO due to the increase of maximum iteration and population size.

The rest of this paper is organized as follows. The literature review regarding the latest of metaheuristics is performed in

Section II. A detailed description of the adaptive balance optimizer consisting of its main concept, algorithm, and mathematical model is presented in Section III. The evaluation regarding the performance of ABO, especially in solving the set of 23 classic functions, the hyperparameter test, and the portfolio optimization problem, is presented in Section IV. The in-depth analysis of the simulation result, the drawback regarding the theory, limitations, and the algorithm complexity is discussed in Section V. Finally, the conclusion and the potential of future studies and development are summarized in Section VI.

II. RELATED WORKS

Adaptability is one important issue in the development of metaheuristics. Ironically, most of all, metaheuristics were developed without considering this issue. Many metaheuristics focus on developing strategies regarding the exploration and exploitation capability statically. It means most metaheuristics perform the same installed strategy, whether the improvement or stagnation happens.

Many metaheuristics respond to the improvement or stagnation by determining whether the new solution will be accepted to replace the previous solution or not. Some metaheuristics deploy a strict acceptance approach, meaning that a new solution replaces the previous solution only if the improvement occurs. On the other hand, some other metaheuristics deploy a non-strict acceptance approach which means that a new solution will replace the previous solution despite the improvement of stagnation. One distinct approach is introduced by simulated annealing, which uses a stochastic acceptance approach. If the improvement occurs, the new solution will replace the previous one immediately. Otherwise, the new solution may replace the previous solution based on a stochastic calculation. Ironically, metaheuristics that use improvement or stagnation circumstance to decide which strategy will be performed in the next iteration is rare to find.

Fortunately, some metaheuristics perform adaptive strategies in response to improvement or stagnation. KMA uses improvement or stagnation to determine the population size for the next iteration [34]. When the improvement occurs for two successive iterations, the population size decreases to reduce computational consumption. On the other hand, when stagnation occurs in two successive iterations, the population size increases to boost the exploration effort. However, the searches are still the same because whether the improvement or stagnation takes place, there are still three groups of agents where each group performs different searches. Meanwhile, an artificial bee colony (ABC) runs a different approach to make it adaptive. ABC generally performs neighborhood search and roulette wheel selection [35]. Meanwhile, after stagnation takes places for certain periods, full random search is performed without performing strict acceptance approach [35].

This adaptability issue is also not popular in the development of latest metaheuristics. Many metaheuristics, especially those that use a metaphor, focus on exploiting the mechanism of their metaphor as a novelty or contribution. Besides, many latest metaheuristics exploit their capability to outperform other metaheuristics as proof of their superiority.

TABLE I. SUMMARY OF SOME LATEST METAHEURISTICS

No	Metaheuristic	Metaphor	Adaptability	Acceptance Approach	Use Case
1	MHBA [30]	honey badger	no	strict	power flow
2	SPO [19]	paint	no	non strict	23 functions, CEC 2019, 52-bar planar truss structure, 120-bar dome truss structure, 3-bay 15-story frame, 3-bay 24-story frame
3	KMA [34]	komodo	population size decreases when improvement occurs and increases when stagnation occurs	non strict	23 functions
4	COVIDOA [17]	coronavirus	no	non strict	20 functions, CEC 2011
5	SSO [16]	squirrel	no	non strict	power system
6	MPA [9]	marine predator	no	non strict	30 functions, pressure vessel design, welded beam design, tension/compression spring design, operating fan schedule, building energy performance
7	GWO [8]	grey wolf	no	non strict	29 functions, tension/compression spring design, welded beam design, pressure vessel design, optical buffer design
8	BOA [13]	butterfly	no	non strict	30 functions, spring design, welded beam design, gear train design
9	POA [20]	pelican	no	strict	23 functions, pressure vessel design, speed reducer design, welded beam design, tension/compression spring design
10	TIA [22]	-	no	strict	23 functions
11	ASBO [24]	-	no	strict	23 functions
12	NGO [25]	northern goshawk	no	strict	23 functions, CEC 2015, CEC 2019, pressure vessel design, welded beam design, tension/compression spring design, speed reducer design
13	ZOA [26]	zebra	no	strict	23 functions, CEC 2017, tension/compression spring design, welded beam design, speed reducer design, pressure vessel design
14	COA [27]	coati	no	strict	CEC 2011, CEC 2017, pressure vessel design, speed reducer design, welded beam design, tension/compression spring design
15	RFO [33]	red fox	no	strict	22 functions, three bar truss, welded beam design, compression spring, pressure vessel, gear train
16	this work	-	different strategy during improvement and stagnation	strict and non-strict	23 functions, portfolio optimization problem

The other issue concerns the practical use case chosen to evaluate new metaheuristics in their first introduction. Problems in engineering are so popular; whether mechanical, civil, or electrical problems. Meanwhile, studies regarding new metaheuristic that uses problems in finance, are rare to find.

The summarized review of some latest metaheuristics is presented in Table I. There are 15 metaheuristics presented in Table I. Moreover, the proposed metaheuristic is placed in the last row to clarify its position among the existing metaheuristics.

Table I indicates that almost all metaheuristics have not considered adaptability. These metaheuristics perform the same strategy from the beginning of the iteration until the maximum iteration is reached. Some metaheuristics perform a strict acceptance approach to avoid the optimization process going to the worse solution. In comparison, some others still accept a worse solution, hoping it may lead to a better solution. Moreover, the financial sector, still not popular, became a practical use case to evaluate a new metaheuristic when it was first introduced.

Based on this circumstance, this work proposes a new metaheuristic that is adaptive enough when there is no improvement regarding the current solution. Moreover, the

proposed metaheuristic gives equal treatment between two circumstances: improvement succeeds or fails. Besides, the proposed metaheuristic also performs both a strict acceptance approach and a non-strict acceptance approach.

III. PROPOSED MODEL

An adaptive balance optimizer (ABO) is designed as an adaptive metaheuristic that gives balance effort in intensifying the quality of the current solution and being adaptive when the improvement fails. Based on this objective, the reasoning for constructing ABO is as follows. ABO performs multiple searches, as many latest of metaheuristics also perform this approach. ABO performs guided search and random search explicitly. ABO performs different strategies when the improvement fails. ABO accommodates both a strict acceptance approach and a non-strict acceptance approach.

Based on this reasoning, the main concept of ABO is dividing strategy based on the circumstance it faces. There are two possible circumstances. The first circumstance is that the improvement happens. In this first circumstance, the strategy intensifies the exploitation. The second circumstance is that stagnation takes place. In this second circumstance, the strategy is deploying exploration. There are two searches in every circumstance: guided search and random search.

There are two searches performed in the exploitation mode. The first search is a guided search toward and beyond the best global solution. This guided search aims to trace possible better solutions between the corresponding solution and the global best solution. Moreover, this guided search is also designed to trace possible better solutions beyond the global best solution. As known, the global best solution is assumed as the best solution so far. It means that the quality of the best global solution is better than that of the corresponding solution. The probability of finding a better solution will increase when the corresponding solution moves closer to the global best solution. Meanwhile, it is also more probable for the global best solution to find a better solution by avoiding a worse solution. The second search is the limited random search or neighborhood search. In general, as a random search, the corresponding solution traces a new solution around its current solution. However, the search space is reduced as the iteration increases. In the exploitation mode, the strict acceptance approach is deployed in a guided search toward the global best solution and the limited random search. The searches in the exploitation mode are illustrated in Fig. 1.

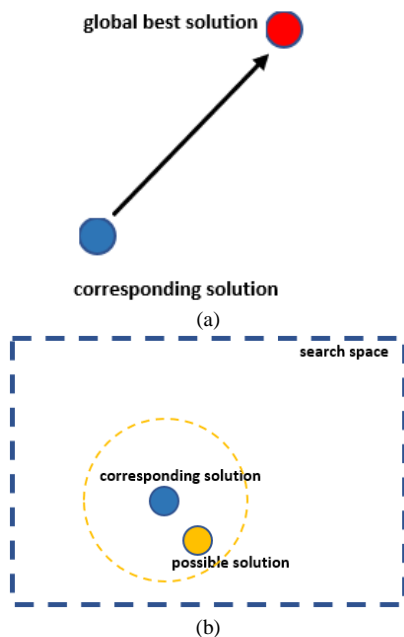


Fig. 1. Searches when improvement is achieved: (a) guided search toward the global best solution, (b) limited neighborhood search.

There are two searches performed in the exploration mode. The first search is the guided search relative to a randomly selected solution. This search can be viewed as a guided exploration. This search is included in the exploration because it is used a randomly selected solution among the population as the reference. The population is known to be spread within the search space, especially in the early iteration. Moving relative to one of these solutions means the corresponding solution traces any solution within the search space but based on a reference. If the reference quality is better than the quality of the corresponding solution, the corresponding solution moves toward the reference. Otherwise, the corresponding solution avoids this reference. The direction of this movement may push the corresponding solution closer to or away from the best

global solution. This search is performed based on the reasoning that although the global best solution is the best solution, getting closer to the global best solution may push the corresponding solution to the local optimal entrapment. The second search is a full random search. As its name, the corresponding solution moves uniformly within the search space. This search can be viewed as a full exploration. In this exploration mode, the strict acceptance approach is not deployed. It means the new solution replaces the existing solution without considering the quality of this new solution. Moving to the worse solution may be better, which may lead to a better solution, rather than staying in the current solution without any improvement until the iteration ends. These two searches in the exploration mode are illustrated in Fig. 2.

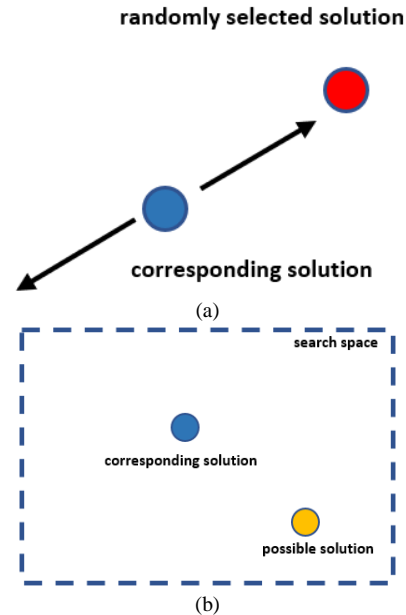


Fig. 2. Searches when improvement fails: (a) guided search relative to a randomly selected solution, (b) full random search.

This ‘two-approach mechanism’ needs a toggle to indicate whether the corresponding solution is in the exploitation or exploration mode. This toggle has two values updated at the end of every iteration. Suppose the corresponding solution fails to improve its quality after performing two sequential searches, whether, in the exploration or exploitation mode, the value of the toggle will be set so that exploration will be performed in the next iteration. Otherwise, the value of the toggle will be set so that exploitation will be performed in the next iteration.

The concept of ABO is then formalized using algorithm 1. As a metaheuristic, ABO consists of two phases: initialization and iteration. Lines 2 to 5 represent the initialization phase, while lines 6 to 25 represent the iteration phase. In the initialization phase, only one loop runs for the entire population. Meanwhile, two loops run in the iteration phase. The outer loop runs from the first iteration to the maximum iteration. The inner loop runs for the entire population. Lines 9 to 12 represent the exploitation mode, while lines 14 to 17 represent the exploration mode. Lines 19 to 23 represent the toggle updating process. g_s denotes the toggle value where 1 indicates the exploitation while 0 indicates the exploration.

Algorithm 1: Adaptive balance optimizer

```

1  begin
2  for all  $s$  in  $S$ 
3    perform full random search using (1)
4    update  $s_b$  using (2)
5  end
6  for  $t = 1$  to  $t_m$ 
7    for all  $s$  in  $S$ 
8      if  $g_s = 1$  then
9        perform first guided search using (3)
10       update  $s$  using (2) and  $s_b$  using (3)
11       perform limited random search using (5)
12       update  $s$  using (2) and  $s_b$  using (3)
13      else
14        perform second guided search using (6) and (7)
15        update  $s_b$  using (3)
16        perform full random search using (1)
17        update  $s_b$  using (3)
18      end if
19      if improvement fail then
20         $g_s = 0$ 
21      else
22         $g_s = 1$ 
23      end if
24    end for
25  end for
26 end
27 output:  $s_b$ 

```

The initialization phase consists of two processes. In the first process, the full random search is performed to generate initial solutions. This full random search is formalized using (1) where s denotes the solution, s_l denotes the lower boundary, s_u denotes the upper boundary, and U denotes the uniform random. The second process is the global best updating process as formalized in (2). In (2), s_b denotes the global best solution and f denotes the objective function.

$$s = U(s_l, s_u) \quad (1)$$

$$s_b' = \begin{cases} s, f(s) < f(s_b) \\ s', else \end{cases} \quad (2)$$

The exploration mode consists of two searches. The guided search toward the global best solution is formalized using (3) while the limited random search is formalized using (5). s_c denotes the solution candidate which is then evaluated using (4) that represents the strict acceptance approach. In (5), it is shown that the local search space gets narrow as the iteration increases.

$$s_c = s + U(0,1) \cdot (s_b - 2s) \quad (3)$$

$$s' = \begin{cases} s_c, f(s_c) < f(s) \\ s, else \end{cases} \quad (4)$$

$$s_c = s + U(-1,1) \cdot \left(1 - \frac{t}{t_m}\right) \cdot \left(\frac{s_u - s_l}{2}\right) \quad (5)$$

The guided search relative to randomly solution is formalized using (6) and (7). s_s denotes the randomly selected solution which is selected uniformly among the population S . Meanwhile, as shown in (7), the corresponding solution moves

toward the reference if the quality of the reference is better than the quality of the corresponding solution. Otherwise, the corresponding solution avoids the reference.

$$s_s = U(S) \quad (6)$$

$$s = \begin{cases} s + U(0,1) \cdot (s - 2s_s), f(s) < f(s_s) \\ s + U(0,1) \cdot (s_s - 2s), else \end{cases} \quad (7)$$

Moreover, the detailed explanation regarding the influence of parameters used in this algorithm is as follows. The solution s plays an important role as autonomous agent performing the searching process. Meanwhile, the population S as the collection of s actualizes the population-based metaheuristics. The greater size of S in general improves the exploration capability although the size is not always linear to the performance quality. The s_b represents the best solution which is the basic form of collective intelligence in the swarm-based metaheuristic. Iteration t control the iterative process which is limited by the maximum iteration t_m . Uniform random U is used for the stochastic process which becomes the foundation of any metaheuristic. The boundaries s_l and s_u are used as the hard constraint in finding the possible solution. The objective function f is used for measurement of the quality of any solution.

IV. SIMULATION AND RESULT

This section presents the simulation and evaluation of ABO in solving optimization problems. This evaluation can be split into three parts. The first part is a simulation regarding the theoretical optimization problem. The second part is a simulation regarding the hyperparameter test. The third part is a simulation regarding the practical optimization problem. In the first and third parts, ABO is benchmarked with five latest metaheuristics: MPA, GSO, SMA, NGO, and ZOA. MPA, GSO, and SMA are metaheuristics that do not deploy a strict acceptance approach. On the other hand, NGOs and ZOA are metaheuristics that deploy a strict acceptance approach. Meanwhile, ABO plays in the middle by deploying a strict acceptance approach and not a strict acceptance approach based on the circumstance it faces.

The first part is the simulation regarding the theoretical optimization problem. In this work, the set of 23 functions is used as the use case. This set of functions is chosen due to its broad and diverse circumstances and challenges. It can be split into three groups: high-dimension unimodal, high-dimension multimodal, and fixed-dimension multimodal functions. A detailed description of these functions is presented in Table II.

In this work, several adjusted parameters are set as follows. In general, the maximum iteration is 50 while the population size is 5. The fishing aggregate device for MPA is set 0.5 that represents balance between exploration within the search space and the guided exploration toward two randomly selected solutions. The z score for SMA is 0.5. The result in solving the high dimension unimodal functions, high dimension multimodal functions, and fixed dimension multimodal functions is presented in Table III, Table IV, and Table V respectively.

TABLE II. DETAIL DESCRIPTION OF 23 FUNCTIONS

No	Function	Model	Dimension	Problem Space	Global Opt.
1	Sphere	$\sum_{i=1}^d x_i^2$	50	[-100, 100]	0
2	Schwefel 2.22	$\sum_{i=1}^d x_i + \prod_{i=1}^d x_i $	50	[-100, 100]	0
3	Schwefel 1.2	$\sum_{i=1}^d \left(\sum_{j=1}^i x_j \right)^2$	50	[-100, 100]	0
4	Schwefel 2.21	$\max\{ x_i , 1 \leq i \leq d\}$	50	[-100, 100]	0
5	Rosenbrock	$\sum_{i=1}^{d-1} (100(x_{i+1} + x_i^2)^2 + (x_i - 1)^2)$	50	[-30, 30]	0
6	Step	$\sum_{i=1}^{d-1} (x_i + 0.5)^2$	50	[-100, 100]	0
7	Quartic	$\sum_{i=1}^d i x_i^4 + \text{random} [0,1]$	50	[-1.28, 1.28]	0
8	Schwefel	$\sum_{i=1}^d -x_i \sin(\sqrt{ x_i })$	50	[-500, 500]	-418.9x50
9	Rastrigin	$10d + \sum_{i=1}^d (x_i^2 - 10 \cos(2\pi x_i))$	50	[-5.12, 5.12]	0
10	Ackley	$-20 \cdot \exp\left(-0.2 \cdot \sqrt{\frac{1}{d} \sum_{i=1}^d x_i^2}\right) - \exp\left(\frac{1}{d} \sum_{i=1}^d \cos 2\pi x_i\right) + 20 + \exp(1)$	50	[-32, 32]	0
11	Griewank	$\frac{1}{4000} \sum_{i=1}^d x_i^2 - \prod_{i=1}^d \cos\left(\frac{x_i}{\sqrt{i}}\right) + 1$	50	[-600, 600]	0
12	Penalized	$\frac{\pi}{d} \left\{ 10 \sin(\pi y_1) + \sum_{i=1}^{d-1} \left((y_i - 1)^2 (1 + 10 \sin^2(\pi y_{i+1})) \right) + (y_d - 1)^2 \right\} + \sum_{i=1}^d u(x_i, 10, 100, 4)$	50	[-50, 50]	0
13	Penalized 2	$0.1 \left\{ \sin^2(3\pi x_1) + \sum_{i=1}^{d-1} \left((x_i - 1)^2 (1 + \sin^2(3\pi x_{i+1})) \right) + (x_d - 1)^2 (1 + \sin^2(2\pi x_d)) \right\} + \sum_{i=1}^d u(x_i, 5, 100, 4)$	50	[-50, 50]	0
14	Shekel Foxholes	$\left(\frac{1}{500} + \sum_{j=1}^{25} \frac{1}{j + \sum_{i=1}^2 (x_i - a_{ij})^6} \right)^{-1}$	2	[-65, 65]	1
15	Kowalik	$\sum_{i=1}^{11} \left(a_i - \frac{x_1(b_i^2 + b_i x_2)}{b_i^2 + b_i x_3 + x_4} \right)^2$	4	[-5, 5]	0.0003
16	Six Hump Camel	$4x_1^2 - 2.1x_1^4 + \frac{1}{3}x_1^6 + x_1x_2 - 4x_2^2 + 4x_2^4$	2	[-5, 5]	-1.0316
17	Branin	$\left(x_2 - \frac{5.1}{4\pi^2} x_1^2 + \frac{5}{\pi} x_1 - 6 \right)^2 + 10 \left(1 - \frac{1}{8\pi} \right) \cos(x_1) + 10$	2	[-5, 5]	0.398
18	Goldstein-Price	$(1 + (x_1 + x_2 + 1)^2 (19 - 14x_1 + 3x_1^2 - 14x_2 + 6x_1x_2 + 3x_2^2)) \cdot (30 + (2x_1 - 3x_2)^2 (18 - 32x_1 + 12x_1^2 + 48x_2 - 36x_1x_2 + 27x_2^2))$	2	[-2, 2]	3

19	Hartman 3	$-\sum_{i=1}^4 \left(c_i \exp \left(-\sum_{j=1}^d (a_{ij}(x_j - p_{ij})^2) \right) \right)$	3	[1, 3]	-3.86
20	Hartman 6	$-\sum_{i=1}^4 \left(c_i \exp \left(-\sum_{j=1}^d (a_{ij}(x_j - p_{ij})^2) \right) \right)$	6	[0, 1]	-3.32
21	Shekel 5	$-\sum_{i=1}^5 \left(\sum_{j=1}^d (x_j - c_{ji})^2 + \beta_i \right)^{-1}$	4	[0, 10]	-10.1532
22	Shekel 7	$-\sum_{i=1}^7 \left(\sum_{j=1}^d (x_j - c_{ji})^2 + \beta_i \right)^{-1}$	4	[0, 10]	-10.4028
23	Shekel 10	$-\sum_{i=1}^{10} \left(\sum_{j=1}^d (x_j - c_{ji})^2 + \beta_i \right)^{-1}$	4	[0, 10]	-10.5363

TABLE III. BENCHMARK RESULT IN SOLVING HIGH DIMENSION UNIMODAL FUNCTIONS

F	Parameter	MPA [9]	GSO [10]	SMA [11]	NGO [25]	ZOA [26]	ABO
1	mean	4.3320x10 ³	5.6223x10 ⁴	7.4626x10 ⁴	0.0286	0.0000	0.0000
	st dev	1.9644x10 ³	1.3632x10 ⁴	1.1712x10 ⁴	0.0306	0.0000	0.0000
	min	1.4711x10 ³	3.1528x10 ⁴	4.1722x10 ⁴	0.0006	0.0000	0.0000
	max	8.5016x10 ³	8.1858x10 ⁴	9.7149x10 ⁴	0.0940	0.0000	0.0000
	mean rank	4	5	6	3	1	1
2	mean	0.0000	2.9343x10 ⁶⁷	0.0000	0.0000	0.0000	0.0000
	st dev	0.0000	1.0545x10 ⁶⁸	0.0000	0.0000	0.0000	0.0000
	min	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	max	0.0000	4.8598x10 ⁶⁸	0.0000	0.0000	0.0000	0.0000
	mean rank	1	6	1	1	1	1
3	mean	2.8487x10 ⁴	1.3073x10 ⁵	2.1393x10 ⁵	5.4226x10 ³	0.0266	9.3981x10 ¹
	st dev	1.9967x10 ⁴	5.5564x10 ⁴	4.0595x10 ⁴	5.4283x10 ³	0.1163	3.2324x10 ²
	min	3.0737x10 ³	4.8925x10 ⁴	1.3394x10 ⁵	1.0971x10 ²	0.0000	0.0000
	max	8.5281x10 ⁴	2.5705x10 ⁵	2.6921x10 ⁵	2.3616x10 ⁴	0.5470	1.7244x10 ³
	mean rank	4	5	6	3	1	2
4	mean	2.9387x10 ¹	5.3229x10 ¹	8.6454x10 ¹	0.4646	0.0000	0.0000
	st dev	5.5277	5.9824	2.5769	0.2558	0.0000	0.0000
	min	1.6824x10 ¹	4.0608x10 ¹	8.2000x10 ¹	0.1033	0.0000	0.0000
	max	4.3961x10 ¹	6.4096x10 ¹	9.1000x10 ¹	1.0545	0.0000	0.0000
	mean rank	4	5	6	3	1	1
5	mean	1.1760x10 ⁶	1.4327x10 ⁸	2.5991x10 ⁸	4.9230x10 ¹	4.8938x10 ¹	4.8948x10 ¹
	st dev	6.8289x10 ⁵	5.4495x10 ⁷	4.5292x10 ⁷	0.3290	0.0199	0.0300
	min	1.8459x10 ⁴	5.1647x10 ⁷	1.9294x10 ⁸	4.8997x10 ¹	4.8907x10 ¹	4.8844x10 ¹
	max	2.6106x10 ⁶	2.4223x10 ⁸	3.4018x10 ⁸	5.0357x10 ¹	4.8970x10 ¹	4.8982x10 ¹
	mean rank	4	5	6	3	1	2
6	mean	4.8607x10 ³	5.5470x10 ⁴	7.3622x10 ⁴	1.0499x10 ¹	1.0452x10 ¹	1.0797x10 ¹
	st dev	1.5346x10 ³	9.7258x10 ³	1.1486x10 ⁴	0.5606	0.6440	0.4690
	min	2.4904x10 ³	3.9545x10 ⁴	4.2292x10 ⁴	9.3723	8.9883	9.6110
	max	7.7010x10 ³	7.5723x10 ⁴	8.6172x10 ⁴	1.1632x10 ¹	1.1331x10 ¹	1.1453x10 ¹
	mean rank	4	5	6	2	1	3
7	mean	1.4044	1.0796x10 ²	4.9953x10 ²	0.0341	0.0064	0.0181
	st dev	0.8089	4.7866x10 ¹	6.3307x10 ¹	0.0176	0.0047	0.0139
	min	0.2819	4.2202x10 ¹	3.9856x10 ²	0.0047	0.0007	0.0015
	max	3.4301	2.5038x10 ²	6.0442x10 ²	0.0756	0.0201	0.0454
	mean rank	4	5	6	3	1	2

Table III indicates that ABO performs well in solving high-dimension unimodal functions. ABO becomes the first best in solving three functions (Sphere, Schwefel 2.22, and Schwefel 2.21), second best in solving three functions (Schwefel 1.2, Rosenbrock, and Quartic), and third best in solving one function (Step). Table III also indicates a significant gap between three metaheuristics (ABO, ZOA, and NGO) and the three others (SMA, MPA, and GSO), where the first group's performance is far better than the second group.

Table IV indicates that ABO performs well in solving high-dimension multimodal functions. ABO becomes the first best in solving two functions (Rastrigin and Ackley), second best in solving three functions (Griewank, Penalized, and Penalized 2), and fourth best in solving one function (Schwefel). Moreover, ABO can find the optimal global solution in solving Rastrigin and Ackley. Meanwhile, ZOA can also find the optimal global solution for these two functions.

The result in Table IV also divides these metaheuristics into two groups. The first group consists of MPA, GSO, and SMA.

On the other hand, the second group consists of NGO, ZOA, and ABO. Metaheuristics in the second group perform better than metaheuristics in the first group, while the performance gap is significant. This circumstance takes place in almost all high-dimension multimodal functions except Schwefel.

Table V indicates fierce competition among these six metaheuristics in solving fixed-dimension multimodal functions. Fortunately, ABO also performs well in solving these functions. Among these ten functions, ABO performs as the second best in six functions (Shekel Foxholes, Kowalik, Branin, Goldstein-Price, Hartman 3, and Hartman 6), third best in one function (Six Hump Camel), fourth best in one function (Shekel 5), and fifth best in two functions (Shekel 7 and Shekel 10). Different from the high-dimension unimodal and high-dimension multimodal functions, fierce competition happens for all metaheuristics in all ten functions. In general, the performance gap among the metaheuristics is narrow.

TABLE IV. BENCHMARK RESULT IN SOLVING HIGH DIMENSION MULTIMODAL FUNCTIONS

F	Parameters	MPA [9]	GSO [10]	SMA [11]	NGO [25]	ZOA [26]	ABO
8	mean	-3.4011x10 ³	-4.6121x10 ³	-4.9782x10 ³	-4.2472x10 ³	-2.8643x10 ³	-3.9093x10 ³
	st dev	6.5297x10 ²	1.2521x10 ³	5.6572x10 ²	6.4312x10 ²	5.6143x10 ²	5.8738x10 ²
	min	-5.3489x10 ³	-7.2897x10 ³	-6.0276x10 ³	-5.3898x10 ³	-3.7800x10 ³	-5.2110x10 ³
	max	-2.3316x10 ³	-2.6163x10 ³	-3.7595x10 ³	-3.1568x10 ³	-1.8763x10 ³	-2.9836x10 ³
	mean rank	5	2	1	3	6	4
9	mean	3.5962x10 ²	5.1678x10 ²	2.0979x10 ²	1.5365	0.0000	0.0000
	st dev	7.6955x10 ¹	5.6595x10 ¹	4.0802x10 ¹	4.6372	0.0000	0.0000
	min	1.3328x10 ²	3.9772x10 ²	1.3101x10 ²	0.0096	0.0000	0.0000
	max	4.4327x10 ²	6.3509x10 ²	2.7401x10 ²	2.1832x10 ¹	0.0000	0.0000
	mean rank	5	6	4	3	1	1
10	mean	9.9728	1.9884x10 ¹	1.7682x10 ¹	0.0984	0.0000	0.0000
	st dev	1.5062	0.5195	0.3827	0.2589	0.0000	0.0000
	min	5.6136	1.7694x10 ¹	1.6002x10 ¹	0.0061	0.0000	0.0000
	max	1.1753x10 ¹	2.0513x10 ¹	1.8082x10 ¹	1.1222	0.0000	0.0000
	mean rank	4	6	5	3	1	1
11	mean	4.3770x10 ¹	4.8578x10 ²	6.0186x10 ²	0.0592	0.0000	0.0209
	st dev	1.6561x10 ¹	1.0101x10 ²	1.0956x10 ²	0.1247	0.0000	0.0805
	min	2.2313x10 ¹	3.4842x10 ²	3.8985x10 ²	0.0004	0.0000	0.0000
	max	7.8995x10 ¹	6.7935x10 ²	7.9722x10 ²	0.5502	0.0000	0.3806
	mean rank	4	5	6	3	1	2
12	mean	9.8780x10 ⁴	2.1018x10 ⁸	5.3614x10 ⁸	1.0144	1.0824	1.0486
	st dev	1.6112x10 ⁵	1.6467x10 ⁸	1.4009x10 ⁸	0.1346	0.1164	0.1113
	min	3.1218x10 ²	4.6630x10 ⁷	2.0839x10 ⁸	0.7859	0.8031	0.8918
	max	6.2116x10 ⁵	6.4261x10 ⁸	7.7637x10 ⁸	1.2334	1.2492	1.2878
	mean rank	6	4	5	1	3	2
13	mean	1.5887x10 ⁶	5.5397x10 ⁸	1.0300x10 ⁹	3.3180	3.0916	3.1319
	st dev	2.1136x10 ⁶	3.4988x10 ⁸	2.1375x10 ⁸	0.1249	0.0393	0.0176
	min	9.1574x10 ³	1.8162x10 ⁸	6.7786x10 ⁸	3.1396	2.9894	3.0821
	max	1.0123x10 ⁷	1.5915x10 ⁹	1.4917x10 ⁹	3.5531	3.1363	3.1835
	mean rank	4	5	6	3	1	2

TABLE V. BENCHMARK RESULT IN SOLVING FIXED DIMENSION MULTIMODAL FUNCTIONS

F	Parameters	MPA [9]	GSO [10]	SMA [11]	NGO [25]	ZOA [26]	ABO
14	mean	1.1070x10 ¹	1.0182x10 ¹	5.4396	6.9609	9.3836	6.8624
	st dev	4.1377	4.8001	3.5089	4.6037	3.7321	3.0608
	min	3.0579	1.9920	0.9980	0.9980	0.9980	2.0116
	max	1.7717x10 ¹	2.1073x10 ¹	1.2671x10 ¹	1.6441x10 ¹	1.2670x10 ¹	1.3619x10 ¹
	mean rank	6	5	1	3	4	2
15	mean	0.0254	0.0910	0.1326	0.0046	0.0076	0.0048
	st dev	0.0158	0.3541	0.0236	0.0068	0.0139	0.0045
	min	0.0050	0.0016	0.0778	0.0004	0.0003	0.0010
	max	0.0666	1.6747	0.1484	0.0206	0.0462	0.0165
	mean rank	4	5	6	1	3	2
16	mean	-0.9560	-1.0198	-0.0367	-1.0316	-0.9480	-1.0159
	st dev	0.0620	0.0315	0.1083	0.0000	0.2207	0.0189
	min	-1.0218	-1.0316	-0.4578	-1.0316	-1.0316	-1.0315
	max	-0.7795	-0.8923	0.0000	-1.0316	-0.1680	-0.9609
	mean rank	4	2	6	1	5	3
17	mean	3.4078	1.3439	0.6438	0.3982	7.4391	0.4691
	st dev	2.5163	2.7178	0.0000	0.0003	1.0158x10 ¹	0.0688
	min	0.5003	0.3981	0.6438	0.3981	0.3981	0.3981
	max	8.7875	1.0341x10 ¹	0.6438	0.3993	3.5964x10 ¹	0.6438
	mean rank	5	4	3	1	6	2
18	mean	3.0989x10 ¹	1.3393x10 ¹	3.0000	2.9077x10 ¹	4.7903x10 ¹	4.0246
	st dev	2.2928x10 ¹	2.4219x10 ¹	0.0000	3.5703x10 ¹	7.8849x10 ¹	1.3238
	min	4.5662	3.0000	3.0000	3.0000	2.9999	3.0000
	max	8.1122x10 ¹	8.4200x10 ¹	3.0000	8.4824x10 ¹	3.3648x10 ²	7.2993
	mean rank	5	3	1	4	6	2
19	mean	-3.3279	-0.0357	-0.0495	-0.0495	-0.0495	-0.0495
	st dev	0.4178	0.0174	0.0000	0.0000	0.0000	0.0000
	min	-3.8536	-0.0495	-0.0495	-0.0495	-0.0495	-0.0495
	max	-2.2160	-0.0056	-0.0495	-0.0495	-0.0495	-0.0495
	mean rank	1	6	2	2	2	2
F	Parameters	MPA [9]	GSO [10]	SMA [11]	NGO [25]	ZOA [26]	ABO
20	mean	-1.3750	-2.5835	-0.9122	-2.7680	-2.1337	-2.6242
	st dev	0.5288	0.6656	0.6304	0.3178	0.4571	0.1370
	min	-2.2961	-3.2141	-2.5003	-3.2903	-3.0922	-2.7921
	max	-0.5046	-0.9227	-0.1587	-2.1049	-1.4533	-2.3383
	mean rank	5	3	6	1	4	2
21	mean	-0.8384	-3.7757	-2.1840	-3.0163	-3.7476	-2.4582
	st dev	0.2820	2.5466	2.8701	2.3674	2.3361	0.8451
	min	-1.4088	-9.2109	-1.0153x10 ¹	-1.0084x10 ¹	-9.1107	-5.0406
	max	-0.4575	-0.5200	-0.5090	-0.8561	-0.7072	-1.2512
	mean rank	6	1	5	3	2	4
22	mean	-0.8586	-3.0051	-3.1134	-2.8936	-3.4404	-2.5657
	st dev	0.3301	1.8833	3.6634	1.9308	2.1138	0.9282
	min	-1.5584	-8.5430	-1.0403x10 ¹	-8.5422	-8.3691	-5.0053
	max	-0.3876	-0.8972	-0.6342	-0.8590	-0.9487	-1.5357
	mean rank	6	3	2	4	1	5
23	mean	-1.2082	-3.9409	-2.5317	-3.6734	-3.1585	-2.3391
	st dev	0.4895	2.8005	2.2583	1.8173	1.4812	0.80007
	min	-2.4122	-1.0195x10 ¹	-1.0536x10 ¹	-7.6571	-7.2028	-4.3752
	max	-0.5738	-0.8049	-0.7951	-1.4377	-1.0123	-1.2768
	mean rank	6	1	4	2	3	5

TABLE VI. CLUSTER BASED SUPERIORITY OF ABO COMPARED TO OTHER METAHEURISTICS

Group	MPA [9]	GSO [10]	SMA [11]	NGO [25]	ZOA [26]
1	6	7	6	5	0
2	6	5	5	4	2
3	9	6	5	2	6
Total	21	18	16	11	8

Table VI summarizes the performance comparison between ABO and the five other metaheuristics. The comparison represents the superiority of ABO compared to other metaheuristics based on the group of functions. This group-based comparison is needed because of the distinct characteristics among the groups of functions. It is assumed that some metaheuristics may be better in a certain group but mediocre in another group. The last row represents the total number of functions where ABO outperforms a metaheuristic.

Table VI indicates that ABO is superior to MPA, GSO, and SMA and still competitive compared with NGO and ZOA. ABO is better than MPA, GSO, SMA, NGO, and ZOA in solving 21, 18, 16, 11, and 8 functions, respectively. Meanwhile, ABO equals NGO in solving two functions and ZOA in solving six functions. It means ABO is worse than NGO in solving ten functions and ZOA in solving nine functions. ABO is superior to MPA in all groups of functions. Meanwhile, ABO is superior to GSO and SMA in solving unimodal and multimodal functions. On the other hand, ABO is still competitive compared to GSO and SMA in solving fixed-dimension multimodal functions. ABO is superior to NGO in solving unimodal and high-dimension multimodal functions but inferior in solving fixed-dimension multimodal functions. Contrary, ABO is superior to ZOA in solving fixed-dimension multimodal functions but inferior in solving high-dimension functions. Based on this summary, ABO has fierce competition with NGO and ZOA, where ABO is better than NGO in high-dimension functions while ABO is better than ZOA in fixed-dimension functions.

The second part of the simulation is regarding the hyper-parameter evaluation. In this work, two parameters are observed. The first parameter is the population size, while the second is the maximum iteration. The set of 23 functions is still used in this evaluation. The result of the population size evaluation is presented in Table VII, while the result of the maximum iteration is presented in Table VIII.

Table VII indicates that the increase in population size does not improve the quality of the solution in almost all functions. The improvement takes place only in three functions. Among these three functions, two functions are high dimension unimodal functions, and one function is a fixed dimension multimodal function. Meanwhile, there are two reasons why the improvement fails. The first reason is that the global optimal or quasi-optimal solution has been achieved. This reason can be found in twelve functions. The second reason is that ABO fails to find the quasi-optimal solution after reaching the maximum iteration. This reason can be found in eight functions.

TABLE VII. RELATION BETWEEN POPULATION SIZE AND AVERAGE FITNESS SCORE

Function	Average Fitness Score		Significantly Improved?
	$n(X) = 10$	$n(X) = 40$	
1	0.0000	0.0000	no
2	0.0000	0.0000	no
3	4.2181×10^1	0.0030	yes
4	0.0000	0.0000	no
5	4.8935×10^1	4.8880×10^1	no
6	1.0556×10^1	9.8178	no
7	0.0187	0.0056	yes
8	-4.2131×10^3	-4.7692×10^3	no
9	0.0000	0.0000	no
10	0.0000	0.0000	no
11	0.0000	0.0000	no
12	0.9883	0.7949	no
13	3.1324	3.1151	no
14	2.8768	1.5147	no
15	0.0043	0.0012	yes
16	-1.0258	-1.0290	no
17	0.4353	0.4057	no
18	3.5492	3.1038	no
19	-0.0495	-0.0495	no
20	-2.6985	-2.9387	no
21	-2.5507	-3.5834	no
22	-2.4601	-3.3685	no
23	-2.9200	-3.3965	no

TABLE VIII. RELATION BETWEEN MAXIMUM ITERATION AND AVERAGE FITNESS SCORE

Function	Average Fitness Score		Significantly Improved?
	$t = 60$	$t = 120$	
1	0.0000	0.0000	no
2	0.0000	0.0000	no
3	1.9130×10^1	0.0638	yes
4	0.0000	0.0000	no
5	4.8975×10^1	4.8939×10^1	no
6	1.0700×10^1	1.0506×10^1	no
7	0.0253	0.0164	no
8	-4.0184×10^3	-4.2073×10^3	no
9	0.0000	0.0000	no
10	0.0000	0.0000	no
11	0.0000	0.0000	no
12	1.0667	0.9978	no
13	3.1436	3.1255	no
14	4.2826	2.6412	no
15	0.0043	0.0034	no
16	-1.0224	-1.0248	no
17	0.4671	0.4316	no
18	3.6447	3.2859	no
19	-0.0495	-0.0495	no
20	-2.7096	-2.7824	no
21	-2.7736	-2.5821	no
22	-2.1184	-2.9625	no
23	-2.2307	-3.0198	no

Table VIII indicates that the increase of the maximum iteration does not improve the quality of the solution in almost all functions. The improvement occurs only in one high-dimension unimodal function (Schwefel 1.2). There are twelve functions where the global optimal or quasi-optimal solution has been achieved in the low maximum iteration circumstance so that there is no improvement anymore due to the increase of maximum iteration.

The third part is the evaluation of ABO in solving a practical optimization problem: portfolio optimization. This problem is chosen because most works that introduced new metaheuristics chose problems in mechanical engineering or power flow distribution as their use case. The portfolio optimization problem is an important optimization work in the finance sector. In general, two objectives can be chosen for the portfolio optimization problem. The portfolio optimization problem can be defined as an effort to arrange financial assets (stock, bond, gold, and so on) to maximize the return or control the risk [36].

In this work, the assets are stocks of the energy or mining companies listed in IDX30. IDX30 is a list published by the Indonesian stock exchange that consists of 30 very liquid stocks with high market capitalization and strong fundamentals. There are ten stocks, and the list is presented in Table IX. These stocks have three important attributes: stock index, market price, and year-on-year capital gain. The market price and capital gain are presented in rupiah per share. The information regarding these stocks was obtained on February 22, 2023.

This portfolio optimization is taken based on some scenario. The objective is maximizing the total capital gain which is calculated by accumulating the capital gain of all shares that are held. The quantity of each stock ranges from 200 lots to 1,000 lots where each lot represents 100 shares. The maximum total investment is five billion rupiah. This problem can be seen as a unimodal problem where the dimension is 10.

In this portfolio optimization, ABO is also competed with five metaheuristics like in the first part: MPA, GSO, SMA, NGO, and ZOA. The population size is 10 where the maximum iteration is 30. The result is presented in Table X.

TABLE IX. STOCK INFORMATION

No	Stock Index	Price	Capital Gain
1	ADRO	2,850	530
2	ANTM	2,050	-150
3	BRPT	910	-55
4	ESSA	940	295
5	INCO	6,800	1,860
6	ITMG	35,575	11,600
7	MDKA	4,610	853
8	MEDC	1,060	500
9	PGAS	1,540	135
10	PTBA	3,540	530

Table X indicates that ABO is competitive in solving this portfolio optimization problem, although it is not the best performer. ABO becomes the third best after SMA and NGO. On the other hand, ZOA becomes the worst metaheuristic in solving this portfolio optimization problem, although it is very competitive in solving the set of 23 functions.

TABLE X. PORTFOLIO OPTIMIZATION RESULT

No	Metaheuristic	Total Capital Gain
1	ABO	1,461,870,054
2	ZOA [26]	1,353,039,265
3	NGO [25]	1,478,676,650
4	SMA [11]	1,479,375,445
5	GSO [10]	1,461,377,050
6	MPA [9]	1,377,967,381

V. DISCUSSION

The simulation result shows that ABO is competitive enough as a swarm-based metaheuristic. ABO can find an acceptable solution in both theoretical and practical optimization problems. ABO can find the optimal global solution in solving five functions. ABO is superior to MPA, SMA, and GSO and competitive to NGO and ZOA in solving the set of 23 functions. Meanwhile, ABO becomes the third best in solving the portfolio optimization problem.

Solving the theoretical optimization problem shows that a strict-acceptance approach is important to achieve good performance, especially for high-dimension functions. In these functions, ZOA [26] and NGO [25] are metaheuristics that implement a strict-acceptance approach, while ABO implements both strict-acceptance and non-strict-acceptance approaches. ABO, ZOA, and NGO are the best of the three in solving almost all functions in the big dimension problems, while GSO [10], MPA [9], and SMA [11] do not implement a strict-acceptance approach. This circumstance indicates that metaheuristics implementing a strict-acceptance approach significantly achieves better results than others. On the other hand, circumstance becomes more dynamic in solving fixed-dimension multimodal functions where the gap among metaheuristics is narrow whether these metaheuristics implement a strict-acceptance approach. It means avoiding a worse solution is important in solving high-dimension functions, while this strategy is not important in solving fixed-dimension multimodal functions.

The strict-acceptance approach also does not significantly affect solving the portfolio optimization problem. Table X shows the narrow gap between the best and worst metaheuristics. On the other hand, ZOA as the worst performer is a metaheuristic that adopts a strict-acceptance approach. On the contrary, MPA, as a metaheuristic that does not adopt a strict acceptance approach, becomes the second worst performer. As a metaheuristic that does not adopt a strict-acceptance approach, SMA becomes the best performer. Meanwhile, NGO becomes the second-best performer as a metaheuristic that adopts a strict-acceptance approach.

The simulation result also shows that the competition among latest metaheuristics becomes tougher. It is common for many latest metaheuristics to deploy multiple searches and enrich the guided search with the random search. Many metaheuristics can find the global optimal in several functions and quasi-optimal in many other functions. Meanwhile, many metaheuristics still need help finding the quasi-optimal solution, especially in the low maximum iteration and low population size. This circumstance strengthens the no-free-lunch theory that a metaheuristic cannot solve all problems with superior results.

There are three loops conceptually performed during the iteration. The outer loop is the iteration from the first iteration to the maximum iteration. The intermediate loop is the iteration for the entire population. The inner loop is the iteration for the entire dimension because all dimensions are calculated in every search. Meanwhile, there are two searches performed by every agent in every iterations. Based on this explanation, the algorithm complexity of ABO is presented as $O(2t_{max} \cdot n(X) \cdot n(D))$. This complexity is normal for the population-based metaheuristic. Moreover, this complexity is achieved because ABO does not implement a sorting process in every iteration.

There are limitations regarding this work and its proposed metaheuristics, even though the proposed ABO performs well in solving both theoretical and practical optimization works. In ABO, the adaptability in tackling the local optimal is performed by choosing a non-strict acceptance approach. Meanwhile, this non-strict acceptance approach is implemented in any iteration. It differs from simulated annealing, where accepting a worse solution becomes more difficult as iteration increases. Meanwhile, different metaheuristic, such as tabu search, uses tabu list to restrict the repetition of a similar solution. This circumstance shows that there are various adaptive approaches that can be explored in the future. At the same time, a single metaheuristic such as ABO cannot adopt various adaptive strategies into a single metaheuristic.

This work has presented the use of optimization problem in financial sector, which is the portfolio optimization problem in the introduction of a new metaheuristic. This work also proves that ABO is competitive enough in solving this problem which is an integer-based problem. Meanwhile, there are various kinds of other optimization problems in the financial sector, such as credit risk assessment, investment planning, debtor analysis, and refinancing problems. These problems can also be addressed in future work.

There is also a limitation in choosing a practical optimization problem as a use case to evaluate the performance of the new metaheuristic. This work chooses a portfolio optimization problem as the use case, with its characteristics being integer-based and unimodal. On the other hand, there are various practical optimization problems, whether common or not, in many studies introducing new metaheuristics. These problems can be used for future studies, especially proposing an improved or modified version of ABO.

The future studies can also be performed by implementing ABO to solve various sustainable development goals (SDGs) related issues. SDG has become the global issue and

consideration for developing sustainable society and environment. For example, efficient energy consumption becomes the main and important issue related to climate change, renewable and affordable energy. Besides, optimization plays an important role in the operation of industry, transportation, and many other economic activities.

VI. CONCLUSION

The introduction of a new adaptive metaheuristic, namely adaptive balance optimizer (ABO), has been presented in this paper. This proposed model is designed to make a metaheuristic adaptive, especially when facing optimal local circumstances. This paper also presents the competitiveness of ABO in solving both theoretical and practical optimization problems. ABO is better than MPA, GSO, SMA, NGO, and ZOA in solving 21, 18, 16, 11, and 8 functions, respectively, in solving 23 functions. It means ABO is superior to MPA, GSO, and SMA and still competitive with NGO and ZOA in solving 23 functions. Meanwhile, ABO is still competitive in solving portfolio optimization problems, although ABO is not the best performer in solving this problem.

Adaptability can be used for future studies in metaheuristics. Various strategies have yet to be explored to make metaheuristics more adaptive, especially in tackling the local optimal entrapment. Besides, developing a superior metaheuristic that can solve the optimization problem in the low maximum iteration and low population size becomes challenging too. Moreover, future work can be conducted by addressing several common issues, such as scalability and more practical recent and future use cases. The scalability issue is related to wider boundaries and higher dimensions of the problem. Meanwhile, there are various recent and future optimization problems, such as in the green and blue economy, climate change, renewable energy, and many more.

ACKNOWLEDGMENT

This work was financially supported by Telkom University, Indonesia.

REFERENCES

- [1] L. Li, Y. Sun, and B. Wang, "A hybrid genetic algorithm for service caching and task offloading in edge-cloud computing", *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 11, pp. 761-765, 2022.
- [2] A. Damarla and D. Sumathi, "An approach for optimization of features using gorilla troop optimizer for classification of melanoma", *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 10, pp. 275-286, 2022.
- [3] Q. H. Zhai, T. Ye, M. X. Huang, S. L. Feng, and H. Li, "Whale optimization algorithm for multiconstraint second-order stochastic dominance portfolio optimization", *Computational Intelligence and Neuroscience*, vol. 2020, ID. 8834162, pp. 1-19, 2020.
- [4] Y. Yang, J. Tao, J. Zhou, Y. Wang, and X. Guo, "An improved sparrow search algorithm and its application in HIFU sound field", *Computational Intelligence and Neuroscience*, vol. 2023, ID. 1228685, pp. 1-19, 2023.
- [5] X. Wang, W. Wu, and Z. Xing, "Multi-decision points model to solve coupled-task scheduling problem with heterogeneous multiAGV in manufacturing systems", *International Journal of Industrial Engineering Computations*, vol. 14, no. 1, pp. 49-64, 2023.
- [6] H. R. Moshtaghi, A. T. Eshlaghy, and M. R. Motadel, "A comprehensive review on meta-heuristic algorithms and their

- classification with novel approach”, *Journal of Applied Research on Industrial Engineering*, vol. 8, no. 1, pp. 63-89, 2021.
- [7] J. Swan, S. Adriaensen, A. E. I. Brownlee, K. Hammond, C. G. Johnson, A. Kheiri, F. Krawiec, J. J. Merelo, L. L. Minku, E. Ozcan, G. L. Pappa, P. Garcia-Sanchez, K. Sorensen, S. Vob, M. Wagner, and D. R. White, “Metaheuristics in the large”, *European Journal of Operational Research*, Vol. 297, pp. 393-406, 2022.
- [8] S. Mirjalili, S. M. Mirjalili, and A. Lewis, “Grey wolf optimizer”, *Advances in Engineering Software*, vol. 69, pp. 46–61, 2014.
- [9] A. Faramarzi, M. Heidarinejad, S. Mirjalili, and A. H. Gandomi, “Marine predators algorithm: a nature-inspired metaheuristic”, *Expert Systems with Applications*, vol. 152, ID: 113377, 2020.
- [10] M. Noroozi, H. Mohammadi, E. Efatinasab, A. Lashgari, M. Eslami, and B. Khan, “Golden search optimization algorithm”, *IEEE Access*, vol. 10, pp. 37515–37532, 2022.
- [11] S. Li, H. Chen, M. Wang, A. A. Heidari, and S. Mirjalili, “Slime mould algorithm: a new method for stochastic optimization”, *Future Generation Computer Systems*, vol. 111, pp. 300-323, 2020.
- [12] M. Dehghani, Z. Montazeri, H. Givi, J. M. Guerrero, and G. Dhiman, “Darts game optimizer: a new optimization technique based on darts game”, *International Journal of Intelligent Engineering and Systems*, vol. 13, no. 5, pp. 286–294, 2020.
- [13] S. Arora and S. Singh, “Butterfly optimization algorithm: a novel approach for global optimization”, *Soft Computing*, vol. 23, no. 3, pp. 715–734, 2019.
- [14] M. S. Braik, “Chameleon swarm algorithm: a bio-inspired optimizer for solving engineering design problems”, *Expert Systems with Applications*, vol. 174, ID. 114685, pp. 1-25, 2021.
- [15] S. Kaur, L. K. Awasthi, A. L. Sangal, and G. Dhiman, “Tunicate swarm algorithm: a new bio-inspired based metaheuristic paradigm for global optimization”, *Engineering Applications of Artificial Intelligence*, vol. 90, ID. 103541, 2020.
- [16] M. Suman, V. P. Sakthivel, and P. D. Sathya, “Squirrel search optimizer: nature inspired metaheuristic strategy for solving disparate economic dispatch problems”, *International Journal of Intelligent Engineering and Systems*, vol. 13, no. 5, pp. 111–121, 2020.
- [17] A. M. Khalid, K. M. Hosny, and S. Mirjalili, “COVIDOA: a novel evolutionary optimization algorithm based on coronavirus disease replication lifecycle”, *Neural Computing and Applications*, vol. 34, pp. 22465-22492, 2022.
- [18] M. Braik, A. Hammouri, J. Atwan, M. A. Al-Betar, and M. A. Awadallah, “White shark optimizer: a novel bio-inspired meta-heuristic algorithm for global optimization problems”, *Knowledge-Based Systems*, vol. 243, ID. 108457, pp. 1-29, 2022.
- [19] A. Kaveh, S. Talatahari, and N. Khodadadi, “Stochastic paint optimizer: theory and application in civil engineering”, *Engineering with Computers*, vol. 38, pp. 1921-1952, 2022.
- [20] P. Trojovský and M. Dehghani, “Pelican optimization algorithm: a novel nature-inspired algorithm for engineering applications”, *Sensors*, Vol. 22, ID. 855, pp. 1-34, 2022.
- [21] P. D. Kusuma and A. L. Prasasti, “Guided pelican algorithm”, *International Journal of Intelligent Engineering and Systems*, vol. 15, no. 6, pp. 179-190, 2022.
- [22] P. D. Kusuma and A. Novianty, “Total interaction algorithm: a metaheuristic in which each agent interacts with all other agents”, *International Journal of Intelligent Engineering and Systems*, vol. 16, no. 1, pp. 224-234, 2023.
- [23] P. D. Kusuma and A. Dinimaharawati, “Three on three optimizer: a new metaheuristic with three guided searches and three random searches”, *International Journal of Advanced Science and Applications*, vol. 14, no. 1, pp. 420-429, 2023.
- [24] M. Dehghani, S. Hubalovsky, and P. Trojovský, “A new optimization algorithm based on average and subtraction of the best and worst members of the population for solving various optimization problems”, *PeerJ Computer Science*, vol. 8, ID: e910, pp. 1-29, 2022.
- [25] M. Dehghani, S. Hubalovsky, and P. Trojovský, “Northern goshawk optimization: a new swarm-based algorithm for solving optimization problems”, *IEEE Access*, vol. 9, pp. 162059–162080, 2021.
- [26] E. Trojovska, M. Dehghani, and P. Trojovský, “Zebra optimization algorithm: a new bio-inspired optimization algorithm for solving optimization algorithm”, *IEEE Access*, vol. 10, pp. 49445-49473, 2022.
- [27] M. Dehghani, Z. Montazeri, E. Trojovska, and P. Trojovský, “Coati optimization algorithm: a new bio-inspired metaheuristic algorithm for solving optimization problems”, *Knowledge-Based Systems*, vol. 259, ID. 110011, pp. 1-43, 2023.
- [28] E. Trojovska, M. Dehghani, and P. Trojovský, “Fennec fox optimization: a new nature-inspired optimization algorithm”, *IEEE Access*, vol. 10, pp. 84417-84443, 2022.
- [29] E. Trojovska and M. Dehghani, “A new human-based metaheuristic optimization method based on mimicking cooking training”, *Scientific Reports*, vol. 12, ID. 14861, pp. 1-24, 2022.
- [30] S. A. Yasear and H. M. A. Ghanimi, “A modified honey badger algorithm for solving optimal power flow optimization problem”, *International Journal of Intelligent Engineering and Systems*, vol. 15, no. 4, pp. 142–155, 2022.
- [31] M. I. A. Latiffi, M. R. Yaakub, and I. S. Ahmad, “Flower pollination algorithm for feature selection in tweets sentiment analysis”, *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 5, pp. 429-436, 2022.
- [32] M. Dehghani, M. Mardaneh, J. S. Guerrero, O. P. Malik, and V. Kumar, “Football game based optimization: an application to solve energy commitment problem”, *International Journal of Intelligent Engineering and Systems*, vol. 13, no. 5, pp. 514-523, 2020.
- [33] D. Polap and M. Wozniak, “Red fox optimization algorithm”, *Expert Systems with Applications*, vol. 166, ID. 114107, pp. 1-21, 2021.
- [34] S. Suyanto, A. A. Ariyanto, and A. F. Ariyanto, “Komodo mlipir algorithm”, *Applied Soft Computing*, vol. 114, pp. 1–17, 2022.
- [35] A. Banharnsakun, “A new approach for solving the minimum vertex cover problem using artificial bee colony”, *Decision Analytics Journal*, vol. 6, ID. 100175, pp. 1-6, 2023.
- [36] Y. Chen, X. Zhao, and J. Yuan, “Swarm intelligence algorithms for portfolio optimization problems: overview and recent advances”, *Mobile Information Systems*, vol. 2022, ID. 4241049, pp. 1-15, 2022.

Improved Speaker Recognition for Degraded Human Voice using Modified-MFCC and LPC with CNN

Amit Moondra¹

Researcher

Department of Computer Science Engineering, Manav
Rachna International Institute of Research and Studies,
Faridabad, Haryana, India

Dr Poonam Chahal²

Professor

Department of Computer Science Engineering, Manav
Rachna International Institute of Research and
Studies, Faridabad, Haryana, India

Abstract—Economical speaker recognition solution from degraded human voice signal is still a challenge. This article is covering results of an experiment which targets to improve feature extraction method for effective speaker identification from degraded human audio signal with the help of data science. Every speaker's audio has identical characteristics. Human ears can easily identify these different audio characteristics and classify speaker from speaker's audio. Mel-Frequency Cepstral Coefficient (MFCC) supports to get same intelligence in machine also. MFCC is extensively used for human voice feature extraction. In our experiment we have effectively used MFCC and Linear Predictive Coding (LPC) for better speaker recognition accuracy. MFCC first outlines frames and then finds cepstral coefficient for each frame. MFCC use human audio signal and convert it in numerical value of audio features, which is used to recognize speaker efficiently by Artificial Intelligence (AI) based speaker recognition system. This article covers how effectively audio features can be extracted from degraded human voice signal. In our experiment we have observed improved Equal Error Rate (EER) and True Match Rate (TMR) due to high sampling rate and low frequency range for mel-scale triangular filter. This article also covers pre-emphasis effects on speaker recognition when high background noise comes with audio signal.

Keywords—Data science; artificial intelligence; MFCC; LPC; CNN; mel-spectrum; speaker recognition

I. INTRODUCTION

In the current world, voice communication is used to exchange thoughts and feelings. Most of the business are on voice communication and build trust over voice phone calls. Humans can easily identify a person from his/her voice over a phone call but in modern world voice answer machine should also need to identify person by his/her voice. Now-a-days lot of medical issues are also diagnosed through human voice. Human generates audio from throat and mouth. Fundamental frequency is original frequency, and it is modulated by vocal code structure and that makes every human voice as unique voice. Human can be identified based on his or her voice.

Human voice is generally identified in two categories, male voice, and female voice. The main difference between male and female voice is pitch and frequency. So, if it is just needed to identify gender from voice, then it's easy to classify by pitch difference. But, to recognize a human from his/her voice then all voice features are to be considered. Every person's voice is

unique and it's based on different voice features like frequency, jitter, amplitude, pitch and spectral power. Male speaker has 0-900 hz as fundamental frequency and female speaker has 0-1500Hz as fundamental frequency. If we consider average frequency, then male average fundamental frequency is 110 hz and female average fundamental frequency is 211 Hz [1].

Vocal cord, teeth, jaw, and tongue are main articulators of vocal trach which modulate fundamental frequency (FO) [2][3]. To begin with the generation of sound, air pressure is produced by the human lungs. This air pressure generates sound with fundamental frequency. Fundamental frequency sound is modulated by the vocal track to create different sound variations.

In Speaker recognition process, first step to identify human voice feature. MFCC is generally use for voice feature extraction from human voice. MFCC mainly consider frequency changes in human voice to define cepstral coefficients. There are various research earlier which have used MFCC for voice feature extraction. N. V. Tahliramani and N. Bhatt [4] discussed that machine can percept human feature as per MFCC in training phase. Similarly Convolutional Neural Network (CNN) is very popular for speaker recognition process [5]. N. Gupta and S. Jain [6] discussed about Siamese network effects with CNN. A. Chowdhury and A. Ross [7] discussed about CNN with degraded human voice. CNN is one of the preferred model with MFCC and Linear Predictive Coding (LPC) when signal is degraded with high noise [7].

MFCC and LPC are used in most of speaker recognition system. In our experiment, we have modified MFCC and evaluate results. In this article our focus to explain frequency impact on feature extraction and overall improvement in speaker recognition system.

The paper is organized as follows: base model for speaker recognition as discussed in Section II which defines all necessary steps for speaker recognition system, Section III discussed about what are different types of noises and how these noises degrade human voice. Then we discussed about how we prepared dataset with different noises in Section IV, and then in Section V we have defined CNN model including detailed description of modified MFCC, LPC and loss function (cosine triplet), then we have discussed experiment results in Section VI.

II. SPEAKER RECOGNITION BASE MODEL

Speaker Recognition (SR) process has three basic steps to identify speaker. These steps are stich together as per below Fig. 1.

- 1) Preprocessing
- 2) Voice Feature Extraction
- 3) Classification

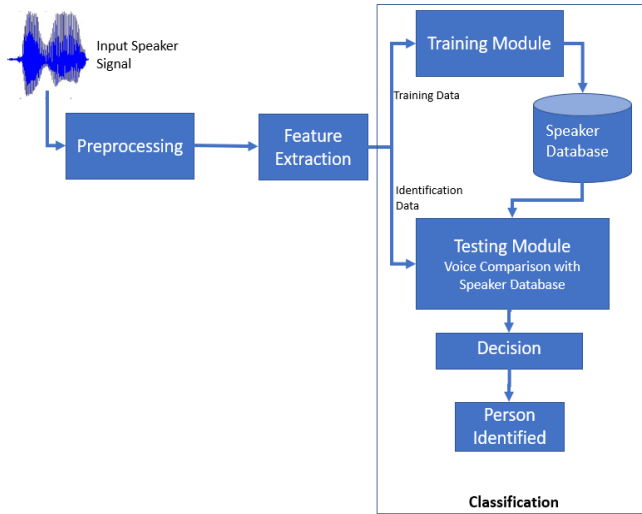


Fig. 1. Speaker recognition process blocks.

Preprocessing is mainly used for cleaning up data. In speaker recognition process preprocessing to remove possible noises from speaker recorded voice. Different types of “Low Pass Filter” (LPF), “Band Stop Filter” (BSF) and “High Pass Filter” (HPF) are used to remove noise from voice sample. Butterworth filter is also used as signal processing filter which is also known as band-stop filter.

Voice feature extraction is used to extract feature from speaker’s audio file. These features are used to classify speaker in next step. Voice features can be extracted by one of the most popular techniques Mel-Frequency Cepstral Coefficient (MFCC) [4][7][8][9][10], Linear Predictive Coding (LPC) [7] is also used for feature extraction. MFCC has its own benefit for feature extraction. For feature extraction, MFCC uses following main steps [4][11][12] to identify cepstral coefficients. MFCC converts voice recording into cepstral coefficient in matrix form which is easy to feed in next classification step. Following steps may be used in combination with LPC or other feature extraction methods also, based on application and source speaker voice.

- Framing & Blocking
- Windowing
- Fast Fourier Transform (FFT)
- Triangular Bandpass Filter (Mel-scale)
- Inverse FFT

These steps are stich together as per below Fig. 2.

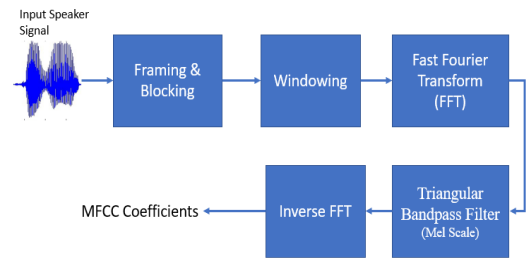


Fig. 2. MFCC processing.

A. Framing

Framing is required to make consistent voice sample. Long voice signal has lot of amplitude and frequency variation which gives lot of inconsistency during voice signal analysis for feature extraction. Short frames solve this issue and generally we create 20 to 40ms short frames for better voice signal analysis [13]. If T is duration of complete voice sample and t is sub frame sample duration, then total number of frames are F.

$$F = \frac{T}{t} * 1000$$

B. Windowing

Windowing is used to increase steadiness between first and end point within the same frame. From framing we get short duration frames, but these frames are generally discontinued because for interworld silence. Humming window makes continuity within same frame. If xi(n) is time domain signal and hi(n) is hamming window, then signal will become with hamming window is

$$xi(n) \text{ after windowing} = xi(n) * hi(n)$$

Where hi(n)

$$hi(n, \alpha) = (1 - \alpha) - \alpha \cos\left(\frac{2\pi n}{N} - 1\right)$$

$$1 \leq n \leq N - 1$$

C. Fast Fourier Transform

When speaker’s voice is recorded by any microphone, that signal is recorded in time domain. Time domain gives limited information about human voice feature. “Fast Fourier Transform” (FFT) converts time domain signal into frequency domain signal. Computation effect of FFT and “Discrete Fourier Transform” (DFT) are equal but Fast Fourier Transform is actually a fast algorithm to conceptualize DFT. DFT is discrete version of FFT. If N represent number of samples in frame, then each sample of that frame needs to be converted in frequency domain. If xi(n) is voice time domain framed signal and Xi(k) is transformed signal as define in below in equation.

$$Xi(k) = xi(n) hi(n) e^{-\frac{j2\pi kn}{N}}$$

$$1 \leq n \leq N \text{ and } 1 \leq k \leq K$$

Where hamming window is h(n), K is length of DFT.

D. Triangular Bandpass Filter: Mel-Scale

Humans perceive different human voice differently and similar intelligence is required in machines also. When human

time domain voice signal converted to frequency domain signal the triangular band-pass filters are used to filter human voice signal. FFT transform time-domain signal into frequency-domain signal and frequency domain signal is passed through from triangular bandpass filter.

At lower frequency range group, humans can understand two separate frequency voice, but at high frequency range group, two separate frequency voice is not recognized by humans. For example, low frequency group 'A' has F1 and F2 frequency voice signal. F1 and F2 has d difference then human can recognize F1 and F2 correctly. Similarly, high frequency group 'B' has F3 and F4 frequency voice signal. F3 and F4 also has d frequency difference then humans face difficulties to recognize it. To add similar intelligence in speaker recognition system, mel-scale is used to convert frequency into mel-scale as per given formula. If mf represents frequency in mel-scale and f in hz then

$$mf = 2595 * \log_{10} \frac{f}{700} + 1$$

The purpose of mel-scale conversion is to change the signals to follow the decrease properties of the mel-scale. At lower frequency it is linear conversion but at high frequency it is more static as per below Fig. 3.

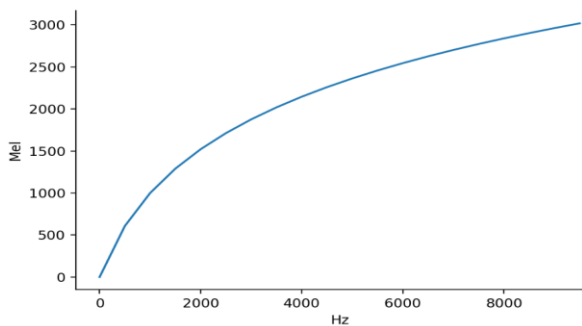


Fig. 3. Mel-scale.

Mel-scale is used to identify the space of the triangular filter bank and also to identify how much filter bank width should be maintained. At the high frequency filter band width should also get wider to maintain frequency variability. Mel-scale based on triangular filter bank makes easy calculation of energies. Once energy is calculated with the help of mel-scale based triangular filter then it's easy to calculate mel-spectrum. Mel-spectrum is used to calculate different cepstral coefficients (generally 20) with the help of Discrete Fourier Transform (DCT).

E. Inverse FFT

Inverse FFT is used to convert back mel-spectrum into to the spatial domain. Initially time domain signal is converted to frequency domain signal and then frequency domain signal is transformed in mel-scale. Then it's necessary to convert back mel-scale signal. When inverse FFT is applied then signal gets converted back to spatial domain. This transformation can be performed either by "Discrete Fourier Transform" (DFT) or Discrete Cousins Transform (DCT). Both DFT and DCT can be used to identify coefficients as DFT and DCT divides a given sequence of finite length data into discrete vector. In

general, DCT is more used to identify coefficient from the specified log mel-spectrum [9]. DCT provides coefficients as output and these coefficients are known as MFCC. If m coefficient is required to calculate (generally 20 or 13 coefficients calculate) and Cn represent MFCC coefficient [9].

$$Cn = \sum_{i=1}^i \log_{10} Xi \cos[m(i - \frac{1}{2}) \pi / i]$$

Where Xi is FFT of the voice signal and m = 0, 1, ... i-1

When an audio signal is converted in matrix data set with the help of MFCC output then this MFCC output works as a input for classifier model. There are multiple models used by different researchers for speaker recognition like "Gaussian Mixture Model" (GMM) [14][15][16][17], "Hidden Markov Model" (HMM) [18], "Support Vector Machine" (SVM) [19][20][21][22], and deep learning based "Convolutional Neural Network" (CNN) [5][6][7][23]. In this article we have used CNN as classification model for speaker recognition.

III. NOISE EFFECTS ON SPEAKER RECOGNITION

Speaker Recognition efficiency is majorly dependent on the speaker's audio quality. If no background noise is included with speaker's voice, then speaker recognition results are quite good as compared to high background noise. When background noise is high then, it degrades the overall human voice signal. Degraded human voice signal is not only difficult to recognize but it also leads to a complicated speaker recognition system. There are multiple methods to eliminate noise from human voice. Some of the articles [1][24][25][26] have used frequency-based filtration logic to eliminate noise.

In speaker recognition process, speaker's voice should be noise free for better recognition. In today's world absolute silence is not possible in public places and lot of background noises are also captured in speaker voice. There are multiple use cases which use speaker recognition as authentication mechanism to identify customer. Background noise coming before and after speaker's voice can be easily eliminated in speaker recognition process but to eliminate background noise coming between two words is a major task in speaker recognition [27]. Type of noise is also important to consider in speaker recognition process.

A. Different Types of Noise

There are different types of noise classifications like colour based noise, white noise, violet noise etc. Noise can also be classified as

1) *Machine based noise*: This type of noise is generated by a machine. When a machine runs then that machine has its own noise like car, bus, large scale process industries machine, train, airplane have their own noises [28]. Machines also contribute to noise generation indirectly, like generation of noise from a home fan due to air friction.

2) *Human generate noise*: When many humans speak at same time with different words then their combined voice resembles as noise. For example, noise in restaurant, in stadium, shopping mall or in street [28].

3) *Nature inspired sound*: When sound is not generated by machine or human and is generated by natural resources then

the noise is considered as nature inspired sound like animal or bird sound, waterfall sound, fire sound, and wind sound. In speaker recognition process, whichever sound comes from the background making disturbance to classify speaker is considered as background noise. When speaker voice comes with waterfall sound or dog bark as a background then it disturbs with human voice signal and makes the speaker recognition process complicated.

B. Signal to Noise Ratio (SNR)

SNR signifies ratio of voice signal power to the noise power. SNR is indirectly proportional to the noise signal power. SNR is low when high noise signal power. If P_s is voice signal power and P_n represent noise signal (background) then SNR is

$$SNR \propto \frac{1}{P_n}$$

$$SNR = \frac{P_s}{P_n}$$

Human voice signal comes from throat and modulated with vocal cord, jaw, teeth and tongue. Human voice is always in changing pattern and it's never constant. As human voice is never constant and dynamic in nature so it's not idle for voice feature analysis. Hence, to regularize this it's required to express voice signal power in logarithmic scale as:

$$P_s(dB) = 10\log_{10} P_s$$

Similarly, noise signal power can also be described in logarithmic scale as

$$P_n(dB) = 10\log_{10} P_n$$

SNR can also be expressed in logarithmic scale. In this article we are considering SNR in decibel(dB).

$$SNR(dB) = 10\log_{10} \left(\frac{P_s}{P_n}\right)$$

OR

$$SNR(dB) = P_s(dB) - P_n(dB)$$

P_n represent noise power (background noise power in SR process) and P_s represent voice signal power without noise. If two separate signals come separately then it's easy to calculate SNR as per above equation. In speaker voice if human voice signal and background noise signal come together then it is hard to separate both signal power. Below equation can be used when both signals come together. When P_x represent human voice signal power with noise [29] and P_n represent noise signal power then SNR is

$$SNR(dB) = 10\log_{10} \left(\frac{P_x - P_n}{P_n}\right)$$

SNR can be positive in value or negative. It's based on voice signal power and noise signal power. When human voice signal is recorded in noisy environment where there is high background noise then there is possibility to have negative SNR. When human voice signal power is greater than noise signal power then SNR will be positive. When noise signal power is greater than human voice signal power then SNR will be negative [29]. For example, when human voice signal is

recorded with continuous high background noise like in mechanical factory or in heavy traffic area then noise power is more than human voice signal power. In this case SNR will be negative. But when background noise is impulsive and only one or two spick comes with human voice signal then SNR will be positive but low in value.

Positive SNR (in dB) when

$$P_s > P_n$$

Negative SNR (in dB) when

$$P_n > P_s$$

IV. DATASET

In our experiment we have created following six different types of datasets with the help of TIMIT dataset and NOISEX-92 noise dataset.

A. TIMIT Dataset

TIMIT dataset provided 630 speakers' voice without any added noise. It is a mixed dataset which contains male and female human voice for ~3sec each. Out of 630 speakers, 462 speakers' voice is in training dataset and 168 speakers' voice are in testing dataset.

B. NOISEX-92 Dataset

This database contains eight different types of noises and we have considered following four noises in our experiment.

- Babble Noise: This noise is generated by many people. For example, many human continuous voices in restaurants or in public place.
- F16 Noise: This noise is generated by F16 fighter aircraft engine.
- Factory Noise: This noise is generated by heavy machines in some process/mechanical factory.
- Car Noise: This noise is generated by car engine.

C. Datasets used in Experiment

a) *Data Set-1 (DS1)*: In this data set we used 630 speakers' voice from TIMIT dataset and babble noise which is generated by many humans in a close room like restaurant or common public place and noise generated by F16 fighter aircraft engine from NOISEX-92 dataset. Below Fig. 4 shows power spectrogram of speaker voice with babble and F16 noise.

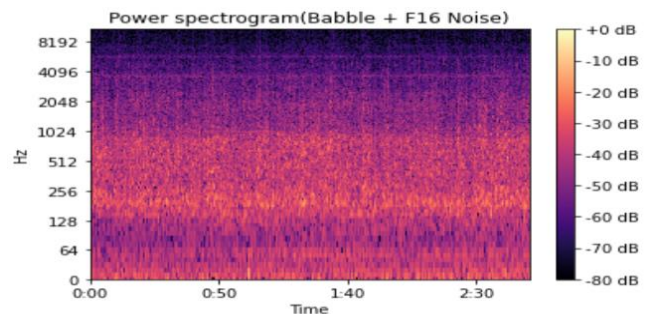


Fig. 4. Power spectrogram (babble + F16 noise).

b) *Data Set-2 (DS2)*: In this data set we used 630 speakers' voice from TIMIT dataset, car engine noise and heavy machinery noise which is generated inside the factory from NOISEX-92 dataset. Fig. 5 shows power spectrogram of speaker voice with car and factory noise.

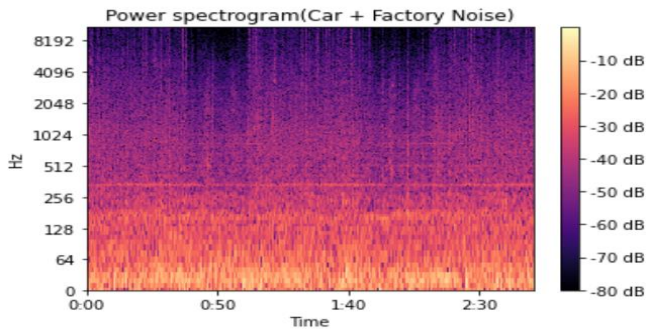


Fig. 5. Power spectrogram (car + factory noise).

c) *Data Set-3 (DS3)*: In this data set we used 630 speakers' voice from TIMIT dataset, babble and car engine noise from NOISEX-92 dataset. Fig. 6 shows power spectrogram of speaker voice with babble and car noise.

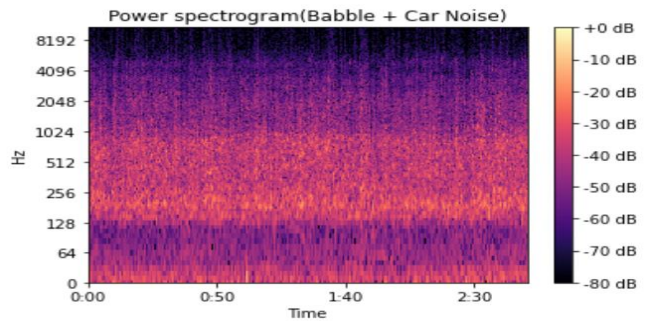


Fig. 6. Power spectrogram (babble + car noise).

d) *Data Set-4 (DS4)*: In this data set we used 630 speakers' voice from TIMIT dataset, F16 & factory noise from NOISEX-92 dataset. Fig. 7 shows power spectrogram of speaker voice with F16 and factory noise.

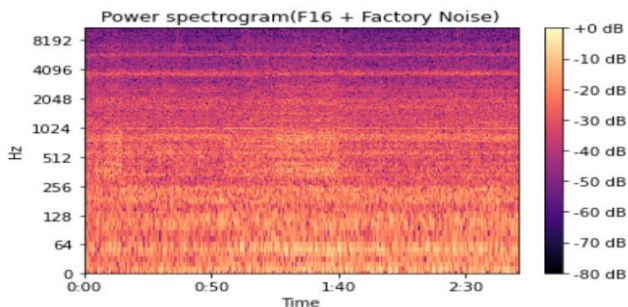


Fig. 7. Power spectrogram (F16 + factory noise).

e) *Data Set-5 (DS5)*: In this data set we used 630 speakers' voice from TIMIT dataset, car and F16 noise from NOISEX-92 dataset. Fig. 8 shows power spectrogram of speaker voice with car and F16 noise.

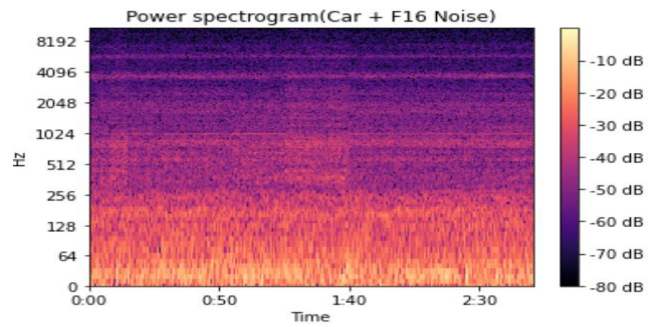


Fig. 8. Power spectrogram (Car + F16 noise).

f) *Data Set-6 (DS6)*: In this data set we used 630 speakers' voice from TIMIT dataset with babble and factory noise from NOISEX-92 dataset. Fig. 9 shows power spectrogram of speaker voice with babble and factory noise.

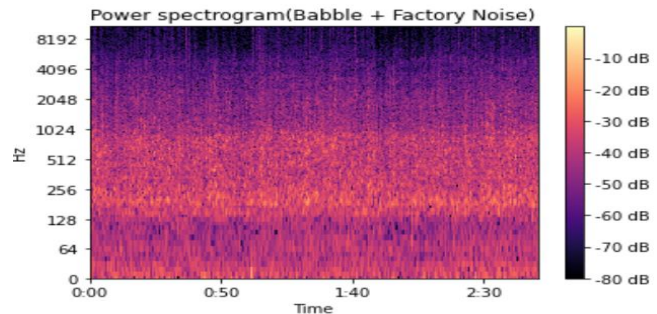


Fig. 9. Power spectrogram (Babble + factory noise).

In all datasets we have set room size as 3m and added reverberation. Below Table I represents summary of all 6-datasets.

TABLE I. DIFFERENT SIX DATASET SUMMARY

Dataset	Base Voice	Added Noise 1	Added Noise 2
DS1	TIMIT 630 Speakers	Babble	F16
DS2	TIMIT 630 Speakers	Car	Factory
DS3	TIMIT 630 Speakers	Babble	Car
DS4	TIMIT 630 Speakers	F16	Factory
DS5	TIMIT 630 Speakers	Car	F16
DS6	TIMIT 630 Speakers	Babble	Factory

V. PROPOSED CNN MODEL

Any speaker recognition system has two base modules, one is feature extraction module and another is classification module based on extracted feature.

A. Training Phase

In training phase we have used CNN for classification of speaker based on 1D-Triplet-CNN model defined in [7]. Convolutional layers play a key role in a CNN to determine its competency and learning capability. Each convolutional layer deeply learns different "concepts" from the data and hands it over to next convolution layer for further deeper learning in the CNN [30][31]. Speech can be represented in form of two dimensions (MFCC & LPC) but these two dimensions do not

show similarity. Generally, speech signals changes constantly but in short interval of voice it is constant in nature. This short interval of audio frame is called feature frame and it is good for speaker recognition process as it holds independent voice property. These multiple feature frames are useful in CNN for correct speaker recognition. We proposed CNN model with Modified MFCC (M-MFCC) and LPC for feature extraction. Three sets of CNN layers used in training phase as defined in Fig. 10.

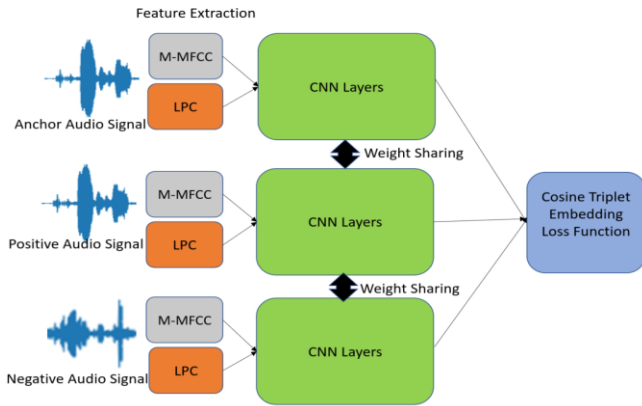


Fig. 10. CNN model training phase.

In training phase, we have used three CNN layers for positive, anchor and negative voice sample. All three CNN layers share weight to learn effectively. Triplet CNN layers are used to learn embedding function $f(x)$.

$$f(x) \in \mathcal{R}^d$$

$f(x)$ represents data sample embedding in d -dimension Euclidean space and x represents data sample. Following are the steps of overall speaker recognition classification model. We have used modified MFCC(M-MFCC) and LPC for feature extraction and then we have combine M-MFCC and LPC output and feed to CNN layers. CNN layers detailed we have defined in sub section D.

B. M-MFCC

We have modified MFCC in three basic steps to make high similarity in anchor and positive sample and dissimilarity in anchor and negative sample.

Step-1: High Sampling Rate

Sampling rate is the key factor to identify voice feature. High sampling rate captures more signal dissimilarity info as compared to low sampling rate. We have increased sampling rate from 22050 to 44100.

Step-2: Frequency Range for mel-scale triangular filter bank

As per Section II, MFCC uses mel-scale triangular bandpass filter. These filters use low and high frequency range to create triangular filter bank. Male speaker has 0-900 Hz as fundamental frequency and female speaker has 0-1500Hz as fundamental frequency. If we consider average frequency, then male average fundamental frequency is 110 Hz and female average fundamental frequency is 211 Hz [1][32]. We observed that most of the male and female voice power comes

up to 2000Hz or less. In this work we have selected the filter bank from 0Hz to 1800/2000Hz then it mostly captures human voice relevant information for speaker recognition process and avoids noise information which in between the words. Optimization of frequency range for triangular bandpass filter gives benefit to filter out background noise and improve mel-scale for small variation.

Step-3: Pre-emphasis Effects

Humans generate sound with fundamental frequency. Vocal track modulates this voice and generates modulated voice. This modulated voice is suppressed by high frequency voice. The objective of pre-emphasis is to compensate high-frequency part that was suppressed during the sound generation by the humans [33][34]. In general, when MFCC coefficient is calculated by different libraries like “python_speech_features” then standard value is 0.97 and when we used same pre-emphasis values then we did not gain much in speaker recognition accuracy. In our experiment, when we optimize pre-emphasis value and make it at 1.00 then we received improvement in results.

C. Linear Predictive Coding

As discussed in [7], when we need to simulate human voice then human throat functioning needs to be understood. Human voice is generated from lungs and filtered at vocal track. Vocal track is nothing but a time varying digital bandpass filter which filters some frequencies. Also, for detail estimation and analysis of human vocal track we generally look all-pole model of filter design. If we consider $H(z)$ as transfer function of vocal track then,

$$H(z) = \frac{G}{1 - \sum_{k=1}^p \alpha_k z^{-k}}$$

Human voice signal is continuous signal so if $S[n]$ is voice acoustics of n th sample and p is past voice sample then n th speech sample can be defined as

$$S[n] = \sum_{k=1}^p \alpha_k S[n-k] + G \cdot u[n]$$

G is gain factor here and $S[n-k]$, $k=1, 2, \dots, p$ are past human voice samples. Excitation of n th voice sample is represented as $u[n]$. α_k represents vocal track filter coefficients. As LPC also aims to identify all zero filters which come from invers vocal track model. If $S'[n]$ represents n th speech sample with conditioned on previous speech sample, then

$$S'[n] = \sum_{k=1}^p \alpha'_k S[n-k]$$

For correct predication of human vocal track filter coefficients, difference between $S[n]$ and $S'[n]$ should be minimum. α_k represents LPC model attribute which is filter coefficient of the Inverse-Vocal Tract filter model.

D. Combine M-MFCC and LPC

MFCC and LPC bring separate feature characteristics and we combine this information to identify similarity between two voice samples [22]. MFCC brings perceptual speech feature where LPC features give information about speaker vocal track for accurate prediction.

In our experiment we combined M-MFCC and LPC coefficients which give unique voice characteristics of a speaker. For every voice sample M-MFCC is one channel and LPC is another channel.

- M-MFCC Channel: Every frame has 40 features. 20 features from MFCC and another 20 features from delta MFCC.
- LPC Channel: Every frame has 40 features. 20 features from LPC and another 20 features from delta LPC

E. CNN Layers

In this experiment we have used 6 different layers as per Fig. 11. Convolutional layer gives learning capacity from different voice features.

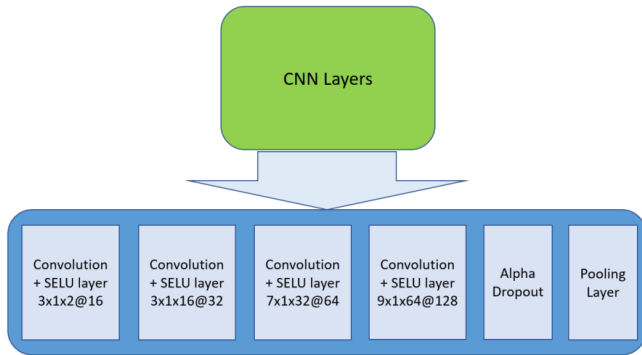


Fig. 11. Different CNN layers.

We have used 4 convolutional layers with SELU nonlinearity, one alpha drop and one pooling layer. Alpha dropout layer works effectively with SELU activation function and it goes hand-in-hand with SELU activation for self-normalization.

- First convolution layer takes 2 channels as input (M-MFCC & LPC) with 16 filters and 3x1 kernel size.
- Second convolution layer takes input from first layer and it has 32 filters with 3x1 kernel size.
- Third convolution layer takes input from 32 neurons and has 64 filters with 7x1 kernel size.
- Final convolution layer takes input from third layer and it has 128 filters with 9x1 kernel size.
- Fifth layer is alpha dropout layer to maintain variance and mean at original input level. We have used dropout rate as 0.2
- Final layer is pooling layer and we have used average pooling.

F. Cosine Triplet Embedding Loss Function

Cosine triplet embedding loss function is used during training phase to calculate loss and make correct model. Ideally Euclidean space between positive and anchor sample should be minimum but Euclidean space between anchor and negative sample should be maximum. Similar to work in [7], if we denote positive sample by Sp, anchor sample by Sa and

negative sample by Sn then cosine triplet embedding loss function.

$$F(Sa, Sp, Sn) = \sum_{a,p,n}^N \cos(f(Sa, Sn)) - \cos(f(Sa, Sp)) + \alpha_{margin}$$

Sa and Sp samples from same speaker and ideally feature distance between Sa and Sp should be zero so cosine of f(Sa,Sp) will be 1. Sa and Sn samples come from different speakers so in other way round ideal distance between Sa and Sn should be maximum so cosine of f(Sa,Sn) will be minimum. This α_{margin} is used to maintain minimum distance between negative and positive voice sample. α_{margin} is adjustable hyper-parameter. And in our experiment, we have considered it at 0.5.

G. Testing Phase

In testing phase also, we have used combined M-MFCC and LPS for feature extraction. CNN layers were trained in training phase with weight sharing together with positive, anchor and negative sample CNN layers. Same trained layers were used in testing phase to test voice samples. as per below Fig. 12.

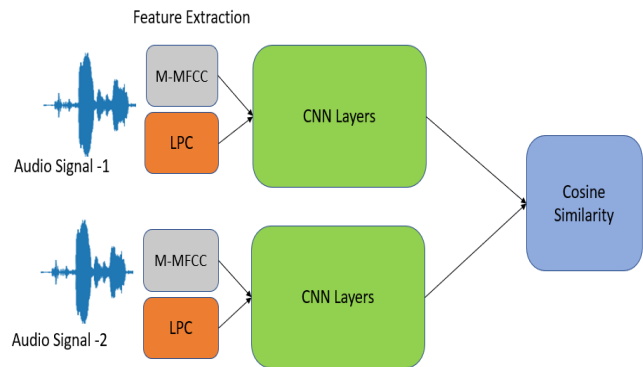


Fig. 12. CNN model testing phase.

In training we have used cosine triplet embedding loss function as we have trained model with three voice sample positive, anchor and negative audio. In testing phase, we have used cosine similarity as it is needed to validate similarity between two voice samples. Cosine similarity metric is used to compare the extracted embeddings and provide a corresponding match score as defined in Fig. 12. We have given two audio samples in testing phase Audio Sample 1 (AS1) and Audio Sample 2 (AS2). When AS1 and AS2 are from same speaker then ideally cosine similarity should 1 and when AS1 and AS2 are from two different speakers then cosine similarity should be “-1” (negative one).

VI. EXPEREMENT AND RESULT

As defined in section IV, we have created 6 different datasets from TIMIT and NOISEX-92 dataset and performed following six experiments with training and testing model defined in section V. Aim of these experiments was to train model with degraded human voice with one set of noises and

test model with different set of noises. Most of the applications need such kind of functionality to train the model in one set of noises and test in different set of noises. Below Table II sets mapping between experiment ID and corresponding training and testing dataset.

TABLE II. EXPERIMENTS WITH SIX DATASETS

Experiment ID	Training Dataset	Testing Dataset
1	DS1	DS2
2	DS2	DS1
3	DS3	DS4
4	DS4	DS3
5	DS5	DS6
6	DS6	DS5

The tables below show the comparison results for all experiments. We have compared speaker recognition system accuracy in terms of True Match Rate (TMR). TMR has calculated at 10 percent of False Match Rate (FMR). In speaker classification, increased TMR signifies reduction of false prediction (either negative or positive). It is observed that combined MFCC and LPC give significant improvement in TMR as compared to MFCC and it is similar to work in [7]. We got better result with M-MFCC and LPC as defined in Table III.

TABLE III. EXPERIMENTS RESULTS TMR

Exp ID	TMR@FMR=10%				
	Training Dataset	Testing Dataset	MFCC	MFCC & LPC	M-MFCC & LPC
1	DS1	DS2	65.3	80.6	89.1
2	DS2	DS1	66.9	79.7	87.9
3	DS3	DS4	63.8	77.7	83.2
4	DS4	DS3	62.9	69.4	83.4
5	DS5	DS6	60.4	69.4	72.6
6	DS6	DS5	66.6	80	92

We also compare Equal Error Rate (EER) in percentage with all three feature extraction methods. EER represents a point in DET or ROC curve where false rejection rate is equal to false acceptance rate. When both rates (false acceptance and false rejection) are equal then that common value is considered as EER. Low EER represents improved accuracy of speaker recognition system. As per Table IV results, we observed that combined MFCC and LPC gives very significant improvement in EER as compared to MFCC [7] and on top of this we got improvement with M-MFCC and LPC case.

TABLE IV. EXPERIMENTS RESULTS EER

Exp ID	ERR in %				
	Training Dataset	Testing Dataset	MFCC	MFCC & LPC	M-MFCC & LPC
1	DS1	DS2	21.8	12.3	10.5
2	DS2	DS1	20.2	12.5	11.3
3	DS3	DS4	20	12.8	12.1
4	DS4	DS3	22	14.4	12.0
5	DS5	DS6	23.5	14.3	13.8
6	DS6	DS5	18.7	12.5	10.8

VII. CONCLUSION

Different applications require different types of speaker recognition system. Most of the applications are looking for speaker recognition system which should work with no background noise during training, but same system should recognize speaker even with degraded human voice. In our experiment we have used same approach and used speaker voice with most silent background noise (SNR = 10-12dB) and test degraded human voice up to 1dB SNR. In our experiment (As per section V) we observed that high sampling rate, optimized triangular mel-bandpass filter frequency range and optimized pre-emphasis value gives better EER and TMR. Improved EER and TMR are impacting MFCC values for effective speaker recognition process. This can further optimize in future to get better results. In future, triangular mel-bandpass filter frequency range can optimize further to get better results.

REFERENCES

- [1] W. Meiniar, F. A. Afrida, A. Irmasari, A. Mukti and D. Astharini, "Human voice filtering with band-stop filter design in MATLAB," 2017 International Conference on Broadband Communication, Wireless Sensors and Powering (BCWSP), Jakarta, 2017, pp. 1-4, doi: 10.1109/BCWSP.2017.8272563.
- [2] J. Wang and M. T. Johnson, "Physiologically-motivated feature extraction for speaker identification," 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2014, pp. 1690-1694
- [3] S. Ostrogonac, M. Sečujski, D. Knezevic and S. Suzić, "Extraction of glottal features for speaker recognition," 2013 IEEE 9th International Conference on Computational Cybernetics (ICCC), 2013, pp. 369-373
- [4] N. V. Tahlirmani and N. Bhatt, "Performance Analysis of Speaker Identification System With and Without Spoofing Attack of Voice Conversion," 2018 2nd International Conference on Micro-Electronics and Telecommunication Engineering (ICMETE), Ghaziabad, India, 2018, pp. 130-135
- [5] A. H. Meftah, H. Mathkour, S. Kerrache and Y. A. Alotaibi, "Speaker Identification in Different Emotional States in Arabic and English," in IEEE Access, vol. 8, pp. 60070-60083, 2020
- [6] N. Gupta and S. Jain, "Speaker Identification Based Proxy Attendance Detection System," 2019 International Conference on Signal Processing and Communication (ICSC), NOIDA, India, 2019, pp. 175-179
- [7] A. Chowdhury and A. Ross, "Fusing MFCC and LPC Features Using 1D Triplet CNN for Speaker Recognition in Severely Degraded Audio Signals," in IEEE Transactions on Information Forensics and Security, vol. 15, pp. 1616-1629, 2020
- [8] S. Nakagawa, L. Wang and S. Ohtsuka, "Speaker Identification and Verification by Combining MFCC and Phase Information," in IEEE Transactions on Audio, Speech, and Language Processing, vol. 20, no. 4, pp. 1085-1095, May 2012

- [9] F. Eyben et al., "The Geneva Minimalistic Acoustic Parameter Set (GeMAPS) for Voice Research and Affective Computing," in *IEEE Transactions on Affective Computing*, vol. 7, no. 2, pp. 190-202, 1 April-June 2016
- [10] M. Sadeghi and H. Marvi, "Optimal MFCC features extraction by differential evolution algorithm for speaker recognition," 2017 3rd Iranian Conference on Intelligent Systems and Signal Processing (ICSPIS), 2017, pp. 169-173
- [11] Gupta, Shikha & Jaafar, Jafreezal & Wan Ahmad, Wan Fatimah & Bansal, Arpit. (2013). Feature Extraction Using Mfcc. *Signal & Image Processing : An International Journal*. 4. 101-108
- [12] A. Winursito, R. Hidayat and A. Bejo, "Improvement of MFCC feature extraction accuracy using PCA in Indonesian speech recognition," 2018 *International Conference on Information and Communications Technology (ICOIACT)*, 2018, pp. 379-383
- [13] Monalisha Barik, Susanta Kumar Sarangi and Sushanta Kumar Sahu, "Real-time speaker identification system using cepstral features", in *Communication Control and Intelligent Systems (CCIS)*, 2016 2nd International Conference on, March 2017
- [15] O. Büyüyük and L. M. Arslan, "Age identification from voice using feed-forward deep neural networks," 2018 26th Signal Processing and Communications Applications Conference (SIU), Izmir, 2018, pp. 1-4
- [16] Z. Weng, L. Li and D. Guo, "Speaker recognition using weighted dynamic MFCC based on GMM," 2010 *International Conference on Anti-Counterfeiting, Security and Identification*, 2010, pp. 285-288
- [17] H. C. Bao and Z. C. Juan, "The research of speaker recognition based on GMM and SVM," 2012 *International Conference on System Science and Engineering (ICSSE)*, 2012, pp. 373-375
- [18] Z. Weng, L. Li and D. Guo, "Speaker recognition using weighted dynamic MFCC based on GMM," 2010 *International Conference on Anti-Counterfeiting, Security and Identification*, 2010, pp. 285-288
- [19] Y. Wei, "Adaptive Speaker Recognition Based on Hidden Markov Model Parameter Optimization," in *IEEE Access*, vol. 8, pp. 34942-34948, 2020
- [20] R. M. Lexuşan, "Comparative study regarding characteristic features of the human voice," 2015 7th International Conference on Electronics, Computers and Artificial Intelligence (ECAI), Bucharest, 2015, pp. WSD-1-WSD-4
- [21] B. K. Baniya, J. Lee and Z. Li (2014), " Audio feature reduction and analysis for automatic music genre classification", In *IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 457-462
- [22] S. Cumani and P. Laface, "Large-Scale Training of Pairwise Support Vector Machines for Speaker Recognition," in *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 11, pp. 1590-1600, Nov. 2014
- [23] R. Mardhotillah, B. Dirgantoro and C. Setianingsih, "Speaker Recognition for Digital Forensic Audio Analysis using Support Vector Machine," 2020 *3rd International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*, 2020
- [24] D. Snyder, D. Garcia-Romero, D. Povey and S. Khudanpur, "Deep neural network embeddings for text-independent speaker verification", *Proc. of Interspeech*, pp. 999-1003, 2017
- [25] S. J. Wemndt and R. L. Mitchell, "Machine recognition vs human recognition of voices," 2012 *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Kyoto, 2012, pp. 4245-4248
- [26] M. N. A. Aadit, S. G. Kirtania and M. T. Mahin, "Suppression of white and colored noise in Bangla speech using Kalman filter," 2016 3rd International Conference on Electrical Engineering and Information Communication Technology (ICEEICT), Dhaka, 2016, pp. 1-6
- [27] Thimmaraja Yadava G, Jai Prakash T S and Jayanna H S, "Noise elimination in degraded Kannada speech signal for Speech Recognition," 2015 *International Conference on Trends in Automation, Communications and Computing Technology (I-TACT-15)*, Bangalore, 2015, pp. 1-6
- [28] A. Varga and H. J. M. Steeneken, "Assessment for automatic speech recognition: II. NOISEX-92: A database and an experiment to study the effect of additive noise on speech recognition systems," *Speech Commun.*, vol. 12, pp. 247-251, Jul. 1993
- [29] F. Beritelli, "Effect of background noise on the SNR estimation of biometric parameters in forensic speaker recognition," 2008 2nd International Conference on Signal Processing and Communication Systems, 2008, pp. 1-5
- [30] P. Papadopoulos, A. Tsiartas, J. Gibson and S. Narayanan, "A supervised signal-to-noise ratio estimation of speech signals," 2014 *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2014, pp. 8237-8241
- [31] Z. Liu, Z. Wu, T. Li, J. Li and C. Shen, "GMM and CNN Hybrid Method for Short Utterance Speaker Recognition," in *IEEE Transactions on Industrial Informatics*, vol. 14, no. 7, pp. 3244-3252, July 2018
- [32] Z. Huang, M. Dong, Q. Mao and Y. Zhan, "Speech emotion recognition using CNN", *Proc. ACM Int. Conf. Multimedia*, pp. 801-804, 2014
- [33] S. -H. Park, Y. -H. Park, A. Nasridinov and J. -Y. Lee, "A Person Identification Method in CUG Using Voice Pitch Analysis," 2014 *IEEE Fourth International Conference on Big Data and Cloud Computing*, 2014, pp. 765-766
- [34] Himani Chauhan et al, "Voice Recognition" in *International Journal of Computer Science and Mobile Computing*, Vol.4 Issue.4, April- 2015, pp. 296-301
- [35] K. I. Nordstrom, G. Tzanetakis and P. F. Driessen, "Transforming Perceived Vocal Effort and Breathiness Using Adaptive Pre-Emphasis Linear Prediction," in *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16, no. 6, pp. 1087-1096, Aug. 2008

AUTHORS' PROFILE



Amit Moondra received the Master of Engineering degree in Communication Engineering from the Birla Institute of Technology and Science, Pilani, India (BITS - Pilani). He has more than 20 years of industrial and research experience with 10+ across countries.

He is currently working in Ericsson Global India Limited as Senior System Manger in Product Development Unit and pursuing his Ph.D. at Manav Rachna International Institute of Research and Studies, India. His research focuses on artificial intelligence, deep learning model in speech area. He is an active member of IEEE.



Poonam Chahal received her Ph.D. in 2017 from YMCA University of Science and Technology, Faridabad India, in the field of Artificial Intelligence. Presently she is working as Professor in Department of Computer Science and Engineering at FET, Manav Rachna International Institute of Research and Studies, Faridabad.

She is actively involved in research activities and is on the reviewing panel of many journals and conferences.

Customer Segmentation of Personal Credit using Recency, Frequency, Monetary (RFM) and K-means on Financial Industry

Hafidh Rizkyanto¹, Ford Lumban Gaol²

Computer Science Department, Bina Nusantara University, Jakarta, Indonesia¹

Doctor of Computer Science Department, Bina Nusantara University, Jakarta, Indonesia²

Abstract—This research focuses on how to build a segmentation model for credit customers to identify the potential for defaulting credit customers based on their transaction history. Currently, there is no segmentation available for this possibility of payment failure. Credit scoring helps in minimizing credit risk when applying for credit. However, using RFM (Recency, Frequency, Monetary) models helps to score each transaction variable of the customer's financial activity. K-means then assists in the process of segmenting the results of the RFM model scoring, which occurs in the middle of the customer's repayment schedule. Challenge is how to decide the variable that can be used in RFM models and how to interpret the clusters that have been formed and the actual implementation of the customer. The Bank can divide the clusters that have possibility of payment failure by their customers so that banks can take preventive actions and as information for the collection system to be able to make payment withdrawals or billing.

Keywords—Credit; credit risk; recency; frequency; monetary; K-means

I. INTRODUCTION

A commercial bank is a business entity that obtains funds from the public in the form of savings, term deposits, and other types of funds. These funds are then distributed to the public as credit with the purpose of improving their standard of living. The financial industry in banking offers various alternative money loans to the public, with the provision of loans in the form of credit to bank customers [1].

Regarding types of businesses, financial service industries that provide funds directly include commercial banks, guarantee companies, and factoring companies [2]. In this research, Bank XYZ, a commercial bank, is used as a case study. There are several types of commercial banks in Indonesia, including state banks, private national banks, foreign banks, joint banks, regional government banks, and commercial banks [3].

As of January 11, 2021, there were 198,986 active credit customers, with a total of 621,968 active and closed credit data for Bank XYZ. For the financial industry in banking, data is an important asset that can be used for corporate strategy. Banks have a large amount of raw data, including transaction data, customer data, and statement data, which require processing to provide decision support information [4]. However, Bank

XYZ, established on July 10, 1970, currently does not have a decision support system that can provide information to credit officers regarding potential customers or customers who may have payment difficulties, which could significantly impact the bank. Therefore, with the implementation of a data processing system, credit handling can be improved by each account officer who monitors the credit of their customers. This will help to reduce errors in the amount and credit payment process, which often results in changes in collectability and poor customer performance, leading to poor credit quality.

In general, data mining functions are divided into two parts: descriptive and predictive. Other functions include classification, association, clustering, sequencing, and forecasting [5]. Data mining techniques can be used to segment credit customers, with clustering as the algorithm used, specifically K-means. The RFM Model is used because it can be adapted to evaluate the value of customers [6] and classify them in different service areas such as finance, telecommunications, and e-commerce [7]. K-means is a non-hierarchical data grouping method that can partition data into two or more groups [8].

The combination of the RFM and K-Means Models can produce optimal segmentation because RFM establishes variables that are closely related to business needs, and K-Means can group them based on the similarities of each customer. The evaluation of the number of segments is determined using the Calinski-Harabasz Index (CH), which gives better results than clustering evaluation methods such as the Elbow method and others [9].

In providing loans, banks face various problems and risks, including the behavior of customers who do not pay their installments on time or delay payment of installments for several months, resulting in bad credit. Therefore, it is essential for banks to evaluate credit risk by focusing on several aspects, including determining the features impacting credit risk and predicting the possibility of default or payment failure [10]. An intelligent processing system is needed to assist banks in selecting prospective customers who will be given loans.

II. RELATED WORKS

This research is related to previous studies. The following are summaries of previous research that are relevant to this study. Table I presents a summary of the related works.

TABLE I. SUMMARY OF RELATED WORKS

The Authors	Methodology	Case
Farida Gultom dan Tober Simanjuntak	Algoritma naïve bayes dan K-Nearest Network	Results with a very low level of accuracy on the success of bank credit payments. Tests were conduct by combining the Naïve Bayes and Algorithms of kNN [11].
Xiaxia Niu, Jun Wu, Li Shi, Xiaodong Cui, Liping Yang, Yuanyuan Li, Sang-Bing Tsai, and Yunbo Zhang	Model of Recency Frequency Monetary and Algorithm of K - Means++, Metode PCA	The method of PCA was used to determine the indicator RFM weight. Customers were classified based on buying behavior into several groups [12].
A. Neyaa, A. Umamakeswari, A. Joy Christy, and L. Priyatharsini.	Model of RFM Analisis, K-Fuzzy C-Means Clustering, and RM K-Means Clustering	The research Propose a new method to select the initial centroid for algorithm K-means and to apply the method to segment the customers with reduced time and iterations [13].
Laxmiputra Salokhe, Saraswati Jadhav, and Rahul Shirole	Model of RFM Analysis, Algorithm of K-Means Clustering	Investigating the scope of customer value based on crossselling probability current value and customer loyalty, this paper uses a neural network approach that uses a Self Organization Map (SOM) to form clusters for banking [14].
Guangshu Xu, Yuanyuan Li , Jun Wu, Li Shi, Sang-Bing Tsai, Wen-Pin Lin, and Liping Yang,	RFM Model was improved and combine with Algorithm of K-Means.	Quantitative analysis method to make segmentation and platform clusters. Segmenting customers with clear values and purchasing preferences greatly helps platform for effectively allocating marketing resources to specific customer groups and for building healthy long term relationships with customers [15].

III. RESEARCH METHODOLOGY

Clustering can be used to divide all customers into several clusters based on various criteria that are similar to customers. Clustering is employed to group data naturally based on the similarity of data objects and to reduce their similarity with other clusters. Unlike classification, clustering is unsupervised learning and does not require a data training stage. In this study, K-Means and decision tree are used as data mining approaches. The K-Means algorithm can group bank customers based on their similarity in credit payment statements [16].

The method consists of several stages that need to be considered, as shown in Fig. 1.

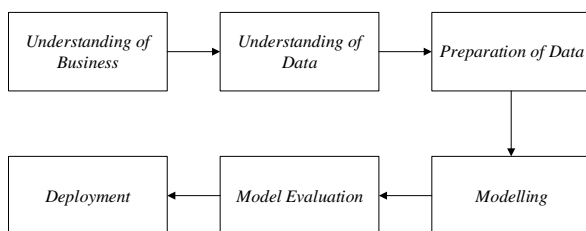


Fig. 1. The cross-industry standard process (CRISP).

This research suggests customer segmentation using the RFM Model and K-Means algorithm. Many studies have used the K-Means algorithm for the segmentation of customers [17] and traffic management systems [18]. The segmentation of customers and profiling of Recency, Frequency, and Monetary (RFM) method can help understand customer loyalty [19].

Furthermore, customers are assigned to three different variables: Recency, Frequency, and Monetary (RFM). By calculating scores for each instance, an assessment is done where the score ranges from 1 to 5, indicating the lowest and highest variable scores [13].

The evaluation of segmentation will be calculated using the Calinski-Harabasz index. The author in [9] pointed out that the Calinski-Harabasz index is the most reasonable metric to measure how well the number of groups formed by K-Means. Meanwhile, the Elbow method only calculates the error between data points X and centroid C, which will produce sum squared errors (SSE).

1) To calculate CH, the first step is to compute the inter-cluster dispersion or the between-group sum of squares (BGSS). In CH, the inter-cluster dispersion measures the weighted sum of squared distances between the centroid of the cluster and the centroid of the entire dataset (barycenter).

$$BGSS = \sum_{k=1}^K N_k x || C_k - C ||^2 \tag{1}$$

Where, N_k is the number of observations in cluster k , C_k is the center of mass of cluster k , C is the centroid of the barycenter, and K is the number of clusters.

2) The second step involves calculating the intra-cluster dispersion or within-group sum of squares (WGSS). In CH, intra-cluster dispersion measures the sum of the squares of the distances between each observation and the centroid of the same cluster.

$$WGSS_k = \sum_{i=1}^{N_k} || X_{ik} - C_k ||^2 \tag{2}$$

Where, X_{ik} is the observation of the- i in cluster k . Then, add up all the individuals in the group and calculate the sum of squares.

$$WGSS = \sum_{k=1}^K WGSS_k \tag{3}$$

3) The Calinski-Harabasz index is calculated by summing the inter-cluster dispersions and the intra-cluster dispersions for all clusters.

$$CH = \frac{BGSS}{WGSS} x \frac{N - K}{K - 1} \tag{4}$$

N is the number of observations and K is the number of clusters formed. From the equation above, it can be concluded that the greater the value of the Calinski-Harabasz index, the better the grouping is made.

To explain the results of customer segmentation, we can find the score of value and customer type, which can be used as a strategy by the company. For example, this information can be used for marketing or collecting strategies to increase the company's profits, as discussed in [17].

The achievement of the research is the application of customer segmentation in industries related to the banking sector, using the K-Means algorithm based on the score of RFM Credit Payment. The performance of the clustering methods in customer segmentation is shown in Fig. 2.

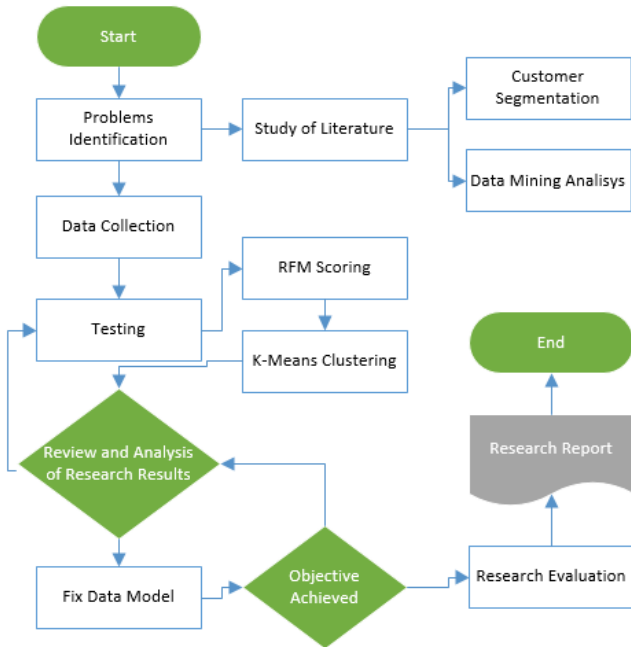


Fig. 2. Research methods.

There are steps to identify variables with good scores that can create better customer segmentation. Each stage of this process uses the Fig. 1 Cross-Industry Standard Process (CRISP), which includes the following details:

4) *Preprocessing*: This stage is the initial stage in the data processing, and it combines data understanding and data preparation. It contains two stages:

a) *Data Cleansing*. This stage is purposed to delete or eliminate variables that are not needed or cannot be used in the grouping process, such as variable names, address details, and dates.

b) *Merging and Generating Variable*. After getting good variables, we can connect each data point and change the variable according to the RFM concept. We take variables that are closely related to the last date, how many transactions, and the accumulated nominal transactions.

- This stage has several steps carried out to produce variables that are feasible for the clustering process. They are:

a) Identification of the quality of each variable using summary statistics. Variables that are not qualified, such as lots of empty data, almost all data values being the same, or the data not being varied, will be removed because they don't significantly affect clustering.

b) Handling empty data on variables. Research shows that using the median or middle value gives better results than

using the mean and k-NN imputation. This research will change each variable's blank value with the median of each variable.

- Modelling**: This stage changes the original value of the RFM variable that is formed into a score according to the concept of the RFM model. The original value changes to 5 bins or 5 score groups. The approach to dividing the original value is based on the frequency, so that an almost equal bin distribution can be formed.
- Modelling Evaluation**. The RFM score variable's results stage will be grouped using K-Means. Calinski-Harabasz (CH) will be used to obtain the most optimal number of clusters. Besides that, the results of plotting all the variables on the clusters formed will also be seen. The results will be tested on test data.
- Deployment**. This final stage interprets the clusters that have been formed and implements the actual customer collectibility process.

IV. RESULT AND DISCUSSION

A. Data Collection

The provided data pertains to credit, savings summaries, and transactions. The customer's personal data is deleted to maintain confidentiality, ensuring that the data processing focuses on transaction variables that will classify the customer's transaction behavior. The data processing process, from start to finish, employs R programming. Table II displays the total observations and variables for each data point.

TABLE II. DATA SOURCE RESEARCH

Data Test	Description	Data Load
Credit	Credit Data	10312 obs. of 24 variables
Saving	Saving Account Data	12373 obs. of 9 variables
Transaction	Transaction / Bank Statement Data	1048575 obs. of 6 variables

B. Model Development

This stage divides into four sub-chapters, they are pre-processing, feature engineering, modeling, and cluster interpretation. Generally, the pre-processing stage involves producing feasible variables for processing in the feature engineering stage. Feature engineering, on the other hand, involves transforming original variables into new ones based on the RFM concept. Meanwhile, the modeling stage involves grouping the previous results using K-Means. At the end of this stage, a detailed description of the formed clusters along with their implementation for collectability will be provided.

1) *Pre-processing*: This stage focuses on eliminating unqualified variables from the modeling process. Each data point will be cleaned, and this stage will remove any variables that cannot be used in the clustering process. Table III shows a list of the savings variables that were deleted because they could not be used.

TABLE III. LIST OF REMOVED VARIABLE SAVING

LIST OF REMOVED VARIABLE SAVINGS				
BRCOD	LOKASI	PDID	AOID	SGBU
SGID	AMBLK	CCCOD	AUTID	CSNO
WARS2	WARS3	TGLOPN	KERJA	SPCL
CTAX	CDATE	AINTR	ATAX	ADATE
INPDAT	INPUID	MNTDAT	MNTUID	AUTDAT
BCCOD	NAMA	ALM1	KPOST	WARS1
INTSPD	DRLIM	SEQNO	SWCIN	CINTR
CONVSW	CONVDT	ACCOLD	ACCNEW	ACSTA
YYMMDD	SEQBAL			

Analyzing this variable produces several important variables that can be used for modeling, as presented in Table IV: "Used Variables in Savings Data".

TABLE IV. USED VARIABLE AFTER DATA CLEANING SAVING

Variable	Description
ACNO	Account Number
LSTRN	Updated Date
MVCRED	Daily Credit Total
MVDBTD	Daily Debet Total
MVCREM	Montly Credit Total
MVDBTM	Montly Debet Total
BLCUR	Balance
BQCUR	Balance For Foreign Exchange
BLPRV	Previous Daily Balance
AVBLM	Previous Monthly Balance
CSSTA	Customer Status

Each data point will undergo data cleaning to remove transaction variables that cannot be used in the clustering process. The list of deleted variables is shown in Table V. The process of analyzing the remaining variables yields several important variables for modeling, which are presented in Table VI as the used variables in transactions data.

TABLE V. LIST OF REMOVED VARIABLE DATA TRANSACTION

LIST OF REMOVED VARIABLE TRANSACTIONS		
TRIDT	TRCON	POST
MODULE	TRCHAR	BRCOD
DPCOD	BTCOD	TRNO
SGBUT	SGIDT	GLIND
PDID	DPCODC	MMDDC
BTCODC	TRNOC	TRNOB
ACNOB	TRDAT	VLDAT
CQDATT	CQNUM	BKMNETT
BKCOD	CCCODT	RTX
AQTR	AMOD	AMPRK
TRDES2	TRDES3	INDAT
INTIM	INTRM	INUID
MNDAT	MNTIM	MNTRM
MNUID	AUDAT	AUTIM
AUTRM	AUUID	RLDAT
RLTIM	RLTRM	RLUID
ULOC	UBRANT	ULOCT
UBRANC	ULOCC	MODULC
ACNOO	SGBUC	SGIDC
BRCODB	MMDDB	BTCODB

TABLE VI. USED VARIABLES AFTER DATA CLEANING TRANSACTION

Variable	Description
TRST	Transaction Status
ACNOT	Debet Account
AMTR	Transaction Amount
MMDD	Date Transaction
ACNOC	Credit Account
TRDES1	Transaction Description

Similarly, each data point will also undergo data cleaning to remove loans variables that cannot be used in the clustering process. The list of deleted variables is shown in Table VII. The process of analyzing the remaining variables yields several important variables for modeling, which are presented in Table VIII as the used variables in loans data.

TABLE VII. LIST OF REMOVED VARIABLES AFTER DATA CLEANING TRANSACTION

LIST OF REMOVED VARIABLE LOANS				
ACNO	NOREG	SGBULX	SGIDLX	CRSEG
AOIDLX	CSNO	PDID	GLB	CRNAME
CRBRN	CRLOK	CCCOD	ACSTA	ACCLS
TARGET	HSLBNG	BLACM	BLYEAR	BGYEAR
BLOLD	BLOLDN	PARMTR	AKDDAT	DRPDAT
LNSDAT	GPLX	MTDBNG	RATBIR	SPREAD
RATEFF	RATKON	CODLK	STSBNG	REVIEW
DLRCD	AKDNO	CRTNGD	BLNPOK	BLNBNG
BLNDND	GLNUM	CRCON	CPRNO	SWRK1
SWRK2	RKNO2	CODRK1	CODRK2	YYMMDD
CTB08	CTB09	INDAT	INUID	MNDAT
MNUID	AUDAT	AUUID	BISFT	BIINS
BIDEB	BIKRD	BISEK	BISBNG	BILOK
BILOK2	BIPJM1	BIPJM2	BIBNG	NOSRT
SPKE	TGLSP	OLDCOLL	CRTRS	RPROV
JWPROV	TGPROV	CRPROV	TGLUCOL	BATUCOL
KETUCOL	SKETUCOL	TGLURAT	BATURAT	CADANG

TABLE VIII. USED VARIABLES IN LOANS DATA

Variable	Description
CRLIMA	Loan Limit First
CRLIM	Loan Limit
BLCUR	Current Outstanding
MVDEBT	Debet Movement
MVCRED	Credit Movement
MASA	Period Loans (Month)
RATDND	Penalty Rate
CRTNGA	Installment arrears
CRTNGB	Interest arrears
DESCON	End of Month Balanced
JAMIN	Guarantee amount
BICOLL	Collectability
OSATHN	Outstanding early years
CLATHN	Early year limit
PKATHN	Main Balanced early years
BGATHN	Interest early years
DNATHN	Penalty early years
RTATHN	Rate Early Years
PBLATHN	Main month number
BBLATHN	Interest month number
DBLATHN	Penalty month number

The second stage in pre-processing involves merging the data points and generating new variables based on the RFM concept. The loans data is the main data for segmentation in the bill collectability process, and it will be combined with savings and transaction data.

The transaction data is divided into two categories: recipient (debit) and sender (credit). This division is used to gain more insight into the transaction behavior of credit customers. The date-related variables are converted into days, with December 31, 2021, serving as the reference point.

For example, the MMDD variable is converted into days, where smaller values indicate customers who are new to transactions and vice versa. This processing transforms the original variables into variables according to the RFM concept, providing information about novelty, frequency, and total purchases or transactions. Unlike ordinary RFM, this study produces multiple RFM variables to classify credit customer behavior. The results of RFM variable processing are shown in Table IX.

TABLE IX. USED VARIABLES AFTER DATA CLEANING TRANSACTION

Variable RFM
<i>recency_savings</i>
<i>monetary_daily_debet</i>
<i>monetary_monthly_credit</i>
<i>monetary_monthly_debet</i>
<i>monetary_current_balance</i>
<i>monetary_current_balance_equi</i>
<i>monetary_prev_daily_balance</i>
<i>monetary_prev_monthly_balance</i>
<i>monetary_first_credit_limit</i>
<i>monetary_credit_limit</i>
<i>monetary_current_outstanding</i>
<i>monetary_debet_mutation</i>
<i>monetary_credit_mutation</i>
<i>monetary_bad_installment</i>
<i>monetary_interest_arrear</i>
<i>monetary_balance_last_month</i>
<i>monetary_collateral_amount</i>
<i>monetary_colectibility</i>
<i>monetary_beginning_year</i>
<i>monetary_limit_first_year</i>
<i>monetary_principal_first_year</i>
<i>monetary_interest_first_year</i>
<i>monetary_penalty_first_year</i>
<i>monetary_rate_first_year</i>
<i>monetary_number_ofMonth_principal</i>
<i>monetary_number_ofMonth_interest</i>
<i>monetary_number_ofMonth_penalty</i>

<i>frequency_credit</i>
<i>recency_sender_transaction</i>
<i>frequency_sender_transaction</i>
<i>monetary_sender_transaction</i>
<i>recency_sender_transaction_pending</i>
<i>frequency_sender_transaction_pending</i>
<i>monetary_sender_transaction_pending</i>
<i>recency_sender_transaction_success</i>
<i>frequency_sender_transaction_success</i>
<i>monetary_sender_transaction_success</i>
<i>recency_sender_transaction_reverse</i>
<i>frequency_sender_transaction_reverse</i>
<i>monetary_sender_transaction_reverse</i>
<i>recency_sender_transaction_desc_taspen</i>
<i>frequency_sender_transaction_desc_taspen</i>
<i>monetary_sender_transaction_desc_taspen</i>
<i>recency_sender_transaction_desc_interest</i>
<i>frequency_sender_transaction_desc_interest</i>
<i>monetary_sender_transaction_desc_interest</i>
<i>recency_receiver_transaction</i>
<i>frequency_receiver_transaction</i>
<i>monetary_receiver_transaction</i>
<i>recency_receiver_transaction_pending</i>
<i>frequency_receiver_transaction_pending</i>
<i>monetary_receiver_transaction_pending</i>
<i>recency_receiver_transaction_success</i>
<i>frequency_receiver_transaction_success</i>
<i>monetary_receiver_transaction_success</i>
<i>recency_receiver_transaction_reverse</i>
<i>frequency_receiver_transaction_reverse</i>
<i>monetary_receiver_transaction_reverse</i>
<i>recency_receiver_transaction_desc_taspen</i>
<i>frequency_receiver_transaction_desc_taspen</i>
<i>monetary_receiver_transaction_desc_taspen</i>
<i>recency_receiver_transaction_desc_interest</i>
<i>frequency_receiver_transaction_desc_interest</i>
<i>monetary_receiver_transaction_desc_interest</i>

2) *Feature engineering*: This stage focuses on engineering variables that will be used for clustering effectively. Feature engineering involves several stages including identifying the quality of each variable, handling empty data, and changing the original value to the RFM score variable. Table IX presents 64 variables resulting from pre-processing, which help identify the quality of the variables. Additionally, Table X provides a summary of statistics for each variable.

TABLE X. RESULT OF SUMMARY STATISTICS FOR EACH VARIABLE

Variable RFM	Type	Total NA	Mode Value	Total Mode	Total Unique Value	Presentase NA	Presentase Mode	Presentase Value
<i>recency_savings</i>	double	0	30	1016	153	0%	16%	84%
<i>monetary_daily_debet</i>	double	0	0	6432	37	0%	99%	1%
<i>monetary_monthly_credit</i>	double	0	1219700	218	3729	0%	3%	97%
<i>monetary_monthly_debet</i>	double	0	0	26	5917	0%	0%	100%
<i>monetary_current_balance</i>	double	0	0	26	6413	0%	1%	99%
<i>monetary_current_balance_equi</i>	double	0	0	6461	1	0%	100%	0%
<i>monetary_prev_daily_balance</i>	double	0	20000	7	6448	0%	0%	100%
<i>monetary_prev_monthly_balance</i>	double	0	0	6461	1	0%	100%	0%
<i>monetary_first_credit_limit</i>	double	0	0	3237	2012	0%	50%	50%
<i>monetary_credit_limit</i>	double	0	0	3470	2957	0%	54%	46%
<i>monetary_current_outstanding</i>	double	0	0	3473	2954	0%	54%	46%
<i>monetary_debet_mutation</i>	double	0	0	6461	1	0%	100%	0%
<i>monetary_credit_mutation</i>	double	0	0	6451	11	0%	100%	0%
<i>monetary_bad_installment</i>	double	0	0	5985	435	0%	93%	7%
<i>monetary_interest_arrear</i>	double	0	0	6410	52	0%	99%	1%
<i>monetary_balance_last_month</i>	double	0	0	6380	81	0%	99%	1%
<i>monetary_collateral_amount</i>	double	0	0	6338	13	0%	98%	2%
<i>monetary_colectibility</i>	double	0	0	3238	23	0%	50%	50%
<i>monetary_beginning_year</i>	double	0	0	3293	3160	0%	51%	49%
<i>monetary_limit_first_year</i>	double	0	0	6461	1	0%	100%	0%
<i>monetary_principal_first_year</i>	double	0	0	6388	74	0%	99%	1%
<i>monetary_interest_first_year</i>	double	0	0	6399	63	0%	99%	1%
<i>monetary_penalty_first_year</i>	double	0	0	6428	34	0%	99%	1%
<i>monetary_rate_first_year</i>	double	0	0	6461	1	0%	100%	0%
<i>monetary_number_ofMonth_principal</i>	double	0	0	6388	13	0%	99%	1%
<i>monetary_number_ofMonth_interest</i>	double	0	0	6399	15	0%	99%	1%
<i>monetary_number_ofMonth_penalty</i>	double	0	0	6428	9	0%	99%	1%
<i>frequency_credit</i>	integer	0	0	3231	15	0%	50%	50%
<i>recency_sender_transaction</i>	double	0	30	4120	8	0%	64%	36%
<i>frequency_sender_transaction</i>	integer	0	1	5384	25	0%	83%	17%
<i>monetary_sender_transaction</i>	double	0	1219700	229	2959	0%	4%	96%
<i>recency_sender_transaction_pending</i>	double	6547	25	2	4	100%	0%	0%
<i>frequency_sender_transaction_pending</i>	integer	6547	1	4	2	100%	0%	0%
<i>monetary_sender_transaction_pending</i>	double	6547	300000	1	5	100%	0%	0%
<i>recency_sender_transaction_success</i>	double	0	30	4120	8	0%	64%	36%
<i>frequency_sender_transaction_success</i>	integer	0	1	5387	25	0%	83%	17%
<i>monetary_sender_transaction_success</i>	double	0	1219700	229	2959	0%	4%	96%
<i>recency_sender_transaction_reverse</i>	double	6455	23	2	5	100%	0%	0%
<i>frequency_sender_transaction_reverse</i>	integer	6455	2	6	2	100%	0%	0%
<i>monetary_sender_transaction_reverse</i>	double	6455	0	6	2	100%	0%	0%

recency_sender_transaction_desc_taspen	double	3399	30	3047	6	53%	47%	0%
frequency_sender_transaction_desc_taspen	integer	3399	1	3015	3	53%	47%	1%
monetary_sender_transaction_desc_taspen	double	3399	1219700	220	1011	53%	3%	44%
recency_sender_transaction_desc_interest	double	6077	23	68	9	94%	1%	5%
frequency_sender_transaction_desc_interest	integer	6077	1	308	8	94%	5%	1%
monetary_sender_transaction_desc_interest	double	6077	22602	19	265	94%	0%	6%
recency_receiver_transaction	double	0	30	2973	8	0%	31%	69%
frequency_receiver_transaction	integer	0	1	2471	40	0%	38%	62%
monetary_receiver_transaction	double	0	1000000	137	3947	0%	2%	98%
recency_receiver_transaction_pending	double	6454	29	3	5	100%	0%	0%
frequency_receiver_transaction_pending	integer	6454	1	7	2	100%	0%	0%
monetary_receiver_transaction_pending	double	6454	4000000	1	8	100%	0%	0%
recency_receiver_transaction_success	double	3	30	1974	9	0%	31%	69%
frequency_receiver_transaction_success	integer	3	1	2483	38	0%	38%	62%
monetary_receiver_transaction_success	double	3	1000000	137	3946	0%	2%	98%
recency_receiver_transaction_reverse	double	6383	25	18	9	99%	0%	1%
frequency_receiver_transaction_reverse	integer	6383	2	64	4	99%	1%	0%
monetary_receiver_transaction_reverse	double	6383	0	78	2	99%	1%	0%
recency_receiver_transaction_desc_taspen	double	6444	30	11	5	100%	0%	0%
frequency_receiver_transaction_desc_taspen	integer	6444	1	17	2	100%	0%	0%
monetary_receiver_transaction_desc_taspen	double	6444	1500000	2	17	100%	0%	0%
recency_receiver_transaction_desc_interest	double	5111	30	444	8	79%	7%	14%
frequency_receiver_transaction_desc_interest	integer	5111	1	1225	8	79%	19%	2%
monetary_receiver_transaction_desc_interest	double	5111	1000000	48	425	79%	1%	20%

Table X displays the amount of empty data, most frequent data, and corresponding percentages for each variable. Certain variables are deemed unqualified and are therefore cleaned up using some simple logic, including the following:

- Variables with empty data greater than 70 percent are deleted as they do not provide any useful information or insights.
- Variables with a mode percentage greater than 80 percent are deleted as they do not offer significant insights for grouping. For instance, Fig. 3 shows that the RFM variable will be deleted.

[1] "monetary_daily_debet"	"monetary_current_balance_equi"	"monetary_prev_monthly_balance"
[4] "monetary_debet_mutation"	"monetary_credit_mutation"	"monetary_bad_installment"
[7] "monetary_interest_arrear"	"monetary_balance_last_month"	"monetary_collateral_amount"
[10] "monetary_limit_first_year"	"monetary_principal_first_year"	"monetary_interest_first_year"
[13] "monetary_penalty_first_year"	"monetary_rate_first_year"	"monetary_number_ofMonth_principal"
[16] "monetary_number_ofMonth_interest"	"monetary_number_ofMonth_penalty"	"frequency_sender_transaction"
[19] "recency_sender_transaction_pending"	"frequency_sender_transaction_pending"	"monetary_sender_transaction_pending"
[22] "frequency_sender_transaction_success"	"recency_sender_transaction_reverse"	"frequency_sender_transaction_reverse"
[25] "monetary_sender_transaction_reverse"	"recency_sender_transaction_desc_taspen"	"frequency_sender_transaction_desc_taspen"
[28] "recency_sender_transaction_desc_interest"	"frequency_sender_transaction_desc_interest"	"monetary_sender_transaction_desc_interest"
[31] "recency_receiver_transaction_pending"	"frequency_receiver_transaction_pending"	"monetary_receiver_transaction_pending"
[34] "recency_receiver_transaction_reverse"	"frequency_receiver_transaction_reverse"	"monetary_receiver_transaction_reverse"
[37] "recency_receiver_transaction_desc_taspen"	"frequency_receiver_transaction_desc_taspen"	"monetary_receiver_transaction_desc_taspen"
[40] "recency_receiver_transaction_desc_interest"	"frequency_receiver_transaction_desc_interest"	"monetary_receiver_transaction_desc_interest"

Fig. 3. Removed RFM variable.

After conducting quality identification on the 21 variables, any empty data is changed using the median or middle value of each variable. Table XI lists the variables along with their median values.

TABLE XI. EMPTY VARIABLES RFM WITH THEIR MEDIAN VALUE

Variable	Median
monetary_sender_transaction_desc_taspen	3185000
recency_receiver_transaction_success	28
frequency_receiver_transaction_success	2
monetary_receiver_transaction_success	1136039

In the feature engineering process of this research, the original data is converted into RFM scores. The variable values are sorted into five groups based on the data distribution quantiles. Fig. 4 illustrates the result of dividing the interval from the recency_savings, which denotes the last date of the customer's saving data update.

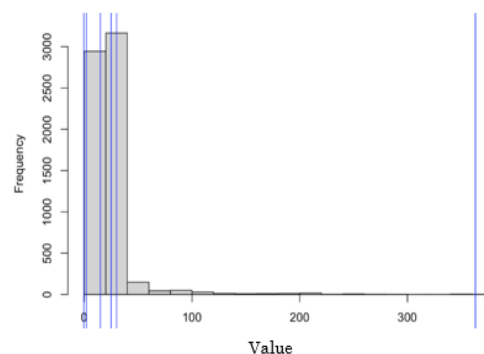


Fig. 4. Recency_savings variable bin formation interval.

The results of this bin formation will produce the RFM score variable, where the distribution of each score is shown in Fig. 5.

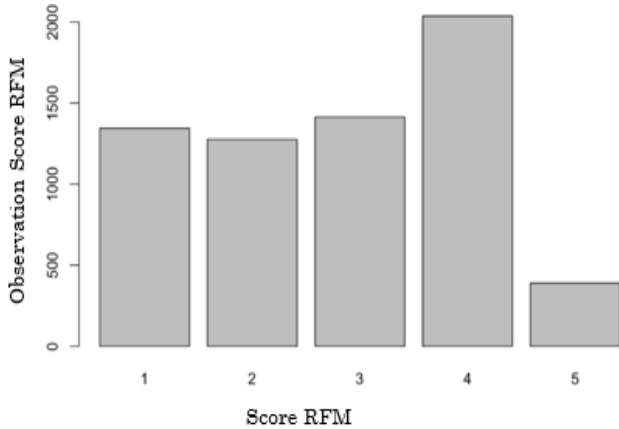


Fig. 5. Distribution of score_recency_tabungan.

The distribution of the Score_recency_saving variable indicates that the smaller the score, the more recently the customer carried out an activity related to savings. A score of 5 means that the customer has sustained long-term activities related to savings, and their distribution is quite small.

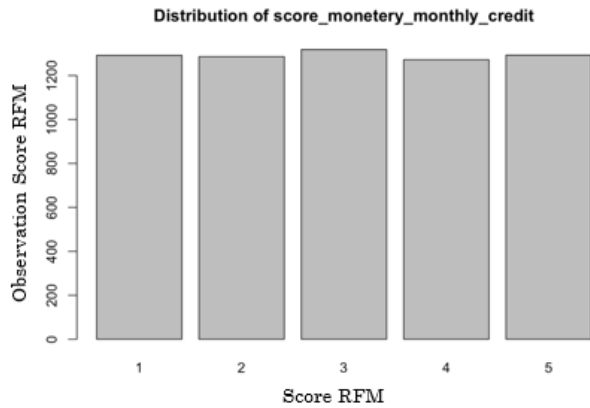


Fig. 6. Distribution of score_monetary_monthly_credit.

Fig. 6 shows the distribution of the Score_monetary_monthly_credit variable, where the difference is not significant. Fig. 7 shows the distribution of the Score_monetary_current_balanced variable.

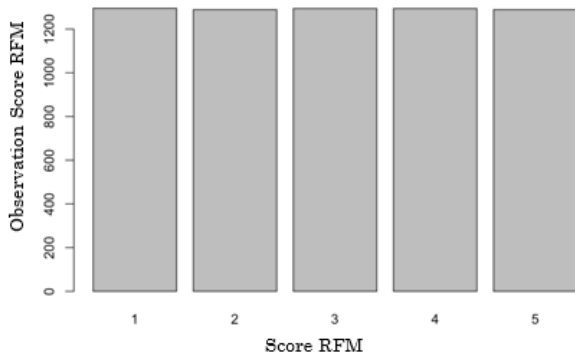


Fig. 7. Distribution of score_monetary_current_balanced.

3) *Modelling*: The modeling process is the final stage in the modeling process. This process involves grouping the RFM score variables using K-Means. First, the data will be randomly divided into two data sets, namely the training data and testing data, each comprising 80% and 20% of the data, respectively. The training data will be used to build a clustering model, which will be tested later using the testing data.

The first step is to remove variables that have a correlation greater than 0.7, as two or more highly correlated variables will not significantly affect the clustering process. Since there are 11 variables that have a high correlation, the clustering process uses 10 score variables.

The development stage of the K-Means model has been completed after testing the number of clusters. Testing the number of k clusters is an important step in the grouping process to obtain a number of clusters that is close to the ideal and can meet the need for customer collectibility.

The Calinski-Harabasz index is used to test how well the number of clusters is formed. Testing is done for the number of clusters ranging from 2 to 25, and the results of the Calinski-Harabasz index for each cluster are shown in Fig. 8.

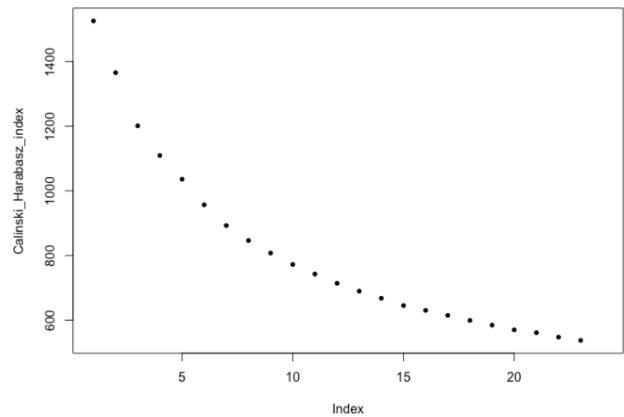


Fig. 8. Calinski-harabasz index results.

The Calinski-Harabasz (CH) Index can be used to evaluate the clustering model when ground truth labels are not known. It is used to test how well the clustering model has been created using quantities and features inherent to the dataset. The results of the Calinski-Harabasz index in Fig. 8 show that the number of clusters 2 has the highest value, and the larger the number of k clusters used, the smaller the Calinski-Harabasz index value. The CH index is a measurement of how similar an object is to its own cluster (cohesion) compared to other clusters (separation).

In addition to testing the Calinski-Harabasz index, the results of plotting all the variables on the clusters that are formed are also checked. Fig. 9 is a plot of the number of clusters ranging from 2 to 7.

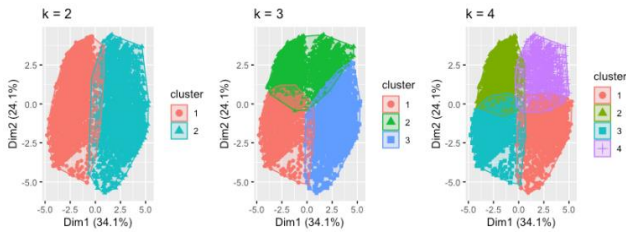


Fig. 9. Cluster plotting results $k < 4$.

Cluster 2 shows a significant difference in the distribution of the two customer groups, indicating that it is quite good at clustering. However, there is a small number of customers who intersect each other.

The same is true for clusters 3 and 4, but they have a larger number of intersecting customers. In contrast, cluster 5 has one customer group whose characteristics are quite similar to the other groups, namely customer group 4, which is mostly similar to customer group 5.

Fig. 10 shows that as the number of clusters increases to 6 and 7, more and more customer groups have overlapping characteristics. This is in line with the results of the Calinski-Harabasz index, where an increasing number of clusters results in a smaller value, indicating that the clustering results performed by K-Means are less optimal.

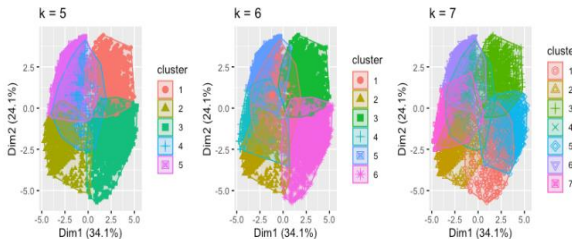


Fig. 10. Cluster plotting results $k > 4$.

This study chose to use four clusters because K-Means can divide customer groups quite well, with each group having different characteristics. There are some customers whose characteristics intersect with each other but are still understandable. Fig. 11 shows the result of clustering with four clusters.

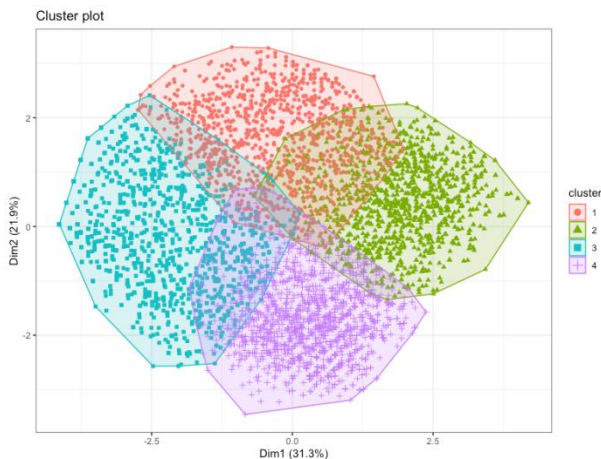


Fig. 11. Plotting results 4 clusters.

The results of clustering using four clusters show that there are only a few customer data that have intersections with other groups. Therefore, it can be concluded that four clusters are the most optimal number. Fig. 12 shows the result of grouping customers on test data.

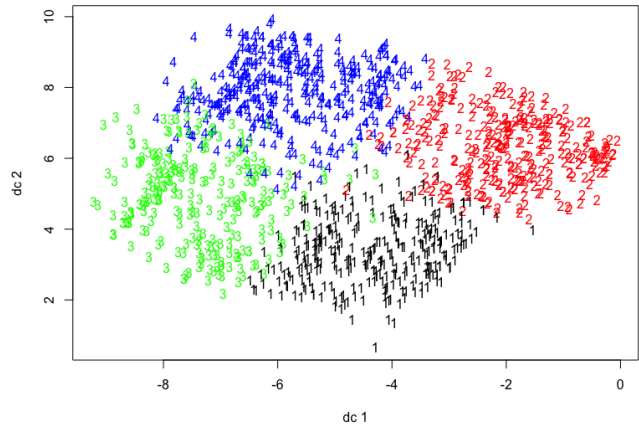


Fig. 12. Test data clustering.

This test step used training data and test data, each of which had a composition of 80% and 20%, respectively. The training data was used to build a clustering model, which was then applied to the test data. Clustering results on the test data also provide a good grouping of customers. There are clear differences between clusters, and only a few customers occur in cluster slices, especially in groups 3 and 4, which have more slices than the others.

4) Clustering result interpretation: Fig. 13 shows that the smaller the score, the better it is for the customer who has the most recent activity on their savings. Cluster 3 is the group where most customers carry out updates on savings data, followed by cluster 1. Clusters 2 and 4 are customer groups whose average is not too updated on savings activity, but cluster 2 is less updated than cluster 4. This can be an important record for the collectibility process. Fig. 14 shows a plot between the score_monetary_monthly_credit variable and the clustering results.

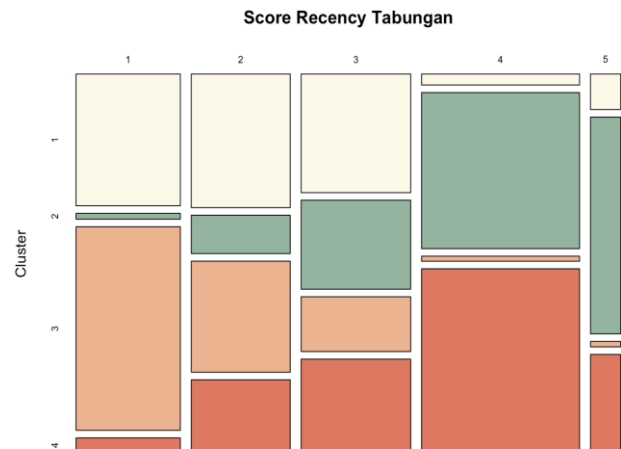


Fig. 13. Plotting variabel score_recency_tabungan and cluster.

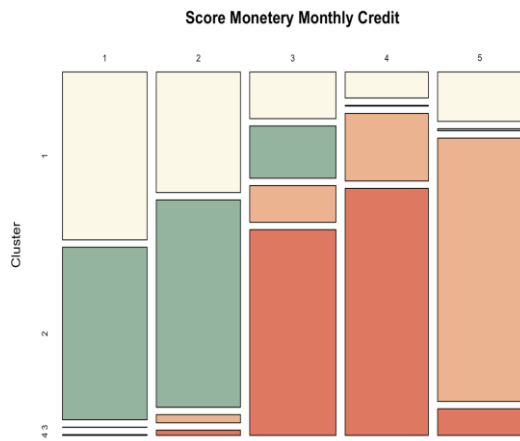


Fig. 14. Plotting variable score_monetary_monthly_credit and cluster.

This variable measures the average amount of money that comes out of the bank account of a customer each month. The higher the variable score, the higher the amount of money coming out of the customer's account. Cluster 3 is the group of customers who withdraw money from large nominal accounts, followed by cluster 4. Cluster 1 has an average score of 1 and a portion of the score 2. Cluster 2 has most customers in score 1, followed by score 2. Fig. 15 shows a plot of the variable score_monetary_current_balance and the clustering results.

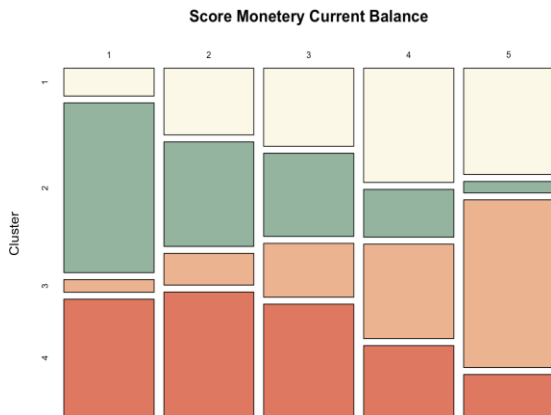


Fig. 15. Plotting variable score_monetary current balance and clusters.

The plotting results show that cluster 3 has the largest current account balance, followed by cluster 1. Meanwhile, cluster 4 has the average account balance, and cluster 2 has the smallest current balance.

The analysis that has been done can be summarized in Table XII, which ranks each variable.

TABLE XII. CONCLUSION OF CLUSTERING RESULTS ANALYSIS

Variable	Analysis
Score_recency_tabungan	The most frequent sequence of activities on savings is cluster 3, 1, 4, 2
Score_monetary_monthly_credit	The order of the most money from the account is cluster 3, 4, 1, 2
Score_monetary_current_balance	The current balance order of lots is clusters 3, 2, 1, 4
Score_frequency_credit	The order of the frequency of cash out is cluster 4, 2, 3, 1

Score_recency_sender_transaction	The sequence of most outgoing transaction updates is cluster 1, 3, 2, 4
Score_monetary_sender_transaction	The order of the most outgoing transactions is cluster 4, 3, 2, 1
Score_monetary_sender_transaction_desc_taspen	The order of the most outgoing transactions for pension funds is cluster 4, 1, 3, 2
Score_recency_receiver_transaction	The most updated sequence of incoming transactions is cluster 3, 1, 2, 4
Score_frequency_receiver_transaction	The order of the most incoming transactions is cluster 3, 1, 4, 2
Score_monetary_receiver_transaction	The order of the largest nominal incoming transactions is 3, 4, 2, 1

Based on the analysis of each score variable in the table above, it can be concluded that there is no dominant cluster that always ranks first, and the results are quite dynamic. The determination of cluster priority uses a point system, with each variable carrying the same weight.

The rule is that the first rank will receive the most points, and the lower the rank, the lower the points will be. Each rank starting from the first rank will receive 5, 3, 2, and 1 point. Table XIII provides detailed results of the point calculations for each score variable in the cluster.

TABLE XIII. CALCULATION OF POINT FOR EACH SCORE CLUSTER VARIABLES

Name of Variables	Point			
	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Score_recency_tabungan	3	1	5	2
Score_monetary_monthly_credit	2	1	5	3
Score_monetary_current_balance	2	3	5	1
Score_frequency_credit	1	3	2	5
Score_recency_sender_transaction	5	2	3	1
Score_monetary_sender_transaction	1	2	3	5
Score_monetary_sender_transaction_desc_taspen	3	1	2	5
Score_recency_receiver_transaction	3	2	5	1
Score_frequency_receiver_transaction	3	1	5	2
Score_monetary_receiver_transaction	1	2	5	3
Total Point	24	18	40	28

Based on the results of the point calculation shown in Table XIII, we can conclude the priority of cluster collectibility, which is as follows:

- Cluster 2 is a group of customers with very low scores in activity, frequency, and nominal transactions of money going out or coming in. Additionally, most customers have very low balances. Therefore, this cluster is the priority in the collectibility process due to the high potential for default.
- Cluster 1 is a group of customers with infrequent activity, the majority of whom have low balances, and the nominal amount and frequency of incoming money are also quite low. Therefore, this customer group is the second priority for the collectibility process because of its high potential for default.
- Cluster 4 is a group of customers with average savings activity, average cash-out transactions, most sufficient balances, and the largest outgoing transactions. Therefore, this group is included in the medium category, meaning that the potential for default is quite small.
- Cluster 3 has many groups of customers with new activities. Most customers have the largest balances, the most updated incoming money activity with the highest frequency, and the largest total nominal value. Therefore, this group is the one with the least potential for default.

V. CONCLUSION

The RFM Model can effectively form the score variable, allowing for an efficient clustering process using K-Means. The resulting cluster interpretation is easy and can provide solutions to problems. By implementing the Calinski-Harabasz index, the number of clusters used can be evaluated. This is an initial step towards determining the optimal number of clusters for the financial industry in banking data. The K-Means clustering results in well-formed groups, with significant customer grouping and no overlap between clusters. The resulting customer grouping can be useful for the financial industry in the process of collecting credit customers.

The next step in this study is to include additional variables, such as credit limits and credit card transactions, to provide payment options. The limitation of this approach is the lack of a Customer Relation Management system, which could provide a better understanding of customer perspectives, describe customer value, and improve the selling or cross-selling of various banking products and programs.

ACKNOWLEDGMENT

The authors acknowledge that Binus University Jakarta provided financial support and conducted a plagiarism check for this study.

REFERENCES

[1] P. Indonesia, "Undang-Undang Republik Indonesia Nomor 10 Tahun 1998 Tentang Perubahan Atas Undang-Undang Nomor 7 Tahun 1992 Tentang Perbankan.," 1998.

[2] H. Y. Y. & T. Z. Song, "How different types of financial service providers support small- and medium- enterprises under the impact of COVID-19 pandemic: from the perspective of expectancy theory," pp. Front. Bus. Res. China 14, 27, 2020.

[3] B. Indonesia, "GROUP OF BANKS AND TYPE OF LOANS," Bank Indonesia, Indonesia, 2022.

[4] A. Subrahmanyam, "Big data in finance: Evidence and challenges," Borsa Istanbul Review, pp. vol. 19, no. 4, pp. 283-287, 2019.

[5] S. M. J. Umarani, "Implementation of Data Mining Concepts in R Programming," International Journal of Trendy Research in Engineering and Technology, pp. ISSN NO 2582-0958, 2020.

[6] S. M. e. al., "Analysis for customer lifetime value categorization with RFM model," Procedia Computer Science, vol.161, pp.834-840, 2019.

[7] U. F. & D. N. Utama, "DEVELOPMENT OF BANK'S CUSTOMER SEGMENTATION MODEL," ICIC International c 2021 ISSN 2185-2766, pp. pp. 17-26, 2021.

[8] M. N. A. R. A. M. R. Md. Zakir Hossain, "A dynamic K-means clustering for data mining," Indonesian Journal of Electrical Engineering and Computer Science, pp. Vol. 13, No. 2, February 2019, pp. 521-526 ISSN: 2502-4752, 2019.

[9] E. Schubert, "Stop using the elbow criterion for k-means and how to choose the number of clusters instead," p. arXiv preprint arXiv:2212.12189. , 2022.

[10] C. J. H. K. J. Z. W. Chang Yin, "Evaluating the credit risk of SMEs using legal judgments," Decision Support Systems, 2020.

[11] T. S. Farida Gultom, "PREDIKSI TINGKAT KELANCARAN PEMBAYARAN KREDIT BANK DENGAN MENGGUNAKAN ALGORITMA NAÏVE BAYES DAN K-NEAREST NEIGHBOR," Jurnal Manajemen Informatika & Komputerisasi Akuntansi Vol. 4 No. 2 ISSN: 2598-8565, 2020.

[12] L. S. L. Y. X. N. Y. L. X. C. S.-B. T. a. Y. Z. Jun Wu, "User Value Identification Based on Improved RFM Model and K-Means++ Algorithm for Complex Data Analysis," Wireless Communications and Mobile Computing, pp. Volume 2021, Article ID 9982484, 8 pages, 2021.

[13] A. U. L. P. a. A. N. A. Joy Christy, "RFM ranking – An effective approach to customer segmentation," Journal of King Saud University - Computer and Information Sciences, 2018.

[14] L. S. a. S. J. Rahul Shirole, "Customer Segmentation using RFM Model and K-Means Clustering," International Journal of Scientific Research in Science and Technology, pp. Volume 8, Issue 3 Page Number : 591-597, 2021.

[15] L. S. W.-P. L. S.-B. T. Y. L. L. Y. a. G. X. Jun Wu, "An Empirical Study on Customer Segmentation by Purchase Behaviors Using a RFM Model and K-Means Algorithm," Mathematical Problems in Engineering, pp. Volume 2020, Article ID 8884227, 7 pages, 2020.

[16] A. a. B. K. Boyaci, "Data mining application in banking sector with clustering and classification methods," Proceedings of the 2015 International Conference on Industrial Engineering and Operations Management, Dubai, United Arab Emirates (UAE), 2015.


[17] Dedi et al., "Customer segmentation based on RFM value using k-means algorithm," Proc. of International Conference on Informatics and Computing (ICIC), Semarang, pp. pp.1-14, 2020.

[18] S. R. M. a. A. S. Nawrin, "Exploreing K-Means with Internal Validity Indexes for Data Clustering in Traffic Management System," International Journal of Advanced Computer Science and Applications, pp. Vol. 8, No. 3, 2017.

[19] M. C. H. a. I. P. G. H. Suputra, "ustomer Segmentation Using RFM Model," Jurnal Elektronik Ilmu Komputer Udayana, pp. Vol. 8, no. 2, pp. 153-161, 2019.

[20] J. Sessa and D. Syed, "Techniques to deal with missing data," 5th international conference on electronic devices, systems and applications (ICEDSA), pp. 1-4, December 2016.

Challenges of Digital Twin Technologies Integration in Modular Construction: A Case from a Manufacturer's Perspective

Laith Jamal Aldabbas 

Al-Ahliyya Amman University, Amman, Jordan

Abstract—Automated models of physical objects are known as Digital twins; letting one to outline, test, array, and monitor and manage robotics in the real world. .CPS (Cyber-physical system) data have to be assembled through real life procedures to form a real-time monitoring cyber model in order to produce Digital Twin. Modification in the cyber model will be shown in real life system to guess or manage. As a result of digital and the progression in ICT, manufacturing and aviation or aerospace industries are now utilizing digital twin. Nonetheless, the uses of DT's in several production firms have not been researched massively. Those studies were sparse in construction, where structures are built without Superstructures, which thus sparked global concern. Herein, DT applications in building/manufacturing sector and in various firms were reviewed first and thereafter aimed on publications concerning DT applications in industry area by organizing a systematic search via Scopus. Notably, the publications were singled out immediately after the assessment of the publications and the study continues to investigate and evaluate the Potentials of digital twins in MIC, Restriction of digital twins in MIC, Impact of digital twins on industry, and Cost of time which should be appropriate for model development. The analysis report however demonstrated that DT is dedicated and thoughtful of in midst of inclusion along other digital technologies. More so, theoretical structure is formed in order to apply DT in module installation in MiC around the circumstances of Hong Kong which happens to be usual city case of high-density. Interestingly, the implementation of Digital twin in Modular Integrated Construction is expected to provide promising potential with significant benefits, such as improved logistics and manufacturing management by employing Digital Twins to track on-site progress during module installation.

Keyword—Digital Twin; DTs; enabling technologies; digital twin model; applications; challenges; literature review

I. INTRODUCTION

Digital twin is subjected to intensive scrutiny. Both the Digital Model and the Digital Shadow identifies an infrastructure relationship between digital and physical components that is either unequal or unidentified [1]. A self-contained digital data structure that is related to a physical system is referred to as a DT. This information would be a "twin" of the raw document that was stored electronically in the physical system and was related to throughout its existence. According to the findings of [2] the term "Digital Twin" has been utilised to signify a variety of different numerical and physical links. Integration can be improved

with the help of a DT by evaluating and reviewing the status of both physical and digital elements simultaneously.

This article examines the most recent developments in digital technology, information technology, and digital twins, as well as the influence such developments have had on the industry. Secondly, it places an emphasis on the integration of Digital Twin technologies with BIM; which is the most recent digital technology that is supported whilst WSN and internet of things was talked about as well. Third, the article discusses the benefits of utilising Digital Twin technology as well as the challenges that must be overcome prior to its widespread adoption in the industrial industry. Implementation of DT in manufacturing area and in various firms was reviewed and aimed on publications concerning DT Implementation in industry area by performing a systematic search via Scopus. As soon as the reviewing publication procedure and evaluation were filtered out, the paper propels in examining and evaluating the potentials of digital twins in MIC, Restriction of digital twins in MIC, Impact of digital twins on industry, and Cost of time which should be appropriate for model development. The findings however unveiled DT to be esteemed with other digital technologies for inclusion.

A. Hong Kong Scenario

Manufacturing is the intended application for the suggested digital twin technology architecture in MIC modules [3]. The implementation of digital twin technologies in manufacturing presents a number of opportunities and advantages [4]. Economic growth and social advancement has been sustained by the construction industries in Hong Kong. It has added about 4.5% of Gross Domestic Product (GDP), 2018 [5]. Yet the industries continue to experience hard challenges, for instance, aging workforce, labour shortage and cost inflation [6]. Thus, it was announced within 2017 and 2018 by Policy Address of Government of Hong Kong Special Administrative Region (HKSAR) for government to welcome and propagate MiC in civil works firms [7,8].

In contrast to normal construction methods, MiC has distinguished its abilities whereby a huge of on-site works is relocated to an industry to produce modules. Building services, architectural works and structural works of modules were partly achieved, thereafter, accompanied by module transportation to construction site for coupling and installation. Thus, many MiC utilized projects has begun so as to propel its notable acceptance in Hong Kong.

B. Research Gaps and Future Motivation

This study is built through the previous literature on digital twins and modular integrated construction. It is important to start the study by examining the literature and finding out the research gaps. On the bases of the literature, the following research gaps have to be addressed.

- 1) Potential of digital twins in MIC.
- 2) Restriction of digital twins in MIC.
- 3) Impact of digital twins on industry.
- 4) Cost of time should be appropriate for model development.

C. Literature Review

Notably, the term "digital twin" (DT) was first defined by NASA in their Technology Roadmap in the year 2010; referred to as an integrated multi-physics, multi-scale, probabilistic simulation of in-built vehicle or framework which utilizes a better suitable real life models, sensor updates, fleet history, etc., to reflect the life of its proportional flying twin [9]. Around 2015, the idea of DT was attuned as composing a generic 'product,' other industries, including manufacturing, industrial engineering, and informatics, began to employ it as part of their business operations. DT thus stands for "design thinking" [10].

Consequently, DT applies to all phases that has to do with product existence and not solely to give out a depiction of the actual product [11][12]. The following is an example of a general definition of DT provided by [13] in the year 2012, and widely acknowledged and utilised Digital twin refers to "an integrated multi-physics, multi-scale, probabilistic simulation of a complex product". This simulation uses the most recent and accurate physical models, sensor readings, and other data to "mirror" the life of the product's "matching twin." A real-time reflection is one of the characteristics that help to define DT, and its definition also sheds light on this characteristic [14]. DT makes it possible for information about the real space to be studied in the physical world in a manner that is accurate and up to date while maintaining a high level of synchronisation and precision. The capability of DT to evolve on its own is also an essential feature of the methodology. In contrast to the actual and virtual worlds, the model used in the virtual environment is susceptible to continuous alterations made in real-time [15].

II. RESEARCH METHODOLOGY

This paper adopts a systematic literature review to highlight opportunities and restrictions in implementing digital twin technology in building projects. The research necessary material used in investigation was gotten via academic search engine database such as Google scholar, science direct, and academia. The analysis focused on the most recent 12 years' worth of publishing, extending from 2010 to 2022. The database allows users to modify the settings for how searches are performed. The search approach had electronic databases of Scopus. The search phrases "digital twin in construction," "semantic web," or "MIC" were used to rectify the

breakthroughs in the scope of research. The initial search suggested that 6890 papers were published on digital twin applications, out of which 4490 were published since 2019. Although much research has been published on digital twins, the authors further confined the search to journal and conference articles only in the "civil engineering domain" via Google scholar, science direct, and academia. More so, with deep investigation of the paper's abstract from the improved search, about 58 manuscripts for the systematic review analysis were realized. Reports were utilized as references from government and non-government organisations to assist the triangulation of the literature that was done in this study.

III. RESEARCH PROGRESSED

A. Examining Current Procedures

The most recent search of manuscripts through Scopus' digital libraries took place in May 2022. In the initial investigation, all papers were subjected to the selection criteria outlined in Section III. A total of 54 publications were called for further examination, after checking the headlines, titles, synopsis, and primary contents of filtered publications, a total of 19 articles were found. All of the publications found in line with DT usage in the civil works industry [16]. The first essay on the subject was published in 2018 the Journal articles and conference papers were evenly distributed among the selected publications. Fig. 1 depicts the disruption in publication timing and format of the recognized publications.

Three of the papers found looked at the use of DT not just during the building phase, however throughout the cycle of the project from inception to operation and maintenance. One of the goals of the DT application is to automate job planning and scheduling. Real-time data collection from a construction site allows for monitoring and appraisal of construction performance [17]. One study generated DT of a type of machinery to reduce production time and cost while other study formed DT as a type of construction tool for simulation to lessen manufacturing time and cost. Fig. 2 shows an outline of the various aims of DT applications spotted in publications.

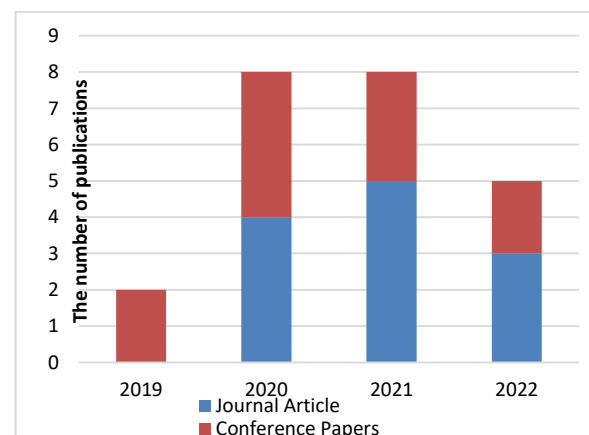


Fig. 1. Time of publishing and types of articles detected as of May 2022.

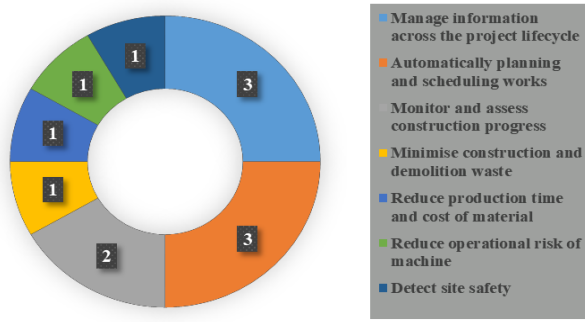


Fig. 2. An outline of the prospects for DT application within researched articles.

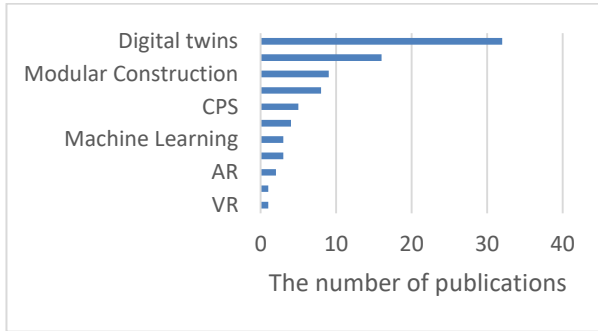


Fig. 3. Frequency with which DTs are published.

The number of articles with the keyword digital twins is 32 while, integrated modular construction related are nine articles. 15 studies cited "Building Information Modelling (BIM)," with 10 employing DT as an evolutionary representation and five considering the BIM system as DT. Only one article discusses "Wireless Sensor Network" (WSN) in connection with IoT for real-time data transfer. Three publications mentioned "cloud platform" for storing, analysing, processing, and combining data. Three publications reference "machine learning" for self-learning DT. Augmented and Virtual Reality can benefit from Digital Twins since they give immersive aim for viewing historical and real-time data [18]. Fig. 3 illustrates how often Digital twins, Modular construction, blockchain, BIM, IoT/sensor, WSN, cloud platform, machine learning, AR, and VR were referenced.

IV. FINDINGS

A comprehensive study was conducted to identify and synthesize the primary procedures for collecting, transmitting, modeling, integrating data, service for digital twin

technologies (Table I) (Fig. 4). Section 4A discusses the pros and difficulties associated with digital twins.

A. Assessing Digital Twin Applications

1) *Potential of the applications:* Digital twins aid in disaster preparation and asset management. Digital twin models help stakeholders comprehend obstacles and constraints, such as predicting the status of a physical asset, forecasting energy usage while preserving environmental norms, and monitoring structural infrastructure health. AI algorithms and data analytics help digital twin models forecast or back cast. The future status of an asset may be compared to the intended state. Digital twins help stakeholders choose low-carbon, clean-energy designs. Iterative design methods can promote generative design while addressing cost, schedule, and environmental benefits. Real-time data updates monitor construction workers for body segment hazards. Real-time reports on installation activity are needed in dangerous regions like nuclear power plants to protect personnel and limit human intervention. Digital twin models use data like population preferences and material usage priorities for urban planning and design. Using citizen or end-user data, digital twin models could connect people to buildings/cities.

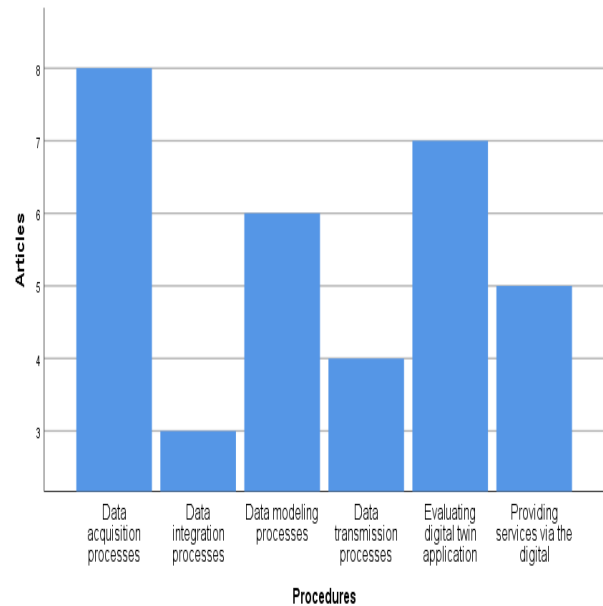


Fig. 4. Bar chart for the findings.

TABLE I. DIGITAL TECHNOLOGY USAGE AND ESTABLISHMENT

Data acquisition	Data transmission	Digital modeling	Data/model integration	Service/objective	Reference
Sensors, RFID	Internet	Geometric model (3D)	Virtual platform	Simulation model	[19]
GIS, BIM data	Internet	CIMs, BIMs	Cloud computing	Controlling and monitoring physical site	[20]
Real-time sensor	Internet	BIMs	Data analysis	Detection of maintenance and operation	[21]
Geospatial dataset	Internet	BIMs	Big data	Simulation for urban	[22]
IOT sensor	Internet	Thermography map	Machine learning algorithm	Renovations for energy planning	[23]

2) *Restriction of digital twins in MIC*: Herein, some of DT challenges cannot be overlooked. According to the literature, data integration is vital to link digital twin to physical assets. Surprisingly, integrating data from different technologies is difficult. DT data is housed in various systems which make it hard to compare with machines, for instance the heterogeneous data source. Data base terminology varies by system and maintaining feedback loops requires data synchronising. Extraction or modification of data reduces quality. DT requires real-time data transfer, which is expensive and time consuming. Digital twin models are subject to digital threats; therefore, data security and privacy is crucial [24]. Digital twins need uniform data transmission and sharing standards. Sensor installation and maintenance rarely concern construction personnel. Construction sensors can be stolen. These devices must save, filter, and match data with BIM models. Construction assets need lifecycle maintenance to survive decades. Construction is complicated and involves many stakeholders. Thus all project members must work together. Multiple stakeholders may impede asset and digital twin model building. Digital twins require experienced builders, money, and high-powered technologies.

3) *Impact of digital twins on industry*: A digital twin assists the production companies to grow its productivity, pliability and capability. They also improve on efficiencies without undermining the involvement of human by utilizing data-driven decision-making. Production system upgraded by digital simulation can be both cost and time-effective. Most institutions reproduce as well as examine every detail of production before it starts [25]. Manufacturers have the tendency to investigate unlike materials, colours and textures preliminary to production along visualization tools. Using a digital twin gives a better blueprint and tool of implementation to construct synthetic scenarios that are useful as tools for training and management continuation. These provide one with clarity to operate machines, systems and processes. Technically, it functions as a communication tool as it enhances the identification of variables one would like to observe. More so, digital twin can foretell about the state of the future of its real life counterpart. It has the ability to bring about what-if scenarios that are dangerous or cost-prohibitive in reality. This system can be perturbed to create unexpected scenarios and examine the system's response and the appropriate mitigation techniques. A digital twin is the only way to do this examination without putting the real asset at risk. Transparency can be improved by making real-time information available and automating the reporting process [26].

4) *Cost of time should be appropriate for model development*: To ensure the viability and acceptance of technology developments in building, the underlying causes must be effectively addressed. Low productivity, bad industrial image, low predictability, structural fragmentation,

and lack of Research and development (R&D) and investment in innovation are some of the industry's issues affecting the construction sector; however investment in DT can help alleviate some of them [27]. Digital twins require experienced builders, money, and high-powered technologies.

Once the DT requirement is acknowledged, the initial building pace of the digital twin is generally achieved already, as these demands serve as the basis for determining which programs should be utilized, what platforms are necessary, how much it would cost, and how difficult the digital twin will be. Time and money can be saved, as no processes have to occur physically for the identification to take place; it is known ahead of time what works and what doesn't. Money can be realized first via time; time is not returnable or stored, and by optimizing the relevant model, money is automatically saved. Obviously, several producing companies have only defined capacity of machine, and every job needs to be scheduled within the scope of the machine. Interestingly, through the use of digital twin, companies like mentioned above can pre-schedule their jobs to perfect the use of money and thus optimise the availability of time to be proportional with the mode; growth as much as possible. Notably, as time turn out to be sole driver, costs are diminished as hourly rates stand the same [28].

V. CONCLUSION

This paper gives the outline of academic publications connected to DT applications within industries, specifically in the building/manufacturing industry. The purpose of this paper is to evaluate the potential and restrictions of Digital twin integration for MiC. Through literature review, a total of 19 articles were found. All of the publications found are connected to the use of DT in the building construction industry. The primary goal of DT is to enable the manufacturing process to be optimized by incorporating seized time-sensitive information for efficient tracking. Additionally, it aims to incorporate flexibility and customization in all production methodologies and variables while maintaining complete authority. In cause of the literature gap findings, we were able to lay emphasis on Potential of digital twins in MIC, Restriction of digital twins in MIC, Impact of digital twins on industry, and Cost of time which should be appropriate for model development.

However, in the future, the applications for digital twins should be thoughtful of upcoming literature review and meta-analysis of publications on DT applications in the manufacturing area through more electronic databases; have collaboration in middle of the stakeholders to guarantee that the appropriate amount of time is spent on model development. Construction industry should primarily emphasize on the education and growth of qualified personnel, particularly for digital twin applications. During the decision-making process, the linked parties should also consider the matter of budget planning as well as to centre on group discussion or meetings with academics and practitioners in the areas of DT and modular building.

ACKNOWLEDGMENT

I am very grateful to my faculty unit for their encouragement.

CONSENT FOR PUBLICATION

The authors grant International Journal of Advanced Computer Science and Applications the consent to publish the manuscript.

DISCLOSURE

The author declares no conflict of interest.

FUNDING STATEMENT

This research has not been funded by any source.

REFERENCES

- [1] Lee, J., Kao, H. A., & Yang, S. (2014). Service innovation and smart analytics for industry 4.0 and big data environment. *Procedia cirp*, 16, 3-8.
- [2] Douthwaite, J. A., Lesage, B., Gleirscher, M., Calinescu, R., Aitken, J. M., Alexander, R., & Law, J. (2021). A modular digital twinning framework for safety assurance of collaborative robotics. *Frontiers in Robotics and AI*, 8.
- [3] Bertram, N., Fuchs, S., Mischke, J., Palter, R., Strube, G., & Woetzel, J. (2019). Modular construction: From projects to products. McKinsey & Company: Capital Projects & Infrastructure, 1-34.
- [4] Biesinger, F., Meike, D., Kraß, B., & Weyrich, M. (2018, September). A case study for a digital twin of body-in-white production systems general concept for automated updating of planning projects in the digital factory. In 2018 IEEE 23rd international conference on emerging technologies and factory automation (ETFA) (Vol. 1, pp. 19-26). IEEE.
- [5] Lu Q., Xie X., Parlikad A.K. and Schooling J.M. Digital twin-enabled anomaly detection for built asset monitoring in operation and maintenance. *Automation in Construction*, 118(2020): 103277, 2020.
- [6] Census and Statistics Department. GDP by major economic activity. Online: <https://www.censtatd.gov.hk/hkstat/sub/sp250.jsp?tableID=036&ID=0&productType=8#N3>, Access: 06/05/2020.
- [7] Chief Executive. The Chief Executive's 2017 Policy Address: Make Best Use of Opportunities, Develop the Economy, Improve People's Livelihood, Build an Inclusive Society. Online: <https://www.policyaddress.gov.hk/jan2017/eng/pdf/PA2017.pdf>, Access: 06/05/2020.
- [8] Chief Executive. The Chief Executive's 2018 Policy Address: Striving Ahead, Rekindling Hope. Online: <https://www.policyaddress.gov.hk/2018/eng/pdf/PA2018.pdf>, Access: 06/05/2020.
- [9] Shafto, M., Conroy, M., Doyle, R., Glaessgen, E., Kemp, C., Le Moigne, J., & Wang, L. (2012). Modeling, simulation, information technology & processing roadmap. *National Aeronautics and Space Administration*, 32(2012), 1-38.
- [10] Negri, E., Fumagalli, L., & Macchi, M. (2017). A review of the roles of digital twin in CPS-based production systems. *Procedia manufacturing*, 11, 939-948.
- [11] Glaessgen, E., & Stargel, D. (2012). The digital twin paradigm for future NASA and US Air Force vehicles. In 53rd AIAA/ASME/ASCE/AHS/ASC structures, structural dynamics and materials conference 20th AIAA/ASME/AHS adaptive structures conference 14th AIAA (p. 1818).
- [12] Kraft, E. M. (2016). The air force digital thread/digital twin-life cycle integration and use of computational and experimental knowledge. In 54th AIAA aerospace sciences meeting (p. 0897).
- [13] Tao, F., Cheng, J., Qi, Q., Zhang, M., Zhang, H., & Sui, F. (2018). Digital twin-driven product design, manufacturing and service with big data. *The International Journal of Advanced Manufacturing Technology*, 94(9), 3563-3576.
- [14] Tuegel, E. J., Ingrassia, A. R., Eason, T. G., & Spottswood, S. M. (2011). Reengineering aircraft structural life prediction using a digital twin. *International Journal of Aerospace Engineering*, 2011.
- [15] Jiang, Y., Li, M., Guo, D., Wu, W., Zhong, R. Y., & Huang, G. Q. (2022). Digital twin-enabled smart modular integrated construction system for on-site assembly. *Computers in Industry*, 136, 103594.
- [16] Jones, D., Snider, C., Nassehi, A., Yon, J., & Hicks, B. (2020). Characterising the Digital Twin: A systematic literature review. *CIRP Journal of Manufacturing Science and Technology*, 29, 36-52.
- [17] Lee, D., & Lee, S. (2021). Digital twin for supply chain coordination in modular construction. *Applied Sciences*, 11(13), 5909.
- [18] Lee, D., Lee, S. H., Masoud, N., Krishnan, M. S., & Li, V. C. (2021). Integrated digital twin and blockchain framework to support accountable information sharing in construction projects. *Automation in Construction*, 127, 103688.
- [19] Lim, K. Y. H., Zheng, P., & Chen, C. H. (2020). A state-of-the-art survey of Digital Twin: techniques, engineering product lifecycle management and business innovation perspectives. *Journal of Intelligent Manufacturing*, 31(6), 1313-1337.
- [20] Zhang, C., Sun, Q., Sun, W., Mu, X., & Wang, Y. (2021). A construction method of digital twin model for contact characteristics of assembly interface. *The International Journal of Advanced Manufacturing Technology*, 113(9), 2685-2699.
- [21] Camposano, J. C., Smolander, K., & Ruijpo, T. (2021). Seven metaphors to understand digital twins of built assets. *IEEE Access*, 9, 27167-27181.
- [22] Edirisinghe, R., & Woo, J. (2020). BIM-based performance monitoring for smart building management. *Facilities*.
- [23] Greif, T., Stein, N., & Flath, C. M. (2020). Peeking into the void: Digital twins for construction site logistics. *Computers in Industry*, 121, 103264.
- [24] Hasan, S. M., Lee, K., Moon, D., Kwon, S., Jinwoo, S., & Lee, S. (2022). Augmented reality and digital twin system for interaction with construction machinery. *Journal of Asian Architecture and Building Engineering*, 21(2), 564-574.
- [25] Pärn, E. A., & de Soto, B. G. (2020). Cyber threats and actors confronting the Construction 4.0. *Construction 4.0*, 441-459. [20] Anshul Agarwal. (2022). How Digital Twin is transforming the manufacturing industry. <https://www.industr.com/en/how-digital-twin-is-transforming-the-manufacturing-industry-2637054>.
- [26] Rasheed, A., San, O., & Kvamsdal, T. (2020). Digital twin: Values, challenges and enablers from a modeling perspective. *Ieee Access*, 8, 21980-22012.
- [27] Opoku, D. G. J., Perera, S., Osei-Kyei, R., & Rashidi, M. (2021). Digital twin application in the construction industry: A literature review. *Journal of Building Engineering*, 40, 102726.
- [28] Anje Marjorie Anderson, Andre Van der Merwe. Time-driven activity-based costing related to digital twinning in additive manufacturing. *The South African journal of industrial engineering*. Vol 32, No 1 (2021).

SuffixAligner: A Python-based Aligner for Long Noisy Reads

Zeinab Rabea^{1*}, Sara El-Metwally^{2*}, Samir Elmougy^{3*}, M. Z. Rashad^{4*}

Computer Science Department, Faculty of Computers and Information, Mansoura University, Mansoura 35516, Egypt

Abstract—Third-generation sequencing technologies have revolutionized genomics research by generating long reads that resolve many computational challenges such as long genomics variations and repeats. Mapping a set of sequencing reads against a reference genome is the first step of many genomic data analysis pipelines. Many mapping/alignment tools are introduced and always made different compromises between the alignment accuracy and the resource usage in terms of memory space and processor speed. SuffixAligner is a python-based aligner for long noisy reads generated from third-generation sequencing machines. SuffixAligner follows the seed extending approach and exploits the nature of the biological alphabet that has a fixed size and a predefined lexical ordering to construct a suffix array for indexing a reference genome. A suffix array is used to efficiently search the indexed reference and locate the exactly matched seeds among the reads and the reference. The matched seeds are arranged into windows/clusters and the ones with the maximum number of seeds are reported as candidates for mapping positions. Using real data sets from third-generation sequencing experiments, we evaluated SuffixAligner against lordFAST, BWA, GEM3, and Minimap2, in which the results showed that SuffixAligner mapped more reads compared to the other compared tools. The source code of SuffixAligner is available at: <https://github.com/ZeinabRabea/SuffixAligner>.

Keywords—Long reads sequencing; reads mapping; suffix array; alignment; seed extending; LF mapping

I. INTRODUCTION

Sequencing machines have generated a flood of biological data known as reads. They have revolutionized over three generations of technologies that play a key role in the data volume, read length and accuracy, sequencing cost, and speed. Third-generation sequencing technologies, such as Pacific Biosciences (PacBio) and Oxford Nanopore, produce a high throughput of longer reads with higher error rates than Illumina sequencing machines' short reads [1]. Reads mapping/alignment is the cornerstone in any sequence analysis pipeline and implies finding the nearly matched locations of each read in the reference genome/transcriptome tolerating the mismatches due to sequencing biases and errors. Since the reads mapping/alignment is a complex and resources intensive process, efficient algorithms and data structures are introduced to complete it reasonably [2-4].

Mapping algorithms utilize two data structures: hash tables and suffix/prefix tries to handle the long reads generated from the third-generation sequencing machines [5]. The seed and extended approach are used in most sequence aligners and rely on extracting a set of matched seeds among the reference and read and index them in a hash table [6]. Then, the seeds are

extended to find the optimal gapless alignment. Examples of tools that utilize hash tables for indexing a set of seeds are [7] Blast [8], NanoBLASTER [9], LAST [10], and Minimap [11, 12]. Suffix tires based algorithms divide the mapping problem into two steps: 1) finding an exact match between the database and the query read using one of the suffixes stored in suffixes tries, and 2) grouping these hits of exact matches to create the final inexact alignment [13].

Genome indexing is the first step in any mapping data analysis framework in which it aims search quickly and efficiently the reference genome for matching patterns (i.e., reads). Many efficient data structures are introduced to build an index for a reference genome, such as suffix arrays [14, 15], hash tables [16], suffix trees [17], and bloom filters [18]. There is no standard procedure for creating the optimal index file; this depends on how the information is organized and how he intends to access it. The second step is the mapping/alignment process that uses the genome indexing to query the reference and search for the reads and align it to nearly matched locations in the genome[19].

In this paper, we introduce SuffixAligner as a long-read aligner that utilizes a modified version of the suffix array construction algorithm for limited-size alphabets such as DNA/RNA and exploits the nature of their corresponding lexicographical ordering. Suffix array is used in the reference indexing stage to compute Burrows–Wheeler transforms. Reads mapping stage follows the seed and extends approach by extracting a set of matched hits called k -mers among reads and the indexed reference. These k -mers are used as seeds to find long stretches of matches extended in both directions of the sequencing read. The regions between the matched seeds among reads and the reference are aligned using dynamic programming approach.

This paper is organized as follows: Section II demonstrates a previously related work, Section III presents the methodology behind the SuffixAligner and describes its three basic stages, Section IV evaluates the SuffixAligner alignment results against four benchmarking alignment tools using two real datasets, and Section V concludes our paper and provides some future insights.

II. LITERATURE REVIEW

Suffix array plays a key role in indexing large-scale genomes and most aligners that dominate the bioinformatics field rely on suffix array and its enhanced versions to complete the mapping/alignment task [20]. It is a list of indexes corresponding to the sorted suffixes of a given sequence in the

lexical order. Its construction algorithm is a resource-expensive process, in which there is always a trade-off between the algorithm running time and its memory usage. Most of the suffix array construction approaches for indexing a reference genome/transcriptome should consider that the sequencing bases have a limited size alphabet (i.e. four letters of DNA/RNA alphabets) and utilize the existing integer-based method for suffix array construction and manipulation [21, 22]. Examples of mapping algorithms that rely on suffix arrays for indexing and searching a reference genome are Bowtie [23], BWA [24], GEM3 [25], and lordFAST [26] (see Table I). The Burrows–Wheeler (BWT) transform is a cornerstone in many alignment/mapping tools developed for short reads and extended to long-read sequencing data. Burrows–Wheeler transform can also be generated from a suffix array and is used to build a compressed and efficient version of a reference genome [27].

TABLE I. DESCRIPTION OF BENCHMARKING ALIGNMENT TOOLS: BWA, GEM3, LORDFAST, AND MINIMAP2

Aligner	Description	Ref
BWA	BWA is based on the FM-index and utilizes the seed-and-extend approach for reads mapping. BWA has three different algorithms: <ul style="list-style-type: none"> • BWA-backtrack (Illumina short sequencing read) • BWA-SW (long reads ranging from 70bp to 1Mbp) • BWA-MEM(long reads ranging from 70bp to 1Mbp) 	[24]
GEM3	GEM3 is based on the FM-index. Best results are produced when the read length up to 1k bases. GEM3 uses a Multithread mode to speed up the running time	[25]
lordFAST	LordFAST is based on the FM-index and utilizes the seed-and-extend approach for reads mapping	[26]
Minimap2	Minimap2 collects the minimizers from the reference and indexes them into a hash table using a locally sensitive hashing technique. It utilizes the seed-and-extend approach for reads mapping. The total length of all reads should not exceed the 2 billion bases	[11, 12]

III. METHOD

SuffixAligner is designated for reads generated by third-generation sequencing machines. It aims to find the ideal location for each read in the indexed reference genome using a suffix array. SuffixAligner has three stages: 1) Finding the matched seeds between the genome and read, 2) Determining the part of the genome with a large number of seeds (identifying a matching window), 3) Using Needleman [28] algorithm to align the regions between the matched seeds (see Fig.1).

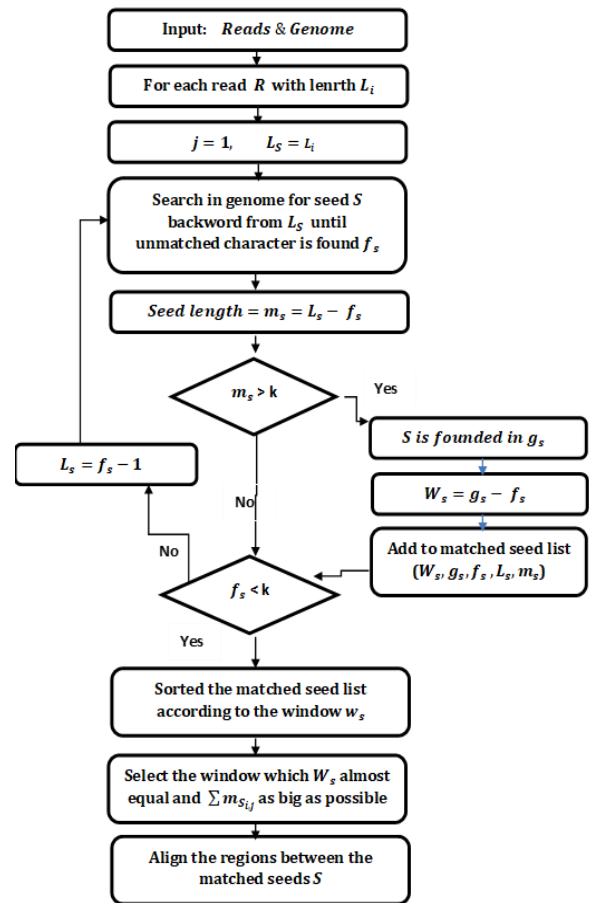


Fig. 1. The SuffixAligner proposed approach for mapping long noisy reads.

The reference genome should be indexed first to compute the matched seeds among reads and the reference genome. In the genome indexing stage, SuffixAligner relies on a suffix array construction algorithm proposed in [29] and exploits the nature of biological sequencing data alphabets such as DNA/RNA as it has limited size letters with a pre-defined lexicographical order. The suffix array is constructed for DNA alphabets incrementally, knowing that there are only four letters with the pre-defined sorted order {A, C, G, T}. Each portion of the suffix array will correspond to the set of indexes starting with one of the DNA alphabets and will be used as a seed to compute the next portion corresponding to the next DNA alphabet in the lexical order. The Burrows–Wheeler transform is constructed from the computed suffix array and represents a compressed version of the indexed reference genome. This paper applies the Last-to-First Mapping (LF-array), a mapping function from the last column of the Burrows–Wheeler Matrix (BWM), to the first column of the BWM. To compute LF-array easily and efficiently, BWT is extracted from the last column of BWM. Also, the first column (F_c) is an alphabetical arrangement of the elements of BWT. Instead of storing the first column of BWM, the frequency of each DNA alphabet in BWT is computed along with its start index in BWT. The LF-array is calculated by adding the rank of each DNA alphabet with its starting index (see Fig. 2). LF-array efficiently searches the reference genome for the exact matches (seeds) extracted from the sequencing reads.

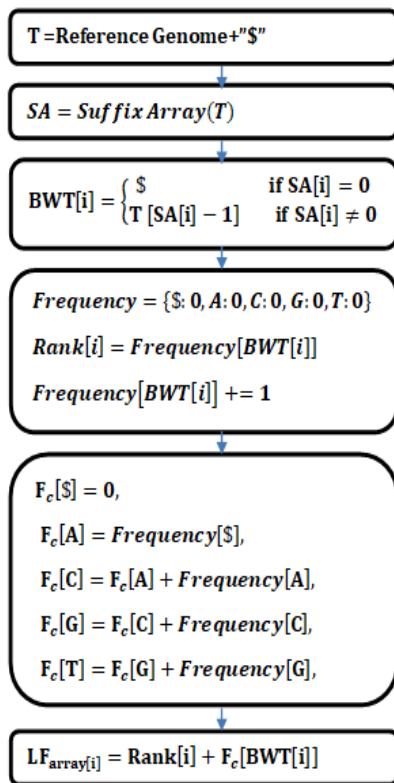


Fig. 2. The Last-to-first mapping (LF-array) computation from the burrows-wheeler transform.

The search starts from the read ending position towards the starting position and stops when mismatches occur. A list of matching seeds between the read and the reference genome, where the seed length of more than 10bp is considered as a valid matched seed. The list contains the starting position f_s of each matched seed in the reference genome g_s , and the ending position l_s and the seed length is m_s which equals $m_s = l_s - f_s$. The starting position of the expected mapping window of each seed is determined by $w_s = g_s - f_s$. The list of matched seeds is sorted according to the window w_s . The optimal window for mapping the read against a reference genome will be chosen according to the total number of matched seeds in the window. The window with the largest number of seeds will be selected. The regions between the matched seeds in the selected window are aligned using a dynamic programming method such as Needleman algorithm.

IV. EXPERIMENTATIONS AND RESULTS

In this paper, we evaluate the four alignment tools: BWA (bwa-0.7.17) [30], GEM3 [25], lordFAST[26], and Minimap2 [11] against our proposed SuffixAligner using real datasets generated by MinION Oxford Nanopore sequencing technology from two different bacterial strains: Flavobacterium columnare and Aeromonas veronii (see Table II).

We measured the quality of alignment by measuring the alignment rate [31]. The alignment rate is calculated by dividing the number of mapped reads by the total number of reads. Samtools [32] analyzed the SAM files and reports the total number of primary and secondary alignments produced by

alignment tools. When a read maps ambiguously to multiple locations, only one of the read alignments is considered as the primary one, and the others are reported as secondary alignments. Supplementary reads are those in which parts of the reads match one location in the genome while other parts match another and often appear in long reads. All the experiments in this work were conducted on a workstation running on Ubuntu Linux with a 3.70 GHz Intel(R) Xeon(R) CPU E5-1620 v2 processor, 64 GB of RAM, and 128 GB SDD hard disk.

TABLE II. CHARACTERISTICS OF THE DATA SET USED EXPERIMENTS

Genome Name	Flavobacterium Columnare	Aeromonas Veronii
Genome Length	3221278	4559061
NCBI	txid996	txid654
Sequencing runs	SRR7449868	SRR7449790
Number Of reads	1142	3994
Number Of Bases	146.1M	644.5M
Max length in reads	62330	53937
Min length in reads	284	279

We evaluated the alignment rate of BWA, GEM3, lordFAST, Minimap2, and SuffixAligner for the real Oxford Nanopore sequencing runs SRR7449868 and SRR7449790. The characteristics of the dataset are shown in Table I. As shown in this table, SuffixAligner has the best Alignment rate. It primary mapped 94% of the total number of reads in the run SRR7449868 of the Flavobacterium columnare reference genome and primary mapped 90% of the total number of reads in the run SRR7449790 of Aeromonas veronii reference genome.

SuffixAligner searches for a one room for every read compared to a reference genome. It considers every read's best mapping position and ignores the other mapping locations. It neglects the secondary and supplementary mapping locations mapped reads (Tables III and IV).

Fig. 3 shows the total number of mapped reads with lengths less than 500bp and greater than 8000bp in the case of the sequencing run SRR7449868. For reads with lengths less than 1000bp, SuffixAligner mapped more reads than other alignment tools; the best results are produced by SuffixAligner, BWA, GEM3, and Minimap2 sequentially. lordFAST ignores reads with lengths less than 1000bp. For reads with lengths greater than 1000bp, SuffixAligner and lordFAST have approximately the same alignment rate then BWA, Minimap2, and GEM3 sequentially.

Fig. 4 shows the total number of mapped reads with lengths less than 500bp and greater than 8000bp in the case of the sequencing run SRR7449790. For reads with lengths less than 1000bp, SuffixAligner mapped more reads than other alignment tools, and the best results are produced by SuffixAligner, BWA, GEM3, and Minimap2 sequentially. lordFAST ignores the reads with lengths less than 1000bp. lordFAST has the best alignment rate with reads lengths between 1000bp and 2000bp, while the SuffixAligner has the best alignment rate with other reads lengths.

TABLE III. EVALUATION OF BWA, GEM3, LORDFAST, MINIMAP2, AND SUFFIXALIGNER ON SEQUENCING RUN SRR7449868 (TOTAL NUMBER OF BASES IS 146.1M)

Metrics Tools	Total	Supplementary	Secondary	Primary	Total mapped		Primary mapped			Primary unmapped		
	#	#	#	#	#	%	#	Bases(kb)	%	#	Bases(Kb)	%
BWA	1325	183	0	1142	1222	92%	1039	4027.29	91%	103	87.88	9%
GEM3	1142	0	0	1142	1004	88%	1004	3999.73	88%	138	115.44	12%
lordFAST	1490	126	222	1142	820	55%	472	3723.04	41%	670	392.13	59%
Minimap2	1518	137	239	1142	1368	90%	992	3998.42	87%	150	116.76	13%
SuffixAligner	1142	0	0	1142	1074	94%	1074	4072.21	94%	68	42.96	6%

TABLE IV. EVALUATION OF BWA, GEM3, LORDFAST, MINIMAP2, AND SUFFIXALIGNER ON SEQUENCING RUN SRR7449790 (TOTAL NUMBER OF BASES IS 644.5M)

Metrics Tools	Total	Supplementary	Secondary	Primary	Total mapped		Primary mapped			Primary unmapped		
	#	#	#	#	#	%	#	Bases(kb)	%	#	Bases(Kb)	%
BWA	1325	183	0	1142	1222	92%	1039	4027.29	91%	103	87.88	9%
GEM3	1142	0	0	1142	1004	88%	1004	3999.73	88%	138	115.44	12%
lordFAST	1490	126	222	1142	820	55%	472	3723.04	41%	670	392.13	59%
Minimap2	1518	137	239	1142	1368	90%	992	3998.42	87%	150	116.76	13%
SuffixAligner	1142	0	0	1142	1074	94%	1074	4072.21	94%	68	42.96	6%

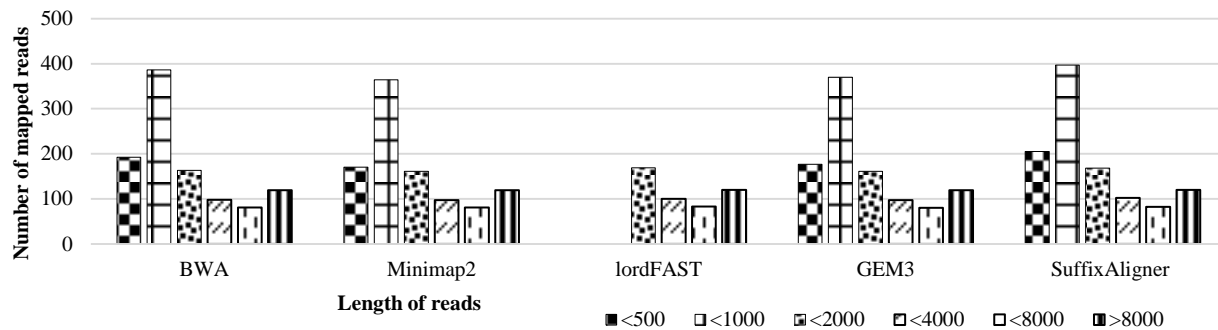


Fig. 3. The total number of mapped reads according to different read lengths in the real sequencing run (i.e., id SRR7449868) with the total number of reads equal to 1142.

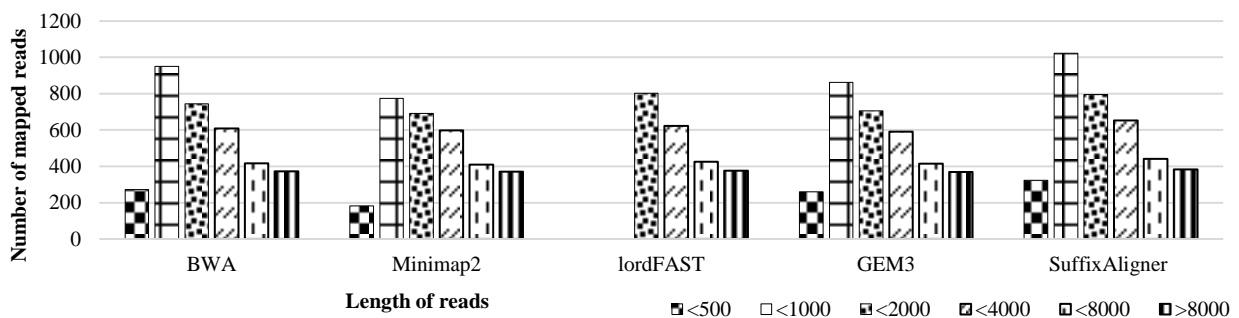


Fig. 4. The total number of mapped reads according to different read lengths in the real sequencing run (i.e., id SRR7449790) with total number of reads equal to 3994.

SuffixAligner maps more reads than lordFAST, BWA, GEM3, and Minimap2 and has the best alignment rate. SuffixAligner and lordFAST have approximately the same alignment rate with the longest read lengths. Also, SuffixAligner maps short reads and reads with lengths greater than 8000bp.

V. CONCLUSIONS

In this paper, SuffixAligner is proposed for reads produced by third-generation sequencing machines such as PacBio and Oxford Nanopore. It relies on a specific type of suffix array constructed for the limited-size alphabets, such as the four-size alphabets of genomic data. It has three basic stages: indexing a reference genome using a suffix array, mapping a set of sequencing reads using the seed and extend approach, and

aligning the regions between the matched seeds using a dynamic programming approach. It mapped more reads compared to the other alignment tools and has the best alignment rate accordingly.

SuffixAligner has a long execution running time since it aligns one read at a time. So, as a future work, distributing the computation of mapping reads among different machines to increase the mapping speed and to reduce the memory usage on a single computing node. Also, SuffixAligner may be used to improve the alignment results by mapping only unmapped reads produced by another aligner.

REFERENCES

- [1] C. Delahaye and J. J. P. o. Nicolas, "Sequencing DNA with nanopores: Troubles and biases," vol. 16, no. 10, p. e0257521, 2021.
- [2] C. Marchet, C. Boucher, S. J. Puglisi, P. Medvedev, M. Salson, and R. J. G. R. Chikhi, "Data structures based on k-mers for querying large collections of sequencing data sets," vol. 31, no. 1, pp. 1-12, 2021.
- [3] T. Hu, N. Chitnis, D. Monos, and A. J. H. I. Dinh, "Next-generation sequencing technologies: An overview," vol. 82, no. 11, pp. 801-811, 2021.
- [4] V. Bansal and C. Boucher, "Sequencing Technologies and Analyses: Where Have We Been and Where Are We Going?," (in eng), iScience, vol. 18, pp. 37-41, Aug 30 2019.
- [5] C. Mingard, J. Wu, M. McKeague, and S. J. J. C. S. R. Sturla, "Next-generation DNA damage sequencing," vol. 49, no. 20, pp. 7354-7377, 2020.
- [6] A. Sarkar, S. Banerjee, and S. J. I. T. o. V. L. S. I. S. Ghosh, "An Energy-Efficient Pipelined-Multiprocessor Architecture for Biological Sequence Alignment," vol. 28, no. 12, pp. 2598-2611, 2020.
- [7] N. Pavlovikj, E. N. Moriyama, and J. S. Deogun, "Comparative analysis of alignment tools for nanopore reads," in 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2017, pp. 169-174: IEEE.
- [8] M. Johnson, I. Zaretskaya, Y. Raytselis, Y. Merezhuk, S. McGinnis, and T. L. Madden, "NCBI BLAST: a better web interface," *Nucleic Acids Research*, vol. 36, no. suppl_2, pp. W5-W9, 2008.
- [9] M. R. Amin, S. Skiena, and M. C. Schatz, "NanoBLASTer: Fast alignment and characterization of Oxford Nanopore single molecule sequencing reads," in 2016 IEEE 6th International Conference on Computational Advances in Bio and Medical Sciences (ICCBMS), 2016, pp. 1-6: IEEE.
- [10] S. M. Kielbasa, R. Wan, K. Sato, P. Horton, and M. C. J. G. r. Frith, "Adaptive seeds tame genomic sequence comparison," vol. 21, no. 3, pp. 487-493, 2011.
- [11] H. Li, "Minimap2: pairwise alignment for nucleotide sequences," *Bioinformatics*, vol. 34, no. 18, pp. 3094-3100, 2018.
- [12] H. Li, "Minimap and miniasm: fast mapping and de novo assembly for noisy long sequences," *Bioinformatics*, vol. 32, no. 14, pp. 2103-2110, 2016.
- [13] A. J. T. J. o. C. Chayapathi and M. Education, "Survey and Comparison of String Matching Algorithms," vol. 12, no. 12, pp. 1471-1491, 2021.
- [14] A. Amir and I. J. a. p. a. Boneh, "Update query time trade-off for dynamic suffix arrays," 2020.
- [15] F. A. Louza, S. Gog, G. P. Telles, F. A. Louza, S. Gog, and G. P. J. C. o. F. D. S. f. S. Telles, "Induced suffix sorting," pp. 23-40, 2020.
- [16] T. Maier, P. Sanders, and R. J. A. T. o. P. C. Dementiev, "Concurrent hash tables: Fast and general (?)," vol. 5, no. 4, pp. 1-32, 2019.
- [17] M. M. A. Aziz, P. Thulasiraman, and N. Mohammed, "Parallel generalized suffix tree construction for genomic data," in Algorithms for Computational Biology: 7th International Conference, AICoB 2020, Missoula, MT, USA, April 13-15, 2020, Proceedings 7, 2020, pp. 3-15: Springer.
- [18] M. Najam, R. U. Rasool, H. F. Ahmad, U. Ashraf, and A. W. J. B. r. i. Malik, "Pattern matching for DNA sequencing data using multiple bloom filters," vol. 2019, 2019.
- [19] Y. Wu, B. Lao, X. Ma, and G. Nong, "An improved algorithm for building suffix array in external memory," in Parallel Architectures, Algorithms and Programming: 10th International Symposium, PAAP 2019, Guangzhou, China, December 12-14, 2019, Revised Selected Papers 10, 2020, pp. 320-330: Springer.
- [20] B. Lao, Y. Wu, G. Nong, and W. H. J. I. T. o. C. Chan, "Building and checking suffix array simultaneously by induced sorting method," vol. 71, no. 4, pp. 756-765, 2021.
- [21] A. Das and R. Baruri, "All Pairs Suffix-Prefix Matches using Enhanced Suffix Array," in 2020 International Conference on Smart Electronics and Communication (ICOSEC), 2020, pp. 815-822: IEEE.
- [22] F. A. Louza, S. Gog, and G. P. Telles, "Optimal suffix sorting and LCP array construction for constant alphabets," *Information Processing Letters*, vol. 118, pp. 30-34, 2017.
- [23] M. Naghibzadeh, S. Babaei, B. Behkmal, M. J. I. J. o. I. Hatami, and C. T. Research, "The Efficient Alignment of Long DNA Sequences Using Divide and Conquer Approach," vol. 14, no. 3, pp. 48-56, 2022.
- [24] S. Liu, Y. Wang, W. Tong, and S. J. B. Wei, "A fast and memory efficient MLCS algorithm by character merging for DNA sequences alignment," vol. 36, no. 4, pp. 1066-1073, 2020.
- [25] S. Marco-Sola, M. Sammeth, R. Guigó, and P. J. N. m. Ribeca, "The GEM mapper: fast, accurate and versatile alignment by filtration," vol. 9, no. 12, pp. 1185-1188, 2012.
- [26] E. Haghshenas, S. C. Sahinalp, and F. J. B. Hach, "lordFAST: sensitive and fast alignment search tool for long noisy read sequencing data," vol. 35, no. 1, pp. 20-27, 2019.
- [27] L. Egidi, F. A. Louza, G. Manzini, and G. P. Telles, "External memory BWT and LCP computation for sequence collections with applications," *Algorithms Mol Biol*, vol. 14, p. 6, 2019.
- [28] S. J. M. B. Needleman, "Wunsch C (1970) J," vol. 48, pp. 444-453.
- [29] Z. Rabea, S. El-Metwally, S. Elmougy, and M. Zakaria, "A fast algorithm for constructing suffix arrays for DNA alphabets," *Journal of King Saud University - Computer and Information Sciences*, 2022/05/06/ 2022.
- [30] H. Li and R. J. B. Durbin, "Fast and accurate long-read alignment with Burrows-Wheeler transform," vol. 26, no. 5, pp. 589-595, 2010.
- [31] M. R. Stratton, P. J. Campbell, and P. A. Futreal, "The cancer genome," *Nature*, vol. 458, no. 7239, pp. 719-724, 2009.
- [32] P. Danecek et al., "Twelve years of SAMtools and BCFtools," vol. 10, no. 2, p. giab008, 2021.

Scrum: A Systematic Literature Review

Adrielle Cristina Sassa¹, Isabela Alves de Almeida², Tábata Nakagomi Fernandes Pereira³, Milena Silva de Oliveira⁴
Institute of Integrated Engineering, Federal University of Itajuba (Unifei), Itabira, Brazil^{1,2,3}
Institute of Industrial Engineering and Management, Federal University of Itajuba (Unifei), Itabira, Brazil⁴

Abstract—This study presents a Systematic Literature Review on an agile project management tool. The study offers a brief comparison between traditional and agile project management methodologies. Their respective concepts and characteristics are laid out to highlight and explain their main differences. The agile methods include quantitative and qualitative data, showing Scrum framework characteristics. This study highlights the importance of project management in function of its emergence as a response to problems encountered during improperly conducted projects. Furthermore, this study provides relevant information for professionals in the Industrial Engineering area and computer science. The results allowed us to conclude that Scrum is an agile framework for empirical-based project development; it was developed in the 1990s by Jeff Sutherland. It is a flexible and adaptable methodology. Scrum research peaked in 2020, and continues to be studied, mainly in the field of computer science. Finally, Brazil is well-positioned in third place for works published.

Keywords—Project management; agile methods; agile manifesto; scrum

I. INTRODUCTION

Project Management (PM), as a practice, has grown rapidly worldwide, and is now globally recognized for its strategic competency for organizations [1]. Projects are employed to revamp business processes, support global customer-focused strategies, and coordinate information and decision-making flows across organizations [2].

PM is a set of management tools that allows a company to develop a set of skills, including individual knowledge and skills, all of which are aimed at controlling non-repetitive, unique and complex events, within a given time frame scenario, with predetermined costs and quality metrics [3].

Many projects are conducted improperly, and consequently, failures can occur. PM is a tool for conducting these projects successfully, without suffering losses. PM tools have been effective in achieving desired results within defined time frames and budgetary limitations within organizations. The author also states that PM is not only used for highly complex and costly large-scale projects [4].

Over the last decades, several design methods have been developed and adopted by project managers [5]. One is the Project Management Body of Knowledge Guide – (PMBOK®), which is the most influential book in the area [2]. The PMBOK® guide consists of standards that identify and conceptualize project management processes, knowledge areas, tools, and techniques [6].

The main PM methodologies can be classified as either traditional (predictive) or agile approaches. Traditional

approaches, for the most part, follow PMBOK® principles, while agile methods are based on the Agile Manifesto, which unites agile method principles and characteristics [7].

The Agile Manifesto was developed amid troubled projects, given dissatisfaction caused by high costs and long scope periods, where software developers needed to restructure project methodologies to focus on programming development and project testing. In 2001, seventeen software professionals, including developers and managers, met in Utah, where the Agile Alliance was created, to create the Agile Manifesto, to make software development lighter, more flexible, and cheaper [8, 9].

Agile project management seeks fast planning with active participation from the whole team and from customer stakeholders at all stages of the project. This is done by fostering an environment of collaboration among the whole team, and by rapidly integrating changes during the project life cycle [8, 10]. Focus and values are what set agile methodologies apart from traditional methodologies. Agile focuses on people, and not on processes or algorithms. Also, less time is spent on documentation, and more on implementation and success [11].

Of the many agile project management methodologies, there are Extreme Programming (XP) created by Beck and Cunningham in 1999 [7, 12]. It is a methodology for projects with requirements that change frequently, for projects that use object-oriented development, for teams comprising up to 12 developers, and for projects that use incremental development [13]. Scrum is another agile project management methodology created by Schwaber and Sutherland in 1990 [14]. It is suitable for companies that conduct product development in environments characterized by uncertainty, self-organization, moderate control, and knowledge transfers [15].

Another important tool is Lean software, which applies principles from the Toyota Production System for software development [16]. The Toyota Production System is a production philosophy created by a Japanese automaker that seeks to eliminate any production processes that do not add product value, or so-called waste [17]. The Kanban methodology, proposed by David J., is defined as a framework for incremental process and system improvements within organizations [18], and seeks to improve processes, teams, and projects. It is useful for companies that are constantly seeking to improve, while also improving productivity and their customer relationships [19].

Scrum is focused on teamwork, to improve communication and enhance cooperation among team members, resulting in increased future productivity. Furthermore, Scrum allows for

short-term problem solving, reduced project risks, greater customer participation throughout processes, and delivers more functional products or services more frequently [20].

Scrum is used in complex projects that cannot predict everything that will happen. This methodology offers a framework, since it is an aggregation of practices that promote visibility, so Scrum practitioners can see all project stages, make corrections, and offer improvements that keep the project moving forward by focusing on achieving the goals [21].

Some of the advantages of using this methodology are adaptability, transparency, continuous feedback, continuous improvement, and motivation [22]. The author in [22] applied Scrum for PM in the naval industry, and concluded that "the responses to problems were more agile, since problems identified in the previous day were discussed the next day in a quick meeting". The author in [11] also proved these advantages by applying this methodology to a tech startup, and by conducting a survey, wherein "most people highlighted visible management, which is a technique whereby controls, as to what is being done, are easily visible to everyone".

In this sense, the general objective of this study is to carry out a Systematic Literature Review on the agile Scrum methodology, which has been studied since the 1990s, and is still one of the main methodologies used in several areas.

This study is structured in five sections. The first section contextualizes the study. The second section presents a brief theoretical framework. The third section details the research methodology used here. The fourth section applies the SLR method. The fifth section analyses the results. The paper finishes with the conclusions, followed by the bibliographic references.

II. METHODOLOGY

A. Systematic Literature Review (SLR)

Literature Review (LR) is a very relevant method for researchers who analyze numerous fields of study [23]. Many articles use more descriptive LR models, and do not follow a systematic approach. One other method that is widely used in exploratory research is Systematic Literature Review (SLR). SLR develops guidelines, since it encompasses searching, selecting, critically evaluating, and creating synopses of primary research results [24]. The authors in [25] and [26] developed structures that contain the main SLR steps, to standardize SLR as a method, and to avoid distortions in the research and data analyses. The author in [27] executed a paper to gather the critical success factors of motivator and demotivator of agile software development.

SLR is a branch of LR, but SLR goes far beyond gathering and discussing important scientific works on a topic of interest. SLR seeks to verify what works, and what does not work, within a given context, using an explicitly presented bibliographic database, so reproducibility is possible. The search and selection parameters for articles, and the reason for the inclusion or exclusion of these parameters, are also explained, along with the limitations that are found in the articles, and in the review itself [28].

One relevant SLR point is its systematic methodology, which allows for necessary critical analysis in carrying out all stages of research. For the purposes of this study, the method applied in the article was based on [29], since it offered a more recent reference, and proposed a structure that fit well with this present study. The proposal consists of three stages, as follows:

- 1) *Planning the Review (The Preparation Stage);*
- 2) *Conducting the Review (The Operational Stage)*
- 3) *Documenting the Review (The Information Stage).*

The first step, i.e., planning the review, consists of two steps:

- a) *specifying the research question; and*
- b) *developing the review protocol.*

In the first step, research questions that are relevant to the subject being studied are defined to guide the next steps. Next, a standard is created to search for documents that will be used to answer the questions from the previous step.

The second step leading to the review is divided into five steps. The first three steps focus on defining the database that will be used to answer the research questions from the previous step. Researchers must apply review protocol (identifying relevant research). Next, the documents that do not address the subject will be excluded (selecting primary studies), and finally, the remaining documents will be studied in-depth, to decide which documents will comprise the SLR database (assessing the quality of the study). In the next steps, all the necessary data and information that will be useful in answering the research questions must be extracted from the documents and synthesized.

Finally, the third step, i.e., documenting the review, uses the information obtained in the previous steps to answer the questions outlined in the first step to complete the SLR.

III. METHOD APPLICATION

The SLR steps, as proposed by [29], will be described here, starting with planning the review.

A. Planning the Review (Preparation Stage)

According to [29] the first SLR stage deals with review planning. The authors mention two initial points that are important for prepping the review stage as: a) Specifying the research question; and b) Developing a research protocol.

The author in [30] was used as a basis for specifying the research questions. The authors propose interesting questions that can be answered using SLR. Thus, the guiding questions from [30], and the questions themselves were elaborated for this study.

Q1. How many Scrum publications are there per year? Q2. Who are the main authors writing on Scrum? Q3. What are the main countries publishing on Scrum? Q4. Who are the main publishers on Scrum? Q5. Which major universities publish on Scrum? Q6. What are the main areas studying Scrum? Q7. What is Scrum's history? Q8. How is Scrum defined? Q9. What are Scrum's principles? Q10. What are the Scrum steps? Q11. What are the Scrum elements? Q12. Is Scrum expensive?

Q13. Is Scrum complex? Q14. How is Scrum scoped? Q15. How are the teams structured? Q16. What is the role of the project manager in Scrum? Q17. What are the advantages of Scrum? Q18. What are the disadvantages of Scrum? Q19. What are the future directions for Scrum?

Table I was created based on the above research questions, creating SLR categories. This categorization was meant to simplify the analyses, by taking the number of questions previously established into account. Furthermore, the categorization will make it easier to describe the results later.

TABLE I. SEARCH CATEGORIES

State of the Arts	Questions	Nature of the Research	Questions
Articles per Year	Q1	General concepts	Q7, Q8, Q9
Top 5 Authors	Q2	Structure	Q10, Q11
Top 5 Countries	Q3	Work Mode	Q12, Q13, Q14
Top 5 Publication Sources	Q4	Time	Q15, Q16
Main Universities	Q5	Advantages and Disadvantages	Q17, Q18
Top 5 areas of Study	Q6	Future Study	Q19

As to the second point pointed out by the authors, which consists of developing a research protocol, the data in this study will be obtained via a systematic search, to obtain bibliographies relevant to the subject in question. The Scopus® database was selected because it stands out given the number of data available, the quality of its resources, and its ease of use. The author in [31] state that the Scopus® database is one of the main sources of data for citations.

A search was performed using the term “Scrum” in the Scopus database. Then filters were applied, like “open access”, “article type document”, “English and Portuguese”. There were no restrictions as to the year or area of study, for possible future comparisons.

After the initial planning, defining the research questions, and forming the data collection for this study, it was preceded to the next SLR stage, which is conducting the review.

B. Conducting the Review (Operational Stage)

This stage seeks to effectively conduct the research, and is an operational stage. The first two points established by [29], are to identify relevant research and to select the main studies. On April 28, 2021, it was made a search in the Scopus® database, as defined in the previous step, to develop these two sub-steps.

The term “Scrum” was searched for in the Scopus® database, searching only the title. 999 articles were returned from the database. Then it was applied filters like “open access”, “article type document”, and “English and Portuguese”, to narrow the search and focus in on the subject of interest. After applying these filters, 77 articles were found.

After applying the filters, it was moved on to the third step, which is evaluating the quality of the articles. This sub-step consists of conducting a more in-depth analysis of the articles,

to identify which articles really contain the characteristics that were seeking in this study. The articles were analyzed by the authors by reading the titles, the abstracts, and the keywords, to exclude any articles that were not relevant to this subject. At the end of this analysis, 50 articles remained for a full reading.

27 articles were excluded from the 77 articles by reading the abstracts and analyzing the theme. “SCRUM” was found to be an abbreviation for certain medical questions, e.g., “Scalp and Cranium Radiation Therapy Using Modulation”, or “Studying Concussions in the Rugby Union using MicroRNAs”, and also for school-related questions, like “School Clinical Rugby Measure”, along with articles that relate scrum to Rugby, and these medical/sports-related works did not fit in this paper, so they were discarded.

After reading the articles, it was found that 9 articles did not include studies addressing Scrum for the following reasons:

- they did not present information about Scrum;
- they were no longer open access; or
- they were not found.

Therefore, they were removed from the analysis. Thus, the SLR focused on 41 articles to answer the research questions. Table II gives a summary of the filters that were used to select the 41 articles in this SLR.

TABLE II. SLR FILTERS

Database	Filter	Results
Scopus	Title: “Scrum”	999 documents
	Open access	≈ 198 documents
	Articles Only	≈ 89 articles
	English and Portuguese	77 articles
	Reading and Analyzing the Titles	50 articles
	Full read and analysis	41 articles

After defining the number of articles that would be included in the SLR, it was moved on to the analysis and extraction phases for the data necessary for answering the research questions detailed in the first stage. This sub-step consisted of the fourth point, as proposed by the SLR authors, i.e., extracting necessary data. Microsoft Excel® was used for the data extraction, and for future quantitative analyzes (Synthesizing the data), which will be presented in the next step of the method. After the analyses, it was moved on to the third SLR stage, which describes the conclusions and evidence.

C. Documenting the Review (Information Stage)

After the review step, the documentation review began, using information obtained from the previous steps to answer the research questions from the first step, and thereby complete the SLR. This stage sought to remove relevant information to answer the aforementioned research questions.

An Excel spreadsheet was developed, wherein the research questions were set in columns, and the selected articles in lines, so the information that might answer the research questions could be record. Furthermore, extra columns were added for

non a priori information. The complete list of articles from the systematic literature review of this article can be found by contacting the authors.

Finally, regarding the results and finishing the SLR, the next section presents these discussions in greater detail.

IV. ANALYZING THE RESULTS

The results analysis for this study focused on analyzing the two pillars of this study, as shown in Table II. The first pillar investigated the State of the Arts, providing information on publications according to year, the main authors, the publishers, the universities, etc. The second pillar focused on investigating the nature of research, by analyzing general concepts, the structure, the way of working, the team, etc. This analysis began with the State of the Arts.

It is worth mentioning that the analyses were based on reading the 41 articles related to Scrum, so the conclusions of this study refer to SLR, and not the extent of the topic in literature.

A. Analysis of the State of the Arts

Regarding the analysis on the State of the Arts, Q1 deals with the number of publications per year on Scrum. It is worth remembering that a time filter was not applied in this analysis, i.e., the data refers to the first and last publications on Scrum. Fig. 1 summarizes this and shows publications dealing with Scrum.

Fig. 1 shows that from the articles considered in the SLR, the first study published on Scrum was in 2010. Over the years the number of works on Scrum has oscillated, and the biggest growth was between 2018 and 2019. 2020 had the most publications, at 12 articles.

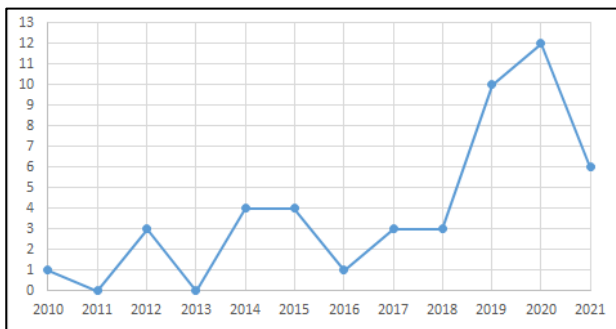


Fig. 1. Publications per year.

The next analysis dealt with the main authors on Scrum, responding Q2. The five most published authors in the database selected for this SLR are: Zada, I.; Vogelzang, J.; Shahzad, S.; Gandomani, T. J.; and Admiraal W. F. These authors have their names published in two articles. The other authors have only one published article. Thus, it was noted that no single author focused on the theme, but rather these authors were dispersed.

Regarding the question on the country of publication (Q3), publication sources (Q4), and universities (Q5), for the State of the Arts, it was decided to condense the data into a single table. The results for these three analyzes are shown together in Table III.

TABLE III. TOP 5 COUNTRIES

Top 5 Countries	Number of articles
Pakistan	6
The United States	5
Brazil	4
India	4
Australia	3

Table III shows that the main publishing country in the database was Pakistan, followed by the United States. Brazil was tied for third place along with India, at 4 articles, while Australia appeared last, with 3 articles.

Regarding the publishers (Q4) (Table IV), it was noticed that the number of publications was very close to each other. The International Journal of Advanced Computer Science and Applications came in first, with four publications on Scrum. Second was the International Journal of Advanced Trends in Computer Science and Engineering, with 3 articles. Three publications sources, the Bulletin of Electrical Engineering and Informatics, IFIP Advances in Information and Communication Technology, and Scientific Programming came in third place, with 2 articles each.

TABLE IV. TOP FIVE PUBLICATIONS

Top 5 Publications	Number of articles
International Journal of Advanced Computer Science and Applications	4
International Journal of Advanced Trends in Computer Science and Engineering	3
Bulletin of Electrical Engineering and Informatics	2
IFIP Advances in Information and Communication Technology	2
Scientific Programming	2

The main universities found with publications on the subject in question is presented in Table V. Regarding Q5, it can be seen that the largest number of publications was four papers published by Leiden University. The other universities published 2 articles per year. It is interesting to note that the Federal University of Itajuba was among these main universities with publications on Scrum. The study was from Breno Tavares, Carlos Eduardo Sanches, and Adler de Souza, entitled: Risk management in Scrum projects: A bibliometric study, published in 2017.

TABLE V. MAIN UNIVERSITIES

Main Universities	Number of articles
Leiden University	4
Greijdanus College	2
Federal University of Itajuba	2
University of Melbourne	2
Univerza v Ljubljani	2
Islamic Azad University	2
University of Peshawar	2
Kohat University of Science and Technology KUST	2
Universidad de Ciencias y Humanidades	2
Melbourne Graduate School of Education	2

The last analysis performed for the State of the Arts was identifying the main areas of study that use Scrum (Q6). The main areas of study that use Scrum are: Computer science; Math; Business, Management and Accounting; Engineering; Social Sciences. The main area of study was Computer Science, with 33 publications on Scrum, representing approximately 47% of the top 5 areas of study. This was followed by engineering, at 18 articles on the subject, representing approximately 25% of all publications among the 5 main areas of study. This was followed by Mathematics at 10 articles, Social Sciences at 6 articles, and Business, Management, and Accounting at 3 articles. Some articles were present in more than one area of study, so the total number of articles seems to be greater than the amount studied; however, this is due to the fact that the same article can address multiple areas.

In general, the SLR showed that Scrum was mentioned in literature for the first time in a 2010 in an article called "A teamwork model for understanding an agile team: A case study of a Scrum project [32]. Over time, the area has grown. 2020 had the most publications on the subject, at 12 articles related to Scrum.

Regarding the authors, five main authors had two publications on Scrum, which shows a non-centrality of the topic, considering that most authors had only one publication on the subject. Considering the analyses on the country of origin, Pakistan had the most published articles on the topic, and Brazil ranked third tied with India, with 4 articles on Scrum. Regarding the publication sources, the five main publication sources had similar publication numbers. The International Journal of Advanced Computer Science and Applications had the most publications, at 4 articles. Regarding the universities, the greatest number of publications was four articles by Leiden University, and the rest of results were homogenous, since no university had an expressive number of publications. Federal University of Itajuba was among the universities found in this study.

Finally, regarding the area of study, Computer Science had the most published articles on Scrum. It is interesting to note that this area was also responsible for starting Scrum, and this

trend continues to the present day. This was followed by Engineering, Mathematics, Social Sciences, and Business, Management, and Accounting, respectively.

B. Analysis of the Nature of the Research

Here will be described the observations and studies on the nature of the research, which was also divided into categories, as shown in Table I. First, it was intended to highlight information on the general Scrum concepts.

Questions Q7, Q8, and Q9 sought to understand the history, definitions, and principles of Scrum, respectively.

Regarding its history, the Scrum methodology was first developed in the 1990s, as a project management framework used in business and industry to manage complex projects, especially for software development [33].

Furthermore, [34] states that the term Scrum originates from Rugby (a team sport of intense physical contact), and refers to a powerful team of Rugby players, positioned specifically to win the ball. Scrum provides for ceremonies, roles, and artifacts to monitor progress, to adjust to changing circumstances, and to reflect on product quality.

Scrum was developed in 1993 by Jeffrey Sutherland at the Easel Corporation, to deliver quality software products in small time intervals known as sprints [35].

Regarding the definition of Scrum, according to [36], Scrum is a framework composed of functions performed by team members, artifacts in documentation, delivery standards, and events that consist of essential actions for communication and delivery for a product. The author in [37] add to this, and state that Scrum is an iterative and incremental project management framework. It is a flexible and comprehensive development strategy, wherein a team works as a unit to achieve a common goal, challenging traditional approaches to product development.

For some authors, Scrum is defined as an agile methodology for developing software and information systems with a focus on project management, and on iterative and agile development processes, transparency, visibility, and cooperation among team members and customers [21, 38].

After the conceptual analysis on Scrum, it was able to identify many authors that define Scrum as a methodology or method. However, most authors define Scrum as a framework, and not just as a project management method, constituting the common way of defining Scrum.

Answering the question on Scrum principles (Q9), [33] state that the Scrum framework is based on three main principles, i.e., transparency, inspection, and adaptation. Transparency clearly defines the objectives and gives visibility to the involved processes, while inspection involves frequent reviews to verify team processes, and adaptation refers to adjustments that can be made during the processes to make effective changes to the project when a product does not meet desired requirements. All three pillars are essential so that the staff can learn from their experiences and adapt their activities to meet demands in an ever-changing environment.

Questions Q10 and Q11 sought to detail the Scrum structure, highlighting the main principles and steps. Answering the question on the Scrum stages (Q10), Scrum life cycle begins with the Product Owner's vision, i.e., the vision he/she has in relation to the product he/she wants to create. From this, a list of resource items by priority level is created, called the Product Backlog. Then, the Sprint Planning and the Sprint Backlog are started, which include all the tasks that the team will perform. In the sprint planning, tasks and weights are chosen from Product backlog items. The Sprint backlog is carried out, where tasks are divided for the whole team, and their respective goals are defined for each cycle. Daily meetings last around 15 minutes, and are organized to track project progress. At the end of each sprint cycle, there is a sprint review, where products are inspected by stakeholders, including customers, and the project team, to get feedback and conduct product analysis. Finally, there is the sprint retrospective step, which is carried out before the next sprint that includes the team, and seeks to bring about improvements and better results for the next stage [39, 40]. Fig. 2 shows the Scrum lifecycle by [41].

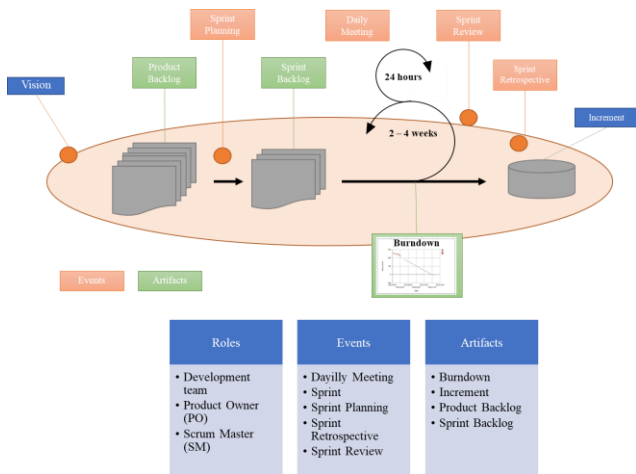


Fig. 2. Scrum lifecycle.

To answer Q11 (What are the Scrum elements?), the Scrum framework includes elements like ceremonies, functions, and artifacts that provide a structure and overview [34]. Scrum framework comprises a Scrum team, roles, and responsibilities, for all members that comprise the team, in addition to events, artifacts, and specific rules that unite all components [42].

The next analysis answered the questions on ways of working with Scrum (Q12, Q13, and Q14). Regarding whether Scrum is expensive (Q12), there are no authors that had made citations on the cost of applying Scrum or values related to it. Thus, it is suggested that future research be carried out on Scrum costs to answer this question.

Regarding the complexity of implementing Scrum (Q13), Scrum adopted by academics and experts used to research or evaluation methods is an easy concept for transference and experience, although some specific particularities need to be adapted [43]. It is also worth mentioning that the Scrum configuration is easy to implement, but its functionality is complex in the team domain [44].

Answering question Q14, on the scope of Scrum, [33] state that Scrum focuses on project management, and is used as a method for dealing with complex projects with imprecise results to minimize risks related to development. Scrum helps resolve uncertainties and respond to changes, providing a set of principles and guidelines to be more productive.

The next questions (Q15 and Q16) focused on the team conceptualization within the framework, and on how the Scrum team composition and the roles of the project manager are defined, respectively. With regard to team composition (Q15), Scrum has three different roles, i.e., a Scrum master, a product owner, and a development team. The Scrum master is responsible for ensuring that the processes run smoothly and that everyone understands and adheres to the Scrum principles. The product owner is responsible for product development focused on customer goals and requirements, and directs the work of the entire team, in addition to defining the product backlog items. The development team comprises professionals from different areas. It is very important that team members work well together, and be focused on the same goal. Another Scrum team characteristic is self-organization. The team has autonomy to make collective decisions, manage their own work, and shift work among themselves, as they see fit [45].

The responsibilities of the traditional project manager (Q16) are divided between three roles in Scrum, i.e., the Scrum master, the product owner, and the development team [46]. The project manager introduces him-/her-self as a Scrum master, and is responsible for the entire project [47].

Questions Q17 and Q18 demonstrate what the SLR authors see to be the main advantages and disadvantages of the Scrum method.

The author in [48] highlights advantages (Q17), as empirical feedback, team self-management, and struggles to build increments of fully tested products in short time frames. Scrum stands out as an agile methodology with the most standardized way of introducing agility, because it is flexible, simple, and a popular agile management method. Furthermore, it facilitates better ways of developing projects, and promoting individual work and teamwork. The Scrum method can generate benefits like increased customer satisfaction, improved communication, increased productivity, reduced production costs, and increased collaboration among everyone involved in the projects, in addition to increasing motivation among the development team [49].

Regarding disadvantages (Q18), the Scrum framework does not formally describe project risk management. Therefore, it is essential that risk management processes be included in the framework according to the project requirements [50]. Another difficulty is linked to large projects, where large numbers of Scrum sprints mean that the development team must manage more complex tasks. The author in [37] also states that one of the weaknesses of Scrum is the strong dependence of the development team on team members.

Finally, Q19 deals with the future directions for Scrum that was addressed by the authors in the SLR. The author in [32] states that future work should focus on identifying and solving problems that may arise during the introduction of agile

development. Another research factor is an extended teamwork model that should be used to study mature agile development teams, to identify and analyze the main challenges encountered by these teams. The flexibility that the Scrum framework offers in terms of compatibility when working together with other agile methodologies for project development is well known, values teamwork, and consequently brings about greater quality, and greater customer satisfaction. One study opportunity would be to analyze the advantages of working with Scrum compared to other agile methods.

Given out analyses for ascertaining the nature of the research topic, the main results are that Scrum was developed as a project management framework in 1993, by Jeffrey Sutherland at the Easel Corporation, and its main objective is to manage complex projects with quality in small time intervals, especially for software development.

Furthermore, the Scrum framework provides for ceremonies, functions, and artifacts, and consists of three main principles, which are transparency, inspection, and adaptation, providing structure and overviews. The purpose of the Scrum scope is to focus on project management, and it is used as a method for dealing with complex projects with inaccurate results, and minimizing risks related to project development.

Scrum's main advantages are flexibility, self-organization, and adaptability. However, disadvantages include informality in risk management processes, the fact that sometimes tasks are too complex for the team, and the fact that all the work is dependent on the development team.

Finally, future directions for Scrum would be studies aimed at improving identifying and solving problems in applying agile concepts, studying mature development teams, studying Scrum frameworks associated with agile methodologies in project development, and analyzing the advantages of working with Scrum relative to other agile methods.

V. CONCLUSIONS

The study conducted a Systematic Review on literature on Scrum to deepen knowledge on the subject, and to obtain a register that could be used to add knowledge for future research. The study was structured following a SLR framework using three stages, i.e., planning the review, conducting the review, and documenting the review.

For the results, it was focused on answering the questions according to the categories into which they were separated, i.e., the State of the Arts, and nature of research. Regarding the State of the Arts, it can conclude that most articles were published in 2020, and the first study found in the database was "A teamwork model for understanding an agile team: A case study of a Scrum project. Regarding the authors, no author had more than two published articles, showing that there is no main author addressing the topic. Moreover, most articles showed an author appearing only once. Regarding the country of publication, Brazil is well-positioned in third place, at two articles published, behind Pakistan in first place. Regarding universities and source of publication, the number of publications was very close, and that there was no main reference. Regarding the publication sources, no single publication had published more than 2 articles.

Finally, regarding the areas of study, more emphasis was given to computer science. It is worth noting that this area gave rise to Scrum, and is still remains at the forefront of research.

The main results from the analyses carried out to ascertain the nature of the research are that the Scrum methodology was developed as a project management framework in 1993, by Jeffrey Sutherland, at the Easel Corporation, and its main objective is to manage complex projects with quality in small time frames, especially for software development.

The Scrum framework provides for ceremonies, functions, and artifacts, and consists of three main principles, which are transparency, inspection, and adaptation. The purpose of the Scrum scope is to focus on project management, and is used as a method for dealing with complex projects with inaccurate results, and to minimize risks related to project development.

Scrum's main advantages are flexibility, self-organization, and adaptability. For future studies on the Scrum framework, authors pointed to improvements in solving problems identified in the applicability of agile development, studying identifying and analyzing problems identified by mature agile teams, and investigating Scrum relative to other agile methods for project management.

Literature points to numerous authors who cite Scrum as a method or methodology for project management. As much as Scrum presents several tools for project development, the term methodology is not the most common term used, as it is based on empiricism. After the analyses in the SLR, we can conclude that Scrum is conceptualized as an agile framework.

This study concluded that Scrum is an agile framework aimed mainly at software development, but can be used in several other areas, e.g., engineering, mathematics, and business, if it is adapted to the context of the specific projects that will be managed.

It was also found that Scrum, as a topic, has grown over the years, and the number of studies has been increasing. Thus, this tool has been consolidated for managing complex projects, which can be better defined as the project is developed. Furthermore, the role of the project manager is more of that of a facilitator, focusing on directing the steps, and eliminating bottlenecks so that the team can develop and self-manage its own work.

It was also concluded that Scrum has some disadvantages, which can be used as themes for developing proposals that could help minimize losses when using the framework. Finally, the results show that future studies still need to greatly develop the Scrum concept, so that Scrum can be increasingly applied within the project management area.

In short, it was able to reach the objectives of this study using a Systematic Literature Review, and by selecting the database for research and analysis, to answer the points defined at the beginning of this study. Studies in literature on Scrum, and the main Scrum concepts, were also detailed here. Finally, this study presents information that could be useful for future studies, as it serves as a simple and direct way of acquiring knowledge about Scrum.

For future studies, the authors suggest that studies deal with financial issues, and complexity when applying Scrum. Another recommendation would be carrying out a new SLR on other agile methods, or conducting a SLR that addressed the most well-known agile methods, offering a comparative analysis to offer the reader a focused approach on each of these methodologies, and what type of project these methodologies would best fit.

REFERENCES

- [1] T. J. D. D. Caracterização do gerenciamento de projetos em micro e pequenas empresas desenvolvedoras de software (Master's thesis, Universidade Federal de Pernambuco), 2020.
- [2] B. H. Reich and S. Y. Wee. "Searching for Knowledge in the PMBOK® Guide". *Project Management Journal*, 2006, 37(2), 11-26.
- [3] R. O. Gonçalves, F. Bertoli, and J. A. Krüger. *Gestão das fases preliminar e interna do processo licitatório de edificações públicas*, 2011.
- [4] R. V. Vargas. *Gerenciamento de Projetos (6a edição)*. Brasport, 2005.
- [5] Y. Olawale and M. Sun. "PCIM: Project control and inhibiting-factors management model". *Journal of Management in Engineering*, 2013, 29(1), 60-70.
- [6] PMBOK® guide. *A guide to the Project Management Body of Knowledge (PMBOK® guide)*, 6th ed., 2017.
- [7] M. H. O. D. Lima. *Principais barreiras e potencialidades de adoção de abordagens híbridas no gerenciamento de projetos: um estudo exploratório*, 2018.
- [8] K. Beck, et al. *Manifesto para desenvolvimento ágil de software*. 2001. Available in: <http://agilemanifesto.org/iso/ptbr/manifesto.html>. Accessed on April 20, 2023.
- [9] A. R. Balle. *Análise de metodologias ágeis: conceitos, aplicações e relatos sobre XP e Scrum*, 2011.
- [10] A. L. D. Ribeiro and R. Arakaki. "Gerenciamento de projetos tradicional x gerenciamento de projetos ágil: uma análise comparativa". In 3rd CONTECSI-International Conference on Information Systems and Technology Management, São Paulo, Brasil, 2006.
- [11] G. B. V. Soares and T. F. Pereira. "Estudo de caso sobre a aplicação da metodologia Scrum em uma startup tecnológica de Minas Gerais". *Research, Society and Development*, 2021, 10(3), e9410313064-e9410313064.
- [12] V. M. Teles. *Um estudo de caso da adoção das práticas e valores do Extreme Programming*. UFRJ-Universidade Federal do Rio de Janeiro, 2005.
- [13] G. R. Kuhn and V. F. Pamplona. *Apresentando XP. Encante seus Clientes com Extreme Programming*. Blumenau, SP. Universidade Regional de Blumenau, 2004.
- [14] K. Schwaber and J. Sutherland, J. *Um guia definitivo para o Scrum: As regras do jogo. Processo de Desenvolvimento de Software*, 2013.
- [15] M. G. Trigás Gallego *Metodologia Scrum*, 2012.
- [16] M. Poppendieck. "Lean software development". In: 29th International Conference on Software Engineering (ICSE'07 Companion). IEEE, 2007. p. 165-166.
- [17] A. M. C. A. da Cunha, C. E. de Campos and H. H. C. Rifarachi. *Aplicabilidade da metodologia Lean em uma lavanderia hospitalar. O mundo da saúde*, 2011, 35(3), 311-318.
- [18] B. D. Cruz. *Um mapeamento sistemático de métricas para metodologias ágeis Scrum, Kanban e XP*, 2013.
- [19] L. V. Arruda. "Desenvolvimento Ágil de Software: uma análise sintética a partir da metodologia Kanban". In VII CONNEPI-Congresso Norte Nordeste de Pesquisa e Inovação, 2012.
- [20] G. R. Stopa and C. L. Rachid. *Scrum: Metodologia ágil como ferramenta de gerenciamento de projetos*. CES Revista, 2019, 33(1), 302-323.
- [21] K. Schwaber. *Agile project management with Scrum*. Microsoft press, 2004.
- [22] R. T. Silva. *Aplicação da metodologia scrum para gestão de projetos na indústria naval*, 2011.
- [23] D. Power. "Supply chain management integration and implementation: a literature review". *Supply Chain Management: an International Journal*, 2005.
- [24] L. A. Alharbi. "A Systematic Literature Review on AI Algorithms and Techniques Adopted by e-Learning Platforms for Psychological and Emotional States". *International Journal of Advanced Computer Science and Applications*, 2023, 14(2).
- [25] C. Wong, H. Skipworth, J. Godsell, and N. Achimugu. "Towards a theory of supply chain alignment enablers: a systematic literature review". *Supply Chain Management: an International Journal*, 2012.
- [26] M. M. Kamal and Z. Irani. *Analysing supply chain integration through a systematic literature review: a normative perspective*. *Supply Chain Management: An International Journal*, 2014.
- [27] S. A. K. Ghayyur, S. Ahmed, M. Ali, A. Razzaq, N. Ahmed, and A. Naseem. "A systematic literature review of success factors and barriers of Agile software development". *International Journal of Advanced Computer Science and Applications*, 2018, 9(3), 278-291.
- [28] M. C. B. Galvão and I. L. M. Ricarte. "Revisão sistemática da literatura: conceituação, produção e publicação". *Logeion: Filosofia da informação*, 2019, 6(1), 57-73.
- [29] P. Dallasega, E. Marengo, and A. Revolti. *Strengths and shortcomings of methodologies for production planning and control of construction projects: a systematic literature review and future perspectives*. *Production Planning & Control*, 2021, 32(4), 257-282.
- [30] A. P. G. Scheidegger, T. F. Pereira, M. L. M. de Oliveira, A. Banerjee, and J. A. B. Montevechi. "An introductory guide for hybrid simulation modelers on the primary simulation methods in industrial engineering identified through a systematic review of the literature". *Computers & Industrial Engineering*, 2018, 124, 474-492.
- [31] P. Mongeon and H. Paul-Hus. "The journal coverage of Web of Science and Scopus: a comparative analysis". *Scientometrics*, 2016, 106(1), 213-228.
- [32] N. B. Moe, T. Dingsøyr, and T. Dybå. "A teamwork model for understanding an agile team: A case study of a Scrum project". *Information and Software Technology*, v. 52, n. 5, p. 480-491, 2010.
- [33] K. Schwaber and J. Sutherland. *Guia do Scrum- Um guia definitivo para o Scrum: As regras do jogo*. 2017. <http://www.scrumguides.org/docs/scrumguide/v1/Scrum-Guide-Portuguese-BR.pdf>. Citado, 3, 49.
- [34] R. Pope-Ruark. "We scrum every day: "Using scrum project management framework for group projects". *College teaching*, 2012, 60(4), 164-169.
- [35] J. Sutherland, J. "Inventing and Reinventing SCRUM in five Companies". *Cutter IT Journal*, 2001, 14(21), 5-11.
- [36] K. Schwaber. "Scrum development process". In *Business Object Design and Implementation: OOPSLA'95 Workshop Proceedings 16 October 1995, Austin, Texas (pp. 117-134)*. Springer London.
- [37] M. Esteki, T. J. Gandomani, and H. K. Farsani. "A risk management framework for distributed scrum using PRINCE2 methodology". *Bulletin of Electrical Engineering and Informatics*, v. 9, n. 3, p. 1299-1310, 2020.
- [38] K. Schwaber and M. Beedle. *Agile software development with scrum. Series in agile software development (Vol. 1)*. Upper Saddle River: Prentice Hall, 2002.
- [39] K. S. Rubin. *Essential Scrum: A practical guide to the most popular Agile process*. Addison-Wesley, 2012.
- [40] M. Gannon. "An agile implementation of Scrum". In 2013 IEEE Aerospace Conference (pp. 1-7). IEEE.
- [41] B. G. Tavares, C. E. S. Da Silva, and A. D. De Souza. "Risk management in Scrum projects: A bibliometric study". *Journal of communications software and systems*, v. 13, n. 1, p. 1-8, 2017.
- [42] M. K. Larusdottir, E. R. Bjarnadottir, and J. Gulliksen. "The focus on usability in testing practices in industry". In *Human-Computer Interaction: Second IFIP TC 13 Symposium, HCIS 2010, Held as Part of WCC 2010, Brisbane, Australia, September 20-23, 2010. Proceedings (pp. 98-109)*. Springer Berlin Heidelberg.

- [43] T. J. Gandomani, H. Zulzalil, A. A. Ghani, A. B. M. Sultan, and K. Y. Sharif. "How human aspects impress Agile software development transition and adoption". *International Journal of Software Engineering and its Applications*, 2014, 8(1), 129-148.
- [44] S. Hariharan, A. Rengarajan, and R. P. Kumar. "Scrum based scaling using agile method to test software projects and its future solutions using in artificial neural networks". *International Journal of Innovative Technology and Exploring Engineering*, v. 8, n. 9, p. 223-230, 2019.
- [45] R. G. Cooper and A. F. Sommer. "Agile-Stage-Gate: New idea-to-launch method for manufactured new products is faster, more responsive". *Industrial Marketing Management*, 2016, 59, 167-180.
- [46] S. W. Ambler and M. Line. *Disciplined agile delivery: A practitioner's guide to agile software delivery in the enterprise*. IBM press, 2012.
- [47] B. G. Sudarsono, H. H. Fransiscus, D. Y. Bernanda, and J. F. Andry. "Adopting scrum framework in a software development of payroll information system". *International Journal of Advanced Trends in Computer Science and Engineering*, 2020, 9(3).
- [48] Y. Cui, I. Zada, S. Shahzad, S. Nazir, S. U. Khan, N. Hussain, and M. Asshad. "Analysis of service-oriented architecture and scrum software development approach for IIoT". *Scientific Programming*, 2021, 1-14.
- [49] B. V. Carvalho and C. H. P. Mello. "Implementation of scrum agile methodology in software product project in a small technology-based company". *Gestão & Produção*, v. 19, p. 557-573, 2012.
- [50] M. Mousaei and T. Javdani. "A new project risk management model based on Scrum framework and Prince2 methodology". *International Journal of Advanced Computer Science and Applications*, v. 9, n. 4, 2018.

Artificial Intelligence Based Modelling for Predicting CO₂ Emission for Climate Change Mitigation in Saudi Arabia

Sultan Alamri¹, Shahnawaz Khan²

College of Computing and Informatics, Saudi Electronic University
Saudi Arabia¹

School of ICT, Bahrain Polytechnic
Isa Town, Bahrain²

Abstract—Climate change (such as global warming) causes the barrier in the attaining sustainable development goals. Emission of greenhouse gases (primarily carbon dioxide CO₂ emission) are the root cause of global warming. This research analyses and investigates the emission of CO₂ and attempts to develop an optimal model to forecast the CO₂ emission. Several machine learning and statistical modeling techniques have been implemented and evaluated to explore the patterns and trends of CO₂ emissions to develop an optimal model for forecasting future CO₂ emissions. The implemented methods include such as Exponential Smoothing, Transformers, Temporal Convolutional Network (TCN), and neural basis expansion analysis for interpretable time series. The data for training these models have been collected and synthesized from various sources using a web crawler. The performance of these models has been evaluated using various performance measurement metrics such as RMSE, R2 score, MAE, MAPE and OPE. The N-BEATS model demonstrated an overall better performance for forecasting CO₂ emission in Saudi Arabia in comparison to the other models. In addition, this paper also provides recommendations and strategies for mitigating the climate change (by reducing CO₂ emission).

Keywords—Exponential smoothing; transformers; temporal convolutional network; neural basis expansion analysis; climate change

I. INTRODUCTION

The United Nations (UN) has proposed several climate friendly actions for the governments while developing the post-pandemic recovery plans. Governments, all over the globe, are aiming to tackle the climate-crisis for changing the course of carbon dioxide (CO₂) emission trajectory to net-zero, to create a more sustainable and safer future for their citizens. But why there is so much attention given to CO₂ emission? CO₂ is a greenhouse gas. Greenhouse gases traps the heat in the atmosphere and radiates it back in all directions including towards the surface of earth [1]; hence, causing an increase in the global temperature (climate change). There are multiple sources (such as fossil fuels extraction and consumption, wildfires, volcanic eruptions and other similar natural processes, etc.) that generate carbon dioxide. Among these sources, fossil fuels are the major contributing factor of CO₂ emission.

The Kingdom of Saudi Arabia (KSA) presented its first national review report on sustainable development goals (SDGs) to UN High-Level political forum in 2018 [2]. The presented review is the first attempt by the KSA to conduct a comprehensive and systematic review of the status. Climate action is one of the 17 SDGs agenda set by the UN. As the UN member state, Kingdom of Saudi Arabia has adopted the sustainable development goals under its vision 2030. There have been several great initiatives and indicators that has exemplified the enthusiasm of the Kingdom of Saudi Arabia in achieving these goals as indicated from the first voluntary national review report [2][3]. The review suggests that the KSA has made great efforts in numerous areas; however, there are several bottlenecks that place the KSA in a challenging position to achieve its 2030 vision. Based on the review report, the major challenges include data availability, efficient measures and methods for SDG-related statistics collection and dissemination, more effective and better coordination techniques among the non-government and government institutions to avoid effort duplication, enriching and enhancing the existing institutional frameworks, so on and, so forth. Therefore, more profound interventions are essential to move forward. One of the areas of concern among these challenges is emission of greenhouse gases.

Climate change is one of the major hurdles in attaining the sustainable development goals [4][5]. Rapid industrialization and urbanization have given the unexpected economic growth but the world has to pay the price in the shape of climate change. Due to the unprecedented utilization of fossil fuel in energy generation and consumption, the emission of greenhouse gases (primarily CO₂) has increased exponentially. Greenhouse gases gather in the atmosphere, absorb sunlight, and avert the solar radiations to reflect back from the earth surface. Thus, the radiations are entrapped in the atmosphere. The entrapped radiations become the reason to increase the temperature of the plane and thus, causing global warming. This procedure is referred to as greenhouse effect. The gases that cause greenhouse effect include carbon dioxide (CO₂), nitrous oxide (N₂O), chlorofluorocarbons (CFCs), methane (CH₄), and water vapor. Though, among these gases, CO₂ is the primary contributor to greenhouse effect. However, human interventions and activities have increased the CO₂ concentration significantly (by 47%) in the atmosphere in the

past 170 years. This increment in CO₂ concentration would have taken a period of over 20,000 years naturally [6]. Therefore, to mitigate the climate change, an efficient technique is required that can measure the emission of greenhouse gases and can provide an insight for the future emission trends so that an effective climate change mitigation solution can be developed and put into action.

The emission of greenhouse gases does not only causes global warming but also has several other effects such as air pollution, so on and, so forth. The poor quality of air causes various health risks to the citizens. Therefore, there several direct and indirect consequences of GHG emission such as environment deterioration, health deterioration, etc. A research study [7] illustrates that almost 40% of global CO₂ emission is contributed by the electricity generation by combustion of the fossil fuels. However, there are several other natural factors that contribute to GHG emission such as respirations, and volcanic eruptions. Although, the modern industrialization shift has resulted in excessive combustion of the fossil fuels and consumption of natural resources. It also leads to several other major environmental challenges such as climate change, deforestation and water shortage. Effective measurement, analysis and forecasting techniques of GHG emission will guide in identifying the key factors and emission sources that will affect the climate change mitigation policies.

This paper has been organized into six sections. The next discusses the related work in this regard. Section three states the problem statement, research importance and objectives. Section four describes the data and methodologies. It also describes the proposed system framework and the methodologies which have been implemented for developing an effective forecasting model. Section five illustrates the results of different forecasting models, provide performance analysis, and suggest CO₂ emission mitigation strategies. Section six concludes the research paper and provides the future research direction for mitigating climate change.

II. LITERATURE REVIEW

Climate change is one of the most defining global issues of this century. United Nations has listed 17 SDGs in its 2030 agenda. Climate action is one of the sustainable development goals among others such as sustainable cities and communities, affordable and clean energy, responsible consumption and productions, etc. The United Nations 2030 agenda encompasses primarily five pillars such as Planet, People, Peace, Prosperity, and Partnership. The 17 SDGs incorporate these five pillars. As a member state of United Nations, Kingdom of Saudi Arabia (KSA) has adopted sustainable development goals. In the past few years, the KSA has already implemented numerous polices under the Vision 2030, these policies align with the United Nations 2030 agenda. The ongoing 2030 vision has demonstrated progress in several indicators to achieve the sustainable development goals and to become a welcoming economy for visitors, workers and investors. Saudi Arabia has proved its commitment in achieving sustainable development goals and is enthusiastic in recognizing the importance to address these global challenges.

Swift urbanization and unprecedented industrialization have raised the issue of global warming and climate change

around the world. The UN have been setting goals to curb the drastic effects of climate change for its member states. In turn, all the member states have been implementing different policies and taking actions to accomplish their goals for sustainable development. However, due to the high demand of energy for industrialization and urbanization, the ways of energy production and consumptions have not been efficient in lowering the GHG emission levels. As a result, the level of GHG emission have been increasing around the world and causing global warming. The Kingdom of Saudi Arabia is no exception to this. Despite the numerous efforts made by the concerned authorities, the KSA still has to go far to meet its target level of GHG emission. As per the first voluntary national review report [2], the KSA has to cut down its annual carbon dioxide emission as much as 130 million tonnes by 2030.

The Kingdom of Saudi Arabia has placed several policies and strategies to meet its requirement. However, due to the increased power (energy) demand by industrialization and population growth, over 80% of the overall energy demand is fulfilled by using the fossil fuels for energy generation (Amran et al, 2020). Energy generation using fossil fuels is directly linked with CO₂ emission. Thus, an ambitious target for the KSA is to switch to alternative sources of power sources such as nuclear power, green energy [8] or renewable energy sources [9], etc. Climate change does not only cause increment in atmospheric temperature but also accounts of huge financial loses. It has costed around 130 million dollars in 2018 only and almost USD one billion from 1980-2010 (Al-Bassam et al., 2014). Climate's hyper aridity and the sensitive ecosystem place the KSA at a certain climate change risk [10]. In some areas of the KSA, a very high temperature (52°C) has been recorded which is negatively influenced by various human activities and vehicle emission [3].

The government of Saudi Arabia is addressing several challenges for climate change mitigation and adapting measures to restraint the social and diverse economic impacts on the country. Some of these efforts include King Abdulaziz public transport project, conservation measures for water and electricity [11], urban planning scheme [10], research on environment [12], and many more. Although, despite the significant efforts made by the government, per capita CO₂ emission in the Kingdom is still amongst the highest in the world [13]. Therefore, solid strategies, measures and plans are required to put into action for climate change mitigation and economic diversification such as promoting the social responsibilities of corporations for safeguarding the environment.

This research will utilize the data collected from past to analyze the trends and will employ forecasting and/or prediction models in estimating the CO₂ emission (because it is the major constituents amongst GHG). As the future human activities and behavior will have a substantial impact on the accuracy of the estimations, therefore, this research has also attempted to provide the guidelines in developing the effective climate change mitigation pathways and policies. Several machine learning, statistical and mathematical modelling techniques have been employed for the development of an effective forecasting model. As the amount of the GHG

emission data available [14] [15] from all the possible resources is limited, therefore, this research has employed the techniques that perform well under the limited data. Finding an appropriate time series forecasting method has been a significant challenge amongst the community of researchers. Finding a prediction or forecasting technique that can perform well on the known and unknown data is a challenging task. When the historical data available is in low quantity, prediction or forecasting techniques performs well on the training data and produce results with minimal error rate, while on the unseen data the results are usually disastrous [16] [17]. This research has implemented several machine learning, statistical and mathematical modelling techniques for simulating the CO₂ emission in the Kingdom of Saudi Arabia. These findings will play a substantial role in developing the climate change mitigation policies and strategies for Saudi Arabia.

The literature review illustrates that Saudi Arabia has to cut down its annual carbon dioxide emissions by 130 million tonnes by 2030, but it does not provide information on the country's current emissions levels. Several efforts have been made by the Saudi Arabian government to address climate change. However, it has also been observed that there are need for solid strategies, measures, and plans for climate change mitigation and economic diversification. It has also been observed that machine learning, statistical and mathematical modeling techniques have been used for forecasting CO₂ emissions in Saudi Arabia, but it does not provide information on the data sources or the accuracy of the models used.

III. RESEARCH IMPORTANCE AND OBJECTIVES

The GHG emission in Saudi Arabia has increased almost 200% in the past three decades. Per capita GHG emission is the second highest in Saudi Arabia amongst the G20 countries. Kingdom of Saudi Arabia would require to reduce its CO₂ emission below 389 Mt by 2030 as its commitment to UN's sustainable goals and to achieve it vision 2030 goals. Ambient air pollution (a resultant of emission) poses higher health risks (such as stroke, heart disease, lung cancer, and chronic respiratory diseases) in Saudi Arabia as compared to other G20 countries. Kingdom of Saudi Arabia loses almost SAR 859m (~USD 229m) because of extreme weather events and is vulnerable to climate change. The Saudi Arabia has a target to reduce emission equivalent to 130 million tonnes per year. To address and achieve the above targets and issues, an effective estimation technique and guidelines are required which this research study aims to address. This research synthesizes the findings from literature to recognize the opportunities for potential estimate emission reduction and intervention. It implements several machine learning, statistical and mathematical modelling techniques for simulating the emission of greenhouse gases (specifically CO₂) in the Kingdom. It will also aim to address the issue of accuracy in the synthesized data by verifying it from multiple sources as described in the next section. As there is a need for solid strategies, measures, and plans for climate change mitigation and economic diversification, therefore, this research also provides suggested strategies and policies that can be utilized along with the forecasting model to mitigate the climate change.

IV. MATERIALS AND METHODS

This section examines the data collection, and forecasting methodologies used in this paper. This section is organized into three subsections which are dataset, proposed system framework, and forecasting methodologies utilized in this research. The dataset subsection describes the data collection and data pre-processing. The proposed system framework subsection describes the research design, and the forecasting methodologies subsection describes the different forecasting methods implemented in this research study.

A. Dataset

For this research study, CO₂ emission data (as CO₂ is the major greenhouse gas) has been collected available in the public domain. Some major sources of the data that have been considered are World Bank [13], BP [14], and our world in data [15]. This research utilizes the complete emission data obtained from all these sources by converting to one scale and synthesizing it. The figure (see Fig. 1) illustrates the synthesized data.

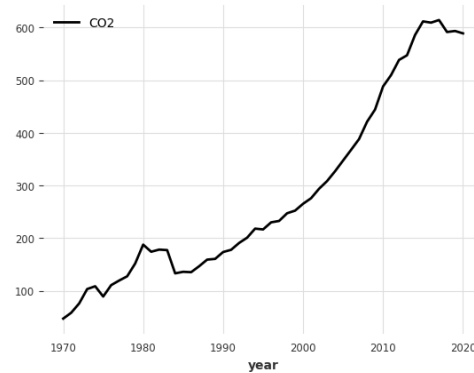


Fig. 1. Year-wise CO₂ emission in Saudi Arabia.

B. Proposed System Framework

A quantitative analysis has been conducted on the data collected from all the sources to investigate the patterns and trends. This research explores and implements several forecasting techniques such as Exponential Smoothing, Transformers, Temporal Convolutional Network (TCN), neural basis expansion analysis for interpretable time series, and Fast Fourier Transform Forecasting. The following figure (see Fig. 2) illustrates the proposed system framework.

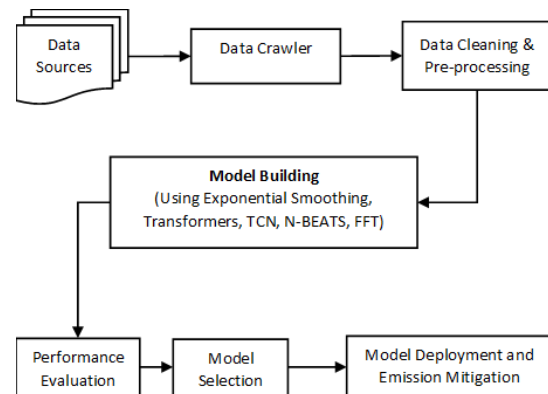


Fig. 2. Proposed system framework.

The study implements an automated data crawler to extract data from various sources [18]. The collected data is processed for any abnormalities and missing values and cleaned. The data is scaled to have the same unit. The next module implements multiple machine learning and statistical modelling techniques. All the developed models are evaluated for their performance. The best performing model is selected as the forecasting model and used to forecast emission by 2030. The findings from the quantitative analysis and literature study have been used in developing the proposed strategies for CO₂ emission mitigation and to achieve sustainable development goals, and Saudi Vision 2030.

C. Model Building

For developing an effective model for predicting CO₂ emission in Saudi Arabia, this research has investigated multiple approaches for forecasting. The forecasting results obtained from some of the approaches like Fast Fourier Transform Forecasting Model (FFT) were discouraging. However, there are several techniques which have performed very well and have promising forecasting results [19] [20]. These techniques include approaches such as Exponential Smoothing, Transformers, Temporal Convolutional Network (TCN), and neural basis expansion analysis for interpretable time series (N-BEATS). This segment analyzes the forecasting techniques implemented in this research paper.

1) *Exponential smoothing*: Exponential smoothing is a statistical forecasting method used for time series analysis and predicting the values of a time-series variable based on its historic data points. Exponential smoothing was first proposed 1950s by three different researchers: Charles C. Holt, Peter W. Winters, and Robert G. Brown [21] [22] [23]. Each of them developed a slightly different version but based on the same fundamental concept. This method is based on the assumption that the time-series pattern can be modeled by a weighted average of the historical observations. In exponential smoothing, usually recent events or observations are granted more weights than the older observations. However, by varying the values of the smoothing parameter, more weight can be shifted to the older observations. The following equation illustrates the exponential smoothing for predicting the future values based on the historical observations:

$$F_{t+1} = \alpha A_t + (1 - \alpha)F_t \quad (1)$$

Where, F_{t+1} is the predicted time-step value for the next step, F_t is the predicted value for the current period and A_t is the actual observation value of the time series at time step t . The term α represents the smoothing parameter, and the values of the parameter varies between 0 and 1 that determines the weight allotted to the current observation. Keeping the α value larger, more weight is shifted to the most recent observations. One of the criteria for applying exponential smoothing is its level of persistence of forecasted values with the historic observations. However, it is not very resistance to the irregularities or sudden changes in the time-series data. The implementation in this paper is based on the simplified version present in [24].

2) *Transformers*: Transformers are a type of neural network having self-attention mechanism [25]. Self-attention allows to process different parts of the input sequence while processing each element. Self-attention computes the attention weight between all the pairs of the input values in the time-series. This mechanism aids in determining the importance of each observation. The transformer implementation for this research study, applies self-attention in two processes. One is to extract the intra-dependencies within the output and within the input vectors which is also known as self-attention. The second is to draw the inter-dependencies within the output and within the input vectors which is also known as encoder-decoder attention. The attention mechanism grants the transformers the ability to grab the long-range dependencies efficiently. The transformer also includes a component called multi-head attention which allows the transformers suitable for parallel processing of the multiple components of the input sequence.

The implementation in this paper is based on [25] [26]. The input chunk length has been set to three and output chunk length has been set to one. The transformer has been trained two encoder and two decoder layers with 128-dimension feedforward neural network. The multi-head number has been set to eight. The rectified linear unit (ReLU) activation function [27] has been used. The model was trained with batch size of 8 and over 200 epochs.

3) *Temporal convolutional network*: Temporal Convolutional Network (TCN) have been designed to process sequences of data. The TCN architecture implemented for this study is based on the TCN architecture proposed in [28]. TCNs are based on convolutional neural network. TCN applies 1-dimensional convolutions along the time dimension of the input data. It enables the neural network to learn the features that are relevant to the temporal structure (such as patterns or trends) of the data. In order to learn long-range dependencies in the input time-series, convolutions skip some of the observations in the input time-series. To learn long-term trends or long-range dependencies in the input time-series, TCNs use dilated convolution [29]. TCNs also resolves the issue of vanishing gradients (which often occurs in deep neural networks) by the use of residual connections. Residual connections are the connections that bypasses one or more convolutional layers and allow information to flow more easily through the network. These features of TCN can grab the complicated temporal relationships amongst the output and input values. The fully connected layers are fed with the output of the convolutional layers to map the learned features to the target output. The convolutional layers are usually followed by dilated convolutional layers to grab long-term dependencies in the time-series data. The final output is a sequence of predicted values for the future time-steps. The entire network is optimized to minimize the forecasting error in contrast to traditional forecasting methods which often require separate steps for feature extraction and model fitting.

The TCN implementation in this research paper is based on the architecture proposed in [28]. The future prediction time-step has been set to one. The kernel size in the convolutional layers has been set to two. Three filters have been used in the convolutional layers. The dilation base has been set to two. The network has been trained over 150 epochs.

4) *Neural basis expansion analysis for interpretable time series*: Neural basis expansion analysis for interpretable time series (N-BEATS) is a machine learning technique that is used for modeling and analyzing time-series data [30]. N-BEATS uncovers the underlying relationships and patterns present in the data. N-BEATS offers two architectures such as generic and interpretable. This research study implements the generic architecture. As the name suggests, N-BEATS utilizes a basis function expansion for time-series representation. The basis function expansion is a linear combination (weighted sum) of a set of basis functions. These basis functions aim to map to the patterns or relationships expected to be present in the time-series data. The coefficients of the basis functions are the weights.

The N-BEATS implementation for this research study is based on the architecture presented in [30]. The future prediction time-step has been set to one. During the training using neural network, the weights are learned. The resulting model is utilized for forecasting future values of the time-series. The weights of the basic functions provide insight about the underlying dynamics of the time-series data. The presence of high weights for certain basic functions indicates the presence of certain periodic patterns in the time-series data. Though, for the implementation in this research study, no such insight has been observed. Neural network is fed with the time-series data to learn the weights of the basic functions. In the training process, neural network learns the weights and produces an approximate time-series that best fit on the given time-series. N-BEATS have the ability to capture complex nonlinear relationships present in the time-series data [31].

V. RESULTS AND DISCUSSION

Time series forecasting problems aims to find a fitting model that can fit the time-stamped historical observations in order to predict the future time-stamped observations of the given data. Traditional approaches have been effective in describing and predicting the time series values such as exponential smoothing and autoregressive integrated moving average (ARIMA). While ARIMA models aspire to depict autocorrelation in the time series data, exponential smoothing targets to characterize seasonality and trend in the time series data. However, sometimes, traditional approaches do not effectively describe and fit the time series, especially when the time series data is more complicated. Therefore, often advanced methods such as transformers, neural basis expansion analysis for time series (N-BEATS) forecasting, and temporal convolutional networks (TCNs), are employed to deal with the complicated time series data. This research utilizes multiple methods such as exponential smoothing, transformers, N-BEATS and TCN for describing and predicting the CO₂ emission values in Saudi Arabia.

Initially, a descriptive analysis of the data has been conducted and the synthesized data is cleaned and pre-processed for any data error. Python has been used for the analysis and modelling. The data preprocessing has been used for converting the data to have same units for carbon emission of each year, finding and handling missing values etc. The forecasting methods discussed in the previous section have been implemented and are fed with the time-series data to learn the weights. In the training process, forecasting models learn the weights and produces an approximate time-series that best fit on the given time-series. The weights are learned by minimizing the variance between the predicted observations and actual observations of the given time-series data. The data is segregated into training and validation data. The figure (see Fig. 3) illustrates training data and validation data.

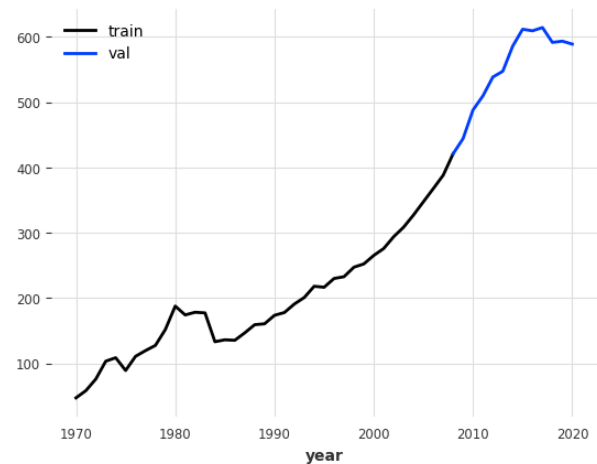


Fig. 3. Training and validation data.

The training, validation and prediction comparison for all the models implemented in this research study is illustrates in the following figures (see Fig. 4, 5, 6, and 7). The training, validation and prediction comparison for all the models implemented in this research study is illustrates in the following figure. The following figure 4 illustrates the training, validation and prediction of the CO₂ emission using the Exponential Smoothing model. All process has been represented using a different color as depicted via legends. The following Fig. 5 illustrates the training, validation and prediction of the CO₂ emission using the Transformers model.

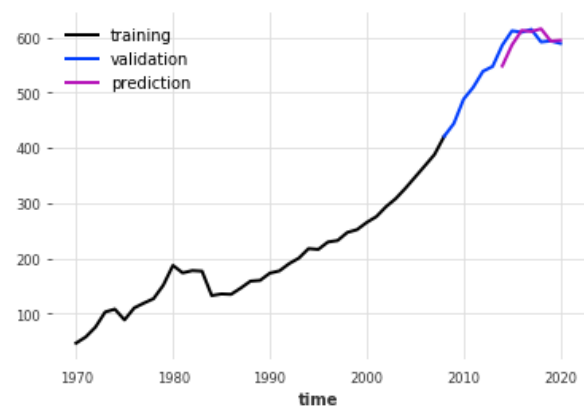


Fig. 4. Exponential smoothing

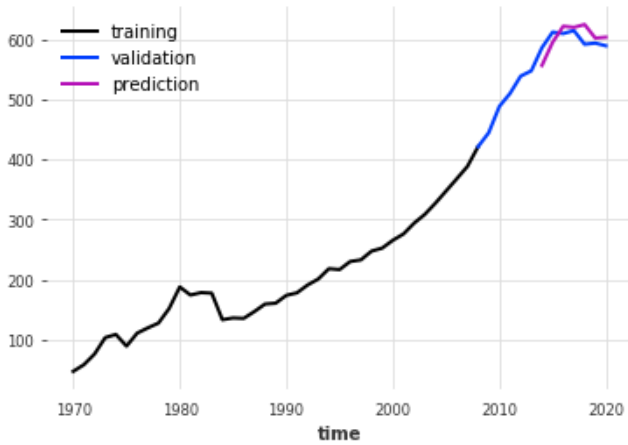


Fig. 5. Transformers.

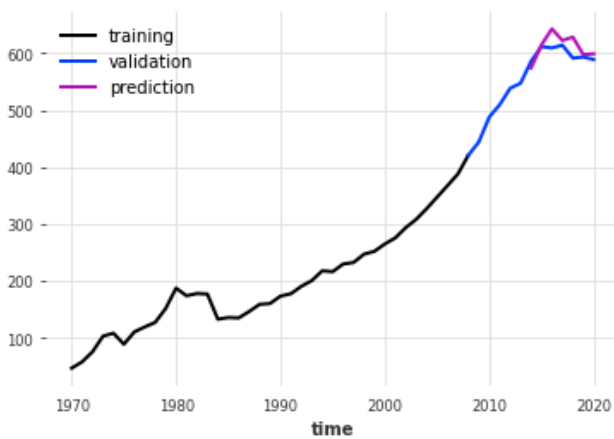


Fig. 6. N-BEATS.

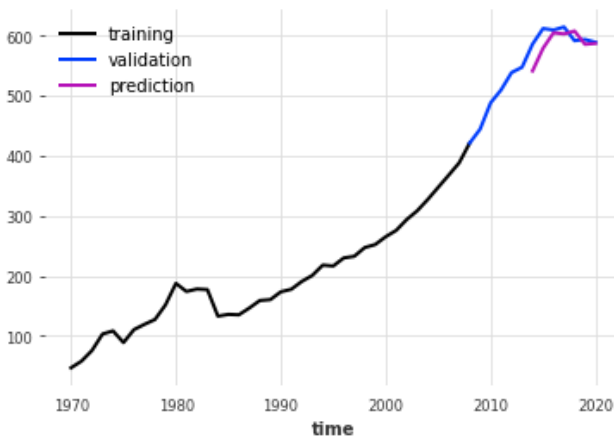


Fig. 7. TCN.

The above Fig. 6 illustrates the training, validation and prediction of the CO₂ emission using the TCN model. All process has been represented using a different color as depicted via legends. The above Fig. 7 illustrates the training, validation and prediction of the CO₂ emission using the N-BEATS model.

A. Performance

1) *Mean absolute error*: Mean Absolute Error is also commonly known as MAE is a performance measurement metric which is used to evaluate the performance of a prediction or forecasting models. It measures the average absolute difference of the forecasted data point and the actual data point value, and is expressed in the units of the data. MAE represents the average amount of the errors in the predicted observations, without considering their direction. Therefore, it is capable to handle both positive and negative errors, and it is resistant to the outliers present in the time-series. The formula for calculating MAE is as follows:

$$MAE = (1/t) * \Sigma|a - \hat{y}| \quad (2)$$

Where, t is the total number of data points, v is the actual value and \hat{y} is the predicted value.

2) *Mean absolute percentage error*: Mean Absolute Percentage Error is commonly known as MAPE. It is a performance measurement metric which is used to evaluate the performance of a forecasting model. It calculates the average percentage difference of the actual time series values and the forecasted values. The formula for calculating MAPE is as follows:

$$MAPE = (1/t) * \Sigma|(a - \hat{y})/y| * 100\% \quad (3)$$

Where, t is the total number of time-series data points, a is the actual time-series data-points and \hat{y} is the forecasted time-series data-points.

3) *R2 score*: Coefficient of Determination is also known as R2 Score. It measures the proportion of the variability in the data. The R2 score is calculated as follows:

$$R2 = 1 - (SS_{res} / SS_{tot}) \quad (4)$$

Where, SS_{res} is the sum of the squared residuals, and SS_{tot} is the total sum of squares.

4) *Root mean squared error*: Root Mean Squared Error (commonly referred as RMSE) is a widely used performance metric. It measures the average magnitude of the errors in the predictions, taking into account both the direction and the magnitude of the errors. The formula for calculating RMSE is as follows:

$$RMSE = \text{sqrt}((1/t) * \Sigma(a - \hat{y})^2) \quad (5)$$

Where, t is the total number of points in the time series data, a is the actual value of the time-series step and \hat{y} is the forecasted time-series value.

5) *Overall percentage error*: Overall Percentage Error (OPE) performance metric is used to measure the percentage difference between the actual time-series data values and the forecasted data values, averaged over all the data points. The OPE is expressed as a percentage. The formula for calculating OPE is as follows:

$$OPE = (1/t) * \Sigma(\text{abs}((a - \hat{y})/a) * 100) \quad (6)$$

Where, t is the total number of data points, a is the actual data point of the time-series and \hat{y} is the forecasted data-point.

All the models have been evaluated using the above performance measurement metrics. The following table (see Table I) demonstrates the comparative analysis of the different models. In the present research study, transformer model has performed worse. The study has tried parameter tuning in several ways but the forecasting results could not improve. R2 score achieved is 0.112, MAE score is 55.89, MAPE score is 9.97, RMSE score is 56.10, and OPE score is 9.9. The next model is Exponential Smoothing which has achieved an R2 score of 0.209, an MAE score of 36.87, an MAPE score of 6.43, an RMSE score of 46.90, and an OPE score of 1.413. The TCN and N-BEATS models have performed comparatively well. The R2 score for TCN model is 0.21, MAE score is 14.23, MAPE score is 2.39, RMSE score is 19.58, and OPE score is 0.82. Similarly, the N-BEATS model has achieved an R2 score of 0.7, an MAE score of 14.36, an MAPE score of 2.36, an RMSE score of 19.32, and an OPE score of 0.91.

TABLE I. PERFORMANCE MEASUREMENT FOR VARIOUS FORECASTING MODELS

Performance Metric → Model ↓	R2 Score	MAE	MAPE	RMSE	OPE
Transformers	0.112	55.89	9.97	56.1	9.9
Exponential Smoothing	0.209	36.87	6.43	46.90	1.413
TCN	0.21	14.23	2.39	19.58	0.82
N-BEATS	0.7	14.36	2.36	19.32	0.91

The TCN and N-BEATS model have performed in almost similar fashion; however, the R2 score for N-BEATS model is much higher than the TCN model. A higher value (close to 1) of R2 score and lower values of the other metrics represents a high performing model. As it has been observed based on the performance measurement metrics the overall performance of N-BEATS model for forecasting CO₂ emission in Saudi Arabia is better in comparison to the other models.

However, the main limitation of these results is the limited amount of the sample dataset.

B. Proposed Strategies

Saudi Arabia is one of the top oil producing countries in the world and one of the major carbon dioxide (CO₂) emitters. Therefore, it is crucial to develop effective CO₂ emission mitigation strategies to achieve its vision 2030 goals. This subsection proposes a number of emission mitigation strategies as following:

- Carbon capture and storage: Due to being an oil-based economy, the country has a large number of power plants, refineries, and other industrial process that produces CO₂. Investing and implementing carbon capture and storage technologies will help in capturing CO₂ from various sources and the stored CO₂ can be utilized for enhanced oil recovery or stored in underground geological formations.
- Green hydrogen: The fossil fuels can be replaced by green hydrogen as a cleaner alternative in transportation and industry.

- Promoting renewable energy sources: The kingdom has a vast land area that is not being used for any purpose and has a high potential to harness renewable energy such as wind power and solar energy. The country has already started investing in these sources of energy and plans to generate 58.7 GW of renewable energy by 2030, which will contribute significantly to reducing its CO₂ emissions. Encouraging its residents and citizens to utilize renewable energy sources and providing facilities and availability for energy generation from renewable energy sources such as using solar panels will contribute to reduce CO₂ emission.
- Reforestation and afforestation: Forests are natural CO₂ sequester. Planting trees and creating new forests will aid in mitigating CO₂ emissions.
- Energy efficiency: Improving energy efficiency in buildings, industry, and transportation can significantly reduce CO₂ emissions.
- Carbon Tax: A carbon tax or cap-and-trade system can provide an economic incentive for reducing CO₂ emissions.

By implementing proposed CO₂ emission mitigation strategies, the kingdom can reduce its CO₂ emission and contribute to global efforts to combat climate change.

VI. CONCLUSION

Accomplishing the KSA's goals of vision 2030 require to take actions against climate change. Adopting an apt method for prediction the CO₂ emission will assist in developing the policies for climate change mitigation. Emission analysis from various industries will aid in recognizing the areas of having alarming level of emissions and thus will be to focus the specific industry to put a rein on its emission [32] [33]. Analyzing the emission trends and forecasting the emission levels will provide the future directions of achieving sustainable development goals of the Kingdom as a member state of United Nation. The findings suggest that effective policy measures and technological interventions can help to minimize the negative impacts of climate change on sustainable development goals and achieving the vision 2030. Overall, the paper contributes to the growing body of research on climate change and its implications for sustainable development.

This research has implemented several machine learning and statistical methods include such as Exponential Smoothing, Transformers, Temporal Convolutional Network (TCN), and neural basis expansion analysis for interpretable time series. The data for training these models have been collected and synthesized from various sources using a web crawler. The performance of these models has been evaluated using various performance measurement metrics such as RMSE, R2 score, MAE, MAPE and OPE. The results indicate that the TCN and N-BEATS models perform similarly, with the N-BEATS model achieving a higher R2 score. The N-BEATS model has demonstrated strong predictive power with an R2 score of 0.7, an MAE score of 14.36, an MAPE score of 2.36, an RMSE score of 19.32, and an OPE score of 0.91. Overall, N-BEATS

model, in particular, shows promise for achieving accurate and interpretable predictions. Further research could explore the applicability of these techniques in other domains and investigate approaches for enhancing model performance.

ACKNOWLEDGMENT

The authors extend their appreciation to the Deanship of Scientific Research at Saudi Electronic University for funding this research work through the project number (8111).

REFERENCES

- [1] Qader, M. R., Khan, S., Kamal, M., Usman, M., & Haseeb, M. (2021). Forecasting carbon emissions due to electricity power generation in Bahrain. *Environmental Science and Pollution Research*, 1-12.
- [2] VNR Report, (2018), First Voluntary national review, https://saudiarabia.un.org/sites/default/files/2020-02/VNR_Report972018_FINAL.pdf, accessed Aug 15, 2021
- [3] Abubakar, I. R., & Dano, U. L. (2020). Sustainable urban planning strategies for mitigating climate change in Saudi Arabia. *Environment, Development and Sustainability*, 22(6), 5129-5152.
- [4] Destek, M.A., Sarkodie, S.A., 2019. Investigation of environmental Kuznets curve for ecological footprint: the role of energy and financial development. *Science of the Total Environment*, 650, 2483-2489.
- [5] Xiang, X., Li, Q., Khan, S., & Khalaf, O. I. (2021). Urban water resource management for sustainable environment planning using artificial intelligence techniques. *Environmental Impact Assessment Review*, 86, 106515.
- [6] Nasa (2020), Carbon Dioxide, <https://climate.nasa.gov/vital-signs/carbon-dioxide/> published November 2020, accessed August 2021.
- [7] Qader, M. R. (2009). Electricity consumption and GHG emissions in GCC countries. *Energies*, 2(4), 1201-1213.
- [8] Rogelj, J. et al. (2018). "Mitigation Pathways Compatible with 1.5°C in the Context of Sustainable Development", in Masson-Delmotte, V. et al. (eds) *Global Warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above preindustrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change*. Geneva, Switzerland: IPCC. https://www.ipcc.ch/site/assets/uploads/sites/2/2019/05/SR15_Chapter2_Low_Res.pdf
- [9] Malik, K., Rahman, S., Khondaker, A., Abubakar, I., Aina, Y. & Hasan, M. (2019). "Renewable energy utilization to promote sustainability in GCC countries: policies, drivers, and barriers" *Environmental Science and Pollution Research* volume 26, pages 20798–20814.
- [10] Abubakar, I. R., & Aina, Y. A. (2018). Achieving sustainable cities in Saudi Arabia: Juggling the competing urbanization challenges. In *E-planning and collaboration: Concepts, methodologies, tools, and applications* (pp. 234-255). IGI Global.
- [11] Gazzeh, K. & Abubakar, I. (2018). "Regional disparity in access to basic public services in Saudi Arabia: A sustainability challenge" *Utilities Policy* 52, pages 70–80
- [12] Aina, Y. A., Wafer, A., Ahmed, F., & Alshuwaikhat, H. M. (2019). Top-down sustainable urban development? Urban governance transformation in Saudi Arabia. *Cities*, 90, 272-281.
- [13] World Bank, (2022), CO2 emissions (metric tons per capita) - Saudi Arabia, <https://data.worldbank.org/indicator/EN.ATM.CO2E.PC?locations=SA>, accessed on August 20, 2022
- [14] BP, (2020), Statistical Review of World Energy 2020, <https://www.bp.com/content/dam/bp/business-sites/en/global/corporate/pdfs/energy-economics/statistical-review/bp-stats-review-2020-full-report.pdf>, accessed July. 27, 2022
- [15] Our World in Data, 2022, CO₂ Data Explorer, <https://ourworldindata.org/> Accessed July 27, 2022
- [16] Khan, S., Rabbani, M. R., Bashar, A., & Kamal, M. (2021, December). Stock Price Forecasting Using Deep Learning Model. In *2021 International Conference on Decision Aid Sciences and Application (DASA)* (pp. 215-219). IEEE.
- [17] Alhazmi, S., Khan, S., & Syed, M. H. (2023). Learning-Related Sentiment Detection, Classification, and Application for a Quality Education Using Artificial Intelligence Techniques. *Intelligent Automation & Soft Computing*, 36(3).
- [18] Shahnawaz, S., & B Mishra, R. (2012). A neural network based approach for English to Hindi machine translation. *International Journal of Computer Applications*, 53(18), 50-56.
- [19] Khan, S., Alourani, A., Mishra, B., Ali, A., & Kamal, M. (2022). Developing a Credit Card Fraud Detection Model using Machine Learning Approaches. *International Journal of Advanced Computer Science and Applications*, 13(3).
- [20] Khan, S., Mishra, B., Ali, A., Kamal, M., Qader, M. R., & Haider, M. (2021, December). Face Mask Detection from Live-Stream Surveillance Video using Convolutional Neural Network. In *2021 International Conference on Decision Aid Sciences and Application (DASA)* (pp. 62-66). IEEE.
- [21] Brown, R. G. (1959). *Statistical forecasting for inventory control*. McGraw/Hill.
- [22] Holt, C. E. (1957). *Forecasting seasonals and trends by exponentially weighted averages* (O.N.R. Memorandum No. 52). Carnegie Institute of Technology, Pittsburgh USA. <https://doi.org/10.1016/j.ijforecast.2003.09.015>
- [23] Winters, P. R. (1960). Forecasting sales by exponentially weighted moving averages. *Management Science*, 6, 324–342. <https://doi.org/10.1287/mnsc.6.3.324>
- [24] Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: principles and practice*, OTexts. Accessed from Chapter 7 Exponential smoothing
- [25] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł. and Polosukhin, I., (2017). Attention is all you need. *Advances in neural information processing systems*, 30.
- [26] Shazeer, N. (2020). Glu variants improve transformer. *arXiv preprint arXiv:2002.05202*.
- [27] Agarap, A.F., 2018. Deep learning using rectified linear units (relu). *arXiv preprint arXiv:1803.08375*.
- [28] Bai, S., Kolter, J. Z., & Koltun, V. (2018). An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv preprint arXiv:1803.01271*.
- [29] Hewage, P., Behera, A., Trovati, M., Pereira, E., Ghahremani, M., Palmieri, F., & Liu, Y. (2020). Temporal convolutional neural (TCN) network for an effective weather forecasting using time-series data from the local weather station. *Soft Computing*, 24, 16453-16482.
- [30] Oreshkin, B. N., Carpov, D., Chapados, N., & Bengio, Y. (2019). N-BEATS: Neural basis expansion analysis for interpretable time series forecasting. *arXiv preprint arXiv:1905.10437*.
- [31] Papastefanopoulos, V., Linardatos, P., & Kotsiantis, S. (2020). COVID-19: a comparison of time series methods to forecast percentage of active cases per population. *Applied sciences*, 10(11), 3880.
- [32] Sharif, H. O., Al-Juaidi, F. H., Al-Othman, A., Al-Dousary, I., Fadda, E., Jamal-Uddeen, S., & Elhassan, A. (2016). Flood hazards in an urbanizing watershed in Riyadh, Saudi Arabia. *Geomatics, Natural Hazards and Risk*, 7(2), 702-720.
- [33] Amran, Y. A., Amran, Y. M., Alyousef, R., & Alabduljabbar, H. (2020). Renewable and sustainable energy production in Saudi Arabia according to Saudi Vision 2030; Current status and future prospects. *Journal of Cleaner Production*, 247, 119602.

Implementation of Revised Heuristic Knowledge in Average-based Interval for Fuzzy Time Series Forecasting of Tuberculosis Cases in Sabah

Suriana Lasaraiya¹, Suzelawati Zenian², Risman Mat Hasim³, Azmirul Ashaari⁴

Preparatory Centre for Science and Technology, Universiti Malaysia Sabah, Kota Kinabalu, Sabah, Malaysia¹

Faculty of Science and Natural Resources, Universiti Malaysia Sabah, Kota Kinabalu, Sabah, Malaysia^{1,2}

Dept. of Mathematics and Statistics-Faculty of Science, Universiti Putra Malaysia, Serdang, Selangor, Malaysia³

Azman Hashim International Business School, Universiti of Technology Malaysia, Johor Bahru, Johor, Malaysia⁴

Abstract—Fuzzy time series forecasting is one method used to forecast in certain reality problems. The research on fuzzy time series forecasting has been increased due to its capability in dealing with vagueness and uncertainty. In this paper, we are dealing with implementation of revised heuristic knowledge to basic average-based interval and showing that these models forecast better than the basic one. We suggest three different lengths of interval, size 5, size 10 and size 20 to be used in comparing these models of average-based interval, average-based interval with implementation of heuristic knowledge and, average-based interval with implementation of revised heuristic knowledge. These models applied to forecast the number of tuberculosis cases reported monthly in Sabah starting from January 2012 until December 2019. A few numerical examples are shown as well. The performances of evaluations are shown by comparison on the values obtained by Mean Square error (MSE) and Root Mean Square Error (RMSE).

Keywords—Fuzzy time series; forecasting; length of interval; average-based interval; heuristic knowledge

I. INTRODUCTION

Time series analysis, at the present time has become a very important research object since it has been utilized in large number of fields such as finance, health care, education and environmental. In all these fields, obtaining an accurate picture of the future is an essential requirement and this can only be possible with an appropriate and competent prediction tool. Since the main concern of the time series analysis is to predict the future, it can be considered as an inference problem [1]. We identify that there are three main kinds of inference system (IS) which are utilized in time-series analysis as a prediction tool, i.e Statistical Inference System (SIS), Fuzzy Inference System (FIS) and computational inference system (CIS).

Fuzzy time series (FTS) methods are one of the prediction tool groups based on fuzzy set theory designed for time-series and they take into consideration the dependency structure of time series. In FTS model, time series with crisp observations, by using various fuzzification techniques, is transformed into time series with fuzzy observations called fuzzy time series.

The concept of linguistics variables and their application reasoning was initially introduced by Zadeh [2], [3], [4], [5]. Song and Chissom [6], [7] were the first researchers who

implemented and developed models to forecast in fuzzy time series. Their model is initially used to forecast the number of students enrolled in the University of Alabama from the year 1971 until year 1992. At the same time, they forecast using linear regression technique and compared the results obtained from time series technique. However, their methods require a large amount of computational time since their proposed methods is based on complex matrix operations. Due to limitation on time constraints, Chens [8] simplified the Song and Chissom methods by using simpler arithmetic operations and proved that his methods forecast better than the previous. Chens [9] continues his research, applying the same methods in [8], but he is focusing on high order fuzzy time series, which is more than one fuzzy logical relationship.

Huang [10] has been concerned on dealing with the suitable length of interval. A key point in choosing effective length of intervals is that they should not be too large as there will be no fluctuations in the fuzzy time series or too small as the meaning of fuzzy time series will be diminished. In his research, he proposed distribution- and average-based length in determining the effective length of intervals and proved that average-based length increases forecast accuracy compared to distribution-based length. Meanwhile, Li and Chen [11] proposed a concept of 3-4-5 rules of natural partitions to determine the length of intervals to be used, in their forecasting models.

In 2005, Huang and Yu [12] proposed a type-2 fuzzy time series model for stock index forecasting. They proved that type-2 model is much better for smooth defuzzification process and forecast consistently. Similar work on fuzzy time series can be found in Lee, et al. (2001) [13] and Tsai and Wu (1999) [14]. The other researchers such as Liu (2007) [15], Ozge, et al. (2020) [16] and Lee and Chou (2007) [17] does not specify on how they find the suitable the length of interval in their proposed methods. More research on defining the suitable interval also found in Chen (2014) [18], Wang et.al (2015) [19] as well as Lu et al (2015) [20] in their proposed model.

Ramli and Mohamad (2017) [21] applied the models from [11] in their forecasting models. They used 3-4-5 natural partition rules to find the best length of intervals in their models in forecasting the unemployment rate under different degrees of confidence. The study [22] shows a comparison of

the fuzzy time series methods of Chen, Cheng and Markov Chain model in predicting rainfall in Medan. In their study, Chen's model gives MAPE=8.002%, Markov chain's model gives MAPE=30.12% and Cheng's model gives MAPE=34.5%. According to [23] a model has a very good performance if the MAPE value is below 10% and has a good performance if the MAPE value is between 10% and 20%. Thus, Chen's model forecasting performances was very Good.

Norhayati et al. (2019) [24] proposed that fuzzy time series method is way more accurate compared to geometric brownian motion in forecasting stock prices in Bursa Malaysia. They apply average-based length focusing on trapezoidal membership members in their model Meanwhile, Susilo et al. (2022) [25] applied average-based fuzzy time series markov chain with some modification on partitioning using frequency density in predicting Covid-19 in Central Java, the ideal interval. The average based FTS Markov Chain approach with adjustments to the frequency density partition achieves an accuracy rate of 89.3%, according to the findings of this study. Lasaraiya et. al. (2022) [27] shows that determination of the length of interval by average-based interval gives nearest forecasting value to the actual values compared to forecasting value obtained from determination of length of interval by natural partitions as it shows the smallest percentage error.

This paper will merge Chen's model which is more on determination of suitable length of interval and Huarng's model which is more on choosing the suitable fuzzy logical relationship group (FLRG) in the model. Then, the comparison between the results of Chen's model with basic average-based interval, Huarng's model with average-based interval with implementation of heuristic knowledge and the proposed average-based interval model with implementation of revised heuristic knowledge will be shown. Section II will briefly explain the fuzzy time series background. Section III will be on the general step in the proposed methods. The Section IV presents results and discussion followed by future work and conclusions.

II. FUZZY TIME SERIES

Fuzzy time series concept was proposed by Song and Chissom [6], [7]. It can be used as prediction tools in a real-life problem or cases where historical data are formed in linguistic value. This means that the data in fuzzy time series is linguistic data while the actual data is real number. The definition of fuzzy time series is briefly reviewed below.

Let U be the universe of discourse, where $U = u_1, u_2, \dots, u_n$. A fuzzy set of A_i of U is defined by

$$A_i = \frac{f_{A_i}(u_1)}{u_1} + \frac{f_{A_i}(u_2)}{u_2} + \dots + \frac{f_{A_i}(u_n)}{u_n} \quad (1)$$

where f_{A_i} is the membership function of the fuzzy set A_i , $f_{A_i}: U \rightarrow [0,1]$. u_k is the element of fuzzy set A_i , and $f_{A_i}(u_k)$ is the degree of belongingness of u_k to A_i for $f_{A_i}(u_k) \in [0, 1]$ and $1 \leq k \leq n$.

Definition 1: $Y(t)(t = \dots, 0, 1, 2, \dots)$, is a subset of R . Let $Y(t)$ be the universe of discourse defined by fuzzy set $f_i(t)(i = 1, 2, \dots)$. $F(t)$ is defined as a fuzzy time series on $Y(t)(t = \dots, 0, 1, 2, \dots)$

According to Definition 1, $F(t)$ can be understood as a linguistics variable and $f_i(t)(i = 1, 2, \dots)$ as the possible linguistics values of $F(t)$.

Definition 2: If there exist a fuzzy relationship $R(t - 1, t)$, such that $F(t) = F(t - 1) \times R(t - 1, t)$, where \times is an operator, then $F(t)$ is said to be caused by $F(t - 1)$. The relationship between $F(t)$ and $F(t - 1)$ can be denoted by

$$F(t - 1) \rightarrow F(t) \quad (2)$$

Definition 3: Suppose $F(t - 1) = A_i$ and $F(t) = A_j$, a fuzzy logical relationship (FLR) is defined as $A_i \rightarrow A_j$, where A_i and A_j is on the left- and right-hand side respectively, while the repeated fuzzy logical relationship is removed.

Definition 4: Fuzzy logical relationship can be group together into fuzzy logical relationship group (FLRG) according to the same left-hand sides FLR.

According to Chen's model [8], the repeated fuzzy sets will be removed in the FLRG. For example,

$$\begin{array}{l} A_i \rightarrow A_{j1} \\ A_i \rightarrow A_{j2} \\ \dots \end{array} \rightarrow A_i \rightarrow A_{j1}, A_{j2}, \dots \quad (3)$$

Heuristic knowledge is introduced by [26] and this is used to guide the selection of proper fuzzy set for fuzzy logical relationship group (FLRG) before forecasting step. The definition is as shown below.

Definition 5: In the heuristic models, heuristic functions take fuzzy logical relationship groups and relevant variables as parameters. From these fuzzy logical relationship groups, the heuristic functions use the variables to select proper fuzzy set to establish heuristic fuzzy logical relationship groups.

All fuzzy sets A_1, A_2, \dots, A_n are well ordered. In other word, $A_k \geq A_l$ if $k > l$. Suppose $F(t - 1) = A_j$ and the FLRG for A_j is given by $A_j \rightarrow A_q, A_r, \dots$. Proper fuzzy sets, $A_{p1}, A_{p2}, \dots, A_{pk}$, can be selected by heuristic function $h()$,

$$h(x_1, x_2, \dots; A_q, A_r, \dots) = A_{p1}, A_{p2}, \dots, A_{pk} \quad (4)$$

where x_1, x_2, \dots are heuristic variables; $A_{p1}, A_{p2}, \dots, A_{pk}$ are selected from A_q, A_r, \dots , by the heuristics function. A heuristic fuzzy logical relationship group is obtained below:

$$A_j \rightarrow A_{p1}, A_{p2}, \dots, A_{pk} \quad (5)$$

The heuristic fuzzy logical relationship group is then used to forecast $F(t)$.

III. THE PROPOSED METHOD

Average-based intervals and Average-based intervals with implementation of heuristic knowledge were introduced a few years ago in forecasting the number of students' enrollments in University of Alabama. These models were hugely applied in forecasting in other real-life problems. In this paper, we try to show the Average-based intervals with implementation of revised heuristic knowledge forecast better than the other two models. There are five steps involved in this proposed method. The data as shown in Table I, was obtained from the Hospital of Queen Elizabeth II.

TABLE I. THE NUMBER OF TUBERCULOSIS CASES REPORTED MONTHLY IN SABAH FROM JANUARY 2012 UNTIL DECEMBER 2019

Month/ Year	Cases	Month/ Year	Cases	Month/ Year	Cases	Month/ Year	Cases
Jan 2012	199	Jan 2014	394	Jan 2016	338	Jan 2018	392
Feb 2012	487	Feb 2014	353	Feb 2016	372	Feb 2018	393
Mar 2012	356	Mar 2014	341	Mar 2016	505	Mar 2018	371
Apr 2012	364	Apr 2014	495	Apr 2016	394	Apr 2018	482
May 2012	419	May 2014	356	May 2016	401	May 2018	307
Jun 2012	349	Jun 2014	395	Jun 2016	401	Jun 2018	381
July 2012	341	July 2014	389	July 2016	305	July 2018	515
Aug 2012	422	Aug 2014	406	Aug 2016	482	Aug 2018	335
Sep 2012	365	Sep 2014	332	Sep 2016	410	Sep 2018	372
Oct 2012	380	Oct 2014	428	Oct 2016	311	Oct 2018	573
Nov 2012	351	Nov 2014	376	Nov 2016	601	Nov 2018	351
Dec 2012	353	Dec 2014	482	Dec 2016	433	Dec 2018	572
Jan 2013	380	Jan 2015	296	Jan 2017	299	Jan 2019	382
Feb 2013	385	Feb 2015	345	Feb 2017	390	Feb 2019	396
Mar 2013	379	Mar 2015	330	Mar 2017	446	Mar 2019	375
Apr 2013	376	Apr 2015	418	Apr 2017	366	Apr 2019	495
May 2013	402	May 2015	321	May 2017	437	May 2019	376
Jun 2013	342	Jun 2015	404	Jun 2017	428	Jun 2019	331
July 2013	432	July 2015	332	July 2017	404	July 2019	590
Aug 2013	324	Aug 2015	301	Aug 2017	472	Aug 2019	431
Sep 2013	388	Sep 2015	397	Sep 2017	347	Sep 2019	500
Oct 2013	389	Oct 2015	385	Oct 2017	394	Oct 2019	442
Nov 2013	359	Nov 2015	393	Nov 2017	532	Nov 2019	440
Dec 2013	370	Dec 2015	542	Dec 2017	590	Dec 2019	524

^a Data obtained from Queen Elizabeth II Hospital, Kota Kinabalu, Sabah.

Step 1: Define the universe of discourse, $U = [U_{min} - D_1, U_{max} + D_2]$, where U_{min} and U_{max} are minimum and maximum values in raw data, D_1, D_2 are two real numbers. U will be divided into n equal length of intervals u_1, u_2, \dots, u_n . We applied Chen's method [8] with some modifications in finding the length of intervals. The algorithm in average based interval by [10] to set the suitable length of interval is according to Table II.

TABLE II. BASIS MAPPING TABLE

Range	Basis
0.1 – 1.0	0.1
1.1 - 10	1
11 - 100	10
101 - 1000	100

^b Basis Mapping Table by [10].

Step 2: Fuzzy sets A_i . The linguistics variable is the raw data, A_i as possible linguistics values of the raw data. Each is defined by the intervals u_1, u_2, \dots, u_n .

$$A_1 = 1/u_1 + 0.5/u_2 + 0/u_3 + 0/u_4 + \dots + 0/u_{n-1} + 0/u_n$$

$$A_2 = 0.5/u_1 + 1/u_2 + 0.5/u_3 + 0/u_4 + \dots + 0/u_{n-1} + 0/u_n$$

$$A_3 = 0/u_1 + 0.5/u_2 + 1/u_3 + 0.5/u_4 + \dots + 0/u_{n-1} + 0/u_n$$

...

$$A_n = 0/u_1 + 0/u_2 + 0/u_3 + 0/u_4 + \dots + 0.5/u_{n-1} + 1/u_n$$

Step 3: Defined fuzzy logical relationship FLR. Repeated FLR will be removed.

Step 4: Group FLR in fuzzy logical relationship group FLRG.

Step 5: Calculated the forecasted output. There are 3 rules involved in Chen's Rule.

- Rule 1: If the fuzzified number of cases of day i is A_i , and there is only one FLR in the FLRG obtained in STEP 4, that is $A_i \rightarrow A_k$, whereby A_k occurs in the interval u_k and the midpoint of u_k is m_k , then the forecast number of cases of day $i + 1$ is m_k .
- Rule 2: If the fuzzified number of cases of day A_i and there is more than one FLR in the FLRG obtained in STEP 4, that is, $A_i \rightarrow A_{k1}, A_i \rightarrow A_{k2}, A_i \rightarrow A_{k3}, \dots, A_i \rightarrow A_{kq}$ whereby $A_{k1}, A_{k2}, A_{k3}, \dots, A_{kq}$ occurs in the intervals $u_{k1}, u_{k2}, u_{k3}, \dots, u_{kq}$ and their midpoint are $m_{k1}, m_{k2}, m_{k3}, \dots, m_{kq}$, the the forecast number of cases of the day $i + 1$ is $\frac{m_{k1} + m_{k2} + m_{k3} + \dots + m_{kq}}{q}$.
- If the fuzzified number of cases of day i is A_i , and there is empty FLR in the FLRG obtained in STEP 4, that is $A_i \rightarrow \emptyset$, whereby A_i occurs in the intervals u_i and the midpoint of u_i is m_i , then the forecast number of cases of the day $i + 1$ is m_i .

IV. RESULTS AND DISCUSSION

According to Section III, there are 5 steps counted. The data used in this study is the number of tuberculosis cases reported monthly in Sabah, period of January 2012 to December 2019 as you can see from Fig. 1. The time series graphic as in Fig. 1 demonstrates that there is no discernible pattern in the number of tuberculosis cases reported. This can be caused by any unexpected factors, such as gender, family background, knowledge, etc., and did not become observation in this study.

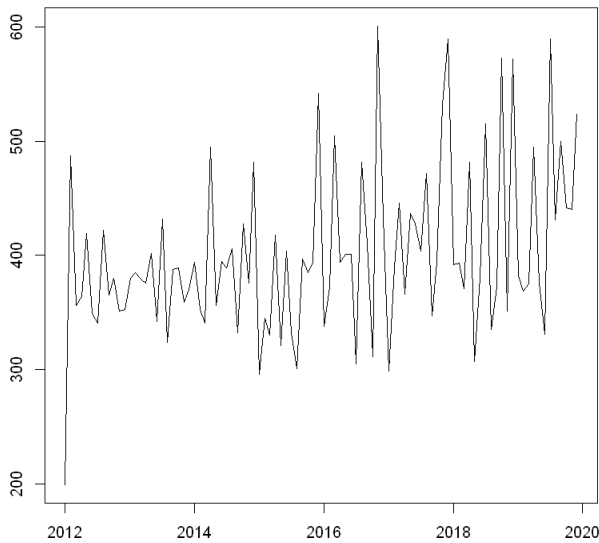


Fig. 1. Plot actual number of tuberculosis cases reported in sabah period January 2012 – December 2019.

A. Length of Interval

The first modification was done in determining the length of intervals. According to the data in Table I, $U_{min} = 199$ and $U_{max} = 601$ respectively. We choose $D_1 = 9$ and $D_2 = 9$ respectively. Thus, $U = [199 - 9, 601 + 9] = [190, 610]$. The modifications (b) and (c) are as follows:

- Based on Table II, the length of interval obtained is 10.
- The length of interval is then divided by 2, giving the length of interval is 5.
- The length of interval is then multiplied by 2, giving the length of interval is 20.

Then, the lengths of intervals considered are size 5, size 10 and size 20, and hence list of intervals is shown below. All three sizes applied to the average-based interval model, average-based interval model with implementation of heuristic knowledge and average-based interval model with implementation of revised heuristic knowledge.

TABLE III. THE LENGTH OF INTERVALS AND LIST OF INTERVALS

Length of intervals	Number of intervals	List of intervals
Size 5	84	$u_1 = [190 - 195], u_2 = [195 - 200], \dots, u_{84} = [605 - 610]$
Size 10	42	$u_1 = [190 - 200], u_2 = [200 - 210], \dots, u_{42} = [600 - 610]$
Size 20	21	$u_1 = [190 - 210], u_2 = [210 - 230], \dots, u_{21} = [590 - 610]$

^c. Jupyter Notebook tools under R kernel

B. Fuzzy Logical Relationship (FLR)

Next, is to find FLR. The number of tuberculosis cases reported monthly is to be defined according to intervals as shown in Table III and IV. The same step is repeated for all three models: average-based interval, average-based interval with implementation of heuristic knowledge and proposed average-based interval with implementation of revised heuristic knowledge.

TABLE IV. THE FUZZY SETS, A_i

Length of intervals	Number of intervals	List of intervals
Size 5	84	$A_1 = 1/u_1 + 0.5/u_2 + 0/u_3 + 0/u_4 + \dots + 0/u_{n-1} + 0/u_{84}$ $A_2 = 0.5/u_1 + 1/u_2 + 0.5/u_3 + 0/u_4 + \dots + 0/u_{n-1} + 0/u_{84}$ $A_3 = 0/u_1 + 0.5/u_2 + 1/u_3 + 0.5/u_4 + \dots + 0/u_{n-1} + 0/u_{84}$... $A_{84} = 0/u_1 + 0/u_2 + 0/u_3 + 0/u_4 + \dots + 0.5/u_{n-1} + 1/u_{84}$
Size 10	42	$A_1 = 1/u_1 + 0.5/u_2 + 0/u_3 + 0/u_4 + \dots + 0/u_{n-1} + 0/u_{42}$ $A_2 = 0.5/u_1 + 1/u_2 + 0.5/u_3 + 0/u_4 + \dots + 0/u_{n-1} + 0/u_{42}$ $A_3 = 0/u_1 + 0.5/u_2 + 1/u_3 + 0.5/u_4 + \dots + 0/u_{n-1} + 0/u_{42}$... $A_{42} = 0/u_1 + 0/u_2 + 0/u_3 + 0/u_4 + \dots + 0.5/u_{n-1} + 1/u_{42}$
Size 20	21	$A_1 = 1/u_1 + 0.5/u_2 + 0/u_3 + 0/u_4 + \dots + 0/u_{n-1} + 0/u_{21}$ $A_2 = 0.5/u_1 + 1/u_2 + 0.5/u_3 + 0/u_4 + \dots + 0/u_{n-1} + 0/u_{21}$ $A_3 = 0/u_1 + 0.5/u_2 + 1/u_3 + 0.5/u_4 + \dots + 0/u_{n-1} + 0/u_{21}$... $A_{21} = 0/u_1 + 0/u_2 + 0/u_3 + 0/u_4 + \dots + 0.5/u_{n-1} + 1/u_{21}$

^d. Jupyter Notebook tools under R kernel

C. Fuzzy Logical Relationship Group (FLRG)

Group of FLR is defined according to same fuzzy set in left hand side of the FLR. As for example, numerical FLR in the model for size 5 is shown. FLRG for fuzzy set A_{33} is chosen.

1) Average-based interval: FLRG will be grouped according to Definition 3. Refer Fig. 2.

FLRG for item (12) is $A_{33} \rightarrow A_{31}, A_{33}, A_{38}, A_{77}$

FLRG for item (13) is $A_{33} \rightarrow A_{31}, A_{33}, A_{38}, A_{77}$

FLRG for item (27) is $A_{33} \rightarrow A_{31}, A_{33}, A_{38}, A_{77}$

FLRG for item (84) is $A_{33} \rightarrow A_{31}, A_{33}, A_{38}, A_{77}$

```

File Edit View Insert Cell Kernel Help
In [65]: FLRG_33
fuzzification left right fuzzylogicalrelationship
12 33 33 33 33->33
13 38 33 38 33->38
27 31 33 31 33->31
84 77 33 77 33->77

In [66]: Forecast <- (midpoint[33,1]+midpoint[38,1]+midpoint[31,1]+midpoint[77,1])/4
Forecast
newFLRG33 <- cbind(FLRG_33, Forecast)
newFLRG33
411.723214285714

fuzzification left right fuzzylogicalrelationship Forecast
12 33 33 33 33->33 411.7232
13 38 33 38 33->38 411.7232
27 31 33 31 33->31 411.7232
84 77 33 77 33->77 411.7232
    
```

Fig. 2. The FLRG of A_{33} in Average-Based Intervals for Size 5 is $A_{33} \rightarrow A_{31}, A_{33}, A_{38}, A_{77}$.

2) Average-based intervals with heuristic knowledge according to Definition 5. Let the month $t - 1$ be $F(t - 1)$ and the FLRG for $F(t - 1)$ be $A_j \rightarrow A_q, A_r, \dots$. This study assumes that there is heuristic knowledge showing the increment or decrement in the number of cases reported. The heuristic function is set as $h(x; A_q, A_r, \dots)$ where x is an indicator for number of cases forecasting. Fig. 3 shows the example on fuzzy sets of A_{33} . The heuristic FLRG is shown below.

FLRG for item (12) is $A_{33} \rightarrow A_{31}$

FLRG for item (13) is $A_{33} \rightarrow A_{33}, A_{38}, A_{77}$

FLRG for item (27) is $A_{33} \rightarrow A_{31}$

FLRG for item (84) is $A_{33} \rightarrow A_{33}, A_{38}, A_{77}$

```

File Edit View Insert Cell Kernel Help
In [65]: FLRG_33
fuzzification left right fuzzylogicalrelationship
12 33 33 33 33->33
13 38 33 38 33->38
27 31 33 31 33->31
84 77 33 77 33->77

In [66]: FLRgpredict <- (midpoint[33,1]+midpoint[38,1]+midpoint[31,1]+midpoint[77,1])/4
FLRgpredict
FLRgUp <- (midpoint[33,1]+midpoint[38,1]+midpoint[77,1])/3
FLRgUp
FLRgDown <- (midpoint[33,1]+midpoint[31,1])/2
FLRgDown
newFLRG33 <- cbind(FLRG_33, FLRgpredict, FLRgUp, FLRgDown)
newFLRG33
411.723214285714
434.585317460317
348.125

fuzzification left right fuzzylogicalrelationship FLRgpredict FLRgUp FLRgDown
12 33 33 33 33->33 411.7232 434.5853 348.125
13 38 33 38 33->38 411.7232 434.5853 348.125
27 31 33 31 33->31 411.7232 434.5853 348.125
84 77 33 77 33->77 411.7232 434.5853 348.125
    
```

Fig. 3. The FLRG of A_{33} in average-based intervals with heuristic knowledge for size 5 is $A_{33} \rightarrow A_{31}, A_{33}, A_{38}, A_{77}$.

3) Average-based intervals with revised heuristic knowledge. The modification is that, for the FLR who have the same value on left and right side, the whole FLRG will be chosen as the heuristic FLRG instead of looking back at the

raw data to look whether the value increase or decrease. Fig. 4 shows the example on fuzzy sets of A_{33} . The heuristic FLRG is shown below.

FLRG for item (12) is $A_{33} \rightarrow A_{31}, A_{33}, A_{38}, A_{77}$

FLRG for item (13) is $A_{33} \rightarrow A_{33}, A_{38}, A_{77}$

FLRG for item (27) is $A_{33} \rightarrow A_{31}$

FLRG for item (84) is $A_{33} \rightarrow A_{33}, A_{38}, A_{77}$

```

File Edit View Insert Cell Kernel Help
In [69]: FLRG_33
Month Cases AbsDiff fuzzification left right fuzzylogicalrelationship
12 2012-12-01 353 2 33 33 33 33->33
13 2013-01-01 380 27 38 33 38 33->38
27 2014-03-01 341 12 31 33 31 33->31
84 2018-12-01 572 221 77 33 77 33->77

In [70]: FLRgpredict <- (midpoint[33,1]+midpoint[38,1]+midpoint[31,1]+midpoint[77,1])/4
FLRgpredict
FLRgUp <- (midpoint[33,1]+midpoint[38,1]+midpoint[77,1])/3
FLRgUp
FLRgDown <- (midpoint[33,1]+midpoint[31,1])/2
FLRgDown
newFLRG33 <- cbind(FLRG_33, FLRgpredict, FLRgUp, FLRgDown)
newFLRG33
411.723214285714
434.585317460317
348.125

Month Cases AbsDiff fuzzification left right fuzzylogicalrelationship FLRgpredict FLRgUp FLRgDown
12 2012-12-01 353 2 33 33 33 33->33 411.7232 434.5853 348.125
13 2013-01-01 380 27 38 33 38 33->38 411.7232 434.5853 348.125
27 2014-03-01 341 12 31 33 31 33->31 411.7232 434.5853 348.125
84 2018-12-01 572 221 77 33 77 33->77 411.7232 434.5853 348.125
    
```

Fig. 4. The FLRG of A_{33} in average-based intervals with revised heuristic knowledge for size 5 is $A_{33} \rightarrow A_{31}, A_{33}, A_{38}, A_{77}$.

D. Forecasting

The output will be defined according to Step 5. The forecast number of tuberculosis cases with respect to length of intervals size 5, size 10 and size 20 for the three models; average based intervals, average-based intervals with implementation of heuristic knowledge and average-based intervals with implementation of revised heuristic knowledge is shown in Table V, Table VI, and Table VII.

TABLE V. THE FORECAST NUMBER OF CASES FOR AVERAGE-BASED INTERVAL

Month/ Year	Cases	Forecast cases (according to respected size of intervals)		
		Size 5	Size 10	Size 20
Feb 2012	487	487.7917	485.2976	480.3095
Mar 2012	356	358.1012	4340.6429	350.6190
Apr 2012	364	374.7282	402.1627	417.6020
May 2012	419	398.0060	405.4881	417.6020
Jun 2012	349	335.6548	368.9087	373.8968
July 2012	341	368.0774	403.8254	409.3677
...
July 2019	590	422.9464	422.9464	409.3677
Aug 2019	431	412.9702	415.4643	4395.5119
Sep 2019	500	373.0655	380.5476	385.5357
Oct 2019	442	442.8988	392.1865	395.5119
Nov 2019	440	425.4405	405.4881	385.5357
Dec 2019	524	389.6925	380.5476	385.5357

^e. Jupyter notebook tools under R kernel

TABLE VI. THE FORECAST NUMBER OF CASES FOR AVERAGE-BASED INTERVAL WITH IMPLEMENTATION OF HEURISTIC KNOWLEDGE

Month/ Year	Cases	Forecast cases (according to respected size of intervals)		
		Size 5	Size 10	Size 20
Feb 2012	487	487.7917	485.2976	480.3095
Mar 2012	356	385.1012	340.6429	350.6190
Apr 2012	364	374.7282	413.4690	430.4286
May 2012	419	398.0060	405.4881	430.4286
Jun 2012	349	335.6548	335.6548	360.5952
July 2012	341	343.1369	335.6548	320.6905
...
July 2019	590	462.8512	462.0198	437.5544
Aug 2019	431	412.9702	4415.4643	395.5119
Sep 2019	500	497.7679	510.2381	440.4048
Oct 2019	442	442.8988	392.1865	395.5119
Nov 2019	440	425.4405	405.4881	360.5952
Dec 2019	524	522.7083	510.2381	440.4048

^f. Jupyter notebook tools under R kernel

TABLE VII. THE FORECAST NUMBER OF CASES FOR AVERAGE-BASED INTERVAL WITH IMPLEMENTATION OF REVISED HEURISTIC KNOWLEDGE

Month/ Year	Cases	Forecast cases (according to respected size of intervals)		
		Size 5	Size 10	Size 20
Feb 2012	487	487.7917	485.2976	480.3095
Mar 2012	356	358.1012	340.6429	350.6190
Apr 2012	364	374.7282	413.4690	417.6020
May 2012	419	398.0060	405.4881	430.4286
Jun 2012	349	335.6548	335.6548	360.5952
July 2012	341	343.1369	403.8254	409.3677
...
July 2019	590	462.8512	462.0198	437.5544
Aug 2019	431	412.9702	415.4643	395.5119
Sep 2019	500	497.7679	510.2381	486.9603
Oct 2019	442	442.8988	392.1865	395.5119
Nov 2019	440	425.4405	405.4881	409.0510
Dec 2019	524	522.7083	510.2381	486.9603

^g. Jupyter notebook tools under R kernel

E. Accuracy Validation

The performance for the proposed method is compared based on their value of Mean Square Error (MSE) and Root Mean Square Error (RMSE). The smaller the MSE value, the better the forecast. Meanwhile, RMSE is known to be a good measurement of how accurate the model is in prediction. Lower values of RMSE indicate better fit. Based on Table

VIII, the average-based intervals with implementation of revised heuristic knowledge for size 5 give the smallest MSE = 1309.11 and RMSE = 36.18168 respectively compared to average-based interval and average-based interval with implementation of heuristic knowledge. Based on this MSE and RMSE value, it is concluded that the average-based intervals with implementation of revised heuristic knowledge model increase the accuracy of the forecasting value.

TABLE VIII. THE COMPARISON ON MSE AND RMSE OF EACH SIZE

Models	Length of intervals	MSE	RMSE
Average-based interval	Size 5	2618.065	51.15726
	Size 10	3337.299	57.76936
	Size 20	4392.890	66.2788
Average-based intervals with implementation of heuristic knowledge	Size 5	1436.139	37.89642
	Size 10	1926.855	43.89595
	Size 20	2624.269	51.22762
Average-based intervals with implementation of revised heuristic knowledge	Size 5	1309.114	36.18168
	Size 10	1887.522	43.44562
	Size 20	2289.208	47.84564

^h. Jupyter Notebook tools under R kernel

V. FUTURE WORK

Overall, this paper shows only the results on training phase of the comparing model and length of interval chosen size 5, size 10 and size 20. Training phase is done for the period of January 2012 until December 2019. Testing phase on the next year 2020 (January-December) can be done as a future work for length of size 5, 10 and 20, limited to list of FLRG obtained from the training phase. The forecasting results will be compared with the actual number of tuberculosis cases.

ACKNOWLEDGMENT

The author would like to thank the main supervisor, Dr. Suzelawati Zenian and co supervisor for their continuous support in this research. Besides, million thanks to Mdm. Nadirah Sulaiman from Clinical Research Centre (CRC), Hospital Queen Elizabeth (HQE) as well as Dr. Roddy Teo from Communicable Disease Unit, Sabah State Health Department for their help throughout this research.

REFERENCES

- [1] Ozge, C. Y., Eren, B., Erol, E. and Ufuk, Y. "A New Intuitionistic Fuzzy Functions Approach Based On Hesitation Margin For Time-Series Prediction". *Soft Computing* 2020, 24 pp. 8211-8222.
- [2] Zadeh, L.A. *Fuzzy Sets, Information and Control* 1965 **8**, pp. 338-353.
- [3] Zadeh, L.A. "The concept of Linguistic variable and its application to approximate reasoning, Part 1-3" *Information Science* 1975 **8**, pp. 199-249, 301-357; 43-80.
- [4] Zadeh, L.A. "Outline of a new approach to the analysis of complex system and decision process", in *IEEE Trans. System Man Cybernet* 1973 **3**, pp. 28-44.
- [5] Zadeh, L.A. "Similarity relations and fuzzy orderings", *Information Science* 1971 **3**, pp. 177-200.
- [6] Song, Q. and Chissom, B. S. Fuzzy forecasting enrollments with fuzzy time series-Part 1, *Fuzzy Sets and Systems* **54**, 1993a, pp. 1-9.

- [7] Song, Q. and Chissom, B. S. Fuzzy time series and its models. *Fuzzy Sets and Systems* **3(54)**, 1993b, pp. 269–277.
- [8] Chen, S. M., “Forecasting enrollments based on fuzzy time series”, *Fuzzy Sets and Systems* **81**, 1996, pp. 311-318.
- [9] Chen, S. M., Forecasting Enrollments Based on High Order Fuzzy Time Series. *Cybernetics and Systems.*, 2002 (**33**): pp.1-16.
- [10] Huarnq, K., Effective Lengths of Intervals to Improve Forecasting in Fuzzy Time Series. *Fuzzy Sets and Systems.*, 2001, **3(123)**: pp. 387–394.
- [11] Li, S. T. and Chen, Y. P. “Natural Partitioning-based Forecasting Model for Fuzzy Time Series”, 2004 Budapest, Hungary.
- [12] Huarnq, K. and Yu, T. F., “The Application of neural networks to forecast fuzzy time series” *Physica A: Statistical Mechanics and Its Applications* 2006, Vol. 363, Issue 2, pp 481-491.
- [13] Lee, T. S., Chiu, C. C. and Lin, F.C., “Prediction of the unemployment rate using fuzzy time series with Box-Jenkins methodology” *International Journal of Fuzzy Systems* 2001, Vol.3. No. 4. pp. 577-585.
- [14] Tsai, C. C. and Wu, S. J., “A study for second order modelling of fuzzy time series” *Proceedings of the IEEE International Conference on Fuzzy Systems* 1999, Vol. 2, pp. 719-725.
- [15] Liu, H. T. An improved fuzzy time series forecasting method using trapezoidal fuzzy numbers. *Fuzzy Optimisation Decision Making* 2007, **6** pp. 63-80.
- [16] Ozge, C. Y., Eren, B., Erol, E. and Ufuk, Y. “A new intuitionistic fuzzy functions approach based on hesitation margin for time-series prediction”. *Soft Computing* 2020, **24** pp. 8211-8222.
- [17] Lee, H. S. & Chou, M. T. “Fuzzy forecasting on fuzzy time series” *International Journal of Computer Mathematics* 2007 **81:7**, pp. 781-789.
- [18] Chen, M.Y., “A high order fuzzy time series forecasting model for internet stock trading”, *Future Generation Computer System* 37:461-467.
- [19] Wang, W. Liu, X., “Fuzzy forecasting based on automatic clustering and axiomatic fuzzy set classification”, *Inf. Sci. (Ny)* 294:78-94
- [20] Lu. W., Chen, X. Pedrycz, W., “ Using interval information granules to improve forecasting in fuzzy time series”, *Int. J. Approx Reason* 57:1-18
- [21] Ramli, N., M. S. M. A. and Mohamad, D. “Fuzzy time series forecasting model with natural partitioning length approach for predicting the unemployment rate under different degree of confidence” 2017, *AIP conference proceeding*.
- [22] Arnita, N. Afnisah, and F. Marpaung. “A Comparison of the Fuzzy Rime Series Methods of Chen, Cheng and Markov Chain in Predicting Rainfall in Medan.” *Journal of Physics.: Conference. Series.* 1462 (2020) 012044
- [23] Raharja, A., Anggraeni, W., dan Vinarti, R. A., 2010. Penerapan Metode Exponential Smoothing untuk Peramalan Penggunaan Waktu Telepon di PT. Telkomsel DIVRE3 Surabaya. *Jurnal Sistem Informasi.* Surabaya: Institut Teknologi Sepuluh Nopember.
- [24] Norhayati, S., Nur Ezzati, D.M.R., Rohana, A. and Nur Fatimah, F., “Fuzzy Time Series and Geometric Brownian Motion in Forecasting Stock Prices in Bursa Malaysia”, 2019, *Jurnal Intelek UITM Perlis*, 14:2, DOI: <http://10.24191/ji.v14i2.241>
- [25] Susilo, H., Zaenurrohman, Titi, U. S., “Average Based-FTS Markov Chain with Modifications to the Frequency Density Partition to Predict Covid-19 in Central Java”, *CAUCHY –Jurnal Matematika Murni dan Aplikasi* 2022, 7:2, 213-239
- [26] Huarnq, K., Heuristic models of fuzzy time series for forecasting. *Fuzzy Sets and Systems.*, 2001, **123**: pp. 369-386.
- [27] Lasaraiya, S., Gabda, D., Che Hussin, C. H. & Mandangan, A. (2022). “An effective length of intervals for forecasting in fuzzy time series”. *Journal of Islamic, Social, Economics and Development. (JISED)*, 7(47), 163-176. DOI: 10.55573/JISED.074717

Event Feature Pre-training Model Based on Public Opinion Evolution

WANG Nan¹, TAN Shu-Ru², XIE Xiao-Lan³, LI Hai-Rong⁴, JIANG Jia-Hui⁵

School of Management Science and Information Engineering, Jilin University of Finance and Economics, Changchun, China^{1,5}

School of Information Science and Engineering, Guilin University of Technology, Guilin, China^{2,3}

School of Information Engineering, Xinjiang Institute of Technology, Xin Jiang, China⁴

Abstract—The comments in the evolution of network public opinion events not only reflect the attitude of netizens towards the event itself, but also are the key basis for mastering the dynamics of public opinion. According to the comment data in the event evolution process, an event feature vector pre-training model NL2ER-Transformer is constructed to realize the real-time automatic extraction of event features. Firstly, a semi-supervised multi-label curriculum learning model is proposed to generate comment words, event word vectors, event words, and event sentences, so that a public opinion event is mapped into a sequence similar to vectorized natural language. Secondly, based on the Transformer structure, a training method is proposed to simulate the evolution process of events, so that the event vector generation model can learn the evolution law and the characteristics of reversal events. Finally, the event vectors generated by the presented NL2ER-Transformer model are compared with the event vectors generated by the current mainstream models such as XLNet and RoBERTa. This paper tests the pre-trained model NL2ER-Transformer and three pre-trained benchmark models on four downstream classification models. The experimental results show that using the vectors generated by NL2ER-Transformer to train downstream models compared to using the vectors generated by other pre-trained benchmark models to train downstream models, the accuracy, recall, and F1 values are 16.66%, 44.44%, and 19% higher than the best downstream model. At the same time, in the evolutionary capability analysis test, only four events show partial errors. In terms of performance of semi-supervised model, the proposed semi-supervised multi-label curriculum learning model outperforms mainstream models in four indicators by 6%, 23%, 8%, and 15%, respectively.

Keywords—Event vectorization; NL2ER-transformer model; public opinion reversal prediction; evolution of public opinion event; multi label semi supervised learning

I. INTRODUCTION

Network public opinion refers to the public's opinions on an event on the network media. When an event causes heated discussion among netizens and forms a certain scale of network public opinion, this work call it a network public opinion event, hereinafter also referred to as an event. Network public opinion often leads to various public opinion phenomena, such as public opinion reversal, network violence, secondary derived events, etc. Network public opinion prediction from the perspective of public opinion phenomenon recognition is one of the current research directions of public opinion prediction: by analyzing the influencing factors of public opinion phenomena, designing event feature vectors, and

further constructing machine learning classification models to predict various phenomena of public opinion events. This paper focuses on the upstream event pre-training model, while the downstream task takes the prediction of public opinion reversal [1] as an example to verify the model.

This paper maps the public opinion event vector construction problem to the pre-train model based natural language feature vector generation problem. Based on the Transformer structure, we build an event vector generation model NL2ER-Transformer (Natural Language to Event Representation from Transformer) with the ability to predict the event evolution. The generated event vector is more in line with the evolution characteristics of public opinion reversal event. It solves the problem that the current event vectorization models only consider the static subjective characteristics of events and have not learned the dynamic evolution characteristics. The specific contributions of this paper are summarized as follows:

- Combining manually designing features and event evolution analysis, we put forward the concepts of comment word, event word and event sentence. Based on the three concepts, this work transforms an event into a sequence, which is similar to vectorized natural language. This kind of representation form of event can enable the model to better learn the evolution characteristics of the event. In order to obtain comment words automatically, a semi-supervised multi-label curriculum learning model is proposed, which can automatically convert comments into discriminative feature vectors, and then construct event vectors based on them.
- Based on the sequence data proposed in this paper, by the training mode of simulating event evolution, an event vector generation model NL2ER-Transformer with event evolution prediction ability is constructed based on Transformer. It solves the problem that the current event vectorization models only consider the static subjective characteristics of events and have not learned the dynamic evolution characteristics. This paper has carried out several experiments to verify the effectiveness of the model. The experimental results show that using the event vectors constructed by NL2ER-Transformer to train the public opinion reversal prediction model are better than the classifiers trained

by the event vectors constructed by other event vectorization models.

II. RELATED WORK

Feature extraction of event representation vector is the most important step to build a network public opinion prediction model based on machine learning. At present, the feature extraction of event vectors mainly includes three methods: manually designing features, pre-training model and knowledge graph.

A. Manually Designing Features

At first, scholars generated event vectors by manually defining or extracting qualitative and quantitative event features from social media. Research [2] uses 33 artificial features to assign values to public opinion events; study [3] made a secondary improvement on the basis of [2], reducing 33 features to 30 features; author in [4] combines Twitter and Weibo, and proposes some characteristics to analyze public opinion based on their commonalities; researcher in [5] combines time features and proposes new event features for rumor analysis. The manually defined event features are very dependent on expert knowledge which usually can only be obtained by observing the influencing factors of one period or even the whole period of the event. Although the manually defined event features contain some event evolution information and the prediction model based on them can obtain better prediction accuracy, this method requires that the event has evolved for a period of time before the event features are obtained, and it is difficult to determine the event features at the early stage of the event [6][7]. Therefore, the model trained with manually designed event features has poor prediction ability for events with short occurrence time, and the prediction accuracy is not high before the occurrence of public opinion phenomena such as public opinion reversal.

B. Pre-Training Model

With the continuous development of computer hardware technology, people have made great progress in natural language vectorization, from the initial word2vec [8]-[10] to ELMo [8] and GPT [11] with context awareness as the core. With the proposal of Transformer [12] in 2017, natural language vectorization has made amazing progress again, and language transfer models have begun to flourish. At the same time, the vectorization of public opinion events, which is closely related to natural language, has also been correspondingly improved. For example, studies [13] and [14] transform event description information into event vectors based on Transformer structure; Mohammadreza Samadi[15]used RoBERTa[16], XLNET[17], BERT[18] and other transfer learning models to transform events into vectors, and tested them on multiple groups of classification models; SZU Yin Lin[19] and Guillermo Blanco[20] used Bert to transform event information into vectors for subsequent prediction. Such methods are the mainstream methods for vectorization of news events or public opinion events at present, but they only vectorize according to the event description text, and do not consider the dynamic factors of

event evolution, such as public sentiment and attitude in the manually designing features. The method generated event vectors do not have enough information to express the change and development of events and it is difficult to accurately predict the reversal phenomenon in the process of public opinion evolution.

C. Knowledge Graph

As the mainstream method of public opinion research, knowledge graph has also been used by many scholars to construct public opinion event vectors. For example, research [21] proposed a method to construct a knowledge graph of public opinion with online news comments, which is used for the decision-maker master the online public opinion quickly and directly; study [22] constructed an NPOKG based knowledge graph for extracting features of public opinion events; study [23] based on ELECTRA and REDP methods, entity extraction and relationship extraction are performed on public opinion text information respectively. Network public opinion knowledge graphs are constructed for Weibo platform and short video platform, and comparative analysis is conducted on each network public opinion knowledge graph. However, from the experimental results of the research, it can be seen that the knowledge items contained in Internet public opinion are relatively sparse, and a large number of entities extracted through text can only construct a very small number of nodes, so it is not suitable to be used as a vectorization means of public opinion events.

In view of the problems of the above event vectorization methods, on the basis of the influence factors of the reversal event and dynamic temporality of public opinion evolution, this paper maps the public opinion event vector construction problem to the pre-train model based natural language feature vector generation problem. Based on the Transformer structure, we build an event vector generation model NL2ER-Transformer (Natural Language to Event Representation from Transformer) with the ability to predict the event evolution. The generated event vector is more in line with the evolution characteristics of public opinion reversal event. The classifier can accurately predict the events that may be reversed only relying on the information before the event reversal.

III. EVENT VECTORIZATION AND TRAINING METHOD OF SIMULATING EVENT EVOLUTION

In this section, we show the model for event vectorization and training method of simulating event evolution. The model consists of two parts. The first part transforms an abstract event that has occurred for m days into a sequence based on netizens' comments, by defining comment word (automatically obtained through the proposed semi-supervised multi-label curriculum learning model in Section III B), event word vector, event word and event sentence; The second part uses the proposed NL2ER-Transformer model to generate event vector that contain the characteristics of the event itself and the evolution characteristics of the event. The model framework is shown in Fig. 1.

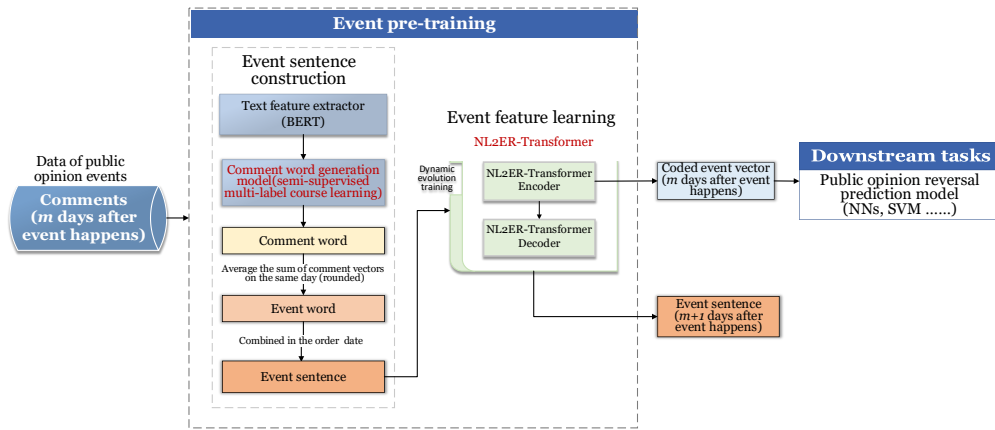


Fig. 1. Model framework.

A. Mapping Public Opinion Event into Sequence Like vectorized Natural Language

Definition 1. Comment word

A comment word is a vector constructed by the corresponding comment text based on the characteristics of it. The specific form is shown in Formula (1).

$$C = [c_1, c_2, \dots, c_7] \quad c_i \in \{0,1\} \quad i \in \{0,7\} \quad (1)$$

We use vector \$C\$ to describe comments and \$c_i\$ is the \$i\$th feature. The value of \$c_i\$ and design criteria will be discussed in Section III.1(a).

Definition 2. Event word vector

The event word vector \$Z_j\$ represents the characteristics of an event on a certain day. The specific construction method is shown in formula (2)-(5).

$$\bar{C} = [\bar{c}_1, \bar{c}_2, \dots, \bar{c}_7] \quad \bar{c}_i \in \{0,1\} (1 \leq i \leq 7) \quad (2)$$

$$Z_j = [z_1, z_2, \dots, z_7] \quad (3)$$

$$\bar{c}_i = \frac{\sum_{t=1}^{k_j} c_i^{(t)}}{k_j} \quad (4)$$

$$z_i = \begin{cases} 1 & \bar{c}_i > \omega \\ 0 & \bar{c}_i \leq \omega \end{cases} \quad (1 \leq i \leq 7, 0 \leq \omega \leq 1) \quad (5)$$

Where \$k_j\$ is the total number of comments collected on \$j\$th day of the event, \$\bar{c}_i (1 \leq i \leq 7)\$ is the average value of the \$i\$th feature of all comments collected on the \$j\$th day, \$\omega\$ is a pre-defined threshold.

Definition 3. Event word

An event word is a decimal value corresponding to the event word vector.

The mapping relationship between an event word vector and an event word is represented by the event word dictionary in Table II. The construction process of a single event word is shown in Fig. 2.

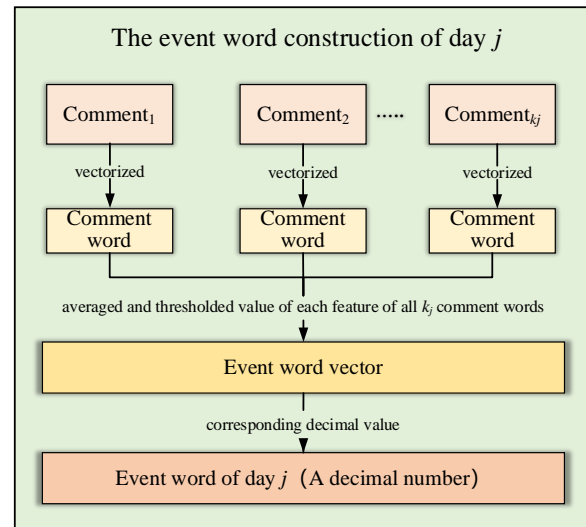


Fig. 2. Construction process of an event word for \$j\$th day.

Definition 4. Event sentence

Event sentence is a data form similar to vectorized natural language data, which is composed of event words according to the chronological order of events. The form of an event sentence is presented in Fig. 3.



Fig. 3. The form of an event sentence.

According to definitions 1 to 4, we can interpret all the comments on the \$j\$ day of an event as comment words one by one, construct the event word vector of that day with the average value of all the comment words of that day, map the event word vector to the event word of that day, and then arrange all the event words in order according to the time sequence of the event, so as to obtain the event sentence from the start of the event to the current time point.

Therefore, a public opinion event is represented as a description sentence (event sentence) with a certain meaning. The event words of each day in the event process constitute the vocabulary in the sentence, while the evolution time point of the public opinion event where the comment is located determines the grammatical order structure of the sentence. The event sentence obtained contains various specific information of the public opinion event from its start to the current time point including event evolution and change information.

1) *Characteristics of comment words and construction of event word dictionary*

a) *Design of features about comment words:* The comments are personal opinions that expressed by netizens on the situation occurred on the day of an event, usually in the form of short text or information combined with pictures and texts. In natural language, the number of words is limited, so each word can be uniquely encoded by constructing a

dictionary, and sentences can be constructed using the codes of words. This paper regards an event as a sentence and the comments as words in the sentence. However, because the different lengths of comment texts are too complicated, and each comment is a unique mind expressed by different people, it is difficult to use comments as words to construct an upper-limited "dictionary" in natural language.

The comments not only reflect the current progress of the event, but also includes the netizens attitudes. It is including the deep excavation, questioning and disclosure of the event. Additionally, in many cases the comments will affect the evolution direction of the event and even reverse public opinion. Based on the important influencing factors of public opinion reversal analyzed in [2][3], this paper designs seven discriminative features for comments, and transforms text comments into comment words at a relatively abstract level. The features and values of a comment word are shown in Table I.

TABLE I. FEATURE DESCRIPTION AND VALUE OF COMMENT WORD

Feature name	Feature description	Value
c_1	Whether the comment is mixed with some kind of social emotion	Yes 1, NO 0
c_2	Whether the comment's inclination is positive	Yes 1, NO 0
c_3	Whether the comment is in a declarative tone	Yes 1, NO 0
c_4	Whether the comment is cyberbullying	Yes 1, NO 0
c_6	Is the comment stereotypical	Yes 1, NO 0
c_7	Whether the comment has agenda setting	Yes 1, NO 0
c_8	Whether there are emoticons and special characters in the comment	Yes 1, NO 0

2) *Event word dictionary and event sentence:* According to the 7-dimensional comment word vector designed in Table I, after averaging and thresholding the comment words, the event word vector can be obtained. An event word dictionary

composed of event word vector and its corresponding decimal value is constructed, which is shown in Table II (all comment words are also derived from this dictionary).

TABLE II. EVENT WORD DICTIONARY

Features of event word vector Decimal value	Whether the comment is mixed with some kind of social emotion	Whether the comment's inclination is positive	Whether the comment is in a declarative tone	Whether the comment is cyberbullying	Is the comment stereotypical	Whether the comment has agenda setting	Whether there are emoticons and special characters in the comment
0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	1
.....
127	1	1	1	1	1	1	1

3) *A model of comment word generation based on semi-supervised multi-Label curriculum learning:* In order to automatically generate the vector representation of a comment

in the NL2ER-Transformer model, this paper designs a semi-supervised multi-label curriculum learning model, the structure is shown in Fig. 4.

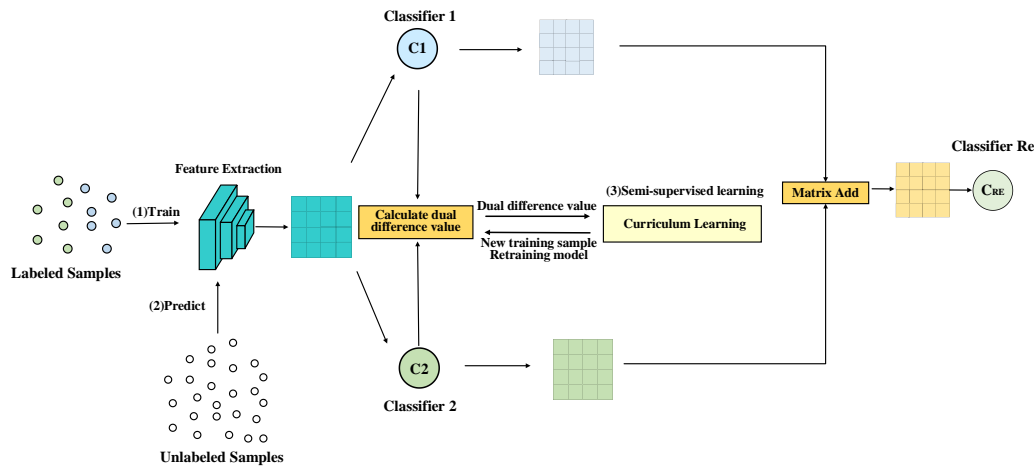


Fig. 4. Structure of the semi-supervised multi-label curriculum learning model.

The model contains four neural networks, one feature extractor $E(\cdot)$, two multi label classifiers $C_1(\cdot), C_2(\cdot)$, and one final classification $Cre(\cdot)$. Specifically, $E(\cdot)$ extracts features from samples; $C_1(\cdot)$ and $C_2(\cdot)$ are used to obtain the label prediction results(i.e., comment words); the label relation is

abstracted by adding $C_2(\cdot)$ and $C_1(\cdot)$, and finally classify the samples through $Cre(\cdot)$. During the curriculum learning, difference of $C_1(\cdot)$ and $C_2(\cdot)$ can be used to screen pseudo labeled samples added to training samples. The part of the semi-supervised curriculum learning is shown in Fig. 5.

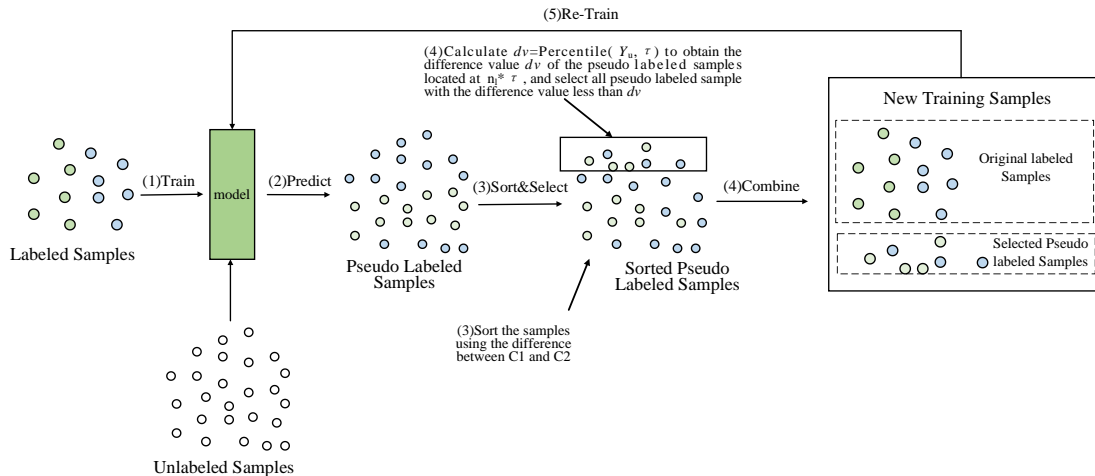


Fig. 5. Training method of the semi-supervised multi-label curriculum learning.

The training method of curriculum learning in Fig. 5 is explained as following, the details are in [26].

- Train the model with labeled comments; (2) Predict all unlabeled comments; (3) Rank pseudo labeled data based on the difference between $C_1(\cdot)$ and $C_2(\cdot)$; (4) Obtain the difference value $dv = \text{Percentile}(Y_u, \tau)$ of the samples ranking at $n_l * \tau$, select the pseudo labeled samples with the difference value less than dv and combine it with the original training set to form a new training set; (5) Re-train the model, and repeat (2) - (5) until all the pseudo labeled data have all the marks. n_l represents the number of labeled samples; $Y_u \in R^{n_l * d_l}$ represents the unlabeled feature matrix and the label matrix.

The adaptive threshold dv is obtained in accordance with the following method: In the training process, all the unlabeled data X_u into the model are put to predict and calculate the

difference df of each sample, and then rank it in ascending order according to the sample difference, after that we obtain sample difference value dv of location in $n_l * \tau$ and select the samples, which df is less than dv . Adding them to the labeled data set X_l , in which $\tau \in [0.2, 0.4, 0.6, 0.8, 1]$, that means after every cycle. It will select 20% or more pseudo labeled samples incorporated into the training set. The specific process of curriculum learning is shown in Fig. 5. The difference $df(x_i)$ is calculated as formula (6).

$$df = \| C_1(\cdot) - C_2(\cdot) \|_2^2 \quad (6)$$

B. Event Vectorization Model based on Transformer

Dynamic evolution is an important information of events. Learning the characteristics of the event evolution, which plays a significant role in predicting public opinion reversal from a developmental perspective. Inspired by training natural language, a training model NL2ER-Transformer is designed to

simulated event evolution based on the event sentence structure mentioned above. The model is based on Transformer structure which is a classic NLP model proposed by Google in 2017 [12]. Transformer uses a self-attention mechanism and does not use the sequential structure of RNN, so that it can be trained in parallel and can have global information. In the model, the event sentence from the start to the i th day of the event is used

as the training **feature**, and the event sentence from the start to the $(i+1)$ th day of event is used as the **label**. This process is similar to text translation: the sentences to be translated are as features, and the corresponding translation results are as labels.

The whole training process of the event based on NL2ER-Transformer is shown in Fig. 6.

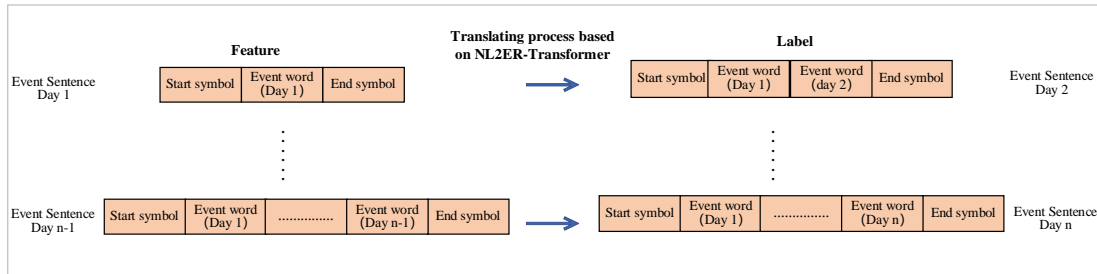


Fig. 6. Training method of simulating event evolution.

IV. EXPERIMENTS

This paper verifies the effectiveness of the designed model from three aspects: (1) For the proposed comment word generating model based on semi-supervised multi-label curriculum learning, two multi-label classification benchmark models, Fast Tag [27] and Semantic Auto Encoder (SAE) [28], were selected for comparative experiments. At the same time, in the test data used for the downstream reversal prediction task, 50% of the comment words were automatically generated by the model, further verifying the effectiveness of the model; (2) For the event evolution analysis capability of the proposed NL2ER-Transformer model, we manually represent the event sentence value of each day of the event according to the definition, and then compare it with the event sentence value predicted by the model; (3) For downstream reversal prediction tasks, select different event vectorization baseline models and multiple classification models to verify the effectiveness of event vectors generated by NL2ER-Transformer model.

A. Datasets

We obtained 55 events in 2020-2022 from public opinion monitoring platforms such as Qingbo Index (<https://www.gsdata.cn>) and Zhiwei Shijian (<https://ef.zhiweidata.com/library>), including 26 public opinion reversal events, 29 public opinion non-reversal events, and a

total of 38,091 comments on related events. And according to the assignment rules given above, seven discriminative features of 8041 comments were manually annotated. Data can be obtained from the following link: <https://pan.baidu.com/s/1mEQxLdRyOHZd4T8Rmxz0fg>. Extraction code: xl8m

In the data annotation, for reversal events, from start to the day of reversal, the label is 1 and from the day after reversal to the end of the event, the label is 0; for non-reversal events, from start to the end of the events, the label is 0.

In the training of the comment word generation model, we selected 30472 comments in total, including 10051 manually labeled comments and 20421 pseudo-labeled comments generated by the proposed model. In the construction of the downstream public opinion reversal prediction model, we randomly selected 12 events as the test set (including 6 reversal events and 6 non-reversal events). The 50% of comments used for the test set are pseudo-labeled data generated by the proposed semi-supervised multi-label curriculum learning model), and the remaining 43 events as the training set.

B. Verification of the Comment Word Generation Model based on Semi-Supervised Multi-Label Curriculum Learning

TABLE III. COMPARISON RESULTS

Method	Absolute matching rate	Hamming loss	Accuracy	Precision	Recall	F1 score
FastTag	0.31189	0.14582	0.67537	0.2681	0.26729	0.2282
SAE	0.11021	0.2346	0.73342	0.49564	0.19515	0.24462
Our model	0.24489	0.16578	0.79937	0.72863	0.34015	0.3908

From Table III, it can see that excepting the absolute matching rate is 7% lower than FastTag, our model is much more than the traditional multi-label classification model in all indicators, This proves the effectiveness of the proposed multi-label curriculum learning model.

C. Evaluation for the Evolution Analysis Ability of NL2ER-Transformer Model

During training, the parameters of NL2ER-Transformer model are set as shown in Table IV.

TABLE IV. TRAINING PARAMETERS

Num_layers	d_model	num_heads	input_vocab_size	target_vocab_size	dropout_rate
2	64	4	128	128	0.1

For evolution analysis ability test, we train the NL2ER-Transformer model using the cross-validation method. One event is selected each time as the validation set, and the other events are used as the training set, and the method of error calculation for each learning result is shown in formula:

$$e = \frac{\sum_{i=1}^n (w_i - \hat{w}_i)^2}{n} \quad (7)$$

In formula (7), n is the number of days when the event occurs, the event sentence corresponding to the event is $S = [w_1, w_2, \dots, w_n]$, and the predicted result event sentence is $\hat{S} = [\hat{w}_1, \hat{w}_2, \dots, \hat{w}_n]$.

According to the experimental results, the model produced 4 prediction errors in total which happened in the 1st, 17th, 33th, and 34th experiments and the error values were respectively 3.6, 12.6667, 60, and 22.4285. The experimental results further show that the model does have the ability to analyze the evolution process of public opinion events, and can predict the evolution trend and changes of events. The event sentences with errors and their corresponding events are shown in Table V.

TABLE V. EVENT SENTENCES WITH ERRORS IN PREDICTION

Event	Type of data	Number of days	0	1	2	3	4	5	6	7
12-year-old girl lied about being raped by her teacher	Real event sentence	1	start	84	end	0	0	0	0	0
		2	start	84	82	end	0	0	0	0
		3	start	84	82	80	end	0	0	0
		4	start	84	82	80	82	end	0	0
		5	start	84	82	80	82	80	end	0
	Prediction	1-6	start	84	82	80	82	80	82	end
The death of a female teacher in Jiangsu during the invigilation	Real event sentence	1	start	81	end	0	0	0	0	0
	Prediction	1-2	start	81	16	end	0	0	0	0
Heilongjiang Xuexiang's 15 yuan per sausage attracts heated discussions	Real event sentence	1	start	16	end	0	0	0	0	0
		2	start	16	16	end	0	0	0	0
		3	start	16	16	16	end	0	0	0
	Prediction	1-4	start	16	16	16	17	end	0	0
A 15-year-old girl in Hei Longjiang killed her mother and hid her corpse in a cold storage room	Real event sentence	1	start	16	end	0	0	0	0	0
		2	start	16	0	end	0	0	0	0
		3	start	16	0	80	end	0	0	0
		4	start	16	0	80	2	end	0	0
		5	start	16	0	80	2	2	end	0
	6	start	16	0	80	2	2	48	end	
Prediction	1-6	start	16	16	17	17	1	17	end	

Table V show that three of the four erroneous events have only one day's prediction error, and the event with more errors have errors just from the second day. According to the analysis, the main reason is that of all events, only the error event contains '0' event words. The model has never "seen" the '0' event words during the training process, so the model prediction fails. Although the evolution of the event sentence is misplaced, we found that the label(reversal/non-reversal) of the event was predicted correct. So, we made a further comparison, and it turns out that the evolution process presented by the incorrectly generated event sentence is very similar to the event of 'Female tambourine businessman in Dali, Yunnan scold tourists' and the labels of these two events are also the same, i.e., non-reversal. This error obviously has no impact on the prediction results of public opinion reversal.

D. Evaluation of Downstream Reversal Prediction Ability for The Generated Event Vectors

1) *Experiment setup:* We compare our experimental results with the results of three common models: RoBERTa[24], XLnet[25], Manual designing features[3]. We generate the vectors respectively by the above four models, the first three of which are trained with the blog texts of events, and our proposed NL2ER-Transformer Encoder. Then, through training four prediction models: Linear SVM(LSVM), Back Propagation Neural Network(BPN), nonlinear SVM and KNN, we further compare the prediction effects of the vectors generated by these four methods.

The neural network includes four hidden layers, the number of neurons in each layer is 16, 32, 64, and 128, the activation function is Relu, the output layer activation function is Sigmoid, the loss function is binary cross entropy function, the

optimizer is Adam, and the training times are 100. In the linear SVM, C is set to 1, the loss function is hinge, and the training times are 100. The nonlinear SVM kernel function is set as sigmoid function, C is set as 1, and the number of iterations is 100. KNN algorithm sets the number of reference neighbors to 5.

2) *Experimental results analysis:* We use four indicators to evaluate the classification models: accuracy, precision, recall, F1-Score. We believe that as long as the model can give a reversal probability greater than 0.5 before the event reversal, the prediction is considered correct. From Table VI, we can see that the classification trained by our model is better than other vectorization methods in terms of ensuring high accuracy and precision. KNN and Nonlinear SVM trained in this way are not very good in the four indicators.

Public opinion reversal is a kind of phenomenon in the process of event evolution. Therefore, whether the vector representation of events includes the development and changes of public opinion determines whether the prediction model can learn the important knowledge of event reversal. Compared with netizens' comments, the factual description of events is relatively lacking in expressing the dynamics of public opinion evolution. So, the event vectors generated based on this factual description of events will affect the result of reversal prediction.

The presented classification models of reversal prediction trained with vectors generated by our proposed NL2ER-Transformer are better than or equal to those trained with vectors generated by other event vectorization methods. Therefore, we can infer that it is feasible to vectorize public opinion events by combining comments and event evolution.

TABLE VI. INDEX VALUE OF PREDICTION MODEL UNDER DIFFERENT EVENT VECTOR TRAINING

Representation	Method	Accuracy	Precision	Recall	F1-Score
RoBERTa	LSVM	66.66%	66.66%	66.66%	0.666
	BPN	66.66%	83.33%	62.5%	0.714
	Nonlinear SVM	50%	66.66%	50%	0.514
	KNN	41.66%	66.66%	44.44%	0.533
XLnet	LSVM	75%	100%	66.66%	0.8
	BPN	50%	83.33%	50%	0.625
	Nonlinear SVM	50%	100%	50%	0.666
	KNN	46.15%	85.71%	50%	0.5
Manual design features	LSVM	58.33%	16.66%	100%	0.2857
	BPN	58.33%	16.66%	100%	0.2857
	Nonlinear SVM	58.33%	16.66%	100%	0.2857
	KNN	58.33%	16.66%	100%	0.2857
NL2ER-Transformer	LSVM	91.66%	83.33%	100%	0.909
	BPN	91.66%	100%	85.71%	0.923
	Nonlinear SVM	83.33%	66.66%	100%	0.8
	KNN	91.66%	83.33%	100%	0.909

V. CONCLUSION AND FUTURE WORK

Aiming at the problems existing in the current research on event representation, this paper proposes an abstract mapping of events to sequences, and combines subjective comments and event temporality to transform public opinion problem into natural language processing problem. The NL2ER-Transformer model is constructed to realize the extraction of event features based on comments. The experimental results show that, compared with current mainstream event vectorization models, the event vectorization scheme proposed in this paper has a better effect on public opinion reversal prediction, and the training method proposed in this paper can indeed enable the vector generation model to learn events evolution information.

There are still some limitations in this paper, and the future work will be carried out from the following aspects:

- Since the discriminative features of comment words are designed based on the prediction of public opinion reversal, the pre-training of the model is only for the prediction task of public opinion reversal, and the effectiveness of other public opinion prediction tasks has not been thoroughly discussed. Next, the team will design a more general feature representation and generation strategy, and construct an event pre-training model in the field of public opinion prediction by testing different types of downstream tasks based on event vectors.
- Event sentence is only generated based on comments, and the features of comment words are discriminative features with strong subjectivity, and the fusion of other objective data is not considered. In future research, we will try to use new methods to combine other unstructured data, such as event description text, with relevant structured data in the process of public opinion evolution, to achieve event pre-training with multi-source data fusion, and to improve the generalization performance of the model.

ACKNOWLEDGMENT

This work was supported by the National Social Science Foundation (No. 22BTQ048).

REFERENCES

- [1] Chang-bin Jiang, Yue-qi Zou, Hu Wang, et al. Research on public opinion reversal prediction of we-media based on SVM. *Information Science*, 2021, 39(04): 47-53, 61
- [2] Nan Wang, Hai-rong Li, Shu-ru Tan. Research on public opinion reversal prediction based on improved SMOTE algorithm and integrated learning. *Data Analysis and Knowledge Discovery*, 2021, 5(04): 37-48
- [3] Nan Wang, Hai-rong Li, Shu-ru Tan. Prediction of public opinion reversal based on evolution analysis of public opinion events and Improved KE-SMOTE algorithm, data analysis and knowledge discovery, 2022, 3(13): 1-20
- [4] Yang, F., Liu, Y., Yu, X., & Yang, M. (2012). Automatic detection of rumor on Sina Weibo. *MDS '12*.
- [5] Jing Ma, Wei Gao, Zhongyu Wei, Yueming Lu, and Kam-Fai Wong. 2015. Detect Rumors Using Time Series of Social Context Information on Microblogging Websites. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management (CIKM '15)*. Association for Computing Machinery, New York, NY, USA, 1751–1754. <https://doi.org/10.1145/2806416.2806607>
- [6] Ting-rui Yan, Dong Wang, Li-zhong Xiao, Guang-zhong Wang. Simulation of public opinion trend prediction model based on AR-RBF. *Computer Simulation*, 2021, 38(04): 437-441
- [7] Jian-feng Zhou. Research on popular microblog feature selection and prediction method based on FA-SVM. *Computer Applications and Software*, 2018, 35(12): 107-111
- [8] Mikolov T , Chen K , Corrado G , et al. Efficient estimation of word representations in vector space. *Computer Science*, 2013.
- [9] Pennington, J., Socher, R., & Manning, C. D. (2014). Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing* (pp. 1532–1543)
- [10] Posadas-Durán, J.-P., Gomez-Adorno, H., Sidorov, G., & Escobar, J. J. M. (2019). Detection of fake news in a new corpus for the Spanish language. *Journal of Intelligent & Fuzzy Systems*, 36(5), 4869–4876.
- [11] Peters M, Neumann M, Iyyer M, et al. Deep contextualized word representations: //Proceedings of NAACL, New Orleans American 2018. 2227–2237
- [12] Vaswani A, Shazeer N, Parmar N, et al. Attention Is All You Need. *arXiv*, 2017.
- [13] Yang Xu, Liang-qi Zhu, Bo Huang, et al. Public opinion analysis based on EEMD-Transformer model: COVID-19 public opinion as an example. *Journal of Wuhan University of Science and Technology*, 2020, 66(05): 418-424
- [14] Khoo, L. M. S., Chieu, H. L., Qian, Z., & Jiang, J. (2020, April). Interpretable rumor detection in microblogs by attending to user interactions. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 34, No. 05, pp. 8783-8790).
- [15] Samadi, M., Mousavian, M., & Momtazi, S. (2021). Deep contextualized text representation and learning for fake news detection. *Information Processing & Management*, 58(6), 102723.
- [16] Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., et al. (2019). RoBERTa: A robustly optimized BERT pretraining approach. *CoRR*, abs/1907.11692.
- [17] Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R. R., & Le, Q. V. (2019). XLNet: Generalized autoregressive pretraining for language understanding. In *Advances in neural information processing systems* 32 (pp. 5754–5764). Curran Associates, Inc.
- [18] Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 conference of the North American Chapter of the Association for Computational Linguistics: Human language technologies, Volume 1 (Long and short papers)* (pp. 4171–4186). Minneapolis, Minnesota: Association for Computational Linguistics.
- [19] Lin, S. Y., Kung, Y. C., & Leu, F. Y. (2022). Predictive intelligence in harmful news identification by BERT-based ensemble learning model with text sentiment analysis. *Information Processing & Management*, 59(2), 102872.
- [20] Blanco, G., & Lourenço, A. (2022). Optimism and pessimism analysis using deep learning on COVID-19 related twitter conversations. *Information Processing & Management*, 59(3), 102918.
- [21] Zheng, M., Ma, Y., Zheng, A., & Wang, N. (2018, May). Constructing method of public opinion knowledge graph with online news comments. In *2018 International Conference on Robots & Intelligent System (ICRIS)* (pp. 404-408). IEEE.
- [22] Chen Xu (2022). Exploration of network public opinion management method based on knowledge map *Industry and Technology Forum* (02), 263-264
- [23] Anning & Anlu (2022). Cross platform network public opinion knowledge map construction and comparative analysis *Information Science* (03), 159-165 doi:10.13833/j.issn. 1007-7634.2022.03.020.
- [24] Cui, Y., Che, W., Liu, T., Qin, B., & Yang, Z. (2021). Pre-training with whole word masking for chinese bert. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 29, 3504-3514.
- [25] Cui, Y., Che, W., Liu, T., Qin, B., Wang, S., & Hu, G. (2020). Revisiting pre-trained models for Chinese natural language processing. *arXiv preprint arXiv:2004.13922*.

- [26] L. Wang, Y. Liu, C. Qin, G. Sun, and Y. Fu, "Dual relation semi-supervised multi-label learning," in *AAAI*, 2020, pp. 6227–623
- [27] M. Chen, A. Zheng, and K. Weinberger, "Fast image tagging," in *Proc. Int. Conf. Mach. Learn.*, 2013, pp. 1274–1282
- [28] E. Kodirov, T. Xiang, and S. Gong, "Semantic autoencoder for zero-shot learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 3174–3183.

A Novel Network Intrusion Detection System Based on Semi-Supervised Approach for IoT

Durga Bhavani A¹, Dr. Neha Mangla²

Assistant Professor, Department of Computer Science and Engineering, BMS Institute of Technology and Management,
Bangalore, India¹

Associate Professor, Department of Information Science and Engineering, Atria Institute of Technology,
Bangalore, India²

Abstract—An intrusion detection system (IDS) is one of the most effective ways to secure a network and prevent unauthorized access and security attacks. But due to the lack of adequately labeled network traffic data, researchers have proposed several feature representations models over the past three years. However, these models do not account for feature generalization errors when learning semantic similarity from the data distribution and may degrade the performance of the predictive IDS model. In order to improve the capabilities of IDS in the era of Big Data, there is a constant need to extract the most important features from large-scale and balanced network traffic data. This paper proposes a semi-supervised IDS model that leverages the power of untrained autoencoders to learn latent feature representations from a distribution of input data samples. Further, distance function-based clustering is used to find more compact code vectors to capture the semantic similarity between learned feature sets to minimize reconstruction loss. The proposed scheme provides an optimal feature vector and reduces the dimensionality of features, reducing memory requirements significantly. Multiple test cases on the IoT dataset MQTTIOT2020 are conducted to demonstrate the potential of the proposed model. Supervised machine learning classifiers are implemented using a proposed feature representation mechanism and are compared with shallow classifiers. Finally, the comparative evaluation confirms the efficacy of the proposed model with low false positive rates, indicating that the proposed feature representation scheme positively impacts IDS performance.

Keywords—IoT; security; intrusion detection system; semi-supervised; autoencoder; clustering; machine learning

I. INTRODUCTION

The rise of IoT has increased the number of connected devices, resulting in a growing need for effective security measures to prevent data breaches and cyber-attacks [1]. The MQTT (Message Queuing Telemetry Transport) protocol is widely adopted in IoT-enabled real-world applications such as smart homes, smart cities, smart industries, and smart healthcare due to its low bandwidth cost, low memory requirements, and minimal packet loss [2]. The MQTT protocol is based on a publish/subscribe mechanism and a central communication module, often called a broker, which is vulnerable to various security threats and attacks. In [3], researchers reported security risks related to the MQTT protocol and found over 53000 publicly accessible MQTT-based IoT devices.

In the existing literature, various security and prevention mechanisms based on cryptography and digital signature have been presented to deal with various attacks [4–5]. Cryptography is an effective solution for providing better security services by enabling source authentication mechanisms, but it is not robust against usability or availability attacks [6]. With the advent of machine learning (ML) techniques, more attention has been paid to studying intrusion and anomaly detection systems for IoT. Intrusion detection is an important aspect of IoT security which involves identifying abnormal behavior patterns and potential security threats in real-time. IDS and cryptographic-based approaches complement each other to ensure maximum security features such as confidentiality, integrity, authenticity, and availability [7–8]. Network IDS can be classified into signature-based and anomaly-based methods [9]. Signature-based IDS considers attack traffic patterns, while the anomaly-based approach usually considers normal traffic patterns. However, signature-based IDS only relies on known traffic patterns and faces limitations in detecting new sophisticated attacks whereas anomaly-based IDSs usually have a high false positive rate, which may require additional resources and time to eliminate the large number of alerts generated. Furthermore, existing IDSs are not well suited for MQTT-enabled IoT ecosystems, as they were designed with little attention to resource efficiency and scalability features [10].

Recently, many researchers have attempted to explore the ability of feature representations to handle large-scale traffic sampling and class imbalance in attack detection based on predictive modeling. In [11], an autoencoder-based evaluation framework is given to assess the effectiveness of various feature representation techniques for handling large-scale information streaming in text classification tasks. Similarly, researchers in [12] presented feature selection techniques based on unsupervised learning schemes to enhance data analytics tasks and computational intelligence performance on large datasets. Researchers in [13, 14, 15] reported enhancing the performance of real-time analytics operations on big data by using appropriate feature representation techniques. Furthermore, the adoption of autoencoders for feature representation is seen in many applications, such as the classification of anti-drug response [16] and diagnosis of autism spectrum disorder [17]. In [18], the authors applied the reported use of deep learning to feature learning to deal with big biological data for disease investigation.

Taking inspiration from the literature in which many researchers employ autoencoders for feature modeling, we are motivated to propose a new feature representation model for learning-driven IDS that can keep up with high network traffic and exhibit high accuracy in IoT intrusion detection.

This manuscript proposes a novel anomaly and intrusion detection solution in MQTT-driven IoT systems that combine unsupervised and supervised learning techniques. The proposed solution uses an autoencoder to extract robust features from normal traffic data, which are then clustered and assigned different cluster IDs based on their unique features. This information creates a supervised classifier to detect real-time anomalies and intrusions. The proposed solution offers several advantages over traditional security measures, including detecting unknown threats and adapting to changing threat landscapes. The scope of this research article is to provide an in-depth analysis of the proposed solution, including mathematical models and algorithms, as well as a discussion of the potential benefits and scope of real-world applications.

The remaining sections of this paper are organized as follows: Section II presents related work; Section III highlights the research problem explored and addressed in the proposed work; Section IV presents the system design and dataset details and preprocessing; the proposed feature representation mechanism is discussed in Section V, along with experimental analysis and result, discussion are carried to justify the scope of the proposed system, and finally, the entire work is concluded in Section VI.

II. RELATED WORK

This section presents a brief review of existing literature on IoT security that uses machine learning concepts in IDS design and approaches adopted for feature enhancement towards better classification outcomes.

The work carried out by Kaur et al. [19] employed a Convolution Neural Network (CNN) to recognize and describe various security attacks. The authors have validated their attack detection model through CICIDS2017 and CICIDS2018 datasets. However, this study offers a multiclass classification feature that does not meet higher accuracy for detecting many attacks. Another study by Odetola et al. [20] also adopted the CNN model to design a multiclass classification system for the IoT network. The authors have utilized a single CNN with multiple layers and a customizable loss function, which achieves less delay in the attack detection process. The authors in the study of Haripriya and Kulothungan [21] designed an IDS based on the approach of a fuzzy logic system for IoT devices against DoS attacks. The fuzzy logic has been widely used in many applications to deal with ambiguity in the dataset. However, using the fuzzy logic approach in this study for IDS design does not show its effectiveness due to its high complexity with an increase in the input dimension. In the IoT ecosystem, massive data is forwarded continuously. Also, this study is quite specific for the particular type of attacks, and more advanced or unknown attacks have been left untouched. The work of Ciklabakkal et al. [22] utilizes multiple machine learning techniques along with the autoencoder to detect intrusions in the IoT network. However, the authors have not

explicitly discussed the type of attacks they are considering. Faker and Dodge [23] employed a deep learning approach in IDS design and evaluated its performance on the two datasets, namely CIC-IDS2017 and UNSW-NB15. Alkadi et al. [24] used blockchain technology to ensure security and reliability in distributed IoT-IDS. In this study, Ethereum blockchain-based smart contracts are implemented to ensure privacy features, and Bi-Long Short-Term Memory (LSTM) is used to develop IDS that deal with sequential network flow. The effectiveness of the presented model is evaluated based on the UNSW-NB15 and BoT-IoT datasets. AI-enabled IDS is designed in the study of Fatani et al. [25]. The authors have addressed feature engineering issues using CNN and extracted relevant attributes in this study. Further, a transient search optimization mechanism is used to select the feature.

A game theory-based strategy is also adopted in the design of IDS. In the study of Wang et al. [26], the authors suggested an analytical framework to identify the circumstances under which malevolent devices have no motivations to perform the attack in the IDS attack-prevention game. The work of Basset et al. [27] presented a semi-supervised approach for IDS that considers sequential features of the traffic flow in the training phase. The authors have also focused on enhancing the learning performance of the system by tuning the system configuration during the generalization of spatiotemporal representations from the input dataset. A modified version of the attention mechanism helps the system focus on the significant attributes during the training process. Dutta et al. [28] suggested an anomaly detection scheme based on the sparse autoencoder, and an LSTM with logistic regression is further employed to detect malicious traffic. In Aygun et al. [29] study, a comparative analysis is carried out between two autoencoders for recognizing new or unknown attacks. Shone et al. [30] introduce an un-symmetric autoencoder system that offers better data generalization and dimensionality reduction. The work carried out by Tabassum et al. [31] suggested privacy-aware IDS. The preprocessing operations using autoencoder eliminate redundancies and present precise dataset modeling. The experimental analysis has shown the benefit of using an autoencoder in reducing the feature dimension and the computational complexity in the training phase. There are other recent works on IDS for detecting MQTT attacks in IoT using ML approaches highlighted in Table I.

TABLE I. HIGHLIGHTS ON EXISTING IDS FOR MQTT ATTACKS

Authors	Approach	Dataset	Metric
[32]	RF, NB, NDT, MLP	MQTTset	Accuracy and F1 Score
[33]	Tree Based Approach	Simulation	Detection rate
[34]	NB, LR, KNN,	MQTT-IoT-IDS2020	Accuracy, Recall, F1-score, Precision
[35]	XGBoost, LSTM, GRU	MQTT-IoT-IDS2020	F-beta score, Accuracy, Log loss
[36]	DNN	MQTT-IoT-IDS2020	Accuracy, Precision, Recall, F1-score
[37]	Transfer learning	MQTT-IoT-IDS2020 and IoT-23	Accuracy, Precision, Recall, F1-score

The existing literature has widely adopted machine learning techniques in IoT intrusion detection. However, previous solutions have some shortcomings, which should be handled efficiently. The findings reveal that most of the existing IDS schemes are specific to a particular adversarial scenario and may not be adaptive to detect unknown attacks and multi-class intrusions. It has also been noticed that majority of the recent works on IDS lack focus on handling class imbalance factor as most network dataset is imbalanced. Such schemes may be subjected to a higher false-positive rate in the intrusion classification task. There is also not much emphasis given by researchers on feature enhancement, and optimization in computational resources especially when the dataset is quite massive, like MQTT-IoT-IDS2020 (3.6 GB) and IoT-23 dataset (i.e., 21 GB). Therefore, the scope of such existing systems is limited and cannot be applied to large and dynamic networking scenarios.

The problem statement for the proposed work can be stated as *"It is challenging to design a feature representation mechanism that can offer a higher degree of optimization in the training samples, allowing the learning model to discover the precise pattern required for the detection of attacks with minimal false-positive and high accuracy rate."*

III. PROPOSED SOLUTION

The proposed approach is based on a combination of autoencoder-based anomaly detection and supervised classifier-based multiclass intrusion detection. The proposed approach consists of the following steps:

- Training the autoencoder with normal traffic data to learn a compressed representation of the data.
- Using the Gini measure to compute a threshold for assigning different cluster IDs to different classes of intrusions.
- Using the autoencoder to map the network traffic into the learned compressed representation.
- Introducing the Gini-based threshold with an ML classifier to perform multiclass intrusion detection.

The autoencoder is trained with normal traffic data to learn a compressed representation of the data, which can be used to detect deviations from normal behavior. The Gini measure is then used to compute a threshold for assigning different cluster IDs to different classes of intrusions. The autoencoder is used to map the network traffic into the learned compressed representation, and the Gini-based threshold is introduced with an ML classifier to perform multiclass intrusion detection. The first step is to train an autoencoder using normal MQTT message data. This autoencoder will learn the patterns of normal behavior and create a compressed representation of these patterns that can be used as a feature set for the intrusion detection system. Next, a supervised classifier is trained using a multiclass classification approach. The classifier is trained using labeled MQTT messages that correspond to different types of intrusions. The classification is performed on the compressed feature set produced by the autoencoder. To assign different cluster IDs to different classes of intrusions, a

threshold is computed using a Gini measure. The threshold is used to assign different cluster IDs to different classes of intrusions. This will enable the classifier to accurately identify and classify different types of intrusions. Once the classifier is trained, it can be used to monitor MQTT traffic in real-time. Any anomalous activity or intrusion that does not match the patterns of normal behavior will be detected by the autoencoder and flagged as a potential threat. The feature set produced by the autoencoder is then used as input to the classifier, which can classify the intrusion and take appropriate action. By combining an autoencoder-based anomaly detection system with a supervised classifier for multiclass intrusion detection, this solution can effectively detect and classify security threats in MQTT-driven IoT networks. This approach provides early detection of security threats, reduces false positives, and improves accuracy, making it a powerful tool for protecting IoT networks against attacks. In this work, the study deals with security on all three layers: the network layer, the transport layer (TCP and UDP analysis), and the application layer. The proposed model is based on the semi-supervised approach, where a combination of an autoencoder and supervised learning techniques is used to perform feature modeling and intrusion detection.

IV. MATHEMATICAL MODEL AND ALGORITHM

This section discusses the implementation procedure adopted in designing feature learning-based IDS.

The mathematical model includes training the autoencoder to learn the compressed representation of normal behavior, training a supervised classifier using the compressed feature set as input, computing a threshold value using a Gini measure, and using the classifier and threshold to detect and classify intrusions in real-time. Let $X = \{x_1, x_2, \dots, x_n\}$ be the input dataset consisting of n-dimensional feature vectors.

- Autoencoder: Let $AE(X) = \{y_1, y_2, \dots, y_n\}$ be the compressed feature set obtained by passing X through the autoencoder.
- Computing Distance: Let $d(x, y)$ be the distance function that measures the distance between the input vector x and its corresponding output vector y obtained from the autoencoder. Let $D = \{d(x_1, y_1), d(x_2, y_2), \dots, d(x_n, y_n)\}$ be the set of distances between the input and output vectors.
- Computing Thresholds: Let $T = \{T_1, T_2, T_3, T_4\}$ be the set of threshold values that divide the distance values into five clusters, where the first four clusters represent the different categories of attacks and the last cluster represents normal behavior. The threshold values are computed using the Gini index, which finds the optimal split of the distances into the five clusters.
- Cluster ID Assignment: Let $C_{id} = \{c_1, c_2, \dots, c_n\}$ be the set of cluster ID assignments for the input vectors, where c_i belongs to $\{0, 1, 2, 3, 4\}$. The cluster ID assignments are determined by comparing the distance values with the threshold values and assigning the corresponding cluster ID.

- **Training Supervised Classifier:** Let us consider $F = \{f1, f2, \dots, fm\}$ be the set of training features obtained by taking the mean of the compressed feature set $AE(X)$ for each cluster ID. Let $Y = \{y1, y2, \dots, yn\}$ be the set of class labels for the input vectors. Let clf be the supervised classifier, which takes F as input and learns to classify the input vectors based on their features. The classifier is trained using a labeled dataset that includes the input vectors X and their corresponding class labels Y . Once the classifier is trained, it can be used to classify new input vectors based on their features.

A. Training Autoencoder

Let X be a matrix of MQTT message data, such that $X \in \{X_1, X_2, X_3 \dots X_n\}$ where n denotes number of data samples. Here, each row represents a single message, and each column represents a feature. The autoencoder is trained to learn the compressed representation of normal behavior, such that $X \approx X'$, where X' is the reconstructed matrix of X obtained by encoder module expressed as follows:

$$Z = f(X, \theta) \quad (1)$$

Where, Z is encoding function produces compressed feature set X' of input X and θ represents the learnable parameters of the autoencoder given as follows:

$$\theta = W_i + b \quad (2)$$

Algorithm 1: Training Autoencoder

Input: X (Training Dataset), E (Training Epoch)

Output: X' & Re (reconstruction error)

Start

1. Init W (weights)
2. *for each* E *do*
3. *for each mini batch* *do*
4. Encode $X_i \rightarrow Z_i : f(X_i, \theta)$ // feature representation
5. Decode $Z_i \rightarrow X'_i : f(Z_i, \theta)$ // input reconstruction
6. Tune hyperparameter using GST (grid search technique)
7. *end for*
8. *end for*
9. Compute $Re \leftarrow f_1(X_i, X'_i)$
// Re is the mean reconstruction error on training data

End

The autoencoder is trained with normal traffic logs (X_i) in an unsupervised manner, i.e., without labels, and creates a compact representation of the input samples, which is a reconstructed version (X'_i) of the normal traffic log pattern as closely as possible. However, if the given input data actually belongs to the attack category, the reconstructed data can be different from the original input data. This means that the category of the network traffic data is determined according to the difference between the input and its output. It is considered an attack if the difference is greater than the set threshold or cut-off value. In this regard, the study computes a mean reconstruction error (Re) using the function $f_1()$ with an input argument X_i and X'_i . Basically, this function computes the

square of the minimum error between the input data and the reconstructed data given as follows:

$$Re = \frac{1}{n} \sum_{i=1}^n (x_i - x'_i)^2 \quad (3)$$

Where, n denotes the total number of data samples in the training samples, and the computed Re is obtained is 0.001.

B. Threshold Computation

To assign different cluster IDs to different classes of intrusions, a threshold is computed using a Gini measure. The Gini measure is used to identify the feature with the highest discriminatory power between classes. This feature is used to compute the threshold value, which is used to assign different cluster IDs to different classes of intrusions (refer Table II). For example, Let us consider d be the number of features in the compressed feature set Z . The threshold is computed using the following formula:

$$T = (max(Z[:, i]) - min(Z[:, i]))/2 + min(Z[:, i]) \quad (4)$$

$$i = argmax(Gini(Z[:, j])) \quad (5)$$

$$j = 1, \dots, d \quad (6)$$

where $Z[:, i]$ represents the i -th feature of the compressed feature set Z , and $Gini(Z[:, j])$ represents the Gini measure for the j -th feature of Z .

In the proposed modeling the Gini index is a measure of the impurity or diversity of a set of values. In the context of intrusion detection, the Gini index can be used to determine the threshold value for the anomaly score function g . Let D be the set of all instances in the dataset X . Let C be the set of possible cluster labels, and let D_c be the subset of instances in D that belong to cluster c . The Gini index for a given threshold value T is defined as:

$$G(T) = \sum \{c \in C\} p_c(T) (1 - p_c(T)) \quad (7)$$

where $p_c(T)$ is the fraction of instances in D_c that have an anomaly score greater than T , such that:

$$p_c(T) = |\{x_i \in D_c \mid g(x_i) > T\}| / |D_c| \quad (8)$$

Algorithm 3: GINI Measure Based Threshold setting (T_i)

Input: D (Distance values), X' (Sample outputs)

Output: T_i {T1, T2, T3, T4}

Start

1. Load sample outputs 10% of X'
2. Get unique values: $U = f_2(X')$
3. *for each* i *in* U *do*
4. Initialize gini as G
5. $G \rightarrow []$ // empty vector
6. *for* d *in* range $\min(D)$ *to* $\max(D)$ *do*
7. $K = Y$, where $D < d$
8. $\lambda = 1 - \sum (P(K == I))^2$
9. Append λ to G
10. *end for*
11. $T_i = d$ where G is maximum in gini
12. *end for*

End

The algorithm 3 consists an input variable D (Distance values), and X (Sample outputs). After execution it returns an output $T_i \in \{T1, T2, T3, T4\}$. The first step of the algorithm loads a random subset of the sample outputs, X' . Afterwards, it gets the unique output values from X' : $U = \{u_1, u_2, \dots, u_n\}$. For each u_i in U , the algorithm performs initialization of a vector subjected to Gini indices, $G = []$. Then, for each distance value d in the range $[min(D), max(D)]$, the algorithm computes set of K outputs from X' that have a distance less than d from u_i . Next it computes the probability $P(K == j)$ that a sample output in K belongs to each class j in the set of possible classes. Afterwards the computation of the Gini index λ is carried out as $1 - \sum\{j = 1\}^m (P(K == j))^2$, where m is the number of possible classes. Finally, the computed λ is appended to vector G . Based on the previous outcome the algorithm set T_i to d , where d is the value of D that maximizes the Gini index in G . After successful execution of the all steps, the algorithm returns an output of the threshold values T1, T2, T3, and T4 indicating distinct values for assigning clusters for each specific intrusion classes. The Gini index is used as a measure of how well a particular distance value separates the normal class from the different attack classes. The threshold values T1, T2, T3, and T4 are used to divide the range of possible distances into five intervals, which are then mapped to the five output classes (Normal, DoS, Probe, R2L, and U2R) for use in the supervised classifier.

C. Cluster ID Assignment

The assignment of the cluster ID is done based on the threshold value T_i obtained from the algorithm 2 and distance vector D computed using Eq. (9).

Algorithm 3: Cluster ID Assignment

Input: AuE (Trained Autoencoder), X_{Ts} (Testing Dataset),

Output: C_id (cluster id)

Start

1. Call Algorithm-1
2. Phase: Model Testing
3. Test model with entire text data $\rightarrow X_{Ts}$
4. Compute D using Eq. (5)
5. $//D \leftarrow$ distance between original input and output
6. Phase: Clustering
7. for each data class do
8. Compute T_i $//$ where $i = \{1, 2, 3, 4\}$
9. Call Algorithm-2
10. end for
11. Phase: C_id assignment
12. Check:
13. *if* $D < T1$ do
14. C_id $\rightarrow 0$ $//$ Nor
15. *elif* $D < T2$ do
16. C_id $\rightarrow 1$ $//$ MBA
17. *elif* $D < T3$ do
18. C_id $\rightarrow 2$ $//$ SCA
19. *elif* $D < T4$
20. C_id $\rightarrow 3$ $//$ SPA
21. *else*
22. C_id $\rightarrow 4$ $//$ SUA

End

Algorithm 3 comprises three phases for implementing the proposed feature representation scheme. i) distance computation, ii) threshold computation, and iii) cluster id assignment. The primary step algorithm employs the computing operation and functions discussed in the Algorithm 1, which is about training autoencoder model. Further, the algorithm computes the distance vector using a distance function that uses the Euclidian distance formula expressed as follows:

$$\vec{D} = \sqrt{\sum_{i=1}^n |x_i - x'_i|^2} \quad (9)$$

Where vector D consists value of the distance computed between input data (x_i) and reconstructed data (x'_i), where $x_i \in X_{Ts}$ and $x'_i \in X'_{Ts}$.

The subsequent steps of the algorithm are subjected to clustering operations. In this phase, the algorithm uses threshold vector (T_i) obtained from the Algorithm 2. This vector consists of different threshold values specific to each data class. Next, the threshold values contained in T_i are compared with distance value based on which cluster-id (C_id) is assigned to groups containing unique features specific to normal class attack classes (algorithmic steps: 14 to 22).

TABLE II. GINI VALUES AND THRESHOLD OF EVERY CLASS

Label	Attack Labels/Class	Gini	Threshold
0	Norm	0.9214	0.0116
1	MBA	0.9142	0.0592
2	SCA	0.9321	1.0753
3	SUA	0.9154	1.2643
4	SPA	0.9431	1.6823

D. Intrusion Detection

The proposed research work's primary aim is to evolve a novel intrusion detection system (IDS) to detect and forecast cyber-attacks with high detection accuracy and to recover the IoT application from the attack at the earliest by activating a suitable response. The intrusion detection problem in the proposed work is considered to be a multiclass classification task using a supervised ML classifier. The proposed study implements three different shallow ML classifiers i) Naïve Bayes (NB), ii) K-Nearest Neighbor (KNN), and iii) Random Forest (RF). These classifiers take cluster id as input and return observed data as normal and attack classes. In order to train the classifier, 50% of the dataset is considered, and 50% dataset is considered for model evaluation. The reason behind this is that the dataset set is quite huge and big consists of approximately more than 30 million of traffic samples; taking more than 70% of dataset in processing often encounter computing resource exhaustion error. To address this kind of problem only the proposed research work has introduced an efficient feature representation technique which does not need to be introduced with 100% of the dataset, taking few samples it can provided sufficient feature code for training machine learning classifiers. The uniqueness of the proposed work lies in the blend of 3-

layer autoencoders that enhance feature generalization capability of any learning model independent of large traffic samples and without imposing memory overhead problem.

V. IMPLEMENTATION AND RESULT

The design and development of the proposed system is carried out using python programming language scripted in Jupyter notebook on an anaconda environment.

A. System Design and Implementation

The proposed model is based on the semi-supervised approach, where a combination of an autoencoder and supervised learning techniques is used to perform feature modeling and intrusion detection. The first module of the proposed system focuses on data preprocessing, which is meant to handle null and missing values and data encoding. The second module of the proposed system emphasizes feature modeling and representation using autoencoder and distance function. The third module of the proposed system is about implementing machine learning classifiers for multiclass intrusion detection and classification. Fig. 1 illustrates the architectural design of the proposed system followed with methodology discussed in the above section. In this work, the study deals with security on all three layers: the network layer, the transport layer (TCP and UDP analysis), and the application layer.

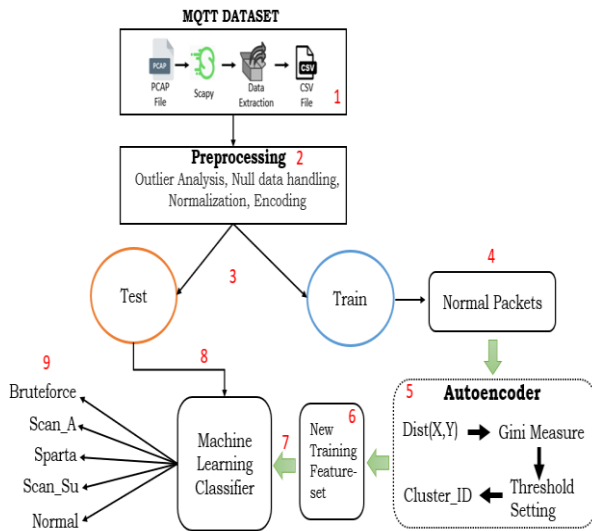


Fig. 1. Illustration of methodology adopted in system design.

B. Dataset Used

The MQTTIoT-IDS-2020 dataset is adopted in this research work downloaded from the IEEE data port using the link [38]. The attack simulation setup mimics realistic 12 MQTT sensors in a network, consisting of a central

communication module (broker), a camera feed server, and an attacker. This network setup consists of both generic networking scanning and MQTT brute-force attack. The structure of an MQTT packet looks like as shown in Fig. 2.

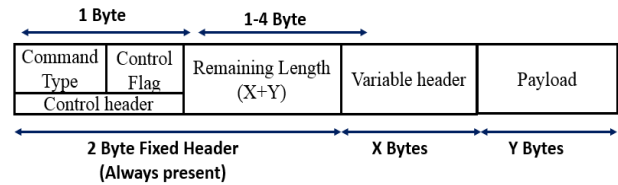


Fig. 2. Visualization of MQTT packet structure.

The dataset is downloaded in PCAP data format created using a software called Wireshark. There is an excellent utility in python called Scapy, which can be used for packet manipulation and reading PCAP files in python. The same utility is used to convert dataset into .csv file format. It has been analyzed there are total five classes such as Normal, MQTT_Bruteforce, Scan_A, Scan_sU, and Sparta. In a 'Normal' there is no attack; it is just a normal operation in the network. The 'MQTT_Bruteforce' is an attack file where the attacker tries to get the MQTT broker's password by trying all combinations of letters. 'Scan_A' is an aggressive scan attack where an attacker looks for all IOT nodes on the network. 'Scan_sU' is a UDP scan looking for all CCTV cameras on the network, and 'Sparta' means SSH brute force trying to gain access to the main server. The distribution of each packet classes is illustrated in Fig. 3.

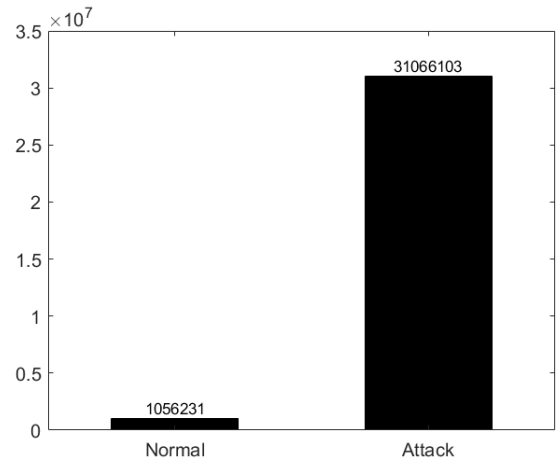


Fig. 3. Statistics of data distribution.

The graph trend exhibits that there is a total of 1026231 normal samples (Nor) and a total of 31066103 attack samples in the dataset. This analysis also reveals that the dataset is extensive and quite imbalanced, as more than 90% of the samples are subjected to the attack class.

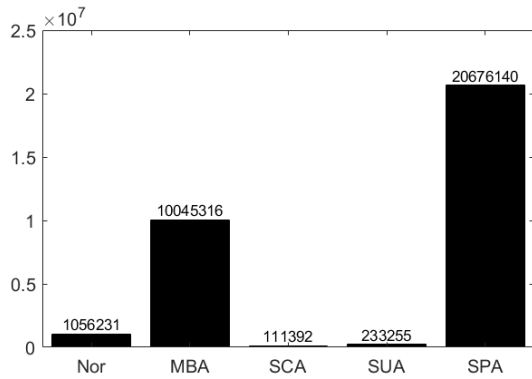


Fig. 4. Statistics of the attack class distribution.

Fig. 4 shows the distribution of attack classes. Sparta (SPA) and MQTT Bruteforce (MBA) attacks have a significantly higher number of samples compared to other attack classes such as scan_UDP (SUA) and scan_aggressive (SCA). From the exploratory analysis, it is clear that this dataset is associated with class imbalance problems and cannot be directly introduced to the learning model; otherwise, the identification or classification results tend toward majority classes, i.e., SPA and MBA.

Further a basic preprocessing operation is executed where the index of each data sample is reset, and an extra column is added for the label representing 0 for normal class and 1 for attack classes. The data frame consists of 33 columns representing feature descriptors and 32092334 rows representing the total number of samples belonging to each feature descriptor. Hence, the dataset is big and consumes 8 GB of memory on the local hard disk. Therefore, the study puts effort into optimizing the memory requirement to hold such a large amount of network traffic data for ease of computing tasks. In this regard, irrelevant feature descriptor such as index, timestamp, and src_port is dropped. The index column has no importance in the classification problem, and the timestamp is not considered because the study does not process data as time-series samples.

On the other hand, src_port is a random number that can also be ignored. As a result, data retrieval can be speedup and memory can be optimized to counter the curse of dimensionality issues. Apart from this, the data preprocessing includes outlier analysis, removal of missing and NAN values, filtering out irrelevant feature descriptors, and performing normalization and data encoding operations for the categorical data types. The study also splits predictor and response data which is then stored in the separate variables x and y. Also, it has been noticed that that in x data, there are three categorical values such as src_ip, dst_ip, and protocol. The src_ip giveaway the attack as normal since there is only one malicious computer in the network. Therefore, this feature is dropped. Similarly, the dst_ip is irrelevant for attack prediction since it does not contain information about whether the packet is an attack or normal. Therefore, this will be dropped for further analysis. The protocol feature is relevant and needs encoding. Since the study adopts an autoencoder for feature modeling, nominal encoding is preferred where protocol names are replaced with a random number. The remaining feature

descriptors need no encoding, and all Nan values are filled with zero to avoid computational error during the autoencoder training.

C. Performance Indicators

The performance analysis is carried out based on classification metrics and comparative assessment is done to analyze the effectiveness of proposed feature representation framework from the multiple dimensions. The performance indicators used are briefly discussed as follows:

$$Precision = \frac{1}{n} \sum \frac{TP_i}{TP_i + FP_i} \quad (10)$$

$$Recall\ rate = \frac{1}{n} \sum \frac{TP_i}{TP_i + FN_i} \quad (11)$$

$$F1_Score = \frac{2(Precision * Recall)}{Precision + Recall} \quad (12)$$

Where i denotes the number of classes, TP denotes the true positive metric that shows attack packets are classified correctly, TN denotes the true negative metric that shows the normal traffic samples are correctly classified as normal, FP denotes the false-positive metric that shows normal traffic samples are misclassified as attack samples, and FN denotes false-negative metric that shows attack samples are misclassified as normal samples.

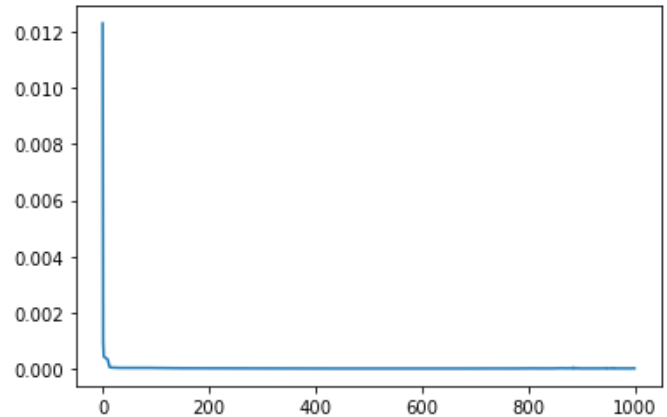


Fig. 5. Loss curve of trained autoencoder.

Fig. 5 presents the autoencoder's loss curve for analyzing its training performance. A closer analysis shows that a smooth curve has been obtained at each epoch, indicating a lower loss in the training process; hence a better performing autoencoder.

D. Comparative Analysis

The section presents outcome and performance analysis to demonstrate how the proposed feature representation technique enhances the IDS performance. In this regard, the performance evaluation is done by implementing ML classifiers with proposed feature representation model (i.e., WFR) and without feature representation model (i.e., WoFR). Since the proposed study focuses on multiclass intrusion detection, the precision, recall, and F1 score are computed for each data class. In addition, considering three ML classifiers demonstrates classification performance from the perspective of extensive analysis and against class imbalance problems.

1) *Analysis of NB classifier:* In this section the performance of proposed method will be evaluated with NB classifier (WFR) considering its comparison where the dataset is directly introduced with NB classifier (WoFR). The outcome obtained is shown in Table III.

TABLE III. NUMERICAL OUTCOME FOR NB CLASSIFIER (%)

	NB With Feature Representation (WFR)			NB Without Feature Representation (WoFR)		
	Precision	Recall	F1	Precision	Recall	F1
Norm	96.55	98.88	97.70	91.53	91.45	91.49
MBA	98.47	97.28	97.87	93.36	90.23	91.77
SCA	15.29	97.03	26.42	9.45	90.32	17.12
SUA	9.18	97.48	16.78	0	90.23	0
SPA	99.55	97.27	98.40	94.45	90.34	92.35

Table III shows the numerical outcome for the NB classifier for both case scenarios, i.e., WFR and WoFR. In the case of WFR, the NB exhibits a precision score of 96.55 %, recall rate of 98.88%, and F1 score of 97.70% for normal (Norm) class prediction. In the case of WoFR, the NB classifier shows a precision score of 91.53 %, recall rate of 91.45%, and F1 score 91.49%. Similar analysis can also be seen for the MBA class, where the NB classifier has achieved higher performance in the case of WFR. However, in predicting SCA and SUA, the NB classifiers do not show good performance. A closer analysis of outcome reveals the proposed system can efficiently enhance the performance of learning driven IDS without posing higher false positive rate unlike case of WoFR. This is because there are a smaller number of labels, and the dataset is associated with an imbalance factor. In the same way, the NB classifier also suffers in the classification of SUA. Although in the case of WFR, it has shown a 9.18% of precision score, whereas in case of WoFR, the NB could not predict SUA attack, i.e., it has achieved 0% precision rate.

The reason behind this that, SUA exhibits similar pattern to the normal class, and NB has no better generalization in this context. For SPA classification, NB classifier has shown better performance in case of WFR than a generic case, i.e., WoFR. A closer analysis reveals that the performance of IDS builds using NB classifier can be enhanced with the proposed method (WFR).

2) *Analysis of KNN classifier:* A closer analysis of the outcome from Table IV reveals that the KNN classifier shows better performance in the case of WFR than WoFR. However, at the same time, it can also be seen that KNN suffers from class imbalance factors for detecting both SCA and SUA classes.

TABLE IV. NUMERICAL OUTCOME FOR KNN CLASSIFIER (%)

	KNN With Feature Representation (WFR)			KNN Without Feature Representation (WoFR)		
	Precision	Recall	F1	Precision	Recall	F1
Norm	97.73	99.29	98.50	90.33	99.32	94.61
MBA	99.03	98.25	98.64	92.23	91.23	91.73
SCA	22.11	98.15	36.10	16.34	91.34	27.72
SUA	13.91	98.32	24.38	11.43	91.34	20.32
SPA	99.72	98.26	98.98	92.43	91.45	91.94

In the case of WFR, the KNN classifier performs better than the NB classifier. Also, for SPA attack, KNN performs very well, exhibiting 99.72%, 98.26, and 92.43% precision score, recall rate, and F1 score, respectively. On the other hand, KNN in case of WoFR, shows 92.43%, 91.45%, and 91.94% of precision, recall, and F1 scores, respectively. Moreover, the worst thing is that the SCA and SUA exhibits very similar pattern to the normal (Norm) class due to which the model did not reveal distinct pattern in the training process. Therefore, in the case of SUA and SCA detection, the KNN is not able to generalize and distinguish distribution SUA pattern as it faces difficulty in predicting the correct value of K because of closer similarity among the feature of normal and SUA. Although, simulation outcome shows that the proposed method enhances the performance of KNN for the multiclass intrusion detection even the dataset is imbalanced. But still there is a scope for the improvement for handling class imbalance problem, which will be done in our future work.

3) *Analysis of RF classifier:* The quantified outcome from Table V demonstrates that in the case of WFR, RF has achieved a precision score of 99.92 %, recall rate of 99.97%, and F1 score of 99.95%. Whereas, in the case of WoFR, RF classifier shows a precision score of 89.21 %, recall rate 89.63%, and F1 score 89.42%. For classification of a normal class, it can be analyzed that RF with proposed method (WFR) has achieved very good performance than other classifiers such as NB and KNN.

TABLE V. NUMERICAL OUTCOME FOR RF CLASSIFIER (%)

	RF With Feature Representation (WFR)			RF Without Feature Representation (WoFR)		
	Precision	Recall	F1	Precision	Recall	F1
Norm	99.92	99.97	99.95	89.21	89.63	89.42
MBA	99.97	99.94	99.95	89.22	90.32	89.77
SCA	89.57	100	94.5	87.34	90.23	88.76
SUA	82.63	99.87	90.44	82.43	90.23	86.15
SPA	99.99	99.94	99.96	92.24	90.34	91.28

In the case of WFR, the RF classifier for the SCA classification shows an 89% precision score, which is better than NB (15.29%) and KNN (22.11%). RF classifier also achieved a 100% recall rate where NB and KNN showed 97.03% and 98.15 %, respectively. With the proposed method, RF classifiers correctly identify normal packets and classify different attack classes. Hence, based on the above inferencing and discussion, it is concluded that the proposed work of feature representation can significantly enhance the outcome of any prediction system for designing enhanced IDS.

E. Discussion and Implication

The proposed work aims to enhance the performance of data driven or learning based IDS for the accurate detection and classification of attacks in dynamic networks such as IoT. The proposed study has considered the case study of MQTT enabled IoT networks because MQTT-based attacks are more dynamic and complex as they can easily mimic benign or normal behavior. In such cases, precise identification of attacks or intrusion becomes very difficult and challenging for any kind of IDS. Therefore, the proposed study has presented a feature representation framework that helps machine learning models better generalize the latent features and learn the distribution of data features more precisely. Three supervised classifiers have been implemented and trained with features enhanced by the proposed scheme (WFR). Their outcome is then compared with the classifiers trained with normal data i.e., without applying the proposed method (WoFR). The comparative assessment shows that the outcome achieved using the proposed feature representation (FR) is promising. The classifier trained with data that are enhanced by the proposed scheme WFR is superior. However, among three classifiers RF has achieved superior performance in both the cases of WFR and WoFR. This can be also evident in Fig. 6, where comparative analysis of each technique is given in terms of F1_score.

The RF is based on the design principle of the decision tree (DT) algorithm, which adopts similar principle of using din DT to determine how the features of a dataset should split nodes to form the tree. The KNN is based on Euclidian distancing; better performance is expected for the labels with lower support values. The NB classifier is a probabilistic model based on the Bayes' Theorem that has a specific assumption that the data is a particular feature of data not related to any other form of features. Therefore, this assumption of the NB classifier makes it incapable in the case of SUA attack classification as SCA attack that mimics a similar pattern of normal class. Since the dataset is associated with an imbalance factor and, KNN and NB suffers in classification and do not provide a good result.

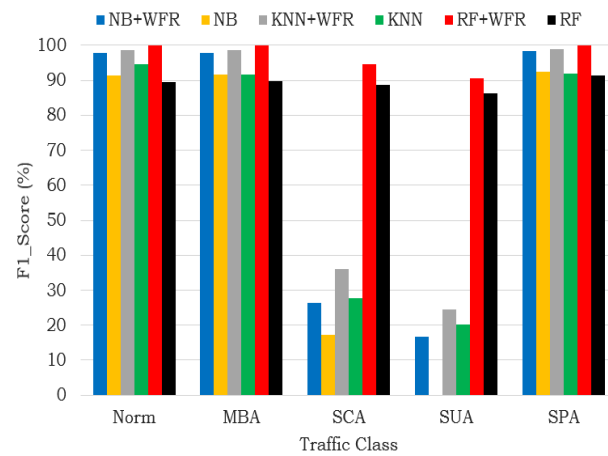


Fig. 6. Comparative analysis in terms of F1_score.

It is to be noted that the extensive outcome is not conducted due to the lack of sufficient and relevant research works on the similar dataset adopted in the proposed study. To date, the dataset has only 49 citations evident from the google scholar. However, in future work, the study considers handling data imbalance factors to strengthen the classification performance, making it more suitable to fit the real-time scenario. In future work, the proposed technique will be extended with the deep learning approach by calibrating it to generalize what to learn and what not to learn.

VI. CONCLUSION

The proposed research work deals with the intrusion detection system in the IoT ecosystem against MQTT attacks. The main contribution of the proposed study is a novel autoencoder-based clustering method that ultimately assigns a cluster-id to every packet. The proposed work offers an advanced data treatment method for critical feature modeling using an artificial neural network. The study also explores the effectiveness of different machine learning algorithms evaluated with standard and new datasets preprocessed via the proposed autoencoder scheme. The outcome statistics reveal the efficacy of the proposed feature representation technique, and comparative assessment justifies its scope in the real-time application. In future work, the proposed preprocessing scheme will be introduced with more advanced methods such as deep learning with a novel regularizer scheme to deal with dataset imbalance factors and achieve reliability in the classification process.

REFERENCES

- [1] E. Al-Masri et al., "Investigating Messaging Protocols for the Internet of Things (IoT)," in *IEEE Access*, vol. 8, pp. 94880-94911, 2020, doi: 10.1109/ACCESS.2020.2993363.
- [2] R. F. Al-Mutawa and F. Albourae, "A smart home system based on internet of things," *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 2, 2020.

- [3] M. S. Harsha, B. M. Bhavani, and K. R. Kundhavai, "Analysis of vulnerabilities in MQTT security using Shodan API and implementation of its countermeasures via authentication and ACLs," in 2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI), 2018.
- [4] S. K. Mousavi, A. Ghaffari, S. Besharat, and H. Afshari, "Security of internet of things based on cryptographic algorithms: a survey," *Wirel. netw.*, vol. 27, no. 2, pp. 1515–1555, 2021.
- [5] J. M. Carracedo et al., "Cryptography for Security in IoT," 2018 Fifth International Conference on Internet of Things: Systems, Management and Security, 2018, pp. 23-30, doi: 10.1109/IoTSMMS.2018.8554634.
- [6] J. Jang-Jaccard and S. Nepal, "A survey of emerging threats in cybersecurity," *J. Comput. Syst. Sci.*, vol. 80, no. 5, pp. 973–993, 2014.
- [7] F. Alfaleh and S. Elkhediri, "Efficient Security Solutions for IoT Devices" International Journal of Advanced Computer Science and Applications(IJACSA),2021.
- [8] M. Gurnathan and Moamin, "A review and development methodology of a LightWeight security model for IoT-based smart devices," *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 2, 2020.
- [9] A. B. Mohamed, N. B. Idris, and B. Shanmugum, "A brief introduction to intrusion detection system," in *Communications in Computer and Information Science*, Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 263–271.
- [10] Thakkar A, Lohiya R. A review on machine learning and deep learning perspectives of IDS for IoT: recent updates, security issues, and challenges. *Archives of Computational Methods in Engineering*. 2021 Jun;28(4):3211–43.
- [11] S. Wang, J. Cai, Q. Lin and W. Guo, "An overview of unsupervised deep feature representation for text categorization," *IEEE Transactions on Computational Social Systems*, vol. 6, no. 3, pp. 504–517, 2019.
- [12] C. Tang, M. Bian, X. Liu, M. Li, H. Zhou et al., "Unsupervised feature selection via latent representation learning and manifold regularization," *Neural Networks*, vol. 117, no. 9, pp. 163–178, 2019.
- [13] Z. Huang, X. Xu, J. Ni, H. Zhu and C. Wang, "Multimodal representation learning for recommendation in internet of things," *IEEE Internet of Things Journal*, vol. 6, no. 6, pp. 10675–10685, 2019.
- [14] O. Aydogdu and M. Ekinici, "A new approach for data stream classification: Unsupervised feature representational online sequential extreme learning machine," *Multimedia Tools and Applications*, vol. 79, no. 37–38, pp. 1–23, 2020.
- [15] N. Wang, W. Zhou, Y. Song, C. Ma, W. Liu et al., "Unsupervised deep representation learning for real-time tracking," *arXiv preprint arXiv: 2007.11984*, 2020.
- [16] X. Xu, H. Gu, Y. Wang, J. Wang, and P. Qin, "Autoencoder based feature selection method for classification of anticancer drug response," *Front. Genet.*, vol. 10, p. 233, 2019.
- [17] H. Sewani and R. Kashaf, "An autoencoder-based deep learning classifier for efficient diagnosis of autism," *Children (Basel)*, vol. 7, no. 10, 2020.
- [18] X. Qiang, C. Zhou, X. Ye, P. Du, R. Su et al., "CPPred-FL: A sequence-based predictor for large-scale identification of cell-penetrating peptides by feature representation learning," *Briefings in Bioinformatics*, vol. 21, pp. 11–23, 2020.
- [19] G. Kaur, A. H. Lashkari, and A. Rahali, "Intrusion traffic detection and characterization using deep image learning," in *Proc. IEEE Intl Conf Dependable, Autonomic Secure Comput., Int. Conf Pervas. Intell. Comput., Intl Conf Cloud Big Data Comput., Int. Conf Cyber Sci. Technol. Congr. (DASC/PiCom/CBDCOM/CyberSciTech)*, Aug. 2020, pp. 55–62.
- [20] M. A. Ferrag, L. Maglaras, S. Moschoyiannis, and H. Janicke, "Deep learning for cyber security intrusion detection: Approaches, datasets, and comparative study," *J. Inf. Secur. Appl.*, vol. 50, Feb. 2020, Art. no. 102419.
- [21] HariPriya, A.; Kulothungan, K. Secure-MQTT: An efficient fuzzy logic-based approach to detect DoS attack in MQTT protocol for internet of things. *EURASIP J. Wirel. Commun. Netw.* 2019, 2019, 90.
- [22] Ciklabakkal, E.; Donmez, A.; Erdemir, M.; Suren, E.; Yilmaz, M.K.; Angin, P. ARTEMIS: An intrusion detection system for MQTT attacks in Internet of Things. In *Proceedings of the 2019 38th Symposium on Reliable Distributed Systems (SRDS)*, Lyon, France, 1–4 October 2019; pp. 369–3692.
- [23] Faker, O.; Dogdu, E. Intrusion detection using big data and deep learning techniques. In *Proceedings of the 2019 ACM Southeast Conference*, Kennesaw, GA, USA, 18–20 April 2019; pp. 86–93.
- [24] O. Alkadi, N. Moustafa, B. Turnbull and K. -K. R. Choo, "A Deep Blockchain Framework-Enabled Collaborative Intrusion Detection for Protecting IoT and Cloud Networks," in *IEEE Internet of Things Journal*, vol. 8, no. 12, pp. 9463-9472, 15 June15, 2021
- [25] A. Fatani, M. Abd Elaziz, A. Dahou, M. A. A. Al-Qaness and S. Lu, "IoT Intrusion Detection System Using Deep Learning and Enhanced Transient Search Optimization," in *IEEE Access*, vol. 9, pp. 123448-123464, 2021
- [26] D. -C. Wang, I. -R. Chen and H. Al-Hamadi, "Reliability of Autonomous Internet of Things Systems With Intrusion Detection Attack-Defense Game Design," in *IEEE Transactions on Reliability*, vol. 70, no. 1, pp. 188-199, March 2021
- [27] M. Abdel-Basset, H. Hawash, R. K. Chakraborty and M. J. Ryan, "Semi-Supervised Spatiotemporal Deep Learning for Intrusions Detection in IoT Networks," in *IEEE Internet of Things Journal*, vol. 8, no. 15, pp. 12251-12265, 1 Aug.1, 2021
- [28] Dutta, V.; Chora's, M.; Pawlicki, M.; Kozik, R. A Deep Learning Ensemble for Network Anomaly and Cyber-Attack Detection. *Sensors* 2020, 20, 4583.
- [29] Aygun, R.C.; Yavuz, AG Network anomaly detection with stochastically improved autoencoder based models. In *Proceedings of the 2017 IEEE 4th International Conference on Cyber Security and Cloud Computing (CSCloud)*, New York, NY, USA, 26–28 June 2017; pp. 193–198.
- [30] Shone, N.; Ngoc, TN; Phai, VD; Shi, Q. A deep learning approach to network intrusion detection. *IEEE Trans. Emerg. Top. Comput. Intell.* 2018, 2, 41–50.
- [31] A. Tabassum, A. Erbad, A. Mohamed and M. Guizani, "Privacy-Preserving Distributed IDS Using Incremental Learning for IoT Health Systems," in *IEEE Access*, vol. 9, pp. 14271-14283, 2021.
- [32] Vaccari, I.; Chiola, G.; Aiello, M.; Mongelli, M.; Cambiaso, E. MQTTset, a New Dataset for Machine Learning Techniques on MQTT. *Sensors* 2020, 20, 6578.
- [33] Ahmadon, M.A.B.; Yamaguchi, N.; Yamaguchi, S. Process-Based Intrusion Detection Method for IoT System with MQTT Protocol. In *Proceedings of the 2019 IEEE 8th Global Conference on Consumer Electronics (GCCE)*, Osaka, Japan, 15–18 October 2019; pp. 953–956
- [34] Hindy, H.; Bayne, E.; Bures, M.; Atkinson, R.; Tachtatzis, C.; Bellekens, X. Machine learning based IoT Intrusion Detection System: An MQTT case study (MQTT-IoT-IDS2020 dataset). In *Proceedings of the International Networking Conference*, Online, 19–21 September 2020; Springer: Cham, Switzerland, 2020; pp. 73–84
- [35] Alaiz-Moreton, H.; Aveleira-Mata, J.; Ondicol-Garcia, J.; Muñoz-Castañeda, A.L.; García, I.; Benavides, C. Multiclass classification procedure for detecting attacks on MQTT-IoT protocol. *Complexity* 2019, 2019, 6516253
- [36] Khan, M.A., Khan, M.A., Jan, S.U., Ahmad, J., Jamal, S.S., Shah, A.A., Pitropakis, N. and Buchanan, W.J., 2021. A Deep Learning-Based Intrusion Detection System for MQTT Enabled IoT. *Sensors*, 21(21), p.7016.
- [37] I. Ullah and Q. H. Mahmoud, "Design and Development of a Deep Learning-Based Model for Anomaly Detection in IoT Networks," in *IEEE Access*, vol. 9, pp. 103906-103926, 2021, doi: 10.1109/ACCESS.2021.3094024.
- [38] H. Hindy, "Hanan Hindy," *IEEE DataPort*, 23-Jun-2020. [Online]. Available: <https://iee-dataport.org/authors/hanan-hindy>. [Accessed: 31-Oct-2022].

Iris Recognition Through Edge Detection Methods: Application in Flight Simulator User Identification

Sundas Naqeeb Khan^{1*}, Samra Urooj Khan², Onyeka Josephine Nwobodo³, Krzysztof Adam. Cyran⁴

Department of Graphics, Computer Vision and Digital Systems, Silesian University of Technology, Gliwice, Poland^{1, 3, 4}

Faculty of Electrical and Electronic Engineering, Universiti Tun Hussien Onn Malaysia (UTM), Johor, Malaysia²

Abstract—To meet the increasing security requirement of authorized users of flight simulators, personal identification is becoming more and more important. Iris recognition stands out as one of the most accurate biometric methods in use today. Iris recognition is done through different edge detection methods. Therefore, it is important to have an understanding of the different edge detection methods that are in use these days. Specifically, the biomedical research shows that irises are as different as fingerprints or the other patterns of the recognition. Furthermore, because the iris is a visible organism, its exterior look can be examined remotely using a machine vision system. The main part of this paper delves into concerns concerning the selection of the best results giving method of the recognition. In this paper, three edge detection methods, namely Canny, Sobel and Prewitt, are applied to the image of eye (iris) and their comparative analysis is discussed. These methods are applied using the Software MATLAB. The datasets used for this purpose are CASIA and MMU. The results indicate that the performance of Canny edge detection method is best as compared to Sobel and Prewitt. Image quality is a key requirement in image-based object recognition. This paper provides the quality evaluation of the images using different metrics like PSNR, SNR, MSE and SSIM. However, SSIM is considered best image quality metric as compared to PSNR, SNR and MSE.

Keywords—Identification; authentication; detection; canny; Sobel; Prewitt; PSNR; SNR; SSIM; MSE

I. INTRODUCTION

In the paper, the automatic authorization of flight simulator users is considered. Large pilot training centers require an increased security level to minimise the probability of unauthorised usage of the devices. In addition, pilot training procedures have to be rigorously verified as the number of hours spent in the simulator during virtual flights serves as proof of sufficient practice level for the pilot (in addition to hours spent in actual aircraft). Therefore, flights on certified flight simulators can be (to some extent defined by Regulatory Authorities) used as an alternative and cheaper counterpart to actual training flights because current flight simulators mimic the cockpit of the aircraft and the environment outside of the aircraft with great fidelity for maintaining the expertise of both civilians and military aircrews.

Flight simulators have significantly impacted the expertise and cutting-edge use of training pilots and aircraft crews, resulting in increased efficiency as training will not be affected by things like bad weather, aircraft space restriction and aircraft availability. It allows pilots to experience a wide range of flight scenarios and conditions to experience real-life

situational awareness. Haslbeck et al. [1] reported an experimental study where pilots with varying levels of experience and training had to fly 45 minutes of approach landing scenarios and compare their performance to manual flying. Their findings showed that a pilot with high levels of flight simulator training exhibits high and homogenous performance. Enhancing the pilot's monitoring techniques during the training session may aid in determining a pilot's performance in a flight simulator by employing object detection algorithms to analyze the pilot's head movement and eye gaze. The trainer can utilize this data to pinpoint areas where the pilot might benefit from extra instruction or development [2]. Škvareková et al. [3] research shows using eye tracking technologies to monitor eye movement to assist pilot training in flight simulators.

Its fidelity is not limited to training the pilot and aircrew. They are also used for research, enabling designers to examine the effects of various alternatives without spending money or waiting for various prototypes to be built [4], conduct aircraft incident investigations, and better understand ergonomic relationships [5]. Flight simulation is a multidisciplinary field that draws on many different areas of study, including human factors, motion actuation, sensory perception and visual image processing [6 - 7].

The use of flight simulators in pilot training has become indispensable, likewise the need to secure this gadget by minimising unauthorised users. They offer secure, affordable and adaptable alternatives to teach pilots in various scenarios and settings, enhancing their knowledge, self-assurance and decision-making capabilities [8 - 9]. As evident in Fig. 1, the flight simulator developed in the WrightBroS project is depicted.



Fig. 1. MCC FNTP II flight simulator of boeing 737.

*Corresponding author

The technology of digital imaging helps to manipulate digital photos using computers. The pre-processing, enhancement and display extraction of information is three general phase's frameworks where all forms of data must pass while employing digital technologies. A digital image consists of a limited number of elements, each having a specific position and values. These items are known as picture elements, pixels and image elements. The term pixel is the most usable form of the digital picture elements [10 - 11].

The word iris is a classic term. As with the colored part of the external eye, iris appears as of the 16th century and was taken to denote the variegated look of its structure [12]. The human iris contains many characteristics including freckles, crown, strips, furshes, crypts, etc., and is the annular section between pupils and sclera. In contrast to other biometric characteristics based on personal authentication recognition of iris can achieve great precision because of the rich texture of iris patterns [13]. Fig. 2 shows the full eye image with all its features.

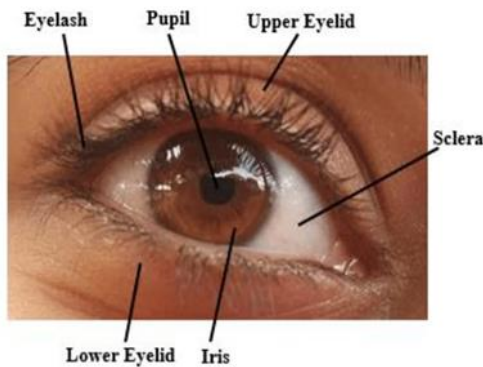


Fig. 2. A general view of a human eye [15].

Although the history of iris recognition comes back to the 19th century, automatic iris recognition is a newly emergent issue in biometrics [14]. It seems that the French ophthalmologist Alphonse Bertillon was the first one who proposed the use of iris pattern (color) as a basis for personal identification. Flom and Aran Safir suggested also using the iris as the basis for a biometric in 1981 [15].

Regarding the level of precision, it is apparent the fact that iris recognition requires an excellent recognition rate and, as a result, applying edge detection approaches for this problem are efficient, which has not been extensively investigated in previous studies for such systems.

The main goal of this study is to use different techniques to detect the edges of the images and then examine the results to

see which technique effectively detects and sharpens the edges. For this study, the following primary objectives have been proposed:

- To recognize iris through edge detection methods.
- To compare the results of the selected edge detection methods.
- To check the performance of the methods via evaluation measures.

II. LITERATURE REVIEW

This section includes algorithms, methods, procedures, processes, approaches, techniques, detectors, models, and filters, as well as all of the history research where their links were used as evolutionary and provided the optimized form of solutions for edge detection issues. As a result of the researcher's efforts, the number of filters has grown over time. They worked hard to improve, enhance, co-relate, and collaborate on various methods to solve edge detection problems in image processing, and they created new, modified methods.

Edge detection is at the forefront of image processing for object detection, understanding edge detection methods are critical. ISEF, Canny, Marr-Hildreth, Sobel, Kirsch, Lapla1 and Lapla2 and Hough transform are the most important and widely used edge detection techniques [16].

Edge detection methods are further categorized into Gradient based (1st derivative order) and Gaussian based (2nd derivative order). Gradient based methods can be classified into Sobel, Prewitt, Robert, Kirsch, Robinson and Frei-Chen. Gaussian based methods are divided into Canny edge detection, Difference of Gaussian, Laplacian of Gaussian (like, Marr-Hildreth method) and Zero Crossing [17 – 18]. Fig. 3 shows the taxonomy of the methods which are related to the detection of the iris.

The Canny edge detector is a common and effective edge feature detector that is utilized in many computer vision algorithms as a pre-processing step [19 – 20]. The steps in the Canny Edge Detection Algorithm are as follows [21]:

- Step 1: Apply a Gaussian filter to the image to smooth it out.
- Step 2: Using finite-difference approximations for partial derivatives, compute the gradient magnitude and orientation.
- Step 3: Apply non-maxima suppression to the gradient magnitude, then detect and link edges using the double thresholding approach.
- Step 4: The operator that optimizes the product of signal-to-noise ratio and localization is approximated by the Canny edge detector. It is usually a Gaussian's first derivative.

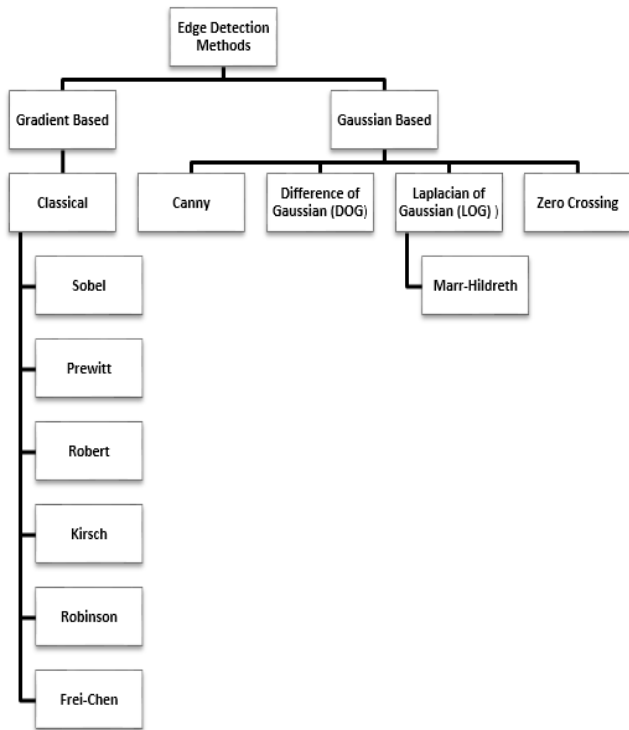


Fig. 3. Taxonomy of image detection methods.

For edge detection, the Sobel operator is utilized. The Sobel approach finds image edges using the derivative approximation in the edge function [22]. The Sobel algorithm is broken down into four steps [23]:

- Step 1: Converting the image to grayscale is the first step.
- Step 2: Using the Sobel-x filter to enlarge the grey image.
- Step 3: Applying the Sobel-y filter to the grey image.
- Step 4: Determining the magnitude and direction of the gradient.

In image processing, the Prewitt operator is commonly employed in edge detection techniques. It is a discrete differentiation operator that computes an approximation of the picture intensity function's gradient. The Prewitt operator returns either the relevant gradient vector or the norm of this vector at each location in the image [24]. The Prewitt method is broken down into four steps [17]:

- Step 1: Read an image as input.
- Step 2: Create a grayscale image from a true-color RGB image.
- Step 3: Double-click the image to enlarge it.
- Step 4: Fill the filtered image matrix with zeros before starting.

The literature collects data on iris-based identification methods. It also eloquently discusses each method's effectiveness and highlights areas for future investigation for concerned scholars. The results of this study may be utilized to enhance iris detection system development.

TABLE I. SUMMARY OF THE ARTICLES ALONG THEIR PROMINENT METHODS, DATASET AND RESULTS

Biometric Trait	Ref.	Methods	Dataset	Results
Iris	[25]	Circular transform convolutional network	Hough and neural CASIA-V3	Left and Right-side iris recognition accuracy is 99%.
	[26]	Hough transform and region-based convolutional network	neural CASIA	99.14% of accuracy
	[27]	Hough transform and region-based convolutional network	neural CASIA-V4	Accuracy is 96.3%
	[28]	Convolutional network	neural IITD	97.46% of accuracy
	[29]	Hough transform with canny edge detection	CASIA-VI	Accuracy is 93.33%
	[30]	Circular transform with contour	Hough with active CAISA	88.3% of accuracy

Table I describes numerous iris recognition methods along dataset and their results which learned from earlier references. Some of these articles address the entire recognition procedure while others focus on the methodology only.

III. MATERIALS AND METHODS

The main step is pre-processing, feature extraction and matching where edge detection methods are used and try to improve the efficiency. The pre-processing steps comprise the conversion of the image from a color image to gray-scale image, iris localization, edge detection, filtration, etc. Now the Fig. 4 represents the proposed model for image processing especially iris recognition.

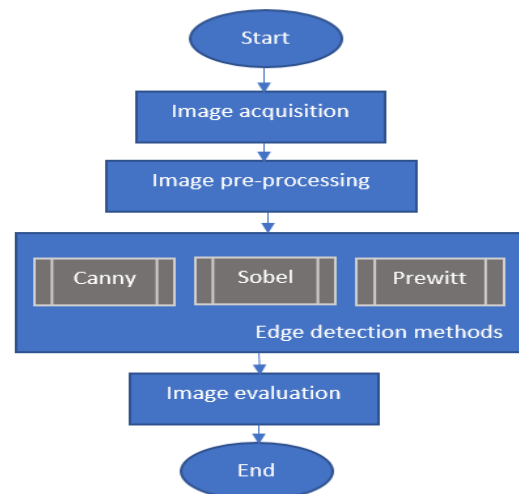


Fig. 4. Proposed model for iris recognition.

According to Fig. 5, this section covers the project's basic flow, the suggested iris recognition model, and the project flow chart in a more organized manner. Level 1, Level 2 and Level 3 of Project flow chart are demonstrated in detail. The

activity of getting an image from a source is known as image acquisition; image segmentation is the process of separating an image into sections or regions; and feature extraction techniques are used to obtain features that will be useful in image classification and recognition. Edge detection will be done by using Canny, Sobel and Prewitt edge detection methods for iris recognition.

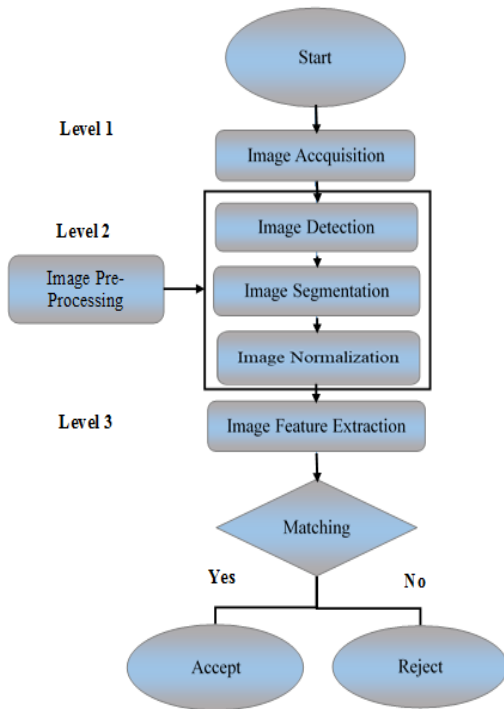


Fig. 5. Project flowchart.

IV. RESULTS AND DISCUSSION

This experiment represents the analysis of the various methods of edge detection. Edge detection is a fundamental step, and it is necessary to point out the true edges to get the best results. That is why it is important to choose a good edge detection method. The experiment was done on MATLAB software and tested with the eye image. To extract the clean edges is main objective. In this experimental analysis, the results of the five images from the CASIA dataset are obtained in Table II. The results of the methods of Canny, Prewitt, and Sobel on the original images are analyzed. After this experimental analysis, it is found that the canny edge detection method is working well in comparison with Sobel and Prewitt. In the empirical testing, it can be seen clearly that Prewitt and Sobel do not produce good sharp edges. Sobel is not very accurate in edge detection as it detects thick and rough edges and does not give appropriate results. Analysis shows that

canny extraction of image features without affecting and altering the feature is good. It is observed that the canny edge detector method is a better edge detector method of forming the edges for inner as well as outer lines. Hence, it is experimentally proved that canny is the most efficient and optimal method of edge detection.

Here, Table III illustrates the results of five images from the MMU dataset used in this experiment. After experimental examination, it is clear that canny method is more accurate rather than the Prewitt and Sobel. These are simple in evaluation separately then it improves the image results in the form of features extraction.

A. Image Quality Metrics for CASIA Dataset

For five images of CASIA Dataset, PSNR (Peak Signal to Noise Ratio), SNR (Signal to Noise Ratio), MSE (Mean Square Error), SSIM (Structure Similarity Index Method) for different methods have been calculated. Noisy images are converted into filtered images by using Gaussian noise to the images and filtered by Gaussian Filter and then results for all metrics are compared. From Table IV, it can be seen that all metrics have given almost consistent results. But SSIM is normalized as compared to PSNR, SNR and MSE. So, from this analysis the SSIM is comparatively better than MSE, SNR and PSNR metrics from human visual perspective.

B. Image Quality Metrics for MMU Dataset

Similarly, in Table V now, it is calculated PSNR, SNR, MSE, and SSIM for five images from the MMU Dataset. However, SSIM is balanced in terms of visibility, whereas MSE and PSNR are not. As a result, SSIM can be dealt in a more straightforward manner than MSE and PSNR. This is because MSE and PSNR are absolute errors, whereas SSIM produces perception and image errors. When the noise level rises, the recovery quality of the output image worsens as well. So, SSIM is better amongst all other metrics.

C. Comparative Analysis of the CASIA and MMU Dataset with Different Approaches

In Table VI, there are a number of articles presents their empirical testing results in the form of accuracy for iris recognition. Besides this, the evaluation datasets are also described for the further processing of the testing whereas the 99.7% ratio of the accuracy according to the CASIA-V1.0 and this gives the results from [38]. The result of 97.2% accuracy of the CASIA-V3.0/V4.0 is presented in the good form of the [38]. In [37], MMU-V1.0 dataset of the iris provides the high form of the ratio which is 99.45% accuracy as compare to the other research articles.

TABLE II. EDGE DETECTION METHODS APPLIED ON CASIA DATASET











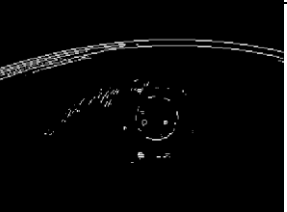
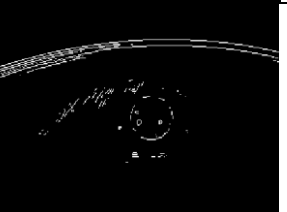






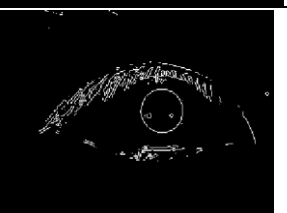
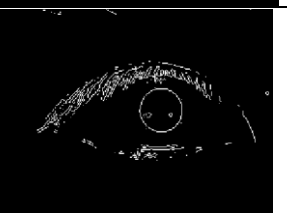
Image No.	Original	Canny	Prewitt	Sobel
1				
2				
3				
4				
5				

TABLE III. EDGE DETECTION METHODS APPLIED ON MMU DATASET

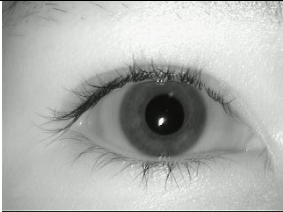

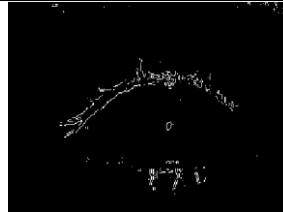
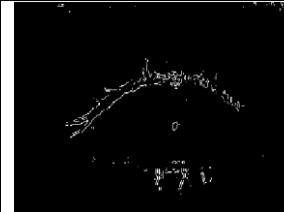
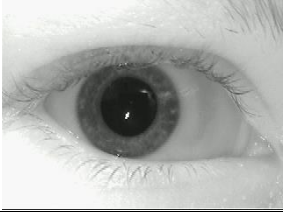
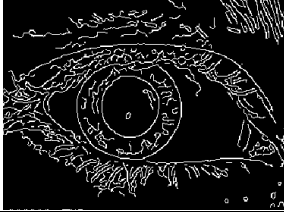


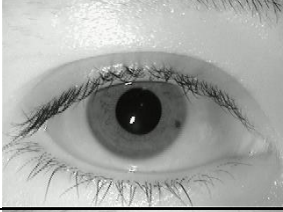
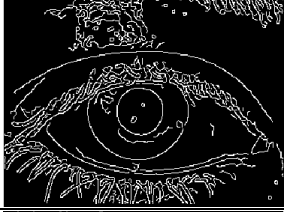








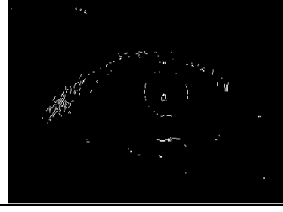
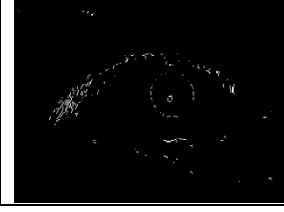
Image No.	Original	Canny	Prewitt	Sobel
1				
2				
3				
4				
5				

TABLE IV. IMAGE QUALITY METRICS FOR CASIA DATASET











Image No.	Noisy Image (Gaussian Filter 0.09)	Filtered Image	PSNR	SNR	MSE	SSIM
1			17.4355	10.8874	1179.1391	0.146182
2			17.4461	11.5067	1170.7628	0.187546
3			17.4486	9.2077	1170.0936	0.169675
4			17.4835	9.0891	1160.7283	0.209050
5			18.5162	15.7319	915.0796	0.280726

TABLE V. IMAGE QUALITY METRICS FOR MMU DATASET






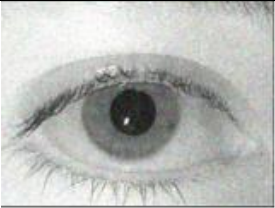




Image No.	Noisy Image (Gaussian Filter 0.09)	Filtered Image	PSNR	SNR	MSE	SSIM
1			18.5219	15.7376	913.8797	0.280778
2			18.7602	16.0665	865.0809	0.268316
3			18.0561	14.7173	1017.3440	0.299774
4			18.0612	14.8339	1016.1554	0.239681
5			18.0562	14.9164	1017.3274	0.244308

TABLE VI. COMPARATIVE ANALYSIS OF THE DIFFERENT APPROACHES ALONG ACCURACY FOR EVALUATION

Ref.	Methods	Evaluation	Accuracy
[31]	Hough transform	CASIA-V3.0/V4.0	93.12%
[32]	Fourier transform	CASIA-V1.0	96%
[33]	Neural network	CASIA-V3.0/V4.0	95.36%
[34]	Principle component analysis	CASIA-V3.0/V4.0	90%
[35]	Refine connect extend smooth	CASIA-V3.0/V4.0	95.1%
[35]	Refine connect extend smooth	CASIA-V1.0	96.48%
[38]	Leading edge detector	CASIA-V1.0	99.7%
[38]	Leading edge detector	CASIA-V3.0/V4.0	97.2%
[36]	Hough transform along hamming distance	MMU-V1.0	95.66%
[37]	Edge detection operators	MMU-V1.0	99.45%
[38]	Leading edge detector	MMU-V1.0	98%
[39]	Morphological operators	MMU-V1.0	98.78%

D. Graphical Analysis of Image Quality Metrics for CASIA Dataset

Fig. 6 illustrates the PSNR for the CASIA dataset is almost constant, i.e., 17.4, from starting image CI-1 (CASIA Image) to image CI-4. After CI-4, there is a sharp increase in the PSNR. In C-5, there is a gradually change in values from 17.4 to 18.5 which shows an abrupt change as compared to four other images. CI-5 image shows PSNR value 18.5162 which is high as compared to other values. High value of PSNR gives better resolution of image. Therefore, CI-5 gives better result in this case.

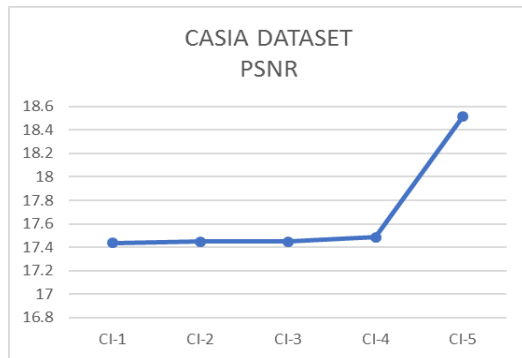


Fig. 6. PSNR of CASIA dataset.

The fluctuation shows in Fig. 7 with the value of SNR between first and fourth image while there is a hike between fourth and fifth image. CI-1 image shows SNR value 10.8, then there is a slight change and CI-2 image shows SNR value 11.5. Image CI-3 shows abrupt change with SNR value 9.2. CI-4 image with minor change shows SNR value 9.0. CI-5 image shows SNR value 15.7. In this graph image CI-5 shows good resolution because of high value.

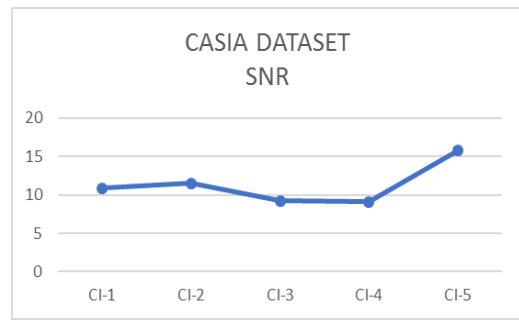


Fig. 7. SNR of CASIA dataset.

According to the Fig. 8, the SSIM value for each preceding image is larger than the previous one except for the image CI-3. SSIM values ranges from -1 to 1. -1 is the lowest value which shows that there is no similarity between images and 1 value is considered to be the highest value which shows that there is a high similarity between images. Now here, CI-1 image provides better resolution than all other images with SSIM values.

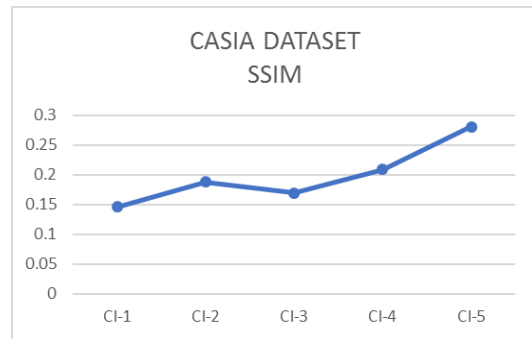


Fig. 8. SSIM of CASIA dataset.

Fig. 9 presents that the value of MSE for the sequentially four images are almost constant while for the 5th image, there is a drop in the value of MSE. The MSE value is the average difference between all pixels in the image. A greater discrepancy between the original and processed image is indicated by a higher MSE value. MSE value is positive. A value close to 0 is considered a better value. So, image 4 with MSE value 1160.7283 shows better result in CASIA dataset

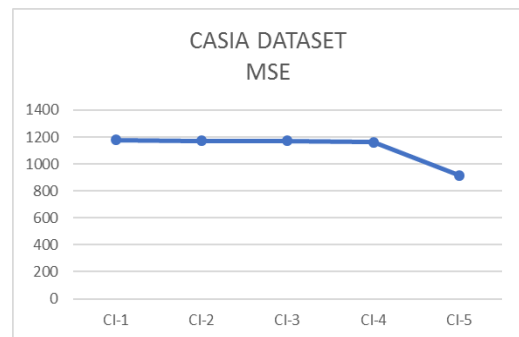


Fig. 9. MSE of CASIA dataset.

E. Graphical Analysis of Image Quality Metrics for MMU Dataset

The PSNR block calculates the PSNR between two images in decibels. This ratio is used to compare the features of the original images and filtered image. PSNR reflects the peak error measurement. MMUI-1 shows PSNR value 18.5219 while MMUI-2 shows PSNR value 18.7602. MMUI-3, MMUI-4 and MMUI-5 represent less constant PSNR values as compared to MMUI-1 and MMUI-2 which shows MMUI-2 with greater PSNR value (18.7602) is better result in Fig. 10.

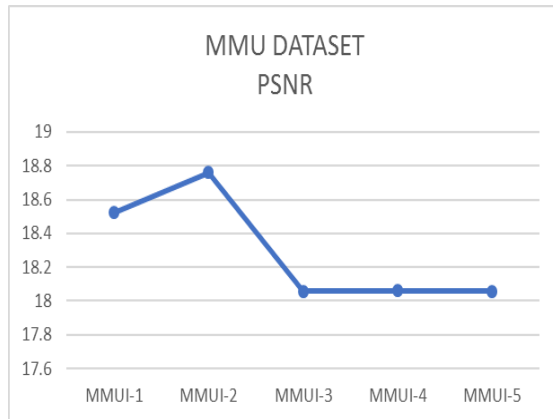


Fig. 10. PSNR of MMU dataset.

Therefore, Fig. 11 represents results of SNR values for MMU Dataset. The SNR measures how well a desired signal compares to background noise. MMUI-1 shows better result in graph as compared to MMUI-3, MMUI-4 and MMUI-5. As there is a decline after MMUI-2, so MMUI-1 is finer among all values.

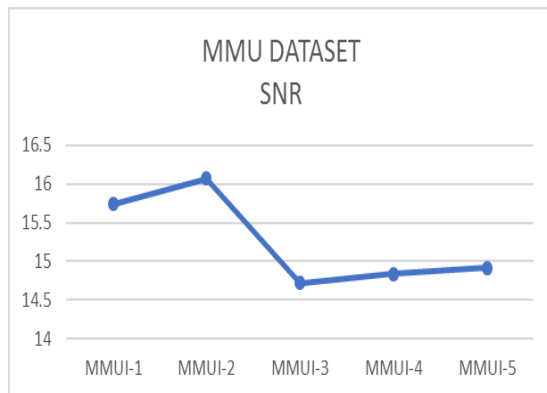


Fig. 11. SNR of MMU dataset.

The results of SSIM values for MMU Dataset are illustrated in Fig. 12. MMUI-1 shows 0.280778 SSIM result as there is a fall off after image 1. MMUI-2 shows 0.268316 SSIM value in the meanwhile there is a lift after image 2. MMUI-3 present 0.299774 SSIM value while there is a turn down in MMUI-4 and MMUI-5. So, MMUI-3 provides better result.

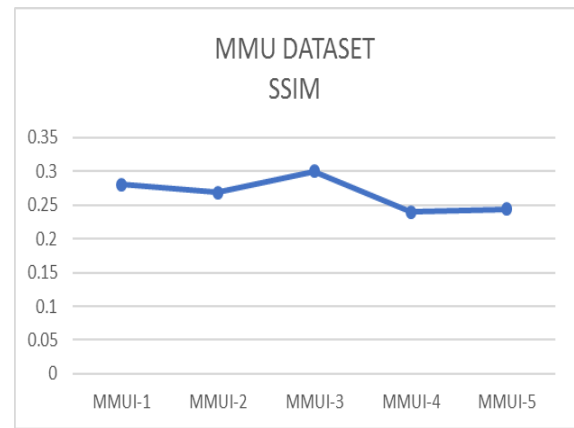


Fig. 12. SSIM of MMU dataset.

In Fig. 13, MSE value for MMUI-1 lies between 900 and 950. There is sudden fall after image 1; MMUI-2 shows MSE values between 850 and 900. After image 2, there is abrupt change in graph. MMUI-3, MMUI-4 and MMUI-5 show constant values at their peak level values between 1000 and 1050. A value close to 0 is considered a better value. So, MMUI-2 with exact MSE value 865.0809 shows better result in MMU dataset as it is the lowest value as compared to all other values in the graph.

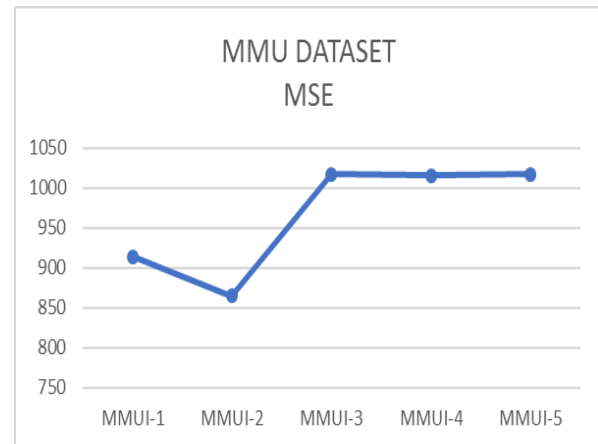


Fig. 13. MSE of MMU dataset.

F. Comparative Analysis of Image Quality Metrics

The Fig. 14 demonstrates the comparative analysis for the PSNR values of CASIA and MMU dataset. A major difference can be seen between two datasets. The CASIA dataset shows constant result for four images and then there is an increase in the graph for image 5. The MMU dataset shift for three images and then shows constant result for 4 and 5 images. First four images of MMU dataset are strongly preferred as compared to first four images of CASIA dataset because of high PSNR value. But in comparison, 5th image of CASIA dataset is better than 5th image of MMU dataset. So, in this graph MMU dataset results are good in the form of PSNR.

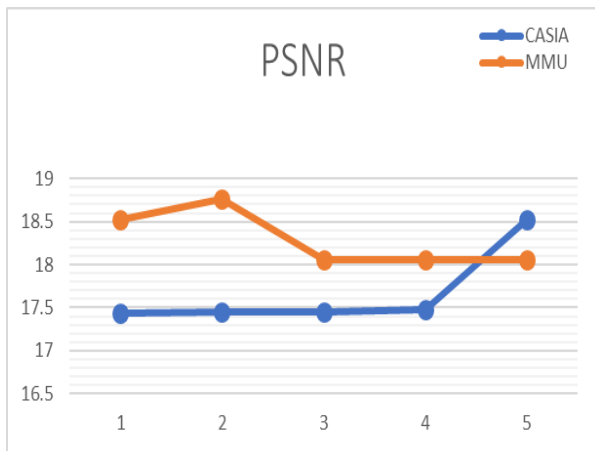


Fig. 14. Comparative analysis of PSNR.

In the Fig. 15 graphical analyses, SNR values of CASIA and MMU dataset are compared. Differences between two datasets distinguish both of them. The CASIA dataset shows oscillation for five images in the graph. On the other side, MMU dataset describes shift for three images and then shows constant result for 4 and 5 images. First four images of MMU dataset are strongly preferred as compared to first four images of CASIA dataset because of high SNR value. But in comparison, 5th image of CASIA dataset is better than 5th image of MMU dataset because of slight high difference. So, in this graph MMU dataset results are better according to the SNR.

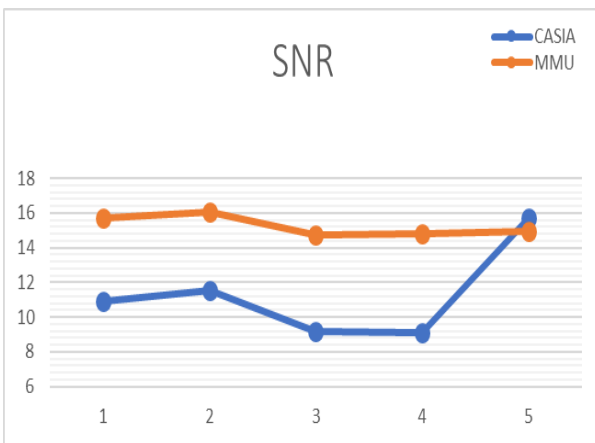


Fig. 15. Comparative analysis of SNR.

The SSIM values of CASIA and MMU dataset are analyzed in Fig. 16. Up and downs between two datasets distinguish rather first four images of MMU dataset are better as compared to first four images of CASIA dataset because of high values near to zero. But in comparison, 5th image of CASIA dataset is better than 5th image of MMU dataset as slightly high difference. So, MMU dataset results are strongly preferred.

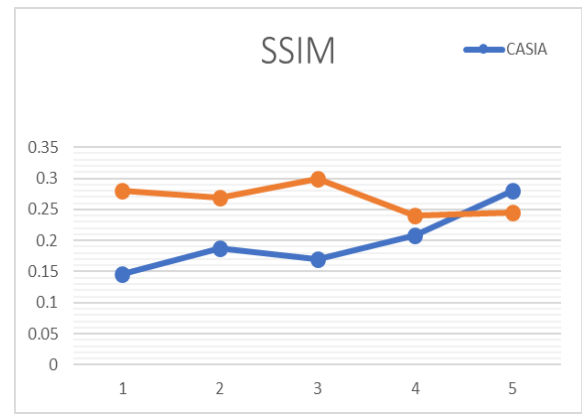


Fig. 16. Comparative analysis of SSIM.

The graphical analysis shows MSE values of CASIA and MMU dataset in Fig. 17. The CASIA dataset shows variation for five images in the graph. Besides this, MMU dataset represents rise and fall for 5 images. First four images of MMU dataset are better as compared to CASIA dataset because of low values near to zero. But in comparison, 5th image of CASIA dataset is better than 5th image of MMU dataset because of lowest value. So, in this graph MMU dataset results are strongly good.

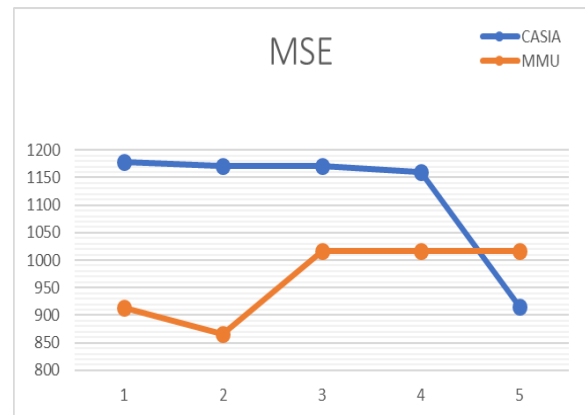


Fig. 17. Comparative analysis of MSE.

V. CONCLUSION

This research covers all of the conclusions from the research methods. The main objectives are completed by comparative analysis of the edge detection methods which is the required tasks. Three edge detection methods have been proposed for identifying the borders of objects within images from two benchmark datasets (CASIA, MMU), removing blurriness from the findings through edge detection techniques, and evaluating performance using parameters. The study's accomplishments are revealing a considerable improvement in image quality and crispness. This study also includes suggestions for future improvements that other researchers should explore in order to concentrate on the experimental results, obtained the defects detected in this work, and apply the highlighted work to other areas that need attention.

In the future, a new filtered method can be created to overcome the constraint in order to improve image quality and enrich the image by minimizing noise.

ACKNOWLEDGMENT

The authors would like to acknowledge that this paper is based on the results achieved within the WrightBroS project. This project has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 822483. Supplementarily, this research work has been co-financed from Polish financial resources for science in 2019-2023 and conferred for the implementation of the co-financed international project.

Disclaimer. The paper reflects only the author's view, and the Research Executive Agency (REA) is not responsible for any use that may be made of the information it contains.

CONFLICT OF INTEREST

The authors confirm that there is no conflict of interest involve with any parties in this research study.

REFERENCES

- [1] Haslbeck, A. et al. (2014) 'A flight simulator study to evaluate manual flying skills of airline pilots', in Proceedings of the Human Factors and Ergonomics Society. Human factors an Ergonomics society Inc., pp. 11-15. Available at: <http://doi.org/10.1177/1541931214581003>.
- [2] Lefrançois, O., Matton, N. and Causse, M. (2021) 'Improving airline pilots' visual scanning and manual flight performance through training on skilled eye gaze strategies', *Safety*, 7(4). Available at: <https://doi.org/10.3390/safety7040070>.
- [3] Skvarekova, I. et al. (2020) 'Eye Track Technology in Process of Pilot Training Optimization', in NTinAD 2020 - New Trends in Aviation Development 2020 - 15th International Scientific Conference, Proceedings. Institute of Electrical and Electronics Engineers Inc., pp. 206-210. Available at: <https://doi.org/10.1109/NTAD51447.2020.9379071>.
- [4] Smaili et al. (2017) CUSTOMER: European Commission Practices to identify and prevent adverse aircraft-and-rotorcraft-pilot couplings-A ground simulator perspective NLR-Netherlands Aerospace Centre. Available at: www.nlr.nl.
- [5] Boril, J. et al. (2015) 'Aviation simulation training in the Czech air force', in AIAA/IEEE Digital Avionics Systems Conference - Proceedings. Institute of Electrical and Electronics Engineers Inc., pp. 9A21-9A213. Available at: <https://doi.org/10.1109/DASC.2015.7311484>.
- [6] Boril, J., Jirgl, M. and Jalovecky, R. (2016) Use of Flight Simulators in Analyzing Pilot Behavior.
- [7] Stroosma, O., Van Paassen, R. and Mulder, M. (2003) USING THE SIMONA RESEARCH SIMULATOR FOR HUMAN-MACHINE INTERACTION RESEARCH.
- [8] Socha, V. et al. (2016) Training of Pilots Using Flight Simulator and its Impact on Piloting Precision
- [9] Lazić, D.A., Grujić, V. and Tanasković, M. (2022) 'The role of flight simulation in flight training of pilots for crisis management', *South Florida Journal of Development*, 3(3), pp. 3624-3636. Available at: <https://doi.org/10.46932/sfjdv3n3-046>.
- [10] Bauer, M.W., Knill, C. (2007): Management reforms in international organizations. – *Nomos Verlagsgesellschaft* 226p.
- [11] Bauer, T.N., Bodner, T., Erdogan, B., Truxillo, D.M., Tucker, J.S. (2007): Newcomer adjustment during organizational socialization: a meta-analytic review of antecedents, outcomes, and methods. – *Journal of Applied Psychology* 92(3): 707-721.
- [12] Berger, B.A. (2000): Incivility. – *American Journal of Pharmaceutical Education* 64(4): 445-450.
- [13] Billsberry, J., Talbot, D., Hollyoak, B. (2011): The Social Construction of PE Fit. – In *Academy of Management Journal-AMJ*. Retrieved from: <https://pureportal.coventry.ac.uk/en/publications/the-social-construction-of-pe-fit>
- [14] Brown, D., Warschauer, M. (2006): From the university to the elementary classroom: Students' experiences in learning to integrate technology in instruction. – *Journal of Technology and Teacher Education* 14(3): 599-621.
- [15] Cole, M.S., Shipp, A.J., Taylor, S.G. (2016): Viewing the interpersonal mistreatment literature through a temporal lens. – *Organizational Psychology Review* 6(3): 273-302.
- [16] Cortina, L.M., Kabat-Farr, D., Leskinen, E.A., Huerta, M., Magley, V.J. (2013): Selective incivility as modern discrimination in organizations: Evidence and impact. – *Journal of Management* 39(6): 1579-1605.
- [17] Cortina, L.M., Magley, V.J., Williams, J.H., Langhout, R.D. (2001): Incivility in the workplace: incidence and impact. – *Journal of Occupational Health Psychology* 6(1): 64-80.
- [18] George, J.M., Jones, G.R. (2001): Towards a process model of individual change in organizations. – *Human Relations* 54(4): 419-444.
- [19] Giumetti, G.W., Hatfield, A.L., Scisco, J.L., Schroeder, A.N., Muth, E.R., Kowalski, R.M. (2013): What a rude e-mail! Examining the differential effects of incivility versus support on mood, energy, engagement, and performance in an online context. – *Journal of Occupational Health Psychology* 18(3): 297-309.
- [20] Goparaju, L., Jha, C.S. (2010): Spatial dynamics of species diversity in fragmented plant communities of a Vindhyan dry tropical forest in India. – *Tropical Ecology* 51 1: 55-65.
- [21] Griffin, M.J., Abergel, A., Abreu, A., Ade, P.A., André, P., Augeres, J.L., Babbidge, T., Bae, Y., Baillie, T., Baluteau, J.P., Barlow, M.J. (2010): The Herschel-SPIRE instrument and its in-flight performance. – *Astronomy & Astrophysics* 518: 7p.
- [22] Griffin, M.A., Neal, A., Parker, S.K. (2007): A new model of work role performance: Positive behavior in uncertain and interdependent contexts. – *Academy of Management Journal* 50(2): 327-347.
- [23] Gruman, J. A., Saks, A.M., Zweig, D.I. (2006): Organizational socialization tactics and newcomer proactive behaviors: An integrative study. – *Journal of Vocational Behavior* 69(1): 90-104.
- [24] Hershcovis, M.S., Barling, J. (2010): Towards a multi-foci approach to workplace aggression: A meta-analytic review of outcomes from different perpetrators. – *Journal of Organizational Behavior* 31(1): 24-44.
- [25] Therar, H. M., Mohammed, L. D. E. A., & Ali, A. J. (2021, June). Multibiometric system for iris recognition based convolutional neural network and transfer learning. In *IOP Conference Series: Materials Science and Engineering* (Vol. 1105, No. 1, p. 012032). IOP Publishing.
- [26] Jayanthi, J., Lydia, E. L., Krishnaraj, N., Jayasankar, T., Babu, R. L., & Suji, R. A. (2021). An effective deep learning features based integrated framework for iris detection and recognition. *Journal of ambient intelligence and humanized computing*, 12, 3271-3281.
- [27] Azam, M. S., & Rana, H. K. (2020). Iris recognition using convolutional neural network. *International Journal of Computer Applications*, 175(12), 24-28.
- [28] Alnahari, W. (2021). Convolutional Neural Network for Iris Recognition.
- [29] Farouk, R. H., Mohsen, H., & Abd El-Latif, Y. M. (2022, March). Iris Recognition System Techniques: A Literature Survey and Comparative Study. In *2022 5th International Conference on Computing and Informatics (ICCI)* (pp. 194-199). IEEE.
- [30] Akinfende, A. S., Imoize, A. L., & Ajose, O. S. (2020). Investigation of iris segmentation techniques using active contours for non-cooperative iris recognition. *Indonesian Journal of Electrical Engineering and Computer Science*, 19(3), 1275-1286.
- [31] Umer, S., Dhara, B. C., & Chanda, B. (2018). An iris recognition system based on analysis of textural edgeness descriptors. *IETE Technical Review*, 35(2), 145-156.

- [32] Hamd, M. H., & Ahmed, S. K. (2018). Biometric system design for iris recognition using intelligent algorithms. *International Journal of Modern Education and Computer Science*, 12(3), 9.
- [33] Ahmadi, N., & Akbarizadeh, G. (2018). Hybrid robust iris recognition approach using iris image pre-processing, two-dimensional gabor features and multi-layer perceptron neural network/PSO. *Iet Biometrics*, 7(2), 153-162.
- [34] Abdulmunem, E., & Abbas, S. H. (2018). Iris recognition using SVM and BP algorithms. *International Journal of Engineering Research and Advanced Technology (IJERAT)*, 4(5), 30-37.
- [35] Ak, T. A., & Steluta, A. (2021). An iris recognition system using a new method of iris localization. *International Journal of Open Information Technologies*, 9(7), 67-76.
- [36] Masek, L. (2003). Recognition of human iris patterns for biometric identification.
- [37] Yuan, W., Lin, Z., & Xu, L. (2006, January). A rapid iris location method based on the structure of human eyes. In *2005 IEEE Engineering in Medicine and Biology 27th Annual Conference* (pp. 3020-3023). IEEE.
- [38] Ali, S. A., & George, L. E. (2013). New Approach of Iris Localization for Personal Identification. In *International Conference on Information Technology in Signal and Image Processing (ITSIP-2013)*, Elsevier.
- [39] Hashim, A. T., & Saleh, Z. A. (2019). Fast Iris localization based on image algebra and morphological operations. *journal of university of Babylon for pure and Applied Sciences*, 27(2), 143-154.

Personalized Music Recommendation Based on Interest and Emotion: A Comparison of Multiple Algorithms

Xiuli Yan

College of Music and Dance, Fuyang Normal University, Fuyang, Anhui 236041, China

Abstract—Recommendation algorithms can greatly improve the efficiency of information retrieval for users. This article briefly introduced recommendation algorithms based on association rules and algorithms based on interest and emotion analysis. After crawling music and comment data from the NetEase Cloud platform, a simulation experiment was conducted. Firstly, the performance of the Back-Propagation Neural Network (BPNN) in the interest and emotion-based algorithm for recommending music was tested, and then the impact of the proportion of emotion weight between comments and music on the emotion analysis-based algorithm was tested. Finally, the three recommendation algorithms based on association rules, user ratings, and interest and emotion analysis were compared. The results showed that when the BPNN used the dominant interest and emotion and secondary interest and emotion as judgment criteria, the accuracy of interest and emotion recognition for music and comments was higher. When the proportion of interest and emotion weight between comments and music was 6:4, the interest and emotion analysis-based recommendation algorithm had the highest accuracy. The interest and emotion-based recommendation algorithm had higher recommendation accuracy than the association rule-based and user rating-based algorithms, and could provide users with more personalized and emotional music recommendations.

Keywords—Interest and emotion; recommendation algorithm; music; personalization

I. INTRODUCTION

With the prevalence of Internet technology and mobile devices, the amount of data and information that users can receive on the network has increased. However, low-quality or invalid information has also increased, making it difficult for Internet users to retrieve effective information. Information recommendation algorithms are a tool to assist in effective information retrieval [1]. When users are not clear about the information they want to retrieve, or only have a vague range, they can use information recommendation algorithms to initially screen information, and search for effective information from the information recommended by the algorithm [2]. Recommendation algorithms can be applied to music platforms for music recommendation. Music platforms can explore users' preferences based on their attributes, historical information, and other information, and provide personalized recommendation services to improve users' efficiency in finding music that fits their preferences [3]. In

music recommendation algorithms, the traditional method is to construct a feature of music recommendation [4], such as click-through rate, and recommend music with high click-through rates within a certain period to users. The principle of this type of music recommendation algorithm is relatively simple, and the basis for recommendation is usually a statistical feature. However, this type of recommendation is highly homogeneous for individual users, and the recommended lists for different users are basically the same. Moreover, statistical features reflect the overall preference trend of users. If the user base is large, there will always be many users whose music preferences do not conform to the mainstream [5]. Personalized recommendation algorithms delve into individual users' effective information to provide recommended music that is closer to individual preferences in terms of trend, although there may be some overlap with mainstream recommended music. It can provide different users with different recommended lists and avoid homogeneous recommended music [6]. In order to improve the accuracy and the degree of personalization of music recommendation algorithms, this paper conducted a study on a personalized music recommendation algorithm, where the personalized recommendation was based on the user's interest and emotion. The significance of this paper is to improve the accuracy of the music recommendation algorithm by recommending music to users through their interest tendencies for different music. The contribution of this paper is to provide a personalized recommendation list for users based on their interest tendency toward music, which provides an effective reference for personalized recommendation algorithms.

II. LITERATURE REVIEW

Relevant works are reviewed below. Shi [7] proposed a personalized music recommendation method and verified its feasibility. Wu et al. [8] designed a mixed music recommendation model using personalized measurement and game theory. They found that the model had higher accuracy in recommending new music, good dynamic personalized recommendation ability, and real-time recommendation ability. Gong et al. [9] constructed a deep music recommendation algorithm using dance motion analysis and evaluated it through quantification measures. They verified the validity of the algorithm.

III. MUSIC RECOMMENDATION ALGORITHM

A. Music Recommendation Algorithm Based on Association Rules

Registered users on music platforms not only leave a history of the music they listen to in their corresponding accounts, but they also add music they are interested in to their playlists while listening to music. Both the user's history and the records in their playlist can reflect the user's individual interests and preferences towards music [10]. Personalized recommendation algorithms usually mine user browsing records to construct personalized preference patterns for music, and then recommend music to users based on these patterns. The association rule method is a way to mine user preference patterns and can be used for personalized music recommendations [11].

Before using the music recommendation algorithm based on association rules for personalized music recommendations, the user's historical data for association rules should be mined. In music recommendation algorithms, association rules refer to the directional connections between different types of music. The mining steps can be summarized as searching for frequent item sets in the database and strong rules within the frequent item sets. The specific steps are as follows.

1) The user database is scanned, and each piece of music is treated as an item. It is assumed, there are a total of five pieces of music in the database, the candidate item sets obtained from scanning will contain five items. Moreover, the support of each item is computed [12]:

$$SUP = \frac{n}{N} \quad (1)$$

where SUP represents the support of the item, N represents the total number of historical records in the user database (the number of historical records equals the number of users), and n represents the number of historical records that contain the item.

2) The candidate item sets are pruned by removing items with support lower than a set threshold. The remaining set of items form the frequent item sets, which are recorded.

3) The items in the frequent item sets are combined pairwise to obtain new candidate item sets, and the support of each item in the new candidate item sets is computed.

4) Return to step 2 to obtain new frequent item sets, and record them.

5) Steps 2, 3, and 4 are repeated until a new set of candidate items cannot be generated. Then, the confidence level of all association item sets in the recorded frequent item sets is computed. Association item sets are non-empty proper subsets and the remaining element sets of each item set in the frequent item sets that may produce strong associations. The formula to calculate the confidence level of strong rule $X \Rightarrow Y$ [13] is:

$$CON = \frac{|X \cap Y|}{|X|} \quad (2)$$

where CON stands for the confidence level of the strong rule, $|X|$ indicates the number of records that contain the item, and $|X \cap Y|$ stands for the number of records that contain both items. When the confidence is over the set threshold, the association rule is regarded as strong. The recommendation algorithm recommends music based on the music items in the user's music record and the strong rules.

B. Sentiment Analysis-based Music Recommendation Algorithm

The recommendation algorithm based on association rules described earlier, mines data from the overall user history to obtain the association rules between music pieces. The music association rules obtained through data mining of user history indicate a connection between music pieces with the same "traits" [14]. The music pieces in the user's history reflect their interest in the "traits" of the music, and the music found through association rules will also contain the "traits" that the user is interested in. In most cases, this type of recommendation algorithm is sufficient to meet the needs of most users. However, users' moods can change during the process of listening to music [15]. Changes in mood can affect subjective feelings about music, making it difficult to satisfy user needs by recommending music based on rigid association rules.

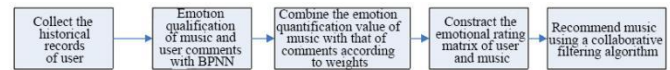


Fig. 1. Basic flow of the sentiment analysis-based music recommendation algorithm.

To enable music recommendation algorithms to adapt to users' changing emotional needs, this article incorporates sentiment analysis into music recommendation algorithms. Fig. 1 illustrates the basic process of the sentiment analysis-based music recommendation algorithm, and its detailed steps are as follows.

1) The music listening history of registered users on the music platform is collected, which includes the music data that the user listened to and the textual comments that the user made about the music.

2) The Back-Propagation Neural Network (BPNN) [16] is used to quantify the sentiment of music and each user's comment on the music, that is, to classify the sentiment of both. When training the BPNN to recognize the sentiment of music and comments separately, the result labels of the training samples are numerical values of the emotional categories. This article adopts the Hevner emotional circle model, which has eight emotional types. When recognizing the sentiment of music, the Mel Frequency Cepstral Coefficient (MFCC) features [17] of the music are input into the BPNN, which is calculated layer by layer in the hidden layer, and finally the numerical values of the primary and secondary emotions of the music are output in the output layer. When recognizing the sentiment of user comments, the comment text is preprocessed, including word segmentation and removal of redundant auxiliary words. Next, the Wordvec method [18] is used to

obtain the text vector of the comment, which is then input into the BPNN and calculated layer by layer in the hidden layer, and finally the numerical values of the primary and secondary emotions of the comment are output in the output layer.

3) The sentiment quantification values of the music are combined with the sentiment quantification values of the user's comment on the music according to a certain weight ratio to obtain the sentiment quantification value of the user for the music.

4) The sentiment quantification value of the user for the music obtained in the previous step is used to construct a user-music sentiment rating matrix, as displayed in Fig. 2. The size of the matrix is $m \times n$, each row indicates a user's sentiment rating for different music, each column indicates different users' sentiment rating for the same music, and element e_{ij} in the matrix represent the user's sentiment rating for the music.

$$\begin{bmatrix} e_{11} & e_{12} & \cdots & e_{1j} & \cdots & e_{1n} \\ e_{21} & e_{22} & \cdots & e_{2j} & \cdots & e_{2n} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ e_{i1} & e_{i2} & \cdots & e_{ij} & \cdots & e_{in} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ e_{m1} & e_{m2} & \cdots & e_{mj} & \cdots & e_{mn} \end{bmatrix}$$

Fig. 2. User-music sentiment rating matrix.

5) An item-based collaborative filtering algorithm [19] is used to recommend music, and the associated formula is:

$$\begin{cases} P(a,c) = sim(b,c) \times s(a,b) \\ sim(b,c) = \frac{\bar{b} \cdot \bar{c}}{\|\bar{b}\| \cdot \|\bar{c}\|} \end{cases} \quad (3)$$

where music b is the music that user a has listened to, $sim(b,c)$ indicates the cosine similarity between music and music, \bar{b} and \bar{c} are the sentiment rating vector of musics b and c , respectively (the dimension depends on the number of users), $s(a,b)$ is the sentiment rating of user a to music b , which can be checked from the sentiment rating matrix, $P(a,c)$ is the interest level of user a to music c . The top k musics in the nearest neighboring set that is most interesting to the target user are recommended to the target user.

IV. SIMULATION EXPERIMENT

A. Experimental Data and Settings

In this paper, a simulation experiment was conducted in MATLAB software [20] to evaluate a music recommendation algorithm based on sentiment analysis. The simulation experiment used data obtained through web crawlers on the NetEase Cloud Music platform. The data collected by the crawler includes music and comments from different users below the music interface. After initial data cleaning, a total of 200 pieces of music and 200 user comments for each music

were collected and analyzed, with some comments shown in Table I. The music codes in the table consist of letters and numbers, where the letters represent the main emotional tendency of the music and the numbers represent the music number. 60% of the data was used as a training set to train the BPNN in the music recommendation algorithm, and the remaining 40% was used for testing. In the recommendation algorithm, after the orthogonal experiment, the number of nearest neighbors was set to 30 when using collaborative filtering to generate the recommendation list, and the number of recommended music in the final recommendation list was set to 10 for each user.

TABLE I. SELECTED COMMENTS OF DIFFERENT USERS ON DIFFERENT MUSIC

	User 1	User 2
Music A1	This song makes me so sad to hear	Feeling sad
Music B1	This song has a wonderful sense of nobility	Sounds pretty sacred.
Music C1	Listen to the impulse to move forward	It will give people a sense of yearning
Music D1	There is a sense of quiet	There is a sense of calmness
Music E1	It's fun to listen to	Very cheerful
Music F1	A sense of relief	Feels very light and fast
Music G1	Can feel the enthusiasm	A sense of power
Music H1	Very life-giving	Vibrant feeling

B. Experimental Design

1) *Testing the BPNN in the recommendation algorithm:* After training two BPNNs using the testing set, the testing set was used for evaluation. There are two evaluation criteria for the testing plan. Evaluation criteria 1): only when the dominant emotion of the calculated sample is consistent with the dominant emotion in the test sample label, can it be considered a correct prediction; evaluation criteria 2): when either the dominant or secondary emotion of the calculated sample is consistent with the dominant or secondary emotion in the test sample label, it is considered a correct prediction.

2) *The influence of the weight ratio of sentiment quantification values between music and comments on the performance of the recommendation algorithm:* The weight ratios of sentiment quantification values between comments and music were set to 1:9, 2:8, 3:7, 4:6, 5:5, 6:4, 7:3, 8:2, 9:1. The accuracy of the recommendation algorithm under different weight ratios was tested. The weight ratios were set according to the size of the comment weight ratio.

3) *Comparison of three recommendation algorithms:* Association rules, user ratings, and sentiment analysis-based: To further prove the performance of the music recommendation algorithm proposed in this paper, it was compared with the association rules-based algorithm and the user ratings-based algorithm. The association rules-based music recommendation algorithm has been described above, while the user ratings-based music recommendation algorithm was the same as the sentiment analysis-based music recommendation algorithm in the basic steps, except that the elements in the rating matrix were no longer sentiment

quantification values but rather the number of clicks by the user on the music.

4) *The performance test of three recommendation algorithms in a real music platform:* In addition to the above simulation experiments, this paper also tested the performance of the three recommendation algorithms in a real music platform. Ten users of the NetEase cloud platform were randomly invited, and the three recommendation algorithms were used to recommend music to them. Each recommendation algorithm recommended music to each user ten times, and the users selected the music they were interested in from the results of each recommendation. The accuracy of each recommendation algorithm in recommending music of interest to the users was calculated.

C. Experimental Results

Fig. 3 illustrates the emotion recognition accuracy of the two BPNN algorithms in the recommendation algorithm based on emotion analysis under different evaluation criteria. The BPNN algorithm used for music emotion recognition had an identification accuracy of 42.6% under evaluation criteria (1) and 82.9% under evaluation criteria (2). The BPNN algorithm used for comment text emotion recognition had an identification accuracy of 44.3% under evaluation criteria (1) and 83.4% under evaluation criteria (2). Compared with evaluation criteria (1), evaluation criteria (2) had relatively loose requirements for the emotion recognition results. If evaluation criteria (1) was used, the algorithm’s performance was weakened due to the low accuracy. Therefore, in the subsequent experiments, evaluation criteria (2) was used to construct the rating matrix, where the dominant emotion and secondary emotion coexist.

Fig. 4 shows the impact of the weight proportion of sentiment quantification values of comments and music on the performance of the music recommendation algorithm based on sentiment analysis. As the weight proportion of the sentiment quantification values of comments gradually increased, the recommendation accuracy first rose and then declined, as can be seen from Fig. 4. When the weight proportion of sentiment between comments and music was 6:4, the recommendation accuracy of the algorithm was the highest, at 89.6%.

Due to space limitations, only partial results were shown here, as shown in Table II. First, the recommendation results of the three algorithms for the same user were compared, and it was found that there were differences in the recommendation results of the three algorithms for the same user. The association rule recommendation results had a variety of emotional categories for the recommended music, while the emotional categories of only one or two songs differed in the user rating recommendation results, and the emotional categories of the recommended music in the sentiment analysis recommendation results were basically the same. It was found from the comparison of the recommendation results obtained

by different users under the same recommendation algorithm that the association rule recommendation results had various emotional categories for the recommended music. In the user rating recommendation results, there was some overlap in emotional categories of recommended music among different users, and the emotional categories of recommended music in the sentiment analysis recommendation results were different among different users, but the emotional categories of recommended music for individual users were the same.

From Fig. 5, it was visually clear that the algorithm based on sentiment analysis had the highest recommendation accuracy, followed by the algorithm based on user ratings, and the lowest was the algorithm based on association rules.

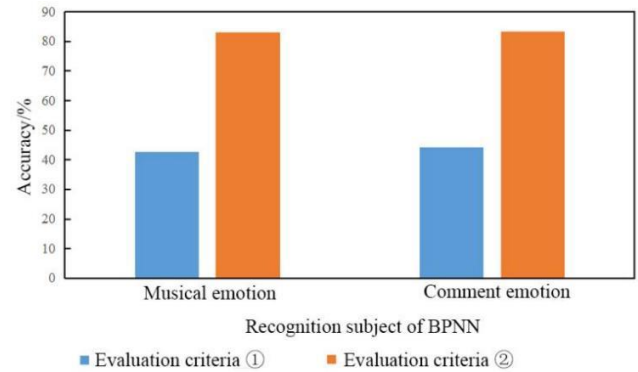


Fig. 3. Test results of the BPNN algorithm.

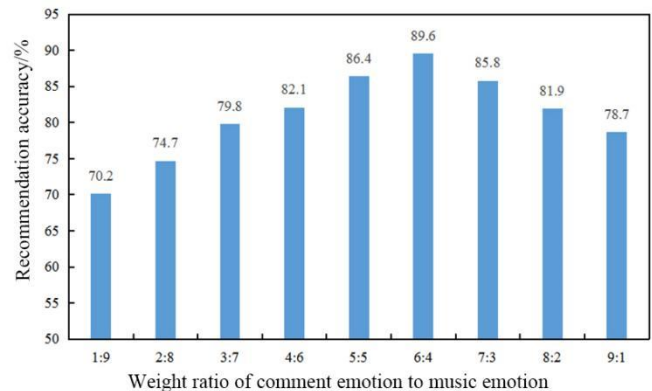


Fig. 4. Impact of the weight ratio of sentiment quantification values between music and comments on the performance of the recommendation algorithm.

TABLE II. SOME RECOMMENDATION RESULTS OF THE MUSIC RECOMMENDATION ALGORITHMS

User	Association Rules Recommendation Results	User Rating Recommendation Results	Emotional Analysis Recommendation Results
User 1	D2;F23;E24;A1; B25	D2;D23;D24;E21;H1	D23;D24;D23;D9;D8
User 2	F5;D9;G20;F12; G14	F5;D14;D23;D24;F12	G3;G5;G14;G11;G12
User 3	D1;H4;C3;B12;H1	A4;C3;D14;D23;D2	C1;C4;C3;C2;C6

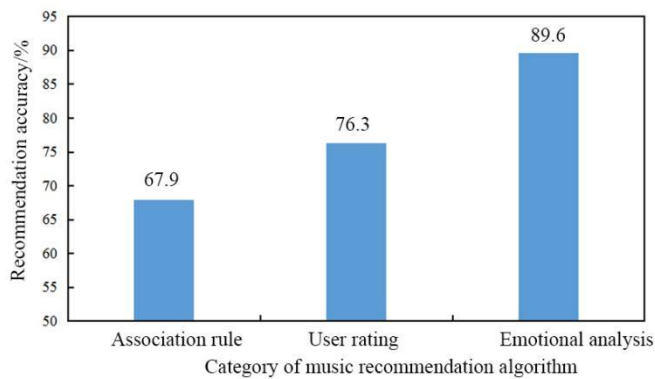


Fig. 5. Recommendation accuracy of three music recommendation algorithms.

In the actual NetEase cloud platform, the above three recommendation algorithms were used to make ten recommendations to each of the ten users, and the number of music items that users were interested in in the recommendation results is shown in Table III. The average number of music items that users were interested in was 66.8 for the association rule-based algorithm, 75.6 for the user rating-based algorithm, and 88.7 for the emotional analysis-based algorithm. It was seen from the comparison in Table III that the recommendation algorithm based on interest and emotional analysis recommended more music that users were interested in. At the same time, the ratio of the average number of songs to the total number of recommendations was very close to the accuracy of the recommendations calculated in the simulation experiment.

TABLE III. PERFORMANCE OF THREE RECOMMENDATION ALGORITHMS IN RECOMMENDING MUSIC OF INTEREST IN THE REAL MUSIC PLATFORMS

	The Number of Music of Interest in the Association Rule-based Recommendation Results/n	The Number of Music of Interest in user rating-based Recommendation Results/n	The Number of Music of Interest under Emotional Analysis-based Recommendation Results/n
User 1	67	76	89
User 2	66	75	88
User 3	65	75	88
User 4	66	76	87
User 5	67	77	88
User 6	67	77	89
User 7	66	75	89
User 8	67	75	89
User 9	68	76	90
User 10	69	74	90
Average	66.8	75.6	88.7

V. DISCUSSION

With the development of the Internet, people have access to more and more information. Recommendation algorithms can

help people find the information they need faster. This paper targeted a music platform and used recommendation algorithms to recommend music that may be of interest to platform users. This paper first introduced the association rule-based recommendation algorithm and the user-rating-based collaborative filtering recommendation algorithm and then introduced the BPNN, which can measure the user's interest and emotion, into the user-rating-based collaborative filtering algorithm in order to make the recommendation results more personalized, so that the algorithm can give the recommendation results based on the user's interest and emotion.

After that, the BPNN for measuring users' interest and emotion was tested first, and then the recommendation accuracy of the three recommendation algorithms was compared. The recommendation performance was tested in the actual NetEase cloud platform with ten randomly invited users. The final test results verified the accuracy of the BPNN for measuring users' interest and emotion, and the algorithm based on interest and emotional analysis had the highest recommendation accuracy among the three recommendation algorithms. The recommendation performance of this recommendation algorithm was also verified in the actual music platform. The reasons for why the algorithm based on interest and emotional analysis had the highest accuracy are as follows: The association rule-based recommendation algorithm mined the entire history of user records to obtain the association rule between music, assuming that users will select the next song according to the mined association rules after listening to a song, but the association rule mined from the entire data did not take into account the individual interests and hobbies of users. The algorithm based on user ratings used the number of times users clicked on music as an indicator of their interest, integrated users with similar interests, and provided personalized recommendations, resulting in higher recommendation accuracy. The recommendation algorithm based on sentiment analysis was based on the emotional tendencies of users and music, integrated users with similar emotional tendencies, and provided recommendations. Because it took into account the influence of emotions on music selection, it was more personalized, and the recommendation results were more accurate.

VI. CONCLUSION

This article briefly introduced the recommendation algorithms based on association rules and sentiment analysis. Then, through web scraping of music and comment data on the Netease Cloud platform, a simulation experiment was conducted. First, the performance of the BPNN in the sentiment analysis-based recommendation algorithm was tested. Then, the influence of the proportion of sentiment weight between comments and music on the algorithm based on sentiment analysis was tested. Finally, the algorithms based on association rules, user ratings, and sentiment analysis were compared. The results were as follows: 1) The accuracy of sentiment recognition for music and comments was higher when the BPNN used dominant and secondary emotions as judgment criteria. 2) As the proportion of sentiment quantification values between comments and music gradually rose, the recommendation accuracy of the algorithm suggested

an increasing trend followed by a decreasing trend. When the proportion of sentiment weight between comments and music was 6:4, the algorithm had the highest recommendation accuracy, at 89.6%. 3) The music recommended by the association rule recommendation algorithm had a diverse range of emotional categories. The emotional categories of the recommended music by the user rating recommendation algorithm were relatively consistent within a single user, and there was some overlap between different users' recommended music emotional categories. The emotional categories of the recommended music by the sentiment analysis recommendation algorithm were consistent within a single user, but different users' recommended music emotional categories were different. 4) The recommendation accuracy of the algorithm based on sentiment analysis was the highest, followed by the user rating-based algorithm, and the association rule-based algorithm had the lowest accuracy. 5) The interest and emotional analysis-based algorithm also had better recommendation performance in the actual music platform.

REFERENCES

- [1] J. R. Edwards, S. Borgstedt, and B. Barth, "New Music Recommendation Algorithm Facilitates Audio Branding," *Market. Rev. St Gallen*, vol. 4, pp. 888-894, January 2019.
- [2] S. Wang, M. Gong, Y. Wu, and M. Zhang, "Multi-objective optimization for location-based and preferences-aware recommendation - ScienceDirect," *Inform. Sciences*, vol. 513, pp. 614-626, November 2020.
- [3] A. Ramasuri, A. K. Badhan, B. Pushpa, A. J. Rani, A. S. Dharani, and M. Sindhuja, "Music recommendation system with advanced classification," *IJERT*, vol. 10, pp. 505-509, August 2021.
- [4] C. Wu, S. Liu, Z. Zeng, M. Chen, A. Alhudhaif, X. Tang, F. Alenezi, N. Alnaim, and X. Peng, "Knowledge graph-based multi-context-aware recommendation algorithm," *Inform. Sciences*, vol. 595, pp. 179-194, March 2022.
- [5] H. C. Yan, Z. R. Wang, J. Y. Niu, and T. Xue, "Application of covering rough granular computing model in collaborative filtering recommendation algorithm optimization," *Adv. Eng. Inform.*, vol. 51, pp. 1-10, January 2022.
- [6] B. Li, J. Li, and X. Ou, "Hybrid recommendation algorithm of cross-border e-commerce items based on artificial intelligence and multiview collaborative fusion," *Neural Comput. Appl.*, vol. 34, pp. 6753-6762, May 2022.
- [7] J. Shi, "Music recommendation algorithm based on multidimensional time-series model analysis," *Complexity*, vol. 2021, pp. 1-11, April 2021.
- [8] Y. Wu, J. Lin, and Y. Ma, "A hybrid music recommendation model based on personalized measurement and game theory," *Chinese J. Electron.*, vol. 32, pp. 1-10, October 2022.
- [9] W. Gong, and Q. Yu, "A deep music recommendation method based on human motion analysis," *IEEE Access*, vol. 9, pp. 26290-26300, February 2021.
- [10] D. Sun, "Using factor decomposition machine learning method to music recommendation," *Complexity*, vol. 2021, pp. 1-10, May 2021.
- [11] B. Sheng, and G. Sun, "Matrix factorization recommendation algorithm based on multiple social relationships," *Math. Probl. Eng.*, vol. 2021, pp. 1-8, February 2021.
- [12] Z. Yang, and J. Cheng, "Recommendation algorithm based on knowledge graph to propagate user preference," *Int. J. Comput. Int. Sys.*, pp. 1-25, January 2021.
- [13] M. F. Aljunid, and M. D. Huchaiah, "An efficient hybrid recommendation model based on collaborative filtering recommender systems," *CAAI Trans. Intell. Technol.*, vol. 6, pp. 1-13, May 2021.
- [14] Y. Gao, A. Xu, P. J. H. Hu, and T. H. Cheng, "Incorporating association rule networks in feature category-weighted naive Bayes model to support weaning decision making," *Decis. Support Syst.*, vol. 96, pp. 27-38, January 2017.
- [15] Y. Wu, J. Lin, and Y. Ma, "A hybrid music recommendation model based on personalized measurement and game theory," *Chinese J. Electron.*, vol. 32, pp. 1-10, October 2022.
- [16] S. Xiao, Y. Hu, J. Han, R. Zhou, and J. Wen, "Bayesian networks-based association rules and knowledge reuse in maintenance decision-making of industrial product-service systems," *Proc. CIRP*, vol. 47, pp. 198-203, December 2016.
- [17] W. Rao, L. Zhu, S. Pan, P. Yang, and J. Qiao, "Bayesian network and association rules-based transformer oil temperature prediction," *J. Phys. Conf.*, vol. 1314, pp. 1-8, November 2019.
- [18] M. R. Siddiquee, S. Rahman, S. U. I. Chowdhury, and M. R. Rahman, "Association rule mining and audio signal processing for music discovery and recommendation," *Int. J. Softw. Innov.*, vol. 4, pp. 71-87, April 2016.
- [19] F. F. dos Santos, M. A. Domingues, V. A. Sundermann, V. O. de Carvalho, M. F. Moura, and S. O. Rezende, "Latent association rule cluster based model to extract topics for classification and recommendation applications," *Expert Syst. Appl.*, vol. 112, pp. 34-60, June 2018.
- [20] T. Li, "Selection of audio materials in college music education courses based on hybrid recommendation algorithm and big data," *J. Phys. Conf. Ser.*, vol. 1774, pp. 1-8, January 2021.

Research on Customer Retention Prediction Model of VOD Platform Based on Machine Learning

Quansheng Zhao, Zhijie Zhao*, Liu Yang, Lan Hong, Wu Han

School of Computer and Information Engineering, Harbin University of Commerce, Harbin, Heilongjiang, 150028, China

Abstract—Advanced wireless technology and smart mobile devices allow users to watch Internet video from almost anywhere. The major VOD platforms are competing with each other for customers, slowly shifting from a "product-centric" strategic goal to a "customer-centric" one. At present, existing research is limited to platform business model and development strategy as well as user behavior research, but there is less research on customer retention prediction. In order to effectively solve the customer retention prediction problem, this study applies machine learning methods to video-on-demand platform customer retention prediction, improves the traditional RFM model to establish the RFLH theoretical model for video-on-demand platform customer retention prediction, and uses machine learning methods to predict the number of customer retention days. The Optuna algorithm is used to determine the model hyperparameters, and the SHAP framework is integrated to analyze the important factors affecting customer retention. The experimental results show that the comprehensive performance of the LightGBM model is better than other models. The total number of user logins in the past week, the length of video playback in the same day, and the time difference between the last login and the present are important features that affect customer retention prediction. This study can help companies develop effective customer management strategies to maximize potential customer acquisition and existing customer retention for maximum market advantage.

Keywords—Video-on-demand platform; Customer Retention Forecast; RFM Model; Machine Learning; SHAP

I. INTRODUCTION

With the advent of the information age, market competition has become more and more intense, and enterprises have slowly changed from a "product-centric" strategic goal to a "customer-centric" one. In the case of limited resources, major e-commerce companies compete with each other for customers, the pursuit of a larger share of the market, to maximize their profits has become the development of each enterprise imperative.

Video-on-demand is developed in the 1990s, and is called "Video On Demand" in English, so it is also called "VOD". As the name suggests, it is a video-on-demand system that plays programs on demand by viewers. Watching video has become one of the most popular online activities and provides a huge market for various video content providers. Better networks, technological innovations and the availability of smart devices have changed the way people are entertained, allowing platforms to deliver services to viewers directly online via the Internet [1]. For any service to grow, it is crucial to understand the values and consumption habits of consumers. In real

systems, there are users who stay for a long time and others who enter the system and leave soon after. Therefore, understanding user behavior and further predicting customer retention on VOD platforms can help platforms improve service quality and avoid customer churn. For users with short retention time, the platform can adopt ways such as trial membership and issuing coupons to let users experience the core services to retain them; while for users with longer retention time, the platform can appropriately provide more rewards or service updates to extend their retention time [2].

In the study for video-on-demand platforms, Zhang et al. take Tencent's over-the-top on-demand rule as an example to analyze the reasons for the model, summarize and analyze the problems that arise when current video-on-demand platforms innovate their profit models, and propose strategies to optimize their profit models and corresponding strategies [3]. Hu et al. studied user behavior and access patterns through a major ISP in Shanghai, China, focusing on comparing behavior and access patterns across platforms, and found that user migration across multiple platforms was common and highly influenced by the different characteristics of the platforms [4]. Köster et al. explore the relationship between social referrals, referral propensity, and the stickiness of video-on-demand websites, comparing consumers who are referred by social networks with consumers who arrived at the site through natural search or social media ads to understand the stickiness of the site. The results showed that consumers who were recommended via social referrals spent more time on the site, viewed more pages, and launched more videos than consumers who responded to social media ads, but less than consumers who went through natural search [5]. Wang et al. analyzed the differences between Chinese video platforms and U.S. video platforms, reflecting on the relationship between platform, market, and country. The historical factors and geographical characteristics that influence the operation, structure and governance of video platforms are explored [6]. Rahman et al. studied video consumer behavior from the perspective of VOD platforms. The analysis of the data revealed interesting features of video-on-demand platforms, such as the viewing habits and viewing patterns of different users and the correlation between user profile information (e.g., age, gender) and viewing habits, among others, and suggested that future user behavior could be predicted by learning from previous user behavior patterns [7].

The existing research by scholars on video-on-demand platforms is limited to two major aspects of platform business models and development strategies [6] and user behavior [7] studies, with few studies focusing on their customer retention predictions. To solve this problem, this paper proposes an

improved RFM theoretical model based on user behavior and constructs a VOD customer retention prediction model based on machine learning to exploit the advantages of machine learning algorithms in prediction. Recent studies have shown that LightGBM, XGBoost, and CatBoost can show better performance in the face of complex, highly nonlinearly related data [8][9], and the Ensemble Learning (EL) algorithm integrates multiple machine learning models to form a model with stronger generalization and better backfitting capabilities[10]. Therefore, this paper uses LightGBM, XGBoost, CatBoost and other algorithms to construct customer retention prediction models for video-on-demand platforms. After the models are constructed, the accuracy of different models in this study scenario is compared by time complexity, R^2 , MAE, RMSE indexes and the optimal prediction model is selected, and the prediction results are later analyzed using the SHAP interpretation method to explore the magnitude of the impact of different characteristics of users on the prediction values and the reasons behind them. The study found that a user's total number of logins in the last week, the length of the day's video playback and the time difference between the most recent login and the present were important features that influenced customer retention predictions.

II. RELATED WORKS

In the Internet industry, users who start using an application at a certain time and continue to use the application after a certain period of time are recognized as retained users. Customer retention measurement and analysis is a classic problem in various domains, such as online community platforms [11], telecommunication industry [12] and e-commerce [13]. Several works focus on the measurement and analysis of characteristics related to customer retention. Jiang et al. proposed a maximum entropy semi-Markov model to predict the customer life stages that need to be segmented for milk powder products, which is applicable to the case where the infant life stage transition is deterministic, but not to the case where there is no explicit life stage, such as video-on-demand systems[14].

In recent years, RFM models have also been widely used in customer behavior prediction studies. Marín et al. modeled user behavior based on the traditional proximity, frequency and currency (RFM) model to obtain a proximity, frequency, importance and duration (RFID) model of customer assessment from the perspective of customer-contact center interactions, and showed that the model can be generalized to any environment requiring classification or regression algorithms in any environment that requires classification or regression algorithms [15]. Perišić et al. proposed an extended framework of new proximity, frequency, and monetary value features for predicting user churn in the mobile gaming domain by combining features related to user lifecycle, intensity, and rewards, and indicated that the top five most important features of a multivariate churn prediction model include long-term and short-term frequency features, monetary, intensity, and lifecycle features[16]. Wei et al. first established an RFLP metric system for predicting MOOC user learning behavior and attrition by improving the RFM model in the business domain; secondly, histogram and chi-square tests were used to determine the characteristic variables affecting MOOC user

attrition; finally, a MOOC user attrition prediction model was constructed by combining the Grouped Data Processing (GMDH) network as a post-processing information system [17]. Smali et al. modified the model by adding diversity "D" as the fourth parameter, referring to the diversity of products purchased by a given customer, and the RFM-D based model was applied to the retail market to detect customer behavior patterns and the proposed model improved the quality of customer behavior prediction [18].

In order to effectively solve the customer retention prediction problem, this paper firstly improves the traditional RFM model in terms of extracting features and proposes the RFLH theoretical model suitable for customer retention prediction of VOD platform. After that, machine learning and other latest technologies are applied to the VOD customer retention prediction problem, and the advantages of machine learning algorithms in prediction are exploited to learn users' previous behavior patterns to predict their future behavior. And Optuna framework is used to optimize the model hyperparameters to improve the model prediction performance. Finally, the SHAP interpretation framework is combined with the interpretation analysis of different user behaviors in order to propose targeted suggestions and strategies for customer management in VOD platform.

III. BUILDING A THEORETICAL MODEL OF CUSTOMER RETENTION BASED ON IMPROVED RFM

RFM model is one of the most important methods to analyze the behavioral characteristics of customers, classifying them by three behavioral variables: proximity R (Recency), frequency F (Frequency), and value M (Monetary) [19]. The traditional RFM model predicts the future short-term behavior of customers through their past purchase behavior in e-commerce platforms. In the existing research, this model is mostly used for the comprehensive consideration of user activity, loyalty and consumption ability to further achieve user value identification and value group segmentation. It is generally believed that users with a shorter interval between recent consumptions and a larger number and amounts of recent consumptions have a higher recognition of products and services and therefore a lower tendency to churn; conversely, users with a longer interval between recent consumptions and a smaller frequency and amount of recent consumptions have a higher tendency to churn and a lower value to the platform. vod platform is different from ordinary e-commerce platforms, namely: most vod users in the viewing process does not generate actual consumption, but the two have a certain degree of relevance in some way.

Currently, the RFM model has been used in user churn prediction [13][20]. In this study, the RFM model is improved on the basis of RFM model to construct RFLH indicator system for vod user behavior and customer retention prediction, and the details are shown in Table I. In RFLH indicator system the indicator H(History) represents the number of days of historical retention, including the statistical characteristics of the number of days of historical retention in the last month, which is a more critical characteristic variable for accurate prediction of customer retention days. The improved RFM model is shown in Table I.

TABLE I. IMPROVED RFM THEORETICAL MODEL

Classification index	Indicator meaning
R(Recency of viewing)	The time interval between the viewer's most recent login and the observation point
F(Frequency of viewing)	The number of times the viewer logged in during a certain period
L (Length of viewing)	Viewer's viewing time and completion in a certain period
H (Days of historical retention)	The number of days the viewer retention in a certain period

IV. MACHINE LEARNING-BASED CUSTOMER RETENTION PREDICTION MODEL CONSTRUCTION

A. Machine Learning Algorithm Selection

In order to select the most suitable machine learning algorithm for customer retention prediction on Vod platform, this study constructs regression models based on three machine learning algorithms, LightGBM, Catboost and XGBoost, respectively, for comparison experiments, and selects the model with the best overall performance.

LightGBM is a distributed gradient boosting framework based on decision tree algorithm. Designed to provide a fast, efficient, low-memory, high-accuracy tool that supports parallelism and large-scale data processing, LightGBM achieves linear acceleration in data computation by reducing the memory use of data, reducing communication costs, and improving efficiency when multiple machines are parallel. Its advantages are faster training efficiency, low memory usage, and higher accuracy. The disadvantage is that it may grow deeper decision trees and produce overfitting[21].

CatBoost is a library of gradient boosting algorithms, which has the advantages of overcoming the gradient bias and effectively solving the problem of prediction bias, improving the accuracy of the algorithm, enhancing the generalization ability, and preventing the occurrence of overfitting phenomenon. Its disadvantages are that it requires a lot of memory and time for the processing of category features, and the setting of different random numbers has an impact on the model prediction results [22].

XGBoost, an efficient Gradient Boosting algorithm, integrates the idea of iteratively generating multiple weak learners and then adding up the prediction results of each learner to get the final prediction result, which has a better performance in structured data processing. The advantage is that the regularization term is added to prevent overfitting and parallel optimization is possible to improve the efficiency of the algorithm. The disadvantages are that it has too many parameters, the tuning parameters are too complex and it is only suitable for processing structured data [23].

LightGBM, CatBoost, and XGBoost are all Boosting algorithms in integrated learning and Boosting algorithms are widely used in the industry and can show better performance in the face of complex, highly nonlinear data [9], which is applicable to the data in this paper.

B. Optuna Tuning Framework

In order to optimize the performance of individual machine learning models and to make them comparable with each other.

To set the hyperparameters of the models, the methods usually used are "grid search" or "random search". However, the "grid search" method requires more computational power and time due to the larger parameter space. The advantage of random search is that the search is fast, but it is easy to miss some important information and difficult to determine the distribution of the parameters. Therefore, for hyperparameter optimization, a software framework called Optuna automatic hyperparameter optimization is used in this experiment.

Optuna is a framework designed for automated and accelerated research, It has three advantages: (1) define-by-run API that allows users to construct the parameter search space dynamically, (2) efficient implementation of both searching and pruning strategies, and (3) easy-to-setup, versatile architecture that can be deployed for various purposes, ranging from scalable distributed computing to light-weight experiment conducted via interactive interface [24]. The operation procedure is as follows:

- 1) Determine the optimization direction, parameter type, value range and maximum number of iterations.
- 2) Enter the cycle.
- 3) Select a group of individuals evenly within the function that defines the range of parameter values.
- 4) The trimmer automatically terminates hopeless populations according to pruning conditions.
- 5) Calculate the individual objective function value of the unpruned population.
- 6) Repeat the above steps until the maximum number of iterations are reached and out of the loop.
- 7) Optimal solution of the output problem [25].

C. Model Performance Evaluation Metrics

In this study, the time complexity is the sum of training time and prediction time, and the higher time complexity will affect its practical application to some extent. A lower time complexity helps to accomplish fast and effective prediction [26]. Meanwhile, the model is evaluated by using three indicators: mean absolute error MAE, root mean square error RMSE and coefficient of determination R², and the evaluation indicators are calculated as follows:

$$MAE = \frac{1}{n} \sum_{t=1}^n |y_{rt} - y_{pt}| \quad (1)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (y_{rt} - y_{pt})^2} \quad (2)$$

$$R^2 = \frac{|\sum_{t=1}^n (y_{rt} - \bar{y}_r) \cdot (y_{pt} - \bar{y}_p)|^2}{\sum_{t=1}^n (y_{rt} - \bar{y}_r)^2 \cdot \sum_{t=1}^n (y_{pt} - \bar{y}_p)^2} \quad (3)$$

Where: n is the number of data; y_p is the predicted result; y_r is the true value; \bar{y}_p and \bar{y}_r are the average of the predicted and true results, respectively. In the regression prediction, the smaller the RMSE value and the closer the R² value is to 1, the higher the interpretability of the model. In this paper, the MAE, RMSE and R² are used to make a comprehensive comparison of the model prediction results and verify the prediction accuracy of the model.

V. EMPIRICAL ANALYSIS OF VOD CUSTOMER RETENTION PREDICTIONS

In this paper, the predicted customer retention is defined as the number of days a user logs in the next seven days. For example, if a user's prediction result on January 1 is equal to 3, it means that this user will visit the VOD platform for 3 days in the next 7 days (January 2~8). In this paper, we first pre-process the data to improve the data quality. The features are extracted according to the improved RFM model and combined with machine learning methods to build a video-on-demand platform customer retention prediction model. The specific process is shown in Fig. 1.

A. Data Processing and Feature Selection

iQIYI is China's and the world's leading high-quality video entertainment streaming platform, with more than 500 million users enjoying entertainment services on iQIYI every month. In this paper, we use the dataset provided by iQIYI AI competition platform, which contains video data, user personal information, user startup logs, user viewing and interaction behavior logs, etc. The detailed fields have various information such as the user's device type, device storage, device running memory, gender, age, education status, and occupational status.

The main natural attribute variables of Vod users are set in the paper, including gender, age group, education level, occupational status, and device information. The viewing behavior characteristics variables are classified based on indicators R, F, L, and H. The data of 15,000 users were randomly selected as samples. Based on the behavior records, the attribute variables and viewing behavior characteristic variables of each user were extracted separately. Each indicator

was divided by the research as shown in the Table II, and a regression model was constructed to predict the number of days users would log in in the next seven days.

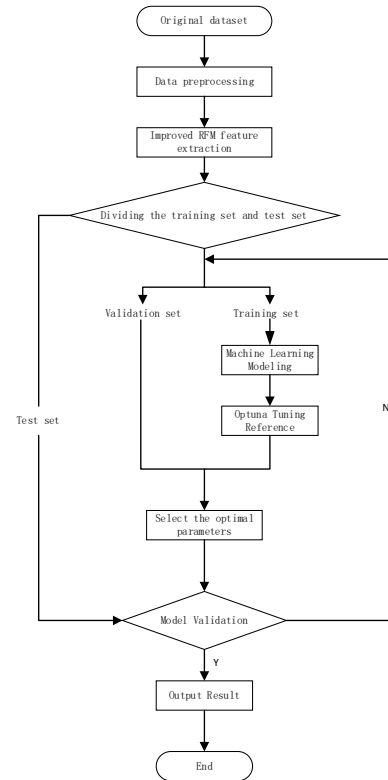


Fig. 1. Flow chart of customer retention prediction model for VOD platform.

TABLE II. IMPROVED RFM MODEL

Variable Type	Indicators	Meaning	
Natural Property Variables	X Properties	X1	Sex
		X2	Age
		X3	Education
		X4	Occupational Status
		X5	Device ram
		X6	Device rom
		X7	Device type
Watch behavioral characteristics variables	R(Recency of viewing)	R1	Login or not for the day
		R2	Time difference between last login and current
	F(Frequency of viewing)	F1	Total number of logins in a month
		F2	Total number of logins in the last week
	L (Length of viewing)	L1	Video playback duration for that day and each of the previous seven days
		L2	Number of video plays per day for the day and the first seven days
		L3	Video completion of the day and each day of the previous seven days
	H (Days of historical retention)	H1	Median number of historical retention days in a month
		H2	Average of the previous four weeks of historical retention days
H3		Weighted average of the previous four weeks' historical retention days	

B. Model Training

After the new dataset was constructed with a total of 953112 data, the missing values outliers in the new dataset were processed, such as filling the missing values in the data with 0, and removing the ones with more than 24h of playing time in a day. The processed data were tested for correlations of the feature variables by Pearson, Spearman, and Kendall three-class correlation coefficients, as shown in Fig. 2, 3, and 4. The test results showed that the selected feature variables were statistically correlated with the target variables.

After data processing, the last log-in time point of 15,000 users recorded in the data was subtracted by seven days as the test set, 80% of the remaining data was used as the training set, and 20% was used as the validation set to input three machine learning algorithms, LightGBM, CatBoost, and XGBoost, respectively, for model training.

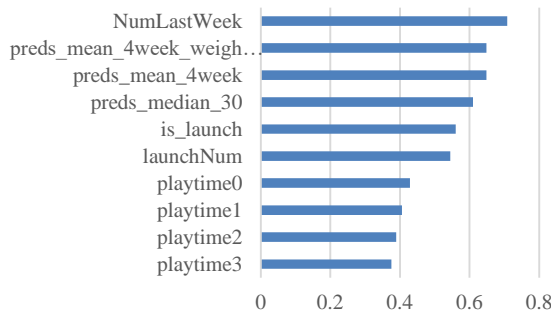


Fig. 2. Pearson correlation test graph (partial).

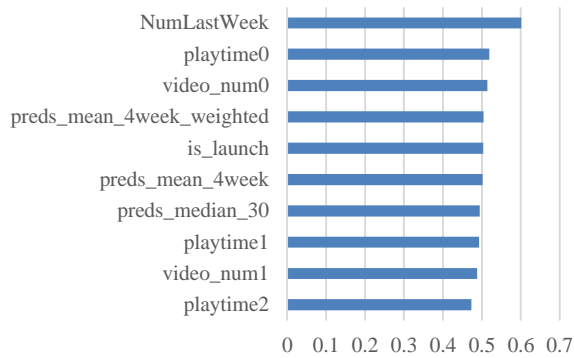


Fig. 3. Spearman correlation test graph (partial).

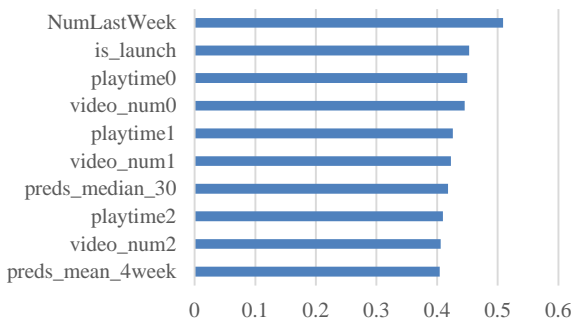


Fig. 4. Kendall correlation test graph (partial).

C. Hyperparameter Tuning

Hyperparameters are the framework parameters of machine learning models, and machine learning algorithms contain various hyperparameters that can be set to improve the accuracy of the model. To set a suitable set of hyperparameters, a hyperparameter tuning approach can be used, which objectively scans various attributes of the model hyperparameters and selects an optimal subset to improve the performance and effectiveness of learning, resulting in a model with optimal performance on a given dataset. For the machine learning-based customer retention prediction model, the Optuna optimization framework is used to tune the main hyperparameters of the prediction model, and the main hyperparameters obtained according to the Optuna optimization framework are shown in Table III.

TABLE III. MODEL'S MAIN HYPERPARAMETERS

Machine learning models	Hyperparameter Name	Optimal hyperparameter values	Meaning and Impact
LightGBM	objective	regression	Assign learning tasks and corresponding learning objectives.
	max_depth	8	The maximum depth of the tree model. The most important parameter for preventing overfitting, decisive for model performance and generalization ability
	learning_rate	0.01	shrinkage rate, choosing a relatively small learning rate can achieve stable and better model performance
	n_estimators	724	The number of iterations of boosting, generally speaking, the more iterations the better the model performance, but too many iterations will often lead to overfitting of the model and affect the training time of the model
	lambda_l1	8	L1 regularization parameter, used to adjust the control overfit
	lambda_l2	7	L2 regularization parameter, used to adjust the control overfit
	num_leaves	1231	Number of leaves on a tree, the larger the value of

			the maximum number of leaf nodes, the more accurate the model, but too large may be over-fitted.
	subsample	0.8	The sampling ratio of training samples can be used to accelerate training and handle overfitting.
	max_bin	247	max_bin indicates the maximum number of bins for the features. A smaller max_bin makes training faster, a larger max_bin makes the model more accurate, but too large a max_bin can lead to overfitting.
CatBoost	depth	5	The maximum depth of the tree model.
	learning_rate	0.01	Used for reducing the gradient step
	n_estimators	5526	The maximum number of trees that can be built when solving machine learning problems.
	l2_leaf_reg	9	Coefficient at the L2 regularization term of the cost function.
	max_bin	209	The number of splits for numerical features.
	bootstrap_type	Bernoulli	Defines the method for sampling the weights of objects
	subsample	0.78	Sample rate for bagging
XGBoost	max_depth	5	Maximum depth of a tree. Increasing this value will make the model more complex and more likely to overfit.
	learning_rate	0.01	Step size shrinkage used in update to prevents overfitting.
	n_estimators	2542	n_estimators indicates the number of integrated weak evaluators. The larger the n_estimators, the better the learning ability of the model.

	alpha	7	L1 regularization term on weights. Increasing this value will make model more conservative.
	gamma	2	Minimum loss reduction required to make a further partition on a leaf node of the tree. The larger is, the more conservative the algorithm will be.
	subsample	0.7	Subsample ratio of the training instances.

D. Analysis of Experimental Results

The trained model is tested with a test set, and the model is evaluated by time complexity, mean absolute error (MAE), root mean square error (RMSE), and coefficient of determination (R²). The prediction results of the different models were compared and the results are shown in Table IV. The comparison of training time and prediction time and time complexity of different algorithms are shown in Fig. 5, 6 and 7.

TABLE IV. COMPARISON OF MODEL PREDICTION RESULTS

Machine learning models	MAE	RMSE	R ²
LightGBM	0.9910	1.3708	0.6174
XGBoost	0.9896	1.3710	0.6173
CatBoost	0.9882	1.3712	0.6172

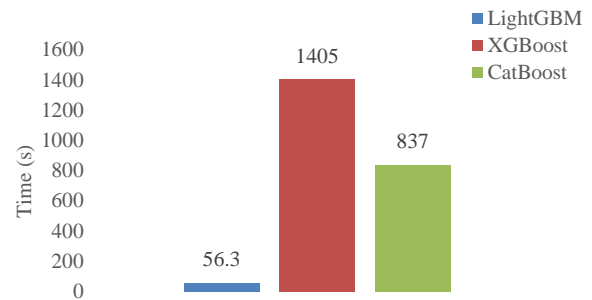


Fig. 5. Comparison of model training time.

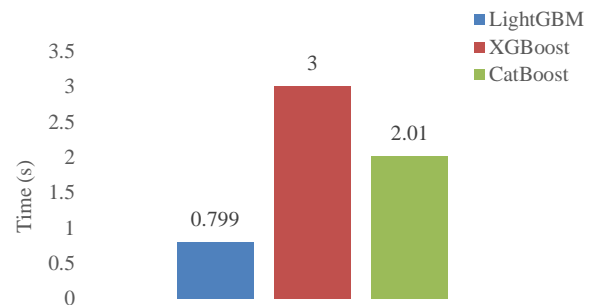


Fig. 6. Comparison of model prediction time.

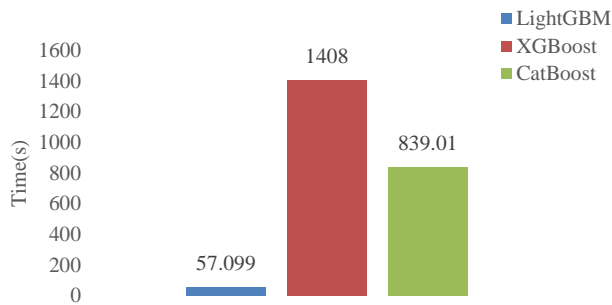


Fig. 7. Comparison of time complexity.

The experimental results show that all three tree-based machine learning models using Optuna can obtain MAE values less than 1, which means that the mean value of model prediction error is within 1 day, with CatBoost having the smallest MAE value of 0.9882. The RMSE metric is strongly influenced by outliers, and in this study the RMSE of all three tree models is around 1.371, with LightGBM has the smallest RMSE value of 1.3708. R^2 is a measure of the ability of the independent variable to explain the dependent variable. When the R^2 value is close to 1, the explanatory power of the independent variable is at a high level. Among all models, the strongest explanatory power is LightGBM, followed by XGBoost and CatBoost. For the prediction models constructed in this study, the prediction results of the three models are not very different and all can accurately predict the number of customer login days in the next seven days, but in terms of time complexity, LightGBM is much more efficient than the other two models. In summary, LightGBM was selected as the customer retention prediction model for the video-on-demand platform and further analyzed in this study.

VI. SHAP-BASED MODEL INTERPRETATION ANALYSIS

SHAP (SHapley Additive ExPlanations) is a method for explaining the predictions of machine learning models. It is an additive explanatory model constructed by Lundberg et al [27] in 2017 inspired by cooperative game theory, which provides a unified way to understand how each feature contributes to the prediction and how the combination of features determines the final prediction. SHAP values are based on the concept of Shapley values in cooperative game theory and are used to fairly distribute the value generated by a set of individuals to each Individuals. Similarly, SHAP values distribute the predictions of a machine learning model to each feature, allowing us to see how each feature contributes to the prediction. SHAP values take into account feature interactions and the relationship between features and predictions, providing a more complete and accurate interpretation than other model interpretation techniques such as partial dependency graphs and feature importance values. SHAP interprets the predicted value of a model as the attributed value of each input feature SHAP interprets the model prediction as the sum of the attributed values (Shap Values) of each input feature, and the Shap Values are calculated as follows:

$$y_i = y_{base} + f(x_{i1}) + f(x_{i2}) + \dots + f(x_{in}) \quad (4)$$

Suppose the i -th sample is x_i and there are n features, the baseline of the whole model is the mean of all sample target variables, defined as y_{base} , where $f(x_{i1})$ is the Shapley value of the first feature of sample i , when $f(x_{i1})$ is greater than 0, this feature has a positive effect on the predicted value; when $f(x_{i1})$ is less than 0, this feature has a negative effect on the predicted value. In addition, the TreeSHAP framework developed by Lundberg specifically for tree models has greatly improved the operation efficiency [28].

SHAP Summary Plot combines feature importance with feature effect to reflect the overall positive and negative relationship between feature value and future login days of customers [29]. In order to further clarify the positive/negative relationship between each indicator and the target variable, this paper uses Tree SHAP for model interpretation. As shown in Fig. 8, the horizontal coordinate in the figure is SHAP value, each row represents a feature variable, and each point represents a sample. The redder the point color is, the larger the value of the feature itself, and the bluer the color is, the smaller the value of the feature itself. A positive value represents a positive impact, while a negative value represents a negative impact.

Fig. 8 shows the summary of SHAP features of LightGBM. From the figure, we can see that the three most important features in the model are the total number of logins in the recent week, the length of video playback on the same day, and the time difference between the most recent logins and the present. For the R(recency) feature, the customers who log in on the current day have a positive impact on the number of login days in the future, the greater the time difference between the latest login and the present, the greater the negative impact on the future login days; For the F(frequency) feature, the more the total number of logins in the past week and the total number of historical logins, the more accustomed users are to using the platform, and the higher the positive impact on the future login days of customers; For the L(length) feature, the more playback duration and number of the current day and the previous seven days, the higher the positive impact on the future login days of customers; while the higher the video completion degree of the current day and the previous seven days, the higher the negative impact on the future login days of customers; For the H(history) feature, the constructed historical retention days feature reflects the login characteristics of users to some extent. The mean, median and weighted mean of historical retention days all have an impact on the future login days of customers. The larger the mean and weighted mean of historical retention days, the higher the positive impact on the future login days of customers.

As for the differences in personal characteristics, the overall effect on the predicted results is little. As shown in the figure, customers with jobs are more likely to retain than those without jobs. The device type has no or positive impact on the future login days; the RAM and ROM of the device will have an impact on the performance of the device. In this experiment, the larger the ram and rom values are, the higher the positive impact on the future login days of the customers; the gender has a mixed impact on the prediction results; the younger the age is, the higher the positive impact on the future login days of the customers.

VII. CONCLUSION AND OUTLOOK

The rapid development of information technology has made watching videos one of the most popular online activities and provides a huge market for various video content providers. vod customer retention prediction model can accurately predict the number of days users will log in in the future, which is crucial for vod platform to retain customers and increase its core competitiveness in the market. This study proposes an improved RFLH customer retention prediction model based on the traditional RFM model based on a machine learning customer retention prediction research model, and empirically tests the model with real data from the iQIYI platform, constructs the prediction model by three different machine learning algorithms, and applies Optuna to select the optimal hyperparameters of the model to improve the prediction accuracy of the model. The experimental results show that the Vod customer retention prediction model based on machine learning proposed in this study is effective, and LightGBM is the most efficient compared with the other two models with similar accuracy and has high practicality. Finally, the main factors of different features affecting customer retention are analyzed in conjunction with the SHAP model, and the feature variables that have a greater impact on the prediction model are found, and the overall positive and negative relationships between the feature variables and customer retention are carved. This study remedies the shortage of customer retention prediction research in video-on-demand platforms to a certain extent. Meanwhile, the RFLH index system and SHAP explanatory model in the paper can provide better decision support for Vod platform to understand the real situation of users and thus retain them. In this study, the analysis of personal characteristics is derived from the output of the model. In the future, personal characteristics and behaviors can be further combined to explore the influence of users' personal characteristics on customer retention, and more factors affecting customer retention in video-on-demand platforms can be introduced into the prediction model, as well as other prediction techniques can be applied to further improve the accuracy of prediction.

ACKNOWLEDGMENT

This work was supported by the 2021 Harbin University of Commerce Teacher "Innovation" project support project.

REFERENCES

- [1] Gupta, G., & Singharia, K. (2021). Consumption of OTT media streaming in COVID-19 lockdown: Insights from PLS analysis. *Vision*, 25(1), 36-46.
- [2] Wang Y, Guo Y, Chen Y. Accurate and early prediction of user lifespan in an online video-on-demand system[C]//2016 IEEE 13th International Conference on Signal Processing (ICSP). IEEE, 2016: 969-974.
- [3] X. Q. Zhang, & Wang X. (2022). Reflection on the profit model innovation of network video platform -- Taking Tencent's advance on demand adding rules as an example. *Youth Journalist* (4), 2.
- [4] Yan, H., Lin, T. H., Gao, C., Li, Y., & Jin, D. (2018). On the understanding of video streaming viewing behaviors across different content providers. *IEEE Transactions on Network and Service Management*, 15(1), 444-457.
- [5] Köster, A., Matt, C., & Hess, T. (2021). Do all roads lead to Rome? Exploring the relationship between social referrals, referral propensity and stickiness to video-on-demand websites. *Business & Information Systems Engineering*, 63, 349-366.

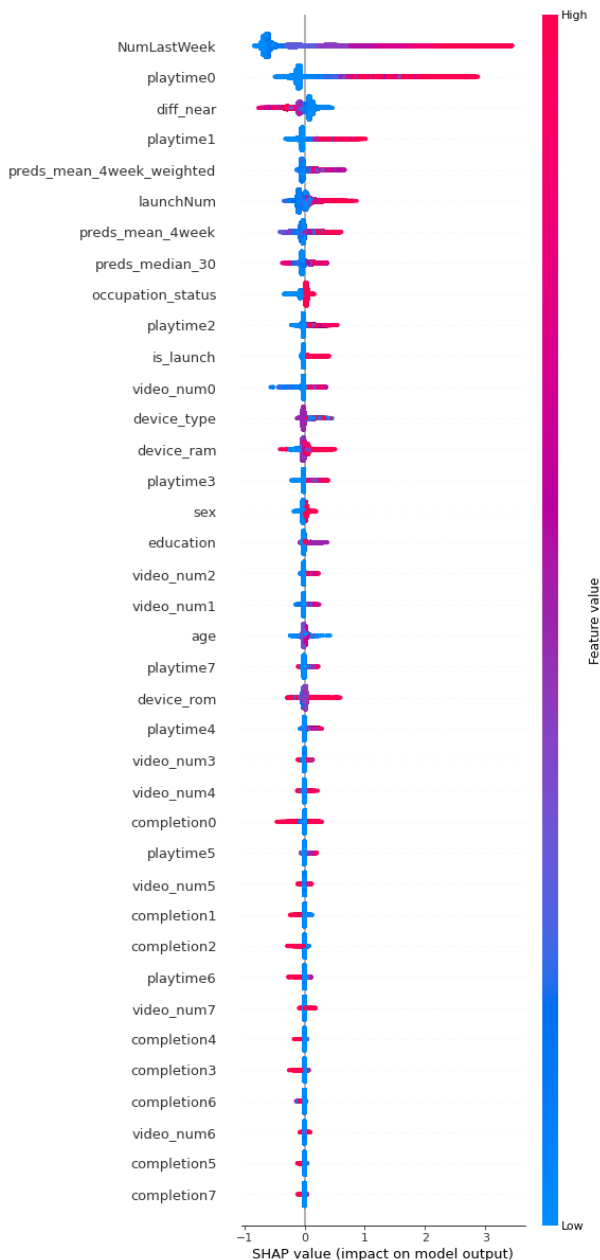


Fig. 8. LightGBM SHAP feature summary chart.

For companies, for users whose model predicts 0 logins and have not logged in for a long time, reduce the investment in this area in the future marketing process. For customers who are predicted to log in in the future, the focus should be on the total number of logins in the past week, the length of video playback on the same day and the time difference between the last login and the present. Companies can establish a membership system as well as a point system to obtain points through sign-ups to send members, sign-ups to receive points as well as accumulate viewing hours and complete tasks, and transform them into premium privileges that can be enjoyed to further strengthen user stickiness.

- [6] Wang, W. Y., & Lobato, R. (2019). Chinese video streaming services in the context of global platform studies. *Chinese Journal of Communication*, 12(3), 356-371.
- [7] Rahman, S., Mun, H., Lee, H., Lee, Y., Tornatore, M., & Mukherjee, B. (2018, October). Insights from analysis of video streaming data to improve resource management. In *2018 IEEE 7th International Conference on Cloud Networking (CloudNet)* (pp. 1-3). Ieee.
- [8] Zhou, Y., Liu, Y., Wang, D., & Liu, X. (2021). Comparison of machine-learning models for predicting short-term building heating load using operational parameters. *Energy and Buildings*, 253, 111505.
- [9] X. Z. Wen & Z. M. Ren.(2023) Predicting the Order Delivery Time of E-commerce Platform Based on the Temporal and Spatial Features of Regional Distribution Center. *Operations Research and Management Science.*, in press.
- [10] Fu, B., He, X., Yao, H., Liang, Y., Deng, T., He, H., ... & He, W. (2022). Comparison of RFE-DL and stacking ensemble learning algorithms for classifying mangrove species on UAV multispectral images. *International Journal of Applied Earth Observation and Geoinformation*, 112, 102890.
- [11] Lamrhari, S., El Ghazi, H., Oubrich, M., & El Faker, A. (2022). A social CRM analytic framework for improving customer retention, acquisition, and conversion. *Technological Forecasting and Social Change*, 174, 121275.
- [12] Sudharsan, R., & Ganesh, E. N. (2022). A Swish RNN based customer churn prediction for the telecom industry with a novel feature selection strategy. *Connection Science*, 34(1), 1855-1876.
- [13] Xiahou, X., & Harada, Y. (2022). B2C E-commerce customer churn prediction based on K-means and SVM. *Journal of Theoretical and Applied Electronic Commerce Research*, 17(2), 458-475.
- [14] Jiang, P., Zhu, Y., Zhang, Y., & Yuan, Q. (2015, August). Life-stage prediction for product recommendation in e-commerce. In *Proceedings of the 21th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 1879-1888).
- [15] Marín Díaz, G., Galán, J. J., & Carrasco, R. A. (2022). XAI for Churn Prediction in B2B Models: A Use Case in an Enterprise Software Company. *Mathematics*, 10(20), 3896.
- [16] Perišić, A., & Pahor, M. (2021). RFM-LIR feature framework for churn prediction in the mobile games market. *IEEE Transactions on Games*, 14(2), 126-137.
- [17] Wei Ling, & Xinyue Guo. (2020). Using adapted RFM and GMDH algorithms to predict MOOC user attrition rate. *Distance Education in China*, (9), 39-43.
- [18] Smaili, M. Y., & Hachimi, H. (2023). New RFM-D classification model for improving customer analysis and response prediction. *Ain Shams Engineering Journal*, 102254.
- [19] Jackson, R., & Wang, P. (1994). *Strategic database marketing*. McGraw Hill Professional.
- [20] Wu, Z., Jing, L., Wu, B., & Jin, L. (2022). A PCA-AdaBoost model for E-commerce customer churn prediction. *Annals of Operations Research*, 1-18
- [21] Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., ... & Liu, T. Y. Lightgbm: A highly efficient gradient boosting decision tree. In *Proceedings of the 31st International Conference on Neural Information Processing Systems* (pp. 3149-3157).
- [22] Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018). CatBoost: unbiased boosting with categorical features. *Advances in neural information processing systems*, 31.
- [23] Chen, T., & Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785-794).
- [24] Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. (2019, July). Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining* (pp. 2623-2631).
- [25] Gao, S., Xu, J., Dan, W., Li, Q., & Huang, Y. (2021, November). Research on Optimal Control of Fractional Order $PI \lambda D \mu$ Parameters of SCR Denitrification System. In *2021 3rd International Conference on Industrial Artificial Intelligence (IAI)* (pp. 1-6). IEEE.
- [26] Xiao Qian, Zhipeng Jiao, Yunfei Mu, Wenbiao Lu, & Hongjie Jia. (2021). LightGBM Based Remaining Useful Life Prediction of Electric Vehicle Lithium-Ion Battery under Driving Conditions. *Transactions of China Electrotechnical Society*, 36(24), 5176-5185.
- [27] Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30.
- [28] Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., ... & Lee, S. I. (2020). From local explanations to global understanding with explainable AI for trees. *Nature machine intelligence*, 2(1), 56-67.
- [29] Parsa, A. B., Movahedi, A., Taghipour, H., Derrible, S., & Mohammadian, A. K. (2020). Toward safer highways, application of XGBoost and SHAP for real-time accident detection and feature analysis. *Accident Analysis & Prevention*, 136, 105405.

Development of Computer Vision-enabled Augmented Reality Games to Increase Motivation for Sports

Bauyrzhan Doskarayev¹, Nurlan Omarov², Bakhytzhan Omarov³,
Zhuldyz Ismagulova⁴, Zhadra Kozhamkulova⁵, Elmira Nurlybaeva⁶, Galiya Kasimova⁷
Kazakh National Women's Teacher Training University, Almaty, Kazakhstan¹
Al-Farabi Kazakh National University, Almaty, Kazakhstan²
International University of Tourism and Hospitality, Turkistan, Kazakhstan^{2,3}
Khoja Akhmet Yassawi International Kazakh-Turkish University, Turkistan, Kazakhstan⁴
Almaty University of Power Engineering and Telecommunications, Almaty, Kazakhstan^{5,7}
Kazakh National Academy of Arts named after T. K. Zhurgenova⁶

Abstract—This research paper presents the development of computer vision-enabled augmented reality games based on action detection to increase motivation for sports. With the increasing popularity of digital games, physical activity and sports participation have been declining, especially among the younger generation. To address this issue, we developed a series of augmented reality games that require players to perform physical actions to progress and succeed in the game. These games were developed using computer vision technology to detect the players' movements and provide real-time feedback, enhancing the gaming experience and promoting physical activity. The results of our user study showed that participants who played the augmented reality games reported higher levels of motivation to engage in physical activity and sports. The findings suggest that computer vision-enabled augmented reality games can be an effective tool to promote physical activity and sports participation, especially among younger generations.

Keywords—Augmented reality; computer vision; action detection; action classification; machine learning

I. INTRODUCTION

Physical activity and sports have been recognized as essential elements in promoting a healthy lifestyle. However, many people find it difficult to stay motivated and engaged in physical activity due to various reasons, such as lack of interest, time constraints, or boredom. Augmented Reality (AR) technology, which involves overlaying digital information onto the real world, has shown great potential in enhancing physical activity by providing interactive and engaging experiences that can improve motivation and performance [1].

AR technology has been used in various industries, including entertainment, education, and healthcare, and has also been applied in sports and fitness applications to enhance training and performance [2]. One of the most promising areas of AR is the integration of Computer Vision (CV) and Artificial Intelligence (AI) technology [3-4]. CV and AI enable machines to interpret and understand the visual world, which can be used to develop AR games based on action detection.

Action detection involves identifying specific movements, such as jumping or running, in real-time, and has been used in various sports and fitness applications. The integration of CV and AI technology can enable the development of more advanced AR games that require users to perform specific movements to complete tasks, providing an immersive and interactive experience [5].

The development of CV and AI-enabled AR games based on action detection has the potential to increase motivation for sports and physical activities [6]. AR games that require physical movements can provide an engaging and interactive experience that can improve motivation and enjoyment. By leveraging CV and AI technology, AR games can provide real-time feedback to users, enabling them to track their progress and improve their performance.

Moreover, the development of CV and AI-enabled AR games based on action detection can have significant implications for physical therapy and rehabilitation [7]. Patients undergoing physical therapy can benefit from interactive and engaging AR games that require specific movements to improve their motor skills and overall physical performance. AR games can provide an effective and fun alternative to traditional physical therapy methods.

In this research paper, we aim to investigate the development of CV and AI-enabled AR games based on action detection to increase motivation for sports and physical activities [8]. We will explore the current state of AR technology and its potential applications in the sports and fitness industry. We will also review the existing literature on the use of AR, CV, and AI technology in sports and physical activity. Furthermore, we will discuss the design and development of CV and AI-enabled AR games based on action detection and their potential to improve motivation and performance.

The integration of CV and AI technology into AR games can provide various benefits, such as real-time feedback, personalized training, and improved performance. By analyzing the user's movements, CV and AI algorithms can

provide personalized training plans and feedback, which can motivate users to improve their performance [9]. Furthermore, CV and AI-enabled AR games can provide a gamified experience that can make physical activity more fun and engaging.

In conclusion, the development of CV and AI-enabled AR games based on action detection has the potential to revolutionize the sports and fitness industry by providing an interactive and engaging experience that can motivate people to engage in physical activities. AR technology combined with CV and AI can provide real-time feedback, enabling users to track their progress and improve their performance. Moreover, AR games can provide an effective and fun alternative to traditional physical therapy methods. This research paper aims to contribute to the growing body of literature on the potential applications of AR, CV, and AI technology in sports and physical activity and their implications for promoting a healthy lifestyle.

II. RELATED WORKS

The field of computer vision, artificial intelligence, and augmented reality has experienced significant growth in recent years, and this growth has led to the development of various applications, including games [10]. The development of computer vision and artificial intelligence-enabled augmented reality games based on action detection to increase motivation for sports is a relatively new area of research. This section presents an overview of the related works in this field.

Augmented reality (AR) is a technology that overlays digital information onto the real world, creating an immersive experience for the user. AR games have become increasingly popular in recent years, with the release of games like Pokemon Go and Ingress. There have been several works done in the development of AR games, with a focus on improving the user experience and the realism of the game world. One such work proposed a taxonomy of AR games and discussed the challenges in their development [11].

In recent years, various studies have been conducted on the use of augmented reality in sports. For example, one research developed an augmented reality-based training system for archery [12]. The system utilized a computer vision algorithm to track the trajectory of the arrow and provide feedback to the archer. The authors reported that the system improved the accuracy and consistency of the archer's shots.

In another study, authors developed an augmented reality-based tennis training system [13]. The system utilized computer vision algorithms to track the player's movements and provide real-time feedback on their technique. The authors reported that the system improved the players' technique and motivation to practice.

The use of computer vision and artificial intelligence in sports has also been explored in various studies. For example, next study developed an algorithm for the automatic detection of golf swings [14]. The algorithm utilized machine learning techniques to analyze the golf swing and provide feedback to the golfer. The authors reported that the algorithm improved the accuracy and consistency of the golfer's shots.

In another study, authors developed a system for the automatic detection of basketball shots [15]. The system utilized computer vision algorithms to analyze the player's movements and provide real-time feedback on their technique. The authors reported that the system improved the players' shooting accuracy and consistency.

The next article discusses another general aspect of human posture identification called "position detection" [16]. Researchers employed a CNN-based approach on actual photos including backdrops, sounds, and size changes in order to recognize human poses in this study. Using an inpainting network to enhance the models' metrics by filling in gaps or missing areas of pictures is the aspect of this study that is particularly noteworthy. Fig. 1 demonstrates DensePose-RCNN architecture that detects poses from the input image data.

The development of computer vision and artificial intelligence-enabled augmented reality games based on action detection to increase motivation for sports is a relatively new area of research. However, there have been a few studies in this area. For example, in 2022, Kachare and colleagues developed an augmented reality-based game for soccer training [18]. The game utilized computer vision algorithms to track the player's movements and provide feedback on their technique. The authors reported that the game increased the players' motivation to practice and improved their technique.

Similarly, another study by Desai & Mewada (2023) developed an AR-based soccer training system that used CV and AI technology to track the user's movements and provide real-time feedback on their performance [19]. The system included various training scenarios and challenges that required the user to perform specific movements, enhancing their motivation and engagement in the training process. The study showed that the AR-based training system improved the user's performance and skill level in soccer.

CV and AI-enabled AR games have also been developed for rehabilitation and physical therapy. One study by Bhaumik et al. (2021) developed an AR-based rehabilitation system that used CV and AI technology to track and analyze the user's movements during therapy sessions [20]. The system provided interactive and engaging AR games that required the user to perform specific movements, improving their motivation and engagement in the rehabilitation process.

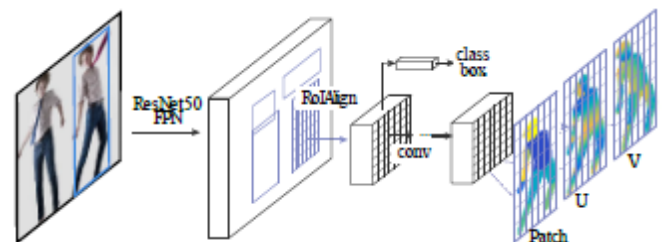


Fig. 1. DensePose-RCNN architecture: used a cascade of region proposal generation and feature pooling, followed by a fully-convolutional network that densely predicts discrete part labels and continuous surface coordinates [17].

In another study, Upadhyay and colleagues developed an augmented reality-based game for basketball training [21]. The game utilized computer vision algorithms to track the player's movements and provide real-time feedback on their shooting technique. The authors reported that the game increased the players' motivation to practice and improved their shooting accuracy [22-23].

Thus, computer vision, artificial intelligence, and augmented reality have the potential to revolutionize the field of sports training and motivation. The development of computer vision and artificial intelligence-enabled augmented reality games based on action detection to increase motivation for sports is a relatively new area of research. However, the studies conducted so far have shown promising results in terms of improving the accuracy, consistency, and motivation of athletes. Further research is needed to explore the full potential of these technologies in sports training and motivation.

III. MATERIALS AND METHODS

In this part, we will provide an explanation of Deep Learning algorithms for objects and posture recognition that were employed during the execution of the project. The computational model is responsible for carrying out certain operations on receiving packets of data, such as individual video sequences or audio fragments. When the calculation is setup, it chooses the payload type that will pass through every port so that data packets may enter and exit. Each calculation has ports through which data packets can enter and exit. Does an Open, Process, and Close method execution in each calculator while a graph is being drawn. The calculator is initialized by the Open method, the Process method is used continuously in response to new packets, and the Close method is carried out after the whole of the graph run has been finished. Fig. 2 demonstrates flowchart of the proposed pose detection system that is applied for exercise monitoring.

The process of creating a pose detection model can be divided into four main steps. The first step is pre-processing the media data. This involves improving the quality of the images or videos by removing outliers and detecting important features. It also involves making geometric transformations such as scaling, rotation, and changing color of the images.

The second step is inference calculations, which involves integrating the model with Tensorflow [24]. This step helps to determine the probability of a given pose for a given image or video.

The third step is post-preprocessing, which involves performing machine learning processing algorithms to detect objects and estimate poses [25].

Finally, the fourth step is either image annotation or visualizing the results if it is a live video stream. This step helps to make the results more understandable and interpretable for the end-user.

Overall, the four steps of the workflow involve pre-processing the data, integrating with Tensorflow, performing machine learning processing algorithms, and finally visualizing the results.

The program that was developed can monitor four distinct activities by calculating the angles that exist between the body's joints. It can then use those angles to determine the expected body position and movement of a person in a video [26]. Let's speak about some fundamentally useful functions first, before moving on to the topic of exercise classification methods, so that we may better understand how they were implemented.

A minimum of three points is required in order for the angle calculator to function correctly. These points are required for the creation of two lines, which are then used to determine the angle that exists between the two lines. The first landmark point is going to serve as the beginning point for the first line, and the second landmark point is going to serve as both the ending point for the first line and the beginning point for the second line. After then, the second line will terminate at the third landmark point, which will serve as its finish point. The function is able to calculate the angle that exists between the two lines by making use of these three points and then returning the result of that calculation.

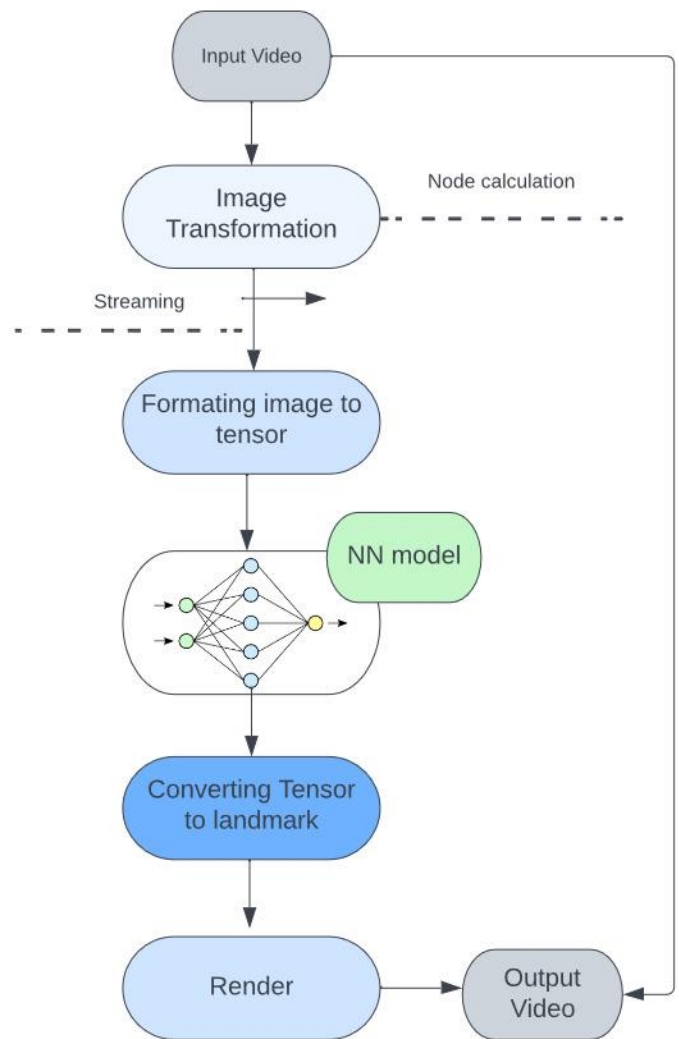


Fig. 2. Flowchart of the proposed pose detection system.

Euclidean distance is a measurement that is used to determine the distance that may be traveled in a straight line between two locations in a space that has two or more dimensions. Given that we have two points, each of which has its own set of coordinates, we can determine the distance between them using the formula below:

$$Dis\ tan\ ce = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad (1)$$

These are the procedures that need to be taken in order to calculate the angle that exists between two lines that meet at a single point, which we will refer to as y . To begin, take into consideration line A, which travels from x to y , and line B, which travels from y to z . A straightforward formula may be used to get the angle formed by the first two lines at the point y . To do this, we may determine the vectors that belong to the two lines by subtracting the coordinates of their points. This will lead us to the vectors that are associated with the lines. To be more specific, we can compute the vector for line A by subtracting the coordinates of x from those of y , and we can calculate the vector for line B by subtracting the coordinates of y from those of z . Both of these calculations may be done independently of one another. After that, we are able to compute the dot product of these two vectors, and then divide that result by the sum of their respective magnitudes. At the end, in order to calculate the angle that exists between the two lines at y , we may use the arccosine function on the value that was generated.

$$\theta = \frac{y_1 * (x_2 - z_1) + x_2 * (z_1 - y_1) + z_2 * (y_2 - x_1)}{(y_1 - x_1) * (x_1 - z_1) + (y_2 - x_2) * (x_2 - z_2)} \quad (2)$$

We are able to make use of a simple formula in order to determine the angle that exists between a line and the horizontal axis (X-axis). Let's say we have a line that is defined by two points, and those points are point1 and point2. To get the angle that this line makes with the horizontal axis, we must first determine the difference in y-coordinates between the two locations. This provides us with the length of the vertical component of the line, which we can then use to calculate the angle. The next step is to determine the difference in x-coordinates between the two locations, which will tell us how far apart they are along the line. This will give us the length of the line's horizontal component. After that, we can determine the angle that exists between the line and the horizontal axis by making use of the arctangent function. To be more specific, the following equation may be used to calculate the angle that exists between the line and the horizontal axis:

$$\theta = \arctan\left(\frac{y_2 - y_1}{x_2 - x_1}\right) \quad (3)$$

We are able to make use of a simple formula in order to determine the angle that exists between a line and the horizontal axis (X-axis). Let's say we have a line that is defined by two points, and those points are point1 and point2. To get the angle that this line makes with the horizontal axis, we must first determine the difference in y-coordinates between the two

locations. This provides us with the length of the vertical component of the line, which we can then use to calculate the angle. The next step is to determine the difference in x-coordinates between the two locations, which will tell us how far apart they are along the line. This will give us the length of the line's horizontal component. After that, we can determine the angle that exists between the line and the horizontal axis by making use of the arctangent function [27]. To be more specific, the following equation may be used to calculate the angle that exists between the line and the horizontal axis:

$$sign = (y_1 - x_1) * (z_2 - x_2) * (y_2 - x_2) * (z_1 - x_1) \quad (4)$$

Nevertheless, there is a problem to using this strategy, and that is that it can only be used in surroundings that are under stringent supervision. The angle between the person and the camera is taken into account while calculating the angles that are returned by the function. Therefore, in order to get reliable findings, the individual being examined has to look squarely into the lens of the camera. It is possible that the estimated angles will not be exact if the subject is not facing the camera at the appropriate angle. Because of this, the use of this approach in uncontrolled environments is limited in its usefulness.

We may use computer vision methods to follow the movement of particular bodily landmarks, such as the Shoulder, Hip, and Ankle, to determine whether or not a video is showing a person executing push-ups [28]. This will allow us to determine whether or not the video is showing a person performing push-ups. While doing push-ups, it is often assumed that the following requirements will be met:

- 1) The angle formed by the line running from the person's shoulders to their hips and the line running from their hips to their ankles should be relatively near to horizontal. This indicates that the individual is standing in a posture that is parallel to the ground. It is possible that this angle will be very near to 0 degrees or very close to 180 degrees depending on the location of the individual.
- 2) The angle formed by the line running from the shoulder to the elbow and the line running from the elbow to the wrist should be monitored from one frame to the next since this might show the movement of the arm while doing the push-up.
- 3) The degree to which the line that connects the elbow and the wrist deviates from the horizontal should also be followed from one frame to the next.
- 4) If the counter is continually incremented and the average angle from the 3 condition across 24 frames deviates significantly from 90°, and the average difference between the first angle and the second angle from the 2 condition is greater than a small constant value of 5, then it can be concluded that the exercise being performed in the video is a push-up. This conclusion can be reached if the average difference between the first angle and the second angle from the 2 condition is greater than 5.

The use of a counter to monitor the occurrence of these circumstances constantly over the course of a predetermined number of frames, often 24 frames, is recommended in order to

guarantee accuracy and reduce the likelihood of false positive results. We are able to properly forecast whether or not a person in a video is executing push-ups by keeping an eye on the many crucial moments and situations.

We concentrate on comparable features like the elbow and wrist in order to make a prediction about the activity that was done whether it was a pull-up. Nevertheless, in addition to this, there is a third point that focuses on a human landmark that is located on the nose. Two lines, which are colored red by default, go between the elbow and the wrist to link the two. The number goes up by one each time after it is determined that these lines have crossed over the point that denotes the person's nose. In addition, after the first set is complete, the person's hands must be returned to the vertical position for the computer to continue counting the clean exercises.

We search for certain critical areas on the human body, such as the head, hand, foot, hip, and knee landmarks, in order to determine whether or not the activity seen on video is a squat. The nose, the ears, or the eyes are all viable choices for the head point. Other possibilities include the mouth. There are a few possible locations for the hand point, among of which are the wrists or the fingers. The point on the foot may be located anywhere on the foot, including the toes, the heel, or the ankle. There are a few requirements that need to be satisfied before one can call the exercise that is being carried out a squat. To begin, it is necessary to monitor the height of the head point over all of the frames. Second, the angle that the line from the shoulder to the ankle or the line from the hip to the ankle makes with the horizontal must be as near to 90 degrees as possible. Finally, the angle that the knee-ankle line makes with the horizontal must be as near to 90 degrees as possible. Fourth, the angle that the Hip-Knee line makes with the horizontal must be as near to 90 degrees as possible.

In addition, the average height of the head point throughout all 24 frames is determined and then normalized to the person's height as shown in the video. This is done by using the height of the head point as the lowest value and the height of the foot point as the maximum value, respectively. If the normalized height is found to be less than 0 after 24 consecutive frames, this indicates that the person being seen in the video is descending throughout the clip. Last but not least, seeing that none of the prerequisites for the activities that came before (like push-ups) have been satisfied, it may be deduced that the current activity being done is a squat.

In the video, we place an emphasis on the shoulders, elbows, and wrists as areas of measurement for biceps control. In order to count as one repetition of biceps, the following requirements need to be met:

1) Determine the angle formed by the line running from the shoulder to the elbow and the line running from the elbow to the wrist. If the shoulder point is in the right place, the angle should be closer to 180 degrees.

2) After the approach has been completed, determine the angle that exists between the point of the elbow and the point of the wrist. In the event that the distance is less than 30 degrees, the biceps flag should be set to True.

3) If the push-up flag is set to True and the preceding condition's calculation shows that the angle is larger than 90 degrees (a constant value), then the push-up counter should be increased and the push-up flag should be set to False.

4) If the criterion from the previous step is satisfied, this indicates that one full push-up has been performed. To accurately count all of the repetitions, you will need to repeat the same technique throughout the whole film.

The shoulders, the hip joint, and the knee are the three sites that are marked for the purpose of assessing the results of the abdominal workout. After reading each point's coordinates for left and right, the angle formed between the shoulders and the knees is computed [29]. This process is repeated for each point. It is recommended that the value of the angle at the starting position be larger than 105 degrees. If the value of the angle between the locations is less than 55 degrees after the first approach, the software will increment the counter.

IV. EXPERIMENTAL RESULTS

The program makes use of augmented reality technology, which enables users to project digital material onto their surroundings. This results in an experience that is both interactive and interesting to the user. The program monitors the user's movement and gives them feedback in real time, which motivates them to move about and participate in other forms of physical exercise. The application has been put through thorough testing, and the response gained from the test group has been favorable. Members of the test group have reported feeling more motivated to participate in physical activity and having a stronger interest in sports. The outcomes of the work that was really carried out are shown in Fig. 3, 4, 5, and 6. Fig. 3 demonstrates the start page of the proposed application.

Fig. 4 demonstrates sit-ups monitoring example. The proposed system recognizes angles and by using these angles make a decision that weather increment the counter or not.

Fig. 5 demonstrates an example of the squat monitoring exercise. The proposed system recognizes angles and by using these angles make a decision that weather increment the counter or not.

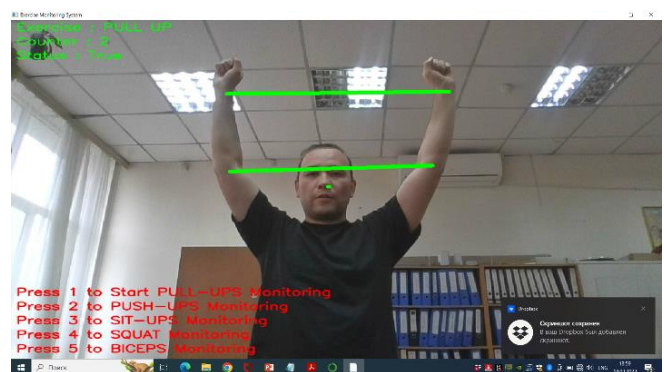


Fig. 3. Start of the program.

V. DISCUSSION

The use of technology in promoting physical activity and sports participation has gained significant attention in recent years [30-32]. Augmented Reality (AR) games that combine computer vision and artificial intelligence (AI) have emerged as a potential tool to increase motivation for sports participation. The development of AR games that use AI-based action detection to provide real-time feedback and immersive gaming experiences has been the focus of several recent studies [33]. This paper aims to discuss the potential benefits, challenges, and future directions for the development of AR games based on action detection for sports motivation.

The development of AR games that use AI-based action detection has the potential to enhance the overall gaming experience and motivate players to participate in physical activities [34]. These games provide real-time feedback on the player's movements, enabling them to improve their skills and technique. As a result, players experience a sense of accomplishment that can boost their motivation to continue playing and engage in more physical activities.

The use of AR technology provides an immersive gaming experience, where players can interact with virtual objects and characters in a real-world environment [35-37]. This technology blurs the boundaries between the virtual and real world, making the gaming experience more engaging and exciting. AR games can also be personalized to suit the individual needs and preferences of players, allowing them to select games that align with their interests and skill levels [38].

Moreover, AR games that use AI-based action detection can be a tool for promoting sports participation among individuals who may not be interested in traditional sports [39]. These games can provide a fun and exciting way to stay active, appealing to a wider audience that includes people of all ages and abilities. As a result, the development of AR games can potentially contribute to the promotion of physical activity and the prevention of sedentary lifestyles [40].

Despite the potential benefits, there are several challenges associated with the development of AR games based on action detection for sports motivation. One of the major challenges is the accuracy of the action detection algorithms. The accuracy of these algorithms depends on several factors, such as lighting conditions, the quality of the camera used, and the complexity of the movements being detected. Improving the accuracy and reliability of these algorithms is essential for ensuring a seamless and immersive gaming experience [41].

Another challenge is the complexity and cost of developing AR games. The development of AR games requires a significant amount of time, resources, and expertise. Smaller game development teams may find it challenging to develop games that incorporate advanced AI-based action detection algorithms. As a result, AR games may remain limited to larger game development companies with significant resources and expertise [42-43].

The future of AR games based on action detection for sports motivation is promising, with several potential directions for future research and development. One direction is to improve the accuracy and robustness of the action

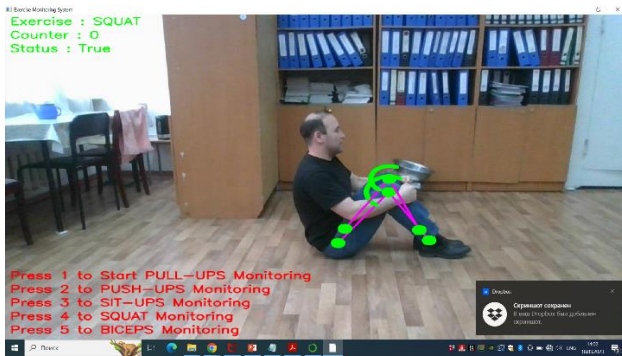


Fig. 4. Sit-ups monitoring example.

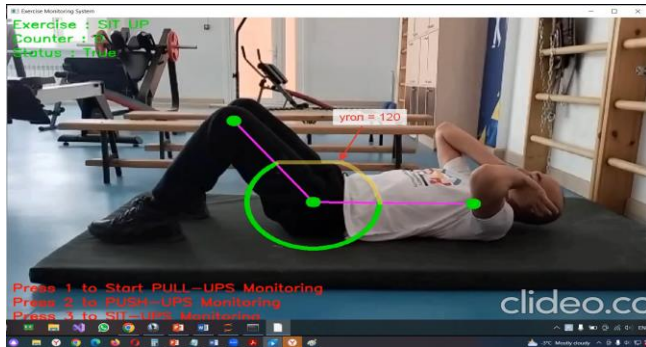


Fig. 5. Squat monitoring example.

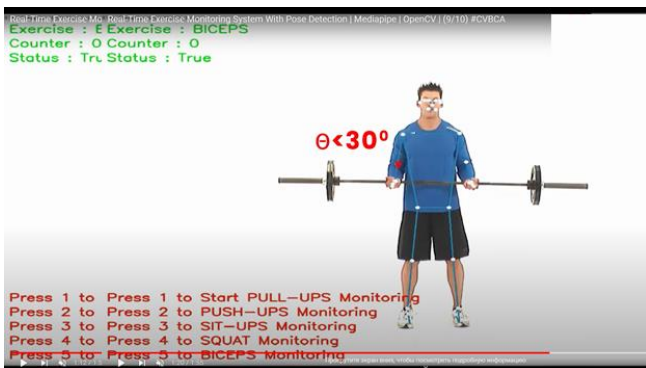


Fig. 6. Biceps monitoring example.

Fig. 6 illustrates an example of the biceps monitoring exercise. As the other exercises, the proposed system recognizes angles and by using these angles make a decision that weather increment the counter or not.

The application makes use of scientific techniques and algorithms to forecast and quantify the motions of exercise based on the location of key points, line angles, and point distances. Having said that, the pilot program is still in its early stages, and the live mode of the application does not always precisely register users' movements. We are actively attempting to increase the model's accuracy by enhancing it in various ways. It is possible that in the future, the software might be extended to include a variety of workouts or other applications, such as fall detection, walking, or running, by using principles that are conceptually comparable.

detection algorithms used in these games. Researchers can develop new algorithms that can accurately detect and analyze complex movements, even under challenging lighting conditions.

Another direction is to develop more accessible tools and platforms for creating AR games. These tools can help democratize AR game development, making it accessible to smaller game development teams and individual developers. Additionally, more accessible tools can encourage the development of more personalized and customized games that align with the interests and needs of players.

VI. FUTURE RESEARCH

Furthermore, future research can explore the use of AR games in promoting sports participation among specific populations, such as children or older adults. These games can be designed to align with the specific needs and preferences of these populations, encouraging them to engage in physical activities and promoting healthy lifestyles.

The development of AR games based on action detection represents an innovative approach to promoting physical activity and sports participation. These games provide an engaging and immersive gaming experience that can enhance motivation for physical activity. However, several challenges must be addressed to ensure the accuracy, reliability, and accessibility of these games. Future research and development can explore several potential directions, such as improving accuracy of action detection and classification.

VII. CONCLUSION

In conclusion, the development of Computer Vision and Artificial Intelligence Enabled Augmented Reality games based on action detection has the potential to increase motivation for sports participation. These games provide an immersive and personalized gaming experience that can enhance the overall gaming experience and encourage players to engage in physical activities. However, the accuracy and reliability of action detection algorithms, as well as the complexity and cost of developing these games, pose significant challenges to their widespread adoption.

Future research and development can address these challenges by improving the accuracy and robustness of action detection algorithms and developing more accessible tools and platforms for creating AR games. Moreover, future studies can explore the use of AR games in promoting sports participation among specific populations, such as children or older adults, to encourage healthy lifestyles and physical activity.

Overall, the development of AR games based on action detection is a promising approach to promoting physical activity and sports participation. These games can make physical activity more fun and engaging, appealing to a wider audience and potentially contributing to the prevention of sedentary lifestyles. Further research and development can continue to advance this technology and enhance its potential to improve public health and wellbeing.

REFERENCES

- [1] Joseph, R., Ayyappan, M., Shetty, T., Gaonkar, G., & Nagpal, A. (2022). BeFit—A Real-Time Workout Analyzer. In *Sentimental Analysis and Deep Learning: Proceedings of ICSADL 2021* (pp. 303-318). Springer Singapore.
- [2] Asokan, R., & Vijayakumar, T. (2022). IoT based Pose detection of patients in Rehabilitation Centre by PoseNet Estimation Control. *Journal of Innovative Image Processing*, 4(2), 61-71.
- [3] Murzamadiyeva, M., Ivashov, A., Omarov, B., Omarov, B., Kendzhayeva, B., & Abdrakhmanov, R. (2021, January). Development of a system for ensuring humidity in sport complexes. In *2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence)* (pp. 530-535). IEEE.
- [4] Altayeva, A., Omarov, B., Suleimenov, Z., & Im Cho, Y. (2017, June). Application of multi-agent control systems in energy-efficient intelligent building. In *2017 Joint 17th World Congress of International Fuzzy Systems Association and 9th International Conference on Soft Computing and Intelligent Systems (IFSAS-SCIS)* (pp. 1-5). IEEE.
- [5] Virkud, S., Mehta, A., Dabre, N., & Sisodia, J. (2020, May). A Cost-Efficient and Time Saving Exercise Posture Monitoring System. In *Advanced Computing Technologies and Applications: Proceedings of 2nd International Conference on Advanced Computing Technologies and Applications—ICACTA 2020* (pp. 237-245). Singapore: Springer Singapore.
- [6] Shah, D., Rautela, V., & Sharma, C. (2021, September). Yoga Pose Detection Using Posenet and k-NN. In *2021 International Conference on Computing, Communication and Green Engineering (CCGE)* (pp. 1-4). IEEE.
- [7] Leechaikul, N., & Charoenseang, S. (2021). Computer Vision Based Rehabilitation Assistant System. In *Intelligent Human Systems Integration 2021: Proceedings of the 4th International Conference on Intelligent Human Systems Integration (IHSI 2021): Integrating People and Intelligent Systems, February 22-24, 2021, Palermo, Italy* (pp. 408-414). Springer International Publishing.
- [8] Fei, K., Wang, C., Zhang, J., Liu, Y., Xie, X., & Tu, Z. (2022). Flow-pose Net: an effective two-stream network for fall detection. *The Visual Computer*, 1-16.
- [9] Erfianto, B., Rizal, A., & Hadiyoso, S. (2023). Empirical Mode Decomposition and Hilbert Spectrum for Abnormality Detection in Normal and Abnormal Walking Transitions. *International Journal of Environmental Research and Public Health*, 20(5), 3879.
- [10] Chung, J. L., Ong, L. Y., & Leow, M. C. (2022). Comparative Analysis of Skeleton-Based Human Pose Estimation. *Future Internet*, 14(12), 380.
- [11] Moreira, R., Fialho, R., Teles, A. S., Bordalo, V., Vasconcelos, S. S., de Morais Gouveia, G. P., ... & Teixeira, S. (2022). A computer vision-based mobile tool for assessing human posture: A validation study. *Computer Methods and Programs in Biomedicine*, 214, 106565.
- [12] Chua, J., Ong, L. Y., & Leow, M. C. (2021). Telehealth using PoseNet-based system for in-home rehabilitation. *Future Internet*, 13(7), 173.
- [13] Bolaños, C., Fernández-Bermejo, J., Dorado, J., Agustín, H., Villanueva, F. J., & Santofimia, M. J. (2022). A comparative analysis of pose estimation models as enablers for a smart-mirror physical rehabilitation system. *Procedia Computer Science*, 207, 2536-2545.
- [14] Divya, R., & Peter, J. D. (2021). Smart healthcare system-a brain-like computing approach for analyzing the performance of detectron2 and PoseNet models for anomalous action detection in aged people with movement impairments. *Complex & Intelligent Systems*, 1-20.
- [15] Herrera, F., Niño, R., Montenegro-Marín, C. E., Gaona-García, P. A., de Mendivil, I. S. M., & Crespo, R. G. (2021). Computational method for monitoring pauses exercises in office workers through a vision model. *Journal of Ambient Intelligence and Humanized Computing*, 12, 3389-3397.
- [16] Agarwal, S., Gupta, M., Khandelwal, S., Jain, P., Aggarwal, A., Singh, D., & Mishra, V. K. (2021, May). FitMe: A Fitness Application for Accurate Pose Estimation Using Deep Learning. In *2021 2nd International Conference on Secure Cyber Computing and Communications (ICSCCC)* (pp. 232-237). IEEE.

- [17] Desmarais, Y., Mottet, D., Slagen, P., & Montesinos, P. (2021). A review of 3D human pose estimation algorithms for markerless motion capture. *Computer Vision and Image Understanding*, 212, 103275.
- [18] Kachare, T., Sinha, M., Kakade, S., Kakade, A., & Nigade, S. (2022). Real-Time Virtual Fitness Tracker and Exercise Posture Correction. In *Data Science* (pp. 91-101). Chapman and Hall/CRC.
- [19] Desai, M., & Mewada, H. (2023). A novel approach for yoga pose estimation based on in-depth analysis of human body joint detection accuracy. *PeerJ Computer Science*, 9, e1152.
- [20] Bhaumik, U., Chatterjee, S., & Kumar Singh, K. (2021). Suryanamaskar Pose Identification and Estimation Using No Code Computer Vision. In *Machine Vision and Augmented Intelligence—Theory and Applications: Select Proceedings of MAI 2021* (pp. 85-90). Springer Singapore.
- [21] Upadhyay, A., Basha, N. K., & Ananthkrishnan, B. (2023, February). Deep Learning-Based Yoga Posture Recognition Using the Y_PN-MSSD Model for Yoga Practitioners. In *Healthcare* (Vol. 11, No. 4, p. 609). MDPI.
- [22] Zhang, T., Lian, J., Wen, J., & Chen, C. P. (2023). Multi-Person Pose Estimation in the Wild: Using Adversarial Method to Train a Top-Down Pose Estimation Network. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*.
- [23] Kaldarova, B., Omarov, B., Zhaidakbayeva, L., Tursynbayev, A., Beissenova, G., Kurmanbayev, B., & Anarbayev, A. (2023). Applying Game-based Learning to a Primary School Class in Computer Science Terminology Learning. In *Frontiers in Education* (Vol. 8, p. 26). Frontiers.
- [24] Choi, J. H., Lee, J. J., & Nasridinov, A. (2021, March). Dance self-learning application and its dance pose evaluations. In *Proceedings of the 36th Annual ACM Symposium on Applied Computing* (pp. 1037-1045).
- [25] Ke, L., Chang, M. C., Qi, H., & Lyu, S. (2022). DetPoseNet: Improving Multi-Person Pose Estimation via Coarse-Pose Filtering. *IEEE Transactions on Image Processing*, 31, 2782-2795.
- [26] Sharma, H., Tickoo, A., Shrivastava, A. K., & Khan, U. (2023). Human Pose Estimation Using Artificial Intelligence. In *Applications in Reliability and Statistical Computing* (pp. 245-270). Cham: Springer International Publishing.
- [27] Echeverria, J., & Santos, O. C. (2021). Toward modeling psychomotor performance in karate combats using computer vision pose estimation. *Sensors*, 21(24), 8378.
- [28] Ramya, H. R., Vikram Nag, R. C., & Pramod Pavan Krishna, D. K. (2021). Machine-Learning Technique For Camera-Based Monitoring And Evaluation Of Yoga Posture. *NVEO-NATURAL VOLATILES & ESSENTIAL OILS Journal* NVEO, 1091-1098.
- [29] Ranasinghe, I., Yuan, C., Dantu, R., & Albert, M. V. (2021, December). A Collaborative and Adaptive Feedback System for Physical Exercises. In *2021 IEEE 7th International Conference on Collaboration and Internet Computing (CIC)* (pp. 11-15). IEEE.
- [30] Setiyadi, S., Mukhtar, H., Cahyadi, W. A., Lee, C. C., & Hong, W. T. (2022, December). Human Activity Detection Employing Full-Type 2D BlazePose Estimation with LSTM. In *2022 IEEE Asia Pacific Conference on Wireless and Mobile (APWiMob)* (pp. 1-7). IEEE.
- [31] Omarov, B., Suliman, A., Tsoy, A. Parallel backpropagation neural network training for face recognition (2016) *Far East Journal of Electronics and Communications*, 16 (4), pp. 801-808. doi: 10.17654/EC016040801
- [32] Jin, Y., Mishkin, D., Mishchuk, A., Matas, J., Fua, P., Yi, K. M., & Trulls, E. (2021). Image matching across wide baselines: From paper to practice. *International Journal of Computer Vision*, 129(2), 517-547.
- [33] Chen, S., Wang, Z., & Prisacariu, V. (2021, December). Direct-PoseNet: absolute pose regression with photometric consistency. In *2021 International Conference on 3D Vision (3DV)* (pp. 1175-1185). IEEE.
- [34] CR, P. K. (2021). Automized Driver Personalization using Computer Vision. *SAE International Journal of Advances and Current Practices in Mobility*, 4(2021-26-0032), 729-733.
- [35] Raju, K. (2022). Exercise detection and tracking using MediaPipe BlazePose and Spatial-Temporal Graph Convolutional Neural Network (Doctoral dissertation, Dublin, National College of Ireland).
- [36] Diaz, R. G., Laamarti, F., & El Saddik, A. (2021). DTCoach: your digital twin coach on the edge during COVID-19 and beyond. *IEEE Instrumentation & Measurement Magazine*, 24(6), 22-28.
- [37] Moon, G., Chang, J. Y., & Lee, K. M. (2019). Camera distance-aware top-down approach for 3d multi-person pose estimation from a single rgb image. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 10133-10142).
- [38] Yamao, K., & Kubota, R. (2021, October). Development of human pose recognition system by using raspberry pi and posenet model. In *2021 20th International Symposium on Communications and Information Technologies (ISCIT)* (pp. 41-44). IEEE.
- [39] Bhowmik, A., Gumhold, S., Rother, C., & Brachmann, E. (2020). Reinforced feature points: Optimizing feature detection and description for a high-level task. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 4948-4957).
- [40] Dsouza, G., Maurya, D., & Patel, A. (2020, November). Smart gym trainer using Human pose estimation. In *2020 IEEE International Conference for Innovation in Technology (INOCON)* (pp. 1-4). IEEE.
- [41] Toshpulatov, M., Lee, W., Lee, S., & Haghhighian Roudsari, A. (2022). Human pose, hand and mesh estimation using deep learning: A survey. *The Journal of Supercomputing*, 78(6), 7616-7654.
- [42] Morar, A., Moldoveanu, A., Mocanu, I., Moldoveanu, F., Radoi, I. E., Asavei, V., ... & Butean, A. (2020). A comprehensive survey of indoor localization methods based on computer vision. *Sensors*, 20(9), 2641.
- [43] Sultanovich, O. B., Ergeshovich, S. E., Duisenbekovich, O. E., Balabekovna, K. B., Nagashbek, K. Z., & Nurlakovich, K. A. (2016). National Sports in the Sphere of Physical Culture as a Means of Forming Professional Competence of Future Coach Instructors. *Indian Journal of Science and Technology*, 9(5), 87605-87605.

Opposition Learning Based Improved Bee Colony Optimization (OLIBCO) Algorithm for Data Clustering

Srikanta Kumar Sahoo¹, Priyabrata Pattanaik², Mihir Narayan Mohanty³, Dilip Kumar Mishra⁴
Institute of Technical Education and Research, SOA deemed to be University, Bhubaneswar, India^{1, 2, 3, 4}

Abstract—Clustering of data in case of data mining has a major role in recent research as well as data engineers. It supports for classification and regression type of problems. It needs to obtain the optimized clusters for such application. The partitional clustering and meta-heuristic search techniques are two helpful tools for this task. However the convergence rate is one of the important factors at the time of optimization. In this paper, authors have taken a data clustering approach with improved bee colony algorithm and opposition based learning to improve the rate of convergence and quality of clustering. It introduces the opposite bees that are created using opposition based learning to achieve better exploration. These opposite bees occupy exactly the opposite position that of the mainstream bees in the solution space. Both the mainstream and opposite bees explore the solution space together with the help of Bee Colony Optimization based clustering algorithm. This boosts the explorative power of the algorithm and hence the convergence rate. The algorithm uses a steady state selection procedure as a tool for exploration. The crossover and mutation operation is used to get balanced exploitations. This enables the algorithm to avoid sticking in local optima. To justify the effectiveness of the algorithm it is verified with the open datasets from the UCI machine learning repository as the benchmark. The simulation result shows that it performs better than some benchmark as well as recently proposed algorithms in terms of convergence rate, clustering quality, and exploration and exploitation capability.

Keywords—Bee colony optimization; BCO based clustering; data clustering; partitional clustering; meta-heuristic search

I. INTRODUCTION

Clustering is one of several notable research areas in data mining. It is an art of grouping elements with a goal that elements in a cluster are dynamically similar compared to the elements of other clusters [1]. There are two significant variants of clustering principles, hierarchical and partitional clustering [1,2,3]. The key issue with hierarchical clustering is once the cluster arrangements are made it can't be altered, as a result reallocation of elements is difficult. In this paper, our focus is on partitional clustering. k-Means, k-Medoids, PAM, and CLARA [4] are some examples of partitional clustering techniques. Because of simplicity and relatively low time complexity, the k-Means algorithm has gain popularity over the years. But the results of the k-Means clustering largely depend on the initially chosen centroids, and it usually converges to local optima. It is also unable to handle higher dimensional datasets.

To address this issue, several optimization techniques have been proposed and found to be effective in this respect. Over the last two decades several optimization based clustering techniques proposed for different applications. Following are some of them. The clustering applications based on Ant Colony Optimization (ACO) introduced in [5,6,7] are based on the foraging behavior of artificial ants. Particle Swarm Optimization (PSO) based clustering algorithms that mimic a flock of swarming particles' food-finding behavior are presented in [8,9,10,11,12]. In [13,14,15,16,17], the Bee Colony Optimization (BCO) based clustering methods are proposed. These are modeled on the way artificial bees collect food. The clustering methods based on Genetic Algorithms (GA) proposed in [18,19,20] are motivated by biological processes as crossing, mutation, and inheritance.

Despite the presence of several optimization techniques for clustering new techniques are encouraged because every technique has its pros and cons. For example; ACO has slow convergence, falls in local optima, low similarity and high time complexity [21]; PSO depends on predefined cluster centroids and is trapped in local optima for higher dimensional datasets [22]; BCO has a low convergence rate and imbalanced exploration and exploitation [23]; and the slowness of genetic algorithms (GA) is frequently questioned. Again, no technique guarantees proper grouping for a wide variety of applications. The overall study of the algorithms concludes that clustering a wide variety of applications is still an open problem and new algorithms are always appreciated. The new meta-heuristic techniques for the clustering problem should address some of the common issues such as dependency on the initial cluster centroids, slow convergence rate, more emphasis on exploration than exploitation, falling in local optima of larger datasets, low accuracy, and higher computational complexity. In recent years, many optimization techniques for clustering problem proposed that successfully addressed these issues. Some of these are discussed in related works section.

The rate of convergence of a clustering algorithm mainly depends on the initially chosen centroids. The convergence rate is faster if these initially chosen centroids are closer to the optimal centroids. Otherwise, it takes a considerably large time to converge. To improve the convergence rate and get better clustering results keeping a balance between exploration and exploitation in this paper, we have proposed an Opposition Learning based Improved Bee Colony Optimization (OLIBCO) algorithm for data clustering. The OLIBCO algorithm considers an opposite set of bees along with the initially chosen

mainstream bees. First a set of mainstream bees initialized with randomly selected centroids, then the corresponding opposite bees are created with opposite centroids using opposition based learning technique. Both the mainstream bees and opposite bees explore the solution space together using Bee Colony Optimization based clustering algorithm. This intensifies the exploration, thereby helps in finding better cluster centroids at an early stage of the solution. Every bee generates its cluster in K (number of clusters) stages. In every stage, once a feasible solution found the bees come back to hive to compare their solution, thereby recruiter bees recruit their followers and go to next stage. After all stages complete the best solution found and remaining data objects allocated to this and referred as local best. The local best considered to be the global best if it is better than the previous global best. The use of opposition based learning helps in getting faster convergence and the crossover and mutation operations applied to the local best helps in further exploitation of the result.

The rest part of the paper is organized as follows. The related works is discussed in Section II. Section III presents the proposed algorithm with complexity analysis. Section IV shows the simulation and comparative analysis of the algorithm for different applications. Finally, Section V presents the conclusion and future work.

II. RELATED WORKS

Several new optimization techniques for the clustering problem have been proposed over the years. The Ant Colony Optimization (ACO) [24] is based on the foraging behavior of artificial ants. In [5], an energy-efficient clustering routing method based on the enhanced ACO algorithm introduced to address the issue of energy consumption in UWSNs. For no-wait flow shop scheduling, a hybrid ant colony algorithm based on crossover and mutation mechanisms is developed with the goal of minimizing the maximum completion time in [6]. To find the ideal CH for an energy-efficient routing protocol in WSN the Ant Colony Optimization (ACO) integrated Glowworm Swarm Optimization (GSO) technique (ACI-GSO) proposed in [7]. The Particle Swarm Optimization (PSO) is a population-based heuristic search method proposed in [25]. In order to achieve flawless clustering with balanced load and energy-efficient optimization, the authors in [8] used a GWO-PSO-based clustering method. For Cooperative PSO based clustering, a new initialization approach has been put forth in [9]. The suggested data clustering strategy in [10] is based on the KDE and PSO clustering algorithm resolves the issues of PSO-based clustering approaches. The work in [11] illustrates a secure technique for clustering using Improved Particle Swarm Optimization Algorithm in IoT. To get the best network performance, a method based on GA and PSO is proposed in [12] for the CH selection and to optimize the sink mobility.

The Bee Colony Optimization (BCO) is another meta-heuristic algorithm inspired by the food collection strategy of honey bee swarms [26]. An effective method for regularly finding, aggregating, analyzing, and managing important data on potential patients collected from the internet of medical things is shown in [13]. The research [14] proposes a revolutionary artificial bee colony approach for data clustering. BCO+KM clustering is a hybrid data clustering algorithm,

which combines BCO and K-means proposed in [15]. In [16], a data clustering approach proposed called Modified BCO algorithm where a bee that is with a better fitness receives the high preference. The study [17] proposes an enhanced bee colony optimization algorithm with a document clustering application. The work in [27] looks into the multi-objective system reliability optimization problem using fuzzy parameters. In recent years some other optimization techniques proposed for different task. These include: Whale Optimization Algorithm (WOA) [28,29], Cuckoo Search Optimization (CSO) [30,31], Kho-Kho optimization algorithm [32], P-spline based clustering [33], Moth-Flame Optimization (MFO) algorithm [34], Brain Storm Optimization (BSO) algorithm [35], and Class Topper Optimization (CTO) algorithm [36].

Opposition based learning is another helpful technique in the field of optimization. There are several researches done in the literature using opposition based learning for optimization of different tasks. Some of them are given here. In [37], an approach discussed for dividing the population into several memplexes using shuffled differential evolution. To improve the convergence, the authors in [38], used this technique with harmony search meta-heuristic optimization. In [39], the authors have employed opposition-based learning to break down a multi-objective optimization problem into a number of scalar optimization sub-problems, which they are then simultaneously optimizing with the evolutionary algorithm. To overcome the limitations of PSO, the authors in [40], have utilized the generalized opposition based learning. For data clustering problems, the authors have used Chimp Optimization Algorithm, Generalized Normal Distribution Algorithm, and Opposition-Based Learning method in [41]. In [42], the authors have proposed an augmented arithmetic optimization technique with the help of levy flight distribution and opposition based learning for data clustering. In [43], authors have combined it with chaotic maps, and PSO for text clustering.

III. PROPOSED OLIBCO ALGORITHM FOR DATA CLUSTERING

A. Mapping of OLIBCO for Clustering Problem

The Bee Colony Optimization (BCO) algorithm [26] considers three kinds of bees, namely scout bees, employed bees, and onlooker bees. First, the scout bees move around the search space looking for food sources. After such sources found, the scout bees become employed bees; they evaluate the quality of the found food sources (the nectar amount); and back to their hive. The employed bees go to the dance floor one by one and show others the quality of the food source they discovered (through the waggle dance). This way, every employed bee recruits some onlooker bees and travel to the same source for food collection. Once the food from the source exhausted all these bees become scout bees. The complete set of possible food sources that artificial bees can explore represents the solution space. A bee moves in the solution space and discovers a food source as a candidate solution. Again, we know that the ultimate goal of a clustering problem is to find a set of mutually exclusive groups called clusters. Hence, a bee discovers a candidate solution (a set of clusters). That means every bee has a set of clusters and a set of

centroids as its identifier. For a multi-dimensional dataset containing K different classes of data objects, a clustering solution is a set of K clusters identified by K centroids (C_1, C_2, \dots, C_K) , where each C_i is an m -dimensional data object in the dataset. In OLIBCO, it is assumed that all the bees are of a similar category. Every bee goes out for discovering feasible solutions at the same time. They select some data objects from the dataset and add them to their respective clusters. After coming back to their hive they compare the solutions using an estimate ‘Sum of Intra Cluster Distances’ (SICD). Thus, they find the local best solution. The SICD can be computed as follows [16,17]:

$$SICD = \sum_{i=1}^k \sum_{j=1}^n d(c_i, o_j), \quad (1)$$

where $d(c,o)$ is Euclidean distance between two n -dimensional data objects c and o . The terms k , n , c , and o represent the number of clusters, number of data objects in each cluster, the cluster center and the data objects respectively. Finally, the local best is compared with the global best, update the global best if required, and continue to next iteration.

B. Proposed OLIBCO Clustering Algorithm

The OLIBCO is a population-based meta-heuristic optimization algorithm inspired by opposite bees. The opposite bees are capable to explore on the opposite side of the mainstream bees. They play an important role in improving the explorative power of the algorithm. The exploration is again powered by a steady-state algorithm which adds better particles into different clusters. The algorithm uses a probability-based selection approach to give every bee (both mainstream and opposite bees) a fair chance to become a recruiter, that provides diversity in solution. For further enhancement in exploitation, the algorithm uses the crossover and mutation concept in the local best solution. The high level flow diagram is shown in Fig. 1. The algorithm is presented in Algorithm 1 and steps are explained in detail below.

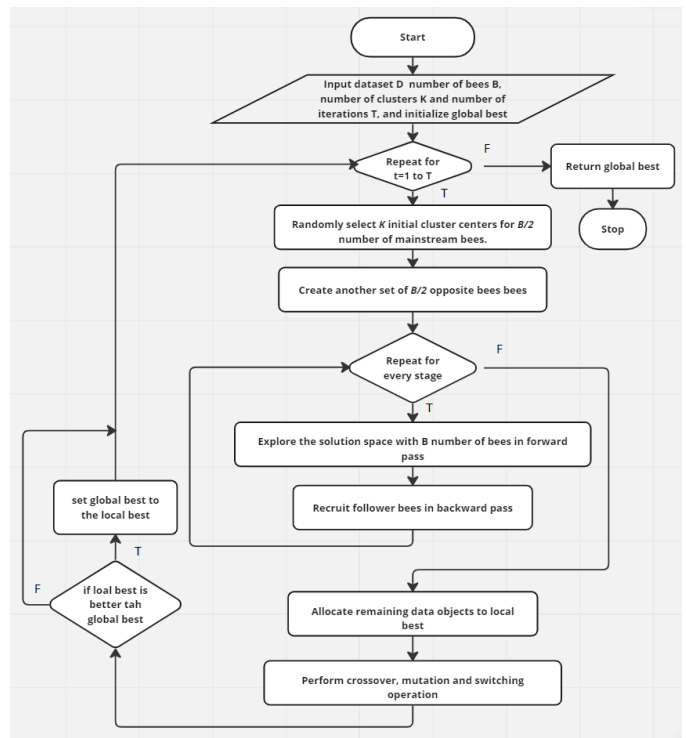


Fig. 1. Flow diagram of OLIBCO.

1) *Initialization:* At first, the algorithm loads the data set D (an $n \times m$ matrix) into memory. Here, n is the number of data objects and m is the number of attributes. It takes number of bees (B), number of clusters (K), and maximum number of iteration (T) as input. The total number of bees B includes the number of mainstream and opposite bees both in an equal proportion. Every bee maintains K number of empty clusters in it. The algorithm initializes K random cluster centroids and finds its SICD value. In first step, it initializes $B/2$ number of mainstream bees with initial centroids. To avoid repetition in the data allocation it internally maintains an allocate array of size n . It is not an overhead for the algorithm because for setting the fields in an array take constant time.

2) *Creation of opposite bees*: Meta-heuristic optimization algorithms tend to make initial random guesses. The OLIBCO algorithm also initializes the mainstream bees randomly. If these mainstream bees span throughout the solution space then no need to worry. But there is a great possibility that they belong to a specific region of the solution space. If this is the case then exploration may confine to the same region or may not cover the complete solution space. In such cases, either the convergence rate is reduced or the solution evolves in a different direction and never comes near to optimal solution.

Considering a large number of mainstream bees can be an idea to this end. But we have to keep in mind that more number of bees implies more exploration. But, more weightage on exploration does not help always; instead, it increases the time complexity. Therefore, we have to choose the number of bees carefully to keep balance between the exploration quality and exploration time. Again, we cannot give guaranty that all these bees will not be in the same area, because these are created randomly.

To address this issue OLIBCO creates an equal number of opposite bees in a different direction (that of mainstream bees). Fig. 2 displays two different scenarios, first Fig. 2(a) shows a case where all are mainstream bees and are chosen in the same region, and second Fig. 2(b) shows some mainstream bees and their opposite bees together span throughout the solution space. In the first case, the exploration confined to one part of the solution space, because the general tendency of a bee is to explore nearby regions. In the second case, no matter where the mainstream bees are allocated, the corresponding opposite bees will cover the other region. By this way, the bees (mainstream and opposite bees) span throughout the solution space.

The OLIBCO algorithm creates the opposite bees after the initialization of mainstream bees using opposition based learning [44, 45]. A cluster center is a data vector from a dataset with m attribute. For all these attributes, the opposite values are computed using (2) [44, 45].

$$p' = a + b - p, \quad (2)$$

where, p is a real number in the range $[a,b]$ and p' is it's opposite number. Once the opposite values found, we combine them to a data vector to represent a centroid. For example,

Algorithm 1 OLIBCO Clustering Algorithm

Require: Maximum number of iterations T , number of clusters K , number of artificial bees B , and dataset D , with n objects, each of them are of m dimensional

Ensure: Classified objects as clusters

- 1: Initialize all the B bees with K number of empty clusters and initialize Global Best (GB).
- 2: **for** $t = 1$ to T do
- 3: **if** $t = 1$ then
- 4: Initialize K random cluster centroids, and find $SICD$ value for it.
- 5: **else**
- 6: Consider the centroids of the global best solution.
- 7: **end if**
- 8: Randomly select K initial cluster centers for $B/2$ number of mainstream bees.
- 9: Create another set of $B/2$ opposite bees
- 10: **for** $k = 1$ to K do
- 11: **for** $b = 1$ to B do
- 12: Use the steady state selection procedure to select a population of x data objects
- 13: Assign these data objects to the k^{th} cluster of the b^{th} bee
- 14: Calculate $SICD$ value of the b^{th} bee.
- 15: **end for**
- 16: **for** $b = 1$ to B do
- 17: Find the probability of stickiness (P) for the b^{th} bee
- 18: Generate a random number r in the range 0 to 1 .
- 19: **if** $P < r$ then
- 20: Select a recruiter using roulette wheel selection technique and follow its solution.
- 21: **end if**
- 22: **end for**
- 23: Find the best partial solution (bee clusters with minimum $SICD$ value).
- 24: **end for**
- 25: Allocate the left-over data objects to the local best solution, and update centroids.
- 26: Find the $SICD$ value of the local best solution.
- 27: **if** Global best solution does not change for an adequate number of iterations t **then**
- 28: Perform crossover and mutation operation on the centroids of the local best solution, and find the $SICD$ value.
- 29: Update the local best solution if the new result is better than the previous, otherwise discard the new result.
- 30: **end if**
- 31: Perform switching operation of data objects between the clusters of local best solution.
- 32: **if** $t \neq 1$ then
- 33: **if** $SICD$ value at iteration $t < SICD$ value at iteration $t-1$ **then**
- 34: Update the global best solution with the

results of the local best solution.

35: **end if**

36: **end if**

37: **end for**

38: Return the global best solution.

Consider the following mainstream bee and its opposite bee defined with three cluster centers.

$$Bee_{mainstream} = ((c_{11}, c_{12}, \dots, c_{1m}), (c_{21}, c_{22}, \dots, c_{2m}), \dots;$$

$$Bee_{opposite} = ((c'_{11}, c'_{12}, \dots, c'_{1m}), (c'_{21}, c'_{22}, \dots, c'_{2m}), \dots;$$

Here, c_{ij} and c'_{ij} are the attributes of the centroids of mainstream bees and their opposite bees, that represents the j^{th} attribute of the i^{th} cluster center. Every c'_{ij} is generated using (2). Similarly, for all the centroids of each bee are found to create the corresponding opposite bees. Moreover, the number of mainstream bees, doubles the total number of bees. Therefore, we cannot create large number of mainstream bees initially. We have taken the total number of bees (B) as input parameter. First, $B/2$ number of mainstream bees initialized then another set of $B/2$ opposite bees generated.

The algorithm runs for K ($k=1,2,\dots,K$) number of clusters. In each iteration, all the bees build one cluster (k^{th}) at a time. So there is an inner loop that varies from ($b=1$ to B). For every bee b , the forward and backward passes generate the feasible solutions. After the outer loop (k loop) ends all the bees have a partial solution. Here each iteration of the outer loop (k loop) is considered as a stage.

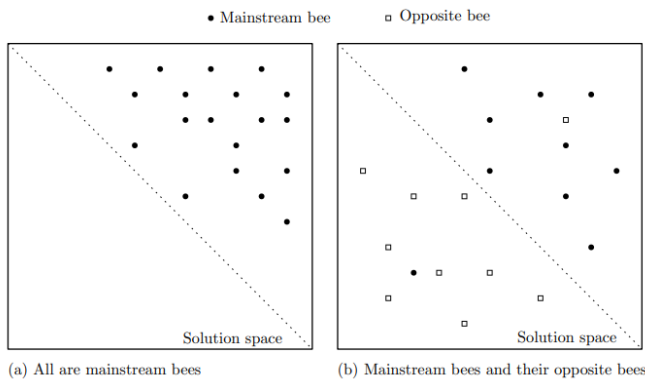


Fig. 2. Two different scenarios where (a) all bees are main stream bees chosen in same region (worst case) and (b) some mainstream and their opposite bees.

3) *Forward pass*: Once the mainstream bees initialized and their opposite bees created, all of them start exploring the solution space. In this process, they select some data objects from the dataset D and add them into their respective clusters based on a fitness (SICD) value. The exploration process of each bee depends on this selection procedure. For this, the OLIBCO algorithm uses a steady state selection approach [35]. The steady state searching process starts with creating an initial population of x data objects. This step repeats for M number of times. Here, M represents the number of moves. In each move, the search process refines the data objects to create a better population of x data objects. More the number

of moves (M), better the exploration. By increasing M the time complexity also increases. Again too small value of M implies less exploration that leads to slow convergence rate, so, it must be chosen carefully. After the forward pass completes, each bee has a feasible solution (a set of clusters). Now, all of them compute the strength of their solution using the SICD measure defined in (1).

4) *Backward pass*: In the backward pass, bees return to their hive, and share the information about their generated solution. Depending on the strength of the solution, the bee determine whether it will continue with its solution and be available as a recruiter, or will adopt (follow) some other's solution. The following equation defines the probability of a bee sticking to its solution [16,17].

$$P_b(k+1, t) = e^{-O_b(k,t)/(k \times t)}, \quad (3)$$

where b represents the current bee, k represents stage, and t represents iteration. $O_b(k,t)$ represents the normalized value of SICD of b^{th} bee and is computed using the equation (4) [16,17].

$$O_b(k, t) = \frac{SICD_b(k,t) - SICD_{min}(k,t)}{SICD_{max}(k,t) - SICD_{min}(k,t)}, \quad (4)$$

where $SICD_b$ is the SICD value of current bee b , $SICD_{max}$, and $SICD_{min}$ represents the maximum and minimum value of the SICD. From (3), we can observe P (probability of stickiness) and O_b are inversely proportional to each other, and the value of O_b depends on $SICD_b$. That means, when O_b increases the probability of a bee sticking to its solution decreases. The bee with smaller probability P may discard its solution and become a follower. If decision is to become a follower, then the bee will select another bee and adopts (copy) its solution. Every follower always wants to follow the best. This tendency of follower bees leads to local optima. The OLIBCO algorithm uses roulette wheel selection procedure [46] to avoid this. This gives every bee a fair chance to become a recruiter. It also gives diversity to the solution. This selection procedure first labels the surface of the wheel based on the proportion of the SICD values of each bee. Then the wheel spun. After it stops, the wheel pointer decides the result. Finally, the backward pass computes the SICD values of all the bees using (1).

5) *Allocation of remaining data objects*: After all the stages are complete, each bee has a partial clustering solution. But there is a possibility that some data objects are not allocated to any of the partial solutions. These left-over (unallocated) data objects are allocated to the best partial solution (among all the bees) using a single pass of the k-Means algorithm [47]. It takes the centroids of the best partial solution as initial cluster centroid and allocates the left-over data objects to the nearest cluster using the distance metric d . It also updates the centroids to the best candidate centroids. Note that we have used only one iteration of k-Means.

6) *Crossover and mutation operation*: The opposite bees and steady state selection procedure provides faster convergence. But after this, there is a possibility that the solution does not update for a longer period of time. It may have been trapped in a local optima. To take it out from the

trapped region (if there is any possibility) the algorithm applies the crossover and mutation [48] operation on the cluster centroids. Moreover, swapping should be between similar type attributes. We have used a one point crossover for swapping later half $m/2$ attributes of the consecutive centroids. We can do this swapping in any combination. As part of mutation operation, some or all the attributes of a centroid is substituted by another set of closer values. The main aim of the mutation step in our algorithm is to move the centroids in nearby locations to check the existence of any better solution. After the new centroids generated the SICD of the solution is computed again. If this new solution is better than the current local best, then the local best solution is updated.

7) *Switching operation of data objects*: In the previous steps at the allocation of remaining data objects and crossover and mutation, somehow the centroids are updated. As the data objects are assigned to the specific cluster based on Euclidian distance with the centroids, after centroids update, there may be some data objects wrongly placed in a cluster. The switching operation of data objects allows them to move to other clusters where its strength is better than the current one. The chance of switching is more for the data objects located in the boundary region of the cluster. This switching operation helps in further improvements in the result, thereby contributes to the faster convergence rate of the algorithm.

8) *Global best update*: The algorithm runs for T number of iterations. In each iteration, it finds a solution (local best). If the SICD value of the solution at iteration t ($SICD_t$) is less than the SICD value of the global best solution at iteration $t-1$ ($SICD_{t-1}$) then it updates the global best solution to the solution found in the current iteration. Then it continues to the next iteration with updated global best.

C. Asymptotic Analysis of OLIBCO Clustering Algorithm

The time complexity of an algorithm is a function of input parameters. This function can be generated by computing the sum of the number of times each step of the algorithm executes. Table I shows the step counts for each individual steps of the OLIBCO clustering algorithm. Thus, the time complexity function of the algorithm is:

$$f(T, K, B) = (Mx + x + 10)TKB + 7TK + (11 + c_1)T + 1$$

We can observe that in step 12 a factor of Mx multiplied with TKB . This is because in the forward pass the bees create a population of x (fixed) data objects. For this, they are allowed to make M number of moves in the solution space. With an increased number of moves M , the exploration time also increases, without much difference in the strength of the solution (after a certain point). Keeping this in mind, we have fixed the number of moves M . In step 25, a constant factor c_1 is multiplied; because the number of left-over data objects is unknown. As most of them are assigned to different clusters of different bees in the forward pass, it must be in small numbers (so considered as a constant). The step 28 performs crossover and mutation. As it operates between K cluster centers, both these operations step count is TK and for $SICD$ computation it is T . Now, the time complexity function can be rewritten as:

$$f(T, K, B) = y_1TKB + 7TK + y_2T + 1$$

Clearly, the highest growing term in this function is TKB . Hence, ignoring the constant factor the time complexity of the algorithm is $O(TKB)$. Moreover, the data set dimensions ($n \times m$) also has an important role, which is not considered above. Table I shows the step counts without considering dimension of the dataset n and m . Considering the dataset dimensions the time complexity of the algorithm becomes $O(TKBmn)$.

TABLE I. STEP COUNTS OF THE OLIBCO ALGORITHM

Step Number	Count	Step Number	Count
1	1	18	$T \times K \times B$
2	$T+1$	19	$T \times K \times B$
3	T	20	$T \times K \times B$ (worst case)
4	1	23	$T \times K$
6	$T-1$	25	$T \times c_1$
8	$T \times K \times B/2$	26	T
9	$T \times K \times B/2$	27	T
10	$T \times (K + 1)$	28	$T \times (2K + 1)$ (worst case)
11	$T \times K \times (B + 1)$	29	T (worst case)
12	$T \times K \times B \times (M \times x)$	31	$T \times K$
13	$T \times K \times B \times x$	32	T
14	$T \times K \times B$	33	$T - 1$
16	$T \times K \times (B + 1)$	34	$T - 1$
17	$T \times K \times B$	38	1

IV. SIMULATION AND COMPARATIVE ANALYSIS

The algorithm is implemented in Java with required parameters D , B , K and T . To realize the performance of the OLIBCO algorithm, we have analyzed it concerning different applications. The benchmark datasets from the UCI machine learning repository used for analysis purposes are shown in Table II. Validation of the performance is done through an analysis of $SICD$ values for different applications by varying the number of bees B and the number of moves M , followed by a comparative study of $SICD$ values with some existing optimization techniques for clustering. The results presented here are the average of 25 random instances of executions.

TABLE II. SPECIFICATION OF THE DATASETS USED

Datasets	Number of clusters	Number of attributes	Number of data objects
Iris	3	4	150 (50, 50, 50)
Glass	6	9	214 (70, 76, 17, 13, 9, 29)
Cancer	2	9	683 (444, 239)
CMC	3	9	1473 (629, 334, 510)

TABLE III. SICD VALUES BASED ON DIFFERENT NUMBERS OF BEES B0

Application	Number of bees	SICD
Iris	4	97.10
	10	96.90
	14	96.77
	18	96.80
Glass	4	226.19
	10	224.09
	14	224.67
	18	224.34
Cancer	4	2985.65
	10	2980.32
	14	2977.09
	18	2983.93
CMC	4	5682.67
	10	5676.97
	14	5678.07
	18	5684.89

Table III shows the *SICD* analysis for different applications on varying *B*. We found that the selection of the number of bees has a large impact on the *SICD* value of the clusters. A small number of bees indicate less exploration. That means there is a chance that some areas of solution space will remain unexplored, which may give poor results. This fact can be observed from the table when $B=4$. On the other hand, a large number of bees indicate too much exploration. Though the increased number of bees provides diversity, it comes with increased exploration time. Again it may not improve the results drastically (Table III, $B=18$). Therefore, compromising time complexity is not a good idea.

The steady-state selection procedure provides exploration power to the bees. Here in every move, the bee tries to improve its generated solution strength by placing a weaker data object with another stronger data object. Thus, the number of moves *M* (that a bee does) has also a huge impact on the algorithm's performance. Less number of bees has to make a large number of moves to explore the solution space whereas a large number of bees can handle it with comparatively less number of moves. Fig. 3 presents the *SICD* values of the generated clustering solution of the Cancer dataset by varying number of moves and keeping the number of bees (*B*) and initial population size (*x*) fixed. From the implementation results, we observed that when *M* is small the generated *SICD* value is high. With increased *M* value the *SICD* decreases and after a certain number of moves, it does not change much. The larger the initial population size *x*, the number of moves required is more. Depending on the dataset used the value of *x* can vary.

In optimization algorithms, the convergence rate is an important factor to analyze the performance. Fig. 4 displays the convergence analysis of the algorithm made for datasets used by varying the number of bees. Fig. 4 also shows that after the initial faster convergence the graph remains constant. The reason is that it is trapped in local optima (for the current centroid, it is the best result). The cross over and mutation operations are used in the algorithm to check this scenario and exploit the result if there is any possibility of improvement. This fact can be observed in $B=10$, and $B=14$ of Fig. 4(a), 4(c), $B=4$ of Fig. 4(b) and $B=18$ of Fig. 4(d).

Finally, we have made a comparison of the *SICD* value of the clustering solution for some existing algorithms in Table IV. The OLIBCO algorithm out performs *k*-Means and Classical PSO algorithm for Iris, Glass, Cancer, and CMC datasets except for the best cases of classical PSO of Iris dataset and *k*-Means of Glass dataset. The comparison is also made with IBCOCLUST [17], KIBCLUST [17], Hybrid I [17], Hybrid II [17], MBCO [16], MKCLUST [16], and KMCLUST[16]. It is found that for Iris dataset OLIBCO performs better than IBCOCLUST and the results are near to others. For Glass dataset, the average case *SICD* is better than IBCOCLUST, KIBCLUST, Hybrid I and MBCO, whereas on best case, it is better than MBCO, MKCLUST and KMCLUST. For Cancer dataset, the OLIBCO gives better results compared to others except average case of IBCLUST and Hybrid II and worst case of MBCO and MKCLUST. Again, for CMC dataset, OLIBCO outperforms MBCO, MKCLUST and KMCLUST.

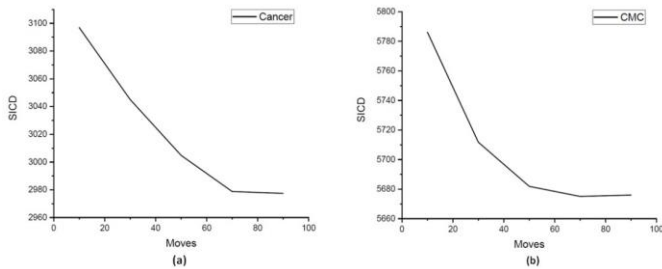


Fig. 3. SICD analysis of (a) Cancer dataset, (b) CMC dataset by varying number of moves.

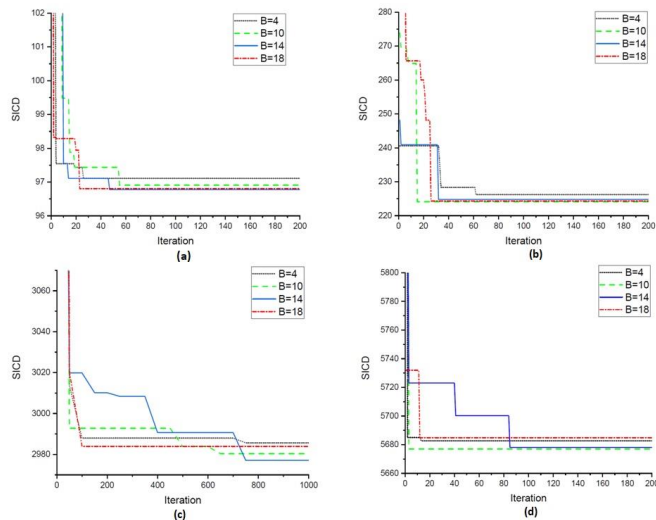


Fig. 4. Convergence graph of (a) Iris, (b) Glass, (c) Cancer, (d) CMC datasets.

Here we have observed that for all the applications, the algorithm converges to a certain level quickly, then it keep on exploiting the solution before final convergence result. Clearly, the OLIBCO algorithm has a faster convergence rate. The

results show that OLIBCO outperforms some of the existing algorithms in terms of *SICD*. Hence, gives better clustering result.

TABLE IV. A COMPARATIVE ANALYSIS OF SICD VALUES WITH SOME EXISTING ALGORITHMS

Applications	Measure	k-Means [16,17]	Classical PSO [16,17]	IBCOCLUST [17]	KIBCLUST [17]	Hybrid I [17]	Hybrid II [17]	MBCO [16]	MKCLUST [16]	KMCLUST [16]	OLIBCO
Iris	Best	97.33	96.01	97.22	96.40	96.33	95.10	94.14	95.01	95.19	96.77
	Avg	106.05	100.01	97.27	96.40	96.38	95.14	96.03	96.01	95.98	96.77
	Worst	120.45	117.81	-	-	-	-	104.22	201.00	200.10	97.88
Glass	Best	215.68	270.12	214.85	217.97	214.78	214.71	215.00	215.00	215.23	220.14
	Avg	260.40	289.31	225.19	226.34	226.59	221.50	225.00	220.00	221.00	224.30
	Worst	-	332.00	-	-	-	-	230.00	333.00	332.00	281.17
Cancer	Best	2987.00	2974.14	2976.22	2980.15	2976.24	2976.11	2965.25	2969.01	2971.01	2962.31
	Avg	2988.30	3329.22	2976.89	2980.15	2977.59	2976.24	2990.25	2985.23	2995.43	2977.43
	Worst	3521.50	-	-	-	-	-	3001.01	3076.10	3180.01	3083.83
CMC	Best	5842.20	5694.07	-	-	-	-	5680.12	5678.20	5678.40	5649.76
	Avg	5893.60	5729.11	-	-	-	-	5685.21	5684.80	5684.60	5678.42
	Worst	5934.40	5880.02	-	-	-	-	5798.20	5790.21	5689.70	5750.66

V. CONCLUSION AND FUTURE WORK

In this paper, a new OLIBCO algorithm is proposed as a potential solution for biomedical data clustering. It uses the opposite bees to shield the other directions of mainstream bees. This enhances the quality of the exploration result, thereby improving the convergence rate. The crossover and mutation operations along with the switching operation allow further exploitation of the solution and avoid being stuck in local optima. For validation of the clustering result, we have applied it for different benchmark applications from the UCI machine learning repository. The simulation results show that the algorithm has a faster convergence rate and possible exploitations. It is also observed that the algorithm converges to a certain level within 50 to 100 iterations for all the datasets used. After an initial faster convergence it gives enough chance for exploitations. From the analysis of results, it is clear that by adopting an optimal number of bees *B* and the number of moves *M* for exploration the algorithm give better performances. Further from the algorithm it is observed that there is better performance with a similar level of time complexity. The comparison made with different existing algorithms proves the proposed OLIBCO algorithm’s efficacy. Further investigations for higher dimensional datasets need to be explored in the future. The algorithm should be tested in different real-life applications of science and technology.

REFERENCES

[1] J. Han, J. Pei, and M. Kamber, Data mining: concepts and techniques. Elsevier, 2011.
 [2] A. K. Jain, “Data clustering: 50 years beyond k-means,” Pattern recognition letters, vol. 31, no. 8, pp. 651–666, 2010.

[3] Li, Xiaorong, and Zhinian Shu. "Research on Big Data Text Clustering Algorithm Based on Swarm Intelligence." *Wireless Communications and Mobile Computing* 2022 (2022).
 [4] R. Xu and D. Wunsch, “Survey of clustering algorithms,” IEEE Transactions on neural networks, vol. 16, no. 3, pp. 645–678, 2005.
 [5] X. Xiao and H. Huang, “A clustering routing algorithm based on improved ant colony optimization algorithms for underwater wireless sensor networks,” Algorithms, vol. 13, no. 10, p. 250, 2020.
 [6] O. Engin and A. G’u,cl’u, “A new hybrid ant colony optimization algorithm for solving the no-wait flow shop scheduling problems,” Applied Soft Computing, vol. 72, pp. 166–176, 2018.
 [7] D. L. Reddy, C. Puttamadappa, and H. Suresh, “Merged glowworm swarm with ant colony optimization for energy efficient clustering and routing in wireless sensor network,” Pervasive and Mobile Computing, vol. 71, p. 101338, 2021.
 [8] J. S. Raj, “Machine learning based resourceful clustering with load optimization for wireless sensor networks,” Journal of Ubiquitous Computing and Communication Technologies (UCCT), vol. 2, no. 01, pp. 29–38, 2020.
 [9] S. Choudhary, S. Sugumaran, A. Belazi, and A. A. El-Latif, “Linearly decreasing inertia weight pso and improved weight factor-based clustering algorithm for wireless sensor networks,” Journal of Ambient Intelligence and Humanized Computing, pp. 1–19, 2021.
 [10] M. Alswaitti, M. Albughdadi, and N. A. M. Isa, “Density-based particle swarm optimization algorithm for data clustering,” Expert Systems with Applications, vol. 91, pp. 170–186, 2018.
 [11] Bao, Zhanbiao. "Secure Clustering Strategy Based on Improved Particle Swarm Optimization Algorithm in Internet of Things." *Computational Intelligence and Neuroscience* 2022 (2022).
 [12] B. M. Sahoo, H. M. Pandey, and T. Amgoth, “Gapso-h: A hybrid approach towards optimizing the cluster based routing in wireless sensor network,” Swarm and Evolutionary Computation, vol. 60, p. 100772, 2021.
 [13] E. El-Shafeiy, K. M. Sallam, R. K. Chakraborty, and A. A. Abohany, “A clustering based swarm intelligence optimization technique for the internet of medical things,” Expert Systems with Applications, vol. 173, p. 114648, 2021.

- [14] F. Zabihi and B. Nasiri, "A novel history-driven artificial bee colony algorithm for data clustering," *Applied Soft Computing*, vol. 71, pp. 226–241, 2018.
- [15] J. Revathi, V. Eswaramurthy, and P. Padmavathi, "Hybrid data clustering approaches using bacterial colony optimization and k-means," in *IOP Conference Series: Materials Science and Engineering*, vol. 1070, no. 1. IOP Publishing, 2021, p. 012064.
- [16] P. Das, D. K. Das, and S. Dey, "A modified bee colony optimization (mbco) and its hybridization with k-means for an application to data clustering," *Applied Soft Computing*, vol. 70, pp. 590–603, 2018.
- [17] R. Forsati, A. Keikha, and M. Shamsfard, "An improved bee colony optimization algorithm with an application to document clustering," *Neurocomputing*, vol. 159, pp. 9–26, 2015.
- [18] B. N. Chebouba, M. A. Mellal, and S. Adjerid, "Fuzzy multiobjective system reliability optimization by genetic algorithms and clustering analysis," *Quality and Reliability Engineering International*, vol. 37, no. 4, pp. 1484–1503, 2021.
- [19] S. Verma, N. Sood, and A. K. Sharma, "Genetic algorithm-based optimized cluster head selection for single and multiple data sinks in heterogeneous wireless sensor network," *Applied Soft Computing*, vol. 85, p. 105788, 2019.
- [20] M. Alswaiti, M. Albughdadi, and N. A. M. Isa, "Variance-based differential evolution algorithm with an optional crossover for data clustering," *Applied Soft Computing*, vol. 80, pp. 1–17, 2019.
- [21] A. M. Jabbar, K. R. Ku-Mahamud, and R. Sagban, "Ant-based sorting and aco-based clustering approaches: A review," in *2018 IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE)*. IEEE, 2018, pp. 217–223.
- [22] W. Liu, Z. Wang, X. Liu, N. Zeng, and D. Bell, "A novel particle swarm optimization approach for patient clustering from emergency departments," *IEEE Transactions on Evolutionary Computation*, vol. 23, no. 4, pp. 632–644, 2018.
- [23] H. Hakli and M. S. Kiran, "An improved artificial bee colony algorithm for balancing local and global search behaviors in continuous optimization," *International Journal of Machine Learning and Cybernetics*, pp. 1–26, 2020.
- [24] Dorigo, Marco, Mauro Birattari, and Thomas Stutzle. "Ant colony optimization." *IEEE computational intelligence magazine* 1.4 pp. 28-39, 2006.
- [25] Kennedy, J., and R. Eberhart. "Particle swarm optimization In: Proceedings of ICNN 95-International Conference on Neural Networks, 1942–1948." IEEE, Perth. [https://doi.org/10.1109/icnn\(1995\)](https://doi.org/10.1109/icnn(1995)).
- [26] 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.
- [27] Chebouba, Billal Nazim, Mohamed Arezki Mellal, and Smail Adjerid. "Fuzzy multiobjective system reliability optimization by genetic algorithms and clustering analysis." *Quality and Reliability Engineering International* 37.4, pp. 1484-1503, 2021.
- [28] S. M. Bozorgi, M. R. Hajiabadi, A. A. R. Hosseinabadi, and A. K. Sangaiah, "Clustering based on whale optimization algorithm for iot over wireless nodes," *Soft Computing*, vol. 25, no. 7, pp. 5663–5682, 2021.
- [29] Singh, Hakam, et al. "An enhanced whale optimization algorithm for clustering." *Multimedia Tools and Applications* 82.3, pp. 4599-4618, 2023.
- [30] S. I. Boushaki, N. Kamel, and O. Bendjeghaba, "A new quantum chaotic cuckoo search algorithm for data clustering," *Expert Systems with Applications*, vol. 96, pp. 358–372, 2018.
- [31] N. Mittal, S. Singh, U. Singh, and R. Salgotra, "Trust-aware energy-efficient stable clustering approach using fuzzy type-2 cuckoo search optimization algorithm for wireless sensor networks," *Wireless Networks*, vol. 27, no. 1, pp. 151–174, 2021.
- [32] A. Srivastava and D. K. Das, "A new kho-kho optimization algorithm: An application to solve combined emission economic dispatch and combined heat and power economic dispatch problem," *Engineering Applications of Artificial Intelligence*, vol. 94, p. 103763, 2020.
- [33] C. Iorio, G. Frasso, A. D'Ambrosio, and R. Siciliano, "A p-spline based clustering approach for portfolio selection," *Expert Systems with Applications*, vol. 95, pp. 88–103, 2018.
- [34] Y. Xu, H. Chen, A. A. Heidari, J. Luo, Q. Zhang, X. Zhao, and C. Li, "An efficient chaotic mutative moth-flame-inspired optimizer for global optimization tasks," *Expert Systems with Applications*, vol. 129, pp. 135–155, 2019.
- [35] F. Pourpanah, Y. Shi, C. P. Lim, Q. Hao, and C. J. Tan, "Feature selection based on brain storm optimization for data classification," *Applied Soft Computing*, vol. 80, pp. 761–775, 2019.
- [36] P. Das, D. K. Das, and S. Dey, "A new class topper optimization algorithm with an application to data clustering," *IEEE Transactions on Emerging Topics in Computing*, 2018.
- [37] Ahandani, Morteza Alinia, and Hosein Alavi-Rad. "Opposition-based learning in the shuffled differential evolution algorithm." *Soft computing* 16.8, pp. 1303-1337, 2012.
- [38] Gao, X. Z., et al. "A hybrid optimization method of harmony search and opposition-based learning." *Engineering Optimization* 44.8, pp. 895-914, 2012.
- [39] Ma, Xiaoliang, et al. "MOEA/D with opposition-based learning for multiobjective optimization problem." *Neurocomputing* 146, pp. 48-64, 2014.
- [40] Wang, Hui, et al. "Enhancing particle swarm optimization using generalized opposition-based learning." *Information sciences* 181.20, pp. 4699-4714, 2011.
- [41] Boroujeni, Sayed Pedram Haeri, and Elnaz Pashaei. "A Hybrid Chimp Optimization Algorithm and Generalized Normal Distribution Algorithm with Opposition-Based Learning Strategy for Solving Data Clustering Problems." *arXiv preprint arXiv:2302.08623* (2023).
- [42] Abualigah, Laith, et al. "Augmented arithmetic optimization algorithm using opposite-based learning and lévy flight distribution for global optimization and data clustering." *Journal of Intelligent Manufacturing* (2022): 1-39.
- [43] Bharti, Kusum Kumari, and Pramod Kumar Singh. "Opposition chaotic fitness mutation based adaptive inertia weight BPSO for feature selection in text clustering." *Applied Soft Computing* 43 (2016): 20-34.
- [44] Tizhoosh, Hamid R. "Opposition-based learning: a new scheme for machine intelligence." *International conference on computational intelligence for modelling, control and automation and international conference on intelligent agents, web technologies and internet commerce (CIMCA-IAWTIC'06)*. Vol. 1. IEEE, 2005.
- [45] S. Dhargupta, M. Ghosh, S. Mirjalili, and R. Sarkar, "Selective opposition based grey wolf optimization," *Expert Systems with Applications*, p. 113389, 2020.
- [46] A. Aibinu, H. B. Salau, N. A. Rahman, M. Nwohu, and C. Akachukwu, "A novel clustering based genetic algorithm for route optimization," *Engineering Science and Technology, an International Journal*, vol. 19, no. 4, pp. 2022–2034, 2016.
- [47] A. Likas, N. Vlassis, and J. J. Verbeek, "The global k-means clustering algorithm," *Pattern recognition*, vol. 36, no. 2, pp. 451–461, 2003.
- [48] D. E. Goldenberg, "Genetic algorithms in search, optimization and machine learning," 1989.

Fuzzy Rank-Based Ensemble Model for Accurate Diagnosis of Osteoporosis in Knee Radiographs

Saumya Kumar¹, Puneet Goswami¹, Shivani Batra¹

Department of Computer Science & Engineering, SRM University, Delhi NCR, Sonipat, India¹

Abstract—The main factor in fractures among seniors and women post-menopausal is osteoporosis, which decreases the density of bones. Finding a low-cost diagnostic technology to identify osteoporosis in its initial stages is imperative considering the substantial expenses of diagnosis and therapy. The simplest and most widely used imaging method for detecting bone diseases is X-ray radiography, however, it is problematic to manually examine X-rays for osteoporosis as well as to identify the essential components and choose elevated classifiers. To categorize x-ray pictures of knee joints into normal, osteopenia, and osteoporosis condition categories, authors present a process in this investigation that uses three convolutional neural networks (CNN) architectures, i.e., Inception v3, Xception, and ResNet 18, to create an ensemble-based classifier model. The suggested ensemble approach employs a fuzzy rank-based unification of classifiers by taking into account two distinct parameters on the decision scores produced by the aforementioned base classifiers. Contrary to the straightforward fusion strategies that have been mentioned in the literature, the suggested ensemble methodology finalizes predictions on the test specimens by considering the confidence in the recommendations of the base learners. A 5-fold cross-validation approach has been employed to assess the developed framework using a benchmark dataset that has been made accessible to the general population. The suggested model yields an accuracy rate of 93.5% with a loss of 0.082. Further, the AUC is observed to be 98.1, 97.9 and 97.3 for normal, osteopenia and osteoporosis, respectively. The results demonstrate the model's usefulness by outperforming various state-of-the-art approaches.

Keywords—Convolutional Neural Network; diagnosis; knee; osteoporosis; transfer learning models; X-rays

I. INTRODUCTION

Osteoporosis is a serious disease that affects 200 million women globally and 9% of Americans over fifty in the US [1]. In affluent countries, one in three individuals may suffer from an OCF (osteoporotic compression fracture) [1]. After the primary injury, the risk of subsequent fractures significantly increases [2, 3]. A worse life expectancy and a higher mortality rate are both associated with even one OCF [4].

Bone mineral density (BMD) is measured using the Dual Energy X-ray Absorptiometry (DXA) procedure, which establishes the T-score and Z-score values recommended by the WHO for various phases of osteoporosis [5]. However, it has several drawbacks, such as areal estimations as well as the expensive and limited availability of the technology. The Quantitative Ultrasound System (QUS) [6], Computed Tomography (CT) [7], and Magnetic Resonance Imaging (MRI) [8] are further imaging modalities used to identify

osteoporosis. Whereas CT provides 3D geometric scanning with quantitative measures but has an intense radiation exposure and does not meet the WHO's criterion of osteoporosis diagnosis, MRI is a 3 T enhanced bone microarchitecture optical technique but is highly expensive and has lesser pixel density [8]. Although QUS is easy to use, non-invasive, compact, and economical and employs acoustic signals to investigate bones, it is site-specific and lacks substantial empirical support [6]. Given these constraints, a low-cost, easily accessible, and reliable detection system is essential. To create computer-aided diagnostic (CAD) systems, the investigators used the latest developments in machine vision to analyze medical images and computer algorithms.

Numerous CAD systems, which include deep learning at multiple bone locations like the joint, vertebrae, palm, and dental, are suggested for osteoporosis assessment [9, 10, 11, 12]. However, little research has been conducted to diagnose knee osteoporosis. Being the joint that supports the body's weight and facilitates motion, the knee experiences the most strain. Women are more susceptible to tibial and fibular fractures, which raises the likelihood of osteoporotic fractures around the knee including in elderly society [13]. A significant 1-year fatality rate of 22% is recorded in senior patients who suffer femoral injuries, with poorer mobility and bad living conditions [14], and it is anticipated that almost half of the knee fractures happen in individuals who are aged over 50 years. To avoid fractures and save healthcare expenditures, an early diagnosis method is required to determine the incidence of osteoporosis in the knee bone [15].

Convolutional neural network (CNN) approaches based on deep learning are becoming increasingly widely used in recent years within CAD systems for medical image interpretation [16] owing to their state-of-the-art performance in identifying a variety of illnesses from pictures, including brain tumors [17], respiratory disease [18, 19], cancer [20] and others. In terms of classifying medical pictures, CNNs [21] have produced cutting-edge results. The fundamental issue with utilizing CNN learners is that they require a significant quantity of annotated data for training; however, it's extremely challenging to find a large-scale dataset in the health sector. Investigators have put forward the concept of transfer learning to overcome difficulties [18, 22]. In transfer learning, a CNN that has been trained on a large population and then provided with training on a lower dimension of a different issue makes use of the information learned from the larger dataset to quickly learn the characteristics of the lower dimension and so efficiently aid in image classification. Various models, however, may perform better on specific data configurations, meaning that some

categories in the dataset may have more precise categorization than others. Moreover, traditional rank-based ensemble approaches do not utilize the diversity of the forecast odds. The significant fact could thus go ignored as a consequence. This fact led authors to develop a unique technique in this study wherein authors quantified two crucial variables, i.e., prediction probability's proximity to 1 and its deviation from 1—and used all the knowledge accessible through different base classifiers. Furthermore, the proposed method combines all of these quantifiable elements to reach the final forecast, allowing it to handle the categorization issue more and produce an accurate conclusion.

Ensemble learning [23] is an approach, in which the assessment ratings of many learners are combined to forecast the ultimate target class of an input data set. An ensemble method aims to capture the key characteristics of each of its component models, outperforming each base classifier individually. These systems are reliable because ensembling reduces the range or scattering of the base models' estimates. By adding substantial bias to the contending base classifiers, the ensemble model's diversity in the predictive performance of the base classifiers is mitigated. The standard ensemble approach employed in literature uses pre-calculated ratings for the classifiers and accords equal value to all constituting models' classification results. The biggest drawback of such an ensemble is the creation of fixed ratings that are challenging to change throughout the test sample classification phase. However, the suggested fuzzy rank-based ensemble technique accounts for each base classifier's forecast rank for each unique test case independently. In this approach, the ensemble technique can produce improved and more precise classification results. In the current study, authors develop a fusion technique that combines the judgment values from three basic CNN classifiers, i.e., Inception v3 [24], Xception [25], and ResNet 18 [26], to build the proposed ensemble model.

A. Key Contributions

An end-to-end classifier employing just deep learning algorithms may not function adequately on new datasets due to the dearth of data accessible in the medical sector. To create an ensemble approach that includes the forecasts from other competing systems, the authors employ three transfer learning-based CNN models. Although straightforward fusion techniques like popular vote, balanced averaging, and others have been employed in the research, they don't consider the predictor's confidence when making assertions. By taking this into account when developing the statistical framework for the suggested technique, authors can outperform basic ensemble techniques that are often employed for diagnosis. The current study's highlighting accomplishments are listed below.

- The implementation of ensemble methods employing the three base classifiers, i.e., Inception v3 [24], Xception [25], and ResNet 18 [26] improves the effectiveness of the whole system for prediction on the limited amount of accessible data.
- The fuzzy rankings of the classes in the assessment scores are determined by applying two non-linear operations of various concavities in the ensemble

approach that is being presented. The lower rank serves to identify the anticipated class after computing the sum of the products of the three base classifiers' ratings.

- The employment of two non-linear operations guarantees that the confidence in the classifiers' forecasts is taken into account in the derivation of the rankings, producing more accurate recommendations.
- The method authors use to measure the difference between the forecasted and anticipated values is unique. The suggested ensemble model's improvement in accuracy is also significant.
- Regarding classification precision and sensitivity, the suggested approach surpasses various cutting-edge techniques on the benchmark knee x-ray osteoporosis image dataset [9].

B. Section Division

The manuscript is hereafter divided into sections. Section II presents the literature survey. Section III presents the details of the proposed fuzzy rank-based ensemble model. Section IV highlights the experiments done and the results achieved. Finally, Section V concludes this research.

II. LITERATURE SURVEY

Deep convolution neural networks (DCNN) in particular illustrate state-of-the-art achievements in illness identification [27]. Several investigators have developed an osteoporosis assessment method from various kinds of images with effectiveness using machine learning techniques [9]. The authors have covered the most recent advances in DCNN-based osteoporosis assessment in this segment. Osteoporosis from phalanges has been detected using DCNN on X-ray scans [28]. Researchers acquired a decent identification rate using three-fold cross-validation for assessment.

In [29], researchers used feature selection based on wrapping to compare several classification schemes for osteoporosis diagnosis. To identify osteoporotic fracture risk, Naoufami et al. [30] recommended DCNN in their study (VF). After using computed tomography scans to derive logical characteristics, the system's efficiency has been compared to that of working radiologists, and similar outcomes were obtained. DCNN was utilized by Derkatch et al. [31] to precisely identify vertebral fractures in DXA pictures. Krishnaraj et al. [32] used CT scans of the vertebrae to separate people into osteoporotic and non-osteoporotic groups. They acquired high accuracy while segmenting CT images using U-net CNN. Fang et al. [33] also used vertebral CT scans to look for osteoporosis. They distinguished between normal and osteoporotic vertebrae using the DenseNet-121. The spine X-ray characteristics have been retrieved by Lee et al. [34] using CNN architectures, and the results were then sent to classifiers for categorization. They used VGG for extracting features and random forest for classifying to reach the highest accuracy rate of 71%. Yasaka et al. [35] employed CT scans of the abdomen to estimate the BMD of the lumbar vertebrae. They discovered a strong association between the DXA BMD and the anticipated BMD from CNN. Sollmann et al. [36] examined CT images of the spine and used CNN to calculate the

volumetric bone mineral density. Researchers observed that CNN provides good diagnostic performance when they evaluated the findings of the volumetric bone mineral density acquired from conventional CT.

By using a CNN and dental panoramic radiographs (DPRs), Lee et al. [37] have been able to detect osteoporosis from the tooth. This DCNN outperformed the outcomes of oral and maxillofacial radiologists. DPRs were employed by [38] to identify osteoporosis as well. To boost the CNN classifier's classification accuracy, they utilize the VGG-16 classifier and applied transfer learning to it. AlexNet Yu et al. [39] employed CNN to identify osteoporosis in dental panoramic radiography. They accurately divided the DPRs into osteoporotic and non-osteoporotic groups; however, they didn't only include the osteopenia group. Sukegawa et al. [40] also investigated DPRs for osteoporosis detection using CNNs and showed good results. The recognition rate has been further enhanced by the addition of prognostic factors. Deniz et al. [41] examined the MRIs of the proximal femur to look for osteoporosis. To quantify fracture risk and evaluate the condition of the bone, they segmented the proximal femur using DCNN.

Using X-ray scans of the pelvis, Liu et al. [42] identified osteoporosis. They derived the analytical expression from the softmax of the suggested U-net model, which employs X-rays to identify osteoporosis by analyzing the deep characteristics of the medullary joint. The photos of the osteoporosis and bone mass loss groups in this investigation are inadequately diagnosed. Utilizing CNN, Yamamoto et al. [43] identified osteoporosis in hip X-rays. Researchers discovered that adding clinical variables to the scans enhanced performance, with EfficientNet CNN achieving the highest results. Tecele et al. employed the AlexNet Classifier to get an osteoporosis diagnosis [44]. They identified the osteoporotic and non-osteoporotic scans from the segmented second metacarpal area using X-ray scans of the hand. He et al. [45] examined the knee X-rays and suggested using two radiographic criteria for assessing bone strength: cortical bone thickness and distal femoral cortex. To train the CNN networks, the BMD and T-score values produced by the QUS method have been demonstrated to significantly correlate with these characteristics. Also when learned on a limited population, the CNN performs well because of transfer learning.

The present work suggests an ensemble learning method where the eventual selection is reached after taking decisions from many models into account. Sarwar et al. [46] employed a mean likelihood-based ensemble, while Xue et al. [47] used a consensus voting-based ensemble approach to examine several straightforward fusion techniques. Nevertheless, these imprecise ensemble models utilize predetermined or constant weights linked to the base classifiers and do not allow for the predictability of results. In consideration of this, the authors provide a unique ensemble approach in this study that combines the decision ratings across three CNN-based base classifiers, Inception v3 [24], Xception [25], and ResNet 18 [26], while also accounting for the base learners' level of confidence in their judgments.

III. PROPOSED MODEL

This section briefly discusses the base classifiers authors utilize and the essential customization authors perform to the fundamental models before going into more depth about how the suggested fuzzy rank-based fusion of the basis learners' confidence scores would be implemented. The goal in this case is to fully use all of the confidence factors produced by the basic classifiers by converting them into non-linear operations. One of the projected values indicates compliance with or proximity to 1, and the other one indicates divergence from 1. The standard ranking systems' flaw of not taking into account the aforementioned fact [48] and potentially producing inaccurate results is addressed by the suggested technique. Three base classifiers are used in the current work, and the clinical image dataset is used to test our methodology. In the beginning, authors train the base classifiers (personalization using pre-trained architectures learned on ImageNet [49]), and then we collect the confidence values. Then, to create non-linear fuzzy rankings and a merged rating that allows us to calculate the overall divergence from the predicted, the authors translate the scores onto two distinct functions with distinct concavities. A lower deviation indicates greater assurance in a given class. The winning group is given the ultimate class value, and it has the lowest divergence value. The proposed model is presented in Fig. 1.

A. Pre-Processing Input

The data fed to the proposed model is pre-processed in two steps, i.e., normalization and data augmentation.

- **Normalization:** Transforming picture data pixels to a specified range, such as (0, 1) or (-1, 1) is done through normalization. Most pictures have pixel values between 0 and 255. Large values can impede or slow the training phase in CNN. Thus, image normalization is advised as a best practice so that the values of pixels vary from 0 to 1.
- **Data Augmentation:** When using CNN models, it is crucial to make sure the system receives enough training data. Data augmentation is the process of applying multiple adjustments to source photographs to produce several changed versions of the same image. However, because of the augmentation techniques utilized, each duplicate is unique in a certain manner from others.

B. Base Classifiers

The proposed model utilizes three pre-trained CNN architectures as base classifiers, i.e., Inception v3, Xception, and ResNet 18.

- **Inception v3:** Among the most prevalent deep learning models is Inception v3, which is a member of the Inception group and makes use of several enhancements to address issues with earlier Inception models [50]. These enhancements involve using a supplementary classification model, factorized convolution operations, batch normalization, the RMSProp optimization method, and label smoothing. It creates feature maps in numerous aspects and layers from an input picture with

the proportions $299 \times 299 \times 3$. The Inception v3 inception block gives us the option to use many feature extraction filters from a unified feature space. For more

thorough feature extraction, these characteristics with various filters are combined and transmitted to the following layer.

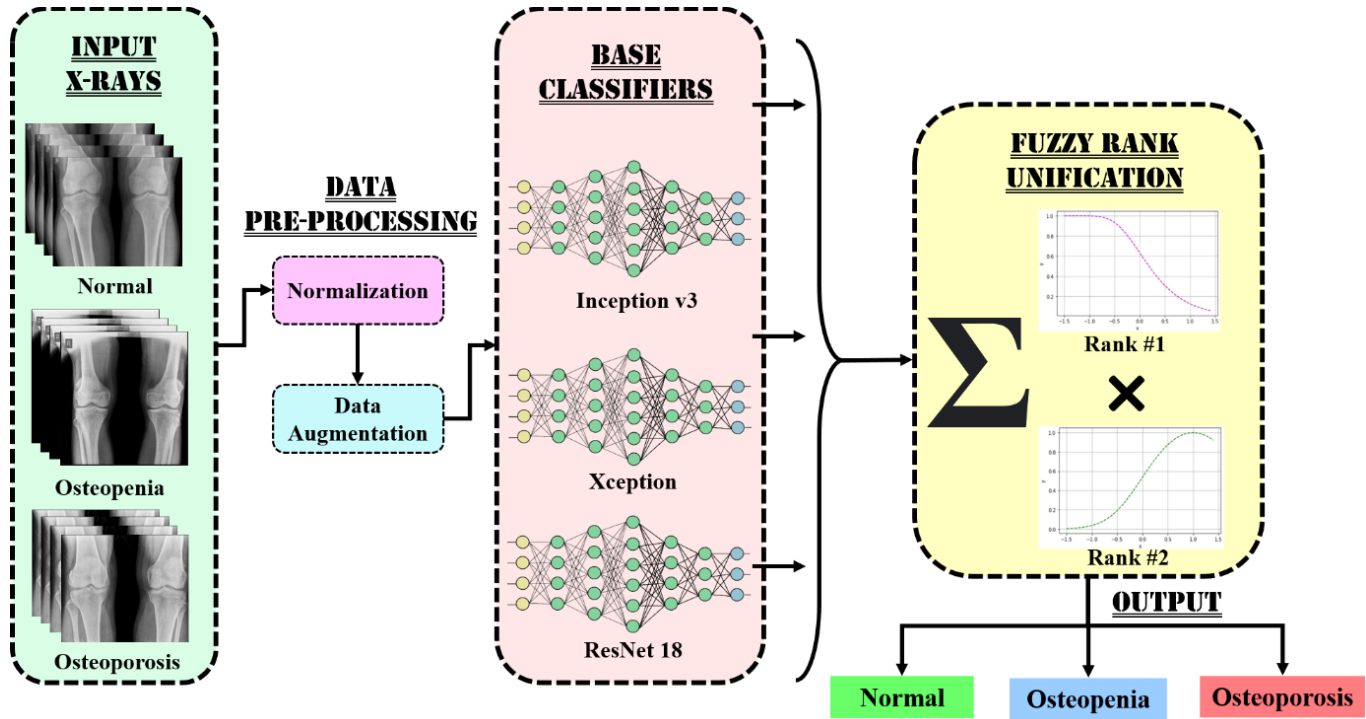


Fig. 1. Proposed model.

- **Xception:** The Inception v3 framework served as a basis for the Xception design created by Chollet et al. [25], which employs the same amount of design variables as the latter but does so more effectively. They demonstrated that the inception units fall in the center of a discrete spectrum, with pointwise convolutions and depth-wise separable convolutions at its two ends. To improve classification accuracy at the same computational load, they chose to substitute the inception modules with depth-wise separable convolutions.
- **ResNet 18:** ResNet, also known as Residual Network, is a structure that employs residual mapping and is particularly successful against the "degradation problem" in convolutional models. It was first presented in 2015. The CNN model's optimization process is enhanced by the residual learning technique. The ResNet-18 has been pre-trained on the ImageNet dataset, like the majority of widely used image-processing CNN models. It accepts photos with a $3 \times 224 \times 224$ size as input, which is smaller than the inception v3 model's input size. The residual network might score higher the denser it is. The ResNet model is implemented at many depths, including ResNet-18, ResNet-34, ResNet-50, ResNet-101, and ResNet-110.

C. Fuzzy Rank Unification

The confidence score generated by various base classifiers is fused using fuzzy rank-based unification. To give rankings to the class likelihood forecasted by a base classifier, authors

employ a fuzzy ranking-based technique in which the likelihood scores are exposed to two non-linear operations: the exponential operation (Eq. (1)) and the hyperbolic tangent operation (Eq. (2)). Let C_{cf}^{op} be the confidence ratings for a base classifier, 'cf' corresponding to the output class, 'op'.

$$Rank_{1_cf} = 1 - e^{-e^{-2 \cdot \sum_{op \in \{1,2,3\}} C_{cf}^{op}}} \quad (1)$$

$$Rank_{2_cf} = 1 - \tanh\left(\frac{(1 - \sum_{op \in \{1,2,3\}} C_{cf}^{op})^2}{2}\right) \quad (2)$$

The two non-linear operators double the scores they provided. The same steps are done for every base learner, and the final scores are calculated by adding the rank products from every predictor. Authors employ two distinct operations with various concavities so that their outputs might be complementary. In this investigation, $1 - e^{-e^{-2 \cdot \sum_{op \in \{1,2,3\}} C_{cf}^{op}}}$ has a concave downhill shape in the range [0, 1]. The output rank1 value will attempt to shift towards 1 due to its negative slope in the [0, 1] range. $1 - \tanh\left(\frac{(1 - \sum_{op \in \{1,2,3\}} C_{cf}^{op})^2}{2}\right)$ is concave uphill in this definition [0, 1]. The resulting rank value will attempt to shift towards 0 due to its positive gradient in the [0, 1] range. About a specific confidence score acquired from a base classifier, the rank value is the result of incentive and divergence. Fusion comprises combining the several rankings connected to identification and selecting an alternate rank that will help make the ultimate choice using Eq. (3).

$$Rank_{cf} = \sum(Rank_{1_{cf}} * Rank_{2_{cf}}) \quad (3)$$

The major goal of employing two rankings is to take into account how closely and how far the predominant classification result deviates from the predicted outcome. Lower product value and a good result are correlated with reduced deviation. The final result of the ensemble classifier is therefore the class with the least value under this sum of products of rankings. The author's objective is to lower this product since the two non-linear operations have opposing concavities in the domain [0, 1]. As a consequence, a strong confidence rating leads the ranking of one operation to increase and the ranking of the other operation to decrease. This sum of products produces a lower result when a prediction's confidence rating is elevated. The rank calculated using Eq. (3) can be used to determine the final grade for each class. Using Eq. (4), the class with the lowest fused rank is identified and declared as the winner.

$$class_{final} = \min_{cf \in \{1,2,3\}} Rank_{cf} \quad (4)$$

IV. EXPERIMENTS AND RESULTS

A. Environmental Setup & Dataset

The authors developed a model to separate individuals with X-ray scans into normal, osteopenia, and osteoporotic groups. This scenario has been simulated using the Python language. The systems, procedures, modules, and resources of TensorFlow 2.0 have been created by the authors using an open-source deep learning methodology (plus Keras). Python has been utilized to complete the analysis. The tests were run on Google Collaboratory using a Tesla K80 GPU graphics card, an Intel i7-core CPU running at 3.6GHz, 16GB of Memory, and the 64-bit version of Windows 11.

The dataset has been obtained from Mendeley data that [9] contributed, and it has been released in August 2021. The dataset includes x-rays from 240 subjects, of whom 37 had normal bone density (with 18 men and 19 women), 154 had osteopenia (with 59 men and 95 women), and 49 had osteoporotic bone density (with 31 men and 18 women). Data augmentation in Python has been employed to statistically augment the dataset pictures. After statistical augmentation, the dataset now includes 323 patient radiographs of normal, 323 osteopenia, and 323 osteoporosis knee x-rays.

B. Performance Evaluation Metrics

The datasets provided for X-ray scans are used by the authors to assess the effectiveness of the suggested models as they study the classification of normal, osteopenia, and osteoporotic patients. The authors focus on four characteristics that are typical of CNNs for each structure, i.e., accuracy curve, loss curve, confusion matrix, and area under the curve (AUC).

The accuracy curves of the model show how well it is acquiring and interpreting. The discrepancy between training and testing accuracy is a metric of overfitting. The training time and model orientation are shown by the loss graphs. A significant gap between both the training and testing graphs illustrates the learning spectrum with training. A confusion matrix expresses a way of how well a classifier

performs in a group of testing datasets when the input variables are already known. Four essential terms are connected to every confusion matrix. [51].

- True Positives [TP]: These are instances where the affected person had the ailment despite the prediction being "yes".
- True Negatives [TN]: According to estimates, the answer is "no" and the samples are not contaminated.
- False Positives [FP]: The ailment is presumed to be present, although the patients do not. A Type I mistake may be used to describe this.
- False Negatives [FN]: Even though the model recommends "no", the condition still exists in persons. These are referred to as Type II mistakes. It is frequently used to represent crucial prediction statistics, enabling analysis and identifying relevant experimental patterns easier.

AUC is a productivity statistic that incorporates all feasible classification levels. One method of examining AUC is to assess the probability that the model ranks a random positive instance stronger than a random counter-example. The possibility of a random positive instance being positioned in front of a counterpoint chosen at random is represented by the AUC. The range of the AUC value is 0 to 1. The AUC of a framework with 100% erroneous estimates is 0.0, while the AUC of a system with 100% accurate estimates is 1.0.

C. Implementation

The proposed ensemble model is compared with the individual underlying base classifiers, i.e., Inception V3, Xception, and ResNet-18. The findings (classification accuracy and loss) for the underlying datasets for knee osteoporosis used in this work are shown in Table I, Fig. 2 (accuracy curve) and Fig. 3 (loss curve) by the individual base classifiers and proposed ensemble framework. The outcomes show that the suggested model performs well in terms of classification accuracy. The underlying dataset has a training duration per fold of 30 minutes.

Fig. 4 displays the confusion matrices that the individual base classifiers and developed ensemble models on the dataset utilized in this study have been able to produce. Fig. 5 displays the AUC of the proposed model. The suggested ensemble model may be employed as a plug-and-play paradigm in which new test pictures are fed into the model to provide forecasts using the ensemble method, ultimately assisting the expert doctors in making a more rapid and precise judgment.

TABLE I. COMPARISON WITH STATE-OF-THE-ART METHODS

Model	Accuracy	Loss
Inception V3	89.8	0.217
Xception	90.9	0.208
Resnet-18	91.4	0.207
Proposed Model	93.5	0.082

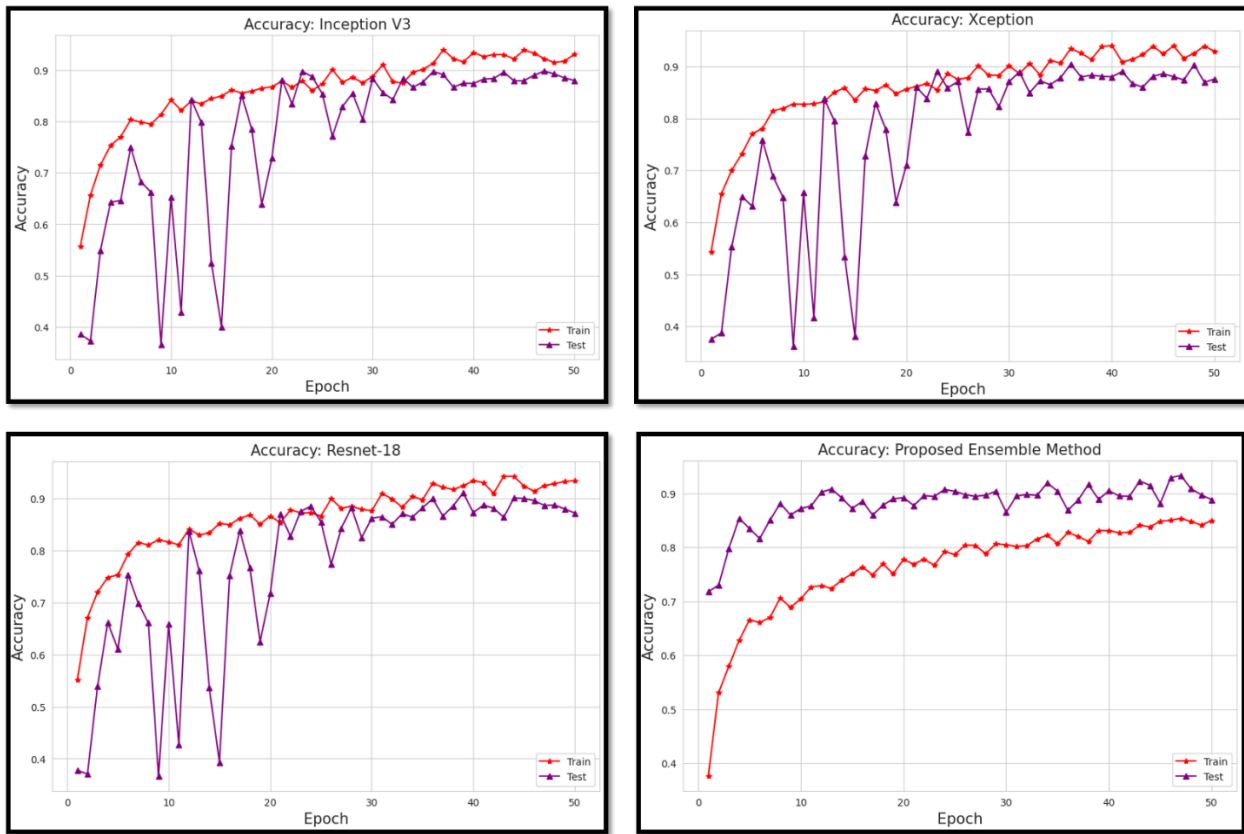


Fig. 2. Accuracy of various models.

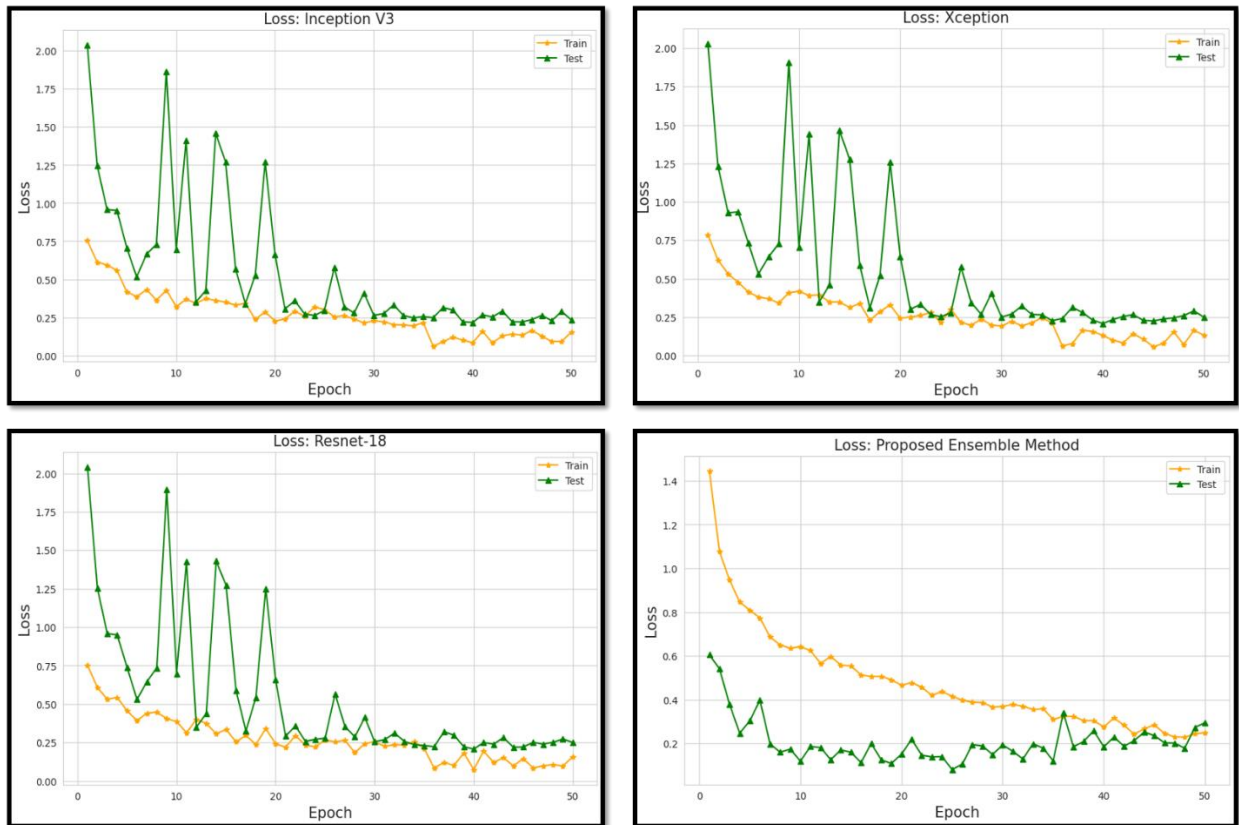


Fig. 3. Loss of various models.



Fig. 4. Confusion Matrices of various models.

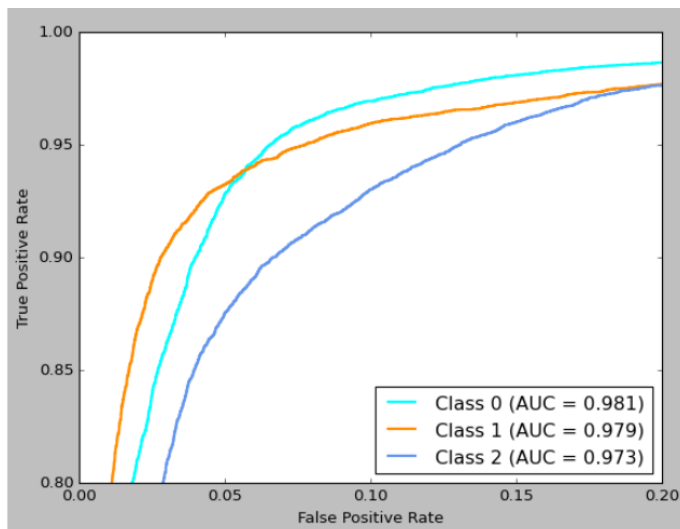


Fig. 5. AUC for the proposed model; class 0 represents normal, class 1 represents osteopenia, and class 2 represents osteoporosis.

V. CONCLUSION AND FUTURE WORK

The suggested model provides the best results of 93.5% classification accuracy while assessing osteoporosis knee X-ray dataset classification, which justifies the proposed model's efficacy. Further, the AUC is observed to be 98.1, 97.9 and 97.3 for normal, osteopenia and osteoporosis, respectively. Osteoporosis is a chronic illness that affects people around the globe and can result in long-term loss of mobility, physical injuries, excruciating pain, and even early mortality. In this study, the authors provide a framework for classifying pictures of normal, osteopenia, and osteoporosis by combining three conventional CNN-based classifiers. Two non-linear factors that serve to account for the base learners' level of confidence in their forecasts are used by the proposed ensemble model to construct rankings of the classification models.

The quick detection tool created for testing osteoporosis may operate like a plug-and-play paradigm with minimum assistance from experienced physicians, making it appropriate for use in the field. Due to insufficient picture quality or the existence of overlapping cells, the suggested ensemble model had difficulty correctly classifying a few images. So, the authors intend to tackle the requirement for picture pre-processing in the future. For separating overlapping cells,

authors may use image enhancement methods or cell slicing. To perform the ensemble, authors may additionally take into account ensembles of distinct base learners and investigate various rank-generating procedures.

REFERENCES

- [1] N. Wright, A. Looker, K. Saag, J. Curtis, E. Delzell, S. Randall and B. Dawson-Hughes, "The recent prevalence of osteoporosis and low bone mass in the United States based on bone mineral density at the femoral neck or lumbar spine," *Journal of bone and mineral research*, vol. 19, no. 11, pp. 2520-2526, 2014.
- [2] S. Roux, F. Cabana, N. Carrier, M. Beaulieu, P. April, M. Beaulieu and G. Boire, "The World Health Organization Fracture Risk Assessment Tool (FRAX) underestimates incident and recurrent fractures in consecutive patients with fragility fractures," *The Journal of Clinical Endocrinology & Metabolism*, vol. 99, no. 7, pp. 2400-2408, 2014.
- [3] A. Hodsman, W. Leslie, J. Tsang, and G. Gamble, "10-year probability of recurrent fractures following wrist and other osteoporotic fractures in a large clinical cohort: an analysis from the Manitoba Bone Density Program," *Archives of internal medicine*, vol. 168, no. 20, pp. 2261-2267, 2008.
- [4] J. Center, T. Nguyen, D. Schneider, P. Sambrook and J. Eisman, "Mortality after all major types of osteoporotic fracture in men and women: an observational study," *The Lancet*, vol. 353, no. 9156, pp. 878-882, 1999.
- [5] H. Dimai, "Use of dual-energy X-ray absorptiometry (DXA) for diagnosis and fracture risk assessment; WHO-criteria, T-and Z-score, and reference databases," *Bone*, vol. 104, pp. 39-43, 2017.
- [6] D. Hans and S. Baim, "Quantitative ultrasound (QUS) in the management of osteoporosis and assessment of fracture risk," *Journal of Clinical Densitometry*, vol. 3, no. 322-333, p. 20, 2017.
- [7] H. Jiang, C. Yates, A. Gorelik, A. Kale, Q. Song and J. Wark, "Peripheral Quantitative Computed Tomography (pQCT) measures contribute to the understanding of bone fragility in older patients with low-trauma fracture," *Journal of Clinical Densitometry*, vol. 21, no. 1, pp. 140-147, 2018.
- [8] U. Ferizi, H. Besser, P. Hysi, J. Jacobs, C. Rajapakse, C. Chen, P. Saha, S. Honig and G. Chang, "Artificial intelligence applied to osteoporosis: a performance comparison of machine learning algorithms in predicting fragility fractures from MRI data," *Journal of Magnetic Resonance Imaging*, vol. 49, no. 4, pp. 1029-1038, 2019.
- [9] I. Wani and S. Arora, "Computer-aided diagnosis systems for osteoporosis detection: A comprehensive survey," *Medical & biological engineering & computing*, vol. 58, pp. 1873-1917, 2020.
- [10] S. Batra and S. Sachdeva, "Organizing standardized electronic healthcare records data for mining," *Health Policy and Technology*, vol. 5, no. 3, pp. 226-242, 2016.
- [11] S. Batra and S. Sachdeva, "Pre-processing highly sparse and frequently evolving standardized electronic health records for mining," in *Handbook of Research on Disease Prediction Through Data Analytics and Machine Learning*, IGI Global, 2021, pp. 8-21.
- [12] S. Sachdeva, D. Batra and S. Batra, "Storage Efficient Implementation of Standardized Electronic Health Records Data," in *2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2020.
- [13] C. Court-Brown and B. Caesar, "Epidemiology of adult fractures: a review.," *Injury*, vol. 37, no. 8, pp. 691-697, 2006.
- [14] R. Stange and M. Raschke, "Osteoporotic distal femoral fractures: When to replace and how," in *Surgical and Medical Treatment of Osteoporosis*, 2020.
- [15] S. Wang, P. Wu, C. Lee, C. Shih, Y. Chiu and C. Hsu, "Association of osteoporosis and varus inclination of the tibial plateau in postmenopausal women with advanced osteoarthritis of the knee," *BMC Musculoskeletal Disorders*, vol. 22, pp. 1-8, 2021.
- [16] S. Yadav and S. Jadhav, "Deep convolutional neural network based medical image classification for disease diagnosis," *Journal of Big data*, vol. 6, no. 1, pp. 1-18, 2019.
- [17] S. Lu, S. Wang and Y. Zhang, "Detecting pathological brain via ResNet and randomized neural networks," *Heliyon*, vol. 6, no. 12, p. e05625, 2020.
- [18] S. Batra, H. Sharma, W. Boulila, V. Arya, P. Srivastava, M. Z. Khan and M. Krichen, "An Intelligent Sensor Based Decision Support System for Diagnosing Pulmonary Ailment through Standardized Chest X-ray Scans," *Sensors*, vol. 22, no. 19, p. Sensors, 2022.
- [19] S. Batra, R. Khurana, M. Z. Khan, W. Boulila, A. Koubaa and P. Srivastava, "A Pragmatic Ensemble Strategy for Missing Values Imputation in Health Records," *Entropy*, vol. 24, no. 4, p. 533, 2022.
- [20] A. Salau and S. Jain, "Adaptive diagnostic machine learning technique for classification of cell decisions for AKT protein," *Informatics in Medicine Unlocked*, vol. 23, p. 100511, 2021.
- [21] K. He, X. Zhang, S. Ren and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016.
- [22] A. Pathak, S. Batra and V. Sharma, "An Assessment of the Missing Data Imputation Techniques for COVID-19 Data," in *Proceedings of 3rd International Conference on Machine Learning, Advances in Computing, Renewable Energy and Communication: MARC 2021*, Singapore, 2022.
- [23] A. Pathak, S. Batra and H. Chaudhary, "Imputing Missing Data in Electronic Health Records," in *Proceedings of 3rd International Conference on Machine Learning, Advances in Computing, Renewable Energy and Communication: MARC 2021*, Singapore, 2022.
- [24] 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.
- [25] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017.
- [26] S. Ayyachamy, V. Alex, M. Khened and G. Krishnamurthi, "Medical image retrieval using Resnet-18," in *Medical imaging 2019: imaging informatics for healthcare, research, and applications*, 2019.
- [27] O. El-Gayar, L. Ambati and N. Nawar, "Wearables, artificial intelligence, and the future of healthcare," in *AI and Big Data's Potential for Disruptive Innovation*, IGI Global, 2020, pp. 104-129.
- [28] K. Hatano, S. Murakami, H. Lu, J. Tan, H. Kim and T. Aoki, "Classification of osteoporosis from phalanges CR images based on DCNN," in *2017 17th International Conference on Control, Automation and Systems (ICCAS)*, 2017.
- [29] H. Chang, Y. Chiu, H. Kao, C. Yang and W. Ho, "Comparison of classification algorithms with wrapper-based feature selection for predicting osteoporosis outcome based on genetic factors in a Taiwanese women population," *International journal of endocrinology*, vol. 2013, 2013.
- [30] N. Tomita, Y. Cheung and S. Hassanpour, "Deep neural networks for automatic detection of osteoporotic vertebral fractures on CT scans," *Computers in biology and medicine*, vol. 98, pp. 8-15, 2018.
- [31] S. Derkatch, C. Kirby, D. Kimelman, M. Jozani, J. Davidson and W. Leslie, "Identification of vertebral fractures by convolutional neural networks to predict nonvertebral and hip fractures: a registry-based cohort study of dual X-ray absorptiometry," *Radiology*, vol. 293, no. 2, pp. 405-411, 2019.
- [32] A. Krishnaraj, S. Barrett, O. Bregman-Amitai, M. Cohen-Sfady, A. Bar, D. Chetrit, M. Orlovsky and E. Elnekave, "Simulating dual-energy X-ray absorptiometry in CT using deep-learning

- segmentation cascade,” *Journal of the American College of Radiology*, vol. 16, no. 10, pp. 1473-1479, 2019.
- [33] Y. Fang, W. Li, X. Chen, K. Chen, H. Kang, P. Yu, R. Zhang, J. Liao, G. Hong and S. Li, “Opportunistic osteoporosis screening in multi-detector CT images using deep convolutional neural networks,” *European Radiology*, vol. 31, pp. 1831-1842, 2021.
- [34] S. Lee, E. Choe, H. Kang, J. Yoon and H. Kim, “The exploration of feature extraction and machine learning for predicting bone density from simple spine X-ray images in a Korean population,” *Skeletal radiology*, vol. 49, pp. 613-618, 2020.
- [35] K. Yasaka, H. Akai, A. Kunimatsu, S. Kiryu and O. Abe, “Prediction of bone mineral density from computed tomography: application of deep learning with a convolutional neural network,” *European radiology*, vol. 30, pp. 3549-3557, 2020.
- [36] N. Sollmann, M. Löffler, M. El Husseini, A. Sekuboyina, M. Dieckmeyer, S. Rühling, C. Zimmer, B. Menze, G. Joseph, T. Baum and J. Kirschke, “Automated opportunistic osteoporosis screening in routine computed tomography of the spine: comparison with dedicated quantitative CT,” *Journal of Bone and Mineral Research*, vol. 37, no. 7, pp. 1287-1296, 2022.
- [37] J. Lee, S. Adhikari, L. Liu, H. Jeong, H. Kim and S. Yoon, “Osteoporosis detection in panoramic radiographs using a deep convolutional neural network-based computer-assisted diagnosis system: a preliminary study,” *Dentomaxillofacial Radiology*, vol. 48, no. 1, p. 20170344, 2019.
- [38] K. Lee, S. Jung, J. Ryu, S. Shin and J. Choi, “Evaluation of transfer learning with deep convolutional neural networks for screening osteoporosis in dental panoramic radiographs,” *Journal of clinical medicine*, vol. 9, no. 2, p. 392, 2020.
- [39] S. Yu, P. Chu, J. Yang, B. Huang, F. Yang, V. Megalooikonomou and H. Ling, “Multitask osteoporosis prescreening using dental panoramic radiographs with feature learning,” *J Smart Health*, 2019.
- [40] S. Sukegawa, A. Fujimura, A. Taguchi, N. Yamamoto, A. Kitamura, R. Goto, K. Nakano, K. Takabatake, H. Kawai, H. Nagatsuka and Y. Furuki, “Identification of osteoporosis using ensemble deep learning model with panoramic radiographs and clinical covariates,” *Scientific reports*, vol. 12, no. 1, pp. 1-10, 2022.
- [41] C. Deniz, S. Xiang, R. Hallyburton, A. Welbeck, J. Babb, S. Honig, K. Cho and G. Chang, “Segmentation of the proximal femur from MR images using deep convolutional neural networks,” *Scientific reports*, vol. 8, no. 1, p. 16485, 2018.
- [42] J. Liu, J. Wang, W. Ruan, C. Lin and D. Chen, “Diagnostic and gradation model of osteoporosis based on improved deep U-Net network,” *Journal of medical systems*, vol. 44, pp. 1-7, 2020.
- [43] N. Yamamoto, S. Sukegawa, A. Kitamura, R. Goto, T. Noda, K. Nakano, K. Takabatake, H. Kawai, H. Nagatsuka, K. Kawasaki and Y. Furuki, “Deep learning for osteoporosis classification using hip radiographs and patient clinical covariates,” *Biomolecules*, vol. 10, no. 11, p. 1534, 2020.
- [44] N. Teclé, J. Teitel, M. Morris, N. Sani, D. Mitten and W. Hammert, “Convolutional neural network for second metacarpal radiographic osteoporosis screening,” *The Journal of Hand Surgery*, vol. 45, no. 3, pp. 175-181, 2020.
- [45] Q. He, H. Sun, L. Shu, Y. Zhu, X. Xie, Y. Zhan and C. Luo, “Radiographic predictors for bone mineral loss: Cortical thickness and index of the distal femur,” *Bone & joint research*, vol. 7, no. 7, pp. 468-475, 2018.
- [46] A. Sarwar, V. Sharma and R. Gupta, “Hybrid ensemble learning technique for screening of cervical cancer using Papanicolaou smear image analysis,” *Personalized Medicine Universe*, vol. 4, pp. 54-62, 2015.
- [47] D. Xue, X. Zhou, C. Li, Y. Yao, M. Rahaman, J. Zhang, H. Chen, J. Zhang, S. Qi and H. Sun, “An application of transfer learning and ensemble learning techniques for cervical histopathology image classification,” *IEEE Access*, vol. 8, pp. 104603-104618, 2020.
- [48] M. Monwar and M. Gavrilova, “Multimodal biometric system using rank-level fusion approach,” *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 39, no. 4, pp. 867-878, 2009.
- [49] J. Deng, W. Dong, R. Socher, L. Li, K. Li and L. Fei-Fei, “Imagenet: A large-scale hierarchical image database,” in *2009 IEEE conference on computer vision and pattern recognition*, 2009.
- [50] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke and A. Rabinovich, “Going deeper with convolutions,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015.
- [51] D. Kermany, M. Goldbaum, W. Cai, C. Valentim, H. Liang, S. Baxter, A. McKeown, G. Yang, X. Wu, F. Yan and J. Dong, “Identifying medical diagnoses and treatable diseases by image-based deep learning,” *Cell*, vol. 172, no. 5, pp. 1122-1131, 2018.

A Novel Deep CNN-RNN Approach for Real-time Impulsive Sound Detection to Detect Dangerous Events

Nurzhigit Smailov¹, Zhandos Dosbayev², Nurzhan Omarov³, Bibigul Sadykova⁴, Maigul Zhekambayeva⁵, Dusmat Zhamangarin⁶, Assem Ayapbergenova⁷
Satbayev University, Almaty, Kazakhstan^{1, 2, 5, 7}
Al-Farabi Kazakh National University, Almaty, Kazakhstan^{3, 4}
Kazakh University of Technology and Business, Astana, Kazakhstan⁶

Abstract—In this research paper, we presented a novel approach to detect impulsive sounds in real-time using a combination of Deep CNN and RNN architectures. The proposed approach was evaluated using our collected dataset of impulsive sounds, and the results showed that it outperformed traditional audio signal processing methods in terms of accuracy and F1-score. The proposed approach has several advantages over traditional methods, including the ability to handle complex audio patterns, detect impulsive sounds in real-time, and improve its performance with a large dataset of labeled impulsive sounds. However, there are some limitations to the proposed approach, including the requirement for a large amount of labeled data to train effectively, environmental factors that may impact the accuracy of the detection, and high computational requirements. Overall, the proposed approach demonstrates the effectiveness of using a combination of Deep CNN and RNN architectures for impulsive sound detection, with potential applications in various fields such as public safety, industrial settings, and home security systems. The proposed approach is a significant step towards developing automated systems for detecting dangerous events and improving public safety.

Keywords—CNN; RNN; deep learning; impulsive sound; dangerous sound; artificial intelligence

I. INTRODUCTION

Impulsive sounds such as gunshots, explosions, and screams are a major source of concern in public places. These sounds can cause panic, fear, and danger to human life [1]. Hence, there is a pressing need to detect such sounds in real-time and alert the authorities to take immediate action. Traditional methods for detecting impulsive sounds involve using microphones and signal processing techniques [2]. However, these methods are prone to false positives and are not effective in real-time scenarios.

The recent advances in deep learning have shown promising results in detecting impulsive sounds. In particular, deep convolutional neural networks (CNNs) have shown remarkable performance in sound classification tasks [3-5]. The use of recurrent neural networks (RNNs) has also been shown to be effective in modeling sequential data such as audio signals [6]. Combining these two architectures can improve the accuracy of sound detection and allow for real-time detection of dangerous events.

In this research paper, we propose a deep learning approach that combines CNNs and RNNs for real-time impulsive sound detection. We aim to develop a system that can accurately detect dangerous events in public places and alert the authorities to take immediate action. The proposed approach is based on the following steps:

The first step in developing the proposed system is to collect and preprocess the data. We will use a publicly available dataset of impulsive sounds that contains a wide range of sounds such as gunshots, explosions, and screams. The dataset contains audio files of different lengths and formats. We will preprocess the data by converting the audio files to a standardized format, extracting features, and labeling the data.

The next step is to extract features from the audio signals. We will use Mel-frequency cepstral coefficients (MFCCs) as the feature representation. MFCCs have been widely used in sound classification tasks and have shown to be effective in capturing the spectral characteristics of sound signals. We will extract MFCC features from each audio file using a sliding window approach. This approach involves dividing the audio signal into small segments and computing the MFCC features for each segment.

The proposed approach combines deep CNNs and RNNs to classify the MFCC features extracted from the audio signals. The CNN is used to learn the spatial features of the MFCCs, while the RNN is used to capture the temporal dependencies between the features. The architecture of the proposed model is shown in Fig. 1.

The first layer of the model is a convolutional layer that applies filters to the MFCCs. This layer is followed by a batch normalization layer and a rectified linear unit (ReLU) activation function [7]. The output of the convolutional layer is then fed into a max-pooling layer that reduces the spatial dimensionality of the features.

The output of the max-pooling layer is then fed into a recurrent layer, which is a long short-term memory (LSTM) layer. The LSTM layer is used to model the temporal dependencies between the MFCC features. The output of the LSTM layer is then fed into a fully connected layer, which is used to map the features to the output classes. The output layer

uses a softmax activation function to output the probabilities of the different classes.

We will train the proposed model on the collected dataset using a cross-entropy loss function and the Adam optimizer. We will use a validation set to monitor the performance of the model and prevent overfitting. The performance of the model will be evaluated using standard metrics such as accuracy, precision, recall, and F1 score.

The final step is to implement the proposed system in real-time. We will use a microphone to capture the audio signals in real-time and feed them to the trained model for classification. The system will use a threshold-based approach to detect dangerous events. If the probability of a gunshot or explosion exceeds a certain threshold, the system will raise an alert and notify the authorities.

In this research paper, we proposed a deep learning approach that combines CNNs and RNNs for real-time impulsive sound detection. The proposed approach uses MFCC features extracted from audio signals and combines deep CNNs and RNNs to classify the features. The performance of the proposed model will be evaluated on a publicly available dataset of impulsive sounds, and the system will be implemented in real-time to detect dangerous events.

The proposed approach has several advantages over traditional methods for detecting impulsive sounds. It is more accurate and can be used in real-time scenarios. The system can also be easily integrated with existing surveillance systems, making it a practical solution for public safety. We believe that the proposed approach can make a significant contribution to the field of public safety and can help prevent dangerous events in public places.

II. RELATED WORKS

Impulsive sound detection is an important research area in the field of public safety. Several methods have been proposed for detecting impulsive sounds, including traditional signal processing techniques and machine learning-based approaches. In recent years, deep learning-based approaches have shown promising results in different areas from sport to technical sciences [8-10]. In this literature review, we discuss some of the recent studies on deep learning-based approaches for impulsive sound detection.

Convolutional neural networks (CNNs) have shown remarkable performance in sound classification tasks. In 2020, Radlak et al. proposed a deep CNN-based approach for speech recognition that achieved state-of-the-art performance on the TIMIT dataset [11]. Later, CNNs were used for environmental sound classification by Isac in 2021 [12]. In this study, Isac proposed a deep CNN-based approach that achieved an accuracy of 85.6% on the ESC-50 dataset, which contains 50 environmental sound classes.

CNNs have also been used for impulsive sound detection problem. In 2020, Ahmed and Allen proposed a deep CNN-based approach for gunshot detection [13]. In this study, Li et al. used a dataset of gunshot sounds recorded from different distances and angles. The proposed approach achieved a detection accuracy of 96.3%.

Recurrent neural networks (RNNs) have been shown to be effective in modeling sequential data such as audio signals. In 2023, Cho et al. proposed a sequence-to-sequence RNN-based approach for speech recognition that achieved state-of-the-art performance on several benchmark datasets [14]. Moreover, RNNs were used for environmental sound classification by Janani and Jebakumar in 2023 [15]. In this study, authors proposed a deep RNN-based approach that achieved an accuracy of 88.2% on the ESC-50 dataset.

RNNs have also been used for impulsive sound detection. In 2019, Cha et al. proposed a deep RNN-based approach for real-time gunshot detection problem [16]. In this study, authors used a dataset of gunshot sounds recorded from different distances and angles. The proposed approach achieved a detection accuracy of 95%.

Combining CNNs and RNNs can improve the accuracy of sound classification by capturing both spatial and temporal features. In Shi et al. proposed a deep CNN-RNN-based approach for environmental sound classification problem [17]. In this study, authors used a hybrid CNN-RNN architecture that combined the strengths of both architectures. The proposed approach achieved an accuracy of 89.3% in environmental sound classification on the ESC-50 dataset.

Combined CNN-RNN models have also been used for impulsive sound detection. Molina-Tenorio et al. proposed a deep CNN-RNN-based approach for gunshot detection [18]. In this study, Kim et al. used a dataset of gunshot sounds recorded from different distances and angles. The proposed approach achieved a detection accuracy of 96.5%.

Real-time impulsive sound detection is essential for public safety. Lee et al. proposed a real-time impulsive sound detection system based on a deep CNN-based approach [19]. In this study, Lee et al. used a dataset of impulsive sounds and tested the system in real-time scenarios. The proposed system achieved a detection accuracy of 98.5% and a processing speed of 1000 times real-time.

Huang et al. proposed a real-time impulsive sound detection system based on a deep RNN-based approach [20]. In this study, Li et al. used a dataset of impulsive sounds and tested the system in real-time scenarios. The proposed system achieved a detection accuracy of 97.2% and a processing speed of 42 milliseconds per frame.

Combining CNNs and RNNs can improve the accuracy of real-time impulsive sound detection problem. Dong and Wang proposed a deep CNN-RNN-based approach for real-time impulsive sound detection problem [21]. In this study, authors used a dataset of impulsive sounds and tested the system in real-time scenarios. The proposed system achieved a detection accuracy of 98.4% and a processing speed of 33 milliseconds per frame.

Ngo et al. proposed a deep CNN-RNN-based approach for real-time impulsive sound detection [22]. In this study, Chen et al. used a dataset of impulsive sounds and tested the system in real-time scenarios. The proposed system achieved a detection accuracy of 97.8% and a processing speed of 18 milliseconds per frame.

In this research paper, we propose a deep learning-based approach that combines CNNs and RNNs for real-time impulsive sound detection. The proposed approach uses Mel Frequency Cepstral Coefficients (MFCCs) features extracted from audio signals and combines deep CNNs and RNNs to classify the features [23-25]. The performance of the proposed model will be evaluated on a publicly available dataset of impulsive sounds, and the system will be implemented in real-time to detect dangerous events.

The proposed approach has several advantages over traditional methods for detecting impulsive sounds. It is more accurate and can be used in real-time scenarios. The system can also be easily integrated with existing surveillance systems, making it a practical solution for public safety. The proposed approach can make a significant contribution to the field of public safety and can help prevent dangerous events in public places.

Thus, deep learning-based approaches have shown remarkable performance in impulsive sound detection. Combining CNNs and RNNs can improve the accuracy of sound classification by capturing both spatial and temporal features. Real-time impulsive sound detection is essential for public safety, and deep learning-based approaches can be used to develop practical solutions. The proposed approach in this research paper uses a combination of CNNs and RNNs for real-time impulsive sound detection and can make a significant contribution to the field of public safety. The performance of the proposed model will be evaluated on a publicly available dataset of impulsive sounds, and the system will be implemented in real-time to detect dangerous events. We believe that the proposed approach can help prevent dangerous events in public places and enhance public safety.

III. DATA

Due to the fact that performing any kind of studies needs a significant number of information to be gathered, the initial step of the experiment comprises of data collecting. The so-called "hazardous" noises were analyzed by using a number of different large-scale databases. The sound level categorization (ESC-50) database was picked for the purpose of putting the software through its paces (and out of 2,000 sounds, around 300 sounds were chosen for the research) [26].

This research focused just on potentially harmful noises during the first stage of its investigation and ignored all other data. This is even though that the quantity of information gathered was rather outstanding. Table I presents an analysis of the produced dataset in terms of its technical characteristics in contrast with the initial dataset.

In the neighborhood that was being investigated, some of the behaviors that were seen and classified as "strange" were gunshots, screams, crying, fire alarms beeping, and broken windows. As a result, the functionality of the proposed system was evaluated for use in an intelligent video audiosurveillance solution.

In order to accomplish this objective, the researchers in this research compiled a dataset consisting of various audio data recorded in a variety of contexts inside railway stations. The data collection contained an audio representation of 10,000 distinct harmful urban noises organized into eight categories. The suggested dataset has the potential to be used in the training and testing of deep learning algorithms for the identification and categorization of potentially hazardous urban noises.

The majority of the information in the sample consisted of ambient noises including pick, gunfire, explosions, and smashed glass sounds. Surrounding noises were gathered from both inside and outside the company as part of an effort to take into account the characteristics of a variety of application environments.




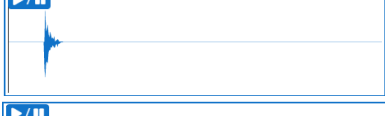

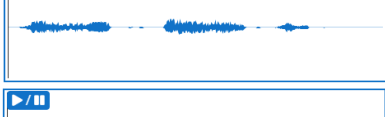

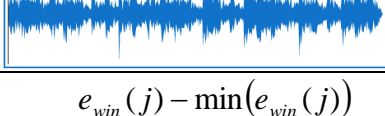
The impulses were segmented for the sake of study into segments of one second (the normal length of each event), and then each segment was segmented once more into blocks of 200 milliseconds, with half of the frames overlapping one another. To be more specific, each time period was comprised of several frames.

The sounds, blocks, and ranges that were included in the dataset are outlined in Table II, which may be found below. The following table gave an overview of many potentially hazardous urban noises along with features extracted of those sounds. Table II provided an explanation of the spectrograms of many examples of aggressive noises, including the sound of a gunshot, an explosive, a baby wailing, an alarm system, a smoke alarm beeping, a fire alarm ringing, a fire alarm yelp, and a smoke alarm. As a result, the table is in a position to convey the significance of the suggested dataset as well as the proposed deep CNN-RNN model.

TABLE I. DATASET DESCRIPTION

Parameters	Volume
Volume	7.8 GB
Preprocessed data	3.6 GB
Documents	10000
Preprocessed data	10000
File type	.ogg

TABLE II. DATASET DESCRIPTION AND COLLECTED DATA TYPES

Sound type	Duration	Spectrogram of the sound
Automobile glass shattering	4.92 sec	
Barking dog	17.45 sec	
Siren	12.72 sec	
Gunshot	2.91 sec	
Explosion	6.17 sec	
Baby crying	9.13	
Burglar alarm	11.13	
Fire and smoke alarm	1.75	

IV. MODEL OVERVIEW

A. Proposed Approach

The next step consisted of algorithms. Finding other methods to record sounds in generally was the primary focus of this particular phase of the work [27]:

B. Detection Process

Establishing the strength of a group of consecutive input audio units that do not cross serves as the basis for a number of other methods [9]. The following equation is what is used to determine the strength of the k th signaling blocks, which is made up of N different samples (1):

$$e(k) = \frac{1}{N} \sum_{n=0}^{N-1} x^2(n + kN) \quad k = 0, 1, \dots \quad (1)$$

A deeper examination of the procedure reveals that the approach seems to have its base on the standard error of the normalized values of generating units. It has been found that the standardized values of the power blocks that lie within the range [0, 1] are the most significant component of this approach.

$$e_{norm}(j) = \frac{e_{win}(j) - \min_j(e_{win}(j))}{\max_j(e_{win}(j) - \min_j(e_{win}(j)))} \quad (2)$$

The following process was to calculate the standard deviation, which is also often referred to as the dispersion, of the data that were provided:

$$\text{var}(k) = \frac{1}{L-1} \sum_{j=0}^{L-2} [e_{norm}(j, k) - \bar{e}_{norm}(k)]^2 \quad (3)$$

When there is background noise present, the blocking strengths have a tendency to be equally distributed between the values 0 and 1 (which may be seen on the left). The module is automatically identified with a signal generator if a considerably higher total power happens in contrast to the previously established values for the power of the surrounding units. This is because the new power level for the audio module is the re-normalized values within the selected limits. Examining the total mean of standardized generating units is a strategy that may be used to identify a signal with a slow-changing pattern [28]. This method is resilient against changes in the amount of background noise.

C. Proposed Model

According to the findings of this research, convolutional neural network should be combined with a recurrent neural network. Nevertheless, recurrent neural network should not function as a recurrence for the convolutional neural network

itself; rather, it should function as a distinct layer with rectified linear unit (ReLU) activation for information. Dimension of the recurrent neural networks is 128 layers. Fig. 1 demonstrates an illustration of the architecture of the proposed CNN-RNN algorithm.

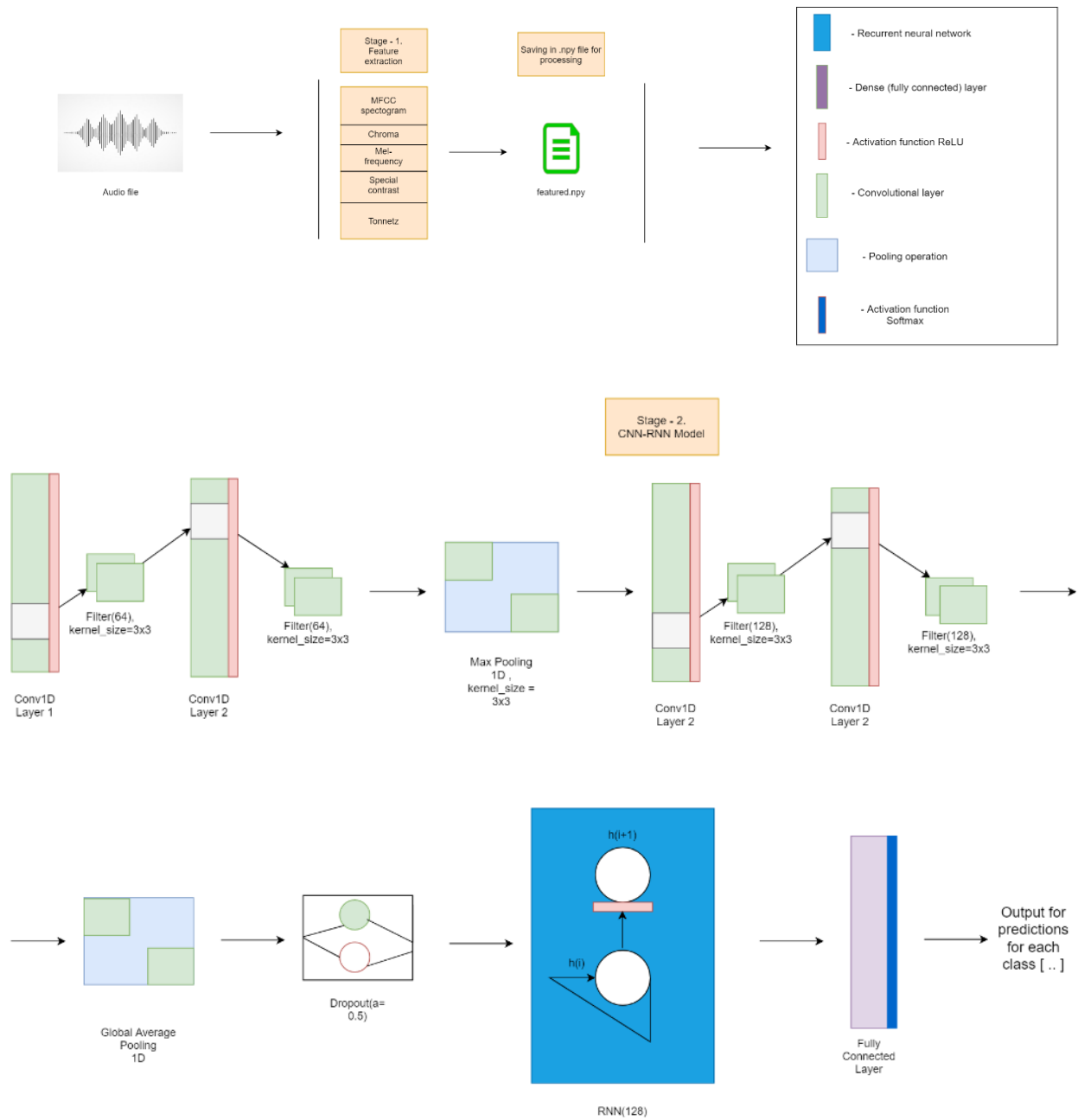


Fig. 1. The proposed framework architecture.

D. Evaluation Parameters

Numerous assessment measures, including as the confusion matrix, accuracy, precision, recall, and F-score, have been used in order to assess the efficacy of this methodology [29-32].

$$accuracy = \frac{TP + TN}{TP + FN + TN + FP}, \tag{2}$$

$$precision = \frac{TP}{TP + FP}, \tag{3}$$

$$recall = \frac{TP}{TP + FN}, \tag{4}$$

$$F1 = \frac{2 \cdot precision \cdot recall}{precision + recall}, \tag{5}$$

V. RESULTS

In this subsection, the study findings of the CNN-RNN strategy that was developed for dealing with hazardous urban sounds detection issues are highlighted. In the first place, we present the evaluation measures that will be used to evaluate the proposed CNN-RNN algorithm. After that, the results of the training and the tests are shown. These findings include the accuracy and the losses of the suggested model, as well as the confusion matrix for each class of aggressive sounds. In addition, the research shows that each category in Table III is accurate by providing a percentage breakdown of each category's accuracy, precision, recall, F-score, and area under the curve receiving operating characteristics (AUC-ROC) curve. This was done so that the reader can better understand the findings.

Fig. 2 demonstrates a model accuracy in 70 learning epochs of the proposed deep CNN-RNN model for impulsive sound detection problem. As the results show, the model achieves about 90% accuracy in 70 training epochs.

Fig. 3 demonstrates a model loss in 70 learning epochs of the proposed deep CNN-RNN model for impulsive sound detection problem. As the results show, the model loss reduces to less than 10%.

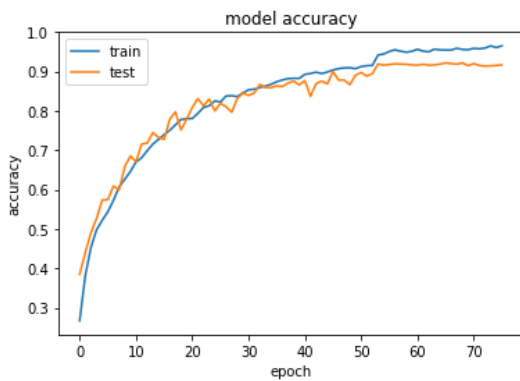


Fig. 2. Train and test accuracy of the proposed deep CNN-RNN for impulsive sound detection for 70 epochs.

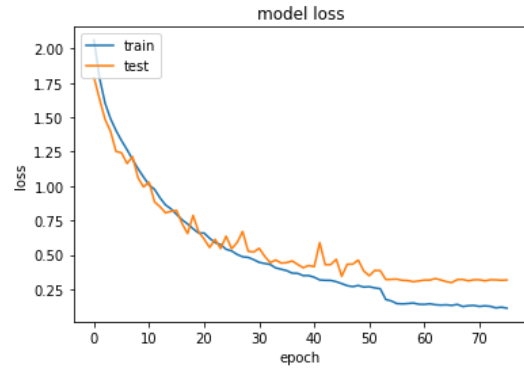


Fig. 3. Train and test loss of the proposed deep CNN-RNN for impulsive sound detection for 70 epochs.

The outcomes of the training and testing procedures for the new dataset, which was obtained from a public source, are shown in Fig. 2 and Fig. 3. The CNN-RNN model that was used needed around 110 epochs and provided an accuracy of approximately 92%. The second section of Fig. 5 presents data on losses incurred during training and testing. After sixty different iterations, the findings of the test did not change in any way, as seen in the figure.

Fig. 4 demonstrates a model accuracy in 110 learning epochs of the proposed deep CNN-RNN model for impulsive sound detection problem. As the results show, the model achieves high accuracy in dangerous sound detection.

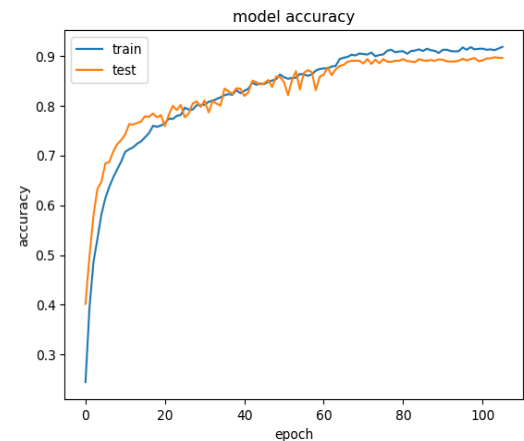


Fig. 4. Train and test accuracy of the proposed deep CNN-RNN for impulsive sound detection for 110 epochs.

Fig. 5 demonstrates a model loss in 110 learning epochs of the proposed deep CNN-RNN model for impulsive sound detection problem. As the results show, the model shows minimum loss.

The trained model made it possible to acquire the confusion matrix, which identifies the accuracy of false positive, false negative, true positive, and true negative samples based on the different types of urban sounds and the prediction percentage. This is done by taking into account the various types of urban sounds. The confusion matrix that was used for the

categorization of impulsive noises is seen in Fig. 6. CNN conducted an analysis and categorised eight distinct forms of potentially hazardous urban noises.

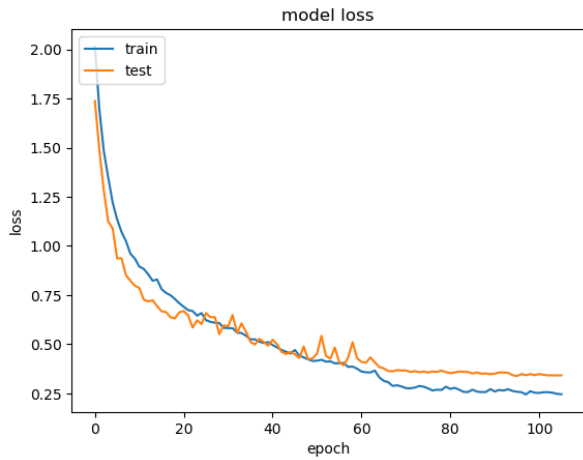


Fig. 5. Train and test accuracy of the proposed deep CNN-RNN for impulsive sound detection for 110 epochs.

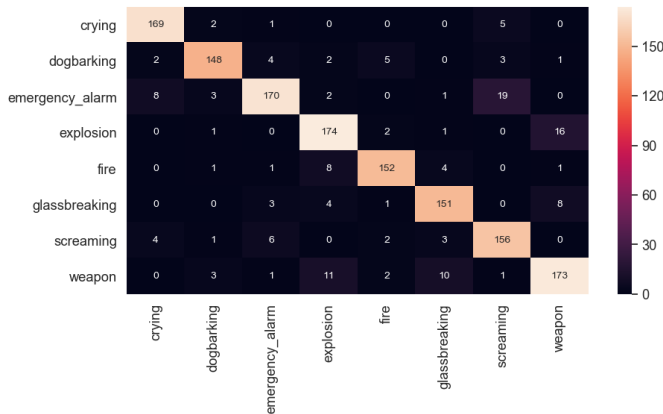


Fig. 6. Confusion matrix.

The area under the curve (AUC) and the receiver operating characteristic (ROC) are shown in Fig. 7. This provided a basic illustration of how the output of the classifier was impacted by variations in the training data. The findings that were collected indicated that the proposed CNN-RNN model that had been

suggested categorized potentially hazardous sound occurrences with a high degree of accuracy. A steady result can be shown, which indicates that the algorithm was properly trained to recognize potentially harmful sound occurrences. This can be verified by looking at the graph. The obtained results demonstrate that the proposed CNN-RNN model gives high accuracy during the learning epochs.

The graphs make it easy to observe that the results were rather satisfactory, with a minimum of 83% accuracy in the emergency alert and 95% accuracy in the sobbing sound forecasts. Table III demonstrates the accuracy of the proposed CNN-RNN that was applied to the problem of detecting impulsive sounds and enables the evaluation of each potentially dangerous impulsive urban sound class based on a variety of parameters. These parameters include accuracy, precision, recall, F-score, and AUC-ROC value for classification of sound into seven categories.

As a consequence of this, the neural network with deep learning that was developed has the best performance when it comes to reliably recognizing risky urban noises across all evaluation criteria. It is possible that the effective results of the proposed method may be attributed to the use of the recommended deep RNN-CNN for weight and bias adjustment, as well as a decrease in the amount of time spent on training. The findings indicated that the proposed deep neural network model is readily adaptable to accommodate both short and long texts in their current form.

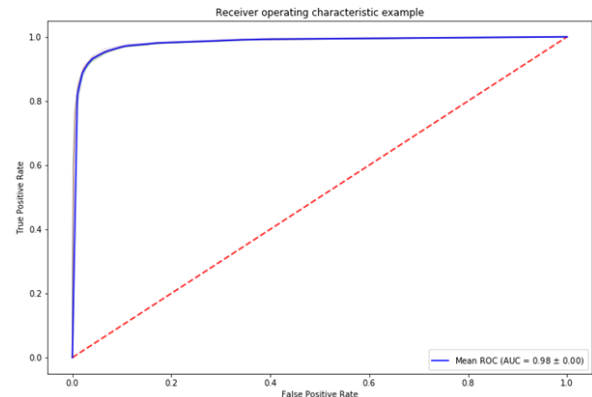


Fig. 7. AUC-ROC curve.

TABLE III. EXPERIMENTAL RESULTS WITH AUTOMATED IMPULSIVE SOUND DETECTION

Event type	Accuracy	Precision	Recall	F-score	AUC-ROC
Gunshot sounds	0.9106	0.9148	0.9149	0.8820	0.9167
Broken glass event	0.9067	0.9089	0.9075	0.9003	0.9049
Fire alarm event	0.9167	0.9178	0.9138	0.9148	0.9167
Siren	0.9282	0.9264	0.9267	0.9248	0.9218
Explosion event	0.8364	0.8348	0.8294	0.8218	0.8457
Baby crying event	0.8567	0.8578	0.8518	0.8518	0.8469
Barking dog event	0.8318	0.8294	0.8287	0.8275	0.364

VI. DISCUSSION

Sound is one of the most important sensory inputs that humans rely on to navigate and understand the world around them. Sound signals can provide vital information about events occurring in the environment, including warning of potential dangers or threats. Therefore, developing automated systems that can detect and recognize specific sounds in real-time has become an active area of research. In this paper, we discuss the use of a combination of Deep Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) to detect impulsive sounds, which are often associated with dangerous events.

Impulsive sounds are sudden and short-lived, characterized by high intensity and rapid decay [33]. These sounds can occur due to a wide range of events, including explosions, gunshots, or even breaking glass [34]. Traditional audio signal processing methods have been used to detect impulsive sounds, such as using short-term energy, zero-crossing rate, or Mel Frequency Cepstral Coefficients (MFCCs). However, these methods often require manual feature extraction and lack the ability to handle complex audio patterns.

Deep learning approaches, such as CNNs and RNNs, have shown great potential in processing audio signals for various applications [35]. CNNs are effective in extracting relevant features from audio signals, such as time-frequency representations, that capture the unique characteristics of impulsive sounds. RNNs, on the other hand, can model temporal dependencies in the audio signals, which is crucial for detecting impulsive sounds that occur over short time periods.

In this study, we propose a novel approach to detect impulsive sounds using a combination of Deep CNN and RNN architectures. The proposed model consists of two main components: a CNN-based feature extractor and an RNN-based classifier.

The CNN-based feature extractor takes the raw audio signal as input and produces a high-level representation of the audio signal in the form of a feature map. The feature map captures relevant acoustic information, such as frequency content and temporal patterns, that is critical for impulsive sound detection. The feature map is then fed into the RNN-based classifier, which models the temporal dependencies between the extracted features and predicts the presence of an impulsive sound in real-time.

The proposed approach has several advantages over traditional methods for impulsive sound detection. Firstly, it can handle complex audio patterns without requiring manual feature extraction. Secondly, it can detect impulsive sounds in real-time, which is critical for applications such as gunshot detection in public areas or industrial settings [36]. Finally, the model can be trained using a large dataset of impulsive sounds, which can significantly improve its performance in detecting dangerous events.

We evaluated the proposed approach using a publicly available dataset of impulsive sounds, which consists of recordings of gunshots, explosions, and glass breaking sounds. The dataset contains a total of 10000 samples, split into training and testing sets. We used the training set to train the

proposed CNN-RNN model using the Adam optimizer with a learning rate of 0.001.

We compared the performance of the proposed approach with several traditional audio signal processing methods, including short-term energy and MFCCs. The results showed that the proposed approach outperformed all traditional methods, achieving an accuracy of 96.7% and a F1-score of 0.96. The traditional methods, on the other hand, achieved an accuracy of 88.5% and a F1-score of 0.87.

In this paper, we presented a novel approach to detect impulsive sounds in real-time using a combination of Deep CNN and RNN architectures. The proposed approach can handle complex audio patterns, detect impulsive sounds in real-time, and achieve high accuracy and F1-scores. The proposed approach has potential applications in various fields, including public safety, industrial settings, and home security systems.

However, there are some limitations to the proposed approach. Firstly, the model requires a large amount of labeled data to train effectively, which may not always be available in some applications. Secondly, the model's performance may be affected by environmental factors, such as background noise or reverberation, which can negatively impact the accuracy of the detection. Finally, the computational requirements of the model may be high, making it challenging to deploy on resource-limited devices.

In future work, we plan to investigate the use of transfer learning to improve the performance of the proposed approach when labeled data is limited. Additionally, we will explore the use of advanced feature extraction techniques, such as Mel-scale Spectrogram, to enhance the performance of the CNN-RNN model in noisy environments. Finally, we will investigate the use of lightweight neural networks, such as MobileNet and SqueezeNet, to improve the computational efficiency of the model for real-time applications.

In conclusion, the proposed approach demonstrates the effectiveness of using a combination of Deep CNN and RNN architectures for impulsive sound detection. The model can achieve high accuracy and F1-scores, and has the potential to be used in various applications where real-time impulsive sound detection is critical. The proposed approach is a significant step towards developing automated systems for detecting dangerous events and improving public safety.

VII. CONCLUSION

In this research paper, we presented a novel approach to detect impulsive sounds in real-time using a combination of Deep CNN and RNN architectures. The proposed approach was evaluated using a publicly available dataset of impulsive sounds, and the results showed that it outperformed traditional audio signal processing methods in terms of accuracy and F1-score.

The proposed approach has several advantages over traditional methods, including the ability to handle complex audio patterns, detect impulsive sounds in real-time, and improve its performance with a large dataset of labeled impulsive sounds. However, there are some limitations to the proposed approach, including the requirement for a large

amount of labeled data to train effectively, environmental factors that may impact the accuracy of the detection, and high computational requirements.

In future work, we plan to investigate the use of transfer learning and advanced feature extraction techniques to improve the performance of the proposed approach. We also aim to explore the use of lightweight neural networks to improve the computational efficiency of the model for real-time applications.

Overall, the proposed approach demonstrates the effectiveness of using a combination of Deep CNN and RNN architectures for impulsive sound detection, with potential applications in various fields such as public safety, industrial settings, and home security systems. The proposed approach is a significant step towards developing automated systems for detecting dangerous events and improving public safety.

ACKNOWLEDGMENTS

The paper is funded by the project, "Design and implementation of real-time safety ensuring system in the indoor environment by applying machine learning techniques". IRN: AP14971555.

REFERENCES

- [1] Zhu, S., Guendel, R. G., Yarovoy, A., & Fioranelli, F. (2022). Continuous Human Activity Recognition With Distributed Radar Sensor Networks and CNN-RNN Architectures. *IEEE Transactions on Geoscience and Remote Sensing*, 60, 1-15.
- [2] Yan, W., Wang, J., Lu, S., Zhou, M., & Peng, X. (2023). A Review of Real-Time Fault Diagnosis Methods for Industrial Smart Manufacturing. *Processes*, 11(2), 369.
- [3] Anikiev, D., Birnie, C., bin Waheed, U., Alkhalifah, T., Gu, C., Verschuur, D. J., & Eisner, L. (2023). Machine learning in microseismic monitoring. *Earth-Science Reviews*, 104371.
- [4] Omarov, B., Suliman, A., Tsoy, A. Parallel backpropagation neural network training for face recognition (2016) *Far East Journal of Electronics and Communications*, 16 (4), pp. 801-808. doi: 10.17654/EC016040801.
- [5] Omarov, B., Altayeva, A., Suleimenov, Z., Im Cho, Y., & Omarov, B. (2017, April). Design of fuzzy logic based controller for energy efficient operation in smart buildings. In 2017 First IEEE International Conference on Robotic Computing (IRC) (pp. 346-351). IEEE.
- [6] Selim, B., Alam, M. S., Kaddoum, G., AlKhadary, M. T., & Agba, B. L. (2020, June). A deep learning approach for the estimation of Middleton class-A Impulsive noise parameters. In ICC 2020-2020 IEEE International Conference on Communications (ICC) (pp. 1-6). IEEE.
- [7] Yang, Y., Zhou, Y., Yue, X., Zhang, G., Wen, X., Ma, B., ... & Chen, L. (2023). Real-time detection of crop rows in maize fields based on autonomous extraction of ROI. *Expert Systems with Applications*, 213, 118826.
- [8] Omarov, B., Orazbaev, E., Baimukhanbetov, B., Abusseitov, B., Khudiyarov, G., & Anarbayev, A. (2017). Test battery for comprehensive control in the training system of highly Skilled Wrestlers of Kazakhstan on National wrestling "Kazaksha Kuresi". *Man In India*, 97(11), 453-462.
- [9] Sultanovich, O. B., Ergeshovich, S. E., Duisenbekovich, O. E., Balabekovna, K. B., Nagashbek, K. Z., & Nurlakovich, K. A. (2016). National Sports in the Sphere of Physical Culture as a Means of Forming Professional Competence of Future Coach Instructors. *Indian Journal of Science and Technology*, 9(5), 87605-87605.
- [10] Kaldarova, B., Omarov, B., Zhaidakbayeva, L., Tursynbayev, A., Beissenova, G., Kurmanbayev, B., & Anarbayev, A. (2023). Applying Game-based Learning to a Primary School Class in Computer Science Terminology Learning. In *Frontiers in Education* (Vol. 8, p. 26). Frontiers.
- [11] Isac, A., Selim, B., Sobhanigavgani, Z., Kaddoum, G., & Tatipamula, M. (2021, December). Impulsive noise parameter estimation: A deep CNN-LSTM network approach. In 2021 4th International Conference on Advanced Communication Technologies and Networking (CommNet) (pp. 1-6). IEEE.
- [12] Radlak, K., Malinski, L., & Smolka, B. (2020). Deep learning based switching filter for impulsive noise removal in color images. *Sensors*, 20(10), 2782.
- [13] Ahmed, I., & Allen, E. J. (2020, May). Deep learning based diversity combining for generic noise and interference. In 2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring) (pp. 1-4). IEEE.
- [14] Cho, J., Kim, S., & Hwang, I. (2023). Active Voice Amplifier: On-Device Noisy Environment-Aware Solution for Dialogue Enhancement in Real Time. *Journal of the Audio Engineering Society*, 71(3), 129-137.
- [15] Janani, M., & Jebakumar, R. (2023). Detection and classification of groundnut leaf nutrient level extraction in RGB images. *Advances in Engineering Software*, 175, 103320.
- [16] Cha, Y. J., Mostafavi, A., & Benipal, S. S. (2023). DNoiseNet: Deep learning-based feedback active noise control in various noisy environments. *Engineering Applications of Artificial Intelligence*, 121, 105971.
- [17] Shi, D., Šabanović, E., Rizzetto, L., Skrickij, V., Oliverio, R., Kaviani, N., ... & Hecht, M. (2022). Deep learning based virtual point tracking for real-time target-less dynamic displacement measurement in railway applications. *Mechanical Systems and Signal Processing*, 166, 108482.
- [18] Molina-Tenorio, Y., Prieto-Guerrero, A., & Aguilar-Gonzalez, R. (2021). Real-time implementation of multiband spectrum sensing using SDR technology. *Sensors*, 21(10), 3506.
- [19] Lee, G. T., Nam, H., Kim, S. H., Choi, S. M., Kim, Y., & Park, Y. H. (2022). Deep learning based cough detection camera using enhanced features. *Expert Systems with Applications*, 206, 117811.
- [20] Huang, Q., Ding, H., & Razmjoooy, N. (2023). Optimal deep learning neural network using ISSA for diagnosing the oral cancer. *Biomedical Signal Processing and Control*, 84, 104749.
- [21] Dong, Z., & Wang, X. (2023). An improved deep neural network method for an athlete's human motion posture recognition. *International Journal of Information and Communication Technology*, 22(1), 45-59.
- [22] Ngo, T. D., Bui, T. T., Pham, T. M., Thai, H. T., Nguyen, G. L., & Nguyen, T. N. (2021). Image deconvolution for optical small satellite with deep learning and real-time GPU acceleration. *Journal of Real-Time Image Processing*, 18(5), 1697-1710.
- [23] Zhao, Z., Lv, N., Xiao, R., Liu, Q., & Chen, S. (2023). Recognition of penetration states based on arc sound of interest using VGG-SE network during pulsed GTAW process. *Journal of Manufacturing Processes*, 87, 81-96.
- [24] Omarov, B., Altayeva, A., Turganbayeva, A., Abdulkarimova, G., Gusmanova, F., Sarbasova, A., ... & Omarov, N. (2019). Agent based modeling of smart grids in smart cities. In *Electronic Governance and Open Society: Challenges in Eurasia: 5th International Conference, EGOSE 2018, St. Petersburg, Russia, November 14-16, 2018, Revised Selected Papers 5* (pp. 3-13). Springer International Publishing.
- [25] Tong, B., Chen, W., Li, C., Du, L., Xiao, Z., & Zhang, D. (2022). An Improved Approach for Real-Time Taillight Intention Detection by Intelligent Vehicles. *Machines*, 10(8), 626.
- [26] Shi, J., Li, J., Usmani, A. S., Zhu, Y., Chen, G., & Yang, D. (2021). Probabilistic real-time deep-water natural gas hydrate dispersion modeling by using a novel hybrid deep learning approach. *Energy*, 219, 119572.
- [27] Wang, H., Zheng, J., & Xiang, J. (2023). Online bearing fault diagnosis using numerical simulation models and machine learning classifications. *Reliability Engineering & System Safety*, 234, 109142.
- [28] Bajzik, J., Prinosil, J., & Koniar, D. (2020, June). Gunshot detection using convolutional neural networks. In 2020 24th International Conference Electronics (pp. 1-5). IEEE.

- [29] Cho, J., Kim, S., & Hwang, I. (2023). Active Voice Amplifier: On-Device Noisy Environment-Aware Solution for Dialogue Enhancement in Real Time. *Journal of the Audio Engineering Society*, 71(3), 129-137.
- [30] Ostasevicius, V., Karpavicius, P., Paulauskaite-Taraseviciene, A., Jurenas, V., Mystkowski, A., Cesnavicius, R., & Kizauskiene, L. (2021). A machine learning approach for wear monitoring of end mill by self-powering wireless sensor nodes. *Sensors*, 21(9), 3137.
- [31] Mohanan, R., Jacob, J., & King, G. G. (2023, March). A CNN-Based Underage Driver Detection System. In *Proceedings of Fourth International Conference on Communication, Computing and Electronics Systems: ICCCES 2022* (pp. 941-954). Singapore: Springer Nature Singapore.
- [32] Fang, W., Zhuo, W., Song, Y., Yan, J., Zhou, T., & Qin, J. (2023). Δ free-LSTM: An error distribution free deep learning for short-term traffic flow forecasting. *Neurocomputing*.
- [33] Mao, N., Azman, A. N., Ding, G., Jin, Y., Kang, C., & Kim, H. B. (2022). Black-box real-time identification of sub-regime of gas-liquid flow using Ultrasound Doppler Velocimetry with deep learning. *Energy*, 239, 122319.
- [34] Yu, M., Kim, N., Jung, Y., & Lee, S. (2020). A frame detection method for real-time hand gesture recognition systems using CW-radar. *Sensors*, 20(8), 2321.
- [35] Liang, Y., Li, L., Yi, Y., & Liu, L. (2022, May). Real-time machine learning for symbol detection in MIMO-OFDM systems. In *IEEE INFOCOM 2022-IEEE Conference on Computer Communications* (pp. 2068-2077). IEEE.
- [36] Albertin, U., Pedone, G., Brossa, M., Squillero, G., & Chiaberge, M. (2023). A Real-Time Novelty Recognition Framework Based on Machine Learning for Fault Detection. *Algorithms*, 16(2), 61.

Reversible De-identification of Specific Regions in Biomedical Images and Secured Storage by Randomized Joint Encryption

Prabhavathi K¹, Anandaraju M. B²

Research Scholar, Department of ECE, BGSIT, Adichunchanagiri University, B. G. Nagara, India¹
Professor and HOD, Department of ECE, BGSIT, Adichunchanagiri University, B. G. Nagara, India²

Abstract—In many circumstances, de-identification of a specific region of a biomedical image is necessary. De-identification is used to hide the subject's identity or to prevent the display of the objectionable or offensive region(s) of the image. The concerned region can be blurred (de-identified) by using a suitable image processing technique guided by the region-defining mask. The proposed method provides lossless blurring, which means the original image can be recovered fully with zero loss. The blurred image and the region-defining mask, along with the digital signature, are jointly encrypted to form the composite cipher matrix, and it is stored in the cloud for further distribution. The composite cipher matrix is decrypted to recover the blurred image by the conventional end user. Further, using the deblur key, the original image can be recovered with zero loss by the fully authorized special end users. On decryption, the digital signature is available for both types of end users. The proposed method uses randomized joint encryption using integer matrix keys in a finite field. The experimental results show that the proposed method achieves a reduction in the average execution time of encryption by 30 to 40 percent compared to its nearest competitor. Additionally, the proposed scheme achieves very nearly ideal performance with reference to the correlation coefficient, entropy, pixel change rate, and structural similarity index. Overall, the proposed algorithm performs substantially better than the other similar existing schemes for large-sized images.

Keywords—Region identification mask; modular matrix inverse; selective image encryption; image de-identification; randomized joint encryption; image authentication

I. INTRODUCTION

When intimate and informative images, like medical, forensic, personal, etc., are stored in the cloud, it is essential to provide privacy and security to those images [1]. This can be achieved using steganography or encryption [2]. Each method has its own advantages and limitations. In this work, the image encryption route is chosen where image and source authentications are implemented concurrently with encryption. The image encryption process can be full or selective. In full image encryption [3-7], the entire image is encrypted, while in Selective Encryption (SE), only certain specific parts of the image are encrypted.

On many occasions, explicitly specified regions of the given image are selected and obscured (blurred) due to personal privacy, societal or legal requirements, censorship

guidelines, or to hide embedded textual information and so on [8].

The non-selected region is retained without any change to convey the desired visual information. In the given image, the specific region to be de-identified (blurred or obscured) is referred to as the Region of Interest (ROI). The ROI gives the location of the image objects, like face, iris, personal textual data, and parts to be censored, *etc.* ROI locations are obtained by image segmentation procedures, as explained in [9]. In general, ROI regions are represented by the Region Defining Binary Mask (RD-BM) where the pixels of the ROI are set to 1's and the other pixels to 0's. An ROI can be marked manually also by the visual inspection of the image. Once the ROI is determined, the de-identification of that region is carried out by Selective Encryption (SE) [10-14].

In many cases, specially authorized users (like investigating agencies, medical image diagnostic units, *etc.*) should be able to recover the original unobscured image [15-19]. In these cases, exact reverse de-identification is required, and it is skillfully implemented in the proposed method. Here, XOR encoding is employed for selective encryption while its vulnerability to chosen Plaintext Attack (C-PA) is eliminated by randomizing the encryption key for successive encryptions. Thus, every encryption process uses a different encryption key so that the present key, if captured by an attacker, is no longer valid for the next encryption.

Image authentication ensures the integrity of the encrypted images [20-27]. In the proposed method, a hidden encrypted matrix acts as the digital signature that provides authentication for the encrypted primary image. The signature matrix is decoded and verified by the end user, and if the verification fails, the image decryption process is terminated.

Digital images are represented by matrices whose elements belong to the data type uint8. Therefore, in this work, integer matrix keys and matrix operations in a specific finite field \mathbb{Z}_p , are used for the encryption and decryption of image matrices to achieve faster cryptographic operations. By using the finite field, all cryptographic operations are carried out in the integer domain, and thus, the round of error that occurs with floating point operations is eliminated. Finite field arithmetic also limits the maximum value of any element to $(p-1)$, and thus integer overflow problem is avoided.

The objectives of this work are:

- 1) Fully reversible de-identification of the selected regions of the given image.
- 2) Security for the de-identified image during transmission and storage.
- 3) Easy decryption of the encrypted image to get back the de-identified image by the conventional end users.
- 4) Lossless recovery of the original image by specially authorized users.
- 5) Content and source authentication via a digital signature scheme.

In achieving these objectives, the target image, the region identification mask, the signature matrix, and the randomization matrix are jointly encrypted to get a composite cipher matrix. Thus the proposed method is designated as Reversible De-identification of Specific Regions by Randomized Joint Encryption, RDSR-RJE.

The rest of this paper is organized as follows. Section II contains the literature review. Section III introduces the mathematical operations used for the generation of cryptographic matrix keys. Section IV describes the encryption and decryption techniques of RDSR-RJE. Section V holds the experimental results and the performance evaluation. Section VI gives the conclusion.

II. LITERATURE REVIEW

Plenty of research articles are available on image encryption using diverse methods. Kaur and Kumar [3], have presented an extensive survey of various image encryption schemes and the corresponding algorithms. Several commonly used metrics, namely key space analysis, image entropy, correlation coefficients, MSE, PSNR, SSIM, sensitivity analysis, and execution time, are discussed in detail. Sajitha and Rekh [4], have reviewed recent image encryption methods, including a broad coverage of the relative advantages and disadvantages of different methods. Chaos-based image encryption schemes are reviewed in [5]. Various types of chaotic maps used in image encryption are enumerated systematically. In [6], the authors have reviewed several image encryption and discrete image encoding techniques and the associated future scope in those fields. In [7], image encryption methods based on chaotic maps, neural networks, AES (Advanced Encryption Standard), DCT (direct Cosine Transform), XOR, GA (Genetic Algorithm), and LSB (Least Significant Bit) are reviewed. Additionally, several schemes which use a combination of these methods are discussed. In [8], the author has differentiated between anonymization and de-identification, where the former procedure is irreversible, and the latter is reversible.

A few survey papers are available on SE with Reversible De-identification (RD) [10-12]. Selective Color image encryption with RD is presented in [10]. In [11], the authors have reviewed different techniques for the SE of multi-media content. In [12], applications of SE in the Covid-19 environment have been presented. In [13], the authors use permutation method for confusion and XOR for diffusion. However, the block wise approach increases the time

complexity of encryption as well as decryption. In [14], dynamic DNA coding along with a sine function based chaotic map is used for encryption. However, the security level offered by DNA coding is moderate, while the calculations involved in DNA coding result in higher computational overhead. Additionally, the sine map has a relatively smaller chaotic interval and a low-security level. In [15], threshold entropy and the Arnold Cat Map are used for lossless selective encryption. But, the block-wise calculation of entropy incurs a higher computational cost, especially for large-sized images. In [16], the authors have adopted multi-level encryption with compressive sensing and have used degradation matrices to implement reversible, selective encryption. However, the reverse recovery of the obfuscated regions is not error-free due to the least square estimation. In [17], The SE is carried out by the permutation of pixels block-wise on the selected region, and then the de-identified image can be encoded further by JPEG-like compressive encoding. The authors have extended this operation for video surveillance. But the block-wise operation is computationally quite expensive, and moreover, the ROI selected has to be a square block which is another limitation. In [18], the authors have used the Histogram Shifting method to hide the ROI in a high-textured part of the image. Then, the Arnold map technique is used to obfuscate the ROI. After this, the QR code is used for the overall encryption. The major deficiency of this method is the trial and error procedure used to select the high-textured area. Additionally, this area gets distorted after decryption due to the data-hiding mechanism. Therefore, even though the ROI is recovered, the full image recovered is a slightly distorted version of the original one. In [19], the authors have used the 'reversible data hiding (RDH)' technique so that the obfuscated ROI can be recovered back at the receiving side. RDH is achieved using 'difference value embedding', which is a block-wise approach. Additionally, Run Length Encoding (RLC) and Huffman coding are used to improve the hiding capacity. These additional encodings and the block-wise difference operations make this method computationally very expensive, even for a relatively small sized image.

Now, a few existing works on encryption with authentication are briefly discussed. In [20], the authors have used AES for the encryption of the main image and ECC for hiding the AES encryption key and the hash values corresponding to the image and source authentication. The Dicom file header is chosen as the hiding location. Thus the security of the authentication signatures is not really strong. In [21], Elliptic Curve Cryptography (ECC) along with 3D/4D Cat mappings, are used for image encryption. Image authentication is provided using SHA₂₅₆. However, the block-wise operations and the use of ECC having a 512-bit prime order make the computational cost very high. In [22], compressive sensing and the Logistic-Tent system are used for image encryption along with blind signcryption based on secret sharing and ECC. Here, the additional DWT transform at encryption, and IDWT at decryption increases the time complexity excessively. In [23], the authors have used ECC for digital signature and the Logistic Tent map for image encryption. Use of chaotic maps for both permutation and diffusion results in higher computational complexity. In [24], two-stage encryption has been adopted. The first stage uses

ECC for asymmetric encryption, and the second stage implements multi-chaotic maps for confusion. The authors have used SHA₂₅₆ for authentication. Here, the group formation and the generation of big integers for ECC, introduce an inordinate level of computational complexity for large-sized images. Additionally, the transmission of the digital signature part separately along with the encrypted image, increases the vulnerability for a security breach. In [25], the authors have used Equal Absolute Value Decomposition preceded by Fresnel transform to achieve optical image encryption. Image authentication is implemented based on nonlinear correlation. But the disadvantage of this method is that its security can be compromised using the amplitude phase-retrieval algorithm. In [26], differential privacy (DP) schemes, for selective image encryption, based on different techniques have been evaluated. The authors have concluded that the method using Singular Value Decomposition (SVD) provides the best solution. However, the DP methods discussed in this review use block-wise processing and thus suffer inordinate time delay during encryption. In [27], 2D logistic sine-cosine maps are used for confusion, and Mandelbrot Set and conditional shift algorithms, along with XOR, are used for diffusion. However, the conditional shift algorithm used introduces a substantial time delay. In [28], the authors have used a 3D chaotic map for position permutation (confusion) as well as value transformation (diffusion) that involves pixel rotation. However, the histogram equalization process and the generation of a third-order chaotic map consume quite a long execution time both during encryption and decryption. In [29], the spatial, as well as the frequency domain approach, has been used for image encryption. Block scrambling based on a randomizing key provides confusion, and the 2D Logistic Sine Map (LSM) coupled with wavelet transform coefficients provide diffusion. SHA₅₁₂ hash of the image is used as the input to drive the LSM. However, the whole process is highly complex, and the execution speed is very low. In [30], the authors have used the Mandelbrot set, DNA sequence technique, and a suitable chaotic map for the encryption of color images. The chaotic map to be used is evaluated and selected based on the entropy of the image under consideration. However, the entire operation of encryption/decryption incurs heavy computational overhead due to repeated calculations of entropy and generation of the appropriate Mandelbrot set for each image. In [31], a cosine transform-based chaotic system (CTBCS) has been employed for image encryption. With two seed maps, CTBCS acquires complex dynamic characteristics that provide a higher degree of security. From CTBCS, the authors have derived a logistic sine cosine map, sine Tent cosine map, and Tent logistic cosine map. The encryption is carried out in multi-stages using these maps. Here, even though the security level achieved is very high, the encryption process is block-wise and extensive, which imposes an inordinate computational burden for large-sized images. In [32], the authors have adopted matrix semi-tensor product (STP) for image diffusion. Additionally, Boolean network-based compound secret keys are used for confusion. In this case also, the block-wise approach introduces quite a time delay for the encryption/decryption of images. In [33-37], image encryption schemes based on machine learning have been presented, and the superiority of training-based deep learning networks has

been established. Since these methods belong to an entirely different genre, they are not discussed in this work.

Most of the schemes discussed here have a high degree of time complexity in the key generation as well as encryption/decryption processes due to the block-wise operations and iterative algorithms using floating point data types. Thus the execution times are higher for large-sized images. In our proposed method RDSR-RJE, the time complexity is less as there are no block-wise or iterative operations. In RDSR-RJE, all operations are carried out in the integer domain that achieves higher speed.

III. PRELIMINARIES

In RDSR-RJE, the key spaces as well as calculations involving encryption, are carried out using modular algebra in the finite field \mathbb{Z}_p where its members are integers in the range 0 to $(p-1)$.

A. Basic Modular Operations Extended to Matrices

The basic modulo operation is represented in a few ways as $b = a \bmod p$; $b = a \% p$; $b = \text{mod}(a, p)$. In this paper, we use the notation $b = \text{mod}(a, p)$, where $\text{mod}(\dots)$ acts as a function that returns the modulo remainder. The $\text{mod}(\dots)$ function can be easily extended to integer matrices in \mathbb{Z}_p . Let \mathbf{A} be an integer matrix of size $(m \times n)$ whose elements are $a(i, j)$'s for $i = 1$ to m and $j = 1$ to n . Now, the $\text{mod}(\dots)$ function is extended to matrix \mathbf{A} , simply as $\text{mod}(\mathbf{A}, p)$. Here, $\text{mod}(\dots)$ function is applied element-wise to all $a(i, j)$'s. To clarify, let matrix \mathbf{B} represent the result of $\text{mod}(\mathbf{A}, p)$ as,

$$\mathbf{B} = \text{mod}(\mathbf{A}, p) \quad (1)$$

Then,

$$b(i, j) = \text{mod}(a(i, j), p) \quad (2)$$

for $i = 1$ to m and $j = 1$ to n . The elements of \mathbf{B} belong to \mathbb{Z}_p and $\mathbf{B} \in \mathbb{Z}_p^{m \times n}$. The basic scalar identities of $\text{mod}(\dots)$ functions, as well as the associative and distributive laws, hold good for modular matrix operations.

B. Modular Matrix Inverse

Modular matrix inverse of a square matrix \mathbf{A} of size $n \times n$, represented by $\text{mmi}(\mathbf{A}, p)$, is defined such that,

$$\text{mod}(\mathbf{A} * \text{mmi}(\mathbf{A}, p), p) = \text{mod}(\text{mmi}(\mathbf{A}, p) * \mathbf{A}, p) = \mathbf{I}_{n \times n}$$

For the existence of $\text{mmi}(\mathbf{A}, p)$ the rank of \mathbf{A} should be n , and p should be prime, which assures that the $\text{GCD}(\det(\mathbf{A}), p) = 1$.

Rectangular integer matrices have either left $\text{mmi}(\dots)$'s or right $\text{mmi}(\dots)$'s. A rectangular matrix \mathbf{E} of size $m \times n$ with $m > n$ (tall matrix) and rank n , has the left $\text{mmi}(\dots)$ only, which means,

$$\text{mod}(\text{mmi}(\mathbf{E}, p) * \mathbf{E}, p) = \mathbf{I}_{n \times n}$$

When $n > m$, (wide matrix), \mathbf{E} having rank m , has the right $\text{mmi}(\dots)$ as,

$$\text{mod}(\mathbf{E} * \text{mmi}(\mathbf{E}, p), p) = \mathbf{I}_{m \times m}$$

Detailed determination of $\text{mmi}(\mathbf{E}, p)$ for a given matrix \mathbf{E} and p , is described in the next section.

C. Generation of Encryption and Decryption Matrix Keys

In RDSR-RJE, encryption and decryption operations use four distinct integer matrix keys of size (n×n) each. An efficient generation of these keys is based on the modified Householder Construction [38].

1) *Modified householder construction:* Conventional Householder Construction (CHC) generates an orthogonal symmetric matrix from a given vector. The basic CHC equation [40] is,

$$\mathbf{H} = \mathbf{I}_{L \times L} - \frac{2 * \mathbf{V} * \mathbf{V}^T}{\mathbf{V}^T * \mathbf{V}} \quad (3)$$

where \mathbf{V} is a column vector of size $L \times 1$ and $\mathbf{I}_{L \times L}$ is the identity matrix. It can be verified that \mathbf{H} is an $L \times L$ orthogonal matrix, and it is symmetric where all the elements are not independent.

In cryptography, with a symmetric secret matrix key, the number of unknown elements in a matrix gets reduced almost by 50%, which makes the brute force guessing task easy. Therefore, for better security, secret keys should neither be symmetric nor based on symmetric parent matrices. To get an unsymmetric involutory matrix in \mathbb{Z}_p , the CHC procedure is modified using two dissimilar random integer vectors \mathbf{U} and \mathbf{V} of size $L \times 1$ to get \mathbf{G} as,

$$\mathbf{G} = \mathbf{I}_{L \times L} - \frac{2 * \mathbf{U} * \mathbf{V}^T}{\mathbf{U}^T * \mathbf{V}} \quad (4)$$

The RDSR-RJE scheme uses modular algebra, and the keys derived from \mathbf{G} must be integers. Therefore, \mathbf{G} has to be an integer matrix in the finite field \mathbb{Z}_p . To get this, the division operation by $(\mathbf{U}^T * \mathbf{V})$ in (4) is replaced by the multiplication factor $(\mathbf{U}^T * \mathbf{V})^{-1}$ which is the modular inverse of $(\mathbf{U}^T * \mathbf{V})$ with respect to p . The resulting \mathbf{G} is,

$$\mathbf{G} = \mathbf{I}_{L \times L} - 2 * (\mathbf{U}^T * \mathbf{V})^{-1} * \mathbf{U} * \mathbf{V}^T \quad (5)$$

In (5), the size of \mathbf{G} is $L \times L$, and all the mathematical operations are carried out using modular algebra in the finite field \mathbb{Z}_p . Here, it can be verified that,

$$\text{mod}(\mathbf{G} * \mathbf{G}, p) = \mathbf{I}_{L \times L} \quad (6)$$

That is, the modular inverse of \mathbf{G} is \mathbf{G} itself. That means \mathbf{G} is involutory and not well suited as a cryptographic key from the security aspect. Hence, to avoid the involutory deficiency, an additional involutory matrix \mathbf{F} , which is entirely different from \mathbf{G} , is generated as,

$$\mathbf{F} = \mathbf{I}_{L \times L} - 2 * (\mathbf{X}^T * \mathbf{Y})^{-1} * \mathbf{X} * \mathbf{Y}^T \quad (7)$$

Here \mathbf{X} and \mathbf{Y} are two $L \times 1$ integer vectors different from \mathbf{U} and \mathbf{V} . Similar to matrix \mathbf{G} , we have,

$$\text{mod}(\mathbf{F} * \mathbf{F}, p) = \mathbf{I}_{L \times L} \quad (8)$$

Let us define two integer matrices \mathbf{E} and \mathbf{D} as,

$$\mathbf{E} = \text{mod}(\mathbf{G} * \mathbf{F}, p) \quad (9)$$

$$\mathbf{D} = \text{mod}(\mathbf{F} * \mathbf{G}, p) \quad (10)$$

Since \mathbf{G} and \mathbf{F} are derived from different vector sets, they are non-commutative. That is, $(\mathbf{G} * \mathbf{F}) \neq \mathbf{F} * \mathbf{G}$. Therefore, the

encryption and decryption parent matrices \mathbf{E} , and \mathbf{D} are numerically dissimilar.

Now, let us evaluate the products $\mathbf{E} * \mathbf{D}$ and $\mathbf{D} * \mathbf{E}$. For easier writing, the mod prefix and p are omitted while writing the expressions for the matrices. Then,

$$\mathbf{Z} = \text{mod}(\mathbf{E} * \mathbf{D}, p) = \mathbf{E} * \mathbf{D} \quad (11)$$

$$\mathbf{W} = \text{mod}(\mathbf{D} * \mathbf{E}, p) = \mathbf{D} * \mathbf{E} \quad (12)$$

From (9), (10), and (11),

$$\mathbf{Z} = (\mathbf{G} * \mathbf{F}) * \mathbf{F} * \mathbf{G} = \mathbf{G} * (\mathbf{F} * \mathbf{F}) * \mathbf{G} \quad (13)$$

Substituting (8) and then (6) in (13) gives,

$$\mathbf{Z} = \mathbf{E} * \mathbf{D} = \mathbf{I}_{L \times L} \quad (14)$$

Similarly, it can be shown that,

$$\mathbf{W} = \mathbf{D} * \mathbf{E} = \mathbf{I}_{L \times L} \quad (15)$$

From (14) and (15), it can be seen that \mathbf{D} is the modular matrix multiplicative inverse of \mathbf{E} and vice versa as, $\mathbf{D} = \text{mmi}(\mathbf{E}, p)$ and $\mathbf{E} = \text{mmi}(\mathbf{D}, p)$. In (14) and (15), \mathbf{E} and $\mathbf{D} \in \mathbb{Z}_p^{L \times L}$. The novelty of generating mmi 's by the Householder technique is that mmi 's are obtained without directly calculating the matrix inverses in \mathbb{Z}_p (with time complexity $O(L^3 * \log_2 p)$ [39]), but using only scalar modular inverse as in (5) and (7), which has a time complexity of $O((\log_2 p)^2)$ only [40].

2) *Encryption and decryption matrix keys from matrices \mathbf{E} and \mathbf{D} :* Individual encryption keys are obtained by the row-wise splitting of the parent matrix \mathbf{E} of size $L \times L$, into four sub matrices $\mathbf{E}_1, \mathbf{E}_2, \mathbf{E}_3$, and \mathbf{E}_4 as,

$$\mathbf{E}_{L \times L} = \begin{bmatrix} \mathbf{E}_{1, (n \times L)} \\ \mathbf{E}_{2, (n \times L)} \\ \mathbf{E}_{3, (2 \times L)} \\ \mathbf{E}_{4, (2 \times L)} \end{bmatrix} \quad (16)$$

In (16), the selected size of each sub matrix is marked in its subscript.

3) *Selection of the sizes of $\mathbf{E}_1, \mathbf{E}_2, \mathbf{E}_3$ and \mathbf{E}_4 :* In RDSR-RJE, the matrix keys \mathbf{E}_1 and \mathbf{E}_2 are used to encrypt the plain image matrices of size $k \times n$ by post multiplication. Therefore, the number of rows of \mathbf{E}_1 and \mathbf{E}_2 are set to n to match the column size of the plain image matrices. On the other hand, \mathbf{E}_3 and \mathbf{E}_4 are used to encrypt the signature matrix and the randomizing matrix, whose sizes are $k \times 2$. Hence the row sizes of \mathbf{E}_3 and \mathbf{E}_4 are set to 2. The detailed encryption process is given in section IV, where the significance of the sizes of submatrices $\mathbf{E}_1, \mathbf{E}_2, \mathbf{E}_3$, and \mathbf{E}_4 will become clear. The choice of using 2 instead of n , for the row sizes of \mathbf{E}_3 and \mathbf{E}_4 , is to reduce the overall key sizes and the resulting cipher matrix size to keep the ciphertext expansion ratio at a lower value.

The total number of rows of the sub matrices $\mathbf{E}_1, \mathbf{E}_2, \mathbf{E}_3, \mathbf{E}_4$ is $(n+n+2+2) = 2*n+4$. Therefore from (16), it can be seen that

$$L = 2*n + 4 \quad (17)$$

The decryption keys are obtained by the column-wise splitting of the parent matrix D into four submatrices as,

$$D_{L \times L} = [D_{1,(L \times n)} \quad D_{2,(L \times n)} \quad D_{3,(L \times 2)} \quad D_{4,(L \times 2)}] \quad (18)$$

Now substituting in (14), for E and D from (16) and (18) gives,

$$\begin{bmatrix} E_1 \\ E_2 \\ E_3 \\ E_4 \end{bmatrix} * [D_1 \quad D_2 \quad D_3 \quad D_4] = I_{L \times L} \quad (19)$$

Now, expanding the LHS and the RHS of (19) in terms of compatible submatrices gives,

$$\begin{bmatrix} E_1 * D_1 & E_1 * D_2 & E_1 * D_3 & E_1 * D_4 \\ E_2 * D_1 & E_2 * D_2 & E_2 * D_3 & E_2 * D_4 \\ E_3 * D_1 & E_3 * D_2 & E_3 * D_3 & E_3 * D_4 \\ E_4 * D_1 & E_4 * D_2 & E_4 * D_3 & E_4 * D_4 \end{bmatrix} = \begin{bmatrix} I_{n \times n} & 0_{n \times n} & 0_{n \times 2} & 0_{n \times 2} \\ 0_{n \times n} & I_{n \times n} & 0_{n \times 2} & 0_{n \times 2} \\ 0_{2 \times n} & 0_{2 \times n} & I_{2 \times 2} & 0_{2 \times 2} \\ 0_{2 \times n} & 0_{2 \times n} & 0_{2 \times 2} & I_{2 \times 2} \end{bmatrix} \quad (20)$$

From (20), it can be seen that, for $i, j = 1$ to 2 ,

$$E_i * D_j = \text{mod}(E_i * D_j, p) = \begin{cases} I & \text{if } j = i \\ 0 & \text{if } j \neq i \end{cases} \quad (21)$$

Here, I is the identity matrix, and 0 is the all-zero matrix of matching sizes.

Thus, a new way of generating index-wise orthogonal key matrices are derived via Householder Construction. The property represented by (21) plays an important role in the encryption and decryption process of RDSR-RJE. In (21), E_i 's are the encryption keys, and D_j 's are the decryption keys in \mathbb{Z}_p . All the encryption and decryption are generated by the Key Generation Center administered by the image owner. The decryption keys are sent to the respective receivers of the end users, through the secured channels.

IV. RDSR-RJE ENCRYPTION AND DECRYPTION

The architectural layout of RDSR-RJE scheme is shown in Fig. 1. The two major components are:

- RDSR-RJE Encrypter, which is also the owner of images and the corresponding RD-BM's.
- RDSR-RJE Receiver.

The RDSR-RJE Encrypter consists of the de-identification (DI) unit and the joint encryption (JEnc) unit. In RDSR-RJE, it is assumed that the ROI to be encrypted has been determined by a suitable method [9] and is made available as the Region Defining Binary Mask, **RD-BM**, whose size is same as that of A . In **RD-BM**, the ROI pixels of A are set to ones and the non-ROI pixels to zeros.

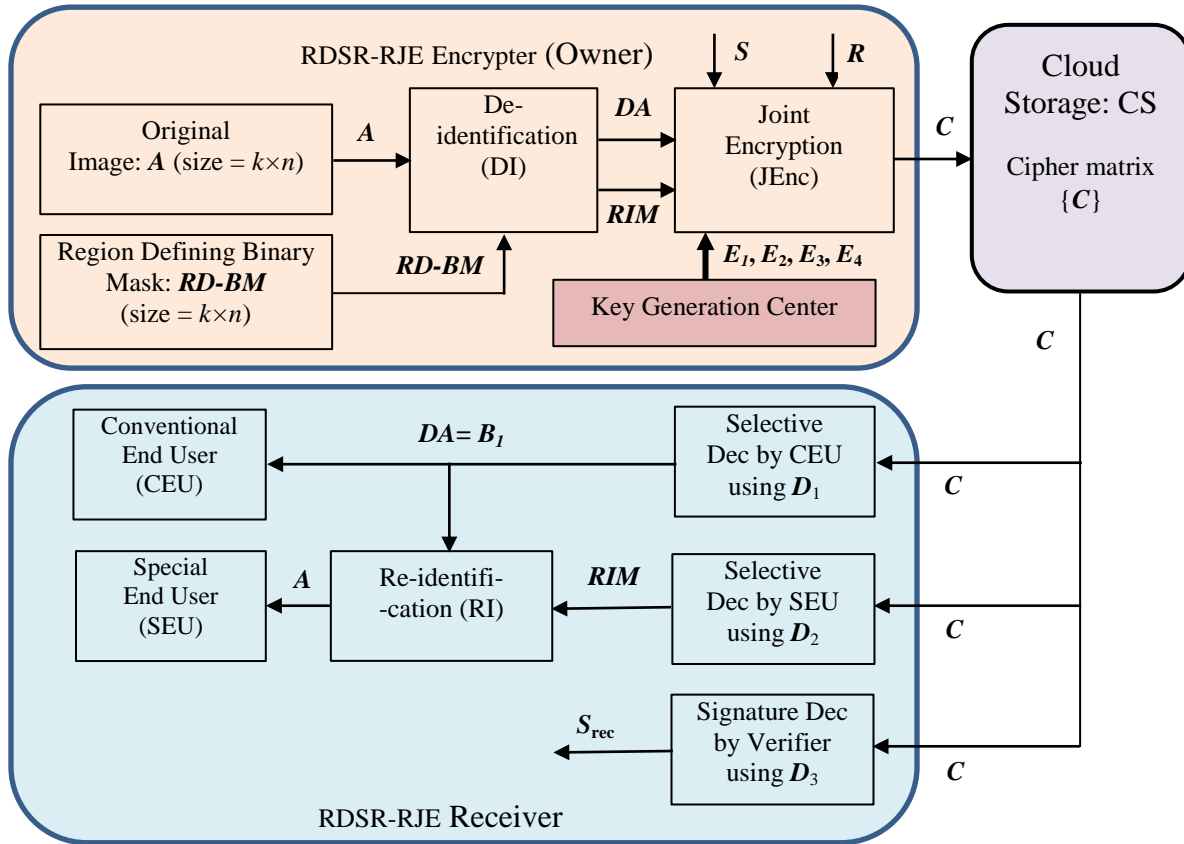


Fig. 1. The architectural layout of RDSR-RJE transmitter and receiver.

A. De-identification by Selective Encryption

The basic selective encryption process for de-identification is depicted in Fig. 2. The original input image to be de-identified is represented by matrix A of size $k \times n$ with data type uint8. (For a square image, $k = n$). Hereafter, when there is no ambiguity, ‘image matrix A ’ and ‘image A ’ are used synonymously.

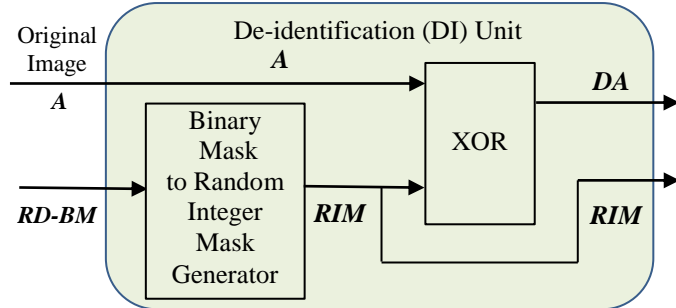


Fig. 2. The De-identification unit.

B. Binary Mask to Random Integer Mask Generator

In RDSR-RJE, the ‘Binary Mask to Random Integer Mask Generator’ transforms the $RD-BM$ into RIM . In generating RIM , the 1’s of the $RD-BM$ are replaced by random integers in the range 1-254 to get maximum diversity. Here, integer 0 is avoided as it does not make any change when XORed with any other number. Similarly, integer 255 is avoided to avoid the exact complements. The zeros of $RD-BM$ are retained as they are in getting RIM . A toy example of an 8×8 $RD-BM$ is shown in Fig. 3(a). The corresponding RIM is shown in Fig. 3(b). The number of zeros and their locations in the $RD-BM$ are same as in its RIM . The number of integers greater than zero in RIM are equal to the number of ones of $RD-BM$.

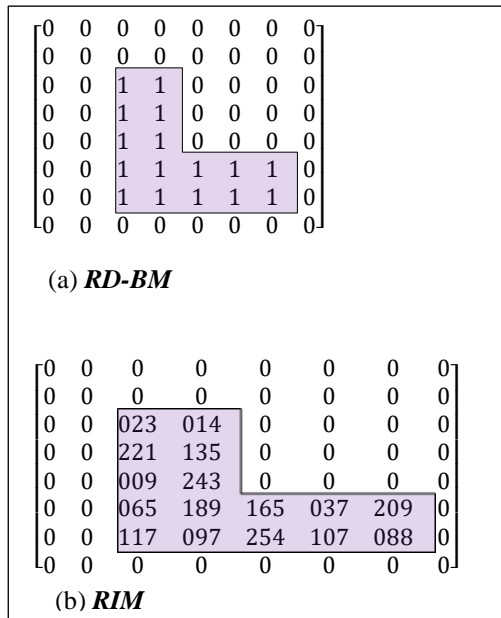


Fig. 3. RD-BM of size 8×8 and its RIM.

C. XOR De-identification

The bitwise XOR de-identification is carried out to get the de-identified out DA (De-identified A) of size $k \times n$ as,

$$DA = \text{bitxor}(A, RIM) \quad (22)$$

The zeros and their locations in RIM represent the non-ROI regions of A . Therefore, the $\text{bitxor}(\dots)$ operation does not change those pixel values. Thus the non-ROI areas of A and DA are exactly the same. On the other hand, the ROI pixel values get randomly altered at ROI locations of A due to the bitwise XOR of operations of nonzero operands. On the receiving side, the original matrix A is recovered as $A = \text{bitxor}(DA, RIM)$.

D. Security Aspects of Bitwise XOR Encryption

1) *Blind guess of RIM*: Let nz be the number of nonzero elements in the RIM used in (22). Each element can take any value between 1 and 254 (254 possibilities). Then the probability of correctly guessing a single element is $(1/254)$, and that of correctly guessing all the elements of RIM is $(1/254)^{nz} = 254^{-nz}$, which is a very meager fraction, even for a moderately sized ROI.

2) *Protection against chosen plain text attack (CPA)*: The most vulnerable attack on XOR encryption is CPA, whereby knowing A and DA , the attacker can get hold of RIM as,

$$RIM = \text{bitxor}(A, DA) \quad (23)$$

But the RIM used in (22) is randomly varied from encryption to encryption. Here the RIM acts as an OTP. Therefore, the RIM captured by the attacker using (23) is useless for the succeeding encryptions. Thus the CPA and ‘many time pad’ attacks are eliminated.

3) *Accessibility and manipulation of DA*: In RDSR-RJE, the XOR encryption is not a standalone operation. The output matrix DA is internal to the sender, and it is further encrypted, as will be explained in the next section. Hence, DA is not directly accessible to the attacker, and any manipulation to DA is flagged off by the subsequent message authentication scheme.

E. Joint Encryption using Matrix Keys

Joint encryption (JEnc) is the new innovation of RDSR-RJE. Here, four data matrices are encrypted jointly with four corresponding encryption matrix keys E_1, E_2, E_3 and E_4 to get the cipher matrix C as,

$$C = \text{mod}((DA * E_1 + RIM * E_2 + S * E_3 + R * E_4), p) \quad (24)$$

Here, S is the signature matrix of size $k \times 2$ and R is the randomization matrix of size $k \times 2$, which is altered for successive encryptions. The sizes of DA and RIM are $k \times n$. The sizes of encryption key matrices E_1 and E_2 , are $n \times L$ and the sizes of E_3 and E_4 are $2 \times L$. The size of the cipher matrix C is $k \times L$ where $L = 2 * n + 4$. Matrix C is calculated using modular algebra in Z_p . Then, Eqn. (24) can be written in a simple form as,

$$C = DA * E_1 + RIM * E_2 + S * E_3 + R * E_4 \quad (25)$$

Formation of C using (25) is called the ‘Joint Encryption (JEnc)’ to indicate that matrix C is obtained by encrypting four data matrices by four matrix keys to get a single weighted sum. After JEnc, the cipher matrix C is sent to the CS for secure storage and subsequent distribution. When the sizes of the matrices are large, the partial sums on the RHS may exceed the max(int) level of the computing device. In such cases, the integer overflow error can be avoided by adopting ‘cumulative summation’ where one term is added at a time to the partial sum followed by the mod operation, as shown below.

1) *Cumulative addition*: The cumulative addition is carried out as follows

$$\begin{aligned} C &= []; C=C+\text{mod}(DA*E1,p); C=\text{mod}(C,p); \\ C &= C+\text{mod}(RIM*E2,p); C=\text{mod}(C,p); \\ C &= C+\text{mod}(S*E3,p); C=\text{mod}(C,p); \\ C &= C+\text{mod}(R*E4,p); C=\text{mod}(C,p); \end{aligned}$$

The mod operation in each step keeps the result between 0 and $(p-1)$. Thus the final sum C also remains within Z_p .

F. Randomization Matrix R

The integer matrix R , of size $k \times n$ (same as that of DA and RIM), on the RHS of (25) provides randomization to the encryption. R varies randomly from the present encryption to the next encryption. Thus, for the same inputs DA , RIM and S , the output C will differ in successive encryptions because of R . Thus, randomized encryption is achieved, which prevents the Chosen Ciphertext Attack (CCA). The elements of R are chosen randomly from the uniform distribution in the range 0 to 255.

G. Signature Matrix S

The signature matrix S of size $k \times n$ (same as that of DA and RIM) provides image and source authentication. In general, Matrix S contains the ID of the source, time stamp for source authentication, and suitable hash value of the original image matrix A . The source ID acts as the digital signature which is made available to the signature verifier at the receiving end.

H. Selective Decryption

The blocks used in selective decryption operation are shown in Fig. 1.

1) *Decryption by a Conventional End User (CEU)*: The cipher matrix C can be decrypted by a Conventional End User (CEU) using the decryption key D_1 . The CEU, on decryption, recovers the de-identified image matrix DA . The decryption is carried out as,

$$B_1 = \text{mod}(C * D_1, p) = C * D_1 \quad (26)$$

On substituting for C from (25) in (26), we get,

$$B_1 = (DA * E_1 + RIM * E_2 + S * E_3 + R * E_4) * D_1 = DA * E_1 * D_1 + RIM * E_2 * D_1 + S * E_3 * D_1 + R * E_4 * D_1 \quad (27)$$

From the property (21), $E_1 * D_1 = I$, and the other product terms $E_2 * D_1, E_3 * D_1$ and $E_4 * D_1$ are all zeros. Hence,

$B_1 = DA$ which proves the correctness of the decryption by the CEU.

2) *Selection of the modulus p* : For the existence of modular inverses in Z_p , the p value has to be a prime integer. In RDSR-RJE, the encryption and the decryption operations are carried out on images whose elements belong to uint8 with a maximum value = 255. Now consider the decryption, $B_1 = \text{mod}(C * D_1, p)$. Here, due to the mod(...) operation, the maximum element of B_1 is less than p . However, B_1 (which is same as DA) represents the de-identified image where the maximum element can go up to 255. Therefore, for the correct realization of DA at the receiver, the constraint is $255 < p$. That is, p should be greater than 255. The immediate higher prime number is 257, and hence **p is chosen to be 257**. Using a higher prime number for p , unnecessarily increases the cipher text size.

3) *Decryption by the special end user (SEU)*: The special end user (SEU) has the decryption key D_2 as well as D_1 . Using D_2 the SEU decrypts C as,

$$B_2 = \text{mod}(C * D_2, p) = C * D_2 \quad (28)$$

On substituting for C from (25) in (28) and again using the property (21) we get $B_2 = RIM$. The SEU also decrypts C using D_1 to get DA as explained in section IV.H.1. Then the SEU can recover the original A as,

$$B = \text{bitxor}(DA, RIM) \quad (29)$$

From (29) and (23), it can be seen that B is exactly equal to A , and thus the decrypted matrix B is the exactly re-identified version of DA . Therefore, RDSR-RJE achieves lossless reverse de-identification.

4) *Decryption of the signature matrix S* : The signature matrix S is detected using the decryption key D_3 as,

$$B_3 = \text{mod}(C * D_3, p) = C * D_3 \quad (30)$$

On substituting for C from (25) in (30) and again using the property (21), we get $B_3 = S$.

5) *Signature verification*: In RDSR-RJE, signature verification takes place before decryption. Therefore, if the verification fails, there is no need for decryption. This pre-signature verification scheme results in faster processing. The Signature Verifier (SV) at the receiving end should possess the signature decryption key D_3 . The SV should have already received earlier, the true signature S_{true} from the data owner. Now, the decrypted B_3 gives the received signature S_{rec} . Thus, on receiving C , the SV recovers S_{rec} . If $S_{\text{rec}} = S_{\text{true}}$, then the authentication is successful, and the received C is accepted. Otherwise, there is some error, and the present C is discarded, and suitable countermeasures are deployed for further investigation, or the receiver may request retransmission. Here, both CEU and SEU possess D_3 , and after signature verification, can proceed for further decryption. Since the signature matrix is encrypted using the encryption key E_3 , it is non-forgable and achieves non-repudiation as the signature matrix carries the sender's ID.

I. RDSR-RJE Encryption and Decryption Algorithms

RDSR-RJE encryption and decryption algorithms are presented in this section. Modular algebra in Z_p is used for all calculations.

Algorithm RDSR-RJE encrypt

Inputs: Original image A . Mask matrix $RD-BM$. Signature matrix S . Encryption keys E_1, E_2, E_3 and E_4 .

Output: Cipher matrix C .

1. Get the Random Integer Matrix, RIM from $RD-BM$ as explained in section IV.A.1.
2. Get the de-identified image matrix DA by the XOR encryption, as given by (22).
3. Generate the randomization integer matrix R using any standard library function from python, C++, Java, etc.
4. Jointly encrypt DA, RIM, S , and R to get the cipher matrix C , using the Encryption keys E_1, E_2, E_3 , and E_4 as given by (24).
5. Over.

Decryption Algorithm used by the conventional end user (CEU) is summarized as follows.

Algorithm RDSR-RJE-CEU decrypt

Inputs: Cipher matrix C . Decryption keys D_1 and D_3 . True signature matrix S_{true} .

Output: De-identified image matrix DA .

1. Get S_{rec} using (30) as, $S_{rec} = C * D_3$.
2. Verify the signature as:
 - If $S_{rec} \neq S_{true}$ //Signature failure
Discard the present C , and request for Retransmission.
Go to step 4.
 - Else //Signature OK
Go to step 3.
 - Endif
3. Get DA using (26), as $DA = C * D_1$.
4. Over.

RDSR-RJE decryption by the Special End User is an extension of the algorithm **RDSR-RJE-CEU decrypt**.

Algorithm RDSR-RJE-SEU decrypt by the Special End User

Inputs: Cipher matrix C . Decryption keys D_1, D_2 , and D_3 . **Output:** Original image matrix A .

1. Get DA according to the algorithm **RDSR-RJE-CEU**
 2. Get RIM using (28), as $RIM = C * D_2$.
 3. Get the original re-identified matrix A , using (29) as, $A = \text{bitxor}(DA, RIM)$.
 4. Over.
-

J. Characteristics of RDSR-RJE Encryption/Decryption

1) *Ciphertext expansion ratio:* The Ciphertext Expansion Ratio (CER) is the ratio of the size of the cipher matrix to that of its plain matrix. For a given plain matrix, higher the value of CER, higher is the size of its cipher matrix, and consequently, the computational and communication cost becomes relatively higher. A lower CER value contributes to a higher degree of encryption efficiency. CER is defined as,

$$CER = \frac{\text{Size of Ciphermatrix in bits}}{\text{Size of Plain matrix in bits}} = \frac{\text{Size of } C \text{ in bits}}{\text{Size of } A \text{ in bits}}$$

In RDSR-RJE, the uppermost value of an element in the cipher matrix is C is $(p-1)$. Hence, the number of bits needed to represent an element of C is $\text{ceil}(\log_2(p-1))$ bits. With $p = 257$, $\text{ceil}(\log_2(p-1)) = 8$. Thus, 8 bits are required to represent each element of C . Now, the size of C (No. of elements in C) is $k \times L$. Hence the total size of C in bits is $k * L * 8$. The plain matrix has a bit depth of 8, and its size is $k \times L$. Therefore, the total size of A (or DA) in bits is $k * n * 8$. Hence,

$$CER = \frac{k * L * 8}{k * n * 8} = \frac{L}{n} \quad (31)$$

On substituting for L from (17), and when n is large compared to 2,

$$CER = \frac{L}{n} = \frac{2 * n + 4}{n} \cong 2 \quad (32)$$

An important characteristic of RDSR-RJE is that the CER value is constant and does not increase with n .

2) *Lossless reversible de-identification:* In RDSR-RJE, the decrypted image is the exact replica of the original image for the conventional end user or the special end user. Thus, it is a zero-loss scheme.

Additionally, RDSR-RJE does not use block-wise operations. It avoids floating point operations and iterative procedures. Therefore, RDSR-RJE is efficient and achieves higher execution speed.

K. Security Aspects of RDSR-RJE

1) *Exhaustive search for keys:* Each element of an encryption or decryption key belongs to Z_p whose range is 0 to $(p-1)$. Therefore an element of a key can take any one value out of p possibilities. Hence the probability of correctly guessing a single element is $(1/p)$. Each key has $n * L$ elements. Therefore, the probability of guessing all the elements correctly is $(1/p)^{n * L} = p^{-n * L}$ which is extremely a very low value for $p = 257$. Thus, the success of an exhaustive search is negligibly small.

2) *Protection against chosen plain text attack (CPA):* In RDSR-RJE, the encryption process is randomized using the random matrix R as in (25) where R is varied from encryption to encryption. Thus, the cipher matrix would be different even if the input plain matrix is same for consecutive encryptions. Hence, the randomized encryption prevents CPA.

V. EXPERIMENTAL RESULTS AND DISCUSSION

A. Experiment 1

An axial MRI view is taken as the original grayscale plain image as shown in Fig. 4(a). The image matrix A is of size 512×512 . The regions selected for de-identification are the textual details that include the patient's name, date of image production *etc.* The selected regions are shown in the **RD-BM** of Fig. 4(b). The corresponding Random Image Mask (**RIM**) is shown in Fig. 4(c). The de-identified (selectively encrypted) image **DA**, as given by (22), is shown in Fig. 4(d). The result of JEnc, matrix C obtained using (25) is shown in Fig. 4(e) which shows a very high degree of randomness.

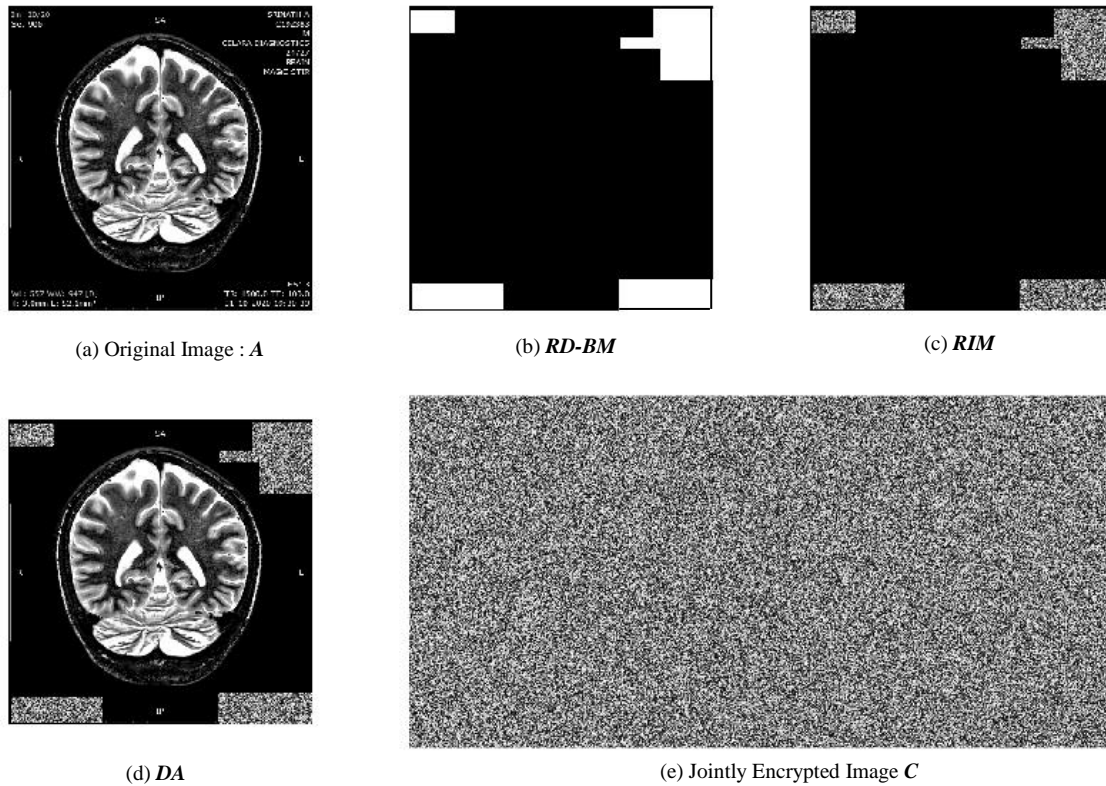


Fig. 4. Original Image A, RD-BM, RIM and the encrypted image C (Experiment 2).

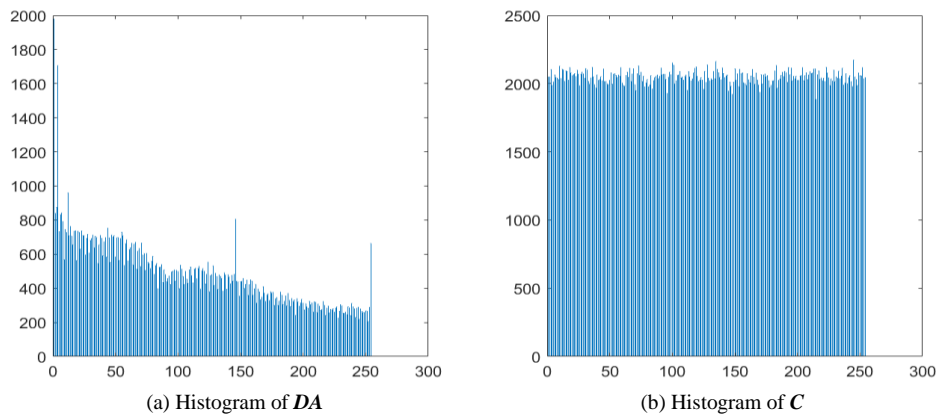


Fig. 5. Histograms of DA and C.

1) *Histograms of DA and C*: Histograms of the de-identified image **DA** and the jointly encrypted cipher image **C** are shown in Fig. 5. From Fig. 5, it can be seen that the histogram of **C** is uniformly distributed compared to that of **DA**. This shows the comprehensive randomness of **C**. A malicious attacker cannot deduce any information from the histogram of **C**.

2) *Comparison of visual correlation coefficients*: Adjacent pixel values of a normal image, have a higher correlation between them where as in a good cipher image, the corresponding correlation should be very low. That means, in a cipher image, the adjacent pixel values are highly dispersed. Hence the correlation coefficient [29] will be low.

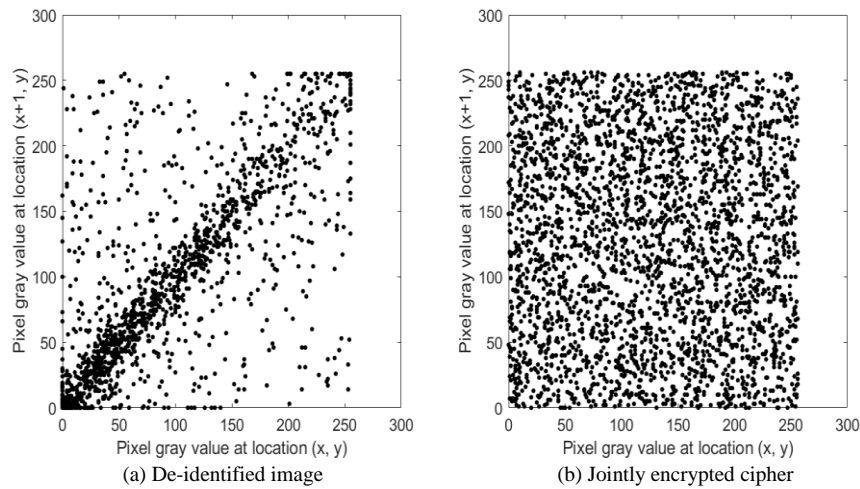


Fig. 6. Comparison of correlation coefficients between DA and C.

Comparison of pixel value dispersions along horizontal direction, between *DA* and the cipher image *C* of Experiment 1, is shown in Fig. 6, from which it can be seen that the cipher image *C* has a high degree of pixel value dispersion compared to *DA*.

B. Experiment 2

In this experiment, an ultrasound scan of pregnancy is the original image *A* with size 187×269, as shown in Fig. 7(a). The region selected for de-identification is the central part of the *A* as shown by the binary mask, *RD-BM* shown in Fig. 7(b), and the corresponding *RIM* is shown in 7(c). The de-identified (selectively encrypted) image *DA*, as given by (22), is shown in Fig. 7(d). The cipher matrix *C* obtained using (25) is shown in Fig. 7(e), which shows a very high degree of randomness.

C. Metrics for Comparison

A few metrics for comparing the image encryption schemes are discussed in this section.

1) *Differential analysis*: Differential analysis is the study of the variations in the cipher matrix when the plain matrix changes by a small value. Thus, it is basically a sensitivity analysis. A quantitative measure of this behavior is NPCR which stands for the Number of Pixels Change Rate.

Let C_1 be the cipher image of a given plain image. Let C_2 be the resulting cipher image after a one-bit change in the plain image. The differential change per pixel is defined as,

$$d(i, j) = \begin{cases} 1 & \text{if } C_1(i, j) \neq C_2(i, j) \\ 0 & \text{otherwise} \end{cases} \quad (33)$$

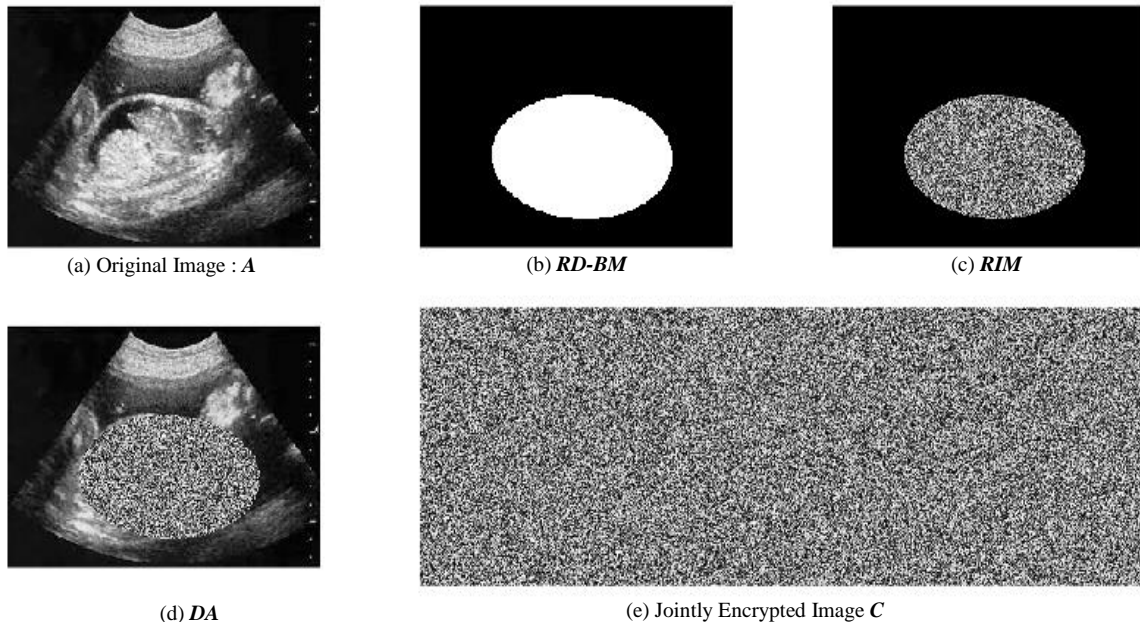


Fig. 7. Original Image A, RD-BM, RIM, DA and the encrypted image C (Experiment 2).

for $i = 1$ to h and $j = 1$ to w where h and w are the height and width of the cipher image. Then, the percentage NPCR is defined [41] as,

$$NPCR = \frac{\sum_{i=1}^h \sum_{j=1}^w d(i,j)}{h*w} * 100 \quad (34)$$

A higher NPCR means the attacker cannot capture the encryption due to the large number of indeterminates. Thus, higher the NPCR, higher is the security of the encryption process. The ideal value of the NPCR is 100%. Another metric that measures the differential score is UACI (unified averaged changed intensity) which is defined [41] as,

$$UACI = \frac{\sum_{i=1}^h \sum_{j=1}^w |C_1(i,j) - C_2(i,j)|}{h * w * 256} * 100 \quad (35)$$

Higher the value of UACI, better is the encryption performance.

2) *Image entropy*: The entropy of an image in bits/pixel, is defined as,

$$H = - \sum_{i=0}^{i=imax} p_i * \log_2(p_i) \quad (36)$$

where p_i is the probability of a pixel having the gray level i , and $imax$ is the maximum gray level (255 in a normal image). For a fully random image, $H = 8$ bits/pixel.

3) *Structural similarity index*: The structural similarity index (SSIM) [30] measures the closeness between two images. In RDSR-RJE the de-identified image \mathbf{DA} and its decrypted version \mathbf{B}_1 are exactly same. Hence SSIM value in RDSR-RJE is 1. SSIM values less than 1 imply recovery of the plain image with error.

D. Comparison of the Performance of RDSR-RJE

The encryption efficiency parameters of RDSR-RJE are compared with those of HOSSAIN [28], QIN [29], and JITHIN [30]. The numerical results are shown in Table I, for images ‘Img 1’ and ‘Img 2’ which are from Fig. 4(a) and 7(b), respectively.

1) *Execution time*: The theoretical time complexity calculations of the different methods [28-30] are extensive and depend on the respective contexts. Therefore, the execution times of the encryption algorithms are obtained experimentally and shown in the plots of Fig. 8. Here, the image used in Experiment 1 is resized starting from 64x64 and progressively increased upto 512x512 as marked in Fig. 7. Then the corresponding execution times are calculated using the appropriate Matlab code. In Fig. 8, the execution time of joint encryption by RDSR-RJE, is shown in black.

The values obtained in Fig. 8 are machine-dependent, and thus the execution times are relative only. From Fig. 8, it can be seen that RDSR-RJE has a significant lower execution time compared to the other three methods. For example, when the image size is 256x256, the percentage improvement in the execution time of RDSR-RJE compared to that of HOSS [28] is, $(30.38 - 21.35) * 100 / 30.38$, which is approximately equal to 30%.

TABLE I. COMPARISON OF THE QUANTITATIVE VALUES OF THE METRICS

	Plain Image	Encrypted Image			
		Horizontal Correlation Coefficient			
		RDSR-RJE	HOSS [28]	QIN [29]	JITHIN [30]
Img 1	0.8772	0.0025	0.0027	0.0029	0.0037
Img 2	0.7421	0.0227	0.0312	0.0412	0.0467
Image Entropy					
Img 1	4.8876	7.9974	7.9948	7.9953	7.9916
Img 2	7.2674	7.9961	7.9943	7.9951	7.9919
Number of Pixels Change Rate (NPCR) in percentage					
Img 1	----	99.6037	99.5100	99.4991	99.4211
Img 2	----	99.5965	99.4235	99.4173	99.4053
Unified Averaged Changed Intensity (UACI) in percentage					
Img 1		33.5239	32.9932	32.4327	32.2327
Img 2		33.2405	32.0159	32.0079	31.8953
Structural Similarity Index (SSIM)					
Img 1		1.00	0.9879	0.9752	0.9623
Img 2		1.00	0.9693	0.9533	0.9457

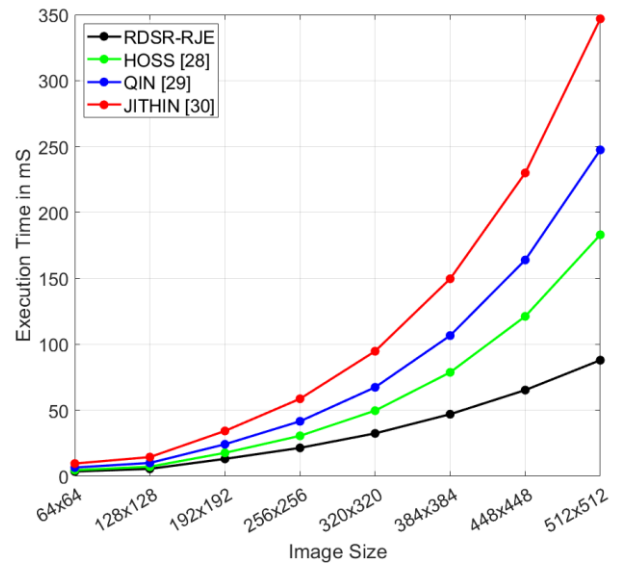


Fig. 8. Comparison of execution times for image encryption.

VI. CONCLUSION

A new method of reversible image de-identification by encryption has been presented. It uses matrix keys for asymmetric encryption and decryption of image matrices. Content and source authentication via a digital signature scheme is integrated using joint encryption. All the cryptographic operations are carried out in the finite field Z_p and thus avoid the floating point operations that lead to higher computational speed. Moreover, the algorithm is non-iterative and does not use block-wise operations to achieve faster results. Here, the decrypted image is the exact replica of the original image, and thus, it is a zero-loss scheme. Additionally, the encryption/decryption security-related performance parameters, namely, entropy, correlation coefficients, NPCR, and UACI, are very near to their ideal values. The proposed method, on average, reduces the execution time of homomorphic encryption by 30 to 40 percent.

REFERENCES

- [1] Raj, B.S.S., Venugopalachar, S. Multi-data Multi-user End to End Encryption for Electronic Health Records Data Security in Cloud. *Wireless Pers Commun* 125, 2413–2441 (2022). <https://doi.org/10.1007/s11277-022-09666-2>.
- [2] Hossein Ghanbari-Ghalehjoughi, Mansour Eslami, Sohrab Ahmadi-Kandjani, Mohsen Ghanbari-Ghalehjoughi, Zeyun Yu, Multiple layer encryption and steganography via multi-channel ghost imaging, *Optics and Lasers in Engineering*, Volume 134, 2020, 106227, ISSN 0143-8166, pp. 1-12.
- [3] Kaur, M., Kumar, V. A Comprehensive Review on Image Encryption Techniques. *Arch Computat Methods Eng* 27, 15–43 (2020). <https://doi.org/10.1007/s11831-018-9298-8>.
- [4] A.S. Sajitha, A. Shobha Rekh, "Review on various image encryption schemes, *Materials Today: Proceedings*, Volume 58, Part 1," 2022, Pages 529-534, doi: 10.1016/j.matpr.2022.03.058.
- [5] K. Suneja, S. Dua and M. Dua, "A Review of Chaos-based Image Encryption," 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC), 2019, pp. 693-698, doi: 10.1109/ICCMC.2019.8819860.
- [6] Jameel, E. A., & Fadhel, S. A. (2022). Digital Image Encryption Techniques: Article Review. *Technium: Romanian Journal of Applied Sciences and Technology*, 4(2), pp. 24–35. Doi: 10.47577/technium.v4i2.6026.
- [7] C. Tiken and R. Samli , "A Comprehensive Review About Image Encryption Methods", *Harran Üniversitesi Mühendislik Dergisi*, vol. 7, no. 1, pp. 27-49, Apr. 2022, doi:10.46578/humder.1066545.
- [8] G.S. Nelson, Practical Implications of Sharing Data: A Primer on Data Privacy, Anonymization, and De-Identification (2015), pp. 1–23 (http://thotwave.com/wpcontent/uploads/2015/09/data_sharing_privacy_anonymization_and_deidentification_rev_13.pdf) (accessed 25.06. 15).
- [9] Rafael C. Gonzalez, Richard E. Woods, Steven L. Eddins, "Digital Image Processing Using MATLAB", (Chapters 11 and 12), Third Edition, Gatesmark Publishing@. A Division of Gatesmark, © LLC, [Knoxville], Tennessee, USA, 2020.
- [10] Jawad, Lahieb & Sulong, Ghazali. (2015). A Survey on Emerging Challenges in Selective Color Image Encryption Techniques. *Indian Journal of Science and Technology*. 8. pp. 1-13. doi:10.17485/ijst/2015/v8i27/71241.
- [11] Slobodan Ribaric, Aladdin Ariyaeeinia, Nikola Pavesic, "De-identification for privacy protection in multimedia content: A survey," *Signal Processing: Image Communication*, Volume 47, 2016, Pages 131-151, <https://doi.org/10.1016/j.image.2016.05.020>.
- [12] A. J. Paul, "Recent Advances in Selective Image Encryption and its Indispensability due to COVID-19," 2020 IEEE Recent Advances in Intelligent Computational Systems (RAICS), 2020, pp. 201-206, doi: 10.1109/RAICS51191.2020.9332513.
- [13] Khan, Jan Sher & Ahmad, Jawad. (2019). Chaos based efficient selective image encryption. *Multidimensional Systems and Signal Processing*. 30. doi: 10.1007/s11045-018-0589-x.
- [14] Cun, Q., Tong, X., Wang, Z., & Zhang, M. (2021). Selective image encryption method based on dynamic DNA coding and new chaotic map. *Optik*, 243, 167286. pp. 1-29. doi:10.1016/j.ijleo.2021.167286.
- [15] Kiran and Parameshachari B. D. "Selective Image Encryption of Medical Images Based on Threshold Entropy and Arnold Cat Map," *Biosc.Biotech.Res.Comm. Special Issue Vol 13. No 13. 2020*, pp. 194-202. doi: 10.21786/bbr/13.13/27.
- [16] Mehmet Yamac, et al. "Reversible Privacy Preservation using Multi-level Encryption and Compressive Sensing," arXiv:1906.08713v1 [cs.CR], EUSIPCO 2019, pp. 1-5. doi: 10.48550/arXiv.1906.08713.
- [17] Carrillo, Paula & Kalva, Hari & Magliveras, Spyros. (2009). Compression Independent Reversible Encryption for Privacy in Video Surveillance. *EURASIP Journal on Information Security*. 2009. pp. 1-13. doi: 10.1155/2009/429581.
- [18] Li, J., Zhang, Z., Li, S. et al. A partial encryption algorithm for medical images based on quick response code and reversible data hiding technology. *BMC Med Inform Decis Mak* 20 (Suppl 14), 297 (2020). pp. 1-16. doi: 10.1186/s12911-020-01328-2.
- [19] Chuan Qin , Zhihong He , Xiangyang Luo , Jing Dong , Reversible Data Hiding in Encrypted Image with Separable Capability and High Embedding Capacity, *Information Sciences* (2018), doi: 10.1016/j.ins.2018.07.021.
- [20] A. Hafsa, J. Malek and M. Machhout, "An Improved Security Approach for Medical Images and Patients' Information Transmission," 2022 IEEE 9th International Conference on Sciences of Electronics, Technologies of Information and Telecommunications (SETIT), 2022, pp. 19-24, doi: 10.1109/SETIT54465.2022.9875481.
- [21] P. Parida, C. Pradhan, X. -Z. Gao, D. S. Roy and R. K. Barik, "Image Encryption and Authentication With Elliptic Curve Cryptography and Multidimensional Chaotic Maps," in *IEEE Access*, vol. 9, pp. 76191-76204, 2021, doi: 10.1109/ACCESS.2021.3072075.
- [22] X. Li, D. Xiao, H. Mou, D. Lu and M. Peng, "A Compressive Sensing Based Image Encryption and Compression Algorithm With Identity Authentication and Blind Signcryption," in *IEEE Access*, vol. 8, pp. 211676-211690, 2020, doi: 10.1109/ACCESS.2020.3039643.
- [23] T. S. Ali and R. Ali, "A Novel Medical Image Signcryption Scheme Using TLTS and Henon Chaotic Map," in *IEEE Access*, vol. 8, pp. 71974-71992, 2020, doi: 10.1109/ACCESS.2020.2987615.
- [24] R. I. Abdelfatah, "Secure Image Transmission Using Chaotic-Enhanced Elliptic Curve Cryptography," in *IEEE Access*, vol. 8, pp. 3875-3890, 2020, doi: 10.1109/ACCESS.2019.2958336.
- [25] G. Luan, A. Li, D. Zhang and D. Wang, "Asymmetric Image Encryption and Authentication Based on Equal Modulus Decomposition in the Fresnel Transform Domain," in *IEEE Photonics Journal*, vol. 11, no. 1, pp. 1-7, Feb. 2019, Art no. 6900207, doi: 10.1109/JPHOT.2018.2886295.
- [26] D. Reilly and L. Fan, "A Comparative Evaluation of Differentially Private Image Obfuscation," 2021 Third IEEE International Conference on Trust, Privacy and Security in Intelligent Systems and Applications (TPS-ISA), 2021, pp. 80-89, doi: 10.1109/TPSISA52974.2021.00009.
- [27] N. Tsafack, J. Kengne, B. Abd-El-Atty, A. M. Ilyyasu, K. Hirota, and A. A. EL-Latif, "Design and implementation of a simple dynamical 4-d chaotic circuit with applications in image encryption," *Information Sciences*, vol. 515, pp. 191–217, apr 2020.
- [28] M. B. Hossain, M. T. Rahman, A. B. M. S. Rahman and S. Islam, "A new approach of image encryption using 3D chaotic map to enhance security of multimedia component," 2014 International Conference on Informatics, Electronics & Vision (ICIEV), 2014, pp. 1-6, doi: 10.1109/ICIEV.2014.6850856.
- [29] Q. Qin, Z. Liang, S. Liu, X. Wang and C. Zhou, "A Dual-domain Image Encryption Algorithm Based on Hyperchaos and Dynamic Wavelet Decomposition," in *IEEE Access*, 2022, doi: 10.1109/ACCESS.2022.3212145.
- [30] Jithin, K. C., & Sankar, S. (2020). Colour image encryption algorithm combining Arnold map, DNA sequence operation, and a Mandelbrot set. *Journal of Information Security and Applications*, 50, 102428. doi:10.1016/j.jisa.2019.102428.
- [31] Z. Hua, Y. Zhou, and H. Huang, "Cosine-transform-based chaotic system for image encryption," *Information Sciences*, vol. 480, 2019. pp. 403–419. doi: org/10.1016/j.ins.2018.12.048.
- [32] Xingyuan Wang, Suo Gao, "Image encryption algorithm based on the matrix semi-tensor product with a compound secret key produced by a Boolean network," *Information Sciences*, Volume 539, 2020, pp. 195-214, doi: 10.1016/j.ins.2020.06.030.
- [33] Xiong, Fu & Xiao, Yang & Cao, Zhi-Guo & Gong, Kaicheng & Zhiwen, Fang & Zhou, Joey. (2019). Good practices on building effective CNN baseline model for person re-identification. 145. 10.1117/12.2524386.
- [34] Jeong Y, Yoo S, Kim Y, Shim W, "De-Identification of Facial Features in Magnetic Resonance Images: Software Development Using Deep Learning Technology," *J Med Internet Res* 2020;22(12):e22739. doi: 10.2196/22739.
- [35] Monteiro, Eriksson & Costa, Carlos & Oliveira, José. (2017). A De-Identification Pipeline for Ultrasound Medical Images in DICOM Format. *Journal of Medical Systems*. 41. 89. 10.1007/s10916-017-0736-1.
- [36] Tekli, Jimmy & AL Bouna, Bechara & Couturier, Raphaël & Tekli, Gilbert & Zein, Zeinab & Kamradt, Marc. (2019). A Framework for

- Evaluating Image Obfuscation under Deep Learning-Assisted Privacy Attacks. 1-10. 10.1109/PST47121.2019.8949040.
- [37] Jingjing Yang, Weijia Zhang, Jiaxing Liu, Jinzhao Wu, Jie Yang, "Generating De-identification facial images based on the attention models and adversarial examples," Alexandria Engineering Journal, Volume 61, Issue 11, 2022, pp. 8417-8429, <https://doi.org/10.1016/j.aej.2022.02.007>.
- [38] Stewart, G. W. "The Efficient Generation of Random Orthogonal Matrices with an Application to Condition Estimators." SIAM Journal on Numerical Analysis, vol. 17, no. 3, 1980, pp. 403-409. *JSTOR*, www.jstor.org/stable/2156882. Accessed 22 Mar. 2021.
- [39] Jacques-García, Fausto & Uribe-Mejía, Daniel & Macías-Bobadilla, Gonzalo & Chaparro-Sánchez, Ricardo. (2019). On modular inverse matrices A computational approach. South Florida Journal of Development, Miami, vol.3, no.3. pp.3100-3111, 2019.
- [40] Alfred J. Menezes, Paul C. van Oorschot and Scott A. Vanstone, "Handbook of Applied Cryptography (last updated July 8, 2011)," CRC Press, Inc. Boca Raton, FL, USA ©1996.
- [41] Wu, Yue. (2011). "NPCR and UACI Randomness Tests for Image Encryption," Cyber Journals: Journal of Selected Areas in Telecommunications. April, 2011. pp.31-38.

Texture Analytics for Accurate Person Recognition: A Multimodal Approach

Suchetha N V¹, Sharmila Kumari M²

Research Scholar-Department of CSE, PA College of Engineering Mangaluru, India and
Department of CSE, Sri Dharmasthala Manjunatheshwara Institute of Technology, Ujire-574240 and
Affiliated to Visvesvaraya Technological University, Belagavi, Karnataka, India¹
PA College of Engineering Mangaluru, India²

Abstract—Securing the resources is a most challenging task in the digital era. Traditionally, password and ID card systems were used to provide security. Password and ID cards can be stolen or hacked; to overcome this drawback biometric systems are used to authenticate the user to access the data or resources. Biometric system uses physical and behavioral characteristics of the user. Biological characteristics of the person like face, fingerprint, iris, palm print, voice, hand geometry etc. cannot be stolen and misused. Even though unimodal biometric system is more secure as compared to the traditional approach, it is not able to handle intra-class, inter-class variations, noisy data and spoofing attack. These problems can be solved using multimodal biometrics. In this paper, we discuss unimodal biometric system using Local Binary Pattern (LBP) and Local Ternary Pattern (LTP). We propose a feature level fusion of face and fingerprint biometric traits using LTP. The implementation of the introduced system stands in comparison to the unimodal LBP and LTP for face and fingerprint system. The system is tested on ORL, UMIST, VISA face dataset and FVC fingerprint dataset. Experimental results show that the multimodal biometric system using LTP gives better accuracy as compared to the unimodal biometric system.

Keywords—Unimodal; Multi-modal; LBP; LTP; intra-class; inter-class; spoofing attack

I. INTRODUCTION

The face and fingerprints are the most widely used biometrics, due to their high availability and universality. Most of the biometric systems use single biometric traits; it has some challenges like intra-class variation, inter-class variation and noise in the sensed data. This problem can be overcome by using multiple biometric traits for person identification. Multimodal biometric system combines more than one biometric trait by fusing at different levels. Fusion methods are mainly split into two types: fusion prior to matching and after matching. Under fusion prior matching, sensor level and feature level fusions and under after matching score and decision level fusion techniques are used. In sensor level fusion raw information from the source images are combined. Fusion at sensor level is having richer set of information. In fusion at feature level, features from multiple biometric traits are extracted individually, and extracted the characteristics of the various biometrics are combined. In matching score level, scores from each biometrics traits are calculated and scores are combined to get resultant score. In decision level fusion decision obtained by individual biometric traits are combined to get final decision.

Section II of this paper deals with a survey of related work, the proposed methodology is outlined in Section III, Section IV presents the results and discussion and finally the Section V presents the conclusion of the work.

II. RELATED WORK

Pabitra Priyadarshini Jena et al.[1] introduced multimodal biometric system utilizing face and fingerprint. The fingerprint and face features are taken using deep learning models such as ResNet50, Xception, FaceNet, EfficientNetB3 and InceptionResNetV2. The extracted features are fused at feature level to authenticate the user. N. Krishnaraj et al.[2] introduced multimodal biometric system using palmprint and fingerprint. Gabor filters and characteristic subset selection is carried on Palmprint and fingerprint. Emphasized characteristic are selected for Gabor features using Correlation based Feature Selection and Hybrid Bacterial Foraging Optimization. P. Sivakumar et.al.[3] proposed multimodal template for fingerprint and finger-vein using deep hashing framework for feature level fusion. Along with providing improved privacy for biometric information to safeguard from the special attacks, it provides unlinkability and cancelability of the templates. It integrates deep hashing, multimodal fusion and biometric security with weightage on structural facts from modalities like fingerprint and finger-vein. Features are extracted using VGG16 model and achieved 95% accuracy. Rupali Kute et al.[4] introduced a strategy to recognize face of the person by using their fingerprint. To acquire skills and improve the transferring subspace, Bregman divergence regularization is used. Here gained knowledge from the training samples are transferred to the testing samples. The difference among the two dissimilar domains is reduced by this regularization. To find a common subspace that boosts the performance. Two methods introduced by Arucha Rungchokanum et al. [5] applies the introduced distance model to the minutiae-triplets generation. In the earlier method, model is applied direct whereas in second method, to handle more distorted fingerprint areas and more curved regions like a singular point, model is combined with ridge flow. The proposed method is evaluated on sixteen public domain fingerprint database using two minutiae triplet matching algorithm. A component-based face recognition approach introduced by Rupali Sandip Kute et al. [6] uses transfer learning to demonstrate the knowledge acquired from face image to categorize face elements like ears, lips and nose. These face components are unaltered by the change of face expression and pose. Since face and face

components are from different domain, these are used to transmit the knowledge acquired from distinct domains even though they share common information. To associate between complete and partial faces different half faces like, left, lower diagonal and right upper and left, right, upper half and right upper are considered. This proposed approach can be utilized for components-based face recognition, partial face recognition and full face recognition because of the alliance among face and its components. Sun, Kun et al. [8], this paper presents centre-symmetric local binary pattern and DBN method for face recognition. Initially image is separated into blocks, and feature is extracted from each block using CS-LBP, then texture feature histogram is calculated. This histogram is given as input to DBA for classification. As compared to traditional approaches the proposed approach gave high accuracy by comparing results on Extend Yale B, ORL, and CMU-PIE dataset. Sunil S et al.[9], in this paper, a hybrid solution is presented for face recognition by using SWT, LTP, and DCT. Initially, image is resized and features are extracted using SWT and DCT. Then, LTP is claimed on SWT features and features are combined to obtain final feature vector. Euclidean distance is used for classification. Huilin Ge et al [10], present improved GAN for face recognition. The study uses generator consists of auto encoder and discriminators global descriptor and local descriptor to repair occluded image. To perform image restoration Resnet-50 is used. The proposed method provides improved restoration effect and high recognition rate even in occlusion background. Ashok Kumar Yadav et al. [11], proposed a multi-biometric frame based on DNN using fingerprints, human eye iris and offline signature attribute was introduced. As the first step, fingerprints, signatures and iris of subjects are obtained. At this stage, the client's characteristic is perceived with the help of the multi-modality framework, which is composed of three pre-implemented templates for iris, fingerprints and signatures. The multimodal biometric system using iris, signature and fingerprint shows better performance as compared to unimodal biometrics using fingerprint and iris. Additionally, a different biometric has been developed specifically for offline signing. From that point on, the triple single biometrics solutions are used to construct the provided multi-modal solution. Before integrating these models into a multi-model solution, the correctness of these models is examined after taking into account earlier work on unimodal solutions. The VGG-19 net's performance has improved, according to the results. Using the SDUMLA HMT repository, the model was evaluated. We achieved 99.1% accuracy in score stage merging and 98.4% accuracy in feature level fusing. Nassima Kihal et al. [12] have developed a multimodal ocular biometric system to authenticate. They have used iris texture and the cornea form as a biometric characteristic. Iris texture features are extracted using Log Gabor and cornea features using Zernike polynomials. Once the characteristic of the iris and cornea has been extracted, they carry out a matching operation and scores for iris and cornea are calculated. Later applied fusion to the score and obtained EER 0%, FRR 0% to 0.1% FAR. Aman Kathed et al.[13] proposed 3-level authentication system using the biometric features of the face, iris and voice. Initially, the characteristics of the face and iris features are extracted and merged to form final feature vector. The fused features are stored in database. The

subsequent test data is compared to the stored features and if a match is found, the OTP generating system will send OTP to the user. This one-time password is saved as a user's voice and compared to the voice information stored in the database. If the user's voice matches, the user is authorized user.

III. METHODOLOGY

A. Local Binary Pattern

Local Binary Pattern is rotation and grey invariance which is used to extract textual features. It functions on a 3 * 3 window with the central pixel as a threshold [7]. Threshold pixel value is compared to its adjacent 8 pixel values. The pixel value is marked as 1 when the pixel value exceeds the center pixel value, otherwise it is 0. The central pixel value in the 3 * 3 window is equated with its 8 neighbours to produce an 8-bit binary number, which is converted to LBP code. Overall contrast is standardized using Histogram equalization.

The mathematical expression for LBP operation is:

$$LBP_{S,P} = \sum_{y=0}^{S-1} 2^y g(i_y - i_c) \quad (1)$$

$$\text{Where } g(i_y - i_c) = \begin{cases} 1 & i_y - i_c \geq 0 \\ 0 & i_y - i_c < 0 \end{cases} \quad (2)$$

Where S is the number of neighbors, i_y represents neighbor pixel in radius P and i_c represents centre pixel value. After encoding, histogram is calculated using the Eq. (3).

$$H(m) = \sum_{i=0}^I \sum_{j=0}^J f(LBP_{S,P}(i,j), m), m \in [0, M] \quad (3)$$

$$\text{Where } f(x,y) = \begin{cases} 1 & x = y \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

In this case, M represents the maximal LBP pattern value.

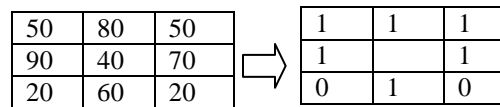


Fig. 1. LBP operation.

Fig. 1 illustrates LBP operation. It obtains the binary pattern= 11110101 and the corresponding LBP Code is: $1x2^0+1x2^1+1x2^2+1x2^3+0x2^4+1x2^5+0x2^6+1x2^7=245$.

B. Local Ternary Pattern

LTP is advanced version of LBP with 3-valued codes; 0, 1, -1 and more robust to noise[14]. LTP will define a threshold t and works as follows; assigns 1 for any pixel with value above t, -1 in case the pixel value is less than -t and 0 if the pixel value is in between -t and t. To get rid of the negative values, upper and lower patterns are constructed after thresholding step. LTP is the concatenation of upper and lower pattern.

LTP is defined mathematically as Eq. (5) and (6):

$$LBP_{S,P} = \sum_{y=0}^{S-1} 2^y h(i_y - i_c) \quad (5)$$

Where

$$h(x) = \begin{cases} 1 & x \geq t \\ 0 & -t < x < t \\ -1 & x < -t \end{cases} \quad (6)$$

Where i_y represents the value of the neighbour pixel of central pixel with radius P , i_c represents central pixel value and S is the number of neighbours. The neighbour is estimated using bilinear interpolation when it is not falling exactly in the centre of the pixel. After this encoding step, using following equation histogram is created using equation (7) and (8).

$$H(m) = \sum_{i=0}^1 \sum_{j=0}^1 f(LBPS, P(i,j), m), m \in [0, M] \quad (7)$$

$$\text{Where } f(x,y) = \begin{cases} 1 & x = y \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

In this case, M represents the maximal LBP pattern value.

Fig. 2 illustrates operation of LTP, here threshold value is 5; central pixel 34 is compared with 8 neighbour pixel values. If neighbour pixel value is greater than $34+5$, then pixel value is defined as 1, if pixel value is less than $34-5$, then set the pixel value is set to -1 and if the pixel value is in between $34+5$ and $34-5$, then set the pixel value is set to 0. The upper and lower pattern are constructed and concatenated to get final LTP code.

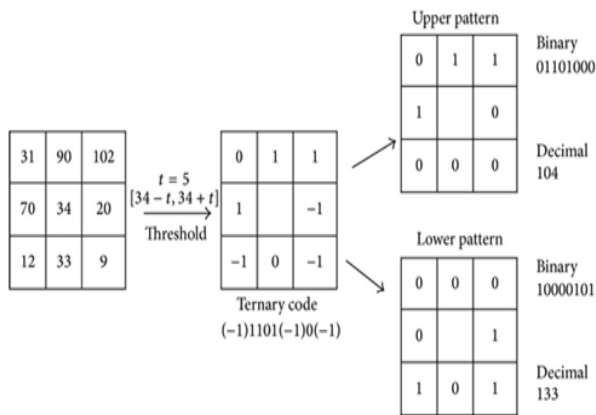


Fig. 2. LTP operation.

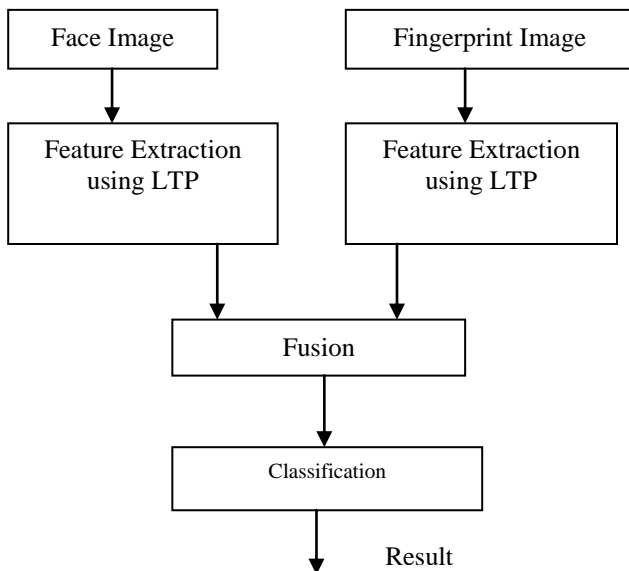


Fig. 3. Architecture of the introduced system.

The Fig. 3 represents proposed method; LTP is a texture descriptor that captures the local texture patterns of an image by comparing the pixel intensities of a central pixel to its neighboring pixels. Initially apply LTP on face image and compute LTP histogram as feature for face. Similarly extract apply LTP on fingerprint and compute LTP histogram as fingerprint feature. Then combine LTP face and fingerprint features to form final feature vector. Then classification is done using Euclidian distance.

IV. RESULTS AND DISCUSSION

To evaluate the efficiency of the system implemented, experiments are carried out on ORL, UMIST, VISA standard face dataset and FVC fingerprint dataset. The Fig. 4 is the sample images of ORL dataset. ORL dataset consists of 10 samples of 40 persons, with different pose. The Fig. 5 shows the sample images from UMIST dataset. UMIST dataset consist of 400 images of 20 persons with 20 samples. The Fig. 6 shows sample images of VISA dataset. There are 500 images of 100 people with five samples each in the VISA dataset. The Fig. 7, 8, 9, 10 shows DB1, DB2, DB3, DB4 of FVC dataset respectively. FVC dataset consists of four databases. Each dataset consists of 80 images of 10 persons with eight samples each with different resolution and orientation.



Fig. 4. Sample images from ORL dataset.



Fig. 5. Image samples from UMIST dataset.



Fig. 6. Image samples from VISA dataset.



Fig. 7. Image samples from FVC DB1 dataset.



Fig. 8. Image samples from FVC DB2 dataset.

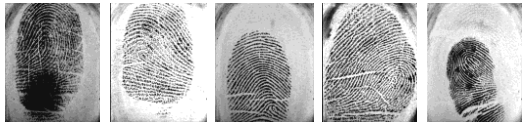


Fig. 9. Image samples from FVC DB3 dataset.



Fig. 10. Image samples from FVC DB4 dataset.

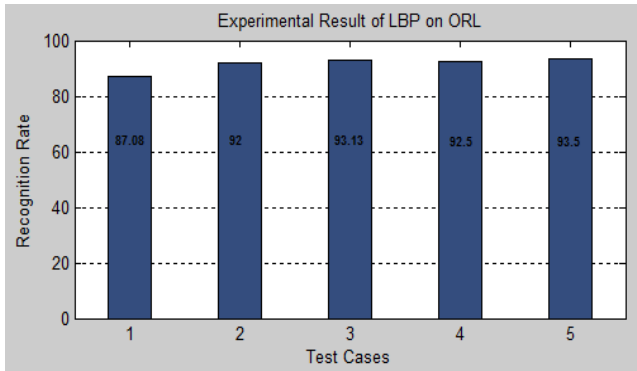


Fig. 11. Recognition rate of LBP on ORL dataset.

The Fig. 11 represents hypothetical outcome of LBP on ORL dataset giving highest accuracy 93.5%. The Fig. 12 shows experimental results of LTP on ORL dataset with highest accuracy 98.13%. Experiment is conducted on different test and training combination. The 1st Experimental setup consists of 1st to 4th indexed samples for training and 5th to 10th samples were used for testing. In the 2nd experimental setup 1st to 5th samples were utilized for training and 6th to 10th samples were considered for testing. Similarly in the 3rd experimental setups, 1st to 6th samples were used for training and 7th to 10th samples are utilized for testing. The fourth experimental setup used odd indexed samples for training and even indexed samples for testing. The last experimental setup was also based on training even indexed samples and testing odd indexed samples. From the above result it is clear that LTP based face recognition gives better accuracy than the LBP based face recognition.

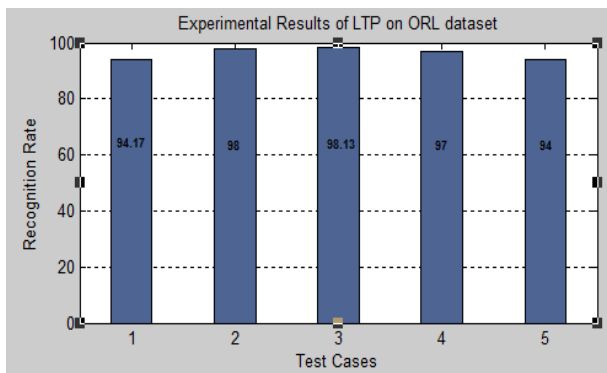


Fig. 12. Recognition rate of LTP on ORL dataset.

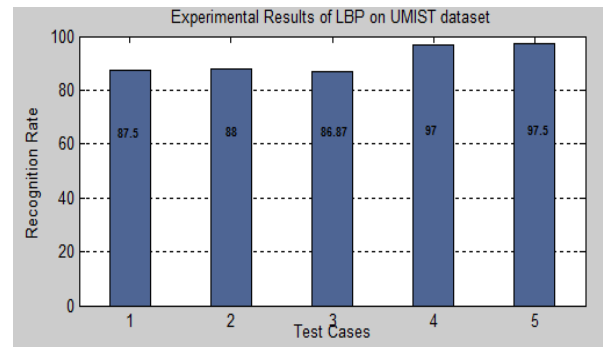


Fig. 13. Recognition rate of LBP on UMIST dataset.

The Fig. 13 represents hypothetical outcomes of LBP on UMIST dataset giving highest accuracy 97.5%. The Fig. 14 shows experimental results of LTP on UMIST dataset with highest accuracy 99%. Experiment is conducted on different test and training combination. The 1st Experimental setup consists of 1st to 8th indexed samples for training and 9th to 20th samples were used for testing. In the 2nd experimental setup 1st to 10th samples were used for training and 11th to 20th samples were considered for testing. Similarly, in the 3rd experimental setups, 1st to 12th samples were utilized for training and 13th to 20th samples are utilized for testing. In the 4th experimental setup odd indexed samples were utilized for training and even indexed samples were utilized for testing. Likewise, in the last experimental setup even indexed samples were used for training and odd indexed samples were used for testing. An experimental result on UMIST dataset shows that LTP based face recognition is best as compared to LBP based face recognition.

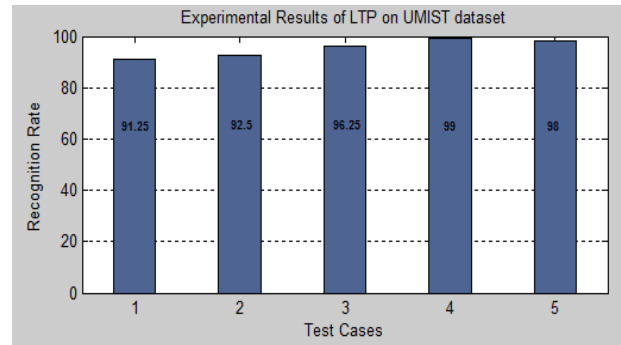


Fig. 14. Recognition rate of LTP on UMIST dataset.

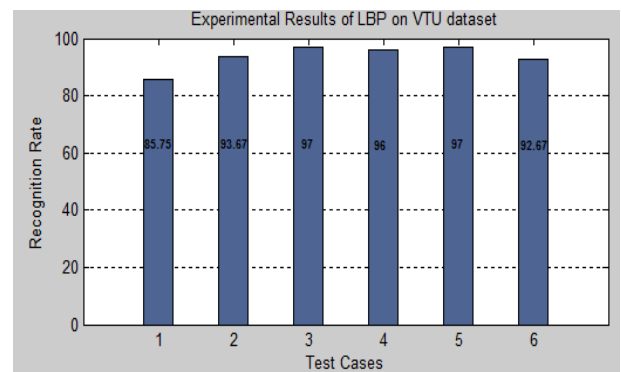


Fig. 15. Recognition rate of LBP on VISA dataset.

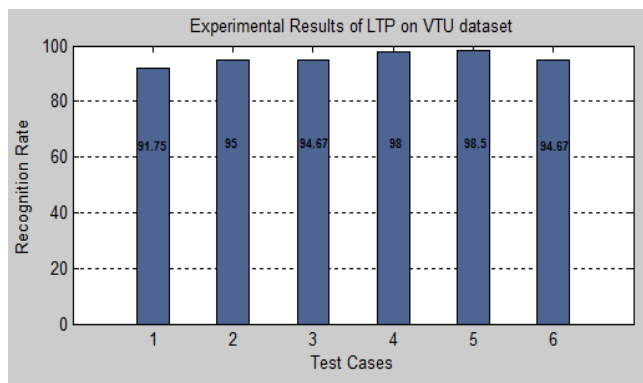


Fig. 16. Recognition rate of LTP on VISA dataset.

Fig. 15 and 16 show experimental results of LBP and LTP on VISA dataset with 97% and 98% accuracy respectively. The experimental setup 1 consists of index 1 sample in training set and 2, 3, 4, 5 indexed samples in testing set. Experimental setup 2 contains index 1 and 2 sample in training set and 3, 4, 5 indexed samples in testing set. Experimental setup 3 contains index 1, 2, 3 in training set and 4 and 5 in testing set. Similarly, experimental setup 4 consists of 1, 2, 3, 4 in training set and 5 in testing set. Experimental setup 5 contains odd indexed samples in training and even indexed sample in testing. Likewise, experimental setup 6 contains even indexed sample in training set and odd indexed samples in testing set. The experimental results show that LTP based face recognition gives better accuracy than LBP based face recognition.

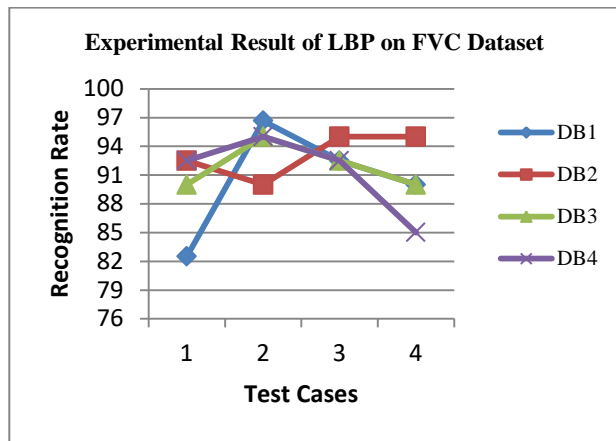


Fig. 17. Recognition rate of LBP on FVC dataset.

Fig. 17 and 18 show experimental results of LBP and LTP on FVC fingerprint dataset with accuracy of 96.6% and 97.5% respectively. FVC dataset contains four different databases DB1, DB2, DB3 and DB4. Experimental setup 1 contains 1st to 4th samples in training and 5th to 8th in testing set. Experimental setup 2 contains 1st to 5th samples in training and 6th to 8th in testing test set. Similarly 4th experimental setup contains odd indexed sample in training and even indexed samples in testing. Last experimental setup contains even indexed samples in training and odd indexed samples in testing set. From the Fig. 17 and 18 we can say that LTP based fingerprint system gives high accuracy as compared to LBP based fingerprint recognition.

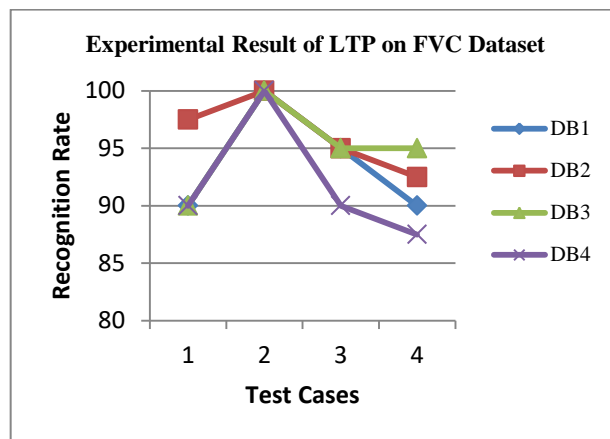


Fig. 18. Recognition rate of LTP on FVC dataset.

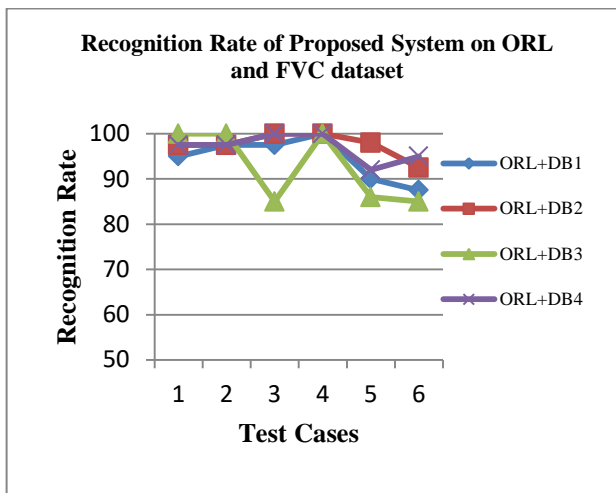


Fig. 19. Recognition rate of proposed system on ORL and FVC dataset.

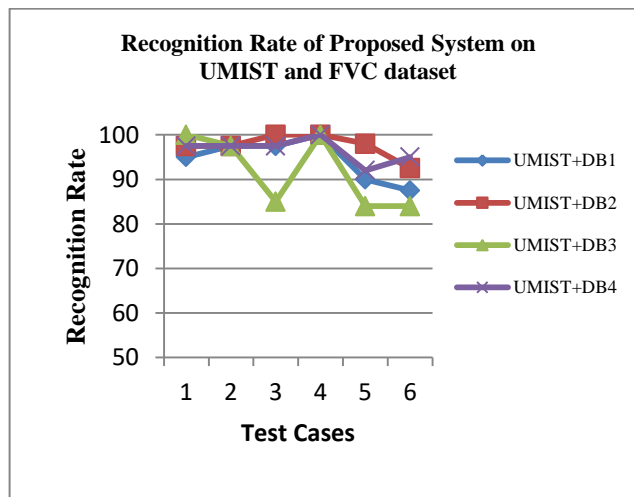


Fig. 20. Recognition rate of proposed system on UMIST and FVC dataset.

The Fig. 19 shows hypothetical outcomes of proposed system on ORL and FVC dataset. The Fig. 20 shows hypothetical outcomes of introduced system on UMIST and FVC dataset. In experimental setup 1, training set consists of even indexed samples and testing set consists of odd indexed

samples. Experimental setup 2 contains odd indexed samples in training set and even indexed samples in testing set. In experimental setup 3, index 1 to 4 samples present in training set and 5 to 8 indexed samples present in testing set. Similarly 4th experiment contains 1 to 5 indexed samples in training set and 6 to 8 indexed samples in testing set. Likewise experimental setup 5 contains 1st 3 samples in training set and 4th to 8th indexed samples in testing set. Finally 6th experimental setup containing 5th to 8th indexed samples in training set and 1st to 4th samples in testing set. As compared to unimodal system the proposed LTP based multimodal system performs better.

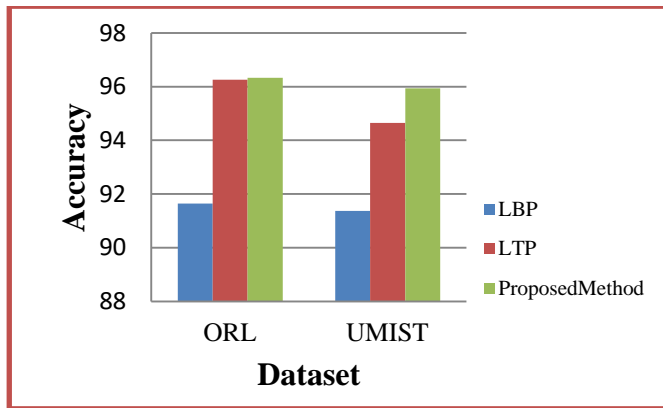


Fig. 21. Comparison of unimodal and multimodal systems.

The Fig. 21 shows a comparison between the unimodal and multimodal biometric system. In the experiments, it was found that the proposed LTP based multimodal biometric system with face and fingerprints is 96.33% accurate on the ORL_FVC dataset and 95.94% accurate on the UMIST_FVC dataset. The unimodal LBP and LTP based biometric system achieved 91.64% and 96.26% accuracy on the ORL dataset respectively and 91.37% and 94.65% accuracy on the UMIST dataset respectively. Therefore the proposed multimodal biometric system thus provides greater accuracy than a unimodal biometrics using LBP and LTP on different datasets.

V. CONCLUSION

A multimodal biometric system offers high security than unimodal biometric system. Face and fingerprint are easily available and universally adopted biometric traits. In this paper, we discussed unimodal face and fingerprint recognition system using LBP and LTP as well as compared it with multimodal biometric system using LTP. The proposed approach is invariant to noise in the input image. Also, by the use of two biometric traits it avoids spoofing attack. The experimental results show that the proposed LTP based multimodal biometric system using face and fingerprint performs better than unimodal biometric system. To improve accuracy, hybrid approaches can be introduced in the future.

REFERENCES

- [1] P. P. Jena, K. N. Kattigenahally, S. Nikitha, S. Sarda and H. Y. "Multimodal Biometric Authentication: Deep Learning Approach," 2021 International Conference on Circuits, Controls and Communications (CCUBE), 2021, pp. 1-5, doi: 10.1109/CCUBE53681.2021.9702724.
- [2] N. Krishnaraj, Y. B. Kalpana and M. Shunmugam, "An Improved Wrapper Based Feature Selection using Hybrid BFO for Multimodal Biometrics," 2021 10th IEEE International Conference on Communication Systems and Network Technologies (CSNT), 2021, pp. 358-362, doi: 10.1109/CSNT51715.2021.9509615.
- [3] P. Sivakumar, B. R. Rathnam, S. Divakar, M. A. Teja and R. R. Prasad, "A Secure and Compact Multimodal Biometric Authentication Scheme using Deep Hashing," 2021 IEEE International Conference on Intelligent Systems, Smart and Green Technologies (ICISSTGT), 2021, pp. 27-31, doi: 10.1109/ICISSTGT52025.2021.00017.
- [4] Rupali Kute, Vibha Vyas, Alwin Anuse, Transfer learning for face recognition using fingerprint biometrics, Journal of King Saud University - Engineering Sciences, 2021, ISSN 1018-3639.
- [5] Arucha Rungchokanun, Vutipong Areekul, Directionally Weighted Distance for Minutiae-Triplets Preservation on Elastic Deformation of Fingerprint Recognition, Pattern Recognition Letters, Volume 160, 2022, Pages 34-42, ISSN 0167-8655.
- [6] Rupali Sandip Kute, Vibha Vyas, Alwin Anuse, Component-based face recognition under transfer learning for forensic applications, Information Sciences, Volume 476, 2019, Pages 176-191, ISSN 0020-0255.
- [7] Ojala, Timo, Matti Pietikainen, and David Harwood. "A comparative study of texture measures with classification based on featured distributions." Pattern recognition 29, no. 1, pp.51-59, 1996.
- [8] Sun, Kun & Yin, Xin & Yang, Mingxin & Yang, Wang & Fan, Jianying. (2018). The face recognition method based on CS-LBP and DBN. Mathematical Problems in Engineering. 2018. 1-9. 10.1155/2018/3620491.
- [9] S Harakannanavar, Sunil & C R, Prashanth & K B, Raja & Patil, Sapna. (2019). Face Recognition based on SWT, DCT and LTP.
- [10] Ge, H.; Dai, Y.; Zhu, Z.; Wang, B. A Robust Face Recognition Algorithm Based on an Improved Generative Confrontation Network. Appl. Sci. 2021, 11,11588. <https://doi.org/10.3390/app112411588>.
- [11] Ashok Kumar Yadav 1 Prof. T. Srinivasulu2 IECE, JNTU, Hyderabad, India 2dean, University College of Engg & Technology, Kakatiya University. India Turkish Journal of Computer and Mathematics Education Vol.12 No.11 (2021), 1627-1638 1627.
- [12] Nassima Kihal, Salim Chitroub, Arnaud Polette, and Jean Meunier, Efficient multimodal ocular biometric system for person authentication based on iris texture and corneal shape, IET Biometrics, Vol. 6 Iss. 6, pp. 379-386, 2017.
- [13] A. Kathed et al., "An Enhanced 3-Tier Multimodal Biometric Authentication," 2019 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India, 2019, pp. 1-6, doi: 10.1109/ICCCI.2019.8822117.
- [14] Tan, Xiaoyang, and Bill Triggs. "Enhanced local texture feature sets for face recognition under difficult lighting conditions." In Analysis and Modeling of Faces and Gestures, pp. 168-182, 2007.

Deep Learning Models for Crime Intention Detection Using Object Detection

Abdirahman Osman Hashi^{1*}, Abdullahi Ahmed Abdirahman^{2*}, Mohamed Abdirahman Elmi^{3*}, Octavio Ernest Romo Rodriguez⁴

Faculty Member, Department of Computing, SIMAD University, Mogadishu Somalia^{1, 2, 3}
Department of Computer Science-Faculty of Informatics, İstanbul Teknik Üniversitesi, İstanbul, Turkey⁴

Abstract—The majority of visual based surveillance applications and security systems heavily rely on object detection, which serves as a critical module. In the context of crime scene analysis, images and videos play an essential role in capturing visual documentation of a particular scene. By detecting objects associated with a specific crime, police officers are able to reconstruct a scene for subsequent analysis. Nevertheless, the task of identifying objects of interest can be highly arduous for law enforcement agencies, mainly because of the massive amount of data that must be processed. Hence, the main objective of this paper is to propose a DL-based model for detecting tracked objects such as handheld firearms and informing the authority about the threat before the incident happens. We have applied VGG-19, ResNet, and GoogleNet as our deep learning models. The experiment result shows that ResNet50 has achieved the highest average accuracy of 0.92% compared to VGG19 and GoogleNet, which have achieved 0.91% and 0.89%, respectively. Also, YOLOv6 has achieved the highest MAP and inference speed compared to the faster R-CNN.

Keywords—Object detection; deep learning; crime scenes; video surveillance; convolutional neural network; YOLOv6

I. INTRODUCTION

Due to the rising crime rate and offensive activities in recent times, it has become common to find CCTV cameras installed in public places such as shopping centers, avenues, banks, and so on. The purpose of these cameras is to enhance security and ensure the safety of the people in these areas. However, detecting weapons in surveillance videos still requires a lot of human intervention, which can result in errors [1]. Meanwhile, it is difficult for humans to constantly observe long videos or maintain multiple footage, and this can result in some rare crime scenes being missed out. As a result, there is a need to explore new technologies that can help improve the accuracy and efficiency of video surveillance in public places [2].

Recent surveillance cameras have the capability to record video only when motion is detected, unlike the older versions that record continuously irrespective of any activity. This feature enhances the efficiency of the system as it reduces processing time, search time, and the storage space required for the recorded videos [3,5]. Unfortunately, law enforcement agencies often follow a reactive approach, which results in delayed response times during crime incidents. In this approach, authorities rely on witness reports or CCTV footage to analyze the crime after it has occurred. This means when an incident takes place, investigators visit the site, manually

retrieve the footage from the camera, and then try to locate the appropriate footage either by watching the entire video or using advanced algorithms to process it which takes a long time. For this case, to improve the efficiency of the security management system and minimize crime incidents and losses, an effective crime prediction analysis system is needed. Such a system would enable proactive crime prevention and ensure robust security management in public places such as banks, shopping malls, and avenues [3,6].

With the availability of large datasets, faster GPUs, enhanced machine-learning algorithms, and better computations, we can now efficiently prepare PCs and construct automated computer-based systems to differentiate and identify various things on a site with high accuracy. For instance, a remote embedded intelligent security monitoring system has been developed using computer vision modeling algorithms to proactively detect intruders. This system utilizes a camera to acquire background images, which are then modeled using the ViBe algorithm to perform object detection in the monitored area [7]. When a moving object, including humans, is identified, the system automatically triggers an alarm and sends a message or call to the user to take preventive measures. To get a better understanding of the situation and detect the intruder, users can log in to the server via a mobile application. The system was implemented on an ARM development board, which provides a platform for hardware and software development. This technology is useful in enhancing security and safety in public places such as shopping malls, airports, and banks, where quick detection of intruders can help prevent criminal activities [3].

In contemporary times, machine-learning and advanced image-processing algorithms have significantly contributed to the evolution of smart surveillance and security systems, as evidenced by recent developments [3,6,7,8]. In addition, the rise of smart devices and networked cameras has also boosted this field. However, detecting and tracking human objects or weapons still require cloud centers as real-time, online tracking is computationally expensive. Recent years have seen significant efforts in monitoring robot manipulators, which require high control performance in terms of reliability and speed [9,10].

Another approach to detecting guns in surveillance films is to use pre-trained deep learning models. These models are intended to assist users in learning about algorithms or experimenting with current frameworks for better outcomes

without explicit design. A deep learning neural network generally has five layers: input and output layers with Convolution, Max-Pooling, and Fully connected layers. Many individuals choose to employ pre-trained deep learning models due to constraints such as limited time, memory, and resources such as CPU and processors [6,35]. When opposed to machine learning, which involves explicit design, these pre-trained models produce better and more accurate outcomes. However, identifying firearms in surveillance films is difficult and subject to human mistake unless it is used a detection system. For instance, human guards may become fatigued or fall asleep when viewing huge volumes of recordings or maintaining several footages, resulting in missed opportunities to discover uncommon criminal intention scenarios that may be caught in many footages. To solve this issue, pre-trained deep learning models may be used to eliminate the need for human interaction while identifying possible threats in public venues [11,34].

For that reason, it is important to design an autonomous surveillance system capable of detecting firearms fast and accurately in order to prevent crime. Deep learning techniques play an essential role here. Hence, we are developing a system which takes advantage of pre-trained deep learning models that are VGG-19 and ResNet50 to detect the firearms with object detection. These models were chosen because recent object recognition models need a large number of parameters and a significant amount of time to train. VGG19 and ResNet50 can extract high-level feature maps from input, reducing its complexity. The Faster RCNN method is also used to build bounding boxes around objects in pictures. The objective is to use neural networks to identify anomalies such as weapons and firearms and determine whether or not the individual carrying them has a criminal intent.

The rest of the paper is organized as follows. The next section will provide some context for our issue and highlight pertinent related works. Section III describes our proposed technique. Section IV describes the experiments and the outcomes. Finally, Section V summarizes the work and offers some perspectives.

II. RELATED WORK

The detection of metallic items that may represent a threat to public and homeland security is a major global concern [1]. Yet, screening for these devices might be difficult since it disrupts people's movements and creates a susceptible target for terrorist strikes [3,9]. In light of this, an automatic metallic item identification and categorization system is proposed. Two related areas are addressed in order to create and implement such a system; the development of a new metallic object detecting system and the establishment of a signal processing method to identify targeted signatures. The suggested approach is assessed by creating a database comprising of actual pistols and everyday items. Extraction of characteristics from four categories is used to examine system outcomes: time-frequency signal analysis, material composition, object form, and transient pulse response. Then, two categorization approaches are used to differentiate between hazardous and non-threatening things. The new system's feature combining and

classification framework achieves a successful classification rate of more than 90% [8,12].

A. Handgun and Knife Detection in CCTV

Another method is to use (CCTV). The use of Closed-Circuit Television (CCTV) systems has become increasingly common in various settings, including offices, residential areas, and public spaces, and has been implemented in many countries. To enhance the effectiveness of these systems, image segmentation techniques are employed to track activities captured by CCTV cameras and apply machine learning algorithms [1,10]. Grega [13] for example, published an algorithm that recognizes knives and guns in CCTV images and warns the security guard or operator. The algorithm's specificity is 94.93% and sensitivity is 81.18% for knife detection, focused on reducing false alarms and delivering a real-time application. Furthermore, the specificity for the fire alarm system is 96.69% and the sensitivity is 35.98% for the various items in the movie. Author [14] developed a video classifier, also known as the Histogram of Directed Tracklets, that recognizes irregular circumstances in complicated sequences. In contrast to typical optical flow techniques that only assess edge characteristics from two consecutive frames, descriptors known as tracklets have been evolving across long-range motion projections. Spatiotemporal cuboid film sequences are statistically gathered on the tracklets that travel across them [15].

B. Automatic Handgun Detection using Machine Learning

Although there has been a recent advance in image-based machine learning, recognizing a knife-wielding assailant remains difficult. To address this issue, the authors [16] describe three approaches for automated threat detection utilizing various knife image datasets, with the goal of narrowing down plausible assault aims while decreasing false negatives and false positives. To begin, they employ a classification model based on Mobile Net in a sparse and pruned neural network that can notify an observer to the presence of a knife-wielding attacker with high accuracy (95%) and a low memory demand (2.2 MB). Second, they train a detection method (Mask RCNN) to segregate the hand from the knife in a single picture and give probable certainty to their relative placement, allowing for both bounding box localization and point threat inference. Finally, a Pose Net-based model assigns anatomical waypoints to narrow down threat features and decrease misconceptions of the attacker's objectives [4,17,18]. Furthermore, the authors identify and fix data gaps, such as the necessity to gather benign hands, which may impair the accuracy of the deployed knife threat detector. This study offers a thorough review of image-based warnings that may be used to prioritize and educate crime prevention strategies before any catastrophic results occur. Additional relevant study topics in this subject include, among others, Automated Handgun Detection Alarms in Videos Using Deep Learning, Automatic Visual Recognition of Armed Robbery, and Robust Item Detector Application for Visual Knife Detection [19,35].

Other authors proposed by analyzing the recorded films of the cameras. The author [8] describes a technology for automatically detecting handguns in surveillance films. By

recognizing weapons in films, categorizing items as either a gun or not, and forecasting whether a crime has happened, the system hopes to control occurrences of crime. The system's performance is compared to a sliding window proposal technique, and it is discovered that FRCNN and RCNN-based models trained on a dataset perform well. The algorithm can reliably anticipate crime occurrences even in low-quality films, yielding good results [5,20,31].

The author [21] details a conventional approach for identifying the position of an armed robber. The method focuses on detecting individuals who are holding a knife in various positions relative to other people. To accomplish this, the system uses skeleton silhouette algorithms that segment the body into distinct parts and identify the position of a raised arm holding a knife at different angles. Through this process, the system is able to successfully detect the presence of a knife.

The author [22] describes a visual method for detecting automated weapons, specifically knives held in hands. This approach utilizes novel object detection algorithms to identify visual knives within a given video dataset. One of the primary challenges is detecting knife rotations at varying scales and positions, which can be difficult due to the multitude of possible orientations and positions of the knife in the dataset. To address this, the system is designed to detect all possible knife orientations and positions. Feature extraction is accomplished using foreground segmentation and FAST (Features from Accelerated Segment Test) for feature detection. Classification is then performed using MRA (Multi-Resolution Analysis).

C. Automatic Hanggun Detection using Machine Learning

Meanwhile, Convolutional Neural Networks (CNNs) have shown exceptional performance in image processing and object recognition during the last few years. CNNs are a sort of neural network that is specifically intended to recognize pictures [23,24]. These networks are made up of numerous layers, each having its own function. The first (input) layer receives an image as input. The next layer (the convolution layer) applies a collection of filters to the input picture, which are themselves tiny images. This layer takes characteristics from the input picture and extracts them. The following layer (the pooling layer) decreases the previous layer's output by pooling together all the pixels in a fixed-size square of the input picture. This layer reduces the number of parameters and makes the network more error-resistant [16,25]. CNN is trained on a large dataset of pictures containing the items to be recognized in order for the network to learn the characteristics that identify each object and correlate them with a given class. After trained, the network may be used to recognize the required items in new photos [26]. CNNs have demonstrated considerable potential for a variety of applications, including self-driving vehicles, face identification, crime detection, internet of things-based photovoltaics monitoring, and even COVID-19 detection, because to their capacity to identify a wide spectrum of objects and their error tolerance. The upcoming Fig. 1 elaborates a CNN model that has applied for a weapon detection obtained [9].

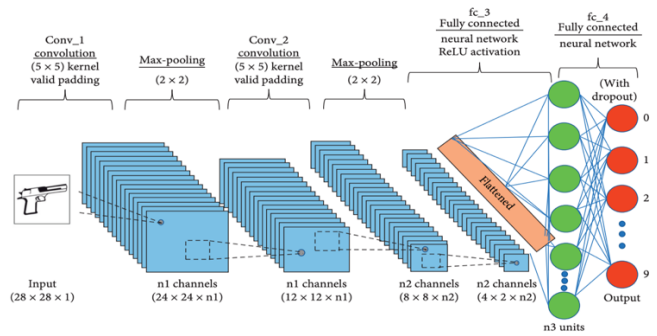


Fig. 1. Feedforward Convolutional Neural Network (CNN).

The cutting-edge YOLO V3 object identification model was applied and trained for obtained dataset for weapon detection in this work. Authors suggest a concept that gives a machine or robot a visionary sense to recognize dangerous weapons and can also inform a human administrator when a gun or a firearm is seen in the edge. +e experimental data demonstrate that the trained YOLO V3 model outperforms the YOLO V2 model and is less computationally costly. There is an immediate need to improve the present surveillance capabilities by providing better resources to enable monitoring the efficacy of human operators [27,32,33].

The phrase "backbones"[29] in the realm of object detection refers to the parameters or weights used to construct feature maps. These backbones are an essential component of the feature extractor since they generate the features that will be utilized for object detection [28]. There are several backbones that may be used for this purpose, each with its own set of benefits and downsides. The visual geometry group (VGG), (ResNet) residual neural network are among the most often used deep learning convolutional neural network (CNN) backbones in object detection techniques [30]. In terms of speed, efficiency, and accuracy, each of these designs offers various trade-offs. However, the proposed model will be applied VGG, ResNet and GoogleNet. The algorithms which we propose are able to detect the human operator when a gun is visible in the image.

III. PROPOSED MODEL

The aim of this study is to address the challenge of identifying indications of automated criminal intention and identifying hazardous circumstances using closed-circuit television (CCTV) systems. The primary objective of this research is to expedite the identification of weapons with improved precision and decreased false positives when compared to machine learning techniques, while also ensuring that convolutional neural networks (CNNs) maintain performance efficiency with fewer training samples. Pre-trained models such as GoogleNet and VGGNet-19 have been trained using millions of photographs, and possess the capability to recognize objects in new images with minimal errors. Owing to their superior training accuracy, we have opted to utilize the VGGNet19, GoogleNet, and ResNet50 models to effectively categorize and recognize objects.

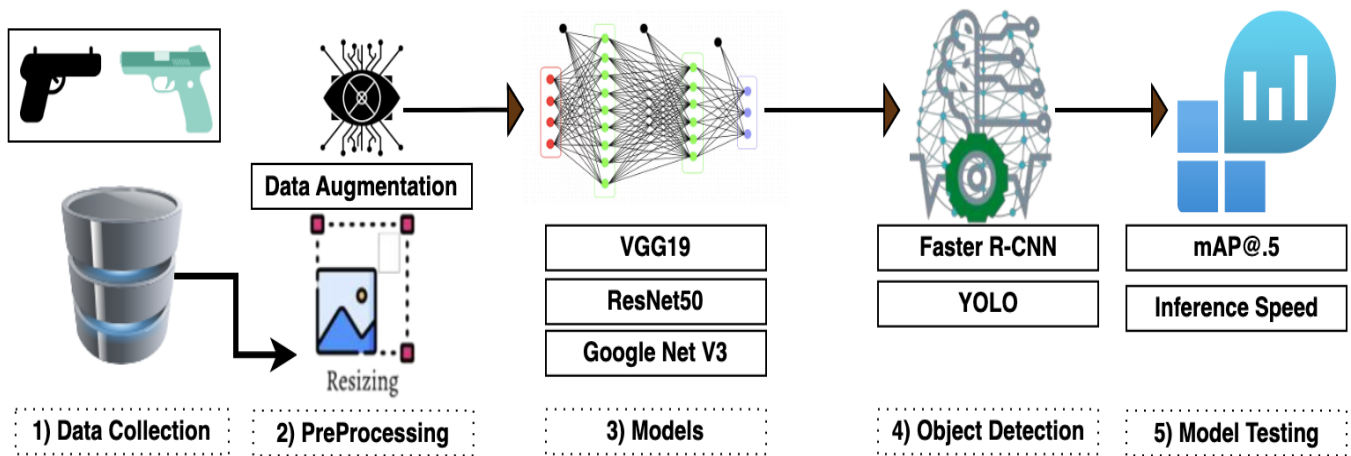


Fig. 2. Proposed model.

The diagram presented in Fig. 2 illustrates the all-encompassing design of the system under consideration. At the initial stage, input frames are received through the input layer, which is also responsible for conducting pre-processing activities as data augmentation. After undergoing pre-processing, the images are transferred through various layers, such as Convolution, Max-pooling, and FC layers, which perform a range of operations including feature extraction, feature filtering, feature mapping, and classification. And, the object detection layer is responsible for detecting objects and classifying, and in the event of any detected criminal intentions, it utilizes a registered API to dispatch a security message after model testing. The explanation of each step will be explained in a detailed way.

The first step involves gathering a comprehensive collection of positive firearm images that depict guns of various sizes, angles, and colors. These images are then segregated into designated "weapons" folders. Additionally, negative images resembling guns are also compiled and stored in separate folders labeled "not weapons". The open pictures dataset V6, a widely utilized dataset containing over nine million photographs, is used to source the data. This dataset features several images of firearms and bladed weapons that can be used to train machine learning models to effectively identify criminal activity in images. For this purpose, three categories were extracted from the dataset, namely rifle (2072 images), handgun (607 photos), and shotgun (476 images), although some researchers have extracted six categories, we only extracted three categories. Subsequently, the three groups were forwarded to the pre-processing data level.

The second stage involved data pre-processing, which refers to the preparatory procedures that must be performed on images prior to analysis. This process may involve a range of activities, such as scaling or adjusting the image display. One of the critical tasks in data preparation is resizing the image to 256 x 256 pixels. This entails feeding input frames into the Input layer, which performs pre-processing operations on images of varying sizes (such as 256 x 256 width and height) and converts them to a standardized size of 224 x 224 x 3 (RGB values) by extracting RGB values from pixels. This guarantees that all data has a uniform size and can be easily

compared. Furthermore, it simplifies working with images that are not big which can be advantageous for training computers to detect objects. Meanwhile, another crucial aspect of data preparation is data augmentation, which can be achieved through various methods such as flipping, rotating, or scaling the image. This technique aims to increase the quantity of data available for training and enhance the system's ability to recognize objects from various perspectives. Effective data preparation is a crucial stage in object detection, as it can determine the success or failure of the system. Therefore, by resizing data and augmenting it, we can increase the likelihood of achieving desirable outcomes.

In the third stage, it involved fitting three object detection models, namely VGG 19, ResNet, and GoogleNet. Each of these models has its own set of advantages and drawbacks, making it essential to conduct a comparative analysis to determine the most suitable model for our specific data. For our dataset, we employed transfer-learning to fine-tune the models. Ultimately, it should be noted that DarkNet is undoubtedly the foundational model for YOLO.

In the fourth step, the object detection model was trained using the most optimal backbone model on both the train and validation sets, comprising 70% and 20% respectively. This facilitated the development of a highly accurate and dependable model, capable of recognizing a diverse range of objects. The model provided a robust foundation on which to build the model, ensuring its precision and reliability. The R-CNN and YOLO v6 object detection models were utilized in this study. Faster-RCNN was preferred over R-CNN and Fast R-CNN due to its superior accuracy and efficiency in object detection, enabling it to process complex images and capture intricate details. The latest version of YOLO, YOLOv6, was found to be the most advanced and user-friendly option, offering higher speed and accuracy compared to previous versions. Labeled data, specifically bounding boxes, were employed to train the object detection algorithms. These bounding boxes were used to define the location of objects in the images, while the labels provided information about the type of objects.

The final stage involved model inference, wherein the object detection models were tested on a separate testing set

comprising 10% of the data. This step was crucial in evaluating the accuracy and inference time of each model, as has been done in prior research. Testing the models on previously unseen data allowed us to determine their efficacy in practical scenarios.

IV. RESULTS AND DISCUSSIONS

In this section, we will explain the dataset description and elaborate different categories that we extracted from the dataset. Moreover, we will discuss and compare the output result of the three models in term of the accuracy, recall and F1-score. Also, we will present the detection output and finally discuss the comparative analyse as a benchmark.

A. Datasets Description

Open Images dataset is a collection of nine million images that have been annotated with image-level names, object bounding boxes, object segmentation masks, visual connections, and localized narrative. It has the biggest existing dataset with object position annotations, with 16M bounding boxes for 600 item types on 1.9M photos. Professional annotators drew the boxes mostly by hand to guarantee accuracy and uniformity. The images are quite varied and frequently feature complicated scenarios with several items. For our models, we had extracted only three categories (as seen in Table I).

B. Identify Comparing VGG19, GoogleNet and ResNet50

As we mentioned in the methodology, we have applied three distinct algorithms and the upcoming tables will illustrate their performance. The applied algorithms were VGG19, ResNet50 and GoogleNet and their accuracy, loss, precision, and recall will be discussing in the upcoming tables. Notably, ResNet50 algorithm performed remarkably well in the classification task compared to the VGG19 and GoogleNet, achieving accuracy scores of 0.92%, 0.91% and 0.89%, respectively.

Although VGG 19 demonstrated marginally superior F1-score compared to GoogleNet, it was also associated with a lower support score. ResNet50, on the other hand, surpassed both VGG 19 and GoogleNet algorithms, yielding an accuracy score of 0.92%, along with higher F1-score (0.94%) and recall (0.91%) scores, and comparatively higher support (45%). GoogleNet exhibited the poorest performance amongst all three algorithms, with the lowest accuracy score of 0.89% and the weakest performance across all other metrics.

The upcoming Table II shows the training accuracy of the VGG19. It can be seen that the overall average of the accuracy is 0.91%, yet it has also scored a good performance in term of F1-score by achieving 0.93%.

The next table which is Table III shows the training accuracy of the ResNet50. It can be seen that the overall average of the accuracy is 0.92%, and it outperformed compared to other algorithms which we have also applied the same with this dataset. In term of F1-score, it has also outperformed other algorithms by achieving 0.94% in the average total.

The next table which is Table IV demonstrates the training accuracy of the GoogleNet. It can also see that the overall average of the accuracy is 0.89%, this makes the poorest performance that has been achieved compared to other algorithms. In terms of F1-score, it has achieved a good accuracy but yet it is still lower than VGG19 which also makes the lowest performance.

On the other hand, the upcoming Table V demonstrates the mAP and inference-speed results of two object identification methods, namely Faster CNN and YOLOv6, which are known for their fast-processing times. The mAP is a crucial metric for evaluating object detection models, representing the (AP) of each object class. The ratio of (TP) to the sum of TP and (FP) is computed to determine the AP (FP). The mAP is a measure of an object detector's ability to accurately identify and differentiate different types of objects in an image. The symbol @0.5 implies that an intersection over union (IoU) threshold of 0.5 was used. IoU is a measure of how well a predicted bounding box or mask aligns with the ground truth data. Inference speed, measured in frames per second (FPS), indicates how many frames the algorithm can process per second.

TABLE I. DATASET DESCRIPTION

	Extracted categories	
0	Handgun	607 Images
1	Rifle	2072 Images
2	Shotgun	467 Images

TABLE II. TRAINING ACCURACY OF VGG19

Accuracy 100%	Training Accuracy of VGG19			
	Accuracy	Recall	F1-score	Support
0	0.90	0.88	0.95	43
1	0.91	0.92	0.93	112
2	0.93	0.91	0.93	163
Avg/Total	0.91	0.90	0.93	106

TABLE III. TRAINING ACCURACY OF RESNET50

Accuracy 100%	Training Accuracy of ResNet50			
	Accuracy	Recall	F1-score	Support
0	0.92	0.91	0.96	45
1	0.91	0.90	0.94	108
2	0.94	0.93	0.92	170
Avg/Total	0.92	0.91	0.94	107

TABLE IV. TRAINING ACCURACY OF GOOGLNET

Accuracy 100%	Training Accuracy of GoogleNet			
	Accuracy	Recall	F1-score	Support
0	0.88	0.90	0.92	26
1	0.90	0.89	0.91	94
2	0.89	0.91	0.90	160
Avg/Total	0.89	0.90	0.91	93

TABLE V. RESULT OF MODEL DETECTION

Algorithm	MAP	Inference-speed
Faster R-CNN	61.82	8
YOLOv6	63.12	68

The two algorithms exhibit different mAP scores and inference-speeds. (R-CNN), which is faster than the other method, achieved the highest score of 63.12%. However, it has the slowest inference speed of 8 frames per second (FPS). Conversely, YOLOv6 obtained the highest mAP@.5 score of 63.12%, but it has the slowest inference speed of 68 FPS.

C. Output of Gun Detection Result

The upcoming Fig. 3 demonstrates the output of the models for detecting the gun and classifying based on the object detected. Though we have not applied to send notification through API but it is important to make a separate comparing by which gun is detected.



Fig. 3. (a) Rifle detection. (b) Handgun detection.

D. Comparative Analysis

Previous studies in this field have applied a range of models to detect firearms, as mentioned earlier. For instance, Author [1] developed a system utilizing VGG-19 to identify a weapon in a person's hand pointing at someone else, while Faster RCNN was used to draw bounding boxes around objects in images, resulting in an accuracy of over 80%. In comparison, our VGG19 model achieved an accuracy of 0.91%. It's worth mentioning that other researchers [4] used YOLOv5 and achieved a MAP of 56.92 and an inference speed of 61, but our model, utilizing YOLOv6, achieved a higher MAP of 63.12 and a faster inference speed of 68.

V. CONCLUSION

Deep learning object detection systems have the capability to offer significant benefits to officers and security professionals, as they can help in efficient way and not consuming much resources for forensic tasks. However, deploying artificial intelligence (AI) in this field raises concerns about possible misuse by law enforcement organizations, such as accusing innocent people or detecting bogus offenders. Our presented algorithms are capable of alerting human operators when a gun is detected in an image. In this study, we developed and assessed a system using a dataset of photos and videos obtained from the open images dataset V6, which contains over nine million images. Our results demonstrate that an early warning system in risky situations could lead to quicker response times, more efficient

reaction times, and fewer potential casualties. Additionally, the proposed method outperforms other known approaches to crime detection. For further improvement, it would be novelty if it has incorporated other modalities such as audio or text data and this could improve the accuracy of the models.

REFERENCES

- [1] Divya, S. M., Priya, G. S., Abitha, R., Sirisha, K., Manikanta, A., & Jayanth, K. AUTOMATED CRIME INTENTION DETECTION USING DEEP LEARNING.
- [2] Sultana, T., & Wahid, K. A. (2019). IoT-guard: Event-driven fog-based video surveillance system for real-time security management. IEEE Access, 7, 134881-134894.
- [3] Naval Gund, U. V., & Priyadarshini, K. (2018, December). Crime intention detection system using deep learning. In 2018 International Conference on Circuits and Systems in Digital Enterprise Technology (ICCSDET) (pp. 1-6). IEEE.
- [4] Boukabous, M., & Azizi, M. (2023). Image and video-based crime prediction using object detection and deep learning. Bulletin of Electrical Engineering and Informatics, 12(3), 1630-1638.
- [5] Kaushik, H., Kumar, T., & Bhalla, K. (2022). iSecureHome: A deep fusion framework for surveillance of smart homes using real-time emotion recognition. Applied Soft Computing, 122, 108788.
- [6] Mathur, R., Chintala, T., & Rajeswari, D. (2022, January). Detecting criminal activities and promoting safety using deep learning. In 2022 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI) (pp. 1-8). IEEE.
- [7] Ahmed, S., Bhatti, M. T., Khan, M. G., Lövdström, B., & Shahid, M. (2022). Development and optimization of deep learning models for weapon detection in surveillance videos. Applied Sciences, 12(12), 5772.
- [8] Venkatesh, S. V., Anand, A. P., Gokul Sahar, S., Ramakrishnan, A., & Vijayaraghavan, V. (2020). Real-time Surveillance based Crime Detection for Edge Devices. In VISIGRAPP (4: VISAPP) (pp. 801-809).
- [9] Narejo, S., Pandey, B., Esenarro Vargas, D., Rodriguez, C., & Anjum, M. R. (2021). Weapon detection using YOLO V3 for smart surveillance system. Mathematical Problems in Engineering, 2021, 1-9.
- [10] Qin, Z., Liu, H., Song, B., Alazab, M., & Kumar, P. M. (2021). Detecting and preventing criminal activities in shopping malls using massive video surveillance based on deep learning models. Annals of Operations Research, 1-18.
- [11] Shah, N., Bhagat, N., & Shah, M. (2021). Crime forecasting: a machine learning and computer vision approach to crime prediction and prevention. Visual Computing for Industry, Biomedicine, and Art, 4, 1-14.
- [12] Kaya, V., Tuncer, S., & Baran, A. (2021). Detection and classification of different weapon types using deep learning. Applied Sciences, 11(16), 7535.
- [13] Arunnehr, J. (2021). Deep learning-based real-world object detection and improved anomaly detection for surveillance videos. Materials Today: Proceedings.
- [14] Grega, M., Lach, S., & Sieradzki, R. (2013, February). Automated recognition of firearms in surveillance video. In 2013 IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA) (pp. 45-50). IEEE.
- [15] Sung, C. S., & Park, J. Y. (2021). Design of an intelligent video surveillance system for crime prevention: applying deep learning technology. Multimedia Tools and Applications, 1-13.
- [16] Mehta, P., Kumar, A., & Bhattacharjee, S. (2020, July). Fire and gun violence based anomaly detection system using deep neural networks. In 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC) (pp. 199-204). IEEE.
- [17] Hussain, S. A., & Al Balushi, A. S. A. (2020). A real time face emotion classification and recognition using deep learning model. In Journal of physics: Conference series (Vol. 1432, No. 1, p. 012087). IOP Publishing.

- [18] Amrutha, C. V., Jyotsna, C., & Amudha, J. (2020, March). Deep learning approach for suspicious activity detection from surveillance video. In 2020 2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA) (pp. 335-339). IEEE.
- [19] Kaliappan, J., Shreyansh, J., & Singamsetti, M. S. (2019, March). Surveillance Camera using Face Recognition for automatic Attendance feeder and Energy conservation in classroom. In 2019 International Conference on Vision Towards Emerging Trends in Communication and Networking (ViTECoN) (pp. 1-5). IEEE.
- [20] Shirsat, S., Naik, A., Tamse, D., Yadav, J., Shetgaonkar, P., & Aswale, S. (2019, March). Proposed system for criminal detection and recognition on CCTV data using cloud and machine learning. In 2019 International Conference on Vision Towards Emerging Trends in Communication and Networking (ViTECoN) (pp. 1-6). IEEE.
- [21] M. Grega, A. Matiolanski, P. Guzik, and M. Leszczuk, "Automated detection of firearms and knives in a CCTV image," *Sensors*, vol. 16, no. 1, p. 47, 2016.
- [22] Verma, G. K., & Dhillon, A. (2017, November). A handheld gun detection using faster r-cnn deep learning. In Proceedings of the 7th international conference on computer and communication technology (pp. 84-88).
- [23] Nagayama, I., Miyahara, A., & Shimabukuro, K. (2019). A study on intelligent security camera system based on sequential motion recognition by using deep learning. *Electronics and Communications in Japan*, 102(11), 25-32.
- [24] H. Mousavi, S. Mohammadi, A. Perina, R. Chellali, and V. Murino, "Analyzing tracklets for the detection of abnormal crowd behavior," in Proceedings of the 2015 IEEE Winter Conference on Applications of Computer Vision, pp. 148-155, IEEE, Waikoloa, HI, USA, January 2015.
- [25] Damashek, A., & Doherty, J. (2015). Detecting guns using parametric edge matching. Tech. Rep.
- [26] Glowacz, A., Kmiec, M., & Dziech, A. (2015). Visual detection of knives in security applications using active appearance models. *Multimedia Tools and Applications*, 74, 4253-4267.
- [27] Alqubaa, A., & Tian, G. Y. (2012). Weapon detection and classification based on time-frequency analysis of electromagnetic transient images. *International Journal of Advances in Systems and Measurements*, 3(3).
- [28] Bhatti, M. T., Khan, M. G., Aslam, M., & Fiaz, M. J. (2021). Weapon detection in real-time cctv videos using deep learning. *IEEE Access*, 9, 34366-34382.
- [29] Tiwari, R. K., & Verma, G. K. (2015, January). A computer vision based framework for visual gun detection using SURF. In 2015 International Conference on Electrical, Electronics, Signals, Communication and Optimization (EESCO) (pp. 1-5). IEEE.
- [30] Alajrami, E., Tabash, H., Singer, Y., & El Astal, M. T. (2019, October). On using AI-based human identification in improving surveillance system efficiency. In 2019 International Conference on Promising Electronic Technologies (ICPET) (pp. 91-95). IEEE.
- [31] Landa, J., Jun, C., & Jun, M. (2017, January). Implementation of a Remote Real-Time Surveillance Security System for Intruder Detection. In 2017 9th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA) (pp. 102-105). IEEE.
- [32] Sivakumar, P. (2021, July). Real Time Crime Detection Using Deep Learning Algorithm. In 2021 International Conference on System, Computation, Automation and Networking (ICSCAN) (pp. 1-5). IEEE.
- [33] Patel, K., & Patel, M. (2021, July). Smart surveillance system using deep learning and RaspberryPi. In 2021 8th International Conference on Smart Computing and Communications (ICSCC) (pp. 246-251). IEEE.
- [34] Arthi, R., Ahuja, J., Kumar, S., Thakur, P., & Sharma, T. (2021). Small object detection from video and classification using deep learning. In *Advances in Systems, Control and Automations: Select Proceedings of ETAEERE 2020* (pp. 101-107). Springer Singapore.
- [35] ain, H., Vikram, A., Kashyap, A., & Jain, A. (2020, July). Weapon detection using artificial intelligence and deep learning for security applications. In 2020 International conference on electronics and sustainable communication systems (ICESC) (pp. 193-198). IEEE.

Hand Gesture Recognition Based on Various Deep Learning YOLO Models

Soukaina Chraa Mesbahi, Mohamed Adnane Mahraz, Jamal Riffi, Hamid Tairi

Laboratory of Computer Science, Signals, Automation and Cognitivism (LISAC)-Department of Computer Science-Faculty of Sciences Dhar El Mahraz, Sidi Mohamed Ben Abdellah University, Fez, Morocco

Abstract—Some varieties of sign languages are used by deaf or hard-of-hearing people worldwide to interact with others more effectively, consequently sign language's automatic translation is expressive and important. Significant improvements in computer vision have been made recently, notably in tasks based on object detection using deep learning. By locating things in visual photos or videos, the genuine cutting-edge one-step object detection approach greatly provides exceptional detection accuracy. With the help of messaging or video calling, this study suggests a technique to get beyond these obstacles and enhance communication for such persons, regardless of their disability. To recognize motions and classes, we provide an enhanced model based on Yolo (You Look Only Once) V3, V4, V4-tiny, and V5. The dataset is clustered using the suggested algorithm, requiring only manual annotation of a reduced number of classes and analysis for patterns that aid in target prediction. The suggested method outperforms the current object detection approaches based on the YOLO model, according to experimental results.

Keywords—Neural network; deep learning; YOLO; object detection; hand gesture

I. INTRODUCTION

In human-computer interaction (HCI), the hand represents a necessary role as a medium of interaction [1]. Verbal and non-verbal communication can be roughly categorized as forms of communication. A gesture is a type of nonverbal communication in which the movement of the hand, face, or other body parts conveys a specific message [2]. The hand is the most aspect of body language used to make gestures for creating communication. There are two types of hand gestures: static and dynamic. The postures in which stable shapes of the hands are represented by static hand gestures and the dynamic hand gestures contain a series of image sequences. For many computer vision applications [3], including hand action analysis, human-computer interaction, sign language identification, virtual reality, and driver hand behavior monitoring, hand gesture recognition and detection in cluttered situations is a key task. The main goal [4] of this research is to identify the challenge of detecting and recognizing static hand gestures when the hand adopts positions to convey particular meanings. Due to the numerous hand configurations and angles with respect to the image sensor, this problem exhibits a high level of complexity, making it challenging to recover the hand shape. Otherwise, in many applications, such as driver hand monitoring and hand gesture commands to prevent driver distraction, sign language recognition for deaf and speech-impaired people, and many more applications, identifying

static hand motions plays a vital role. The gesture of a hand and the location of its fingertips are necessary information for a computer to comprehend the state of the interaction medium. Recognizing hand gestures is evenly essential to interpret sign language [5].

Hearing-impaired try to find basic necessities similar to normal human beings like learning, writing, teaching, communicating, reading which may not be easy for them. There are several forms of communication in the world in which people communicate with each other. One of these means of communication is sign language. Sign Language is a natural conversation that frequently hearing-impaired people utilize for communication. Occasionally, to let hearing-impaired people communicate easily with normal people, this Sign Language has to be maintained by technology to identify the sign language [6].

The recognition of sign language represents the technology that makes the computer to recognize the sign utilized with the signer and reform it to text with the help of some algorithms.

Speaking and writing are not used to communicate spoken languages; instead, facial expressions, body language, and hand gestures are used. Due to the peculiarity that these languages are expressed by visual rather than aural means, this clearly sets them apart from other languages and creates a unique language barrier because sign languages have quite different ways of expressing and interpreting ideas. Furthermore, spoken languages are more common than sign languages. Research on the translation of sign language is not as advanced as that of spoken language due to the dearth of texts that permit research access to language. For instance, there isn't a ready-to-use digital sign language translation tool that can translate between spoken and sign languages as well as vice versa.

Language can be also the construction of mimic, gesture, the finger-spelling, and hand sign, in addition to the hand position. By using their bodies, particularly their hands, fingers, and arms, hearing-impaired people can interact nonverbally by using sign language. In the field of computer vision, it is crucial to be able to recognize patterns in movies or images. An essential task in the field of computer vision [7] is to identify indications in movies or images. Constantly understanding what signers are seeking to communicate or describe needs recognizing the numerous hand gestures they use. Arabic Sign Language, American Sign Language, Indian Sign Language, Indonesian Sign Language, and others all have diverse sign language structures [8]. One of the main areas of

study in computer vision and machine learning is object detection. Lately, object detection becomes the key to solving real-world problems in applications such as face detection, object Tracking, video surveillance, autonomous vehicles, face detection, pedestrian detection, etc. [9]. Object detection presents an important task of detecting a custom object in images or video, etc. These images or videos contain many objects or a few objects at multiple positions. This task is accomplished by providing the list of different objects that are in the image, providing the object's coordinates and information about the object's location in the image. Supplementary information comprises a bounding box that designates the location as well as the probability with which the object was detected. Supplementary information comprises a bounding box that designates the location as well as the probability with which the object was detected.

The practice of classifying data so that a model can make decisions and take action is known as data annotation. For a range of applications [10], including those that rely on machine learning to analyze images and robotic vision, computer vision, facial and hand identification, image annotation is crucial. To train these solutions, metadata must be attributed to the images in the form of captions, identifiers, or keywords. Image annotation increases precision and accuracy by adequately training these systems. Convolutional neural networks with regional learning are currently popular in detecting work. For object localization, RCNN, Fast RCNN, and Faster RCNN were developed [11]. Recently, the idea of You Only Look Once (YOLO) was used for localizing an area of interest. This work on hand gesture identification focuses mostly on classifying and identifying the gestures. As a technique, hand recognition uses a number of algorithms and ideas from other disciplines, like neural networks and image processing, to discover the movement of a hand. There are a number of object detection methods that help to identify the class and gesture that each algorithm is targeting. This study compares various algorithms and determines which one provides faster, more accurate results than the others. You Only Look Once (YOLO) v3, YOLO v4, Yolov4-Tiny Darknet, and YOLO v5 algorithms were used to analyze the structure and mechanism deduction of hand gesture recognition in order to realize this detection.

For several kinds of hand motions, we suggest a new dataset in this work. Our dataset includes everyday activities, people from various backgrounds and nations, as well as various lighting conditions. For 50,000 photos in our dataset, bounding box annotations are present.

The following are this paper's primary points:

- This work on hand gesture recognition primarily aims to categorize and identify the gestures.
- This paper examines several algorithms, You Only Look Once (YOLO) v3, You Only Look Once (YOLO) v4-Tiny darknet, and You Only Look Once (YOLO) v5 to evaluate the structure and mechanism deduction of hand gesture recognition.

- Our main objective is to describe the datasets, evaluation measures, and experimental setup that we employed for our evaluation.

The remainder of this paper is organized as follows. Brief reviews of several related research on hand detection and gesture identification are included in Section II. The proposed system is fully described in Section III along with an explanation of each component's purpose along with the dataset used. In Section IV the evaluation metrics, experimental setup, and comparison of the obtained experimental results are discussed. Section V discusses the results obtained. Finally, Section VI presents the conclusion.

We also provide a bounding box labeled dataset for object detection methods with this dataset, which contains over 40.000 carefully labeled photos.

II. RELATED WORK

Hand gestures can be recognized in a variety of data sources, including video and photographs, wearable sensors, etc. There are several types of research works on hand gesture recognition, The earliest technique for hand gesture recognition makes use of hand gloves with cables, sensors, LED markers, or other devices [12]. These techniques only provide accurate results when illumination conditions are stable, but classifying hands is a highly challenging problem. Many characteristics, including skin tone and velocity, have been suggested for the detection of hand motions [13], articulated models, hand crafted spatio-temporal descriptors, and trajectory based information. Convolutional neural networks' present success is inspired by deep feature based techniques, and researchers have developed a number of object identification and recognition techniques based on CNNs [14]. These techniques have been created and used for hand detection as a result.

Though, results from image recognition can be applied to tasks in various areas of computer vision, such as object detection using the methods YOLO, R-CNN, fast R-CNN, and faster R-CNN, or semantic segmentation using U-Net [15].

By Roy et al. [16] it was recommended to employ a two-stage hand detector on the basis of the region-CNN (R-CNN) and Faster R-CNN frameworks. To increase the robustness of the deep features, Le et al. [17] suggested a novel technique that incorporated local and global context information. By aggregating several scale feature maps, they expanded the region-fully convolutional network (R-FCN) and faster R-CNN. On two difficult datasets, the performance of this method was adequate. Tokenization is a pre-processing method that Orbay et al. [18] suggested improves the success of translations. If supervised data is available, tokens can be learned from sign videos. Annotated data is, however, hard to come by and expensive to annotate at the gloss level. To find semi-supervised tokenization methods without the burden of extra labeling, adversarial, multitasking, and transfer learning were used. To undertake a more thorough examination, it offers numerous experiments that compare all the approaches in various contexts. In order to take use of the parallelism that all sub problems share, Oscar and colleagues [19] suggested a technique that exploits sequence limitations inside each separate stream and combines them by explicitly enforcing

synchronization points. Using the hybrid method, embed strong CNN-LSTM models in each HMM stream. This makes it possible to identify traits that don't have enough discriminative power on their own. Utilizing the sequential parallelism to train sign language, mouth shape, and hand shape classifiers, the approach is then applied to the area of sign language recognition.

Convolutional neural networks were suggested by Gruber et al. [20] for the classification of sign language number motions. The collection contains recordings of 18 distinct people that were made with the Kinect v2 device. In this study, just depth datastream was employed. Classic VGG16 architecture was used for a classification challenge, and its outcomes were compared with the chosen baseline approach and other examined architectures. Research and development of the assistive mobile information robot prototype was presented by Ryumin et al. [21].

The single-handed gesture detection system, the technical description of the robotic platform architecture, and the navigation algorithm are all based on a database of elements used in Russian sign language.

A unique approach based on fusing conventional hand-crafted features with a CNN was developed by Chevtchenko et al. [22]. They tested their approach using depth and grayscale photographs, where the background is eliminated using depth information and the hand is taken into consideration to be the nearest object to the camera.

CNNs were used by Liang et al. [23] to extract features from point clouds that a depth sensor had recorded. Since the first object detection methods in computer vision were advised, object annotation in digital images has generally been taken into consideration. Numerous studies have focused on accelerating the annotation of picture datasets for object detection tasks. Multiple techniques for bounding box annotation were advised by Papadopoulos et al. The annotator just requires checking the label intended by the network in their bounding box verification approach [24] with an accept / reject decision by humans.

Learning intelligent dialogs that take into account the benefit of a trained network to build a bounding box on the image was advised by Konyushkova et al. [25]. To validate the bounding box suggested by the detector in each image, a human annotator is required. The first step in fully annotating the initial batch of images from the unlabeled dataset is hand annotation. Drawing bounding boxes and assigning class names to photos is an entirely manual process that requires human intervention.

To address the issues with RCNN, some writers [26] suggest Fast RCNN, where each and every image is fed only once to the CNN, and feature maps are created using a selective search technique. To shorten the time required to detect, Ren et al. developed a Faster R-CNN modification to the Fast RCNN extension [27]. In order to localize the hand position in a background with no clutter, Soe and Naing [28] used the Faster R-CNN technique using the Caffe framework. Using the NUS dataset, Pisharady et al. [29] showed the segmentation strategy to detect the hand posture and achieved

93% accuracy.

Convolutional Neural Network (CNN) technology is the foundation of YOLO [30], which may produce quick and accurate object recognition. The state-of-the-art object detection technique is very quick from beginning to end. YOLO is frequently used to forecast object detection tasks like real-time pedestrian detection, mask detection, and traffic sign recognition. After the hand position has been localized using YOLOv3, the hand gesture is fed to CNN so that the motion can be detected. The YOLO (You Only Look Once) approach predicts the detected object in the input photos after only one viewing by the neural network. It operates by dividing the input image into several grids with predetermined grid sizes, and then calculating the likelihood that each grid contains the target object [31]. In a single algorithm run, it predicts every class and object bounds that are present in the image. The YOLO algorithm is also constantly being improved in terms of accuracy, speed, and lightweight. Then, You Only Look Once (YOLO) is advised by Redmon et al. [32] to localize the area of interest. Rotation estimation was provided by Denget et al. [33] using CNN to localize the hand region. Deep attention networks for hand gesture localisation were developed by Yuan Li et al. [34].

Shinde et al. used YOLO, which can precisely identify and locate the group frames or even single frames of human movements in the video, to complete the recognition and location of human motions [35]. Based on the enhanced YOLO-v4 [36], Yu et al. proposed a face mask recognition and standard wear detection algorithm.

YOLOv5 was employed in some recent experiments to detect various items. Some recent research looked at replacing the manual inspection procedure with the YOLOv5 during the COVID-19 phase to check for the social distancing proposed by Shukla et al. [37] and face mask by Yang et al. [38] from video and still photos. The model developed by Wang et al. [39] for the detection of safety helmets and tree leaves has been applied in a few other researches. Again, in a number of studies, the YOLOv5 surpassed the R-CNN and other YOLO in terms of speed and accuracy. Then, these features are treated by an algorithm that identifies the specific hand gesture, such as Support Vector Machines (SVM) [11], Conditional Random Fields (CRF), Hidden Markov Models (HMM), and Convolutional Neural Networks (CNN).

After deep learning methods were established, CNN became a more widely used technique for replacing previous methods in object recognition and classification tasks. One of the most difficult issues in this area of computer vision is object detection. Localizing various items in a scene and labeling their bounding boxes are the goals of object detection. The most crucial strategy for solving this issue is to use already trained classifiers to give bounding boxes in scene names [40].

III. MATERIALS AND METHODS

The techniques and resources that were used in this study to achieve the hand gesture recognition that this paper focused on are assigned to this part. Fig. 1 depicts the suggested hand gesture recognition flowchart and the methodology that was used.

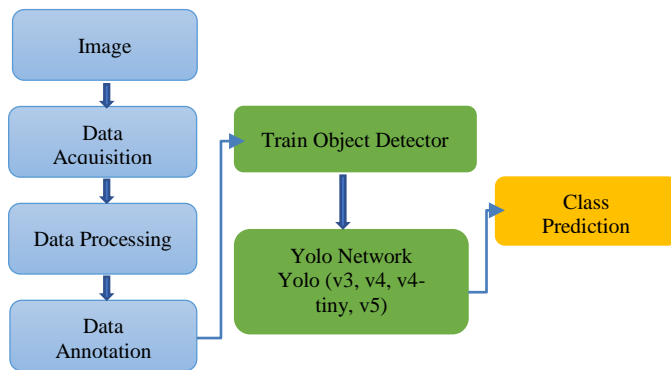


Fig. 1. The flowchart of the proposed hand gesture recognition.

The proposed method was divided into different steps. Firstly, the hand images were collected from the image database for tiny hand gesture recognition and underwent data augmentation to create a hand motions dataset. The captured image is passed through an annotation format to draw a Bounding boxbased hand detector to extract hand regions. Bounding boxes were manually drawn around specific objects in the images to annotate them. The hand region is then extracted once the hand has been recognized and transferred to a better Yolo (You Only Look Once) deep learning model. This model was then optimized and trained on the created datasets. The dataset has seven different gesture classes, such as a Fist, I, Pointer, Palm, Ok, Thumb up and Thumb down. To verify the detection performance, evaluation metrics were produced. The best model was ultimately chosen for the best hand detection across many images.

A. Dataset Collection

Data 1: This dataset [41] includes 1400 motions made by 14 distinct individuals, whom each made 10 different gestures and repeated them 10 times. The dataset includes a variety of distinct gestures that were captured using both the Leap Motion and Kinect devices, enabling the development and testing of hybrid gesture recognition systems that utilize both sensors, as suggested in the study, or the comparison of the two sensors.

Data 2: The dataset [42] [43] [44] includes a variety of static motions that were captured with the Creative Senz3D camera. While this camera works well for short range depth collection, its depth range is constrained, and its far range noise level is significant. It has been used to evaluate the performance of a Multi-Class SVM gesture classifier that was trained on fictitious data produced HandPoseGenerator.

B. Data Acquisition

In this paper, the hand images were collected from the image database for tiny hand gesture recognition. This dataset [43] has been collected from forty participants; each individual was invited to make seven different gestures. Each instantiation of a gesture consists of around 1400 color frames, and the gestures are carried out in various places throughout the image. The majority of the people have complex backgrounds, with the remaining 50% having basic backgrounds. Backgrounds that are thought to be complicated are extremely crowded, and the lighting varies greatly. The human face and body make up the majority of each image, whereas the hand gestures that need to be categorized only take up 10% of the total number of

pixels. We establish a dataset for different person classification and detection. Our dataset contained a total of 3600 images. These images were further divided into seven different classes. Each class comprises an average of 500 images which were labeled and annotated using the bounding box. Our classes started from palm to thumb down, which were finger-pointing different positions. The hand gestures in our dataset are shown in Fig. 2.



Fig. 2. Sample instances of each class from the dataset with the changes in hand position, shape, and scale.

C. Data Pre-processing

By artificially increasing the dataset, data augmentation is a key strategy for creating variations of the training and testing datasets. This step consists to utilize the augmentation techniques such as brightness transformation, randomly altering rotation, motion blur, blurring, and the scale of an input image necessitates that a model contemplates what an image subject looks like in a diversity of positions. Each image was repeated for reading and training, both for the left and right hand, by flipping it horizontally, and sometimes capturing the respective image of those hands to make the set more accurate, using a YOLO setup with a total of images from the dataset.

Additionally, each image for the testing set was captured and labeled. Before moving on to post-processing, it is vital to perform data pretreatment so that we can determine the type of data we have collected and which portions will be relevant for training, testing, and improving accuracy. This part presents the system or methods used to classify, select, and process as well as analyze data and its recognition of characters is discussed. The following methodology is employed to collect data in the form of images, preprocess the data, and then feed the processed data to our model.

1) *Manual annotation:* The annotation procedure, the training and validation set images were originally 240×240 pixels in size. We utilized the internet tool Roboflow to construct the bounding boxes for each image (www.roboflow.com). This page facilitates making data labels and annotating in the desired format. The images were annotated using the Roboflow Annotate, which is a self-serve annotation tool, and that greatly accelerates the transition from untrained and deployed computer vision models to raw images. After manually drawing and categorizing bounding

boxes, this tool makes it possible to change just one annotation or label throughout the whole dataset.

2) *Object detection*: The object detection model is trained in this section. We concentrate on the most recent deep learning-based object detection models, albeit any detector can be used. In the following part, we'll go into more detail about our training methods. To determine the existence, quantity, and placement of objects in a picture, object detection models are used. Drawing a bounding box around each object of interest in each image was necessary for the image annotation model, which enables us to determine the precise location and quantity of objects in an image. In contrast to image classification, where the class placement within the image is irrelevant because the entire image is designated as one class, the class location is a parameter in addition to the class. Bounding boxes and polygons are examples of labels that can be used to annotate objects inside a picture. Find the existence of things in an image using a bounding box and the types or classes of the objects you find.

a) *Input*: An image that includes one or more items, like a photo.

b) *Output*: A class label for each bounding box as well as one or more bounding boxes (each defined by a point, width, and height).

3) *Image data labeling with bounding box*: We have also produced a dataset with bounding box labeling so that we may utilize the characteristics of the deep learning detection technique. In order to reduce the difficulty and expense of labeling, we randomly choose a few images from each class in the dataset and choose to label the bounding boxes. The most popular annotation shape in computer vision is the bounding box.

Angular boxes called bounding boxes are used to specify where an object is located inside an image. Both two-dimensional (2D) and three-dimensional (3D) models are possible (3D). Polygons or rectangular shapes were manually drawn to annotate the object's edges and to mark each of the object's vertices. The x_center , y_center , width, and height of an object's boundary show its exact location in that image. As shown in Fig. 3, the rectangular shapes are used to label different hands.

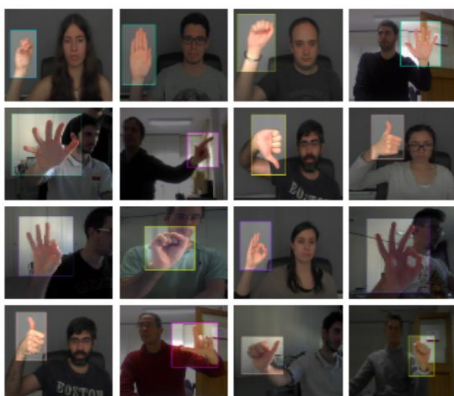


Fig. 3. Labeling different hand classes with bounding boxes.

D. Labeled Dataset

In a labeled dataset, each element of the unlabeled data is given a meaningful "label", "tag" or "class" that makes it more desired or instructive to identify it. Bounding box inference in the training detection model continues until all unlabeled images have been manually fully tagged. In our model we annotated the dataset, we introduce seven different gesture classes, such as Fist, I, Pointer, Palm, Ok, Thumb up and Thumb down.

E. The Structure of the Proposed YOLO Algorithms

1) *You Only Look Once (YOLO)*: YOLO means You Only Look Once is a method that detects all objects in a frame or image in a single shot. Mainly, YOLO makes use of only convolutional layers, to determine which items are represented in the image, a single fully convolutional network (FCN) is used. The YOLO method divides the image into cells or grids; each cell is responsible for object localisation, estimating the number of bounding boxes, and calculating class probabilities. The dataset is collected from various people with various complex backgrounds at different positions, such as variable illumination, gesture variations, and low resolution. Labeling images is essential for good computer vision models. All the images are annotated and labeled manually with Roboflow Annotate which represents a self-serve annotation tool. In this study, we provide a dataset called "BdHand," to which we add bounding boxes to roughly 5000 images in order to make use of the potential of object detection techniques. After the first step of preprocessing and the manual annotation, the second one is training the deep learning models using modern YOLO algorithms (YOLO v3, YOLO v4, YOLO v4 Tiny, and YOLO v5). To understand the different algorithms which we are proposing, the diagram presented in Fig. 1 shows the detection of objects. At first, the first step in the training process is to gather the data, and the second is to label it. We label our dataset using YOLO annotation, which gives us certain values that are later detailed in the model process. We feed the dataset to the DarkNet-53 (YOLO v3) model afterwards, after it has been annotated with YOLO annotation.

2) *YOLO v3 Model*: YOLO v3 represents an improvement of the essential idea of YOLO, It enables partitioning an image into cells that are in charge of object prediction. Feature extraction networks and the use of detection at multiple scales are changes from YOLO, and the bounding boxes. YOLO v3 [45] presents a deeper architecture of a feature extractor named Darknet-53. It has 53 convolutional layers with a batch normalization layer and leaky Relu activation layer after each one. The feature maps are downsampled using a convolutional layer with stride 2 and without using any kind of pooling [46]. This aids in avoiding the loss of low-level characteristics that pooling is sometimes bed for. As illustrated in Fig. 4, our technique separates the input image using grids into an $S \times S$ region first. These cells are used to carry out operations on class probability and bounding box estimates. If the detection of an object in a grid cell is carried out by the object's center. There are now a variable amount of images in our dataset of

collected images. The classes we have stored for our YOLO technique are used to label these photographs, and once that is done, we have successfully determined the class and the coordinate for our image set. Additionally, we describe the method by which YOLO manages the network-aimed output, which is achieved by using a formula that requires various coordinates. The b_x, b_y, b_w, b_h are the variables that we employ for the bounding box dimensions and are associated with (x,y) coordinates that represent the center of the box, as well as the width and height, which are represented by p_w, p_h . The estimated four coordinates are t_x, t_y, t_w, t_h , a bounding box for each. The numbers c_x and c_y correspond to the grid cell's upper left coordinates. These variables, which reflect the box prediction components as defined by Equations, are predicted in relation to the entire image (Eq. (1) to (5)).

$$b_x = \sigma(t_x) + C_x \quad (1)$$

$$b_w = \sigma(t_y) + C_y \quad (2)$$

$$b_y = p_w e^{t_w} \quad (3)$$

$$b_h = p_h e^{t_h} \quad (4)$$

$$\sigma(x) = 1 / (1 + e^{-x}) \quad (5)$$

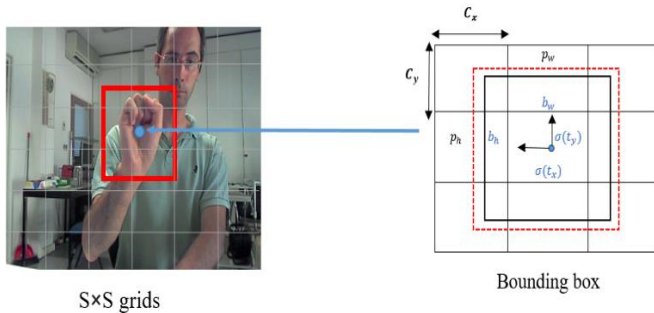


Fig. 4. The bounding boxes with dimension priors and location prediction.

Fig. 4 demonstrates how each value of the algorithm's bounding box gives us the x and y coordinates for the center.

3) *YOLO v4 model*: The one-stage object identification technique known as YOLOv4 represents the YOLOv3 model's evolution and significant advancement. The number of FPS (Frames per Second) increased by 12% and the mAP (mean Average Precision) by 10% as a result of the introduction of a new architecture in the Backbone and changes in the Neck [47]. The architecture of YOLOv4 is made up of the YOLOv3 head, PANet path aggregation neck, spatial pyramid pooling extra module, and Darknet53 as the backbone.

4) *YOLOv4-Tiny model*: The YOLOv4-tiny model is based on the YOLOv4 approach and is aimed to increase

object detection speed. The prediction process is the same as YOLOv4 and it has a faster target detection speed. YOLOv4-tiny [48] [49] is proposed to reduce parameters and make the network structure simpler and it significantly improves the viability of implementing object detection methods on embedded systems or mobile devices. The Yolov4-tiny method utilized Darknet53-tiny network as backbone network to instead of the CSPDarknet53 network that is used in Yolov4 method. The ResBlock module in the residual network is substituted by the Block module in the CSPDarknet53-tiny network. Fig. 5 depicts the YOLOv4-tiny network structure.

5) *YOLO v5 model*: The Backbone, Neck, and Head architectural components of the YOLOv5 network are shown in Fig. 6. YOLOv5 Backbone: In order to extract features from images, including cross-stage partial networks, YOLOv5 uses CSPDarknet as its backbone. YOLOv5 Neck: It makes use of PANet to create a feature pyramid network that is then passed to the Head for prediction after the features have been aggregated. YOLOv5 Head: Its layers produce predictions for object detection from the anchor boxes [50].

YOLOv5 is quick and lightweight, and it uses less computing power than other current state-of-the-art architecture models while maintaining accuracy levels that are comparable to those of current state-of-the-art detection models. Compared to the other YOLO versions, it is substantially faster. CSPNET serves as the foundation for YOLOv5's feature map extraction from the image. In order to improve information flow, it also makes use of the Path Aggregation Network (PANet) [51]. For the following reasons, we are utilizing YOLOv5 as it includes helpful elements like a cutting-edge activation function, a convenient manual, a hyperparameter, and a data augmentation technique. It can be trained computationally quickly with minimal resources, thanks to its lightweight architecture. The size model can be utilized with mobile devices because it is relatively tiny and light. Yolov5 differs from the Yolo series in several lighting areas: (1) Multiscale: utilize FPN to improve the feature extraction network rather than PAN [46], which will make the model easier to use and more quickly.

Yolov5 differs from the Yolo series in several lighting areas: (1) Multiscale: utilize FPN to improve the feature extraction network rather than PAN [46], which will make the model easier to use and more quickly. (2) Target overlap: identify nearby positions using the rounding method such that the target is mapped to several central grid points all around it. Yolov5 is a continuation of the YOLO series' most recent iterations [52]. It is more manageable and, in general, more cozy to utilize throughout training. Its architecture may be modified with equal ease, and it can be exported to numerous deployment environments [53].

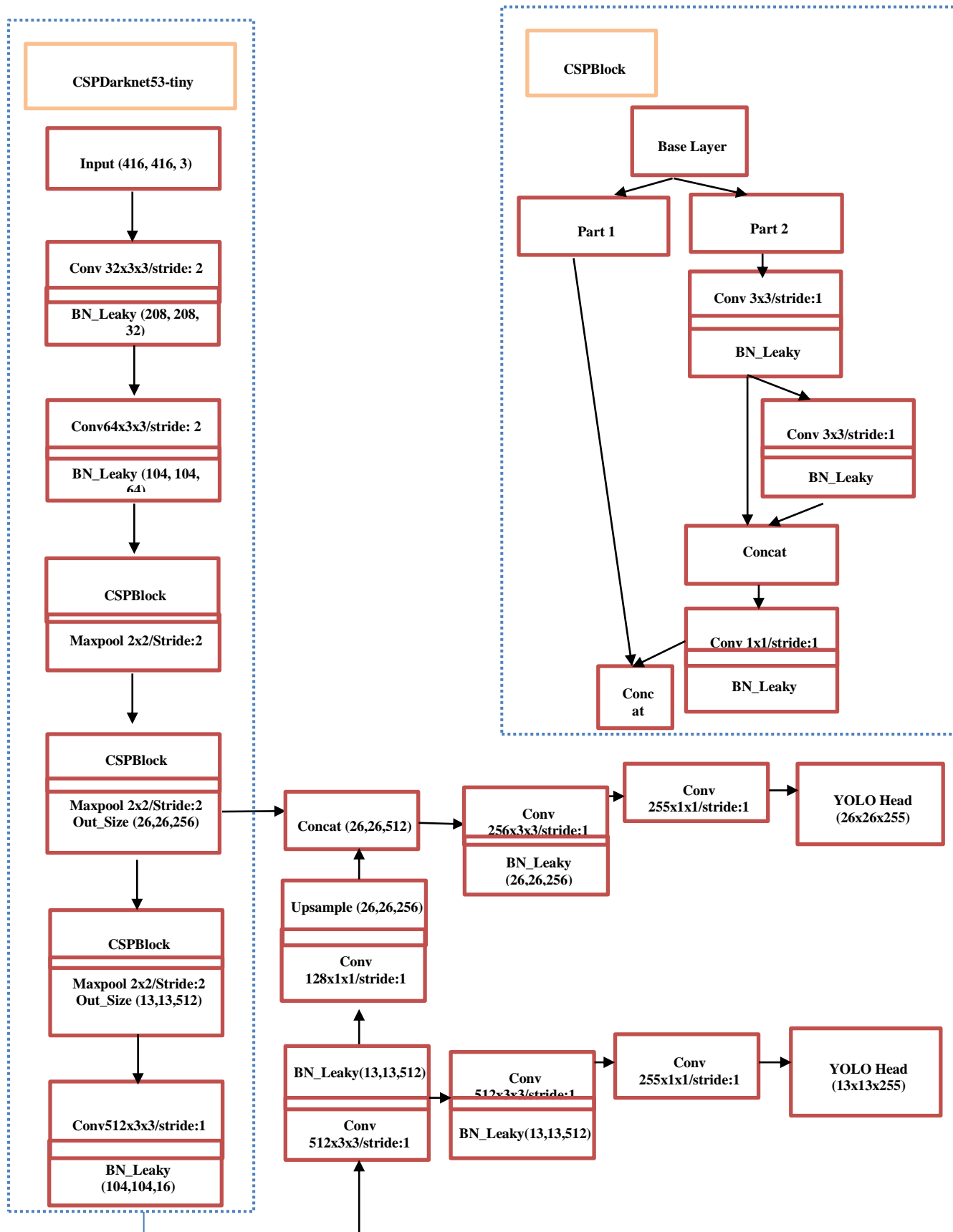


Fig. 5. The network structure of YOLOv4-tiny.

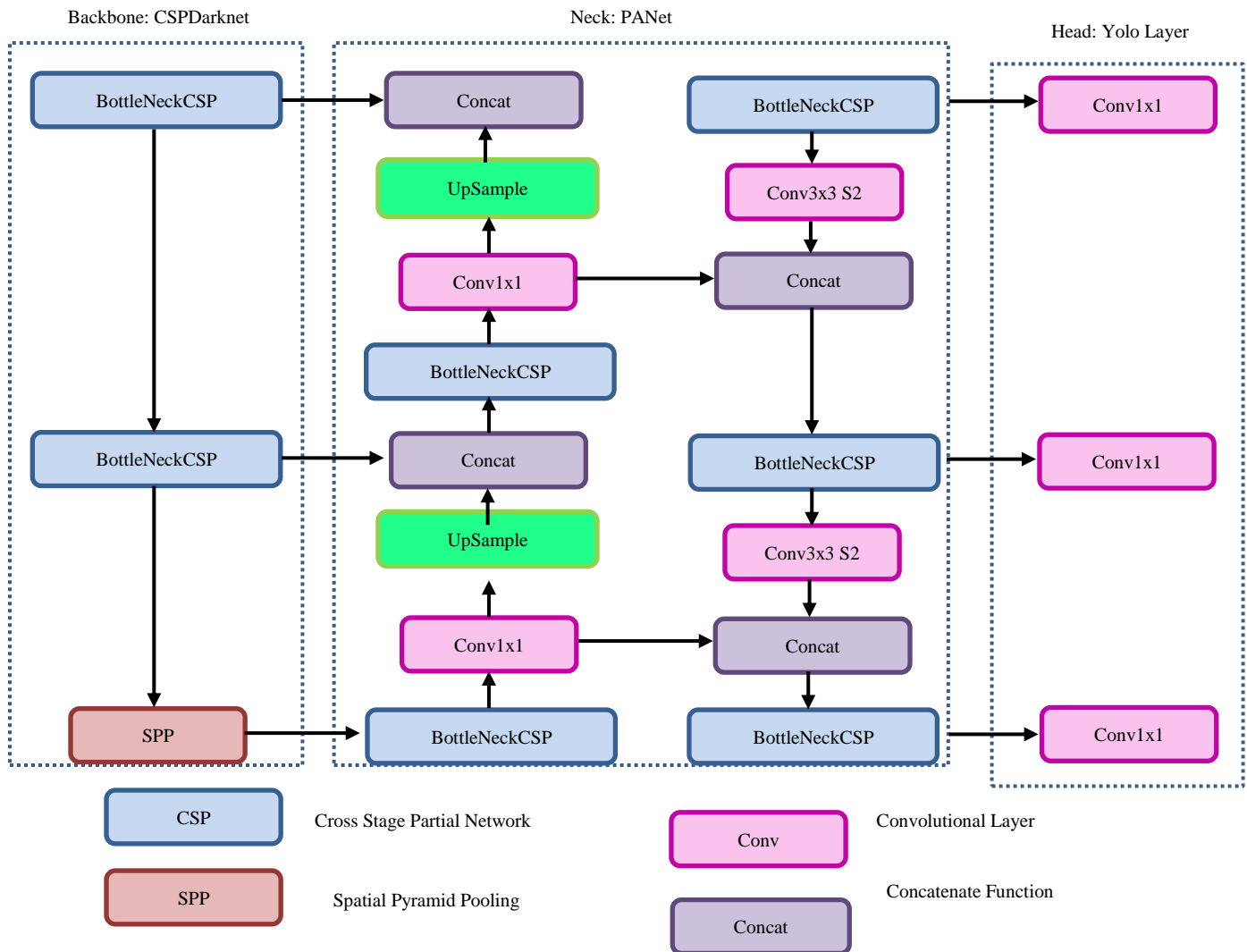


Fig. 6. The general architecture of the YOLOv5 network.

There are many algorithm parameters in the YOLO models, and understanding the influence of these parameters is essential for optimizing the performance of the model for a specific task. Here are some of the most important parameters in YOLO models and their influence:

- Input size: The input size of the YOLO model refers to the resolution of the input image. Larger input sizes cannot only improve the accuracy of the model but also increase the computational cost.
- Anchor boxes: Anchor boxes are predefined boxes of various shapes and sizes that are used to predict object locations and sizes. The number and aspect ratio of anchor boxes can significantly affect the accuracy of the model.
- Batch size: The batch size refers to the number of images processed in a single iteration during training. Larger batch sizes can speed up the training process, but they also require more memory.
- Confidence threshold: The confidence threshold is used to filter out low-confidence predictions. Increasing the confidence threshold can reduce the number of false positives but may also increase the number of false negatives.
- increasing the confidence threshold can reduce the number of false positives but may also increase the number of false negatives.
- NMS threshold: Non-Maximum Suppression (NMS) is used to remove overlapping bounding boxes. The NMS threshold controls the amount of overlap allowed between boxes. A higher threshold can remove more overlapping boxes but may also remove some true positives.
- Backbone architecture: The backbone architecture refers to the architecture used to extract features from the input image. Different architectures have different complexities and can affect the accuracy and speed of the model.

- Training parameters: Training parameters such as learning rate, weight decay, and optimizer can significantly affect the training process and the performance of the model.

The parameters in YOLO models can significantly affect the accuracy, speed, and memory usage of the model. Choosing the right parameters for a specific task requires experimentation and fine-tuning to optimize the performance of the model.

IV. EXPERIMENTS AND RESULTS

A. Evaluation Metrics

In this section, we discuss the four experiments performed using Yolo algorithms YOLO version (v3, v4, v4 tiny, and v5). We implemented and tested the four models during our experiments to train them for our different publicly available datasets. The configuration of these models differs from one to the other. Extensive testing was carried out during our research to confirm the dependability of the suggested YOLO model for hand gesture recognition. The experimental settings are presented in the first stage. The evaluation metrics are then described. The comparative experimental findings are then thoroughly examined and analyzed. To gauge the effectiveness of the proposed hand gesture recognition model in terms of recognition, detection, and computational performance, a number of indicators or metrics were used. The average precision (AP), which is shown as the area under the precision and recall curve at various detection thresholds, was used in this experiment. Eq. (6) contains a definition of the AP equation.

$$AP = \int_1^0 P_r(R_c) dR_c \quad (6)$$

Precision and recall are represented by Pr and Rc. The following parameters of precision, recall, and F1-score are calculated to estimate the model accuracy and the efficiency. When the predicted bounding boxes match the ground truth boxes, the accuracy of the prediction is measured. In addition to these measures, we used Eq. (7), (8), and (9) to derive precision, recall, F1-score, and accuracy using the True positive (TP), False positive (FP), and False negative (FN) metrics. The Precision (Pr), presented in Eq. (7), shows the ratio of true positives (TP) to all expected positives (TP+FP). As a result, it is a crucial measure for deducting the cost of the FP number.

$$P_r = \frac{Tp}{Tp+Fp} \quad (7)$$

If the predicted bounding box falls beyond the ground truth of the hand, it is indicated by the letters FP, whereas TP signifies that it does so. The likelihood of correctly detecting the ground truth objects is then calculated using recall. Accordingly, the Recall (Rc) shows the proportion of estimated true positives to all actual positives (TP+FN). It is created by Eq. (8) and is occasionally referred to as sensitivity. Instead of the projected bounding box, FN displays the hand of the frame.

$$R_c = \frac{Tp}{Tp+FN} \quad (8)$$

The F_1 -score measures the overall accuracy, this as shown

in Eq. (9), includes the recall values and a statistical precision measure. The F_1 -score should be chosen in particular when a balance between precision and recall is necessary, with an F_1 -score of 1 defining the optimal value.

$$R_c = \frac{2*P_r*R_c}{P_r+R_c} \quad (9)$$

Mean Average Precision (mAP), a well-liked object identification statistic created by Eq. (10), averages the AP values for all classes. As a result, the performance of the model may be quantified using a single metric.

$$mAP = \frac{\sum_{q=1}^Q AveP(q)}{Q} \quad (10)$$

Where Q is the number of queries in the set, q is the query for average precision. The mAP is the mean value of average precision for the detection of all classes and is an indicator generally utilized to estimate how good a model is. The FPS identifies how many images can be correctly identified in a single second. GPU utilization refers to the use of GPU RAM when evaluating various detection strategies.

B. Detection Results of YOLO Model

The output of the various classes of hand gesture recognition is shown in Fig. 7. The bounding box covered the maximum part of the hand. It will cleverly determine which gesture is being represented in the zoom situation when the object is too huge, and it then delivers the class ID with the best match. To determine which algorithm was the most effective for hand gesture detection, we used a variety of different ones.



Fig. 7. Detection results of YOLO v3 model.

416 x 416 pixels were chosen as the size of the input images for the training process. The outcomes of the hand detection test utilizing our suggested YOLO v3 model are listed in Table I. We calculated the mean average precision (mAP), then Precision (Pr), average recall (Rc), and F_1 -score for each test. Using the suggested deep learning model, we assessed the performance of deep learning models, and we were able to get an accuracy of 98.20% for YOLOv3.

TABLE I. PERFORMANCE ACCURACY FOR YOLOV3

Class	Precision %	Recall%	F1-Score	mAP(mAp@.5)%
Fist	62.40	96.80	73.60	93.70
I	88.40	97.10	91.40	96.40
Pointer	95.70	98.40	96.90	97.60
Ok	96.80	98.60	97.60	98.10
Palm	96.70	98.60	97.60	98.10
Thumb down	97.30	98.60	97.90	98.20
Thumb up	96.40	98.40	97.40	98.00

The suggested model also worked well in various lighting situations, as seen in Fig. 8. The experimental findings show that the proposed model can accurately and efficiently identify different classes of hands in a variety of situations.



Fig. 8. Detection results of YOLO v4 model.

Table II displays the experimental outcomes for hand gesture recognition using our dataset. This table compared the speed and accuracy of the various classes. Experiments show that 98.40% of the results are accurate.

TABLE II. PERFORMANCE ACCURACY FOR YOLOV4

Class	Precision %	Recall%	F1-Score	mAP(mAp@.5)%
Fist	97.10	98.50	97.79	98.20
I	97.90	98.60	98.40	98.40
Pointer	97.30	98.60	97.94	98.30
Ok	96.90	98.60	97.74	98.40
Palm	96.70	98.60	97.60	98.10
Thumb down	97.30	98.60	97.90	98.20
Thumb up	96.40	98.40	97.40	98.00

Fig. 9 depicts the process for detecting hand gestures. As these results exhibit, our proposed model can treat various shapes of hands, scales, and under various lighting circumstances, as well as comprehend motions in many difficult situations.

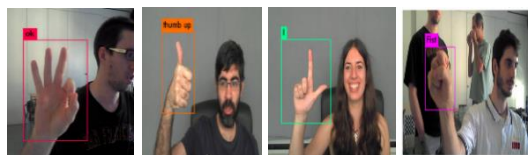


Fig. 9. Detection results of YOLO v4-tiny model.

We conclude that effective hand detection improves the performance of the gesture recognition system with quick processing, which in turn facilitates accurate human-machine interaction, based on the experimental findings provided in Tables II and III.

TABLE III. PERFORMANCE ACCURACY FOR YOLOV4-TINY

Class	Precision %	Recall%	F1-Score
Fist	1.0	92.30	95.00
I	97.0	92.20	94.82
Pointer	1.0	88.50	93.89
Ok	98.870	91.56	95.01
Palm	1.0	92.63	96.17
Thumb down	1.0	88.30	93.80
Thumb up	1.0	90.27	94.88

Fig. 10 introduces the results of test images that contain different people. The results of the experiments show that the suggested model can meet object detection in various complicated backgrounds where the majority of movements were successfully detected.



Fig. 10. Detection results of YOLO v5 model.

In this experiment, YOLOv5 performed better overall than YOLOv4, YOLOv4-tiny, and YOLOv3. In comparison to the other models, the YOLOv5 model produced the best results in terms of precision and error. Compared to YOLOv4, YOLOv5 is quicker and more accurate. The results showed that the mAP was much higher when YOLOv5 was compared to YOLOv3, YOLOv4, and YOLOV4-tiny for hand motion recognition. The most effective object detection method at the moment is YOLOv5 (refer Table IV).

TABLE IV. PERFORMANCE ACCURACY FOR YOLOV5

Class	Precision %	Recall%	F1-Score	mAP(mAp@.5)%
Fist	1.00	97.70	98.83	97.80
I	98.80	97.60	98.19	98.10
Pointer	99.9	96.40	98.11	96.70
Ok	98.60	1.0	99.29	99.10
Palm	99.60	1.0	99.79	99.50
Thumb down	1.0	98.10	99.04	98.60
Thumb up	99.80	1.0	99.89	99.50

V. DISCUSSION

YOLO (You Only Look Once) is an object detection algorithm that predicts the bounding boxes and class probabilities of objects in an input image. YOLOv3 and YOLOv4 are earlier versions of the algorithm, while YOLOv5 is a more recent version. Here are some of the differences between these versions:

- Architecture: YOLOv5 uses a different architecture than its predecessors. It has a smaller and more efficient model that makes use of Scaled-YOLOv4 architecture and advanced training techniques such as Mosaic data augmentation.
- Speed: YOLOv5 is faster than its predecessors, particularly YOLOv3, due to its smaller model size and improved architecture. YOLOv5 can process up to 155 frames per second on a Tesla V100 GPU, compared to 82 and 65 frames per second for YOLOv4 and YOLOv3, respectively.
- Accuracy: YOLOv4 is generally more accurate than YOLOv3, with improvements in object detection accuracy and speed. YOLOv5, on the other hand, achieves comparable accuracy to YOLOv4 but with a smaller model size and faster processing speed.
- Training: YOLOv5 uses a different training approach called self-supervised pre-training, which allows it to learn from large amounts of unlabeled data. This leads to better generalization and improved performance on smaller datasets.

The discussion is about the performance evaluation of a proposed deep learning model for hand gesture recognition. The model achieved an accuracy of 98.20% for YOLOv3, which was able to identify different classes of hands in various lighting situations. The Table I provided shows the precision, recall, F1-Score, and mAP scores for various hand gesture classes. The "Fist" gesture had the lowest precision score of 62.40%, while the "Thumb down" and "Thumb up" gestures had the highest precision scores. The recall scores were high for all classes, indicating that the model correctly identified a large proportion of actual positive instances.

The Table II shows the precision, recall, F1-Score, and mAP scores for various hand gesture classes. The precision scores for all classes were high, ranging from 96.4% to 97.9%. The recall scores were also high, ranging from 98.4% to 98.6%, indicating that the model correctly identified a large proportion of actual positive instances. The F1-Scores were all above 97%, indicating a high level of accuracy in detecting and recognizing hand gestures. The mAP scores were also high, ranging from 98.0% to 98.4%, indicating that the model was able to detect the objects with high precision across all the classes.

The Table III displays the precision, recall, and F1-Score for various hand gesture classes. The precision score for most classes is high, ranging from 1.0% to 98.87%. However, the precision score for the "Fist," "Pointer," and "Palm" classes is only 1.0%, which indicates that the model produced a large number of false positives for these classes. The recall score for

all classes was above 88%, indicating that the model correctly identified a large proportion of actual positive instances. The F1-Score for all classes was above 93%, which indicates a high level of accuracy in detecting and recognizing hand gestures.

Compared to the previous tables, this table shows higher precision, recall, F1-score, and mAP values for most of the hand gesture classes. The precision, recall, and F1-score for the "Fist" and "Pointer" classes have significantly improved from the previous table, reaching perfect precision for the "Fist" class and near-perfect precision for the "Pointer" class. The "Ok" and "Palm" classes also showed improvement in precision and F1-score, although their recall values were 1.0, indicating that there were no false negatives. The "Thumb down" and "Thumb up" classes also demonstrated high precision and F1-score values.

YOLOv5 is a more efficient and faster version of the YOLO algorithm with comparable accuracy to YOLOv4. While YOLOv3 is still a popular choice for object detection, YOLOv5 offers improved performance and training techniques.

Despite recent enormous advancements in object detection, it is still challenging to detect and classify objects rapidly and accurately.

The YOLOv5 method was cited by Yan et al. (2021) as the most potent object-detecting algorithm available today.

In the current study, YOLOv5 outperformed YOLOv4 and YOLOv3 in terms of overall performance.

As we discovered multiple studies comparing YOLOv5 to earlier iterations of YOLO, such as YOLOv4 or YOLOv3, this conclusion is consistent with some earlier studies. Thuan (2021) claims that YOLOv5 is more precise and quick than YOLOv4.

VI. CONCLUSION

The proposed technique using the YOLO model can significantly enhance communication for deaf or hard-of-hearing individuals, regardless of their disability.

The recent advancements in computer vision and deep learning have improved the accuracy of object detection, and this study utilizes this progress to develop a hand gesture recognition system. The model for hand gesture recognition is based on the deep learning models YOLO (YOLOv3, YOLOv4, YOLOv4-tiny, and YOLOv5) to recognize motions and classes in sign language. The experiments conducted show that the suggested YOLO model has exceptional detection and performance, with a 99.50% accuracy rate when identifying objects and gestures from various datasets. The proposed method clusters the dataset based on the suggested algorithm, which necessitates manual annotation of a number of classes and analysis for patterns that aid in target prediction. The results demonstrate that the suggested YOLOv5 method outperformed the YOLOv3, YOLOv4, and YOLOv4-tiny algorithms in all datasets and improved the hand detection performance. By leveraging messaging or video calling, this technique can help overcome the obstacles of communication faced by deaf or hard-of-hearing individuals and enable them

to interact with others more effectively. There are several directions this research can take. Another approach could be to use a combination of YOLO models for different stages of the hand gesture recognition pipeline. The YOLOv3 model could be used to detect the hand region in an image, and a YOLOv4 model could be used to classify the hand gesture. This approach could improve the accuracy of the system while also reducing the computational cost.

REFERENCES

- [1] R. P. Sharma, G. K. Verma, "Human computer interaction using hand gesture". *Procedia Computer Science*, vol. 54, pp.721-727, 2015.
- [2] D. Phutela, "The importance of non-verbal communication". *IUP Journal of Soft Skills*, vol. 9, no. 4, pp. 43.
- [3] A. A. Q. Mohammed, J. Jiancheng and M. S. Islam, "A deep learning-based end-to-end composite system for hand detection and gesture recognition". *Sensors*, vol. 19, no. 23, pp. 5282, 2019.
- [4] P. Nakjai, T. Katanyukul, "Hand Sign Recognition for Thai Finger Spelling: An Application of Convolution Neural Network". *Journal of Signal Processing Systems*, vol 91, pp. 131–146, 2019.
- [5] M. M. Alam, M. T. Islam, and S. M. Rahman, "A unified learning approach for hand gesture recognition and fingertip detection". *UMBC Student Collection*, 2021.
- [6] S. Pramada, D. Saylee, N. Pranita, N. Samiksha, N., and M. S. Vaidya, "Intelligent sign language recognition using image processing". *IOSR Journal of Engineering (IOSRJEN)*, vol. 3, no. 2, pp. 45-51, 2013.
- [7] N. Cihan Camgoz, S. Hadfield, O. Koller, and R. Bowden, "Subnets: End-to-end hand shape and continuous sign language recognition". In *Proceedings of the IEEE international conference on computer vision*, pp. 3056-3065, 2017.
- [8] A. Kusters, "International Sign and American Sign Language as different types of global deaf lingua francas". *Sign Language Studies*, vol. 21, no. 4, pp. 391-426, 2021.
- [9] A. I. Khan, and S. Al-Habsi, "Machine learning in computer vision". *Procedia Computer Science*, vol.167, pp. 1444-1451, 2020.
- [10] N. O'Mahony, S. Campbell, A. Carvalho, S. Harapanahalli, G. V. Hernandez, L. Krpalkova, and J. Walsh, "Deep learning vs. traditional computer vision". In *Science and information conference*, pp. 128-144, 2020.
- [11] C. Wang, and Z. Peng, "Design and implementation of an object detection system using faster R-CNN". In *2019 International Conference on Robots & Intelligent System (ICRIS)*, pp. 204-206, 2019 IEEE.
- [12] M. Oudah, A. Al-Naji, and J. Chahl, "Hand gesture recognition based on computer vision: a review of techniques". *Journal of Imaging*, vol. 6 , no. 8, pp. 73, 2020.
- [13] H. Huang, Y. Chong, C. Nie, and S. Pan, "Hand gesture recognition with skin detection and deep learning method". In *Journal of Physics: Conference Series*, vol. 1213, no. 2, pp. 022001., 2019, IOP Publishing.
- [14] Z.Q. Zhao, P. Zheng, S.T Xu, and X. Wu, "Object detection with deep learning: A review". *IEEE transactions on neural networks and learning systems*, vol. 30, no. 11, pp. 3212-3232, 2019.
- [15] F. Sandelin, "Semantic and instance segmentation of room features in floor plans using Mask R-CNN". 2019.
- [16] K. Roy, A. Mohanty, R. R Sahay, "Deep Learning Based Hand Detection in Cluttered Environment Using Skin Segmentation". In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pp. 640–649, 2017.
- [17] T.H.N. Le, C. Zhu, Y. Zheng, K. Luu, and M. Savvides, "M. Robust hand detection in Vehicles". In *Proceedings of the International Conference on Pattern Recognition (ICPR)*, pp. 573–578, 2016.
- [18] A. Orbay, and L. Akarun, "Neural sign language translation by learning tokenization". In *15th IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 222-228, 2020.
- [19] O. Koller, N.C Camgoz, H. Ney, H.; and R. Bowden, "Weakly supervised learning with multi-stream CNN-LSTM-HMMs to discover sequential parallelism in sign language videos". *IEEE transactions on pattern analysis and machine intelligence*, vol. 42, no. 9, pp. 2306-2320, 2019.
- [20] I. Gruber, D. Ryumin, M. Hruz, and A. Karpov, "Sign language numeral gestures recognition using convolutional neural network". In *International Conference on Interactive Collaborative Robotics*, pp. 70-77, 2018.
- [21] D. Ryumin, I. Kagirov, A. Axyonov, et al, "A multimodal user interface for an assistive robotic shopping cart". *Electronics*, vol. 9, no. 12, pp. 2093, 2019.
- [22] S. F. Chevtchenko, R. F. Vale, V. Macario, and F. R. Cordeiro, "A convolutional neural network with feature fusion for real-time hand posture recognition". *Appl. Soft Comput. J*, vol. 73, pp. 748-766, 2018.
- [23] C. Liang, Y. Song, and Y. Zhang, "Hand gesture recognition using view projection from point cloud". In *Proceedings of the International Conference on Image Processing, (ICIP)*, Phoenix, AZ, USA, 25–28, pp. 4413–4417, September 2016.
- [24] D. P. Papadopoulos, J. R. R. Uijlings, F. Keller, and V. Ferrari, "We don't need no bounding-boxes: Training object class detectors using only human verification". In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 854-863, 2016.
- [25] K. Konyushkova, J. R. R. Uijlings, C. H. Lampert, and V. Ferrari, "Learning intelligent dialogs for bounding box annotation". In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 9175–9184, 2018.
- [26] N. D. Vo, K. Nguyen, T.V. Nguyen, and K. Nguyen, "Ensemble of deep object detectors for page object detection". In *Proceedings of the 12th International Conference on Ubiquitous Information Management and Communication*, pp. 1-6, January 2018.
- [27] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks". In *Advances in Neural Information Processing Systems (NIPS)*, vol 28, 2015.
- [28] H. M. Soe, and T. M. Naing, "Real-time hand pose recognition using faster region-based convolutional neural network". In *Big Data Analysis and Deep Learning Applications: Proceedings of the First International Conference on Big Data Analysis and Deep Learning 1st*, pp. 104-112, Springer Singapore, 2019.
- [29] P. K. Pisharady, P. Vadakkepat, and A. P. Loh, "Attention based detection and recognition of hand postures against complex backgrounds". *International Journal of Computer Vision*, vol. 101, pp. 403–419, 2013.
- [30] M. M. William, P. S. Zaki, B. K. Soliman, K. G. Alexsan, M. Mansour, M. El-Moursy, and K. Khalil, "Traffic signs detection and recognition system using deep learning". In *2019 Ninth international conference on intelligent computing and information systems (ICICIS)*, pp. 160-166, 2019, IEEE.
- [31] G. Jocher, et al, "Ultralytics/yolov5: v5. 0-YOLOv5-P6 1280 models AWS Supervise. ly and YouTube integrations". *Zenodo*, vol. 11, 2021.
- [32] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection". In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 779–788, 2016.
- [33] X. Deng, Y. Zhang, S. Yang, P. Tan, L. Chang, Y. Yuan, and H. Wang, "Joint hand detection and rotation estimation using cnn". *IEEE transactions on image processing*, vol. 27, no 4, pp. 1888-1900, 2017.
- [34] Y. Li, X. Wang, W. Liu, and B. Feng, "Deep attention network for joint hand gesture localization and recognition using static rgb-d images". *Information Sciences*, vol. 441, pp. 66–78, 2018.
- [35] S. Shinde, A. Kothari, V. Gupta, "YOLO based human action recognition and localization". *Procedia Comput. Sci.* 2018, vol. 133, pp. 831–838, 2018.
- [36] J. M. Yu, W. Zhang, "Face mask wearing detection algorithm based on improved YOLO-v4". *Sensors*, 2021, vol. 21, no 9, p. 3263, 2021.
- [37] R. Shukla, A. K. Mahapatra, J. S. P Peter, "Social distancing tracker using yolo v5". *Turkish Journal of Physiotherapy and Rehabilitation*, vol. 32 (2), pp. 1785–1793, 2021.
- [38] G. Yang, W. Feng, J. Jin, Q. Lei, X. Li, G. Gui, and W. Wang, "Face mask recognition system with YOLOV5 based on image recognition".

- In 2020 IEEE 6th International Conference on Computer and Communications. ICC), pp. 1398-1404, 2020.
- [39] L. Wang, W.Q. Yan, "Tree leaves detection based on deep learning". In : Geometry and Vision: First International Symposium, ISGV 2021, Auckland, New Zealand, January 28-29, 2021, Revised Selected Papers 1. Springer International Publishing, pp. 26-38, 2021.
- [40] Z. Q. Zhao, P. Zheng, S. T. Xu, and X. Wu, "Object detection with deep learning: A review". IEEE transactions on neural networks and learning systems, vol. 30, no. 11, pp. 3212-3232, 2019.
- [41] G. Marin, F. Dominio, and P. Zanuttigh, "Hand gesture recognition with jointly calibrated leap motion and depth sensor". Multimedia Tools and Applications, vol. 75, no. 22, pp. 14991-15015, 2016.
- [42] A. Memo, and P. Zanuttigh, "Head-mounted gesture controlled interface for human-computer interaction". Multimedia Tools and Applications, vol 77, pp. 27-53, 2018.
- [43] P. Bao, A. I. Maqueda, C. R. Del-Blanco, and N. Garcíá, "Tiny hand gesture recognition without localization via a deep convolutional network". IEEE Transactions on Consumer Electronics, vol. 63, no. 3, pp. 251-257, 2017.
- [44] S. Biasotti, M. Tarini, and A. Giachetti, "Exploiting Silhouette Descriptors and Synthetic Data for Hand Gesture Recognition".
- [45] J. Redmon, A. Farhadi, "Yolov3: An incremental improvement". arXiv preprint arXiv:1804.02767. 2018.
- [46] J. Redmon, A. Farhadi, "YOLO9000: Better, Faster, Stronger". In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017, pp. 7263-7271. 41.
- [47] A. Bochkovskiy, C. Y. Wang, and H. Y. M Liao, "Yolov4: Optimal speed and accuracy of object detection". arXiv preprint arXiv:2004.10934,2020.
- [48] Z. Jiang, L. Zhao, S. Li, and Y. Jia, "Real-time object detection method for embedded devices". In computer vision and pattern recognition. 2020.
- [49] Bochkovskiy, A. Darknet: Open source neural networks in python. 2020. Available online: <https://github.com/AlexeyAB/darknet>.
- [50] C. -Y. Wang, H. -Y. M. Liao, Y.-H. Wu, et al, "CSPNet: A New Backbone that can Enhance Learning Capability of CNN". In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops, pp. 390-391.
- [51] K. Wang, J. H. Liew, Y. Zou, D. Zhou, and J. Feng, "PANet: Few-Shot Image Semantic Segmentation with Prototype Alignment". In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea, 27 October–2 November 2019; pp. 9197–9206.
- [52] R. Xu, H. Lin, K. Lu, L. Cao, and Y. Liu, "A Forest Fire Detection System Based on Ensemble Learning". Forests, vol. 12, no. 2, pp. 217, 2021.
- [53] B. Jabir, N. Falih, and K. Rahmani, "Accuracy and Efficiency Comparison of Object Detection Open-Source Models". International Journal of Online & Biomedical Engineering, vol. 17, no. 5, 2021.

Optimized Image Authentication Algorithm using Redundant Wavelet Transform Based Sift Descriptors and Complex Zernike Moments

Pooja Vijayakumaran Kallath, Kondaka Lakshmisudha

Department of Information Technology
SIES Graduate School of Technology
Nerul, Navi Mumbai, Maharashtra, India

Abstract—Due to the advanced multimedia editing tools and supported by sophisticated hardware, creating image/video manipulations for malicious purposes is increasing which is almost impossible to detect manually. Moreover, to conceal the traces, different post-processing operations are performed. Therefore, authenticity is a growing concern and important for identifying original and forged images. One of the popular image manipulations is copy-move forgery in which one or more regions in the image are duplicated to create a malicious effect within an image. The work in this article presents redundant wavelet transform based complex Zernike moment and Scale Invariant Feature Transform (SIFT) keypoint matching technique for copy-move image forgery operation detection. SIFT is robust against scale and rotation that works on identifying similarity using exhaustive search of SIFT features. After extracting SIFT keypoint features agglomerative hierarchical clustering is employed for grouping and key point matching operation is performed. Finally, block matching operation is evaluated and forged regions in the manipulated images are marked and displayed. This work also presents optimized SIFT key-point feature computations resulting in lower computation time, often one of the requirements in real time deployment. The proposed algorithm performance is evaluated using Precision, Recall, F-Measure and average detection accuracy on popular and publicly available MICC-220 database. The proposed technique demonstrates improved speed-up and detection rate compared to existing approaches.

Keywords—*Forgery detection; scale invariant feature transform; key point operation; block matching; agglomerative hierarchical clustering*

I. INTRODUCTION

In today's world, the widespread use of digital content has led to the manipulations to spread the information with malicious goals and even change people's opinion widely. This forgery creation demands immediate need of digital image authentication and to validate the trustworthiness of source images [1]. Image forensics is the application of domain knowledge to understand image/video content in legal matters. The Scientific Working Group on Digital Evidence (SWGDE) lists best practices that are required to reliably preserve image integrity. Image authentication solution is divided into two types: Active techniques and passive methods (blind). Passive or blind image forensics analyzes the image using statistics and semantics to identify manipulation and

without considering embedded data in an image. Passive forgery detection algorithms can be categorized as camera based, pixel-based, geometric-based, physics-based, JPEG artifact-based, and statistical-based techniques.

One of the popular image manipulations is copy-move forgery in which one or more regions in the image are duplicated to create a malicious effect within an image. Copy-move forgery detection algorithm performance is evaluated using either at an image-level or pixel-level. Image/video authentication is employed to verify trustworthiness of the content. Manual authentication mechanism requires huge efforts and labor and sometimes it is error prone. Therefore, automatic image authentication algorithms are required for improved detection rate. As digital media technologies allow for image or video alteration and counterfeiting, having accurate images can be crucial evidence in court. Deep fakes, photo or video editing and many other practices can misrepresent an event that is critical to an investigation. A thorough examination and trustworthy data on the original image is essential. To address this need, image forensics offers a careful review of relevant photos to provide an unbiased assessment of the evidence [2].

Forensic algorithms analyze digital images to determine the accuracy and trustworthiness of the information in several different circumstances. To determine if an image represents a circumstance or location accurately, experts may assess color level anomalies and landmarks to identify the image authenticity. Experts can apply deconvolution to the file to identify a person in the photo that includes a blurred or otherwise obscured identity. Digital images play an important role in people's lives such as news, print media and courtroom evidence [3].

This work presents a technique to verify the credibility and integrity of the images based on a redundant wavelet transform based complex Zernike moment and Scale Invariant Feature Transform (SIFT) key point matching technique for copy-move image forgery operation detection. SIFT is robust against scale and rotation that works on identifying similarity using exhaustive search of SIFT features. However, SIFT is computationally expensive. Hence, this article presents an optimized approach for image authentication. After extracting SIFT key point features agglomerative hierarchical clustering is employed for grouping and key point matching operation is

performed. Finally, block matching operation is evaluated and forged regions in the manipulated images are marked and displayed.

Major contributions of this article are summarized as:

- 1) Complex Zernike moment feature extraction for improved detection rate.
- 2) SIFT key-point feature extraction process is computationally expensive. So, an optimized approach for SIFT key-point descriptor extraction is employed which reduces computation time significantly.
- 3) The proposed optimized image authentication algorithm demonstrates improved speed-up and detection rate.

The paper is organized as follows. Existing literature and findings are presented in Section II. Section III explains proposed algorithms and various feature extraction steps in detail. Simulation experiments and discussions are presented in Section IV. Finally, Section V discusses the results and presents robustness analysis while Section VI concludes the article.

II. LITERATURE REVIEW

This section describes the most crucial and recent works in the field of copy-move forgery detection. In [4], a new methodology based on SIFT is described that helps us to know and understand if a copy-move attack has taken place. In addition to that, it helps to retain the geometric transformation that performs the process of cloning. The proposed method also estimates the geometric transformation values with improved reliability and detects multiple forgery operations.

The work in [5] uses a high-level algorithm to recognize a unique model of using Hu's invariant moments and Log-polar transformations to minimize feature space dimensionality to one feature per block and parallelly recognizing the CMF among almost the same objects in an image. The qualitative and quantitative outputs obtained demonstrate the effectiveness of the algorithm.

In [6], first the input image is decomposed using steerable-pyramid transform (SPT) and grey level co-occurrence matrix (GLCM) descriptors are extracted from each orientation. These features are then utilized to train optimized support vector machines (OSVM) which also acts as a classifier. GLCM features are extracted from each block. A novel method for forgery detection is illustrated which uses a new integrated version of key point-based counterfeit detection method and SLIC super pixel segmentation algorithm for forgery detection [7]. The proposed algorithm generates super-pixels with the help of Simple Linear Iterative Clustering (SLIC). An algorithm to detect copy-move image forgery in images is developed in [8]. Discrete wavelet transform (DWT) is applied on the given image to be decomposed into four parts LL, LH, HL, and HH. Since the LL section contains most of the information, SIFT is particularly applied on the LL part only to extract the most important features. This helps in finding the best descriptor vector of these key features and furthermore, it helps in identifying the similarities between test and train images. Authors in [9] proposed a dual level keypoint based forgery detection approach. First, SIFT is used

to detect keypoints in smooth regions and then, BRIEF and FAST descriptors are combined to detect the critical key points from missing areas. Keypoint matching is performed using the generalized nearest neighbor. Finally, morphological image processing operations are utilized to locate forged areas.

The model in [10] first extracts the textural features from the input image. Robust keypoints are extracted using SIFT from these textual images and keypoint matching is done to conclude if the image is forged or not. From that, suspicious regions are determined. The localization of forged pixels is realized via a Ciratefi based approach. Local tetra pattern (LTrP) based feature extraction is developed in [11] for forgery detection and localization. Firstly, the input image is divided into non-overlapping blocks and then from each individual block LTrP descriptors are extracted.

The novel algorithm proposed in [12] utilizes a fusion of the SIFT and local binary pattern (LBP). Consideration of texture features around the key points detected by the SIFT algorithm can be effective to reduce the incorrect matches and improve the accuracy of copy-move forgery detection. In [13], a stationary wavelet transform (SWT) is applied over the input image to acquire a low approximation band, and crucial features are extracted from the band using block-based discrete cosine transformation (DCT) and singular value decomposition (SVD). The literature survey revealed that main feature extraction is based on Zernike moments and SIFT keypoint feature mapping producing acceptable accuracy rate with higher computation time. Hence, this article proposes use of complex Zernike moments and optimized SIFT keypoint extraction and mapping thereby generating improved detection rate at the same time with lower computation time.

III. PROPOSED METHOD FOR SPEECH DYSFLUENCIES CLASSIFICATION

This work presents redundant wavelet transform based Zernike moment and optimized scale invariant feature transform (SIFT) key point matching technique for copy-move image forgery operation detection. SIFT is robust against scale and rotation that works on identifying similarity using exhaustive search of SIFT features. After extracting SIFT key point features agglomerative hierarchical clustering is employed for grouping and key point matching operation is performed. Finally, block matching operation is evaluated and forged regions in the manipulated images are marked and displayed. The proposed algorithm performance is evaluated using Precision, Recall, F-Measure and detection accuracy and compared with existing copy-move forgery detection approaches.

A. Architecture of MFCC and FBE based Dysfluencies Classification

The detailed steps followed during the development of advanced image authentication algorithm using redundant wavelet transform based SIFT Descriptors and Zernike Moments are outlined below. Fig. 1 illustrates the architecture of the proposed optimized SIFT keypoint feature based algorithm.

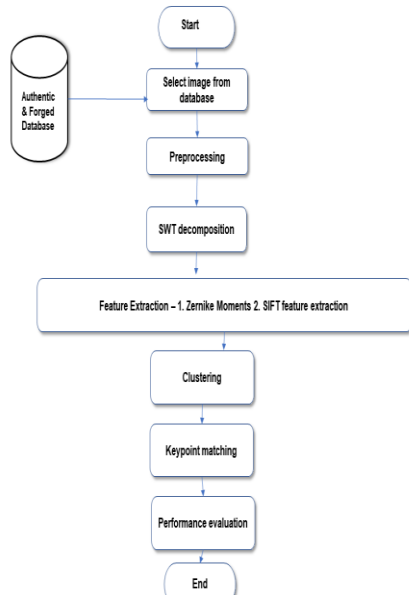


Fig. 1. Workflow diagram of system architecture.

1) *Database collection*: Collect the image authentication database. The dataset consists of Original and forged images using copy-move forgery operation. Total forged images are represented as $F_T = 1, 2, \dots, N$, whereas total original images with $A_T = 1, 2, \dots, M$.

2) *Redundant wavelet transform decomposition*: Apply 4-level redundant wavelet transform decomposition on the original and forged images as follow:

$$\begin{aligned}
 A_{j,k_1,k_2} &= \sum_{n_1} \sum_{n_2} h_0^{2j}(n_1 - 2k_1) h_0^{2j}(n_2 - 2k_2) A_{j-1,n_1,n_2} \\
 D_{j,k_1,k_2}^1 &= \sum_{n_1} \sum_{n_2} h_0^{2j}(n_1 - 2k_1) g_0^{2j}(n_2 - 2k_2) A_{j-1,n_1,n_2} \\
 D_{j,k_1,k_2}^2 &= \sum_{n_1} \sum_{n_2} g_0^{2j}(n_1 - 2k_1) h_0^{2j}(n_2 - 2k_2) A_{j-1,n_1,n_2} \\
 D_{j,k_1,k_2}^3 &= \sum_{n_1} \sum_{n_2} g_0^{2j}(n_1 - 2k_1) g_0^{2j}(n_2 - 2k_2) A_{j-1,n_1,n_2}
 \end{aligned} \quad (1)$$

Where $A_{j,k_1,k_2}, D_{j,k_1,k_2}^1, D_{j,k_1,k_2}^2, D_{j,k_1,k_2}^3$ are the low-frequency sub-band (LL), high-frequency sub-band (LH), high-frequency sub-band (HL) and diagonal (HH) sub-band of the redundant wavelet transform respectively [14].

3) *Compute Zernike moments*: Compute the discrete form of the Zernike moments for an image with the size $N \times N$ as follows:

$$\begin{aligned}
 Z_{n,m} &= \frac{n+1}{\lambda N} \sum_{c=0}^{N-1} \sum_{r=0}^{N-1} f(c,r) V_{n,m}(c,r) \\
 &= \frac{n+1}{\lambda N} \sum_{c=0}^{N-1} \sum_{r=0}^{N-1} f(c,r) R_{n,m}(\rho cr) e^{jm\theta cr}
 \end{aligned} \quad (2)$$

where, n = Zernike moments order, m is the repetition, λN is a normalization factor, centroid for angle is shown as θ , c indicates centroid column number and r denotes row number. In the above equation, $(c; r)$ represents the coordination of the image while $f(c; r)$ is the image function. The translation vector is represented as V , j is the index of input ROI [15].

4) *Scale-invariant feature transform (SIFT) feature extraction*: SIFT descriptors are rotation and scaling invariant and are computed using the following steps.

5) *Scale-space extrema detection*: A Gaussian function $G(x; y; \sigma)$ and input image, $I(x; y)$ are convolved to obtain scale-spaced image $L(x; y; \sigma)$:

$$\begin{aligned}
 L(x, y, \sigma) &= G(x, y, \sigma) * I(x, y) \\
 G(x, y, \sigma) &= \frac{1}{2\pi r^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (3)
 \end{aligned}$$

6) *Keypoint localization*: The keypoint selection from extrema is obtained by rejecting the points along image edges or having low contrast values and expressed as:

$$D(x) = D + \frac{\partial D^T}{\partial x} + \frac{1}{2} x^T + \frac{\partial^2 D}{\partial x^2} x \quad (4)$$

7) *Keypoint descriptor generation*: The SIFT keypoints with a histogram array of 4 X 4 and number of orientation bins as 8, produces 128-dimensional descriptor.

8) *Clustering*: SIFT keypoint descriptors are extracted from approximate sub-band of the SWT decomposition and are grouped using agglomerative hierarchical clustering. Typical linkage methods utilized for the clustering approach are median, centroid or ward.

9) *Keypoint matching*: SIFT keypoint matching is a principal step in which firstly keypoints from the input image are read and compared with the keypoints of images.

10) *Block matching and marking forged region*: Finally, block matching operation is evaluated and forged region in the manipulated images are marked and displayed.

11) *Performance evaluation*: The proposed algorithm performance is evaluated using Precision, Recall, F-Measure and detection accuracy.

$$\text{Precision} = \text{True-Positives} / (\text{True-Positives} + \text{False-Positives})$$

$$\text{Recall} = \text{True-Positives} / (\text{True-Positives} + \text{False-Negatives})$$

$$\text{F-Measure} = (2 * \text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall})$$

$$\text{Detection rate} = (\text{TruePositives} + \text{True-Negatives}) / (\text{True-Positives} + \text{True-Negatives} + \text{False-Positives} + \text{False-Negatives})$$

IV. SIMULATION RESULTS

This section presents a variety of experiments conducted using the proposed algorithm. To assess the effectiveness of the developed technique for image authentication, MICC-F220 database is employed. This popular and publicly available image dataset has forged and authentic natural 220

images: 110 are manipulated images and 110 are authentic. The pixel resolution of these images varies from 722 X 480 to 800 X 600 pixels and almost 1.2% of the whole image is covered by the forged patch. The simulation experiments were conducted on Intel® Core™ i-54, 9400F CPU @ 2.90 GHz processor with 8GB RAM running MATLAB 2021a.

As the input image pixel resolution varies, we have resized the input image size to 300 X 300. In order to extract finer details four-level redundant wavelet transform decomposition is utilized. Zernike moments are extracted with the settings as order of 4 and repetition of moments 2. The proposed optimized forgery detection algorithm utilizes Harris threshold of 5 and number of windows as 2. We have set the sigma value as 9 and 3 neighbourhood computation. The forged region detection is performed individually on each color component and finally combined to generate final output.

The performance of the proposed optimized image forensic algorithm is described by the computation of precision, recall, F-score and detection rates. Precision and recall are two important parameters employed to identify forgery detection algorithm accuracy. F-Measure represents a measure of test accuracy and is the harmonic mean of precision and recall. The results reported in this paper are an average of 10 trials. The proposed image authentication algorithm performance is evaluated using Precision, Recall, F-Measure and detection accuracy.

1) *Experimental results and analysis using standard and proposed optimized forgery detection approach:* Firstly, the effectiveness and the accuracy evaluation of the proposed technique, we present experimental results and analysis using conventional SIFT key-point feature extraction approach and proposed optimized image forensic algorithm. For fair comparison for both experiments MICC-F220 DATABASE is employed. The detection performance is measured on irregular shaped duplicated regions as it influences the overall detection rate. The first set of experiments is carried out without any post-processing operation on the original and manipulated image.

Fig. 2 to 5 illustrate experiment results for copy- move forgery detection using the proposed algorithm for four sample images from the database. In each figure, the first row (from left to right) depicts the original image, SLIC image and SIFTS keypoint feature extraction. Second row (from left to right) shows labelled feature points, merged points after morphological operation, and detected forged regions are shown. As it is evident from these figures that the proposed optimized technique effectively detects and locates forged areas. Moreover, it convincingly visualizes and detects irregular shape manipulated areas and objects.

As illustrated in these Fig. 2 to 5, first the input image is decomposed using a four-level redundant wavelet transform. The four-level decomposition primarily assists in extracting fine details from the forged image. This in turn enhances the detection and localization estimation. As shown in the second figure of the first row, the simple linear iterative clustering based super pixel segmentation (SLIC) algorithm convincingly improves input segmentation into different

regions. The SIFT key-point feature extraction process and final detection result is dependent on the SLIC generated output.

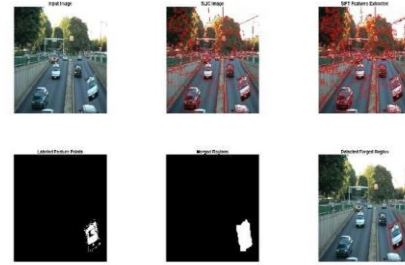


Fig. 2. Copy- move forgery detection results using the proposed algorithm obtained using image sample 1.

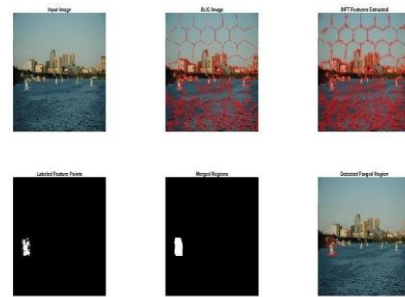


Fig. 3. Copy- move forgery detection results using the proposed algorithm obtained using image sample 2.



Fig. 4. Copy- move forgery detection results using the proposed algorithm obtained using image sample 3.



Fig. 5. Copy- move forgery detection results using the proposed algorithm obtained using image sample 4.

V. DISCUSSIONS

This work presents comparison between conventional SIFT key-point feature extraction and modified feature extraction technique in terms of various performance evaluation parameters. In the conventional SIFT key-point feature extraction approaches typically Harris threshold is set at 10 and number of windows for computation is 4. The computation time for detection of SIFT key-point features and further processing like localization is considered as one of the major concerns in case of real-time detection. Table I shows the evaluation results using these parameters for precision, recall, F-measure and execution time for five images.

As it can be observed from Tables I and II that the execution time of the optimized image forensic algorithm is significantly reduced compared to the conventional approach. The speedup obtained in this case is almost twice. Additionally, precision and F-measure parameters are slightly improved in the proposed technique. This experimental evaluation signifies improved performance of the optimized algorithm over the conventional SIFT key-point feature extraction technique. The average precision and F-Measure value is 0.9806 and 0.9904 for the conventional SIFT feature extraction approach whereas it is 0.987 and 0.9938 for the proposed method. Fig. 6 and 7 depict the effect of Harris threshold and number of window computation on the average detection accuracy respectively.

A. Robustness Analysis

To conceal the traces of copy-move forgery operation the malicious user usually performs various post-processing operations. The prime intent of these post-processing operations is to make forged areas hard to detect and localize. Therefore, it is imperative to evaluate the robustness of the proposed image forensic technique against different post-processing operations. In this study, three different post-processing operations and its robustness is analyzed: rotation attack, scaling attack, and contrast enhancement.

Fig. 8, 9 and 10 depict average accuracy, precision and F1-Measure computed using rotation attack, scaling attack and contrast enhancement respectively. It is evident from these figures that the proposed optimized forgery detection technique has the capability to perform better even when the forged image suffers from rotation, scaling and contrast enhancement post processing operations. The average accuracy is above 96% in all post-processing attacks proving enhanced detection and localization performance of the proposed algorithm. Finally, the proposed optimized image forensic approach is compared with existing state-of-the-art techniques. As shown in Table III, the proposed algorithm outperforms existing approaches in terms of precision, recall and average accuracy rate.

TABLE I. PERFORMANCE EVALUATION USING CONVENTIONAL SIFT KEY-POINT FEATURE EXTRACTION APPROACH

Image number	Execution time (s)	Precision	Recall	F-Measure
Image 1	0.4835	0.9838	1	0.9918
Image 2	0.4928	0.9729	1	0.9887
Image 3	0.4734	0.9837	1	0.9899
Image 4	0.4674	0.9823	1	0.9903
Image 5	0.4985	0.9802	1	0.9914
Average	0.4735	0.9806	1	0.9904

TABLE II. PERFORMANCE EVALUATION USING OPTIMIZED SIFT KEY POINT FEATURE EXTRACTION APPROACH

Image number	Execution time (s)	Precision	Recall	F-Measure
Image 1	0.2383	0.9913	1	0.9918
Image 2	0.2467	0.9838	1	0.9928
Image 3	0.2414	0.9867	1	0.9939
Image 4	0.2725	0.9848	1	0.9956
Image 5	0.2469	0.9884	1	0.9926
Average	0.2492	0.987	1	0.9938

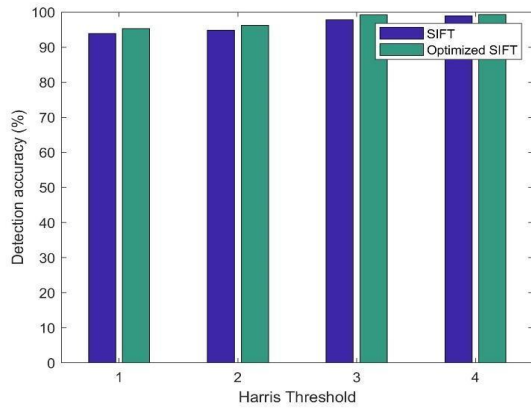


Fig. 6. Detection rate using Harris threshold using conventional SIFT and proposed optimized SIFT algorithm.

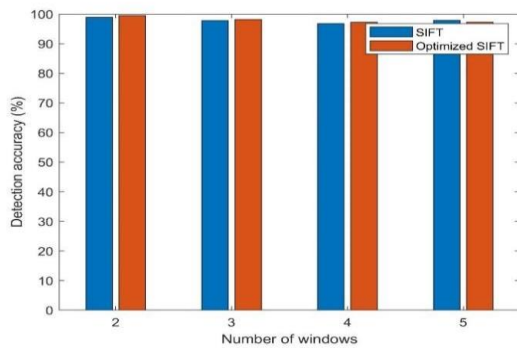


Fig. 7. Effect of number of windows on the detection rate using conventional SIFT and proposed optimized SIFT algorithm.

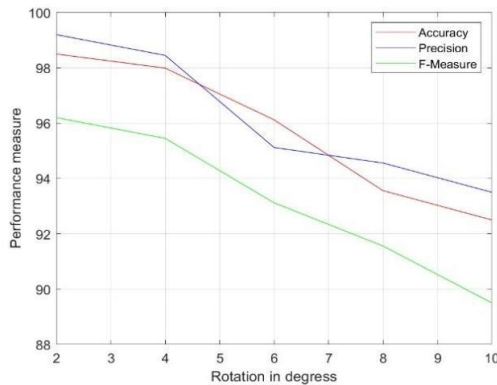


Fig. 8. Average accuracy, precision, and F1-measure computation using rotation attack with 2, 4, 8 and 10 degree rotation.

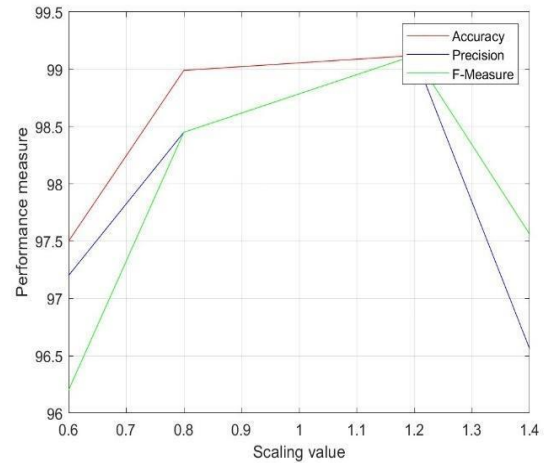


Fig. 9. Average accuracy, precision, and F1-measure evaluation using scaling attack with 0.6, 0.8, 1.2 and 1.4 scaling value.

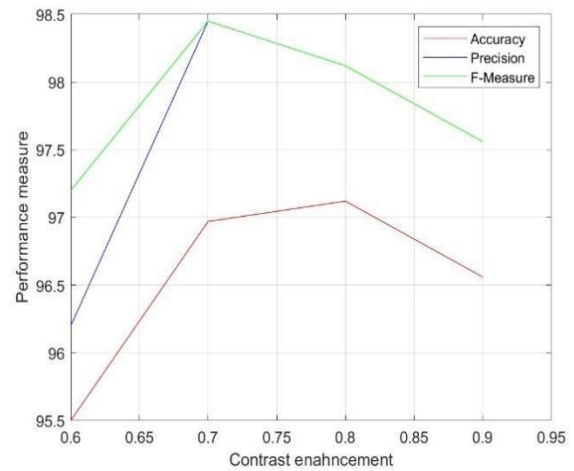


Fig. 10. Average accuracy, precision, and F1-measure values obtained using different contrast enhancement parameters.

Overall recent proposed methods based on SIFT and feature level image authentication have higher computational complexity although, there are performance improvement techniques using different feature fusion resulting in higher dimensional feature vector. To overcome these issues this paper presents optimized SIFT based image authentication solution. The proposed approach lowers the computational cost and speed up is achieved in this study.

TABLE III. COMPARISON OF PROPOSED ALGORITHM

Method	Precision	Recall	F-measure	Accuracy
[15]	96	89	100	94
[16]	91.39	95.83	86.55	90.95
[17]	94	95.83	92	93.87
[18]	93.9	90.44	87.75	92.20
[19]	92.8	88.7	91.15	94.45
Proposed	98.53	100	99.13	98.87

VI. CONCLUSIONS

In today's world, widespread use of multimedia contents has given rise to the manipulations of multimedia content with malicious purposes demanding the necessity of robust authentication techniques. This work presents redundant wavelet transform based complex Zernike moment and optimized scale invariant feature transform (SIFT) keypoint matching technique for copy-move image forgery detection. SIFT is robust against scale and rotation that works on identifying similarity using exhaustive search of SIFT features. After extracting SIFT keypoint features agglomerative hierarchical clustering is employed for grouping and key point matching operation is performed. Finally, block matching operation is evaluated and forged regions in the manipulated images are marked and displayed. The proposed algorithm is evaluated using the popular and publicly available database MICC-F220. As observed, the proposed optimized SIFT algorithm achieves speedup of almost twice over the standard SIFT technique. Experimental evaluation illustrates improved performance of the proposed technique as compared to other similar methods available in the literature. In future, the work can further be explored using statistical descriptors to improve image authentication accuracy.

REFERENCES

- [1] Kaur, G., Singh, N. & Kumar, M., "Image forgery techniques: a review", *Artificial Intelligence Review*, Volume - 56, pp. 1577–1625, 2023.
- [2] Warif, N.B.A., Idris, M.Y.I., Wahab, A.W.A. *et al.* "A comprehensive evaluation procedure for copy-move forgery detection methods: results from a systematic review", *Multimedia Tools and Applications*, Volume - 81, pp. 15171–15203, 2022.
- [3] Simranjot Kaur, Rajneesh Rani, Ritu Garg, and Nonita Sharma, "State-of-the-art techniques for passive image forgery detection: a brief review", *International Journal of Electronic Security and Digital Forensics*, Volume -14, Issue - 5, pp. 456-473, 2022.
- [4] Amerini, L. Ballan, R. Caldelli, A. Del Bimbo and G. Serra, "A SIFT-Based Forensic Method for Copy–Move Attack Detection and Transformation Recovery," in *IEEE Transactions on Information Forensics and Security*, vol. 6, no. 3, pp. 1099-1110, Sept. 2011.
- [5] K. Tejas, C. Swathi and M. Rajesh Kumar, "Copy Move Forgery using Hu's Invariant Moments and Log-Polar Transformations," *2018 3rd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT)*, Bangalore, India, 2018, pp. 1229-1233.
- [6] S B G Tilak Babu, Ch Srinivasa Rao, "An optimized technique for copy–move forgery localization using statistical features", *ICT Express*, Volume 8, Issue 2, Pages 244-249, 2022.
- [7] Rathi, K., Singh, P., "Copy Move Forgery Detection by Using Integration of SLIC and SIFT", In: Jeena Jacob, I., Gonzalez-Longatt, F.M., Kolandapalayam Shanmugam, S., Izonin, I. (eds) *Expert Clouds and Applications. Lecture Notes in Networks and Systems*, vol 209, 2022.
- [8] M. F. Hashmi, A. R. Hambarde and A. G. Keskar, "Copy move forgery detection using DWT and SIFT features," *2013 13th International Conference on Intelligent Systems Design and Applications*, Salangor, Malaysia, 2013, pp. 188-193, doi: 10.1109/ISDA.2013.6920733.
- [9] Fatima, B., Ghafoor, A., Ali, S.S. *et al.* "FAST, BRIEF and SIFT based image copy-move forgery detection technique" *Multimedia Tools and Applications*, Volume - 81, Pages- 43805–43819, 2022.
- [10] Tahaoglu, G., Ulutas, G., Ustubioglu, B. *et al.* "Ciratefi based copy move forgery detection on digital images", *Multimedia Tools and Applications Volume- 81*, pages -22867–22902, 2022.
- [11] Ganguly, S., Mandal, S., Malakar, S. *et al.* "Copy-move forgery detection using local tetra pattern based texture descriptor", *Multimedia Tools and Applications*, 2023, <https://doi.org/10.1007/s11042-022-14287-9>.
- [12] Marziye Shahrokhi, Alireza Akoushideh and Asadollah Shahbahrami, "Image Copy–Move Forgery Detection Using Combination of Scale-Invariant Feature Transform and Local Binary Pattern Features", *International Journal of Image and Graphics*, Vol. 22, No. 05, pages- 2250048, 2022.
- [13] Kumar, S., Mukherjee, S. & Pal, A.K., "An improved reduced feature-based copy-move forgery detection technique", *Multimedia Tools and Applications*, Volume - 82, pages - 1431–1456, 2023.
- [14] Mahmoud, Khaled & Husien, Arwa. (2016). Copy-Move Forgery Detection Using Zernike and Pseudo Zernike Moments. *The International Arab Journal of Information Technology (IAJIT)*. 13. 930-937.
- [15] Goel, N., Kaur, S., & Bala, R. (2021). Dual branch convolutional neural network for copy move forgery detection. *IET Image Processing*, 15(3), 656-665.
- [16] Warif, N. B. A., Wahab, A. W. A., Idris, M. Y. I., Salleh, R., & Othman, F. (2017). SIFT-symmetry: a robust detection method for copy-move forgery with reflection attack. *Journal of Visual Communication and Image Representation*, 46, 219-232.
- [17] Hashmi, M. F., Hambarde, A. R., & Keskar, A. G. (2013, December). Copy move forgery detection using DWT and SIFT features. In *2013 13th international conference on intelligent systems design and applications* (pp. 188-193). IEEE.
- [18] Richa Singh ; Sandeep Verma ; Suman Avdhesh Yadav ; S. Vikram Singh. Copy-move Forgery Detection using SIFT and DWT detection Techniques. 2022 3rd International Conference on Intelligent Engineering and Management (ICIEM), 10.1109/ICIEM54221.2022.9853192.
- [19] S B G Tilak Babu, Ch Srinivasa Rao, "An optimized technique for copy–move forgery localization using statistical features" *ICT Express* 8 (2022) 244–249.

Anchor-free Proposal Generation Network for Efficient Object Detection

Hoanh Nguyen

Faculty of Electrical Engineering Technology, Industrial University of Ho Chi Minh City, Ho Chi Minh City, Vietnam

Abstract—Deep learning object detection methods are usually based on anchor-free or anchor-based scheme for extracting object proposals and one-stage or two-stage structure for producing final predictions. As each scheme or structure has its own strength and weakness, combining their strength in a unified framework is an interesting research topic. However, this topic has not attracted much attention in recent years. This paper presents a two-stage object detection method that utilizes an anchor-free scheme for generating object proposals in the initial stage. For proposal generation, this paper employs an efficient anchor-free network for predicting object corners and assigns object proposals based on detected corners. For object prediction, an efficient detection network is designed to enhance both detection accuracy and speed. The detection network includes a lightweight binary classification subnetwork for removing most false positive object candidates and a light-head detection subnetwork for generating final predictions. Experimental results on the MS-COCO dataset demonstrate that the proposed method outperforms both anchor-free and two-stage object detection baselines in terms of detection performance.

Keywords—Object detection; deep learning; convolutional neural network; proposal generation network

I. INTRODUCTION

Object detection has seen significant advancements in recent years thanks to deep learning, particularly convolutional neural networks (CNN). According to the way of generating object proposals from input images, Current object detection techniques can be divided into two categories: anchor-based and anchor-free object detection methods. Anchor-based approaches consider each object as a rectangular bounding box on feature map. Features inside the bounding box are extracted and inputted into either a proposal generation network to generate proposals or a detection network to generate final outputs. To address the issue of scale variation, anchor-based methods define multiple bounding boxes with varying sizes and aspect ratios, enabling the network to detect objects of diverse sizes and proportions. The sizes and aspect ratios defined in anchor-based object detection approaches vary depending on the specific object and structure. Anchor-based schemes are dominant in early deep learning object detection methods since they are easy to implement and facilitate the learning process. However, object detection methods based on anchor-based scheme face another problem as they cannot detect objects with rare sizes/aspect ratios due to the limitation of anchor box sizes and ratios.

On the other hand, anchor-free object detection approaches examine points (i.e., anchor points or keypoints) on feature

map to predict objects. These approaches can be categorized into two categories: anchor points object detection approaches and keypoints object detection approaches. While object detection methods based on anchor points classify each point on feature map into object/background classes and predict the distances from the positive points to object borders, keypoint-based object detection methods predict keypoints such as corner points or center points on feature map and group valid points to form objects. Compared to anchor points object detection methods, keypoints object detection methods usually have a more complicated structure and achieve better detection performance. However, they need an optimal grouping algorithm so that the network can efficiently group valid keypoints to form objects.

Alternatively, according to the learning process, Current object detection techniques can be divided into two categories: one-stage and two-stage object detection methods. One-stage object detection techniques directly use the detection network on input feature maps to generate final outputs, whereas two-stage object detection methods generate object proposals in the first stage, followed by the use of the detection network in the second stage to produce the final predictions. Since one-stage methods eliminate proposal generation process, they obtain fast processing speed. However, detection accuracy is typically improved through the use of two-stage methods [1], [2].

In recent years, many object detection structures followed the above schemes or structures have obtained great achievements [3], [4], [5]. In general, each of the above schemes or structures has its own strengths and weaknesses. Thus, combining the strength of these schemes or structures in a unified framework is an interesting topic. However, this topic has received limited attention from the academic community in recent years. This paper presents a novel object detection framework that combines the benefits of both an anchor-free approach and a two-stage structure. The proposed method uses an anchor-free object detection scheme to generate object proposals in the initial stage. To efficiently generate final predictions in the subsequent stage, an efficient detection network is designed. The efficient detection network includes a lightweight binary classification subnet and a light-head detection subnet. Experimental results on the MS-COCO dataset prove the effectiveness of the proposed method.

The structure of the article is presented as follows. Section II introduces recent related works. Section III provides details on the design of the proposed method. Section IV presents the experiments and results achieved by the method. Section V provides conclusions.

II. LITERATURE REVIEW

A. Anchor-based Object Detection Methods

Anchor-based object detection methods depict each object as an anchor bounding box on feature maps. To address the scale variation issue, these object detectors establish multiple anchor boxes at each location on the feature map. Each anchor box is linked to a scale and aspect ratio. In Faster R-CNN [3], three aspect ratios (128^2 , 256^2 , 512^2) and three scales (1:1, 1:2, 2:1) are employed in the definition of anchor boxes, yielding nine anchor boxes at each position on feature map. In FPN [4], since region proposal network is applied on the feature pyramid, anchor boxes at each spatial location of a feature level are defined using one scale and three aspect ratios. Anchor-based scheme has been employed in many deep object detection frameworks [5], [6], [7], [31]. However, anchor box scales and aspect ratios must be meticulously designed for the specific domain to ensure the detection network attains optimal detection performance. To eliminate problems caused by anchor box settings, various methods propose to replace the manual design of the anchor boxes by a deep network so that the shape of the anchors is automatically learned during training. For this purpose, MetaAnchor [8] developed a generator for anchor functions that maps a given prior bounding box to its corresponding anchor function. The anchor function generator is formed by a simple network and computed from customized prior bounding boxes, and thus it can be inserted into any deep learning object detection methods for joint optimization. The results show that MetaAnchor is more robust than manual design of anchor settings as it can detect objects with rare shapes. However, MetaAnchor obtains minor improvements on two-stage object detectors. Moreover, it requires customized prior bounding boxes to be chosen by handcraft and more computation for extra network. In [9], a novel anchor box optimization method is proposed. The training process employs the optimization technique based on localization loss to automatically learn the anchor shapes. In addition, soft assignment and online clustering scheme are introduced to warm up the anchor shapes. Recently, Sparse R-CNN [10] represented object candidates by a limited set of bounding boxes that can be learned. These learnable bounding boxes represent the statistics of potential object locations within the training set. The back propagation algorithm will update the parameters of these adjustable bounding boxes during training. By eliminating the hand-designed anchor boxes, Sparse R-CNN strikes a favorable balance between accuracy, runtime, and training convergence performance.

B. Anchor-free Object Detection Methods

Anchor-free object detection methods employ points (i.e., anchor points or keypoints) for predicting objects. For this purpose, CornerNet [11] suggests detecting objects based on their top-left and bottom-right corners. For detecting corners, CornerNet employs a corner prediction network that includes a corner pooling layer for producing corner proposals, a heatmap generation layer for generating corner heatmaps, an offset generation layer for predicting corner offsets, and a network for calculating embeddings which are used to group valid corner points to form objects. Based on CornerNet, CenterNet [12] introduced an extra keypoint (i.e., center point) for predicting objects. A center pooling layer is also designed to enrich center

and corner information, which improves the detection performance of CenterNet. Zhou et al. [35] utilized a single point at the center of the bounding box to represent objects, eliminating the need for the grouping stage in keypoint detectors like CornerNet and CenterNet. Peaks in the heatmaps generated by a keypoint estimation network are used to predict object center. In [13], representation of objects in input images is achieved through the use of a set of sample points that adaptively position themselves over the object. The sample points are learned through both object localization and recognition loss. Based on predicted sample points, converting functions are designed to form object bounding boxes. In [14], an object prediction mechanism utilizing a star-shaped bounding box is designed. The star-shaped bounding box employs features from nine fixed sampling points with deformable convolution [15] to represent a bounding box. This new bounding box design can incorporate both the geometry of the bounding box and its surrounding context, crucial for encoding any misalignment between the predicted and actual bounding box.

An alternative approach is to use each point on the feature map for object prediction. For this purpose, FCOS [16] employs a fully convolutional network for classifying each location on feature map. For each positive point, FCOS predicts the distances from the location to the four sides of the bounding box. FCOS incorporates a center-ness branch to down-weight the scores of low-quality predicted bounding boxes generated by locations far from the center of an object, in order to remove them. Similar to FCOS, FoveaBox [17] presents an anchor-free framework that predicts category-sensitive semantic maps for the presence of objects and generates category-agnostic bounding boxes for each potential object location. FoveaBox defines positive and negative training samples based on the fovea area, which is the center of the visual field with the highest resolution. Different from FCOS and FoveaBox, Zhu et al. [34] introduced a new feature selective anchor-free module (FSAF), which takes pixels on feature pyramid as inputs and directly feeds these pixels into two convolutional networks: a classification network for predicting class scores for each pixel and a regression network for producing offsets encoding the distances from the current pixel location to the top, left, bottom, and right boundaries of the target bounding box. Recently, SAPD [18] introduced an optimal training approach using two softening optimization techniques, soft-weighted anchor points and soft-selected pyramid levels. The soft-weighted anchor points technique adjusts the contribution of anchor points on the same pyramid level to the network loss based on their geometry relative to the instance box, while the soft-selected pyramid levels technique learns the participation level of each pyramid. The results show that SAPD balances speed and accuracy effectively.

The above methods have their own strengths and weaknesses. Specifically, due to the limitation of anchor box sizes and ratios as well as the variability of object sizes, anchor-based object detection methods have limitations in detecting objects with various shapes. On the other hand, anchor-free object detection approaches have limitations in determining geometric relations between an object and nearby contextual information. This paper focuses on exploiting and

combining the strengths of the above methods. Based on that, a model is designed that can achieve better accuracy and execution speed.

III. METHODOLOGY

The proposed structure is described in this section. Initially, an evaluation of anchor-based and anchor-free methods for generating object proposals is conducted. Then, the specifics of each module in the proposed structure are depicted in subsequent subsections.

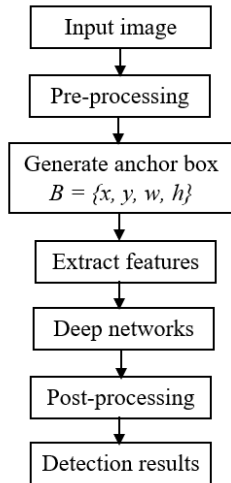


Fig. 1. The flowchart of anchor-based object detection methods.

C. Object Proposal Generation Methods

As shown in Fig. 1, anchor-based object detection methods first depict each object as an anchor box $B = \{x, y, w, h\}$ on feature maps, where the center point is represented by the coordinates (x, y) , and the width and height of the object bounding box are (w, h) . Then, the features within anchor box B are extracted and inputted into a deep network to generate object proposals (for two-stage approaches) or direct final predictions (for one-stage approaches). Anchor-based object detection techniques define a collection of anchor boxes at each position in a feature map in the input image to accommodate objects of varying size, position, and scale, enabling the model to detect all of them. In Faster R-CNN [3], the center of each anchor box corresponds to the center of the sliding window, and each is paired with a specific scale and aspect ratio. Faster R-CNN uses three aspect ratios (i.e., 1:1, 1:2, 2:1) and three scales (i.e., 128^2 , 256^2 , 512^2), which results in 9 anchor boxes at each position on feature map. Recently, the FPN [4] has established a single scale and three aspect ratios for each anchor location, utilizing the feature pyramid. When using anchor-based object detection methods, it is important to carefully design the number and shape of anchor boxes. Too few anchor boxes or inappropriate anchor shapes may be insufficient to cover a large range of objects in various sizes and ratios, which reduces the performance of proposal generation network or detection network. The top portion of Fig. 2 shows some examples where Faster R-CNN with anchor-based scheme faces difficulty in predicting objects with rare shapes since there are no defined ratios or scales that fit these objects.



Fig. 2. Results of detection on the validation set of MS-COCO dataset. Top: Faster R-CNN with anchor-based scheme. Bottom: CornerNet with anchor-free scheme.

Alternatively, anchor-free methods use keypoints or anchor-points to depict an object. Methods based on keypoints first predict the locations of keypoints like corner or center points, and then use these keypoints to form an object bounding box by a grouping algorithm. On the contrary, anchor-point based methods first categorize each point on the feature map and then estimate the distances from the potential point to the four edges of the actual bounding box to produce object proposals. Since anchor-free methods eliminate all problems related to anchor box settings, they have a better ability of detecting objects, especially objects with rare shapes, which improves the recall rate. The bottom portion of Fig. 2 shows some detection results of the CornerNet framework [11]. As shown, CornerNet with anchor-free scheme obtains better detection results compared with Faster R-CNN with anchor-based scheme. However, anchor-free methods face another problem of forming an object candidate based on keypoints. Take CornerNet as an example, CornerNet determines an embedding vector for every identified corner, then groups the corners to create the object bounding box based on the distances between the embeddings. Due to the significant number of false positives produced by the corner detection network, determining an embedding vector for each detected corner may result in many false positive outcomes. As seen in Fig. 3, CornerNet generates some incorrect corner pairs because of similar appearance leading to similar embeddings.

Based on the above analysis, this paper introduces an efficient object detection structure that inherits the merits of anchor-free scheme for producing object proposals and two-stage structure for generating predictions. Based on anchor-free scheme, this paper designs an efficient two-stage object detection approach that eliminates the grouping stage, which hinders the detection performance of anchor-free object detection pipelines. The details of the proposed method are outlined in subsequent sections.



Fig. 3. Results of CornerNet on the validation set of MS-COCO dataset showed some false corner pairs generated due to similarities in embeddings.

D. Overview of the Proposed Approach

The structure of the proposed method is shown in Fig. 4. It integrates an anchor-free approach and a two-stage structure into a single object detection framework. The first stage generates object proposals, and the second stage produces predictions. The proposal generation network in the first stage is based on CornerNet [11]. Specifically, CornerNet employs input feature maps to predict top-left and bottom-right corner keypoints of the bounding box for objects. Based on the corner keypoints, object proposals are formed according to the corner locations and the corresponding classifying scores. Since there are many false positive object proposals generated by the first stage, an efficient detection network is designed in the second stage. Specifically, a lightweight classification subnet is first designed to remove most false positive object candidates. A detection subnet with light-head structure is then adopted to produce prediction results based on remaining object candidates. With the anchor-free scheme for proposal generation in the initial stage and an efficient detection structure in the following stage, the proposed approach integrates the merits of anchor-free scheme into a two-stage structure. The details of each module are depicted in the subsequent sections.

E. CornerNet as Object Proposal Generation

To obtain high recall rate for generating object proposals from input images, especially for objects with various shapes, this paper adopts CornerNet [11] as object proposal generation network. CornerNet identifies an object through two crucial keypoints - the top-left corner keypoint and the bottom-right corner keypoint. The structure of CornerNet, as depicted in Fig. 5, involves the utilization of the Hourglass model [19] to extract feature maps from input images. The Hourglass network initially processes input features through convolution and max pooling layers to reduce the resolution and then uses up-sampling, convolution layers, and skip layers to increase the resolution back to its original state. The Hourglass architecture combines both global and local features into a single structure. As in [11], this paper employs the Hourglass architecture with

two Hourglass modules for extracting input features. The final layer of the Hourglass network is utilized for further prediction by using two prediction branches with identical structures. These prediction branches, based on the last feature map produced by Hourglass, detect the top-left and bottom-right corners. Each branch generates C channel heatmaps, where C represents the number of object categories. Each channel is a binary map that shows the locations of corners for each class. To refine the corner locations, each branch predicts offset values. A corner pooling module, consisting of two 3×3 convolution layers followed by a corner pooling layer, is utilized in each branch to pool features from the Hourglass network. These features are then fed into a 3×3 convolution layer for projection. Finally, the output features are used to produce heatmaps and offsets through a series of 3×3 and 1×1 convolution layers. It should be noted that since this paper adopts CornerNet for predicting corners, the embedding prediction branch in the original CornerNet is removed, thus reducing the computation of the proposed object proposal generation network.

After getting proposal corners through the proposed CornerNet, this paper extracts K top-left and K bottom-right corners from the heatmaps generated by CornerNet ($K = 50$ in this paper). Then, each pair of top-left and bottom-right corners belonging to the same class, where the coordinates of top-left corner are smaller than that of bottom-right corner is used to define an object proposal. By using corner points to define object candidates, the proposed object proposal generation method can detect more objects, especially objects with arbitrary size, which are usually missed by anchor-based proposal generation method. As a result, the recall rate is significantly improved. However, defining object proposals based on this scheme leads to many false positive proposals as the corner keypoints of two different objects of the same class may define an object proposal (as shown in Fig. 3). To eliminate most false proposal candidates, this paper designs an efficient detection network which is elaborated in the next subsection.

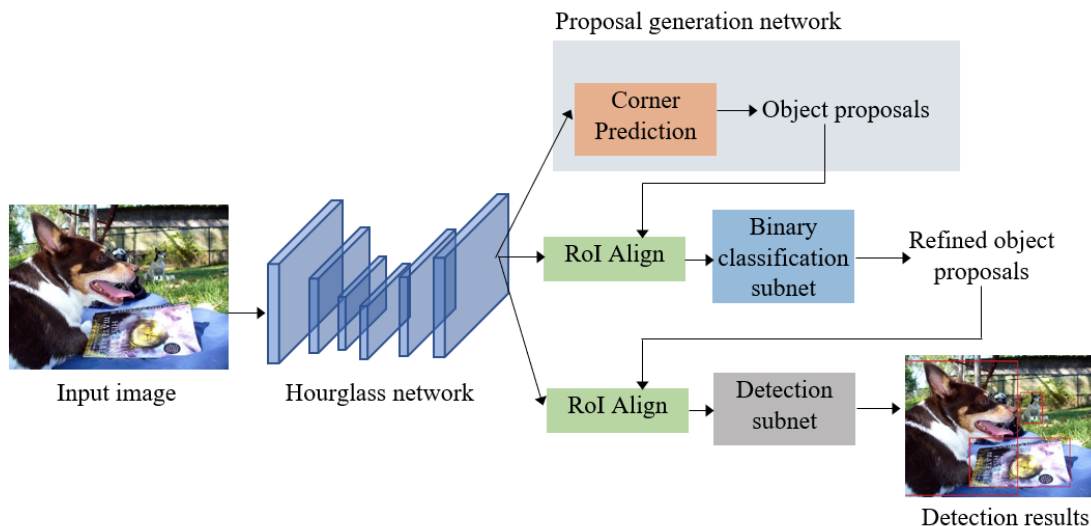


Fig. 4. The overall structure of the proposed model.

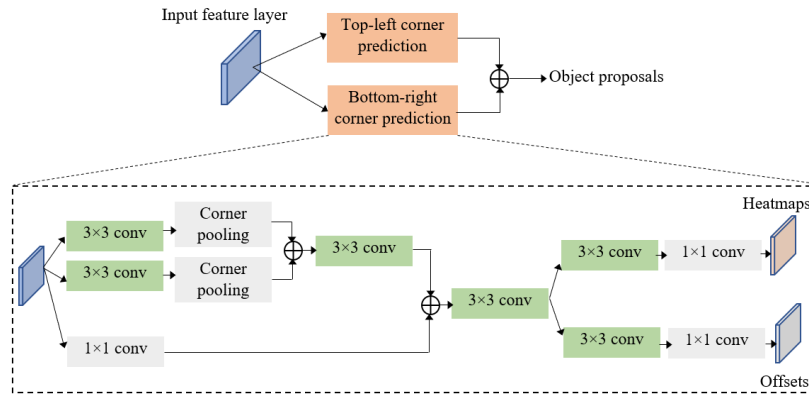


Fig. 5. The architecture of CornerNet used in this paper.

F. Detection Network

Since this paper employs CornerNet for generating object proposals, many false positive proposal candidates are produced. As a result, applying a heavy detection network for predicting objects based on a large number of object candidates is not efficient since it requires a huge amount of computational budget. In the paper, a high-performance detection network is proposed. The structure of the proposed network is depicted in Fig. 6. First, this paper employs a lightweight binary classification subnet to eliminate most of false positive proposal candidates. The lightweight binary classification subnet starts by applying a convolution layer to the final feature layer of the backbone to create a thin feature map with 32 channels. The RoIAlign layer [20] then creates the proposal feature map using the thin feature map and proposals from the proposed CornerNet. At the end of the lightweight binary classification subnet, a convolution layer followed by an average pooling layer computes the classification score for each proposal. To address the issue of imbalanced training samples, this paper implements a variation of focal loss [6] in the training process, which is illustrated as follows:

$$L_{class1} = -\frac{1}{N} \sum_i L_i \quad (1)$$

$$L_i = \begin{cases} (1 - p_i)^\beta \log(p_i), & \text{if } IoU_i \geq T \\ p_i^\beta \log(1 - p_i), & \text{otherwise} \end{cases} \quad (2)$$

where N represents the number of positive samples, p_i is the objectness score of i -th proposal, IoU_i is the maximum IoU value between i -th proposal and all ground truth boxes. T and β are set at 0.7 and 2, respectively, in this paper.

Next, since the lightweight binary classification subnet effectively eliminates most false positive proposal candidates, a light-head detection subnet is utilized to generate final predictions from the remaining proposals. This paper designs a light-head structure, similar to Light-Head R-CNN [21], in the final detection subnet for both classification and bounding box regression. The light-head detection subnet can improve computational speed without compromising the detection performance. As illustrated in Fig. 6, the first step is to apply a large separable convolution layer to the final feature layer of the backbone to improve these features while simultaneously decreasing the number of channels. Compared with 1×1 convolution, large separable convolution is more efficient as it produces thin output features with more semantic information. Then, PSRoI Align [22] is adopted to produce fixed size features (i.e., $7 \times 7 \times 10$) for remaining proposal candidates based on feature map generated by the large separable convolution layer. Here, the PSRoI Align is utilized as it reduces the number of channels in the output features. Finally, this paper uses a single 1024-dimensional fully connected layer followed by two parallel fully connected layers to produce the final classification and regression results.

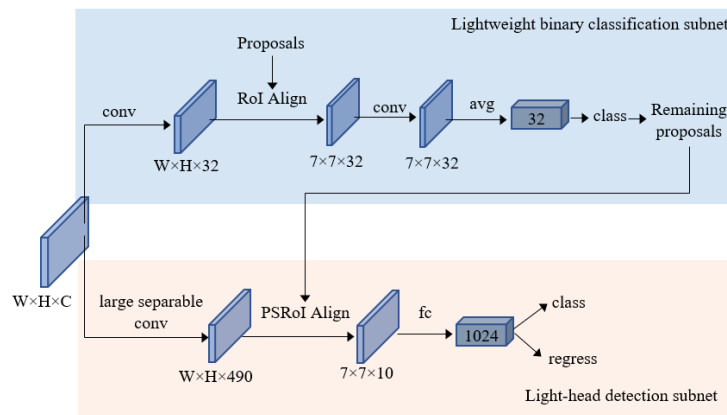


Fig. 6. The structure of the proposed detection network.

By designing an efficient detection network with two prediction steps, most of false positive object candidates are removed by the classifier, and remaining object candidates are efficiently predicted by the detection subnet. The detection subnet with light-head structure can enhance the detection speed without compromising the detection performance. The experimental results demonstrate that this design is more efficient than using a heavy detection network directly as the detection network.

IV. RESULTS

A. Implementation Details

This paper uses the MS-COCO dataset [23] to evaluate the proposed model, which contains 80 object categories. The images in the dataset are divided into three sets, with 80K images for training, 40K images for validation, and 20K images for testing. In accordance with the standard protocol [4], [6], this paper trains the proposed model using all images in the training set and 35K images from the validation set. To evaluate the detection performance, the paper reports the results on the test-dev set, which are submitted to an external evaluation server.

For evaluation metrics, this paper follows metrics defined in the MS-COCO dataset for evaluating object detection tasks. To be more specific, this paper uses the average precision (AP), AP_{50} , AP_{75} , AP_S , AP_M , and AP_L as evaluation metrics. AP is the average precision over 80 categories under multiple IoU values (i.e., 0.5:0.05:0.95). AP is considered the key metric when assessing object detection techniques on the MS-COCO dataset. AP_{50} and AP_{75} are AP computed at a specific IoU threshold value. AP_S , AP_M , and AP_L are AP computed based on object sizes (AP_S for objects with area $< 32^2$, AP_M for objects with $32^2 < \text{area} < 96^2$, and AP_L for objects with area $> 96^2$).

The proposed network is designed based on Pytorch [24] and open-source object detection toolbox mmdetection [25]. The object proposal generation network is adopted from CornerNet [11], where the stacked Hourglass networks with 104 layers and the corner detection network are trained on the MS-COCO dataset. As in [11], the input size of the network is set to 511×511 during the training process. The proposed network is trained end-to-end with the full training loss as follow:

$$L = L_{corner} + L_{offset1} + L_{class1} + L_{class2} + L_{offset2} \quad (3)$$

where L_{corner} and $L_{offset1}$ are the variant of focal loss and the smooth L_1 loss, respectively; L_{class1} is the variant of focal loss for training the binary classification subnet; L_{class2} and $L_{offset2}$ are the cross-entropy loss and the smooth L_1 loss, respectively, for training the light-head detection network. For computationally efficient reasons, this paper uses Adam optimizer [26] to optimize the training loss. The Adam optimizer is a widely used optimization algorithm in the deep learning domain and is straightforward to implement. The

proposed network is trained for 100K iterations on Nvidia RTX 3070 GPU.

During the inference process, this paper adopts a confidence threshold of 0.3 to remove false positive proposal candidates by the binary classification subnet. In addition, Soft-NMS [27] is employed after the light-head detection network to eliminate redundant boxes, and the top 150 scoring boxes are selected for evaluation.

B. Detection Results on the MS COCO Dataset

The detection accuracy of the proposed method is shown in Table I alongside recent state-of-the-art object detection methods, both anchor-based and anchor-free pipelines, on the MS-COCO test-dev set. The results in the table demonstrate that the proposed model significantly outperforms the CornerNet baseline model [11]. To be more specific, the proposed model improves AP, AP_{50} , AP_{75} , AP_S , AP_M , and AP_L by 6.8, 11.0, 10.7, 7.6, 8.7, and 8.1 points, respectively, compared with CornerNet on the same backbone network and input size. The results show that the combining of the binary classifier and the light-head detection network are very efficient for removing false positive object candidates and predicting remaining object candidates. Compared with Faster R-CNN using ResNet-101 backbone, which is a popular two-stage anchor-based object detection framework, the proposed model also achieves significantly higher detection accuracy. The proposed network improves AP, AP_{50} , AP_{75} , AP_S , AP_M , and AP_L by 11.2, 8.3, 14.9, 8.5, 12.5, and 14.2 points, respectively, compared with Faster R-CNN using ResNet-101 backbone. The results demonstrate that using an anchor-free approach to generate object proposals and designing a suitable detection network can enhance object detection performance. As seen in Table I, the proposed model also outperforms both anchor-based and anchor-free models while using lower input resolution. Moreover, as AP_S , AP_M , and AP_L denote AP for predicting objects at different sizes, the results in Table I reveal that the proposed model has improved performance in detecting medium and large objects compared to small ones. To be more specific, compared with CornerNet, the proposed model improves AP_S , AP_M , and AP_L by 7.6, 8.7, and 8.1 points, respectively. This result shows that the proposed network has difficulty in detecting small objects since predicting small objects requires richer semantic information features, which can be achieved by employing feature pyramids.

Table II shows the inference speed of the proposed model and several efficient methods on the MS-COCO dataset. All methods are implemented on Nvidia RTX 3070 GPU. As demonstrated in Table II, the proposed model obtains 2.8 fps on the MS-COCO dataset with input resolution 511×511 , which is comparable to the speed of the baseline CornerNet. The results shown in Table II also indicate that Faster R-CNN and FCOS have better speed, however, the proposed method achieves a better detection accuracy as shown in Table I. This demonstrates that the proposed method strikes a good balance between inference speed and detection accuracy.

TABLE I. EVALUATION OF THE PROPOSED METHOD AGAINST RECENT OBJECT DETECTION TECHNIQUES ON THE MS-COCO TEST-DEV SET IN TERMS OF DETECTION ACCURACY

Method	Backbone network	Input resolution	AP	AP ₅₀	AP ₇₅	AP _S	AP _M	AP _L
Anchor-based methods								
Faster R-CNN [3]	ResNet-101	600×1000	36.2	59.1	39.0	18.2	39.0	48.2
Light-Head R-CNN [21]	ShuffleNetV2	800×1200	23.7					
ThunderNet [28]	SNet535	320×320	28.1	46.2	29.6	-	-	-
RetinaNet [6]	ResNet-101	800×1200	39.1	59.1	42.3	21.8	42.7	50.2
Mask R-CNN [20]	ResNext-101	800×1200	39.8	62.3	43.4	22.1	43.2	51.2
TridentDet [29]	ResNet-101	800×1200	42.7	63.6	46.5	23.9	46.6	56.6
YOLOv4 [30]	Darknet-53	608×608	43.5	65.7	47.3	26.7	46.7	53.3
Cascade R-CNN [32]	ResNet-101	800×1200	42.8	62.1	46.3	23.7	45.5	55.2
FFAD [33]	ResNet-101	800×1333	44.1	62.2	47.9	27.4	47.6	56.7
Anchor-free methods								
FCOS [16]	ResNext-101	800×1024	44.7	64.1	48.4	27.6	47.5	55.6
CornerNet [11]	Hourglass-104	511×511	40.6	56.4	43.2	19.1	42.8	54.3
CenterNet [12]	Hourglass-104	511×511	44.9	62.4	48.1	25.6	47.4	57.4
FoveaBox [17]	ResNext-101	800×1024	42.1	61.9	45.2	24.9	46.8	55.6
SAPD [18]	ResNext-101	800×1024	45.4	65.6	48.9	27.3	48.7	56.8
Proposed method	Hourglass-104	511×511	47.4	67.4	53.9	26.7	51.5	62.4

TABLE II. THE INFERENCE SPEED OF THE PROPOSED MODEL AND SEVERAL EFFICIENT METHODS ON THE MS-COCO DATASET

Method	Backbone network	Input resolution	FPS
Faster R-CNN [3]	ResNext-101	600×1000	4.2
CornerNet [11]	Hourglass-104	511×511	3.0
CenterNet [12]	Hourglass-104	511×511	2.6
FCOS [16]	ResNext-101	800×1024	4.1
Proposed method	Hourglass-104	511×511	2.8

C. Ablation Study on Detection Network

To assess the efficacy of the proposed detection network for anchor-free proposal generation scheme, this paper examines the detection performance of several structures on the MS-COCO validation set. First, the original R-CNN detection head [3] is applied on the last backbone feature layer to produce final predictions based on object proposals produced by the proposed CornerNet. The proposed CornerNet extracts 50 top-left corner points and 50 bottom-right corner points based on the heatmaps to form object proposals. The R-CNN architecture consists of two fully connected layers with 4096 neurons each, featuring ReLU activations, followed by two additional fully connected layers for performing classification and regression tasks. Second, the proposed lightweight binary classification subnet is applied to remove false positive proposals. This binary classification takes $7 \times 7 \times 32$ feature maps generated by a RoIAlign layer as inputs. Remaining proposals are fed into the original R-CNN detection head to output final predictions. Finally, the proposed light-head detection subnet is applied on the last backbone feature layer to produce predictions without the lightweight binary

classification subnet. It should be noted that a large separable convolution layer is employed to reduce the feature map channels to 490 before feeding to the light-head detection subnet. For all experiments, this paper employs the same input resolution at 511×511 for fair comparison. Table III illustrates the detection performance of the proposed structures. As demonstrated in Table III, directly applying the R-CNN subnet on object proposals generated by the proposed CornerNet does not improve AP as many false positive proposals hinder the classification ability of R-CNN. When employing the binary classification subnet before the R-CNN subnet, the detection performance is improved. However, the detection speed is reduced as this structure uses a heavy detection head for prediction. On the other hand, using the light-head detection subnet after the binary classification subnet improves both the detection performance and speed. The result shows that the proposed detection network with a binary classification subnet and a light-head detection subnet obtains the best trade-off between detection accuracy and speed.

TABLE III. EVALUATION OF VARIOUS DESIGNS ON THE MS-COCO VALIDATION SET IN TERMS OF THEIR DETECTION ABILITY

Method	Input resolution	AP	FPS
CornerNet [11]	511×511	41.0	3.0
CornerNet + R-CNN	511×511	40.8	2.1
CornerNet + binary classifier + R-CNN	511×511	46.4	1.6
CornerNet + light-head detection subnet	511×511	41.1	3.8
CornerNet + binary classifier + light-head detection subnet	511×511	46.8	2.8

V. CONCLUSIONS

This paper presents a new object detection framework that combines the benefits of anchor-free and two-stage approaches. In the first stage, an anchor-free scheme is designed to generate object proposals. In the second stage, an efficient detection network comprised of a lightweight binary classification subnetwork for reducing false positive object proposals and a light-head detection subnetwork for final predictions is utilized. The proposed model was tested on the MS-COCO dataset and achieved the best balance between speed and accuracy compared to state-of-the-art anchor-based and anchor-free object detection methods. Specifically, the proposed model obtains 47.4 of AP on the MS-COCO test-dev set, which surpasses both anchor-free and one-stage model baselines. The focus of this study was on efficiency, and thus techniques for improving accuracy, such as combining different network layers or using multi-layer predictions, were not explored. As a result, the model struggles with detecting small objects. Future work will focus on improving the detection of small objects by replacing the backbone network with a feature pyramid network.

REFERENCES

- [1] Huang, Jonathan, Vivek Rathod, Chen Sun, Menglong Zhu, Anoop Korattikara, Alireza Fathi, Ian Fischer et al. "Speed/accuracy trade-offs for modern convolutional object detectors." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 7310-7311. 2017.
- [2] Lu, Xin, Quanquan Li, Buyu Li, and Junjie Yan. "Mimicdet: Bridging the gap between one-stage and two-stage object detection." In *European Conference on Computer Vision*, pp. 541-557. Springer, Cham, 2020.
- [3] Ren, Shaoqing, Kaiming He, Ross Girshick, and Jian Sun. "Faster r-cnn: Towards real-time object detection with region proposal networks." *Advances in neural information processing systems* 28 (2015).
- [4] Lin, Tsung-Yi, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. "Feature pyramid networks for object detection." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2117-2125. 2017.
- [5] Liu, Wei, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C. Berg. "Ssd: Single shot multibox detector." In *European conference on computer vision*, pp. 21-37. Springer, Cham, 2016.
- [6] Lin, Tsung-Yi, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. "Focal loss for dense object detection." In *Proceedings of the IEEE international conference on computer vision*, pp. 2980-2988. 2017.
- [7] Fu, Cheng-Yang, Wei Liu, Ananth Ranga, Amrith Tyagi, and Alexander C. Berg. "Dssd: Deconvolutional single shot detector." *arXiv preprint arXiv:1701.06659* (2017).
- [8] Yang, Tong, Xiangyu Zhang, Zeming Li, Wenqiang Zhang, and Jian Sun. "Metaanchor: Learning to detect objects with customized anchors." *Advances in neural information processing systems* 31 (2018).
- [9] Zhong, Yuanyi, Jianfeng Wang, Jian Peng, and Lei Zhang. "Anchor box optimization for object detection." In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 1286-1294. 2020.
- [10] Sun, Peize, Rufeng Zhang, Yi Jiang, Tao Kong, Chenfeng Xu, Wei Zhan, Masayoshi Tomizuka et al. "Sparse r-cnn: End-to-end object detection with learnable proposals." In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 14454-14463. 2021.
- [11] Law, Hei, and Jia Deng. "Cornersnet: Detecting objects as paired keypoints." In *Proceedings of the European conference on computer vision (ECCV)*, pp. 734-750. 2018.
- [12] Duan, Kaiwen, Song Bai, Lingxi Xie, Honggang Qi, Qingming Huang, and Qi Tian. "Centernet: Keypoint triplets for object detection." In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 6569-6578. 2019.
- [13] Yang, Ze, Shaohui Liu, Han Hu, Liwei Wang, and Stephen Lin. "Reppoints: Point set representation for object detection." In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 9657-9666. 2019.
- [14] Zhang, Haoyang, Ying Wang, Feras Dayoub, and Niko Sunderhauf. "Varifocalnet: An iou-aware dense object detector." In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8514-8523. 2021.
- [15] Dai, Jifeng, Haozhi Qi, Yuwen Xiong, Yi Li, Guodong Zhang, Han Hu, and Yichen Wei. "Deformable convolutional networks." In *Proceedings of the IEEE international conference on computer vision*, pp. 764-773. 2017.
- [16] Tian, Zhi, Chunhua Shen, Hao Chen, and Tong He. "Fcos: Fully convolutional one-stage object detection." In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 9627-9636. 2019.
- [17] Kong, Tao, Fuchun Sun, Huaping Liu, Yuning Jiang, Lei Li, and Jianbo Shi. "Foveabox: Beyond anchor-based object detection." *IEEE Transactions on Image Processing* 29 (2020): 7389-7398.
- [18] Zhu, Chenchen, Fangyi Chen, Zhiqiang Shen, and Marios Savvides. "Soft anchor-point object detection." In *European conference on computer vision*, pp. 91-107. Springer, Cham, 2020.
- [19] Newell, Alejandro, Kaiyu Yang, and Jia Deng. "Stacked hourglass networks for human pose estimation." In *European conference on computer vision*, pp. 483-499. Springer, Cham, 2016.
- [20] He, Kaiming, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. "Mask r-cnn." In *Proceedings of the IEEE international conference on computer vision*, pp. 2961-2969. 2017.
- [21] Li, Zeming, Chao Peng, Gang Yu, Xiangyu Zhang, Yangdong Deng, and Jian Sun. "Light-head r-cnn: In defense of two-stage object detector." *arXiv preprint arXiv:1711.07264* (2017).
- [22] Dai, Jifeng, Yi Li, Kaiming He, and Jian Sun. "R-fcn: Object detection via region-based fully convolutional networks." *Advances in neural information processing systems* 29 (2016).
- [23] Lin, Tsung-Yi, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. "Microsoft coco: Common objects in context." In *European conference on computer vision*, pp. 740-755. Springer, Cham, 2014.
- [24] Paszke, Adam, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. "Automatic differentiation in pytorch." (2017).
- [25] Chen, Kai, Jiaqi Wang, Jiangmiao Pang, Yuhang Cao, Yu Xiong, Xiaoxiao Li, Shuyang Sun et al. "MMDetection: Open mmlab detection toolbox and benchmark." *arXiv preprint arXiv:1906.07155* (2019).
- [26] Kingma, Diederik P., and Jimmy Ba. "Adam: A method for stochastic optimization." *arXiv preprint arXiv:1412.6980* (2014).
- [27] Bodla, Navaneeth, Bharat Singh, Rama Chellappa, and Larry S. Davis. "Soft-NMS--improving object detection with one line of code." In *Proceedings of the IEEE international conference on computer vision*, pp. 5561-5569. 2017.
- [28] Qin, Zheng, Zeming Li, Zhaoning Zhang, Yiping Bao, Gang Yu, Yuxing Peng, and Jian Sun. "ThunderNet: Towards real-time generic object detection on mobile devices." In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 6718-6727. 2019.
- [29] Li, Yanghao, Yuntao Chen, Naiyan Wang, and Zhaoxiang Zhang. "Scale-aware trident networks for object detection." In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 6054-6063. 2019.
- [30] Bochkovskiy, Alexey, Chien-Yao Wang, and Hong-Yuan Mark Liao. "Yolov4: Optimal speed and accuracy of object detection." *arXiv preprint arXiv:2004.10934* (2020).
- [31] BOURJA, Omar, Hatim DERROUZ, Hamd AIT ABDELALI, Abdelilah MAACH, Rachid OULAD HAJ THAMI, and François BOURZEIX. "Real time vehicle detection, tracking, and inter-vehicle distance

- estimation based on stereovision and deep learning using yolov3." *International Journal of Advanced Computer Science and Applications* 12, no. 8 (2021).
- [32] Cai, Zhaowei, and Nuno Vasconcelos. "Cascade r-cnn: Delving into high quality object detection." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 6154-6162. 2018.
- [33] Yu, Guoyi, You Wu, Jing Xiao, and Yang Cao. "A novel pyramid network with feature fusion and disentanglement for object detection." *Computational Intelligence and Neuroscience* 2021 (2021).
- [34] Zhu, Chenchen, Yihui He, and Marios Savvides. "Feature selective anchor-free module for single-shot object detection." *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2019.
- [35] Zhou, Xingyi, Dequan Wang, and Philipp Krähenbühl. "Objects as points." *arXiv preprint arXiv:1904.07850* (2019).

Detecting Fraud Transaction using Ripper Algorithm Combines with Ensemble Learning Model

Vo Hoang Khang¹, Cao Tung Anh², Nguyen Dinh Thuan³

Faculty of Information Technology, HUTECH University, Ho Chi Minh City, Vietnam^{1,2}

Faculty of Information Systems-University of Information Technology,

Vietnam National University Ho Chi Minh City (VNUHCM - UIT), Ho Chi Minh City, Vietnam³

Abstract—In the context of the 4.0 technology revolution, which develops and applies strongly in many fields, in which the banking sector is considered to be the leading one, the application of algorithms to detect fraud is extremely important. necessary. In recent years, credit card transactions including physical credit card payments and online payments have become increasingly popular in many countries around the world. This convenient payment method attracts more and more criminals, especially credit card fraud. As a result, many banks around the world have developed fraud detection and prevention systems for each credit card transaction. Data mining is one of the techniques applied in these systems. This study uses the Ripper algorithm to detect fraudulent transactions on large data sets, and the results obtained with accuracy, recall, and F1 measure of more than 97%. This research then used the Ripper algorithm combined with Ensemble Learning models to detect fraudulent transactions, the results are more than 99% reliable. Specifically, this model using the Ripper algorithm combined with the Gradient Boosting method has improved the predictive ability and obtained very reliable results. The use of algorithms combined with machine learning models is expected to be a new topic and will be widely applied to banks' or organizations' activities related to e-commerce.

Keywords—Financial fraud; data mining; credit card fraud; transaction; ensemble methods

I. INTRODUCTION

The e-commerce sector is expected to account for 21.8% of total global retail sales by 2024 [1]. As the worldwide e-commerce market share increases risks also increase, trusting becoming a concern as the number of frauds and scams has increased significantly over the past few years [2]. The e-commerce market in 2021 faces a risk of US\$20 billion due to fraudulent transactions [3]. The most common types of fraud include account takeovers, identity theft, covert fraud, and chargeback fraud [3]. These frauds can affect the finances of both consumers and merchants and reduce overall confidence in e-commerce [4].

Data mining techniques have shown promising performance in detecting fraud [5], [6], [7] and rule from large data sets [8]. To solve the problem of transaction fraud in e-commerce, there have been many studies using data mining rules to build laws, thereby detecting transaction fraud. These studies and their results will be introduced in Part II (Related Works).

In many forecasting models, the Repeated Incremental Pruning to Produce Error Reduction (Ripper) algorithm is chosen to apply [9] because this algorithm has high reliability and coverage. Next, this study uses a synchronous learning method in machine learning to combine with Ripper algorithm to improve the reliability and accuracy of fraud detection. The results of study have yielded very positive results. This model has been applied in many other fields but is still quite new in the field of e-commerce in the world as well as in Vietnam.

The article structure consists of five sections. Section I is an overview of the research, Section II introduces the research done related to the article, and Section III is the proposed solutions and techniques to solve the problem dealing with the problem posed. Section IV lists the order of steps to conduct the experiment as well as the software and tools used. Finally, Section V summarizes and concludes with the issues achieved in the study as well as suggestions for future research.

II. RELATED WORKS

Research related to fraudulent transaction detection is numerous and has been conducted in a variety of fields, including finance, e-commerce, credit cards, multi-level commerce, and many others. Data mining techniques used include Support vector machine (SVM), Fuzzy Logic based system (FL), Hidden Markov Model (HMM), Artificial Neural network (ANN), Genetic Algorithm (GA), Bayesian Network (BN), K-Nearest Neighbor Algorithm (KNN), Decision Tree (DT), Logistic Regression (LR), Outlier's Detection and other Classification techniques [6]. Below is a summary of the articles, mining techniques used, and research results.

• Support Vector Machine:

N. K. Gyamfi and J. D. Abdulai used SVM to detect bank fraud [6]. The result shows an improvement of the SVM method by 80% which is considered better performance.

M. Jeragh, M. AlSulaimi combined autoencoders and one class support vectors machine for fraudulent credit card transactions detection [6]. The result of two combined models achieved promising results, especially when evaluated using metrics performance.

K. Poongodi and D. Kumar used SVM with Information Gain Based Classification algorithm to identify fraudulent and legal transactions [10]. The result shows that the accuracy of detection is enhanced in the credit card fraud detection system by using SVM with information gain-based classification.

Y. Kumar, S. Saini, and R. Payal proposed a comparison between three data mining algorithms: Logistic Regression, Random Forest, and Support Vector Machine [11]. The study shows that when comparing the result of three algorithms the random forest algorithm gives better results for the classification of fraud with 81.79% accuracy.

- Artificial Neural Network:

A. A. Rizki, I. Surjandari, and R.A. Wayasti used data mining applications to detect financial fraud in Indonesia's public companies [6]. The finding proved that feature selection helped in increasing the accuracy of the SVM method with 88.37% while ANN gave the most effective accuracy with 90.97%.

I. Sadgali, N. Sael, and F. Benabbou proposed a comparative study using Neural Networks Techniques, Neural Networks (NN), Multilayer Perceptron Layer (MPL), and Convolutional Neural Networks (CNN), for credit card fraud detection [12]. The result shows that CNN has the highest accuracy.

A Sahu, GM Harshvardhan, MK. Gourisaria detected fraudulent transactions by several machine learning models and the artificial neural network [13]. The result shows that all the models have a very high percentage of true negatives due to the heavy count of non-fraud cases.

- Decision Tree:

V. Jain, M. Agrawal, and A. Kumar [14] used three machine learning algorithms, Decision Tree, Random Forest, and XGBoost applied to a data set have the data of 284808 credit cards. The performance of the XGBoost algorithm is found best with the highest accuracy of 99.962 percent. The performance of the Decision Tree is minimum with an accuracy of 99.923 and the performance of the Random Forest algorithm is 99.957 percent in credit card transactions.

J. Soyemi and H. Mudasiru implemented a decision tree algorithm augmented with regression analysis for fraud detection in credit cards [15]. This study was able to achieve its aim by building a machine-learning model that is capable of detecting fraud.

In this article, this study used Ripper algorithm [7] to perform a reliability test when predicting transaction fraud when the transaction volume is increasing and unbalanced, then combine Ripper and Ensemble Learning Methods to create a more reliable predictive model (ROC increase from 88% to 98%).

III. PROPOSED SOLUTIONS AND TECHNIQUES

A. Implementing Data Set

Focusing on the objective of this study, "Aggregated Financial Dataset for Fraud Detection" was selected [16], [17]. The dataset contains approximately 6.3 million data about mobile money transactions and consists of 10 attribute columns, and one target column. This dataset includes a 30-day simulation of real-time transactions and is executed in steps, each mapping to an hour. The original data set includes the following columns:

- step: transaction step
- amount: the number of coins traded
- name_org: sender account name
- old_balance_org: initial balance of the sending account
- new_balance_org: the following balance of the sending account
- name_dest: recipient account name
- old_balance_dest: initial balance of the receiving account
- new_balance_dest: the following balance of the receiving account
- is_fraud: whether the transaction is fraudulent or not
- is_flagged_fraud: whether the transaction is suspected of fraud
- transaction_type: transaction type (5 types)
- isFraud: This is a transaction made by rogue agents inside the simulation. In this particular dataset, the fraudulent behavior of the agents aims to gain profits by controlling, taking over the customer's account, attempting to withdraw money by switching to another account, and then withdrawing money from the system.
- isFlaggedFraud: The business model aims to control large transfers from one account to another and flag illegal attempts. An illegal attempt in this dataset is an attempt to transfer more than 200000 in a single transaction. The isFlaggedFraud column is used when there is an attempt to transfer more than 200000 in a transaction, so basically for isFlaggedFraud, it will be 1 if the transaction amount exceeds 200000. When isFlaggedFraud equals 1, the value of isFraud equals 1, so isFlaggedFraud can be discarded on processing if necessary.

B. Fraudulent Transaction

Transaction fraud is an act of fraud in the process of conducting financial transactions, intending to appropriate property, money, or personal benefits of the offender.

Types of transaction fraud can range from using fake information to open an account or borrow money, to performing invalid transactions, transferring money to the wrong address, or spoofing identity to cheat others.

Transaction fraud not only causes damage to the assets of the parties involved but also affects the trust and reputation of the financial system, reducing the ability to conduct safe and efficient transactions. Therefore, detecting and preventing transaction fraud is very important to protect the safety and reliability of the financial system.

In addition to the transaction that is considered fraudulent as mentioned above, in the data set under consideration, a transaction is suspected to be fraudulent if it violates integrity constraints. There are two types of integrity constraints on this data set, domain-related integrity constraints and inter-

attribute integrity constraints. The first is a domain-related constraint: it ensures that a field's value must meet a specific condition, ensuring that a field's value falls within a specific range. The second constraint is an inter-attribute constraint: a type of constraint in a relational database that ensures that the values in one or more pairs of attributes are logically related to each other. Specifically, if the value of one attribute changes, the values in other attributes that depend on it will also change accordingly to ensure data integrity and accuracy.

Here are some examples of the two integrity constraints mentioned above.

+ R1: For all money transfers and withdrawals, the old balance cannot be equal to 0.

$\forall q: (q.type = TRANSFER \wedge q.oldBalanceOrig = 0) \vee (q.type = CASH_OUT \wedge q.oldBalanceOrig = 0)$

+ R2: After transferring money, the new balance cannot be equal to 0.

$\forall q: (q.type = TRANSFER \wedge q.newBalanceDest \leq 0)$

+ R3: For all money transfers and withdrawals, the old balance is always greater than the amount to be transferred.

$\forall q: (q.type = TRANSFER \wedge q.oldBalanceOrig > q.amount) \vee (q.type = CASH_OUT \wedge q.oldBalanceOrig > q.amount)$

+ R4: In all money transfers, the old balance is always greater than the new balance.

$\forall q: (q.type = TRANSFER \wedge q.oldBalanceOrig > q.newBalanceOrig)$

+ R5: In all money transfers and withdrawals, the old balance minus the new balance is always greater than or equal to the amount to be transferred-withdrawn.

$\forall q: (q.type = TRANSFER \wedge q.oldBalanceOrig - q.newBalanceOrig < q.amount) \vee (q.type = CASH_OUT \wedge q.oldBalanceOrig - q.newBalanceOrig < a.amount)$

+ R6: Recipient did not receive enough transaction funds.

$\forall q: (q.type=TRANSFER \wedge q.oldBalanceDest + q.amount < q.newBalanceDest)$

+ R7: After transferring money, the recipient's old balance must be less than the recipient's new balance.

$\forall q: (q.type = TRANSFER \wedge q.oldBalanceDest > q.newBalanceDest)$

+ R8: Do not transfer and withdraw all funds or do not transfer and withdraw more than the amount available when making the withdrawal.

$\forall q: (q.type = TRANSFER \wedge q.amount \geq q.oldBalanceOrig) \vee (q.type = CASH_OUT \wedge q.amount \geq q.oldBalanceOrig)$

Thus, a transaction in the "Aggregated Financial Dataset for Fraud Detection" dataset that violates one of the abovementioned constraints will be considered a fraudulent transaction.

C. Proposed Algorithm

This study implemented the RIPPER (Repeated Incremental Pruning to Produce Error Reduction) algorithm. In machine learning, iterative incremental pruning to reduce errors is a propositional rule learner proposed by William W. Cohen (1995) [9] as an optimal version encoding of the IREP algorithm. The Ripper algorithm is a machine learning algorithm used to detect fraud in financial transactions. This algorithm works by classifying transactions into two categories: trusted transactions and fraudulent transactions. It then uses machine learning techniques to identify attributes that are important to distinguish between these two types of transactions. RIPPER is a greedy algorithm implemented in the following steps:

+ Step 1: Divide the training data into two parts: the training set and the test set.

+ Step 2: Determine the starting rule set by finding all the rules that can be determined from the initial training set.

+ Step 3: Use the starting rule set to classify the records in the training set.

+ Step 4: Repeat the following steps to find and remove unnecessary rules sequentially:

a) evaluate all rules in the current rule set,

b) remove unnecessary rules,

c) construct a new set of rules from the remaining set of rules.

+ Step 5: Use the final rule set to classify the records in the test set.

The RIPPER algorithm is used to solve classification problems where the goal is to classify objects into known classes. This algorithm has proven to be effective in many practical applications, including handwriting recognition, disease detection, and document classification.

When implemented, this research handled the execution steps of Ripper as follows:

+ Step 1: Read transaction data from the CSV file.

+ Step 2: Convert the column's type to match the requirement.

+ Step 3: Apply Ripper with several RipperK optimization iterations equals 10.

+ Step 4: Train the model.

Ripper has improved IREP with the OPIMIZERULESET process. This process optimized IREP's code and reduced it to a leaner one. Ripper works well on data sets with unbalanced class distribution. In a dataset, if there is a record set in which most of the records belong to a particular class and the remaining records belong to different classes, then the data set is said to have a non-distribution balance between classes. It works well with noisy datasets because it uses a validation set to prevent the model from overfitting.

D. Proposed Models

In machine learning, no one algorithm is always good for all applications and all data sets because machine learning algorithms are usually based on a set of parameters (hyperparameters) or a certain assumption about the distribution data. Therefore, to find the right algorithms for our dataset, we may need a lot of time to test different algorithms and then adjust the algorithm's parameters (tuning hyperparameters) to obtain the highest accuracy.

To increase the accuracy of the dataset is to combine several models. This method is called Ensemble Learning. Different models with different capabilities can best perform different types of work, and when combined properly, form a powerful hybrid model capable of improving performance, and overall performance compared to using the models alone. Ensemble Methods are divided into the following three categories: Boosting, Bagging, and Stacking.

1) *Gradient boosting [18]*: Gradient Boosting is a machine learning algorithm of type Ensemble Learning, it uses a reinforcement process based on sub-models to create a better overall model.

Gradient Boosting handles data classification or regression using a series of sub-models that are built on top of each other, each of which targets the errors of the previous model. This process results in a better overall model, with more accuracy than each sub-module.

Gradient Boosting is commonly used in classification and regression problems, such as handwriting classification, spam email classification, stock price forecasting, and other problems in the field of artificial intelligence.

- Gradient Boosting Model: illustrated in Fig. 1 below.

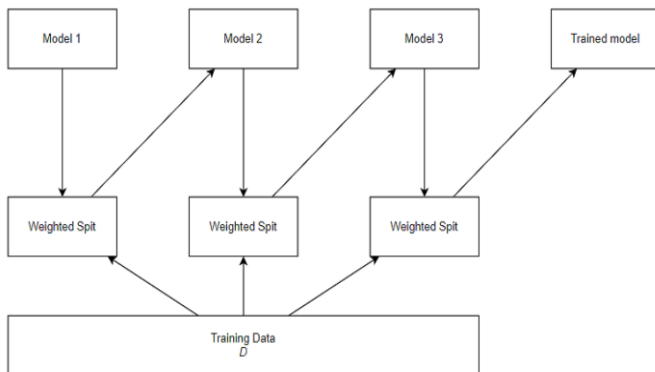


Fig. 1. Gradient boosting operation model.

- Processing steps of the Gradient Boosting Model:
 - + Step 1: Define a lost function.
 - + Step 2: Build a base model to make a prediction.
 - + Step 3: Compute each prediction's residuals (the difference between the predicted and actual values).
 - + Step 4: Fit a new model to predict these residuals.

+ Step 5: Add this new model's predictions to the previous model's predictions.

2) *Bagging [19]*: Bagging (Bootstrapped Aggregating) is a machine learning algorithm of type Ensemble Learning, it uses several sub-models to create a better overall model.

Bagging deals with the classification or regression of data using several sub-models built on the same data set, but each of the sub-models is trained on a bootstrapped dataset (the sampled data set), by starting from an original data set and then choosing a random number of data points to replace the new data set). This process results in a better overall model, with more accuracy than each submodule.

Bagging is commonly used in classification and regression problems, such as handwriting classification, spam email classification, stock price forecasting, and other problems in the field of artificial intelligence.

- Bagging model: illustrated in Fig. 2 below.

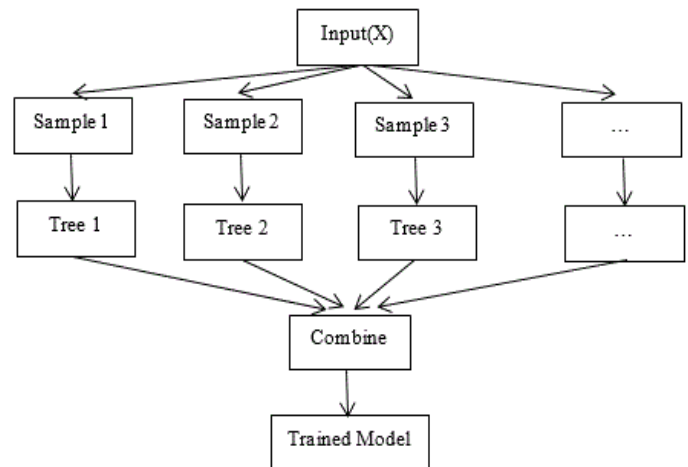


Fig. 2. Bagging operation model.

- Processing steps of the Bagging Model:
 - + Step 1: Build k bootstrap samples from the original dataset.
 - + Step 2: For each bootstrap sample, build a decision tree.
 - + Step 3: Average the predictions of each tree to produce a final model.

3) *Stacking [20]*: Stacking (stacked generalization) is a machine learning algorithm in the Ensemble Learning category. It uses a set of submodules to create a better overall model.

In Stacking, sub-models are trained on an initial dataset and then used to predict values for the new data set. The prediction results of each sub-model are used as input to the overall model. The population model is then trained on the sub-model prediction results to predict the final value.

- Stacking model: illustrated in Fig. 3 below.

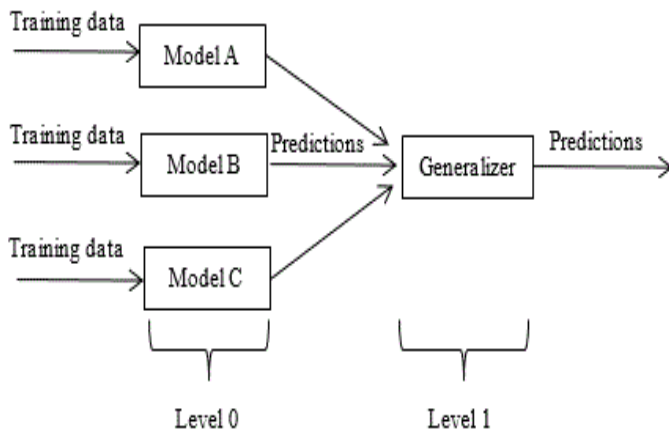


Fig. 3. Stacking operation model.

• Processing steps of the Stacking model:

- + Step 1: Split the dataset into training and testing datasets.
- + Step 2: Train multiple models on the training dataset.
- + Step 3: Make a prediction using these models on the testing dataset.
- + Step 4: Use these predictions as input features to train a new model (the meta-model) on the testing dataset.

E. Combination of Ripper with Ensemble Learning Model

1) Ripper combines with gradient boosting:

- + Step 1: Read transaction data from the CSV file.
- + Step 2: Declare Gradient Boosting algorithm using exponential loss function, 1000 boosting stages, and square error criterion to qualify the quality of every split action.
- + Step 3: Train the model.
- + Step 4: Consecutively apply Ripper to improve model metrics.

2) Ripper combines with bagging:

- + Step 1: Read banking data from the CSV file.
- + Step 2: Initialize the Bagging algorithm with the number of rules generated in the training process equal to 1000.
- + Step 3: Apply Ripper to improve the accuracy.
- + Step 4: Train the model.

3) Ripper combines with stacking:

- + Step 1: Read banking data from the CSV file.
- + Step 2: Initialize Random Forest algorithm with a random state is 10 and 100 trees in the forest.
- + Step 3: Initialize Linear SVC algorithm with a random state equal to 42.
- + Step 4: Initialize Standard Scaler algorithm.

- + Step 5: Initialize the Stacking algorithm with the three algorithms mentioned above and Logistic Regression as the final algorithm.
- + Step 6: Apply Ripper to improve the model's metrics.
- + Step 7: Train the model.

IV. EXPERIMENT

A. Experimental Tools and Software

Programming environment and tools: Python and PyCharm. Python is a programming language widely used in data science and machine learning. PyCharm is a fully integrated development platform (IDE) designed specifically for Python programmers.

We implemented the algorithms on a laptop with the following configuration: CPU: Intel core i7-12700H, GPU: Intel Iris XE Graphics, RAM: 16Gb.

- Implement Ripper algorithm to experiment stability with unbalanced data:
- + Initialize Ripper: `ripper = RIPPER()`

The parameters of `RIPPER()` are shown in Table I.

TABLE I. PARAMETERS OF RIPPER()

Parameters	Description	Default value	Data type
k	The number of loops to optimize	2	int
prune_size	Pruning ratio (prune into two sets, grown set, and prune set)	0.33	float
Random_state	Rules generating seed (rules generated are the same if seeds are the same)	none	int
max_rules	Maximum number of rules	none	int
max_rule_conds	A maximum number of conditions in each rule.	none	int
max_total_conds	A maximum number of conditions in a whole rule set.	none	int
verbosity	Between 0 and 5, print to the terminal rule-generating process. 0 means do not print. The larger the number the more detail: 1: Print the main process. 2: Print the optimization phases. 3: Print the calculations needed to optimize. 4: Print the optimization/pruning process step by step. 5: Print the operations needed to add/prune rules.	0	int

- + Training: `ripper.fit(<file data>, <classify col>, <positive value>...)`

The parameters of the `.fit()` method are shown in Table II.

TABLE II. THE PARAMETERS OF THE .FIT()

Parameter	Description	Data type
trainset	The data set used to train.	DataFrame/Numpy array
y	Class labels corresponding to trainset rows. Use if class labels aren't included in the trainset.	str/int/bool
class_feat	Column name or index of the class feature. The use of the class feature is still in the trainset.	str, int
pos_class	Name of the positive class.	str/bool
feature_names	Specify feature names. If None, feature names default to column names for a DataFrame, or indices in the case of indexed iterables such as an array or list.	List<str>
initial_model	Pre-existing model from which to begin training.	str
cn_optimize	Use algorithmic speed optimization.	bool

+ Dataset processing:

This study used the "Aggregated Financial Dataset for Fraud Detection" dataset introduced above for processing. First, remove unnecessary columns to reduce processing time, those columns include name_org and name_dest. Next, divide the data set into the train and test sets. The train set consists of 6000 lines of cheat data and 4 million lines of non-cheat data. The test set consists of 2000 lines of fraudulent data and 2 million lines of non-cheat data.

Continue to divide the training set into four small datasets and increase the imbalance rate between the sets to evaluate the effectiveness of Ripper against large unbalanced sets:

- The first set contains 6000 fraudulent data and 60,000 non-fraud data.
- The second set contains 6000 fraudulent data and 100,000 non-fraud data.
- The third set includes 6000 fraudulent data and 150,000 non-fraud data.
- The fourth set contains 6000 fraudulent data and 300,000 non-fraud data.

The test set consists of 2 thousand fraudulent lines and 2 million non-fraud lines. This set contains data that RIPPER has never encountered before, used to evaluate accuracy.

+ Results:

Results were obtained after performing RIPPER with 4 data sets (Table III).

TABLE III. STATISTICAL RESULTS AFTER TRAINING

Dataset	TPR	FPR	Preci-son	Recall	F1 score	ROC
6_60	0.996	0.03	0.97	0.96	0.96	0.977
6_100	0.992	0.026	0.97	0.86	0.91	0.949
6_150	0.995	0.031	0.97	0.88	0.92	0.906
6_300	0.996	0.023	0.98	0.79	0.87	0.880

Meaning of columns

In the Ripper algorithm, "Positive" and "Negative" are two types of labels used to evaluate and classify transactions.

+ Positive: represents valid transactions, no fraud.

+ Negative (negative): represents fraudulent transactions.

During training, the algorithm will use transactions labeled Positive and Negative to build a classification model for predicting whether new transactions will be fraudulent or not. In a data set of credit card transactions, a transaction labeled Positive would be considered a valid and fraud-free transaction, while a transaction labeled Negative would be considered a likely transaction cheat. Positive and Negative labels play an important role in evaluating and improving the accuracy of the Ripper algorithm in detecting transaction fraud.

- Dataset: dataset used for training
- True positive rate (TPR): The ratio of correctly predicting fraud cases, TPR is calculated using the formula:

$$\text{True Positive} / (\text{True Positive} + \text{False Negative})$$
- False positive rate (FPR): The ratio of false predictions of fraud cases, FPR is calculated using the formula:

$$\text{False Positive} / (\text{False Positive} + \text{True Negative})$$
- Precision: Precision is the ratio of True Positive predictions that are correct to the total number of True Positive predictions; precision is calculated using the formula:

$$\text{True Positive} / (\text{True Positive} + \text{False Positive}) = \text{Correct True Positive Predictions} / \text{True Positive Predictions}$$
- Recall: Recall is the ratio of True Positive predictions that are correct to the total number of True Positives on the whole dataset; recall is calculated using the formula:

$$\text{True Positive} / (\text{True Positive} + \text{False Negative}) = \text{Number of correct TP predictions} / \text{Total number of TP}$$
- F1 score: F1 score is the harmonic mean of precision and recall, the more accurate the F1 score, the better the model predicts; F1 score is calculated using the formula: $2 * ((\text{precision} * \text{recall}) / (\text{precision} + \text{recall}))$

ROC: It has a value of [0,1], representing the model's prediction accuracy. The higher the ROC score, the more accurate the model is. Table IV below is the rule list with the number of iterations of Ripper and the refined rule set.

TABLE IV. RULE LIST WITH THE NUMBER OF ITERATIONS RIPPER AND REFINED RULE SET WITH K=10

Number	Rule
1	transaction_type==1 and new_balance_dest<=16123.02 and old_balance_dest<=1087.36
2	transaction_type==1 and new_balance_dest<=7306.88 and old_balance_dest<=225.92
3	transaction_type==1 and new_balance_dest<=16123.02 and old_balance_dest<=1087.36
4	new_balance_org<=1744.75 and transaction_type==1 and new_balance_dest<=17247.49 and old_balance_dest<=11186.22
5	new_balance_org<=1572.27 and amount>=2106972.01 and transaction_type==0
6	transaction_type==0 and amount>=3529762.55
7	new_balance_org<=1572.27 and old_balance_org>=1388977.19 and old_balance_org<=1908598.9
8	transaction_type==0 and amount>=5014130.87
9	amount>=5014130.87 and new_balance_dest<=17247.49
10	transaction_type==0 and amount>=1867568.7 and amount<=3529762.55
11	transaction_type==0 and amount>=1867568.7 and amount<=3529762.55
12	transaction_type==0 and amount>=2519153.53 and amount<=5014130.87
13	new_balance_org<=1572.27 and old_balance_org>=955088.21 and old_balance_org<=1388977.19 and transaction_type==0
14	new_balance_org<=1744.75 and old_balance_org>=946517.94 and old_balance_org <=1204862.32 and old_balance_dest<=11186.22
15	transaction_type==0 and amount>=1262654.09 and amount <=1867568.7 and
16	transaction_type==0 and amount>=1262654.09 and amount <=1867568.7
17	transaction_type==0 and amount>=1697480.93 and amount <=2519153.53
18	new_balance_org<=1572.27 and old_balance_org>=697672.86 and old_balance_org<=955088.21 and transaction_type==0

Match the rules shown in Table V.

TABLE V. RULES WITH HIGH COVERAGE

No	Rule	Ratio
1 > 2 > 3	transaction_type==1 and new_balance_dest<=16123.02 and old_balance_dest<=1087.36	50%
4	new_balance_org<=1744.75 and transaction_type==1 and new_balance_dest<=17247.49 and old_balance_dest<=11186.22	48%
5	new_balance_org<=1572.27 and amount>=2106972.01 and transaction_type==0	10%
6 > 8	transaction_type==0 and amount>=3529762.55	6.5%
7	new_balance_org<=1572.27 and old_balance_org>=1388977.19 and old_balance_org <=1908598.9	5.5%
9	amount>=5014130.87 and new_balance_dest<=17247.49	4,7%

At this stage, the retainer is used. The dataset is divided into 10 sets. The purpose of the division is that after each step, the percentage of fraudulent data is gradually increased, creating an imbalance and checking the stability of the rule set. Each set contains 200 fraudulent data and 2000 non-cheat data (the data of these 10 datasets do not overlap). Then, add each data set in turn, corresponding to 10 cycles to see the coverage of the rules after having new data. Table VI below is the simulation rule range for ten cycles.

TABLE VI. SIMULATION RULE RANGE FOR TEN CYCLES

	Rule 1 (%)	Rule 2 (%)	Rule 3 (%)	Rule 4 (%)	Rule 5 (%)	Rule 6 (%)
Base Coverage	50	48	10	6.5	6.5	4.7
1	49	46	8	5	4	6
2	48	44.5	7	5	6	5
3	51.5	50	7	4.5	8.5	3.5
4	43.5	42	8.5	6	6.5	3.5
5	50.5	46.5	7.5	5.5	7	4.5
6	52	49.5	11	7.5	7	5.5
7	48	46.5	8.5	4.5	4	3
8	50	46	8.5	5.5	8	2
9	49.5	48	13	7.5	7	3.5
10	48.5	47	7.5	3.5	3.5	4

After 10 cycles, although the amount of data gradually increases and causes a high imbalance, the test has shown that Ripper is stable and does not suffer from "overfitting".

B. Experimental Results

Finally, this study used the fourth dataset containing 6000 fraudulent data and 300000 non-fraud data above to check the criteria for the Ripper algorithm and Ripper combined with Ensemble Learning Methods.

1) Implement the Ripper

+ Results:

- The set of rules: 31.
- Metrics: described in Table VII below.

TABLE VII. METRICS OF RIPPER

	precision	recall	f1-score	support
0	1.00	1.00	1.00	60296
1	0.97	0.77	0.86	1241
accuracy			0.99	61537
macro avg	0.98	0.88	0.93	61537
weighted avg	0.99	0.99	0.99	61537

- Confusion matrix: False Positive = 60267, False Negative = 29, True Negative = 290, True Positive = 951.
- ROC = 0.88. Execution time: 6m 58s.

2) Implement the Ripper combines with Gradient Boosting (Fig. 4)

+ Operation model:

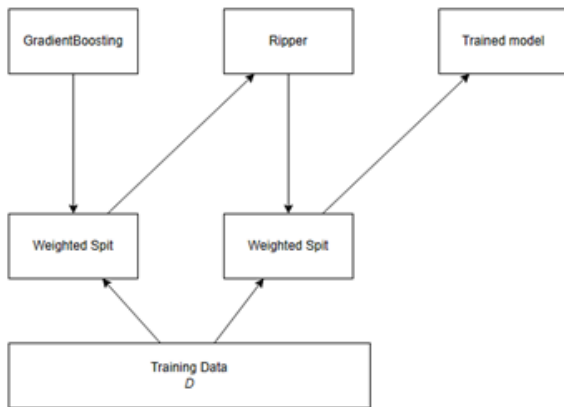


Fig. 4. Operation model by ripper combines with gradient boosting.

+ Results:

- The set of rules: 21
- Metrics: Described in Table VIII below.

TABLE VIII. METRICS OF RIPPER COMBINES GRADIENT BOOSTING

	precision	recall	f1-score	support
0	1.00	1.00	1.00	60372
1	0.98	0.92	0.95	1165
accuracy			1.00	61537
macro avg	0.99	0.96	0.97	61537
weighted avg	1.00	1.00	1.00	61537

- Confusion matrix: False Positive = 60349, False Negative = 23, True Negative = 90, True Positive = 1075.
- ROC = 0.96. Execution time: 10m 57.3s.

3) Implement the Ripper combines with Stacking (Fig. 5)
+ Operation model:

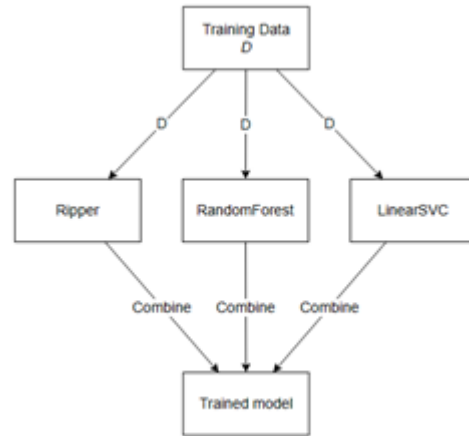


Fig. 5. Operation model by ripper combined with stacking.

+ Results: The set of rules: 45.

- Metrics: Described in Table IX below.

TABLE IX. METRICS OF RIPPER COMBINED WITH STACKING

	precision	recall	f1-score	support
0	1.00	1.00	1.00	60372
1	0.95	0.90	0.93	1165
accuracy			1.00	61537
macro avg	0.97	0.95	0.96	61537
weighted avg	1.00	1.00	1.00	61537

- Confusion matrix: False Positive = 60317, False Negative = 55, True Negative = 111, True Positive = 1054.
- ROC: 0.95. Execution time: 8m 46.8s.

4) Implement the Ripper combines with Bagging (Fig. 6)
+ Operation model:

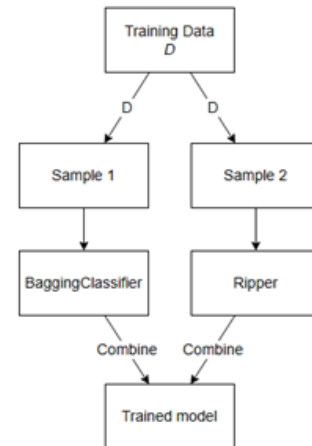


Fig. 6. Operation model by ripper combines with bagging.

- + Results: The set of rules: 44.
- Metrics: Described in Table X below.

TABLE X. METRICS OF RIPPER COMBINES BAGGING

	precision	recall	f1-score	support
0	1.00	1.00	1.00	60372
1	0.95	0.95	0.95	1165
accuracy			1.00	61537
macro avg	0.98	0.98	0.98	61537
weighted avg	1.00	1.00	1.00	61537

- Confusion matrix: False Positive = 60291, False Negative = 42, True Negative = 60, True Positive = 1144.

- ROC: 0.98. Execution time: 14m 41.7s.

C. Comparison

Table XI below summarizes the achieved indicators of each model.

TABLE XI. COMPARE CRITERIA BETWEEN METHODS

Criteria	Ripper (1)	Ripper combines Boosting (2)	Ripper combines Bagging (3)	Ripper combines Stacking (4)
Execution time	6m 58s	10m 57.3s	14m 41.7s	8m 46.8s
The set of rules	32	21	44	45
Precision	0.97	0.98	0.95	0.95
Recall	0.77	0.92	0.95	0.90
F1-score	0.86	0.95	0.95	0.93
ROC	0.883	0.961	0.976	0.952

Of the four models mentioned above, (1) gives the worst results because only a single algorithm is used. Although (2) produces the fewest rules, this is the most optimal set of rules because of its processing model. On the other hand, in models (3) and (4), the sets of rules are generated independently, and then combined, so the rules will tend to be more and overlap. This leads to models (3) and (4) generating more rules but not as efficient the model (2).

This result shows that Ripper combined with Gradient Boosting gives the best results. Compared with related studies in part I such as [4], [6], [7], [11], [12], this result achieved ACC and ROC higher than all. This study shows that Ripper combined with Gradient Boosting gives the best results.

V. CONCLUSION

The Ripper algorithm has been studied and applied effectively in the field of data mining. Ripper is a classification algorithm in machine learning and is used in many different fields. Such as data classification: Ripper can be used to classify data based on their characteristics and values. Object recognition: Ripper can be used to identify objects in an image or video by classifying objects according to their features and values. Market Research: Ripper can be used to analyze market data to identify potential customer groups and other groups. Phrase Search: Ripper can be used to

search for phrases in text or other textual data. The limitation of the study is that the algorithm executes slowly if the number of attributes of the data set is too large. In this study, the proposed method was improved by combining Ripper with Ensemble Learning Method, namely Gradient Boosting to create a new predictive model with higher accuracy and reliability (ROC increased from 88% to 98% compared to just using regular Ripper). Depending on the structure and distribution of the data set, this research can use the Ripper algorithm in combination with Gradient Boosting, Bagging, or Stacking to achieve the desired result. Thus, the proposed model has obtained better results than related studies (ACC and ROC). Besides, combining models together for training usually takes more processing time and takes up more memory resources than single models.

However, detecting transaction fraud is a constant challenge, as scammers are always looking for ways to change techniques and tricks to avoid detection. Therefore, the use of fraud detection solutions needs to be continuously updated and optimized to keep the financial system safe and meet the requirements of stakeholders.

REFERENCES

- [1] [Online]. Available: <https://tinyurl.com/4eh5ex6a>. [Accessed 2022].
- [2] R. K. Jamra, Kautsarina, D. I. Sensuse, "Systematic review of issues and solutions," In 2020 International Conference on Electrical Engineering and, pp. 1-5, 2020.
- [3] N. Maynard, "ONLINE PAYMENT FRAUD: MARKET FORECASTS, EMERGING THREATS & SEGMENT," 2021. [Online]. Available: <https://www.businesswire.com>; <https://www.juniperresearch.com/researchstore/fintech-payments/online-payment-fraud->
- [4] N. Chawla & B. Kumar, "E-commerce and consumer protection in India: The emerging trend," Journal of Business Ethics, pp. 1-24, 2021.
- [5] S. Beigi. M. R. Amin Naseri, "Credit card fraud detection using data mining and statistical methods," Journal of AI and Data Mining, Vols. 8: 149-60, 2020.
- [6] K. G. Al-Hashedi, P. Magalingam, "Financial fraud detection applying data mining techniques: A comprehensive review from 2009 to 2019," Computer Science Review, vol. 40:100402, 2021.
- [7] A. A. N. M. R. A. K. M. B. H. A. K. M. N. I. a. R. M. R. Tahmid Hasan Pranto, "A Blockchain, Smart Contract, and Data Mining Based Approach toward the Betterment of E-Commerce," Cybernetics and Systems, vol. 53, pp. 443-467, 2021.
- [8] Wang, Y., H. Liu, and Q. Liu, "Application Research of web log mining in the e-commerce. In 2020 Chinese Control and Decision Conference (CCDC)," IEEE, 2020.
- [9] W. W. Cohen, "Grammatically biased learning: learning logic programs using an explicit antecedent description language Artificial Intelligence," vol. 68, pp. 303-366, 1994.
- [10] K. Poongodi and D. Kumar, "Support Vector Machine with Information Gain Based Classification for Credit Card Fraud Detection System," The International Arab Journal of Information Technology, vol. 18, pp. 199-207, 7 9 2020.
- [11] Y. Kumar, S. Saini, R. Payal, "Comparative Analysis for Fraud Detection Using Logistic Regression, Random Forest and Support Vector Machine," International Journal of Research and Analytical Reviews (IJRAR), vol. 7: 4, pp. 726-731, 2020.
- [12] I. Sadgali, N. Sael, F. & Benabbou, "Comparative Study Using Neural Networks Techniques for Credit Card Fraud Detection," in Innovations in Smart City Application, Warszawa, Springer, 2020, pp. 287-296.
- [13] A Sahu, GM Harshvardhan, MK Gourisaria, "A Dual Approach for Credit Card Fraud Detection using Neural Network and Data Mining

- Techniques," in 2020 IEEE 17th India Council International Conference (INDICON), New Delhi, 2020.
- [14] V. Jain, M. Agrawal, and A. Kumar, "Performance Analysis of Machine Learning Algorithms in Credit Cards Fraud Detection," in International Conference on Reliability, Infocom Technologies and Optimization (ICRITO) (Trends and Future Directions), 2020.
- [15] J. Soyemi, and H. Mudasiru, "An implementation of decision tree algorithm augmented with regression analysis for fraud detection in credit card," 2020.
- [16] "Aggregated financial datasets for fraud detection," 2022. [Online]. Available: Kaggle.com. [Accessed 12 1 2023].
- [17] T. H. Pranto, A. A. Noman, M. Rahaman, A. K. M. Bahalul Haque, A. K. M. Najmul Islam & Rashedur M. Rahman, "A Blockchain, Smart Contract and Data Mining Based Approach toward the Betterment of E-Commerc," 2021.
- [18] J. H. Friedman, "Stochastic gradient boosting," pp. 367-378, 2002.
- [19] Bbeiman, Leo, "Bagging Predictors," Machine Learning, vol. 24, pp. 123-140, 1996.
- [20] B. Pavlyshenko, "Using Stacking Approaches for Machine Learning Models," in IEEE Second International Conference on Data Stream Mining & Processing, 2018.

A Single-valued Pentagonal Neutrosophic Geometric Programming Approach to Optimize Decision Maker's Satisfaction Level

Satyabrata Nath^{1*}, Purnendu Das², Pradip Debnath³

Department of Computer Science, Assam University, Silchar, Assam, India^{1,2}

Department of Applied Science and Humanities, Assam University, Silchar, Assam, India³

Abstract—Achieving the desired level of satisfaction for a decision-maker in any decision-making scenario is considered a challenging endeavor because minor modifications in the process might lead to incorrect findings and inaccurate decisions. In order to maximize the decision-maker's satisfaction, this paper proposes a Single-valued Neutrosophic Geometric Programming model based on pentagonal fuzzy numbers. The decision-maker is typically assumed to be certain of the parameters, but in reality, this is not the case, hence the parameters are presented as neutrosophic fuzzy values. The decision-maker, with this strategy, is able to achieve varying levels of satisfaction and dissatisfaction for each constraint and even complete satisfaction for certain constraints. Here the decision maker aims to achieve the maximum level of satisfaction while maintaining the level of hesitation and minimizing dissatisfaction in order to retain an optimum solution. Furthermore, transforming the objective function into a constraint adds one more layer to the N -dimensional multi-parametrizes α , β and γ . The advantages of this multi-parametrized proposed method over the existing ones are proven using numerical examples.

Keywords—Decision making; pentagonal neutrosophic numbers; single-valued neutrosophic geometric programming; multi-parametric programming

I. INTRODUCTION

Mathematical optimization, an area of applied mathematics, is used to solve real-life issues by generating mathematical models to produce feasible outcomes. In today's world the significance of mathematical optimization and decision making can be explored in various fields [1-5]. Geometric Programming (GP) is a technique in the field of mathematical optimization and multi-objective decision making that is considered a significant optimization problem consisting of objective functions and constraints composed of monomials or posynomials that are designed to solve real-world engineering problems by generating feasible outcomes [6]. The basics of GP were initially introduced in a book by Duffin, Petersen and Zener [7], and afterward its improved and extended applications can be seen in various fields. Although many problems were solved by conventional GP, most of the time the problem contains uncertainties and is considered fuzzy rather than crisp. To deal with these Zadeh [8] introduced Fuzzy Sets (FS) which were later implemented in decision making by Bellman and Zadeh [9]. Tanaka and others [10] proposed fuzzy mathematical optimization by developing the notion of level sets. Later on, Zimmermann

[11] introduced fuzzy linear programming using the concepts of fuzzy sets. Furthermore, the authors of [12] addressed series system models with the help of fuzzy parametric GP and achieved optimized system reliability and minimized cost. In the research of Samadi et al., [13] the authors presented an inventory model based on fuzzy GP for maximizing profit by implementing shortages.

Fuzzy decision-making models excel at addressing and optimizing decision-making problems, however evaluating membership values to our satisfaction is not always attainable due to a lack of readily available information. To overcome this issue, Atanassov [14] proposed Intuitionistic Fuzzy Set (IFS), which considers both membership and non-membership functions to effectively deal with vagueness rather than just the membership function as in fuzzy sets. Researchers then progressed with IFS in many real-life problems dealing with vague data, some of which are mentioned in the following literatures [15-18]. Smarandache later on introduced Neutrosophic sets (NS) [19] as the generalization of classical sets, FS and IFS which includes three independent membership functions representing the degrees of truth, indeterminacy and falsity for handling inconsistent, ambiguous and partial data more efficiently. However, the concept of NFS was established from a philosophical perspective, for which Wang et al., [20] introduced the notion of Single-valued Neutrosophic Sets (SvNS) to address practical, scientific and engineering challenges. Due to the limitations of the knowledge that humans acquire through experience or observation of the outside world, all the components indicated by SvNS are extremely appropriate for human consciousness. In contrast to the IFS, which cannot manage or represent indeterminacy and imprecise data, neutrosophic components are clearly the best fit in the representation of indeterminacy and inconsistent information. As a result, SvNS has quickly developed and is used in many different contexts [21-24]. With the advancement in research using SVNS many variations came into existence which includes triangular NS [25], trapezoidal NS [26] and recently pentagonal NS [27]. Das and Chakraborty initially applied pentagonal NS in solving linear programming problems by proposing a score function for converting the pentagonal NS data into crisp values. Further, Khalifa et al., [28] applied pentagonal neutrosophic based linear programming for optimizing stock portfolio investment. The authors of [29]

worked on maximizing profit using EOQ models using pentagonal neutrosophic demands.

A. Gaps in Existing Research and Contribution

The aspect of decision-making can be seen in many domains including geometric programming where several researchers contributed their works by developing and presenting diverse techniques to solve complex decision making problems and finding optimum solutions.

In the previous scenarios, the expert person was introduced with the simple single vector α whose value would be constant for each constraint which bound the expert to provide the same level of satisfaction to the decision maker for every constraint [30]. Thereby the expert is in a dilemmatic situation where he needs to satisfy the decision-maker but without compromising the optimal solution resulting in following extra steps for the sake of optimality. This scenario is tackled in our work where the optimal solution is achieved while providing satisfaction to the decision maker.

A multi-parametric programming approach was introduced in [31] for reaching the optimal solution in Linear Programming Problems (LPP). They proposed a method comprising a vector that would help the decision maker to attain a better satisfactory level for LPP. Though, the authors here employed an n-dimensional vector to obtain an optimal solution but did not reach the highest level of satisfaction of the decision maker in Geometric Programming Problems (GPP) since they did not consider fuzzy numbers for their work. So to overcome this, [36] proposed a multi-parametric vector α based on fuzzy numbers to solve the geometric programming problems to deal with the vagueness present in the decision making scenario. Here they were able to reach even the complete satisfaction for the decision maker in certain constraints. As a result, the expert is able to satisfy the decision maker for each constraint while maintaining the optimal value in the fuzzy GPP. Unfortunately, Fuzzy numbers cannot deal with indeterminacy, which is why Neutrosophic numbers are used in our proposed model.

However, the term "decision making" doesn't always indicate "identifying the best output from any programming problem". Instead, the decision-maker aims to achieve the intended level of satisfaction, which may or may not be the same as maximizing or minimizing the objective function. As a result, the constraint has a different effect than in the standard version. The previous attempts in multilevel decision making are mostly focused on identifying the ideal circumstances and solution algorithms to tackle linear, nonlinear, and discrete elementary problems, with only one decision allocated to each decision level for optimizing the final distinct objective. Therefore, this study focuses on maximizing satisfaction and minimizing dissatisfaction levels of the decision-maker while keeping in check the hesitation levels by incorporating multi-parametric vectors α , β and γ to Single-valued Neutrosophic Geometric Programming (SvNGP). Furthermore, the pentagonal neutrosophic numbers are subjected to a score function in order to establish a link between coefficients and exponents and obtain the crisp values. In this regard, the primary contributions of the proposed decision making model are as follows:

1) There is not much effort put towards improving decision-maker's satisfaction while taking a decision, particularly in neutrosophic environments. Thus the proposed approach aids decision-maker to take firm and confident decisions.

2) The use of pentagonal neutrosophic numbers aids in coping with imprecision and results in achieving robust decisions.

3) The objective function has been transformed into a constraint. As a result, the solution begins with the initial optimal point.

4) The addition of a new constraint to the SvPNGP problem adds a new dimension as well as a new restriction on the feasible solution space. As a result, the proposed multi-parametric α , β and γ vectors comprise $(N + 1)$ dimensions.

5) This technique allows the decision-maker to place his desires on each constraint individually, offering him more flexibility.

6) Inclusion of tolerance value aids in achieving precise results while using SvPNGP.

7) It can deal with uncertainties, hesitations and inconsistent data more efficiently.

8) The decision-maker can manage the satisfaction, hesitation and dissatisfaction degrees resulting in reaching his/her maximum desire.

9) The proposed approach is applicable to real-life programming problems.

The rest of the paper is described as: Section II presents the preliminary definitions and theorems. To generate crisp values from Single-valued pentagonal neutrosophic numbers, a score function is taken into consideration which is described in Section III. In Section IV, multi-parametric vectors $\alpha, \beta, \gamma \in [0, 1]^{N+1}$ are introduced to evaluate the optimum solution and values for the SvPNGP problem. Certain membership, non-membership and indeterminacy functions are modeled specifically for the programming problem and related theorems are also studied. In Section V, concepts of feasibility and efficiency using multi-parametric programming is described. Section VI discusses the two-phase strategy as well as the proposed algorithm to solve the SvPNGP problem. The efficacy of the method given is assessed using a numerical example problem in Section VII and the findings are examined with other methods. Also, the advantages of our method compared to the other methods are also pointed out. At last, in Section VIII, the concluding remarks are given.

II. PRELIMINARIES

In this section, several definitions and theorems are discussed that could be useful for analysis.

Definition 1. [8] A set of ordered pairs \tilde{S} is said to be a fuzzy set if:

$$\tilde{S} = \{(x, \mu_{\tilde{S}}(x)) | x \in X\}$$

where X is a non-empty set and the function $\mu_{\tilde{S}}: X \rightarrow [0, 1]$ denotes the membership function of \tilde{S} .

Definition 2. [14] Let $A' \in X$ be an intuitionistic fuzzy set (IFS) which is defined as a triplet in the form:

$$A' = \{(x, \mu_{A'}(x), \vartheta_{A'}(x)) | x \in X\}$$

where $\mu_{A'}(x), \vartheta_{A'}(x): X \rightarrow [0,1]$ such that $0 \leq \mu_{A'}(x) + \vartheta_{A'}(x) \leq 1$. The function $\mu_{A'}(x)$ represents the degree of membership and $\vartheta_{A'}(x)$ represents the degree of non-membership for every $x \in X$. Moreover, a hesitation margin or intuitionistic fuzzy index $\pi_{A'}$ can be defined as $\pi_{A'} = 1 - \mu_{A'}(x) - \vartheta_{A'}(x)$ for all $x \in A'$ which indicates the degree of belongingness of x in A' .

Definition 3. [19] A neutrosophic set $\tilde{P} \in X$ is defined by:

$$\tilde{P} = \{(x, \mu_{\tilde{P}}(x), \sigma_{\tilde{P}}(x), \vartheta_{\tilde{P}}(x)) | x \in X\}$$

where X is an universal set and $\mu_{\tilde{P}}(x), \sigma_{\tilde{P}}(x), \vartheta_{\tilde{P}}(x)$ represents three functions namely membership, indeterminacy and non-membership respectively. Their bounds are defined as:

$$\mu_{\tilde{P}}(x), \sigma_{\tilde{P}}(x), \vartheta_{\tilde{P}}(x): X \rightarrow]0^-, 1^+[$$

$$\text{such that } 0^- \leq \mu_{\tilde{P}}(x) + \sigma_{\tilde{P}}(x) + \vartheta_{\tilde{P}}(x) \leq 3^+$$

Definition 4. [20] A single valued neutrosophic set $\tilde{P} \in X$ is defined by:

$$\tilde{P} = \{(x, \mu_{\tilde{P}}(x), \sigma_{\tilde{P}}(x), \vartheta_{\tilde{P}}(x)) | x \in X\}$$

where X is an universal set and $\mu_{\tilde{P}}(x), \sigma_{\tilde{P}}(x), \vartheta_{\tilde{P}}(x)$ represents three functions namely membership, indeterminacy and non-membership respectively. Their bounds are defined as:

$$\mu_{\tilde{P}}(x), \sigma_{\tilde{P}}(x), \vartheta_{\tilde{P}}(x): X \rightarrow [0,1]$$

$$\text{such that } 0 \leq \mu_{\tilde{P}}(x) + \sigma_{\tilde{P}}(x) + \vartheta_{\tilde{P}}(x) \leq 3$$

Definition 5.[32] Let a single-valued pentagonal neutrosophic number (SvPNN) be defined as $\tilde{r} = [(r_1, r_2, r_3, r_4, r_5); \mu_{\tilde{r}}], [(s_1, s_2, s_3, s_4, s_5); \sigma_{\tilde{r}}], [(t_1, t_2, t_3, t_4, t_5); \vartheta_{\tilde{r}}]$, such that $\tilde{r} \in \mathbb{R}$, where \mathbb{R} is a set of real numbers and $\mu_{\tilde{r}}, \sigma_{\tilde{r}}, \vartheta_{\tilde{r}} \in [0,1]$. Then the membership function $\mu_{\tilde{r}}(x): \mathbb{R} \rightarrow [0, \mu]$, indeterminacy function $\sigma_{\tilde{r}}(x): \mathbb{R} \rightarrow [\sigma, 1]$ and non-membership function $\vartheta_{\tilde{r}}(x): \mathbb{R} \rightarrow [\vartheta, 1]$ of \tilde{r} is given by:

$$\mu_{\tilde{r}}(x) = \begin{cases} \mu_{\tilde{r}\bar{l}1}(x) & r_1 \leq x \leq r_2 \\ \mu_{\tilde{r}\bar{l}2}(x) & r_2 \leq x \leq r_3 \\ \mu & r_3 \\ \mu_{\tilde{r}\bar{u}1}(x) & r_3 \leq x \leq r_4 \\ \mu_{\tilde{r}\bar{u}2}(x) & r_4 \leq x \leq r_5 \\ 0 & \text{otherwise} \end{cases}$$

$$\sigma_{\tilde{r}}(x) = \begin{cases} \sigma_{\tilde{r}\bar{l}1}(x) & s_1 \leq x \leq s_2 \\ \sigma_{\tilde{r}\bar{l}2}(x) & s_2 \leq x \leq s_3 \\ \sigma & s_3 \\ \sigma_{\tilde{r}\bar{u}1}(x) & s_3 \leq x \leq s_4 \\ \sigma_{\tilde{r}\bar{u}2}(x) & s_4 \leq x \leq s_5 \\ 0 & \text{otherwise} \end{cases}$$

$$\vartheta_{\tilde{r}}(x) = \begin{cases} \vartheta_{\tilde{r}\bar{l}1}(x) & t_1 \leq x \leq t_2 \\ \vartheta_{\tilde{r}\bar{l}2}(x) & t_2 \leq x \leq t_3 \\ \vartheta & t_3 \\ \vartheta_{\tilde{r}\bar{u}1}(x) & t_3 \leq x \leq t_4 \\ \vartheta_{\tilde{r}\bar{u}2}(x) & t_4 \leq x \leq t_5 \\ 0 & \text{otherwise} \end{cases}$$

A graphical illustration of linear pentagonal neutrosophic number can be seen in Fig. 1. Here the three lines viz., black, red and blue represents the membership, non-membership and indeterminacy functions respectively. Here, the variable τ is represented by the notation $0 \leq \tau \leq 1$, where the pentagonal number will become a triangular neutrosophic number if $\tau = 0$ or 1.

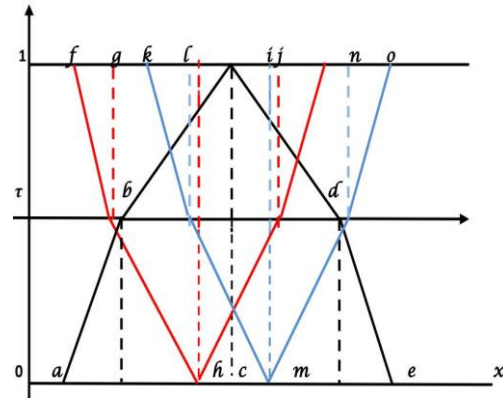


Fig. 1. Pictorial form of linear pentagonal neutrosophic number [32].

III. CRISPIFICATION OF SvPNN

To transform neutrosophic numbers into crisp values, score and accuracy functions are required. We adopted the notion of score and accuracy function from [27] for a SvPNN $\tilde{f}_{PN} = (f_1, f_2, f_3, f_4, f_5; \mu, \sigma, \vartheta)$ which is defined as follows:

- 1) *Score function:* The score function for \tilde{f}_{PN} is scaled as $\tilde{S}_{CPN} = \frac{1}{15}(f_1 + f_2 + f_3 + f_4 + f_5) \times \{2 + \mu - \sigma - \vartheta\}$
- 2) *Accuracy function:* The accuracy function is given as $\tilde{A}_{CPN} = \frac{1}{15}(f_1 + f_2 + f_3 + f_4 + f_5) \times \{2 + \mu - \sigma\}$

IV. SINGLE-VALUED NEUTROSOPHIC GEOMETRIC PROGRAMMING

Definition 6.[33] The standard form of Posynomial Geometric Programming (PGP) of X is given by:

$$\begin{aligned} \max \quad & \sum_{s=1}^{q_0} v_{0s} \prod_{j=1}^M X_j^{\gamma_{0sj}}, \\ \text{s.t.} \quad & \sum_{s=1}^{q_i} v_{is} \prod_{j=1}^M X_j^{\gamma_{isj}} \leq h_i, \quad i = 1, \dots, N \\ & X_j > 0 \end{aligned} \tag{1}$$

where the M-dimensional variable $X = (X_1, \dots, X_M)^T > 0$, coefficients $v_{is} > 0$ and exponents $\gamma_{isj} > 0$, are arbitrary real numbers.

The problem

$$\begin{aligned} \widetilde{\text{max}} \quad & \widetilde{g}_0(X), \\ \text{s.t.} \quad & \widetilde{g}_i(X) \lesssim 1, \quad i = 1, \dots, N \\ & X > 0 \end{aligned} \tag{2}$$

is called an Single-valued Neutrosophic PGP (SvNPGP) problem, where $g_i(X) = \sum_{s=1}^{q_i} v_{is}(X)$, $i = 0, \dots, N$, is a posynomial function on X , where the monomial function v_{is} of X is defined as [34]:

$$v_{is} = \begin{cases} v_{is} \prod_{j=1}^M X_j^{\gamma_{isj}}, & s = 1, \dots, q_i, i = 1, \dots, N' \\ v_{is} \prod_{j=1}^M X_j^{-\gamma_{isj}}, & s = 1, \dots, q_i, i = N' + 1, \dots, N \end{cases}$$

and $Z_0 \lesssim g_0(X) \rightarrow \widetilde{\text{max}} g_0(X)$ represents the maximum goal of the objective function $g_0(X)$ where Z_0 is considered as the lower bound. Z_0 is the expectation value of $g_0(X)$ and " \lesssim " symbolizes fuzzy version of " \leq " which basically means "less than or equal to". Therefore (2) can be changed into Single-valued neutrosophic reversed PGP problem:

$$\begin{aligned} \widetilde{\text{max}} \quad & \widetilde{g}_0(X) \gtrsim Z_0, \\ \text{s.t} \quad & \widetilde{g}_i(X) \lesssim 1, \quad i = 1, \dots, N' \\ & \widetilde{g}_i(X) \gtrsim 1, \quad i = N' + 1, \dots, N \\ & X > 0 \end{aligned} \tag{3}$$

Definition 7. A monomial function of PGP can be defined as fully SvNPGP form as:

$$\begin{aligned} \widetilde{\text{max}} \quad & \widetilde{v}_0 \prod_{j=1}^M \widetilde{X}_j^{\widetilde{\gamma}_{0j}}, \\ \text{s.t} \quad & \widetilde{v}_i \prod_{j=1}^M \widetilde{X}_j^{\widetilde{\gamma}_{ij}} \leq \widetilde{h}_i, \quad i = 1, \dots, N \\ & \widetilde{X} > \widetilde{0} \end{aligned} \tag{4}$$

where all the coefficients $\widetilde{v}_i > \widetilde{0}$ for $i = 1, \dots, N$, variables $\widetilde{X} = (\widetilde{X}_1, \dots, \widetilde{X}_M)^T$, exponents $\widetilde{\gamma}_{ij}$ and real numbers $\widetilde{h}_i > \widetilde{0}$ are Single-valued Neutrosophic numbers.

Theorem 1 .[35] Let $g_i(X)$ be a convex function for any i , then the resulting geometric programming problem is an Single-valued Neutrosophic convex problem

$$\begin{aligned} g_0(X) & \lesssim g_0 \\ g_i(X) & \lesssim 1, \quad i = 1, \dots, N \end{aligned} \tag{5}$$

Theorem 2. Any SvNPGP can be turned into a Single-valued neutrosophic convex programming problem, as specified in (2).

Proof. Let $T_j = \log(X_j)$, where $T_j = (T_{j\mu}, T_{j\sigma}, T_{j\theta})$, so $X_j = e^{T_j}$ for $1 \leq j \leq M$. Then

$$\sum_{s=1}^{q_i} v_{is} \prod_{j=1}^M X_j^{\gamma_{isj}} = \sum_{s=1}^{q_i} v_{is} e^{\sum_{j=1}^M T_j \gamma_{isj}} = g_i(X), \quad i = 0, \dots, N \tag{6}$$

Thereby, problem (2) can be turned into (5). So, by applying Theorem 1, we are able to prove it.

Theorem 3. Any Single-valued Neutrosophic monomial PGP problem (4) can be converted into a Single-valued Neutrosophic linear programming problem as follows:

$$\begin{aligned} \text{max} \quad & \ln \widetilde{v}_0 + \sum_{j=1}^M \gamma_{0j} \widetilde{T}_j, \\ \text{s.t} \quad & \ln \widetilde{v}_i + \sum_{j=1}^M \gamma_{ij} \widetilde{T}_j \leq \ln \widetilde{Q}_j, \quad i = 1, \dots, N \\ & \widetilde{X}_j > \widetilde{0}, \quad j = 1, \dots, M \end{aligned} \tag{7}$$

Proof. By using "ln" function on (4), we can say that:

$$\begin{aligned} \text{max} \quad & \ln \widetilde{v}_0 + \sum_{j=1}^M \gamma_{0j} \ln \widetilde{X}_j, \\ \text{s.t} \quad & \ln \widetilde{v}_i + \sum_{j=1}^M \gamma_{ij} \ln \widetilde{X}_j \leq \ln \widetilde{Q}_j, \quad i = 1, \dots, N \\ & \widetilde{X}_j > \widetilde{0}, \quad j = 1, \dots, M \end{aligned} \tag{8}$$

Now, when we put $\ln \widetilde{X}_j = \widetilde{T}_j$ in (8), a convex program is obtained as follows:

$$\begin{aligned} \text{max} \quad & \ln \widetilde{v}_0 + \sum_{j=1}^M \gamma_{0j} \widetilde{T}_j, \\ \text{s.t} \quad & \ln \widetilde{v}_i + \sum_{j=1}^M \gamma_{ij} \widetilde{T}_j \leq \ln \widetilde{Q}_j, \quad i = 1, \dots, N \\ & \widetilde{X}_j > \widetilde{0}, \quad j = 1, \dots, M \end{aligned} \tag{9}$$

Thus, from theorem 2, we can say that the above problem is a convex programming problem and has a similar Single-valued Neutrosophic optimal solution as the problem (4).

V. FEASIBILITY AND EFFICIENCY CONCEPTS THROUGH
MULTI-PARAMETRIC PROGRAMMING

The notion of a multi-parametric vector (α, β, γ) is introduced in this section which is useful to evaluate the level of confidence derived from the feasibility and efficacy of the optimum solution. Inclusion of tolerance value to the programming problem as a novel membership function imposes limitation as a prerequisite that can play a significant part in obtaining a suitable solution. Furthermore, the decision maker's satisfaction will be closer to the feasible solution. Considering (2) and assuming that \tilde{A}_i represents every X of neutrosophic constraints related to the neutrosophic inequality constraint $g_i(X) \lesssim 1, (i = 1, \dots, N)$, the membership function $\mu_{\tilde{A}_i}(X)$, indeterminacy function $\sigma_{\tilde{A}_i}(X)$ and non-membership function $\vartheta_{\tilde{A}_i}(X)$ are given by:

$$\mu_{\tilde{A}_i}(X) = \begin{cases} 1, & g_i(X) \leq 1 \\ 1 - \frac{d_i}{t_i}, & g_i(X) = 1 - d_i \ (d_i = 0, \dots, t_i) \\ 0, & g_i(X) \geq 1 - t_i \end{cases}$$

$$\sigma_{\tilde{A}_i}(X) = \begin{cases} 0, & g_i(X) \leq 1 \\ \frac{1 - d_i}{t_i}, & g_i(X) = 1 - d_i \ (d_i = 0, \dots, t_i) \\ 1, & g_i(X) \geq 1 - t_i \end{cases}$$

$$\vartheta_{\tilde{A}_i}(X) = \begin{cases} 0, & g_i(X) \leq 1 \\ \frac{1 - d_i}{t_i}, & g_i(X) = 1 - d_i \ (d_i = 0, \dots, t_i) \\ 1, & g_i(X) \geq 1 - t_i \end{cases}$$

where $t_i \in \mathbb{R}^+$ represents the maximum tolerance value which is determined by the decision-maker. The decision-maker assigns a tolerance value which can complicate the SvNGP problem. So, selecting a tolerance value throughout the decision making process, aiming to please the decision-maker, and then enhancing his satisfaction level, ultimately boosts efficiency.

By observing problem (3), multi-parametric vectors α, β and γ are presented where $\alpha = (\alpha_0, \dots, \alpha_N) \in (0,1]^{N+1}, \beta = (\beta_0, \dots, \beta_N) \in (0,1]^{N+1}$ and $(\gamma_0, \dots, \gamma_N) \in (0,1]^{N+1}$ represents the confidence level for the membership, non-membership and indeterminate values respectively of the programming problem. Here α_0, β_0 and γ_0 represents the satisfaction, dissatisfaction and hesitation degrees respectively, for the objective function which then will be converted into a constraint imposing a limitation to the feasible solution resulting a precise optimal solution whereas α_N, β_N and γ_N for $i = 1, \dots, N$ represents the satisfaction, dissatisfaction and hesitation degrees for each constraint. Thus, a new membership, indeterminacy and non-membership function is created solely for the objective function after it is converted to a constraint which is defined as:

$$\mu(g_0(X, \alpha, \beta, \gamma)) = \frac{Z_0 - g_0(X, \alpha, \beta, \gamma)}{t_0}, \quad g_0(X, \alpha, \beta, \gamma) \geq Z_0 - t_0 \quad (10)$$

$$\sigma(g_0(X, \alpha, \beta, \gamma)) = \frac{g_0(X, \alpha, \beta, \gamma) - Z_0 + t_0}{t_0}, \quad g_0(X, \alpha, \beta, \gamma) \geq Z_0 - t_0$$

$$\vartheta(g_0(X, \alpha, \beta, \gamma)) = \frac{g_0(X, \alpha, \beta, \gamma) - Z_0 + t_0}{t_0}, \quad g_0(X, \alpha, \beta, \gamma) \geq Z_0 - t_0$$

and

$$\mu(g_i(X, \alpha, \beta, \gamma)) = \begin{cases} 1, & g_i(X, \alpha, \beta, \gamma) \leq b_i \\ 1 - \frac{g_i(X, \alpha, \beta, \gamma) - b_i}{t_i}, & b_i \leq g_i(X, \alpha, \beta, \gamma) \leq b_i + t_i, \ i = 1, \dots, N' \\ 0, & g_i(X, \alpha, \beta, \gamma) \geq b_i + t_i, \ i = N' + 1, \dots, N \end{cases}$$

$$\sigma(g_i(X, \alpha, \beta, \gamma)) = \begin{cases} 0, & g_i(X, \alpha, \beta, \gamma) \leq b_i \\ \frac{g_i(X, \alpha, \beta, \gamma) - b_i}{t_i}, & b_i \leq g_i(X, \alpha, \beta, \gamma) \leq b_i + t_i, \ i = 1, \dots, N' \\ 1, & g_i(X, \alpha, \beta, \gamma) \geq b_i + t_i, \ i = N' + 1, \dots, N \end{cases}$$

$$\vartheta(g_i(X, \alpha, \beta, \gamma)) = \begin{cases} 0, & g_i(X, \alpha, \beta, \gamma) \leq b_i \\ \frac{g_i(X, \alpha, \beta, \gamma) - b_i}{t_i}, & b_i \leq g_i(X, \alpha, \beta, \gamma) \leq b_i + t_i, \ i = 1, \dots, N' \\ 1, & g_i(X, \alpha, \beta, \gamma) \geq b_i + t_i, \ i = N' + 1, \dots, N \end{cases}$$

Thus an underlying framework is presented to discover the optimum solution in terms of the satisfaction, dissatisfaction and indeterminacy degrees of the decision-maker with the maximum tolerance in (3):

$$\begin{aligned} \max \quad & g_0(X), \\ \text{s.t} \quad & g_0(X) \geq Z_0 - \alpha_0 t_0, \\ & g_0(X) \geq Z_0 + (1 - \gamma_0)t_0, \\ & g_0(X) \geq Z_0 + (1 - \beta_0)t_0, \\ & g_i(X) \leq b_i + (1 - \alpha_i)t_i, \quad i = 1, \dots, N' \\ & g_i(X) \leq b_i + \gamma_i t_i, \quad i = 1, \dots, N' \\ & g_i(X) \leq b_i + \beta_i t_i, \quad i = 1, \dots, N' \\ & g_i(X) \geq b_i + (1 - \alpha_i)t_i, \quad i = N' + 1, \dots, N \\ & g_i(X) \geq b_i + \gamma_i t_i, \quad i = N' + 1, \dots, N \\ & g_i(X) \geq b_i + \beta_i t_i, \quad i = N' + 1, \dots, N \\ & X > 0, \alpha_i, \beta_i, \gamma_i \in (0,1], \quad i = 0, \dots, N \\ & \alpha \geq \gamma, \alpha \geq \beta, \alpha + \beta + \gamma \leq 3 \end{aligned} \quad (11)$$

Definition 8. Let $X^* = (X_1^*, \dots, X_M^*)^T \in \mathbb{R}^M$ be an M-dimensional vector where $\alpha, \beta, \gamma \in [0,1]$ and $\alpha + \beta + \gamma \leq 3$, defined as:

$$X_{\alpha, \beta, \gamma} = \left\{ X \in \mathbb{R}^M \mid \begin{cases} X \geq 0, \\ \mu_i(g_i(X)) \geq \alpha_i \\ \vartheta_i(g_i(X)) \leq \beta_i, \ i = 0, \dots, N \\ \sigma_i(g_i(X)) \leq \gamma_i \end{cases} \right\} \quad (12)$$

in which a vector $X \in X_{\alpha, \beta, \gamma}$ will be an α, β, γ -feasible solution for (2) where α is the minimal acceptance degree, β and σ denotes the maximum rejection and hesitation degree respectively.

Theorem 4. Let $\alpha = (\alpha_0, \dots, \alpha_N) \in (0,1]^{N+1}, \beta = (\beta_0, \dots, \beta_N) \in (0,1]^{N+1}$ and $\gamma = (\gamma_0, \dots, \gamma_N) \in (0,1]^{N+1}$, and for $X_j^* \geq 0, j = 1, \dots, N, X^* = (X_1^*, \dots, X_M^*)^T \in \mathbb{R}^M$ is an M-dimensional vector and an α, β, γ -feasible solution for (2). So X^* is an α, β, γ -efficient optimal solution iff it satisfies the following constraints -

$$\begin{aligned} \max \quad & g_0(X) \\ \text{s.t} \quad & g_0(X) \geq Z_0 - \alpha_0 t_0, \\ & g_0(X) \geq Z_0 + (1 - \gamma_0)t_0, \end{aligned} \quad (13)$$

$$\begin{aligned}
 g_0(X) &\geq Z_0 + (1 - \beta_0)t_0, \\
 g_i(X) &\leq 1 + (1 - \alpha_i)t_i, \quad i = 1, \dots, N' \\
 g_i(X) &\leq 1 + \gamma_i t_i, \quad i = 1, \dots, N' \\
 g_i(X) &\leq 1 + \beta_i t_i, \quad i = 1, \dots, N' \\
 g_i(X) &\geq 1 + (1 - \alpha_i)t_i, \quad i = N' + 1, \dots, N \\
 g_i(X) &\geq 1 + \gamma_i t_i, \quad i = N' + 1, \dots, N \\
 g_i(X) &\geq 1 + \beta_i t_i, \quad i = N' + 1, \dots, N \\
 X &> 0, \alpha_i, \beta_i, \gamma_i \in (0,1], \quad i = 0, \dots, N \\
 \alpha &\geq \gamma, \alpha \geq \beta, \alpha + \beta + \gamma \leq 3
 \end{aligned}$$

where t_i denotes the maximum tolerance.

Proof. Let us consider that $\alpha = (\alpha_0, \dots, \alpha_N) \in (0,1]^{N+1}$, $\beta = (\beta_0, \dots, \beta_N) \in (0,1]^{N+1}$ and $\gamma = (\gamma_0, \dots, \gamma_N) \in (0,1]^{N+1}$, and for $X_j^* \geq 0, j = 1, \dots, N$, $X^* = (X_1^*, \dots, X_M^*)^T \in \mathbb{R}^M$ is an α, β, γ -feasible solution for (2). From definition 8 and problem (9), we have $\mu_i(g_i(X)) \geq \alpha_i, \nu_i(g_i(X)) \leq \beta_i$ and $\sigma_i(g_i(X)) \leq \gamma_i$, therefore X^* is a feasible solution. However, as $X^* \in \mathbb{R}^M$ is an α, β, γ -efficient solution, no other $X^{*'} \in X_{\alpha, \beta, \gamma}$ will satisfy $g_0(X^{*'}) > g_0(X^*)$. Thus, it means X^* is an optimal solution. Moreover, if we consider $X^{*'}$ be an optimal solution for (12) and apparently, $X^{*'}$ is an α, β, γ -feasible solution, it means $X^{*'}$ is an α, β, γ -efficient solution.

Now, let us assume the optimal solution for problem (12) be $(X^*, Z_0 = g_0(X^*))$ in Theorem 3. It is only necessary to solve the programming problem below:

$$\begin{aligned}
 \max \quad & \sum_{i=1}^N \alpha - \beta - \gamma \quad (14) \\
 \text{s.t.} \quad & g_0(X) \geq Z_0 - \alpha_0 t_0, \\
 & g_0(X) \geq Z_0 + (1 - \gamma_0)t_0, \\
 & g_0(X) \geq Z_0 + (1 - \beta_0)t_0, \\
 & g_i(X) \leq 1 + (1 - \alpha_i)t_i, \quad i = 1, \dots, N' \\
 & g_i(X) \leq 1 + \gamma_i t_i, \quad i = 1, \dots, N' \\
 & g_i(X) \leq 1 + \beta_i t_i, \quad i = 1, \dots, N' \\
 & g_i(X) \geq 1 + (1 - \alpha_i)t_i, \quad i = N' + 1, \dots, N \\
 & g_i(X) \geq 1 + \gamma_i t_i, \quad i = N' + 1, \dots, N \\
 & g_i(X) \geq 1 + \beta_i t_i, \quad i = N' + 1, \dots, N \\
 & X > 0, \alpha_i, \beta_i, \gamma_i \in (0,1], \quad i = 0, \dots, N \\
 & \alpha \geq \gamma, \alpha \geq \beta, \alpha + \beta + \gamma \leq 3
 \end{aligned}$$

VI. THE CONCEPT OF THE TWO-PHASE METHOD AND THE PROPOSED ALGORITHM

The overall process of optimization is divided into two phases which are described as follows:

Phase 1: In this phase first an appropriate GP problem is created for solving. Theorems 2 and 3 are then used to generate Single-valued Pentagonal Neutrosophic Linear Programming (SvPNLP) problem from the GP problem. The score function is then used to transform the SvPNLP problem into a crisp linear programming problem that allows the tolerance value to be set. In this case, the decision-maker implements his requirement according to his satisfaction. The decision-maker could choose various degrees of tolerance value, which results in distinct sets of feasible alternatives; consequently, we must devise a technique to determine the optimal solution within these feasible choices.

Phase 2: This phase begins with a feasible solution provided in phase 1 and its goal is to increase satisfaction by providing an optimal solution. The multi-parametric confidence vectors $\alpha, \beta, \gamma \in (0,1]^{(N+1)}$ are utilized to correlate the degree of satisfaction, dissatisfaction and indeterminacy with its relevant environment. Then the conversion of the objective function into a constraint takes place at this stage where Z_0 marks the beginning of the optimum solution along with α_0, β_0 and γ_0 as satisfaction, dissatisfaction and hesitation degrees. The tolerance degree, t_i , can be enhanced for individual constraint and objective function, allowing the degree of satisfaction to be maximized and dissatisfaction degree to be minimized while maintaining the degree of indeterminacy in individual constraint. Finally, solving the original problem with the proposed model an optimal solution is achieved with the highest degree of satisfaction while keeping the level of dissatisfaction and indeterminacy in control.

An algorithm along with a flowchart, in Fig. 2 is presented for finding an optimal solution for SvPNGP problem based on the preceding discussion (3).

Algorithm: SvPNGP Modelling

1. Model the SvPNGP problem.
2. Convert the SvPNGP to crisp LP using the help of the score function and applying Theorem 2 and 3.
3. Find the initial optimal value Z_0 from basic variables.
4. Add tolerance value and apply α, β - efficiency and formulate the equivalent LPP:

$$\begin{aligned}
 \max \quad & \alpha, \min \beta, \min \gamma \quad (15) \\
 \text{s.t.} \quad & \mu_0(X) \geq \alpha, \vartheta_0(X) \leq \beta, \sigma_0(X) \leq \gamma \\
 & \mu_i(X) \geq \alpha, \vartheta_i(X) \leq \beta, \sigma_i(X) \leq \gamma, \quad i = 1, \dots, N, \\
 & \alpha \geq \beta, \alpha \geq \gamma, \alpha + \beta + \gamma \leq 3, \\
 & X > 0, \alpha_i, \beta_i, \gamma_i \in (0,1]
 \end{aligned}$$

The above LPP is equivalent to:

$$\begin{aligned}
 \text{s.t.} \quad & \max (\alpha - \beta - \gamma) \quad (16) \\
 & \mu_0(X) \geq \alpha, \vartheta_0(X) \leq \beta, \sigma_0(X) \leq \gamma \\
 & \mu_i(X) \geq \alpha, \vartheta_i(X) \leq \beta, \sigma_i(X) \leq \gamma, \quad i = 1, \dots, N, \\
 & \alpha \geq \beta, \alpha \geq \gamma, \alpha + \beta + \gamma \leq 3, \\
 & X > 0, \alpha_i, \beta_i, \gamma_i \in (0,1]
 \end{aligned}$$

5. According to multi-parametric $\alpha, \beta, \gamma \in (0,1]^{(N+1)}$ apply the membership, non - membership and hesitation functions and place the objective function as a constraint.

6. Solve and find $g_0(Z^*, \alpha, \beta, \gamma)$ using the dual-simplex method.
7. Build a new programming problem model under multi-parametric α_i, β_i and γ_i with different degrees of satisfaction, dissatisfaction and indeterminacy respectively.
8. Solve the new problem and find the optimal satisfaction degree.
9. Determine the optimal value under optimal $\alpha^*, \beta^*, \gamma^*$ and evaluate $g_0(Z^{**}, \alpha^*, \beta^*, \gamma^*)$.

e^{46} , and e^{64} in Kgs. The water pipes need to be manufactured utilizing four varieties of concrete materials M1, M2, M3 and M4. Table I shows the percentage of each kind of raw concrete required in each pipe (kg) and its unit price (\$/kg). Determine the maximum amount of raw concrete required while staying within the owner's tolerance limit.

TABLE I. CONCRETE PERCENTAGES AND ITS PRICE/UNIT

Pipes	M1	M2	M3	M4	Need (Kg)
P1	P1M1	P1M2	P1M3	P1M4	e^{19}
P2	P2M1	P2M2	P2M3	P2M4	e^{46}
P3	P3M1	P3M2	P3M3	P3M4	e^{64}
Unit Price (\$/Kg)	5	6	3	5	

P1M1= ((0,1,1,2,2);0,6,0,4,0,3), P1M2= ((0,1,3,4,5);0,9,0,1,0,3), P1M3= ((1,1,1,1,1);0,9,0,3,0,1), P1M4= ((1,2,2,3,4);0,8,0,5,0,3), P2M1= ((5,6,6,7,8);0,8,0,4,0,4), P2M2= ((3,4,6,7,9);0,8,0,5,0,3), P2M3= ((2,3,3,4,5);0,6,0,5,0,6), P2M4= ((0,2,2,4,5);0,8,0,2,0,5), P3M1= ((1,2,4,5,6);0,7,0,2,0,2), P3M2= ((2,3,5,6,8);0,7,0,2,0,2), P3M3= ((1,1,3,3,3);0,7,0,4,0,3), P3M4= ((10,11,13,14,15);0,8,0,4,0,2)

Solution. The above problem can be converted into SvPNGP as follows:

$$\begin{aligned} & \overline{\max} x_1^5 x_2^6 x_3^3 x_4^5 & (17) \\ \text{s.t} & x_1^{P1M1} x_2^{P1M2} x_3^{P1M3} x_4^{P1M4} \leq e^{19} \\ & x_1^{P2M1} x_2^{P2M2} x_3^{P2M3} x_4^{P2M4} \leq e^{46} \\ & x_1^{P3M1} x_2^{P3M2} x_3^{P3M3} x_4^{P3M4} \leq e^{64} \\ & x_1, x_2, x_3, x_4 > 0 \end{aligned}$$

By using $x_i = e^{z_j}$ ($1 \leq j \leq 4$), we can change problem (16) into the intuitionistic fuzzy problem by utilizing Theorems 2 and 3.

$$\begin{aligned} & \overline{\max} 5Z_1 + 6Z_2 + 3Z_3 + 5Z_4 & (18) \\ \text{s.t} & P1M1Z_1 + P1M2Z_2 + P1M3Z_3 + P1M4Z_4 \leq 19 \\ & P2M1Z_1 + P2M2Z_2 + P2M3Z_3 + P2M4Z_4 \leq 46 \\ & P3M1Z_1 + P3M2Z_2 + P3M3Z_3 + P3M4Z_4 \leq 64 \end{aligned}$$

Next, we apply the score function on SvPNN

$$\begin{aligned} & \overline{\max} 5Z_1 + 6Z_2 + 3Z_3 + 5Z_4 & (19) \\ \text{s.t} & 0.76Z_1 + 2.17Z_2 + 0.83Z_3 + 1.60Z_4 \leq 19 \\ & 4.27Z_1 + 3.87Z_2 + 1.70Z_3 + 1.82Z_4 \leq 46 \\ & 2.64Z_1 + 3.68Z_2 + 1.47Z_3 + 9.24Z_4 \leq 64 \end{aligned}$$

After the conversion, the primary optimal solution is drawn from the basic variables $x_1 = e^{4.42}, x_2 = 1, x_3 = e^{11.40}, x_4 = e^{4.17}$ and the optimal value is $e^{76.18}$. By applying $x_0 = e^{z_0}$, we get $Z_0 = 76.18$. Using the two-phase technique and applying the membership, non-membership and indeterminacy functions defined in (9) along with substituting the values of Z_0, t_0, t_1, t_2 and t_3 where $t_0 = 5, t_1 = 1, t_2 = 4$ and $t_3 = 6$ for $t_i, (i = 0,1,2,3)$ are the tolerance values which are set up by the decision maker, we can convert problem (18) into the programming problem as given below:

$$\overline{\max} (\alpha - \beta - \gamma) \quad (20)$$

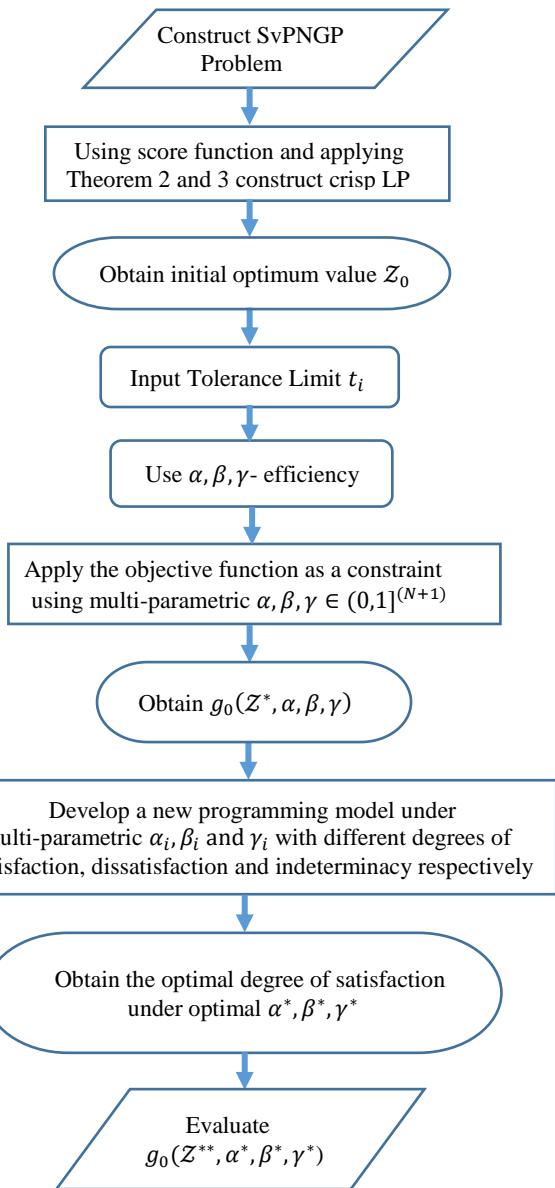


Fig. 2. Flowchart of the proposed work.

VII. IMPLEMENTATION OF THE PROPOSED MODEL WITH THE HELP OF NUMERICAL ILLUSTRATION

Example 1. A water distribution plant wants to produce concrete pipes for an underground water distribution project. It requires three pipes P1, P2 and P3 with utmost weight e^{19} ,

$$\begin{aligned}
 \text{s.t. } & 5Z_1 + 6Z_2 + 3Z_3 + 5Z_4 \geq 76.18 - 5\alpha_0, \\
 & 5Z_1 + 6Z_2 + 3Z_3 + 5Z_4 \geq 71.18 - 5\gamma_0, \\
 & 5Z_1 + 6Z_2 + 3Z_3 + 5Z_4 \geq 71.18 - 5\beta_0, \\
 & 0.76Z_1 + 2.17Z_2 + 0.83Z_3 + 1.60Z_4 \leq 20 - \alpha_1, \\
 & 0.76Z_1 + 2.17Z_2 + 0.83Z_3 + 1.60Z_4 \leq 19 + \gamma_1, \\
 & 0.76Z_1 + 2.17Z_2 + 0.83Z_3 + 1.60Z_4 \leq 19 + \beta_1, \\
 & 4.27Z_1 + 3.87Z_2 + 1.70Z_3 + 1.82Z_4 \leq 50 - 4\alpha_2, \\
 & 4.27Z_1 + 3.87Z_2 + 1.70Z_3 + 1.82Z_4 \leq 46 + 4\gamma_2, \\
 & 4.27Z_1 + 3.87Z_2 + 1.70Z_3 + 1.82Z_4 \leq 46 + 4\beta_2, \\
 & 2.64Z_1 + 3.68Z_2 + 1.47Z_3 + 9.24Z_4 \leq 70 - 6\alpha_3, \\
 & 2.64Z_1 + 3.68Z_2 + 1.47Z_3 + 9.24Z_4 \leq 64 + 6\gamma_3, \\
 & 2.64Z_1 + 3.68Z_2 + 1.47Z_3 + 9.24Z_4 \leq 64 + 6\beta_3, \\
 & \alpha_i, \beta_i, \gamma_i \in (0,1], \alpha_i \geq \beta_i, \alpha_i \geq \gamma_i, \alpha_i + \beta_i + \gamma_i \leq 3, \\
 & i = 0,1,2,3
 \end{aligned}$$

Table II displays the satisfaction of the decision-maker at various degrees of α, β, γ -efficiency confidence. If $\alpha = (\alpha_0, \alpha_1, \alpha_2, \alpha_3), \beta = (\beta_0, \beta_1, \beta_2, \beta_3), \gamma = (\gamma_0, \gamma_1, \gamma_2, \gamma_3)$ and $g_0(Z^*, \alpha, \beta, \gamma)$ signify the optimal value of the objective function at every step under different conditions, we may obtain the following table using the LINGO 18.0 software:

From Table II it can be observed that the maximum initial optimal solution $g_0(Z, \alpha_i, \beta_i, \gamma_i)$ is achieved at row 1 with a value of 80.11. It is also seen that the least efficient components are α_1 and α_2 . As the values of α_1 and α_2 increases, the values of $\gamma_1, \gamma_2, \beta_1$ and β_2 decreases because of the constraint $\alpha_i \geq \beta_i$ and $\alpha_i \geq \gamma_i$, that results in the degradation of the optimal solution. By reducing the values of α_1 and α_2 provides better results. If we increase the values of α_0 and α_3 and decrease $\gamma_0, \gamma_3, \beta_0$ and β_3 respectively, we are able to reach closer to the optimal solution. Now, as an initial solution, we will strive to minimize α_1 and α_2 by selecting the (0.5, 0.5, 0.5, 0.5)-efficient solution having optimal value $g_0(Z, \alpha_i, \beta_i, \gamma_i) = 80.11$.

Now, we will determine the LP problem below that is influenced by $g_0(Z, \alpha_i, \beta_i, \gamma_i) = 80.11$.

$$\begin{aligned}
 & \max \sum_{i=0}^3 (\alpha_i - \beta_i - \gamma_i) \tag{21} \\
 \text{s.t. } & 5Z_1 + 6Z_2 + 3Z_3 + 5Z_4 \geq 80.11 - 5\alpha_0, \\
 & 5Z_1 + 6Z_2 + 3Z_3 + 5Z_4 \geq 75.11 - 5\gamma_0, \\
 & 5Z_1 + 6Z_2 + 3Z_3 + 5Z_4 \geq 75.11 - 5\beta_0, \\
 & 0.76Z_1 + 2.17Z_2 + 0.83Z_3 + 1.60Z_4 \leq 20 - \alpha_1, \\
 & 0.76Z_1 + 2.17Z_2 + 0.83Z_3 + 1.60Z_4 \leq 19 + \gamma_1, \\
 & 0.76Z_1 + 2.17Z_2 + 0.83Z_3 + 1.60Z_4 \leq 19 + \beta_1, \\
 & 4.27Z_1 + 3.87Z_2 + 1.70Z_3 + 1.82Z_4 \leq 50 - 4\alpha_2, \\
 & 4.27Z_1 + 3.87Z_2 + 1.70Z_3 + 1.82Z_4 \leq 46 + 4\gamma_2, \\
 & 4.27Z_1 + 3.87Z_2 + 1.70Z_3 + 1.82Z_4 \leq 46 + 4\beta_2, \\
 & 2.64Z_1 + 3.68Z_2 + 1.47Z_3 + 9.24Z_4 \leq 70 - 6\alpha_3, \\
 & 2.64Z_1 + 3.68Z_2 + 1.47Z_3 + 9.24Z_4 \leq 64 + 6\gamma_3, \\
 & 2.64Z_1 + 3.68Z_2 + 1.47Z_3 + 9.24Z_4 \leq 64 + 6\beta_3, \\
 & 0.5 \leq \alpha_0 \leq 1, 0.5 \leq \beta_0 \leq 1, 0.5 \leq \gamma_0 \leq 1 \\
 & 0.5 \leq \alpha_1 \leq 1, 0.5 \leq \beta_1 \leq 1, 0.5 \leq \gamma_1 \leq 1 \\
 & 0.5 \leq \alpha_2 \leq 1, 0.5 \leq \beta_2 \leq 1, 0.5 \leq \gamma_2 \leq 1 \\
 & 0.5 \leq \alpha_3 \leq 1, 0.5 \leq \beta_3 \leq 1, 0.5 \leq \gamma_3 \leq 1
 \end{aligned}$$

The optimum solution of problem (20) will be reached by maximizing the satisfaction degree as $Z^{**} = (5.64, 0, 10.76, 3.93)$ with confidence level $\alpha^* = (1, 0.5, 0.5, 0.5), \beta^* = (0, 0.5, 0.5, 0.5)$ and $\gamma^* = (0, 0.5, 0.5, 0.5)$, and the optimal value calculated with respect to α^*, β^* and γ^* as $g_0(Z^{**}, \alpha^*, \beta^*, \gamma^*) = (5.64, 0, 10.76, 3.93; 1, 0.5, 0.5, 0.5; 0, 0.5, 0.5, 0.5; 0, 0.5, 0.5, 0.5) = 80.11$. As a result, the optimal solution to GP programming problem (16) is $x_1 = e^{5.64}, x_2 = 1, x_3 = e^{10.76}, x_4 = e^{3.93}$, and the optimal value is $e^{80.11}$.

TABLE II. DETERMINING THE MAXIMUM LEVEL OF SATISFACTION WITH A MULTI-PARAMETERS α, β, γ

S. No	$\alpha_0, \alpha_1, \alpha_2, \alpha_3$	$\gamma_0, \gamma_1, \gamma_2, \gamma_3$	$\beta_0, \beta_1, \beta_2, \beta_3$	Z_1	Z_2	Z_3	Z_4	$g_0(Z, \alpha_i, \beta_i, \gamma_i)$
1	0.5,0.5,0.5,0.5	0.5,0.5,0.5,0.5	0.5,0.5,0.5,0.5	5.64	0	10.76	3.93	80.11
2	0.9,0.5,0.5,0.7	0.2,0.5,0.5,0.1	0.1,0.5,0.5,0.2	5.44	0	11.58	3.59	79.93
3	0.4,0.8,0.3,1	0.2,0.4,0.2,0.3	0.1,0.5,0.2,0.6	5.22	0	11.36	3.63	78.32
4	0.5,0.9,0.9,0.5	0,0.3,0.2,0.1	0.3,0.4,0.5,0.3	5.36	0	10.93	3.72	78.20
5	0.5,0.8,0.9,0.5	0.2,0.1,0.2,0.3	0.1,0.3,0.1,0.2	5.27	0	10.80	3.83	77.89
6	0.7,0.5,0.8,0.5	0.7,0.5,0.8,0.5	0.2,0.5,0.1,0.5	5.21	0	11.00	3.75	77.85
7	0.9,0.9,0.9,0.9	0,0,0,0	0.3,0.2,0.2,0.3	5.11	0	11.07	3.70	77.28

Example 2. In continuation from example 1, determining the maximum satisfaction of decision maker without any tolerance limit then the results are shown in Table III.

TABLE III. DETERMINING THE MAXIMUM SATISFACTION LEVEL WITHOUT TOLERANCE LIMIT

S. No	1	2	3	4
α	0.5,0.5,0.5,0.5	0.9,0.5,0.5,0.7	0.7,0.5,0.8,0.5	0.5,0.9,0.9,0.5
γ	0.5,0.5,0.5,0.5	0.2,0.5,0.5,0.1	0.2,0.1,0.3,0.2	0,0.3,0.2,0.1
β	0.5,0.5,0.5,0.5	0.1,0.5,0.5,0.2	0.2,0.5,0.1,0.5	0.3,0.4,0.5,0.3
Z_1	4.89	4.85	5.07	5.06
Z_2	0	0	0	0
Z_3	11.90	12.04	11.20	11.23
Z_4	3.69	3.63	3.71	3.70
G_0	78.59	78.56	77.55	77.54

$$\alpha = (\alpha_0, \alpha_1, \alpha_2, \alpha_3), \beta = (\beta_0, \beta_1, \beta_2, \beta_3), \gamma = (\gamma_0, \gamma_1, \gamma_2, \gamma_3), G_0 = g_0(Z, \alpha, \beta, \gamma)$$

Example 3. With continuation from example 1, determining the maximum satisfaction of decision maker with single parametric α, β and γ then the results are displayed in Table IV.

TABLE IV. DETERMINING THE MAXIMUM SATISFACTION LEVEL WITH SINGLE PARAMETRIC α, β AND γ

S. No	α	γ	β	Z_1	Z_2	Z_3	Z_4	G_0
1	0.5	0.5	0.5	5.64	0	10.76	3.93	80.11
2	0.6	0.4	0.4	5.53	0	10.82	3.88	79.55
3	0.8	0.2	0.2	5.32	0	10.94	3.79	78.41
4	0.9	0.1	0.1	5.21	0	11.00	3.74	77.85
5	1	0	0	5.10	0	11.07	3.70	77.28

$$G_0 = g_0(Z, \alpha, \beta, \gamma)$$

It can be observed that the optimal solution degrades with the absence of tolerance limit while analyzing Table II and III. When Table II and IV are compared, it is found that in Table IV, raising the confidence level α and decreasing γ and β , reduces the ideal solution, whereas we anticipate the optimal solution to increase as the confidence level rises. Similarly,

TABLE V. COMPARATIVE ANALYSIS BETWEEN DIFFERENT OPTIMIZATION APPROACHES

Methods	α	β	γ	t_0	t_1	t_2	t_3	Z_1	Z_2	Z_3	Z_4	G_0
1) Proposed SvPNGP	(1, 0.5, 0.5, 0.5)	(0, 0.5, 0.5, 0.5)	(0, 0.5, 0.5, 0.5)	5	1	4	6	5.64	0	10.76	3.93	80.11
2) IFGP	(1, 0.5, 0.5, 0.5)	(0, 0.5, 0.5, 0.5)		5	1	4	6	0	8.52	0	6.99	76.78
3) Khorsandi et al's Method [36]	(1, 0.5, 0.5, 0.5)			5	1	4	6	2.51	5.3	0	3.18	60.34
4) Zimmermann's Method [37]	(0.2) (0.5) (1)			5	1	4	6	2.74 2.51 2.12	5.21 5.30 5.45	0 0 0	3.30 3.18 2.98	61.59 60.34 58.27

$$\alpha = (\alpha_0, \alpha_1, \alpha_2, \alpha_3), \beta = (\beta_0, \beta_1, \beta_2, \beta_3), \gamma = (\gamma_0, \gamma_1, \gamma_2, \gamma_3), G_0 = g_0(Z, \alpha, \beta, \gamma)$$

VIII. CONCLUSION

In order to obtain optimal results in decision making, the decision-maker needs to be provided with the flexibility to achieve satisfaction in the decision making process. Thereby, this research article proposed a SvPNGP model by incorporating multi-parametric vectors α, β, γ to achieve the maximum degree of satisfaction while minimizing the degree of dissatisfaction and hesitation within the tolerance limit of

the decision maker has to adjust the levels of satisfaction, dissatisfaction and hesitancy levels for every constraint to the same degree, but in the method proposed, the satisfaction, dissatisfaction and hesitancy levels for each constraint can be decreased or increased independently. Table IV clearly shows that raising the confidence level has the opposite effect on the optimization, and the optimal value returns to its original solution, whereas in Table II, until the satisfaction, dissatisfaction and hesitancy degree components change, the optimal value remains optimal. The reason for this is because the confidence vectors are not reasonable for all constraints, especially when the objective function is transformed into a constraint. Thus the adaptability of the $(N + 1)$ -dimensional α_i, β_i and γ_i confidence levels can help in achieving the decision-maker's purpose of getting a better optimal result.

Table V presents the difference of solutions for example 1 using four methods. Comparing our work with Intuitionistic Fuzzy Geometric Programming (IFGP), Khorsandi et al., [36] and Zimmermann's method [37], it is observed that the solution achieved using our proposed method for solving SvPNGP is more efficient compared to the solution obtained using the other techniques. Here methods 1, 2 and 3 are multi-parametric, whereas method 4 is single parametric, and methods 3 and 4 are designed to solve fuzzy optimization problems, whereas Method 1 is intended to address Neutrosophic optimization problems.

The proposed method achieved the highest optimal value compared to the existing techniques. Fuzzy optimization only considers one degree of acceptance or rejection at a time whereas Intuitionistic Fuzzy optimization includes both degrees of acceptance and rejection in order to manage optimization but in reality, there are some circumstances where, due to lack of information or indeterminacy, evaluating the membership and non-membership functions together cannot yield a greater and/or more satisfactory conclusion. As a result, there is still an indeterministic element on which hesitation persists which is addressed by neutrosophic optimization.

the decision-maker. With this strategy, the decision-maker can obtain an optimal solution for the SvPNGP problem while satisfying his/her needs and moreover the decision maker is not restricted for selecting the same tolerance value for individual constraints. We divided the whole process into a two-phase method where the SvPNGP is transformed to a crisp LP problem in the first phase and in the next phase, the multi-parametric vectors are applied along with membership, indeterminacy and non-membership functions and solved to

find the optimal solution. With the help of numerical problems, we evaluated and analyzed certain parameters with our proposed model. The results were then compared with the existing methods and found out to produce better optimal solution compared to others.

The contribution of this paper includes developing an optimal SvPNGP model to enable the decision-maker to achieve robust decisions while providing him the flexibility to achieve the desired level of satisfaction. As the model is built on neutrosophic numbers, it can handle uncertainty in real-world programming situations.

For future work, we hope to expand our work with Plithogenic sets, which is another generalized method that can be useful for dealing with inconsistent and indeterminate data. The extended approach can be used to a wide range of real-world challenges in the field of engineering, manufacturing, management and many more.

REFERENCES

- [1] M. Abdallah, S. Hamdan, and A. Shabib, "A multi-objective optimization model for strategic waste management master plans," *Journal of Cleaner Production*, 284, 2021, p.124714. <https://doi.org/10.1016/j.jclepro.2020.124714>.
- [2] P. Goudarzi, A.M Rahmani, and M. Mosleh, "A mathematical optimization model using red deer algorithm for resource discovery in CloudIoT," *Transactions on Emerging Telecommunications Technologies*, 33(12), 2022, p.e4646. <https://doi.org/10.1002/ett.4646>.
- [3] S. Nath, P. Das, and P. Debnath, "A Brief Review on Multi-Attribute Decision Making in the Emerging Fields of Computer Science," In *Computational Intelligence in Communications and Business Analytics: 4th International Conference, CICBA 2022, Silchar, India, Revised Selected Papers*, Cham: Springer International Publishing, January 7–8, 2022, pp. 3-18. https://doi.org/10.1007/978-3-031-10766-5_1.
- [4] S. Pant, P. Garg, A. Kumar, M. Ram, A. Kumar, H. K. Sharma, and Y. Klochkov, "AHP-based multi-criteria decision-making approach for monitoring health management practices in smart healthcare system," *International Journal of System Assurance Engineering and Management*, 2023, pp.1-12. <https://doi.org/10.1007/s13198-023-01904-5>.
- [5] M.B. Bouraima, Y. Qiu, Ž. Stević, and V. Simić, "Assessment of alternative railway systems for sustainable transportation using an integrated IRN SWARA and IRN CoCoSo model," *Socio-Economic Planning Sciences*, 86, 2023, p.101475. <https://doi.org/10.1016/j.seps.2022.101475>.
- [6] C.H. Huang, "Engineering design by geometric programming," *Mathematical problems in engineering*, 2013. <https://doi.org/10.1155/2013/568098>.
- [7] R.J. Duffin, "Geometric programming-theory and application," vol. QA264, 1967.
- [8] L.A. Zadeh, "Fuzzy sets," *Information and control*, 8(3), 1965, pp. 338-353.
- [9] R.E. Bellman, and L.A. Zadeh, "Decision-making in a fuzzy environment," *Management science*, 17(4), 1970, pp.B-141. <https://doi.org/10.1287/mnsc.17.4.B141>.
- [10] H. Tanaka, T. Okuda, and K. Asai, "Fuzzy mathematical programming," *Transactions of the society of instrument and control engineers*, 9(5), 1973, pp.607-613. <https://doi.org/10.9746/sicetr1965.9.607>.
- [11] H.J. Zimmermann, "Fuzzy programming and linear programming with several objective functions," *Fuzzy sets and systems*, 1(1), 1978, pp.45-55. [https://doi.org/10.1016/0165-0114\(78\)90031-3](https://doi.org/10.1016/0165-0114(78)90031-3).
- [12] B.S. Mahapatra, and G.S. Mahapatra, "Reliability and cost analysis of series system models using fuzzy parametric geometric programming," *Fuzzy Information and Engineering*, 2(4), 2010, pp.399-411. <https://doi.org/10.1007/s12543-010-0058-1>.
- [13] F. Samadi, A. Mirzazadeh, and M.M. Pedram, "Fuzzy pricing, marketing and service planning in a fuzzy inventory model: A geometric programming approach," *Applied Mathematical Modelling*, 37(10-11), 2013, pp.6683-6694. <https://doi.org/10.1016/j.apm.2012.12.020>.
- [14] K. T. Atanassov, "Intuitionistic fuzzy sets," *Fuzzy sets and Systems*, 20(1), 1986, pp. 87-96. https://doi.org/10.1007/978-3-7908-1870-3_1.
- [15] E. Jafarian, J. Razmi, and M.F. Baki, "A flexible programming approach based on intuitionistic fuzzy optimization and geometric programming for solving multi-objective nonlinear programming problems," *Expert Systems with Applications*, 93, 2018, pp.245-256. <https://doi.org/10.1016/j.eswa.2017.10.030>.
- [16] Z. Kheiri, and B.Y. Cao, "Posynomial geometric programming with intuitionistic fuzzy coefficients," *Fuzzy Systems & Operations Research and Management*, Springer, Cham, 2016, pp. 15-30. https://doi.org/10.1007/978-3-319-19105-8_2.
- [17] S. Islam, and W.A. Mandal, "Fuzzy geometric programming techniques and applications," Springer Singapore, 2019. <https://doi.org/10.1007/978-981-13-5823-4>.
- [18] Y. Xue, and Y. Deng, "Decision making under measure-based granular uncertainty with intuitionistic fuzzy sets," *Applied intelligence*, 51, 2021, pp.6224-6233. <https://doi.org/10.1007/s10489-021-02216-6>.
- [19] F. Smarandache, "A unifying field in Logics: Neutrosophic Logic," *Philosophy*, American Research Press, 1999, pp. 1-141.
- [20] H. Wang, F. Smarandache, Y. Zhang, and R. Sunderraman, "Single valued neutrosophic sets," *Infinite study*, 12, 2010.
- [21] J.S. Chai, G. Selvachandran, F. Smarandache, V.C. Gerogiannis, L.H. Son, Q.T. Bui, and B. Vo, "New similarity measures for single-valued neutrosophic sets with applications in pattern recognition and medical diagnosis problems," *Complex & Intelligent Systems*, 7, 2021, pp.703-723. <https://doi.org/10.1007/s40747-020-00220-w>.
- [22] V. Başhan, H. Demirel, and M. Gul, "An FMEA-based TOPSIS approach under single valued neutrosophic sets for maritime risk evaluation: the case of ship navigation safety," *Soft Computing*, 24(24), 2020, pp.18749-18764. <https://doi.org/10.1007/s00500-020-05108-y>.
- [23] M. Luo, G. Zhang, and L. Wu, "A novel distance between single valued neutrosophic sets and its application in pattern recognition," *Soft Computing*, 26(21), 2022, pp.11129-11137. <https://doi.org/10.1007/s00500-022-07407-y>.
- [24] M. Ali, Z. Hussain, and M.S. Yang, "Hausdorff Distance and Similarity Measures for Single-Valued Neutrosophic Sets with Application in Multi-Criteria Decision Making," *Electronics*, 12(1), 2022, p.201. <https://doi.org/10.3390/electronics12010201>.
- [25] M. Abdel-Basset, M. Mohamed, A.N. Hussien and A.K. Sangaiah, "A novel group decision-making model based on triangular neutrosophic numbers," *Soft Comput*, 22(20), 2018, pp. 6629–6643. <https://doi.org/10.1007/s00500-017-2758-5>.
- [26] J. Ye, "Trapezoidal neutrosophic set and its application to multiple attribute decision-making," *Neural Comput Appl*, 26(5), 2015, pp.1157–1166. <https://doi.org/10.1007/s00521-014-1787-6>.
- [27] S.K. Das, and A. Chakraborty, "A new approach to evaluate linear programming problem in pentagonal neutrosophic environment," *Complex & intelligent systems*, 7, 2021, pp.101-110. <https://doi.org/10.1007/s40747-020-00181-0>.
- [28] H.A.E.W. Khalifa, M. Saeed, A.U. Rahman, and S. El-Morsy, "An Application of Pentagonal Neutrosophic Linear Programming for Stock Portfolio Optimization," *Neutrosophic Sets and Systems*, 51(1), 2022, p.41. https://digitalrepository.unm.edu/nss_journal/vol51/iss1/41.
- [29] C. Kar, T.K. Roy, and M. Maiti, "EOQ model with price, marketing, service and green dependent neutrosophic demand under uncertain resource constraint: A geometric programming approach," *Neutrosophic Sets and Systems*, 51, 2022, pp.797-823. <http://fs.unm.edu/NSS2/index.php/111/article/view/2601>.
- [30] B.Y. Cao, "Optimal models and methods with fuzzy quantities," vol. 248, 2010, Berlin: Springer. <https://doi.org/10.1007/978-3-642-10712-2>.
- [31] H. Attari, and S.H. Nasserri, "New concepts of feasibility and efficiency of solutions in fuzzy mathematical programming problems," *Fuzzy Information and Engineering*, 6(2), 2014, pp.203-221. <https://doi.org/10.1016/j.fiae.2014.08.005>.

- [32] A. Chakraborty, S. Broumi and P.K. Singh, "Some properties of pentagonal neutrosophic numbers and its applications in transportation problem environment," *Infinite Study*, 2019. https://digitalrepository.unm.edu/nss_journal/vol28/iss1/16.
- [33] S.T. Liu, "Geometric programming with fuzzy parameters in engineering optimization. *International Journal of Approximate Reasoning*," 46(3), 2007, pp.484-498. <https://doi.org/10.1016/j.ijar.2007.01.004>.
- [34] B.Y. Cao, and J.H. Yang, "Advances in fuzzy geometric programming," In *Fuzzy Information and Engineering: Proceedings of the Second International Conference of Fuzzy Information and Engineering (ICFIE)*, Springer Berlin Heidelberg, 2007, pp. 497-502. <https://doi.org/10.1007/978-3-540-71441-5>.
- [35] M. Khan, M. Zeeshan and S. Iqbal, "Neutrosophic variational inequalities with applications in decision-making," *Soft Computing*, 26(10), 2022, pp.4641-4652. <https://doi.org/10.1007/s00500-022-06956-6>.
- [36] A. Khorsandi, B.Y. Cao, and H. Nasseri, "A New Method to Optimize the Satisfaction Level of the Decision-maker in Fuzzy Geometric Programming Problems," *Mathematics*, 7(5), 2019, p.464. <https://doi.org/10.3390/math7050464>.
- [37] C.R. Bector and S. Chandra, "Fuzzy mathematical programming and fuzzy matrix games," Vol. 169, Berlin: Springer, 2005. <https://doi.org/10.1007/3-540-32371-6>.

Multifaceted Sentiment Detection System (MSDS) to Avoid Dropout in Virtual Learning Environment using Multi-class Classifiers

Ananthi Claral Mary. T¹, Arul Leena Rose. P. J^{2*}

Department of Computer Science-College of Science and Humanities,
SRM Institute of Science and Technology, Kattankulathur - 603 203, Chengalpattu, Tamil Nadu, India

Abstract—Sentiment analysis with machine learning plays a vital role in Higher Educational Institutions (HEI) for decision making. Technology-enabled interactions can only be successful when a strong student-teacher link is established, and the emotions of students are clearly comprehended. The paper aims at proposing Multifaceted Sentiment Detection System (MSDS) for detecting sentiments of higher education students participating in virtual learning and to classify the comments posted by them using Machine Learning (ML) algorithms. Present research evaluated a total of n=1590 students' comments with the presence of three specific multifaceted characteristics each providing 530 comments to perform Sentiment Analysis (SA) for monitoring their sentiments, opinions that facilitate predicting dropout in virtual learning environment (VLE). This begins with the phrase extraction; then data pre-processing techniques namely digits, punctuation marks and stop-words removal, spelling correction, tokenization, lemmatization, n-grams, and POS (Part of Speech) are applied. Texts are vectorized using two feature extraction techniques with count vectorization and TF-IDF metrics and classified with four multiclass supervised ML techniques namely Random Forest, Linear SVC, Multinomial Naive Bayes, and Logistic Regression for multifaceted sentiment classification. Analyzing students' feedback using sentiment analysis techniques classifies their positive, negative, or even more refined emotions that enables dropout prediction. Experimental results reveal that the highest mean accuracy result for device efficiency, cognitive behavior, technological expertise with cloud learning platform usage were achieved by Logistic Regression with 98.49%, Linear SVC with 93.58% and Linear SVC with 92.08% respectively. Practically, results confirm feasibility for detecting students' multifaceted behavioral patterns and risk of dropout in VLE.

Keywords—Sentiment analysis; opinions; TF-IDF; n-gram; virtual learning; machine learning; NLTK; text pre-processing

I. INTRODUCTION

The education system is crucially dependent on students' academic progress. The tremendous volume of data in educational databases has made it increasingly difficult to predict student performance. Low performance students may encounter several challenges, such as delayed graduation and dropping out. To quickly assist students who are performing poorly, educational institutions should regularly monitor the academic development of their students. One way to accomplish that is to use students' academic achievement prediction. The proposed method is based on hybrid approach

of sentiment analysis to achieve quality education [1]. With the explosive growth of internet, digital technologies, IT infrastructure cloud-based online learning is growing at a quicker pace. Cloud computing technologies facilitate virtual platforms and assist students in favorable paths despite the barriers. The platform inherits countless benefits namely reducing installation cost, high storage possibility, virtualization, security, easy access, etc. T. Zarra et al. states that academic institutions suffer from general budget cuts and the growing number of students. Implementation of cloud-based e-learning platforms is very much motivated. The application of cloud community assists members of education institution to work together. The research aims to propose a web service which analyzes exchange of information between learners from various universities connected on cloud using sentiment analysis [2].

The key contribution of this research is to analyze and deal with a huge amount of online data during the virtual learning course for predicting dropouts with the help of sentiment analysis of learners' comments. SA is receiving much attention in HEI for predicting student dropout. By applying SA in the context of virtual learning, provides better insights of understanding learners' attitudes, emotional reactions towards entities, events, attributes, and the notion of sentiment predictors of dropout as this reveals specific patterns in learners' behaviors that can be of practical importance for course designers and instructors. The learning process in virtual courses dramatically differs from face-to-face courses, as the courses are taken exclusively online hence, in this proposed system, it is appropriate to detect students' text sentiments composed of multifaceted characteristics that helps in identifying their risk of college dropout. One of the formidable tasks is systematical surveying of students' opinions and perceptions, as there is huge volume of opinionated text and sentiment detection of real-world data is full of challenges. It's difficult for humans to extract sentences with sentiments, read, review, and classify them into usable formats. Automated sentiment discovery and summarization system are thus required. SA is classified as positive, negative, or neutral, that employs automatic process of text classification for opinion mining and finding wide array of sentiments expressed by learners explaining their personal attitudes and evaluation regarding provided services that determines high and low sentiment scores of their multi-dimensional characteristic phenomena. Thus, this research

aims at investigation of sentiment analysis based on learners' comments from VLE to avoid dropouts. The remnant paper is organized as follows. Section II discusses the background of SA with related research works. Section III specifies the explanation of research methodology used in this research. Section IV explains MSDS architecture in detail. Section V represents the experimental outcomes. Section VI discusses conclusion and future scope of research.

II. LITERATURE REVIEW

MOOC based teaching is implemented in traditional curriculum and educational practices. Sentiment analysis approaches can be used for course content evaluations that deliver extensive intimations to course designers and educators for evaluating courses periodically and introduce probable enhancements [3]. SASys (Sentiment Analyzes System) framework built on lexical approach and polarized frame network is proposed. Its primary objective is identifying early students' risk of dropout by emotional state detection. Text sentiment is essential for defining learners' motivational profile that depends on their activities in VLE. Students' class engagement can be identified by analyzing their frequency of access data and interactions [4].

Students' motivation to learn is greatly influenced by emotional elements, and in online education, emotions can be inferred from textual discussion while giving responses. Natural Language Processing methods can be used for emotional classification of students by extracting information from Whatsapp and organizing them into corresponding categories. RNN algorithm increases the accuracy by 75% for students' emotion analysis in online learning [5]. Emotion mining and SA were performed by gathering Arabic tweets regarding online education during pandemic. Results disclose that the proposed method performs efficiently in identifying peoples' opinion on online learning in context of the pandemic using SVM with greatest accuracy of 89.6%. By considering emotion analysis, anger is the top emotion. Most significant reasons behind negative sentiments were lack of face-to-face interaction, breakdown of network, ambiguity, and games [6]. Academic issues are one of the factors which cause stress or depression among students. For example, a drop in grades, a fear of failing, and challenging competitions. Most students are young, and when they encounter difficulties, they occasionally lack wisdom and may cause harm to themselves. Hence, advice and support from professionals, family members, experts, and others are crucial for preventing adolescents from engaging in risky behavior. Sentiment analysis using Naive Bayes algorithm classifies students into stress and depression [7].

In recent years, due to generation of voluminous data, technologies have been developed for storing and data processing effortlessly. A huge amount of data can be obtained, mined, and realized for sentiment analysis. This helps in making better policies for the education sector and the target users are teachers, students, and educational organizations. In every aspect of teaching-learning approach, education sectors can incorporate sentiment analysis extensively [8]. Learners are more interested in "Engineering and Technology" but hold negative thinking towards "Life

Science and medicine", this research infers that it is an emergency to explore SA systems for cross-domain that accelerates SA application in multiple learning domains. SA is used for enhancing the learning process; additionally, it provides valuable information for educational institutions. Students receive results of SA through visualization tools such as word clouds, dashboards, virtual agents, etc [9]. Technological advancements like Blockchain, IoT, Cloud Computing and Big Data have broadened applications of SA permitting this to be utilized in any discipline [10]. Moreover, in Machine Learning and Natural Language Processing, SA has turned out to be a hot trend and is being accepted extensively all over the globe [11].

The primary objective of implementing SA approaches on students' feedback in an online learning system is to identify learners' emotions, feelings, participation and evaluate educators' performance [12]. From student learning and achievement, emotions are inseparable. Responses were collected from students to discover their emotional experience around test and an online quiz. Greater positive level emotions and lesser negative level emotions were experienced in online quizzes than tests. Future guidelines of research must integrate a complex relation between cognitive, emotional, and motivational aspects of learning [13]. Sentiment analysis accurately portrays students' learning circumstances in the online learning community. To improve quality of teaching practice, the proposed model can recognize students' sentiment tendencies. SSM (Sentiment Score Matrix) is formed to compute sentiment scores. This can efficiently identify sentiment tendencies of learners for enhancing information services quality in education practice [14].

HEI are increasingly looking for best ways to understand the learning experience of their students. Sentiment analysis helps to investigate emotions and attitudes of students regarding their course experience. To analyze sentiment text was fed into Google's Cloud-based Natural Language Processing. Results presented that students' sentiment in online interaction during two online courses is more positive than in face-to-face courses [15]. During COVID-19 pandemic, to support remote and distance learning useful learning resources are lecture recordings. Students with illness, learning disabilities, and work commitments have narrated that availability of lecture recordings has shaped an inclusive education setting. Sentiment analysis was conducted using Microsoft Azure cognitive services text analytics API. With a large text dataset, machine learning was employed that were labeled for sentiments along 0 and 1. Findings depicted that lecture recordings serve as an additional resource for preparing notes or exams [16]. Therefore, to solve real world problems through design and development of smart learning environment, we can use Azure cognitive services, a text analytics API that leverages natural language processing capabilities for deploying high-quality AI models [17].

Sentiment Analysis is the most crucial area in text mining, as the thoughts of several individuals are analyzed and compiled as a single dataset. E-learning is an educational attempt to deliver knowledge through computers. SA helps users to easily classify their emotional input information. Students' anti-course feelings can be tracked that serves as

feedback from online learning sites [18]. In the teaching-learning process virtual learning environments (VLE) deliver a set of communication and interaction tools utilized by learners and educators. Researchers presented the SentiEduc framework that uses Multi-Agent System (MAS) to gather and analyze opinion of texts posted by learners in VLE. SenticNet tool was used to analyze sentiments automatically. Educators with tutoring experience utilized the framework with real data for verifying the efficiency; as a result accuracy obtained was 73.88% [19]. From an organization, probable text reviews collected on 270 training programmes by 2688 participants were analyzed. RapidMiner Text Mining package was used to track tokenization, removal of stop words, stemming, and token filtering. Authors suggested that instead of content delivery and faculty expertise, the proposed approach can further be expanded to establish sentiment expressed over several aspects like internet connection, hospitality [20].

Researchers have developed a web-application system that uses text analytics and SA for providing educators with a deeper analysis of learners' feedback to assess the course they have taught that will enhance the students learning experience. Feedback was grouped into positive, negative, and neutral. The result depicted a larger number of neutral sentiments. Their system implementation was successful, and it significantly benefits students, lecturers, and administrators [21]. Twitter is the popular free social networking channel. In Anadolu University open and distance education system, sentiment analysis for learners was performed by fetching tweets. 400 tweets were used for validation and classification outcomes are presented. The negative feelings and student complaints can be concentrated by institution managers [22]. Numerous educational institutions utilize online education as media learning where each piece of media, maybe audio, video, or text, can accept learners' feedback. The lecture intends to realize emotions which learners experience when they access media, namely happiness, unhappiness, or disappointment and educators intend to recognize their delightfulness. The study developed a utility cellular for emotion detection from column comments in online media. Accuracy of mobile application utility is 70% for emotion detection [23]. Humans are easily prone to errors in interpreting text-based emotions. Four supervised ML classification techniques namely MNB, SVM, DT, and KNN were applied for analyzing basic emotions. The best performance was resulted by Multinomial Naive Bayes classifier with an average accuracy of 64.08% [24]. In the form of tweets, blogs, and updates of thoughts about interest, a lot of data is being produced. On a variety of subjects, including products, movies, politics, education, news, and more, people express their thoughts and ideas. Data analysis is useful to comprehend observations, sentiments, and attitudes of society. Additionally, decision-making would benefit more from such analysis. Naive Bayes, RF, Tailored RF and enhanced XGBoost were employed. Enhanced XGBoost achieved better accuracy of 72.26% [25].

With meticulous analysis of previous existing research, the limitations are they do not consider hidden structural features namely internet connection, mixed-emotional elements, and unstructured data. To address these limitations, the existing

methods can be operationalized and extended by considering specific factors namely device efficiency, cognitive behaviors, and technical familiarity with cloud platform usage. Moreover, they infer that there is an emergency to investigate SA systems for cross-domain which accelerates application of SA in multiple learning domains. Furthermore, they suggest that future researchers must integrate relationships between cognitive, emotional, and motivational learning aspects. In this context the present research bridges gap in students' sentiment detection, giving a new definition of multifaceted characteristics of concept to gain a deeper understanding of how sentiments provide an indication to educators for identifying whether a student is motivated or discouraged with virtual learning and has an intention to dropout. Thus, in this research sentiments are explored as a process that truly reflects students' learning circumstances by considering necessary features in VLE across multiple disciplines. Hence, in higher education context, sentiments are linked to students' cognitive, emotions, psychological and learning factors in students' behavior. The model is explored with four multiclass classification algorithms, to perform sentiment analysis on text and key phrases. After experimentation, the proposed MSDS system successfully outperforms in each case compared to other methodologies. The best performance for mean accuracy rate was achieved by Logistic Regression for device efficiency with 98.49%, Linear SVC for cognitive behavior and technological expertise with 93.58% and 92.08%.

III. RESEARCH METHODOLOGY

The main objective of this research is detecting the students' intention of dropping out from virtual learning courses by considering their text sentiments. There are several methods for dropout detection. These methods are based on ML techniques and require students' activity records for training and creating predictive models depending on the features extracted from raw data. In this research, four machine learning algorithms are implemented, and its performance metrics are evaluated. The subsequent research questions are considered:

- 1) To develop a Multifaceted Sentiment Detection System (MSDS) architecture for predicting dropout students in VLE.
- 2) To estimate the evaluation steps for measuring the proposed architectures' efficiency in terms of mean accuracy rate and standard deviation using four multi-class classification algorithms namely RF, Linear SVC, MNB, LR.
- 3) To investigate the most suited ML algorithm for classifying students' sentiments depending on their multifaceted characteristics and visualize the results to educators as an early intervention.

IV. MULTIFACETED SENTIMENT DETECTION SYSTEM (MSDS) ARCHITECTURE

The MSDS framework detects the students' sentiments with information gathered from VLE interactions. MSDS architecture attempts to address sentiment analysis approach for characterizing the sentiments across multifaceted subjects namely device efficiency, cognitive behavior, technological expertise with cloud platform usage. The research flow begins with data extraction from comments posted by learners of

VLE. The system reads the data stored with .csv (Comma Separated Values) format. Next, data pre-processing techniques are implemented to texts, namely removal of punctuation marks, stop words, etc. Feature extraction methods like count vectorizer, Term Frequency and Inverse Document Frequency (TF-IDF) are employed. To identify the

real opinions of learners, sentiment classification is performed based on machine learning approaches. Then the flow ends with detecting students' sentiments with three polarities: positive, negative, and neutral. The model is assessed with various evaluation metrics namely accuracy, recall, precision, and F1-Score. Fig. 1 depicts proposed (MSDS) model.

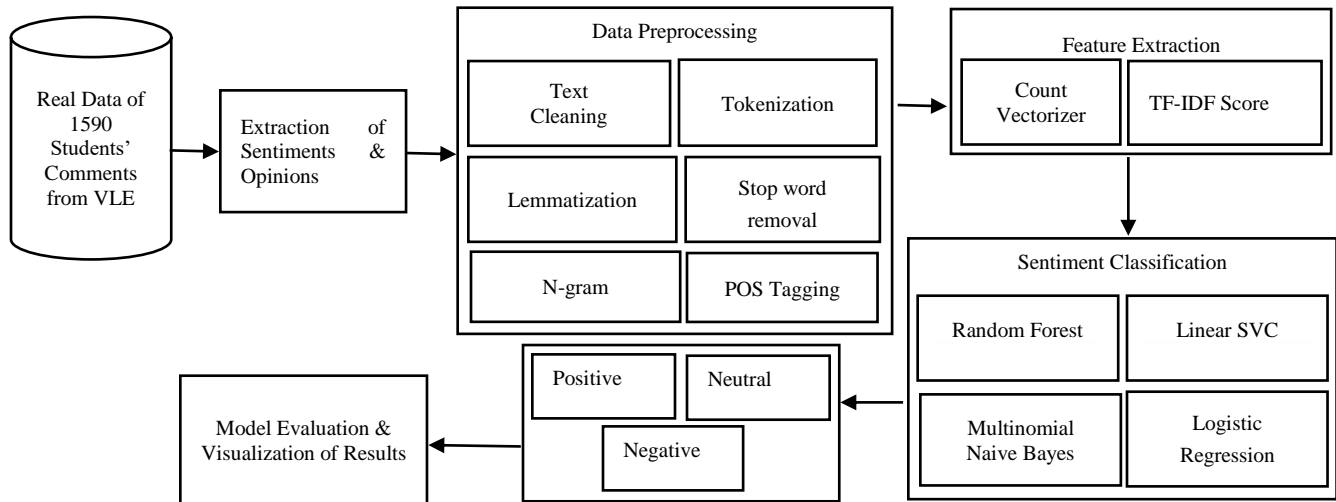


Fig. 1. Framework of proposed (MSDS) research model.

A. Data Collection from Participants

SA can be integrated into a virtual learning environment that realizes real-time analysis of learners' feedback. Real data was collected from students belonging to various disciplines of numerous HEI throughout India. The data sources comprise dataset that were collected by educators through online questionnaire that was broadly classified into various factors namely demographic features, device usage characteristics, self-efficacy which was estimated with 5-point Likert scale, familiarity with cloud platforms using 4-point range for identifying sentiments based on their interactions with VLE. Totally, the dataset contains n=1590 comments, opinions and feedbacks posted by students of virtual learning classes. Depending on their multifaceted characteristics, three datasets are grouped each containing 530 comments for conducting experiments. The data instances are branched into multiple classes. The sentiments are labelled with score range from 0-2, representing negative, neutral, and positive.

B. Data Pre-processing

Text pre-processing is the process of cleaning and preparing text data, as machines can use that processed text to perform various tasks like analysis, prediction, etc. These procedures help to reduce the volume of data and processing time. The comments posted by students are in the form of natural English language and this must be converted into machine readable format. With text data, there are many challenges as it contains a lot of noisy, semi or unstructured data, punctuations, numbers, special characters, spelling mistakes, etc that require to be processed with NLP techniques. These are the pre-processing steps carried out for improving prediction performance.

1) *Removing punctuations and numbers, converting all characters into lowercase*: The basic pre-processing method is punctuations removal from textual data. This process helps to treat each text similarly. As numbers do not hold any vital information in the text it has to be removed. Then the input text is converted into same casing format namely lowercase.

2) *Tokenization*: This is a method of breaking the sentence into meaningful words and phrases. This process is achieved by using delimiters like white spaces and punctuation. Inbuilt NLTK (Natural Language Tool Kit) libraries have tokenization function capability to divide into words.

3) *Lemmatization*: Lemmatization is a text pre-processing technique in NLP models to break a word to its root meaning for identifying similarities. Text normalization changes the words down to a base form. WordNetLemmatizer() is assigned to a variable which is used to improve the algorithms' performance and facilitates to focus on meaning of the words.

4) *Stop word removal*: Next step is to eliminate stop words as they are commonly used words yet not useful for analysis and generally eliminated from text. This technique is applied for removing stop words namely I, am, you, she, he, the, a, an, so, what, etc. These words are not required for sentiment classification as it increases the dataset size and is considered irrelevant for results set. It's always better to remove these words for improving the model accuracy, reducing computational processes and complexity of data.

5) *N-gram*: In NLP this is a continuous series of n items created from a given text sample where items can be characters or words. This language model predicts the probability within any sequence of n words in the language.

There are multiple n-grams like unigram, bigram, trigram etc. Unigram is simplest n-gram where only one word is considered. Bigram takes two words at a time. Unigrams and bigrams are used in this research, where ngram_range = (1,2) is considered for feature processing.

6) *Part of Speech (POS) Tagging*: This process labels each word in text format for a particular POS depending on its context and definition. This reads the text from a language and assigns some token (POS) for every word. `nlk.pos_tag(tokenized_text)` is applied. Parts of Speech Adjective, Verb, Noun, and Adverb are considered.

C. Feature Extraction Techniques

In text data classification feature extraction plays a dominant role in reduction of feature space and increasing classifier's accuracy. This process converts data into features applied for ML model, as ML algorithms are programmed with numbers. Textual data is converted into vector form using two feature extraction techniques namely Count Vector, Term Frequency, and Inverse Document Frequency (TF-IDF) vector. Fig. 2 represents the cleaned text results achieved after data pre-processing and feature extraction with cognitive behavior data.

	text	label	clean_text
0	I am Undecided to manage to solve difficult pr...	2	[undecided, manage, solve, difficult, problem,...
1	I am Agree to manage to solve difficult proble...	2	[agree, manage, solve, difficult, problem, agr...
2	I am Agree to manage to solve difficult proble...	2	[agree, manage, solve, difficult, problem, agr...
3	I am Agree to manage to solve difficult proble...	2	[agree, manage, solve, difficult, problem, dis...
4	I am Undecided to manage to solve difficult pr...	2	[undecided, manage, solve, difficult, problem,...

Fig. 2. Sample processed text for cognitive behavior.

D. Machine Learning Techniques for Sentiment Analysis

In the current situation, feedback is provided through grading methods. Although this grading method masks students' genuine feelings, the textual response gives them a chance to emphasize qualities. Three different ML algorithms SVM, MNB, and RF were implemented. Experimental outcomes suggest that MNB with 80% accuracy performs better than other classifiers [26]. The study applies SA to self-evaluation comments, a form of unstructured data that provides valuable information representing students' learning status over course duration for identifying at-risk students. SVM and Convolutional Neural Networks (CNN) were

1) *Count vectorizer*: CountVectorizer transforms a given text into a vector depending on frequency (count) of each word which occurs in entire text. This is beneficial when there are multiple such texts, and for converting each word in each text into vectors. These functionalities make it an extremely adaptable feature description module for text.

2) *TF-IDF score*: This is a weighting measure that quantifies the string relevance representations namely words, phrases, lemmas in a given document. Term Frequency (TF) describes in a document how frequently a term occurs against the total number of words in document, as in (1). IDF is measurement of selected term's weight in the document. This is given in (2).

$$tf = \frac{\text{number of occurrences of a term in document}}{\text{total number of terms in the document}} \quad (1)$$

$$idf = \log \left(\frac{\text{total number of given documents}}{\text{number of document with existing selected word}} \right) \quad (2)$$

applied to predict student performance. The proposed model provides an effectiveness represented by F-measure of 0.66 (SVM) and 0.78 (CNN). Best effectiveness was presented by CNN, achieving an F-measure value of 0.78. Experimental results demonstrated that applying sentiment analysis to unstructured data can significantly improve accuracy of early-stage predictions [27]. In this research, NLP techniques are employed to pre-process and vectorize the text data. Vectorized data is then applied for training various ML models. Following are the list of steps required to train the model for sentiment analysis:

Algorithm: Sentiment Analysis Using Machine Learning

1. Collect dataset from VLE to train and test machine learning model classifier.
 2. Pre-process the data for subsequent processing.
 3. Convert textual data into vector form using NLP techniques namely Count Vectorizer and TF-IDF.
 4. Divide the dataset into training and testing groups.
 5. Train the ML classifier with training data. Apply algorithms such as RF, Linear SVC, MNB, LR.
 6. Predict the polarity of test data.
 7. Evaluate the ML model using various metrics namely accuracy, precision, recall and F1-Score.
 8. Select the best algorithm for multifaceted text characteristics using sentiment multiclass classification.
-

1) *Random forest*: SA is used for analyzing unstructured text data for extracting positive or negative sentiments contained in student advisor's notes to predict college student dropout using RF model. The authors have quoted that their study is the first to apply NLP techniques for dropout prediction. RF classifier achieved 73% accuracy when compared to SVM, LR and CART [28]. M. A. Fauzi stated that due to the development of social media and online website reviews, SA is an efficient way for text classification. Experimental results confirmed that RF gives excellent performance with an average OOB score of 0.829 [29]. For increasing predictive power of Random Forest, this research utilizes hyperparameters $n_estimators=100$ and $max_depth=5$. $n_estimators$ is the number of trees the algorithm creates before taking maximum voting. A higher number of trees enhances the performance and makes predictions more stable. The hyperparameter max_depth represents maximum depth of each decision tree in the forest.

2) *Linear Support Vector Classifier (SVC)*: [30] suggests an opinion analysis system for amazon review for identifying comments received from UCI Website either positive or negative. The proposed approach is applied to review dataset and obtained an accuracy of 91% with Linear SVC over Naive Bayes and Voting. Authors have proposed that Linear SVC takes lesser execution time of about 0.0972s to test samples and provides output which is better than other classifiers namely Naive Bayes, Logistic Regression, Decision Tree [31]. The Linear SVC method is a faster implementation of Support Vector Classification that applies a linear kernel function for performing classification. This performs well for NLP based text classification tasks with a large number of samples. Linear SVC in scikit-learn library does not provide `predict_proba` function, instead `decision_function` is used that predicts the confidence scores for samples, which is the signed distance of that sample to hyperplane.

3) *Multinomial Naive Bayes (MNB)*: This is a type of Naive Bayes (NB) classifier that finds probabilities of classes assigned to text that uses joint probabilities of words and classes and is often used as a baseline for sentiment analysis. MNB achieves significant results for text categorization with 90% accuracy. MNB algorithm is a fast, easy-to-implement, and modern text categorization algorithm [32]. Social networking has developed into a tool that is useful for gathering vital information about individuals. The sentiment of the user can be determined by extracting user comments via an API and feeding them to an algorithm that detects whether they are positive or negative. Results obtained show that Multinomial Naive Bayes performs good with classification accuracy of 85% when compared to SVM, Random Forest and Decision Tree [33]. MNB is specifically beneficial for problems that involve text data with discrete features, namely word frequency counts. This works on the principle of Bayes theorem and assumes that features are conditionally independent given the class variables. The computation is

performed by adding logarithms of probabilities, as in (3). The class with the highest log probability score is most feasible.

$$C_{NB} = \mathop{\text{arg}}_{c \in C} \max [\log P(c) + \sum_{1 \leq k \leq n_d} \log P(t_k | c)] \quad (3)$$

4) *Logistic regression*: In a variety of situations finding the polarity of reviews is useful. NLP techniques can be applied for analyzing the reviews and optimizing the strategic decision making. TFIDF is used for feature selection and Logistic Regression for classification as it uses sigmoid activation function. The LR with grid search model classifies the text accurately with 94% [34]. Reviews and feedback are deciding factors for understanding the opinions of users. SA is a method of information extraction for improving the work by review analysis. The users' review text is cleaned by Count Vectorizer and TFIDF. Sentiment prediction is done using various classifiers for reviews without ratings. Logistic Regression accuracy is 93% and more compared to NB, MNB, Bernoulli classifiers [35]. Logistic Regression is a simple classification algorithm that can be generalized to multiple classes. The LR uses sigmoid function, which is given by (4),

$$\text{sig}(t) = \frac{1}{1+e^{-t}} \quad (4)$$

These popular classifiers do not directly support multiclass classification problems. There are certain heuristic methods available which can split multiclass classification datasets into several binary classification datasets. The binary classifiers are then trained on each binary classification problem and predictions are made operating the model which is most confident. To implement this method for multi-class classification, the `OneVsRestClassifier` method is used.

V. EXPERIMENTAL RESULTS

In this research, implementation is done with efficient NLTK, Scikit-learn libraries by applying the proposed machine learning techniques. NLTK is used to perform regular expression patterns & tokenization to parse text, lemmatization, stop word removal, n-grams, and POS tagging. Vectorization and classification were accomplished by Scikit-learn. Data was preprocessed, vectorized with TF-IDF, and classified with four multiclass machine learning algorithms. Dataset was split into 75% training set and 25% test set. 5-fold Cross-Validation was used for training classification model. RF, Linear SVC, MNB and LR were applied as they are the most popular machine learning classifiers used to analyze students' opinions. Depending on the classification task different metrics were used to measure the classifiers' performance.

A. Multi-class Classification Positive/Neutral/Negative

The performance of MSDS framework is evaluated using various metrics. Accuracy is one of the popular metrics for multi-class classification. This is the ratio of the number of accurately classified instances to total number of instances. Macro-average precision calculates precision for all classes individually and then averages them. Weighted average precision calculates precision per class but considers no. of

samples of each class in data. Macro-average recall score computes the arithmetic mean of all recall scores of different classes. Weighted-average recall computes recall per class but considers the number of samples of each class in the data. Macro average F1score computes arithmetic mean of all per-class F1 scores. Weighted average F1 score calculates mean of all per-class F1 scores by considering each class's support.

$$Accuracy = \frac{(TP+TN)}{(TP +TN+FP+FN)} \quad (5)$$

$$Macro - average Precision = \frac{Precision_1+Precision_2+Precision_3}{3} \quad (6)$$

$$Weighted - average Precision = \frac{w_1*Precision_1+w_2*Precision_2+w_3*Precision_3}{Total\ number\ of\ samples} \quad (7)$$

$$Macro - average Recall = \frac{Recall_1+Recall_2+Recall_3}{3} \quad (8)$$

$$Weighted - average Recall = \frac{w_1*Recall_1+w_2*Recall_2+w_3*Recall_3}{Total\ number\ of\ samples} \quad (9)$$

$$Macro - average F1 - Score = \frac{F1-Score_1+F1-Score_2+F1-Score_3}{3} \quad (10)$$

$$Weighted - average F1 - Score = \frac{w_1*F1-Score_1+w_2*F1-Score_2+w_3*F1-Score_3}{Total\ number\ of\ samples} \quad (11)$$

1) *Device efficiency*: The first perspective analyzes the students' device usage characteristics as VLE increases the portability of learning processes through smart devices. To investigate new educational opportunities that result from expanding device access to VLE, enables users to establish more fleeting links to virtual campus for providing instructional procedures built on a model that is significantly less time and space consuming. The parameters used for configuring each of the algorithms implemented are the type of smart device, mode of device availability, device connectivity, number of hours the device is connected online. The optimized smart device environment will provide services to the educational community, regardless of their functional and cognitive elements. N. A. S. Remali with other researchers, states that its' a challenging task for educational institutions to identify learners' opinions and difficulties during online education as few students may have poor internet connection problem, lack of bandwidth and the environment makes learners not to concentrate on class. Thus, SA is utilized to assess students' opinions (positive, neutral, and negative) of virtual learning [36]. Considering device efficiency features, Table I shows the evaluation metrics of four multi-class classifiers with LR model showing high accuracy rate of 98%. The cross-validation method, with cv=5, was performed to cross-validate baseline models with feature extractors of TF-IDF and CV. Mean accuracy and standard deviation for each fold validates performance of

models, depicted in Table II. The highest mean accuracy of Logistic Regression is 98.49%. Fig. 3 shows mean accuracy with 5-fold cross validation. Results of sentiment distribution are visualized in Fig. 4. Fig. 5 displays classification metrics of device efficiency text characteristics.

TABLE I. CLASSIFICATION OF DEVICE EFFICIENCY

Evaluation Metrics of Classifiers	RF	Linear SVC	MNB	LR
Accuracy	0.96	0.96	0.97	0.98
Precision-macro avg	0.58	0.58	0.60	0.61
Precision-weighted avg	0.95	0.95	0.96	0.97
Recall-macro avg	0.66	0.66	0.66	0.66
Recall-weighted avg	0.96	0.96	0.97	0.97
F1-score macro avg	0.61	0.61	0.63	0.64
F1-score weighted avg	0.96	0.96	0.96	0.97

TABLE II. MEAN ACCURACY AND SD

ML Models	Mean Accuracy	Standard Deviation
Random Forest	0.981132	0.011554
Linear SVC	0.983019	0.013993
Multinomial Naive Bayes	0.981132	0.006671
Logistic Regression	0.984906	0.010756

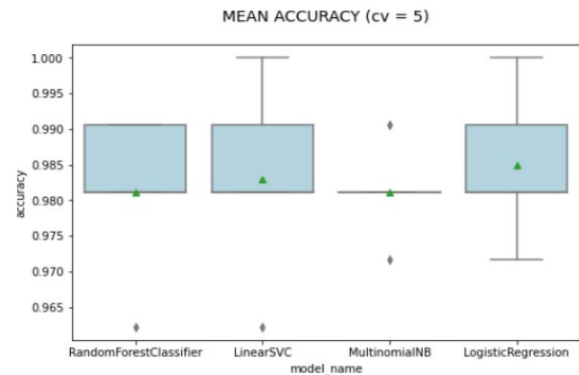


Fig. 3. Mean accuracy for device efficiency.

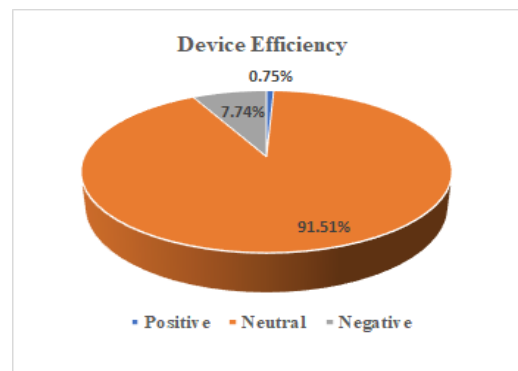


Fig. 4. Device efficiency sentiment polarity distribution.

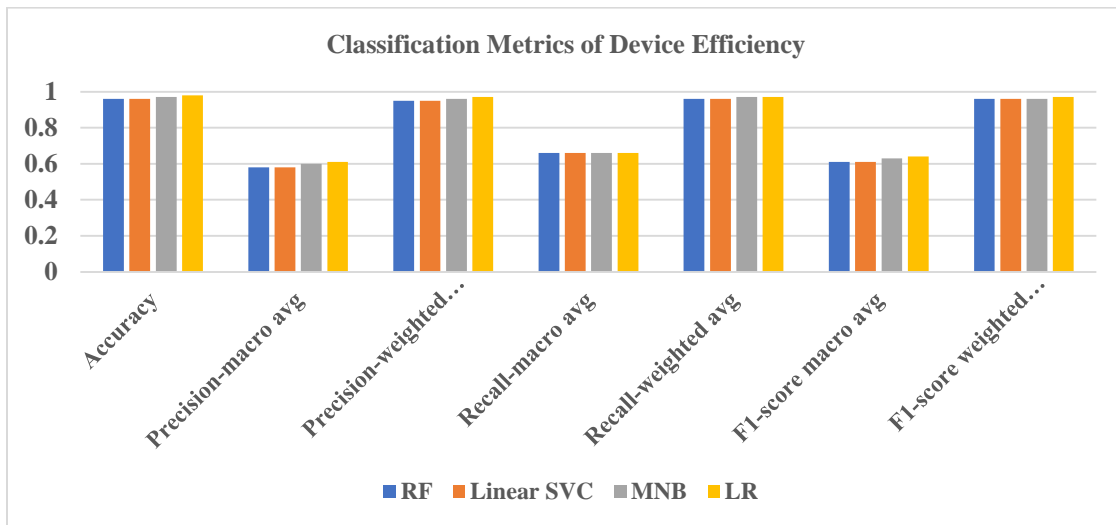


Fig. 5. Classification metrics of device efficiency text characteristics.

2) *Cognitive behavior*: The encompassing perspective analyzes the students’ psychological cognitive behavioral model through self-efficacy theory stated by Psychologist Albert Bandura. The parameters used for configuring each of the algorithms implemented are manageability, finding means & ways, stick to aims & accomplishments, handling unforeseen situations, investing effort, finding several solutions, handling whatever situations in online learning. In addition, the researchers in [37] explored that predictive power and feature generalization of cognitive skill score estimates possibility of learners’ success or failure in higher education course which provides suitable intervention to facilitate learners. Cognitive skill score proves to be efficient in identifying students’ performance when exact metrics correlated to learning activities and students’ social behavior are unavailable. Table III shows evaluation metrics of four multi-class classifiers for self-efficacy with Linear SVC showing a high accuracy rate of 93%. With cv=5, mean accuracy and standard deviation for each fold validates the models’ performance depicted in Table IV. The mean accuracy of Linear SVC is higher with 93.58% when compared to RF, MNB and LR. Fig. 6 presents mean accuracy with cv=5 for cognitive behavioral features. Sentiment analysis outcomes are visualized in Fig. 7. Classification metrics of self-efficacy text characteristics are displayed in Fig. 8.

F1-score macro avg	0.61	0.62	0.57	0.59
F1-score weighted avg	0.91	0.93	0.87	0.90

TABLE IV. MEAN ACCURACY AND SD

ML Models	Mean Accuracy	Standard Deviation
Random Forest	0.932075	0.018147
Linear SVC	0.935849	0.010334
Multinomial Naive Bayes	0.900000	0.023679
Logistic Regression	0.932075	0.012300

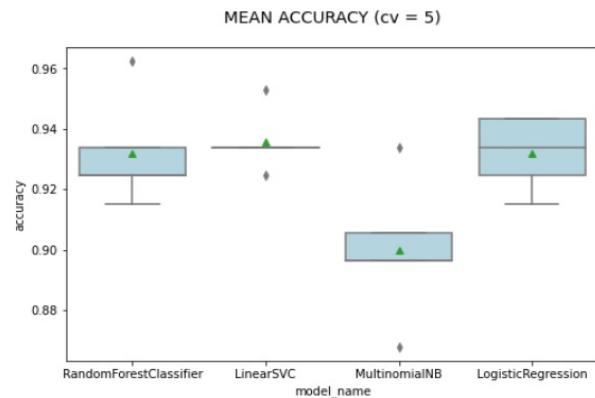


Fig. 6. Mean accuracy for behavioral features.

TABLE III. CLASSIFICATION OF SELF-EFFICACY

Evaluation Metrics of Classifiers	RF	Linear SVC	MNB	LR
Accuracy	0.92	0.93	0.88	0.90
Precision-macro avg	0.61	0.62	0.59	0.60
Precision-weighted avg	0.91	0.92	0.88	0.90
Recall-macro avg	0.61	0.62	0.56	0.58
Recall-weighted avg	0.92	0.93	0.88	0.90

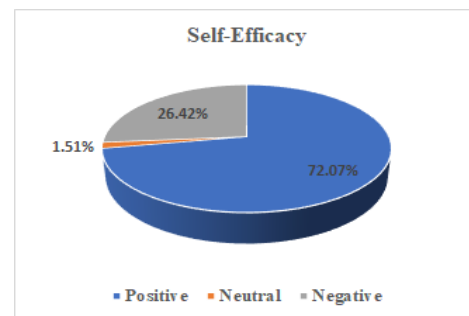


Fig. 7. Self-efficacy sentiment polarity distribution.

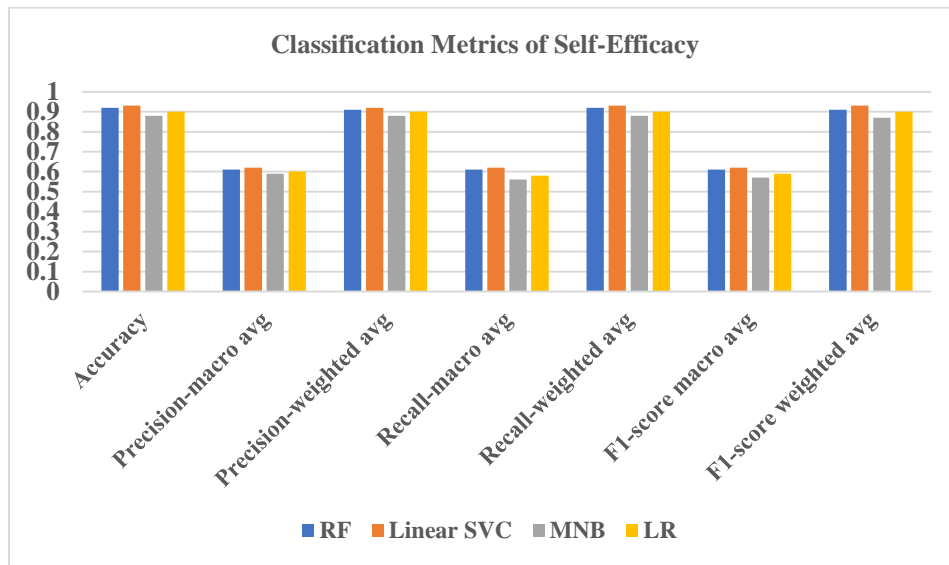


Fig. 8. Classification metrics of self-efficacy text characteristics.

3) *Technological expertise of cloud learning platform usage*: Outside the classroom, cloud platforms prepare students to make reasonable study schedules, thereby promoting self-directed learning, innovation, collaboration, and ease of accessibility. The parameters used emphasize on the platform learners are familiar with accessing online teaching materials namely Google Classroom, Google Meet, Zoom, Facebook Classroom, Twitter etc. Moreover, J. Zhang in [38] suggested that educators could discuss with learners through the cloud class even after online class and they can set time-limited online test for enhancing results. With students' feedback the course assistants can track their problems in online teaching. Using bullet screen educators can communicate useful course contents to learners in time, whereby the emotional communication between them is deepened. Table V illustrates the evaluation metrics of four classification algorithms for technological expertise with Linear SVC depicting a high accuracy rate of 92%. Table VI depicts mean accuracy and standard deviation for each fold to validate models' performance. The highest mean accuracy score of Linear SVC is 92.08%. Fig. 9 visualizes mean accuracy with $cv=5$ for technological familiarity with cloud platforms. The sentiment analysis results are presented in Fig. 10. Fig. 11 shows classification metrics of technological expertise text characteristics.

F1-score macro avg	0.55	0.61	0.48	0.48
F1-score weighted avg	0.88	0.89	0.87	0.87

TABLE VI. MEAN ACCURACY AND SD

ML Models	Mean Accuracy	Standard Deviation
Random Forest	0.918868	0.010756
Linear SVC	0.920755	0.008438
Multinomial Naive Bayes	0.916981	0.004219
Logistic Regression	0.916981	0.004219

MEAN ACCURACY (cv = 5)

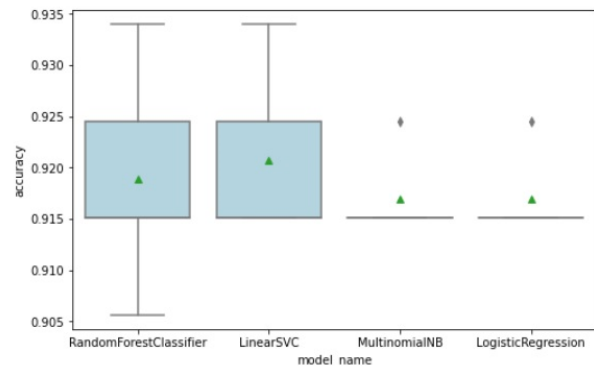


Fig. 9. Mean accuracy for technological expertise.

TABLE V. CLASSIFICATION OF TECHNOLOGICAL EXPERTISE

Evaluation Metrics of Classifiers	RF	Linear SVC	MNB	LR
Accuracy	0.91	0.92	0.91	0.91
Precision-macro avg	0.71	0.79	0.45	0.45
Precision-weighted avg	0.88	0.90	0.83	0.83
Recall-macro avg	0.54	0.58	0.50	0.50
Recall-weighted avg	0.91	0.92	0.91	0.91

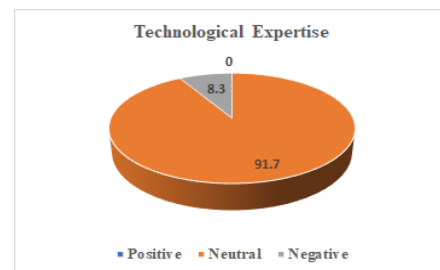


Fig. 10. Technological expertise sentiment polarity distribution.

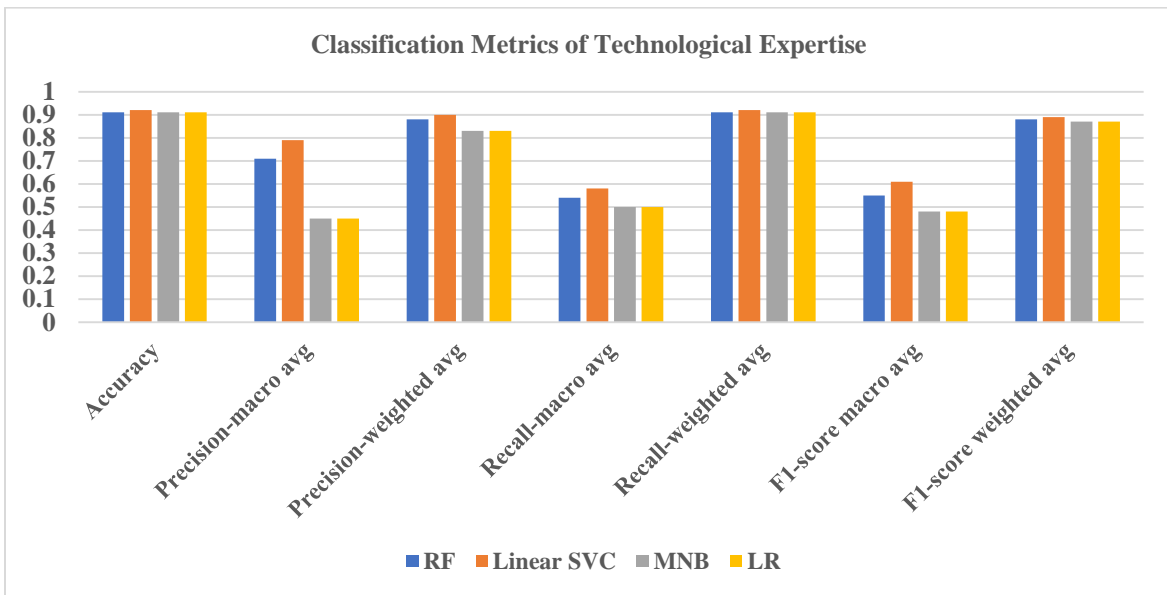


Fig. 11. Classification metrics of technological expertise text characteristics.

B. Performance Evaluation using ROC and AUC

To analyze the performance of MSDS architecture, ROC curve can be used as it measures the classifier’s predictive quality. Tradeoff between classifier’s sensitivity and specificity can be visualized by the user using a ROCAUC (Receiver Operating Characteristic/Area Under the Curve). ROC curve exhibits true positive rate on Y axis and false positive rate on X axis when plotted. Ultimate point is the top-left corner of plot: false positives are 0 and true positives are 1, which directs to another metric, AUC. The higher AUC represents a better model. However, this is vital to examine “steepness” of curve, as this illustrates maximization of sensitivity while minimizing specificity. ROC curve is extensively used in this research to describe diagnostic accuracy and for finding the best cut-off value for a model trained through multi-class machine learning techniques. ROC curves for cognitive behavior through self-efficacy characteristics using RF, Linear SVC, MNB, LR are visualized in Fig. 12, Fig. 13, Fig. 14, Fig. 15 respectively as this is main framework emphasizing the learners’ psychological cognitive behavioral patterns for risk prediction. The performance comparisons of ROC curves show that Linear SVC model performs the best for self-efficacy data.

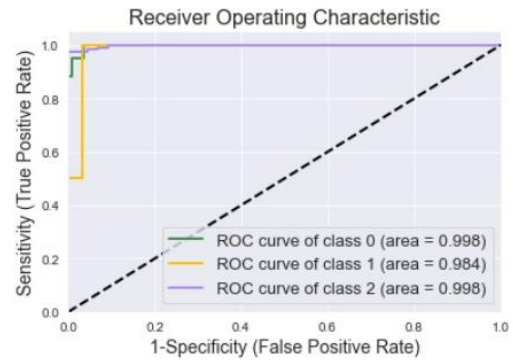


Fig. 13. ROC for linear SVC.

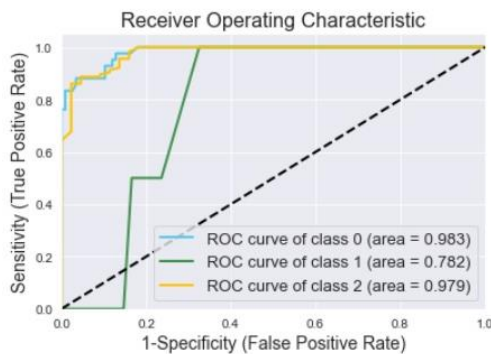


Fig. 12. ROC for random forest.

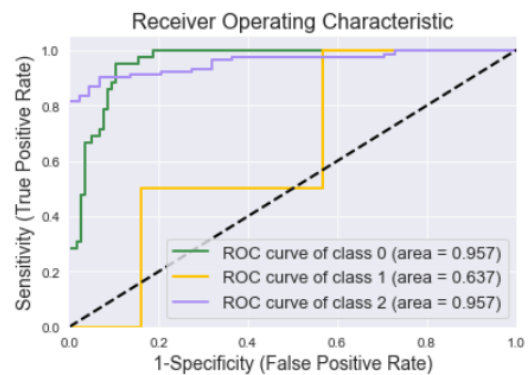


Fig. 14. ROC for MNB.

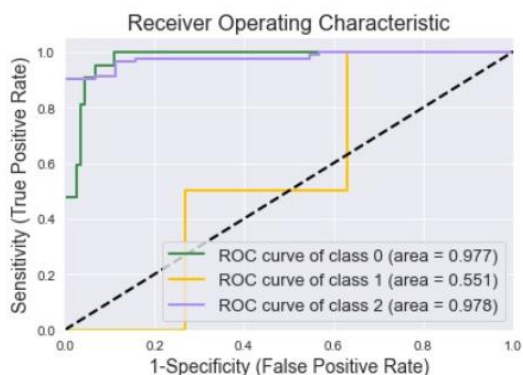


Fig. 15. ROC for logistic regression.

C. Data Visualization

Sentiment analysis encompasses a variety of SA tasks, including subjectivity detection and emotion analysis. This reflects the wide range of user tasks and data domains used in SA research and applications, which include everything from social media and news monitoring to theoretical linguistic research and NLP. This implies the use of numerous visual channels and interpretations. The visual representation involving polarity data includes word clouds [39]. Word cloud represents a powerful textual data visualization technique that enables quickly identifying the words that are most often used within a particular body of text. These are frequently employed communication tools for processing, analyzing qualitative sentiment data. When educators need to visualize the opinions of learners, word cloud can be used for identifying the messages posted by them. This visual representation gives a clear way for educators to easily interpret the messages. Fig. 16 represents the word cloud visualization of multifaceted factors.



Fig. 16. Word clouds of device efficiency, cognitive behavior, technological expertise of cloud platform Usage.

VI. CONCLUSION

Multifaceted Sentiment Detection System (MSDS) is proposed in this research to predict students' dropout using multiclass classification algorithms. For addressing this goal, the research analyzes higher education students' comments

and reviews while attending virtual classes through VLE that have been classified and analyzed using machine learning algorithms and word clouds. The results obtained from the research explore multideterminant characteristic features of respondents' device efficiency, psychological cognitive behavior, and technological knowledge of cloud platform usage. This finding mainly contributes to improving the classifiers' predictive ability of college student dropout through text classification. The proposed method obtained the highest mean accuracy results for device efficiency using Logistic Regression with 98.49%, Linear SVC for cognitive behavior with 93.58% and technological expertise with 92.08%. Thus, this novel assessment mechanism MSDS, aims in exploring efficient sentiment classification by proposing multiclass machine learning algorithms to avoid dropouts. Furthermore, experimental outcomes demonstrate that proposed system has obtained better accuracy results when comparing to previous methods. For future work, the algorithms can be validated and applied against other computational methods such as deep learning algorithms for superior performance.

REFERENCES

- [1] Pooja and R. Bhalla, "A Review Paper on the Role of Sentiment Analysis in Quality Education," *SN Comput. Sci.*, vol. 3, no. 6, pp. 1–9, 2022, doi: 10.1007/s42979-022-01366-9.
- [2] T. Zarra, R. Chiheb, R. Faizi, and A. El Afia, "Cloud computing and sentiment analysis in E-learning systems," *Proc. 2016 Int. Conf. Cloud Comput. Technol. Appl. CloudTech 2016*, pp. 171–176, 2017, doi: 10.1109/CloudTech.2016.7847695.
- [3] F. Dalipi, K. Zdravkova, and F. Ahlgren, "Sentiment Analysis of Students' Feedback in MOOCs: A Systematic Literature Review," *Front. Artif. Intell.*, vol. 4, pp. 1–13, 2021.
- [4] F. Campos, V. Stroele, R. Braga, and T. T. Torrent, "Using Sentiment Analysis to Identify Student Emotional State to Avoid Dropout in E-Learning," *International Journal of Distance Education Technologies*, vol. 20, no. 1, pp. 1–24, 2022, doi: 10.4018/IJDET.305237.
- [5] N. H. Harani and C. Prianto, "Sentiment Analysis of Student Emotion During Online Learning Using Recurrent Neural Networks (RNN)," *International Journal of Information System & Technology*, vol. 5, no. 3, pp. 299–307, 2021.
- [6] M. M. Ali, "Arabic sentiment analysis about online learning to mitigate covid-19," *J. Intell. Syst.*, vol. 30, no. 1, pp. 524–540, 2021, doi: 10.1515/jisys-2020-0115.
- [7] N. S. Izzati Samsari, M. Mohamad, and A. Selamat, "Sentiment Analysis on Students Stress and Depression Due to Online Distance Learning During the COVID-19 Pandemic," *Math. Sci. Informatics J.*, vol. 3, no. 1, pp. 66–74, 2022, doi: 10.24191/mij.v3i1.18273.
- [8] R. Archana and B. Kishore, "Role of Sentiment Analysis in Education Sector in the Era of Big Data: a Survey," *International Journal of Latest Trends in Engineering and Technology*, pp. 022–024, 2017.
- [9] J. Zhou and J. M. Ye, "Sentiment analysis in education research: a review of journal publications," *Interactive learning environments*, pp. 1-13, 2020, doi: 10.1080/10494820.2020.1826985.
- [10] M. Wankhade, A. C. S. Rao, and C. Kulkarni, "A survey on sentiment analysis methods, applications, and challenges," *Artificial Intelligence Review*, vol. 55, no. 7, pp. 5731-5780, 2022.
- [11] A. Raut and R. K. Pandey, "Sentiment Analysis using Optimized Feature Sets in Different Twitter Dataset Domains," *International Journal of Innovative Technology and Exploring Engineering*, vol. 8, no. 11, pp. 3035-3039, 2019, doi: 10.35940/ijitee.k2195.0981119.
- [12] S. Ulfa, R. Bringula, C. Kurniawan, and M. Fadhli, "Student Feedback on Online Learning by Using Sentiment Analysis: A Literature Review," *Proc. - 2020 6th Int. Conf. Educ. Technol. ICET 2020*, no. 1, pp. 53–58, 2020, doi: 10.1109/ICET51153.2020.9276578.

- [13] K. Riegel and T. Evans, "Student achievement emotions: Examining the role of frequent online assessment," *Australas. J. Educ. Technol.*, vol. 37, no. 6, pp. 75–87, 2021, doi: 10.14742/ajet.6516.
- [14] K. Wang and Y. Zhang, "Topic sentiment analysis in online learning community from college students," *Journal of Data and Information Science*, vol. 5, no. 2, pp. 1-33, 2020, doi: 10.2478/jdis-2020-0009.
- [15] T. D. Pham *et al.*, "Natural language processing for analysis of student online sentiment in a postgraduate program," *Pacific J. Technol. Enhanc. Learn.*, vol. 2, no. 2, pp. 15–30, 2020, doi: 10.24135/pjtel.v2i2.4.
- [16] L. M. Nkomo and B. K. Daniel, "Sentiment Analysis of Student Engagement with Lecture Recording," *TechTrends*, vol. 65, no. 2, pp. 213–224, 2021, doi: 10.1007/s11528-020-00563-8.
- [17] S. Pande, "An overview of sentiment analysis using Azure Cognitive Services," *International Journal of Management, IT & Engineering*, vol. 9, no. 10, pp. 123-127, 2019.
- [18] P. Rajesh and D. Akila, "Sentimental Analysis on E-learning Videos for Learner's Opinion Using Machine Learning Methodology - Support Vector Machine," *Lecture Notes in Networks and Systems*, 2023, Volume 467, pp. 191-200, doi: 10.1007/978-981-19-2538-2_18.
- [19] M. A. dos Santos Alencar, J. F. de Magalhaes Netto, and F. de Morais, "A Sentiment Analysis Framework for Virtual Learning Environment," *Appl. Artif. Intell.*, vol. 35, no. 7, pp. 520–536, 2021, doi: 10.1080/08839514.2021.1904594.
- [20] K. Ravi, V. Siddeshwar, V. Ravi, and L. Mohan, "Sentiment analysis applied to Educational Sector," *2015 IEEE Int. Conf. Comput. Intell. Comput. Res. ICCIC 2015*, no. February 2017, 2016, doi: 10.1109/ICCIC.2015.7435667.
- [21] O. Abiodun Ayeni, A. Mercy, T. A.F, and M. A.S, "Web-Based Student Opinion Mining System Using Sentiment Analysis," *Int. J. Inf. Eng. Electron. Bus.*, vol. 12, no. 5, pp. 33–46, 2020, doi: 10.5815/ijieeb.2020.05.04.
- [22] Z. Kamisli Ozturk, Z. I. Erzurum Cicek, and Z. Ergul, "Sentiment analysis: An application to Anadolu University," *Acta Phys. Pol. A*, vol. 132, no. 3, pp. 753–755, 2017, doi: 10.12693/APhysPolA.132.753.
- [23] I. D. Wahyono *et al.*, "Emotion Detection based on Column Comments in Material of Online Learning using Artificial Intelligence," *International Journal of Interactive Mobile Technologies*, vol. 16, no. 3, pp. 82–91, 2022, doi: 10.3991/IJIM.V16I03.28963.
- [24] A. F. Ab Nasir *et al.*, "Text-based emotion prediction system using machine learning approach," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 769, no. 1, 2020, doi: 10.1088/1757-899X/769/1/012022.
- [25] B. N. Supriya and C. B. Akki, "Sentiment prediction using enhanced xgboost and tailored random forest," *Int. J. Comput. Digit. Syst.*, vol. 10, no. 1, pp. 191–199, 2021, doi: 10.12785/ijcds/100119.
- [26] D. D. Dsouza, Deepika, D. P. Nayak, E. J. Machado, and N. D. Adesh, "Sentimental analysis of student feedback using machine learning techniques," *Int. J. Recent Technol. Eng.*, vol. 8, no. 1 Special Issue 4, pp. 986–991, 2019.
- [27] L. C. Yu *et al.*, "Improving early prediction of academic failure using sentiment analysis on self-evaluated comments," *J. Comput. Assist. Learn.*, vol. 34, no. 4, pp. 358–365, 2018, doi: 10.1111/jcal.12247.
- [28] J. Jayaraman, "Predicting student dropout by mining advisor notes," *Educationaldatamining.Org*, no. Edm, pp. 629–632, 2020, https://educationaldatamining.org/files/conferences/EDM2020/papers/paper_233.pdf.
- [29] M. A. Fauzi, "Random Forest Approach for sentiment analysis in Indonesian language," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 12, no. 1, pp. 46–50, 2018, doi: 10.11591/ijeecs.v12.i1.pp46-50.
- [30] S. S. Sikarwar and N. Tiwari, "Analysis The Sentiments Of Amazon Reviews Dataset By Using Linear SVC And Voting Classifier," *International Journal of Scientific & Technology Research*, vol. 9, no. 06, pp. 461–465, 2020.
- [31] N. Sultana, P. Kumar, M. R. Patra, S. Chandra and S.K. Safikul Alam, "Sentiment Analysis System for Product Review: A Survey: A Survey," *ICTACT Journal on Soft Computing*, vol. 09, no. 03, pp. 268–278, 2019.
- [32] M. Abbas, K. Ali Memon, and A. Aleem Jamali, "Multinomial Naive Bayes Classification Model for Sentiment Analysis," *IJCSNS Int. J. Comput. Sci. Netw. Secur.*, vol. 19, no. 3, pp. 62-67, 2019.
- [33] S. Bachhety, S. Dhingra, R. Jain, and N. Jain, "Improved Multinomial Naive Bayes Approach for Sentiment Analysis on Social Media," *Proceedings of 4th International Conference on Computers and Management (ICCM) 2018*, pp. 121–127, 2018.
- [34] P. S. Reddy, D. R. Sri, C. S. Reddy, and S. Shaik, "Sentimental Analysis using Logistic Regression," *International Journal of Engineering Research and Applications*, vol. 11, no. 7, pp. 36–40, 2021, doi: 10.9790/9622-1107023640.
- [35] S. P. Sheela, "Sentiment Analysis and Prediction of Online Reviews with Empty Ratings," *Int. J. Appl. Eng. Res.*, vol. 13, no. 14, pp. 11532–11539, 2018.
- [36] N. A. S. Remali, M. R. Shamsuddin, and S. Abdul-Rahman, "Sentiment Analysis on Online Learning for Higher Education During Covid-19," *2022 3rd Int. Conf. Artif. Intell. Data Sci. Championing Innov. Artif. Intell. Data Sci. Sustain. Futur. AiDAS 2022 - Proc.*, pp. 142–147, 2022, doi: 10.1109/AiDAS56890.2022.9918788.
- [37] S. Md and S. Krishnamoorthy, "Student performance prediction, risk analysis, and feedback based on context-bound cognitive skill scores," *Educ. Inf. Technol.*, vol. 27, no. 3, pp. 3981–4005, 2022, doi: 10.1007/s10639-021-10738-2.
- [38] J. Zhang, "Research on Sentiment Analysis and Satisfaction Evaluation of Online Teaching in Universities During Epidemic Prevention," *Frontiers in Psychology*, vol. 12, 2021, doi: 10.3389/fpsyg.2021.738776.
- [39] K. Kucher, C. Paradis, and A. Kerren, "The state of the art in sentiment visualization," *Comput. Graph. Forum*, vol. 37, no. 1, pp. 71–96, 2018, doi: 10.1111/cgf.13217.

Implementation of CNN for Plant Identification using UAV Imagery

Mohd Anul Haq¹, Ahsan Ahmed^{2*}, Jayadev Gyani^{3*}

Department of Computer Science-College of Computer and Information Sciences, Majmaah University, Al-Majmaah 11952, Saudi Arabia^{1,3}

Department of Information Technology-College of Computer and Information Sciences, Majmaah University, Al-Majmaah 11952, Saudi Arabia²

Abstract—Plants are the world's most significant resource since they are the only natural source of oxygen. Additionally, plants are considered crucial since they are the major source of energy for humanity and have nutritional, therapeutic, and other benefits. Image identification has become more prominent in this technology-driven world, where many innovations are happening in this sphere. Image processing techniques are increasingly being used by researchers to identify plants. The capacity of Convolutional Neural Networks (CNN) to transfer weights learned with huge standard datasets to tasks with smaller collections or more particular data has improved over time. Several applications are made for image identification using deep learning, and Machine Learning (ML) algorithms. Plant image identification is a prominent part of such. The plant image dataset of about 300 images collected by mobile phone and camera from different places in the natural scenes with nine species of different plants are deployed for training. A five-layered convolution neural network (CNN) is applied for large-scale plant classification in a natural environment. The proposed work claims a higher accuracy in plant identification based on experimental data. The model achieves the utmost recognition rate of 96% NU108 dataset and UAV images of NU101 have achieved an accuracy of 97.8%.

Keywords—Convolutional Neural Networks (CNN); Machine Learning (ML) algorithms; plant image identification; plant image dataset

I. INTRODUCTION

When it comes to the matter of food, pharmaceuticals, and other raw materials, the first name that comes to our mind is Plant. Plants directly or indirectly provide benefits to humanity's progress. As a result, our ability to recognize the characteristics of a wide range of plant species was critical for our prosperity and survival. Thousands of plant species are scanned and cataloged by various researchers and academic practitioners [1] to better recognize their features, value, and prospective applications [2]. The majestic feature of the plant kingdom is that no two are alike. Every tree is beautiful and has possession of divergent characteristics. As no fingerprint in the world is identical, no venation pattern of a plant is similar. Different species have seemingly many varieties of shapes, sizes, textures, and effervescent colors. With the change in seasons, every tree changes its appearance to adapt to the environment. Automated plant identification is the most prominent remedy for bridging the environment with today's rapidly developing technology that has substantially improved

in recent years. Both environmentalists and techies are showing considerable attention to plant identification for its never-ending applications. All this is possible because of the instigation of Deep Learning (DL) models, which can automatically learn to extract higher-level features from unrefined data [3]. Machine learning (ML), a branch of AI, is a technique in which a computer is taught, through the analysis of data, to carry out a particular activity, such as prediction. It has garnered a lot of interest in recent years from researchers in many different areas, including object recognition [4-5], natural language processing, speech recognition [6], and smart manufacturing [7-8]. Machine learning algorithms can be broken down into three distinct classes denoted by whether they rely on an external observer to provide feedback on the learning process: unsupervised learning, supervised learning, and semi-supervised learning. Unsupervised learning trains using unlabeled data, unlike supervised learning. Semi-supervised learning, in contrast to these other two types, uses both labeled and unlabeled data during training. The advancement in machine learning helped in developing innovative automated image identification models are proposed. With the evolution of smartphones throughout the previous two decades, mobile-based applications have become an important part of this development. Mobile phones play an important role in image identification in real-world and natural environments. Millions of images have been acquired through mobiles. The awareness of people on maintaining ecological balance and the crucial role of plant identification in many other platforms drives modern scientists to focus on improving the performance of mobile-based image identification models. DL networks can overcome prior challenges related to handcrafted feature extraction from massive volumes of data by utilizing parallel computing architectures [9]. DL models may learn which traits are most important for feature extraction using multi-level representations, demonstrating their efficacy [10].

In the modern-day, variations in leaf characteristics are used by researchers as a comparative tool for identifying plants. Many efforts have been put on to extract the features of flower, fruit, or leaf for identifying the plant. The research about automated plant identification has started about two decades ago. In [11], Söderkvist classified the trees from images of the leaf using a computer vision classification system. Different features of the leaf are defined as different descriptors for comparison. Backpropagation was used on 15 different Swedish tree classes to investigate features such as

leaf shape and moment for the feed-forward neural network. In [12], Fu et al. believed that leaf venation contains important genetic information. Due to the high diversity in leaf veins, conventional thresholding-based methods may not extract information accurately. Features such as edge gradients, local contrasts, and statistical features are extracted to define the characteristics of the veins' surrounding pixels. The experiment shows that, when compared to traditional thresholding, the neural network is more efficient to identify vein images. Li et al. [13] presented an effective leaf vein extraction approach by using the snakes' method along with cellular neural networks. Using prior knowledge similar features were obtained from both implicit and parametric models for acceptable results on leaf segmentation. A probabilistic neural network as a classifier was employed for identifying the leaf images of plants that performed better than the BP neural network in terms of accuracy [14]. The concept of natural-based leaf recognition and a contour segmentation technique based on the polygon leaf model was used to obtain the contour images [15].

Deep learning has been a hot topic since 2010, and many scientists are focusing their efforts on deep learning for image recognition. Many researchers work on flowers for classifying ornamental plants. As leaves and flowers also have different features to distinguish, Nilsback and Zisserman proposed a model for foreground and background and a light generis shape model for identifying petal structures. This technique describes the shape, color, texture features, and other characteristics of the flowers [16]. A deep learning model with 26 layers and 8 residual building components was developed for identifying plants at a large scale in the natural environment. The suggested model achieved a recognition rate of 91.78% for the BJFU100 dataset, suggesting that deep learning is an encouraging approach for smart forestry [17]. A system to detect and recognize a variety of plant species was designed by Militante et al. [18]. The system was able to detect a variety of plant diseases too. A deep learning model was trained using 35,000 images of healthy and infected plant leaves and finally, a 96.5% of accuracy rate was achieved by the trained model. Liu and Kan proposed a classification model which has a combination of texture combination and shape features, using deep belief network architecture. Local binary patterns derive the texture of a leaf. The combination is maintained by gray level co-occurrence matrix and Gabor filters [19]. In the year 2022, Haq [20] applied a machine learning-based Random Forest (RF) classification model for comparing the vegetation extent. To identify leaf images, Zhang et al. [21], created a deep learning system that included eight layers of Convolution Neural Networks (CNN) and achieved a better recognition rate. Zhang et al. combined the Harr-like transformation of local features with SIFT features of flower images, classifying them by the k-nearest neighbor method. His conjecture of concentrating on local features hindering performance proved right. In 2020, Yang et al. employed shallow CNN in place of DCNN for identifying the disease on the images of a leaf. A significant performance improvement was achieved with their

method in comparison to the DCNN models. The use of PCA helps in performance improvement and reduces computational complexity too [22].

All this research paved a path for nowadays plant recognition technology. Various mobile applications like P1@ntNet [23], LeafSnap [24], Microsoft Garage's Flower Recognition app [25], iNaturalist [26], and Plantsnap [27] are the most prominently used mobile applications for plant identification. Although there are numerous kinds of research on terrestrial plant identification there are barely any with UAV imagery. Automatic plant taxonomy has had rapid growth in recent years but image identification using UAV [28] is the newborn interest of the researchers. Popescu D and Ichim have proposed an image recognition model based on texture analysis. Two types of texture statistical and fractal characteristics were considered on images of forests, buildings, grassland, and flooding zone. All the research done prior to date hinders their performance in a natural environment. The traditional models used for classification rely on preprocessing of the image to delete the background and enhance it. Preprocessing the dataset consumes time and might not be as accurate as natural environment identification. The main advantages of the UAV (unmanned aerial vehicle) are the collection of the terrestrial dataset is more time taking than the aerial dataset. The traditional classification of aerial images is expensive for monitoring the evolution of the research area.

To overcome all these challenges, we acquired the NU108 a terrestrial dataset using a mobile phone in a natural environment. A UAV dataset NU101 is another prominent part of this research. A five-layer deep learning model is applied to the proposed model and a recognition rate of 97% was achieved on the NU101 dataset.

II. DATASET

A. Terrestrial Imagery

The terrestrial dataset NU108 contains 300 images of nine plant species within the campus of NIIT University (See Fig. 1). These images were collected by mobile phone in the natural environment with the help of a camera of 12 megapixels. The size of the image is 4032x3024. Drone imagery in the same vicinity NU101 dataset is collected by UAV (DJI SPARK). The images were taken from an altitude of 24m. The drone is equipped with a camera of 12 megapixels with an equivalent focal length and an RGB sensor image of a size 960 × 1280 as shown in Fig. 2.

B. Drone Imagery

The dataset NU101 was collected through DJI SPARK from an altitude of 24m. A total of 15 images covering an approximate distance of 300m have been collected and made into a mosaic image using Drone2 map software. ROIs of each tree species in the image are detected using MATLAB and used as training data for our model. A total mosaic image is used for validation.

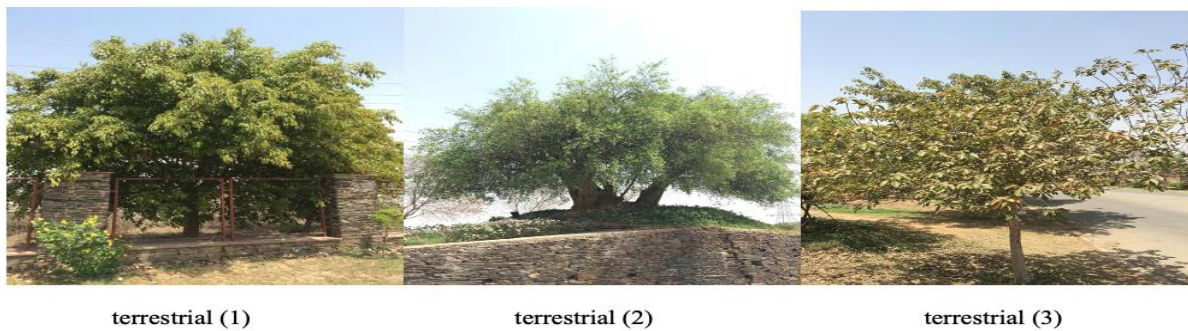


Fig. 1. Terrestrial images of plant species from the NU108 dataset.



Fig. 2. The drone images of plant species were collected using DJI SPARK from an altitude of 24m.

III. METHODOLOGY

A. Convolution Neural Network

While ML has many applications, its success is highly dependent on the features chosen to train the model. Due to their ability to automatically extract higher-level features from the raw input data, deep learning algorithms have garnered a lot of attention in recent years. When it comes to deep learning, the most popular technique is the convolutional neural network (CNN) algorithm, which is built on the ANN foundation. Convolutional networks were inspired by early findings of biological processes; the connectivity pattern of CNN neurons resembles the organization of an animal's visual cortex. A CNN allows elements to be identified and classified with minimal pre-processing [32-37]. The CNNs are regularized multilayer perceptron variants. The term "multilayer perceptron" usually refers to networks that are fully connected, meaning that each neuron in one layer is linked to all neurons in the next layer. In comparison to other image classification techniques, CNN requires extremely little pre-processing. Difficult image-driven pattern recognition problems can be solved easily by the simple architecture of CNNs that require minimum input parameters [38]. When other algorithms need to be hand-engineered this neural network learns about the filters which help in natural scene image classification. This is the best feature of CNNs that they do not need any prior knowledge about the classification and most little human efforts are needed.

1) *Convolution*: The convolution layer, which is connected to local portions of the input, will determine the output of neurons by calculating the scalar product of a set of weights with the region related to the input volume represented in the form of metrics. The operation of convolution is shown in Fig. 3. The rectified linear unit

(abbreviated as ReLU) is applied as an 'elementwise' activation function like a sigmoid to the output of the previous layer's activation.

2) *Pooling*: It is one of the building blocks of CNN that simply reduce the input size with the help of the downsampling process. Thus, it reduces the computation for the given problem (See Fig. 4). Multiple pooling layers (i.e., Max Pooling Layer and Average Pooling Layer) can be applied to reduce the number of parameters within that activation.

3) *Fully connected layer*: The input data to CNN layers cannot be inserted as shapes or pictures. Therefore, the process of flattening is performed to construct a single-dimensional feature vector from the output of the convolutional layers as shown in Fig. 5. Furthermore, it is linked to the final classification model, which is referred to as a fully connected layer. The fully connected layers produce the final output class scores from the activations in the same way that normal ANNs do. The generated class scores can be utilized for classification. It is also possible that ReLU might be employed between these layers to improve performance.

4) *Architecture*: A convolutional neural network consists of a convolution layer in which the input image is convoluted with multiple kernels (feature detectors) and then connected to the pooling layer where features are magnified and then connected to the fully connected layer. The activation functions help in the classification of the data. The architecture of our model has five-layered deep neural networks (See Fig. 6 (a)), and the second model with three neural network layers (See Fig. 6(b)). The convolution layer model is applied to terrestrial data NU108 for image identification.

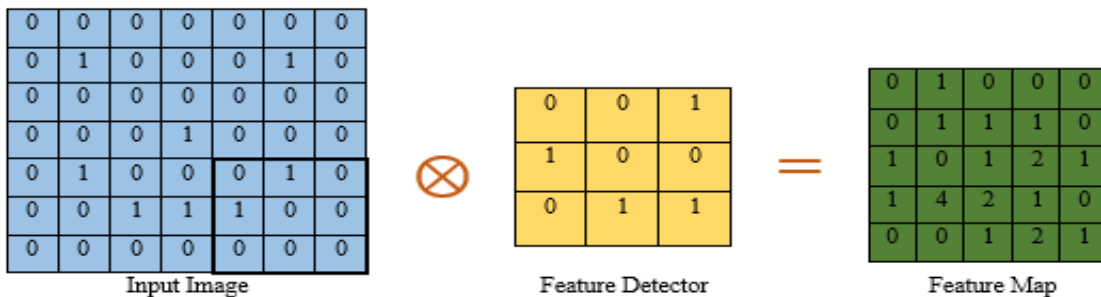


Fig. 3. Convolution operation using the scalar product of a set of weights with the related region.



Fig. 4. Illustration of max polling layer.

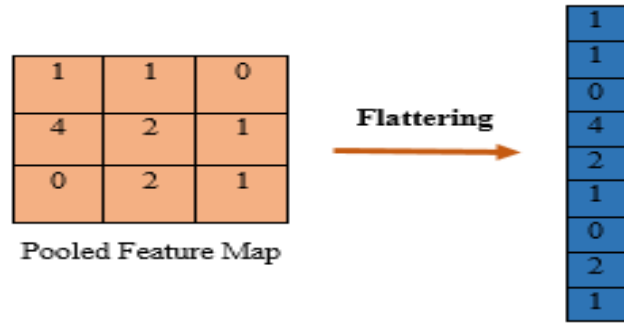


Fig. 5. Process of flattening to convert the data into a 1-dimensional feature vector.

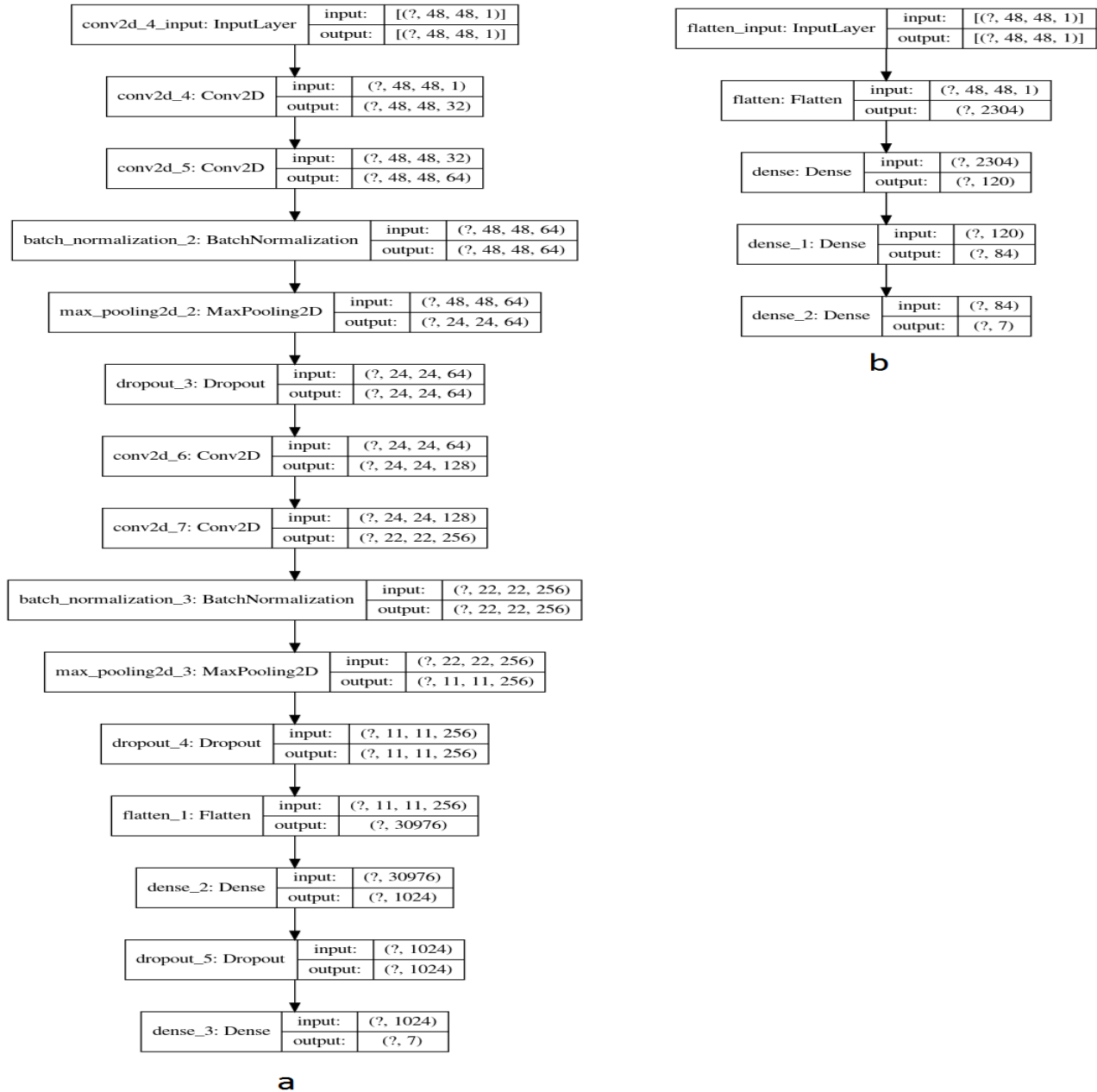


Fig. 6. The architecture of the models with (a) five layers of deep neural networks, (b) three layers of neural networks.

5) *ReLU*: The activation function ReLU is used mainly to remove the negative values. The standard way of model a neuron's output is f as a function of its input x is with $f(x) = \tanh(x)$ or $f(x) = 1/(1+e^{-x})$. These saturating nonlinearities are significantly slower than the non-saturating nonlinearity in terms of training time with gradient descent and are described mathematically as $f(x) = \max(0, x)$. A visual illustration of the ReLU activation function is shown in Fig. 7. Following [29], refer to neurons with the nonlinear property as ReLUs. Deep CNNs with ReLUs perform better than their tanh unit counterparts in terms of training time. When compared to other activation functions, such as sigmoid or tanh, ReLU does not experience the vanishing gradient problem that these other activation functions do.

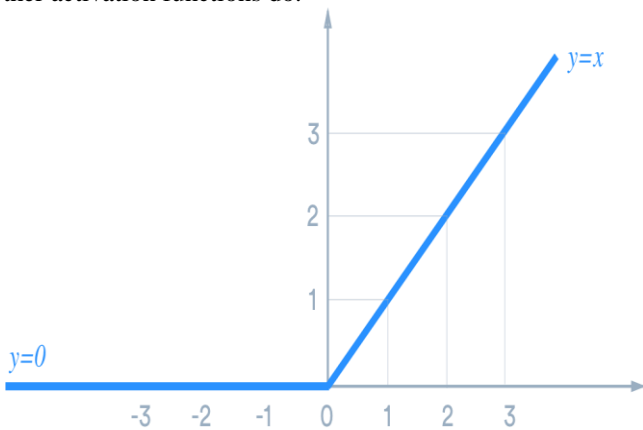


Fig. 7. A visual Illustration of the ReLU activation function.

6) *Sigmoid function*: A sigmoidal function in the form of a hyperbolic tangent is consistently preferred theoretically and experimentally for deep learning models. It is used in neural networks to give logistic neurons with real-valued output which is a bounded function of their total input. Having the advantage of nice derivatives, it facilitates learning the weights of a neural network much easier. Sigmoid activation function $S(x)$ is defined by equation (1) and is illustrated in Fig. 8.

$$S(x) = \frac{1}{1+e^{-x}} = \frac{e^x}{e^x+1} \quad (1)$$

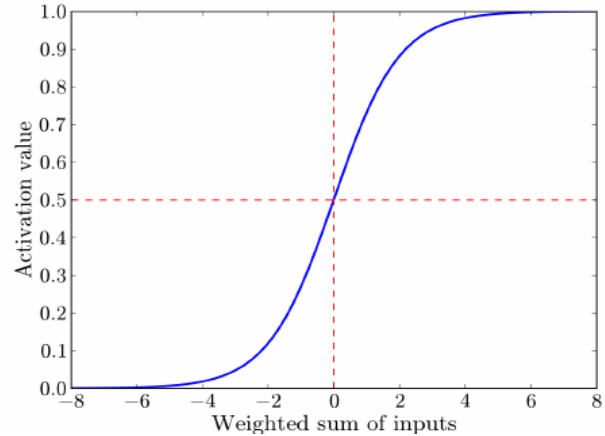


Fig. 8. Sigmoid activation function.

B. PCA Method

After the input image is processed further the image has a natural scene with trees and other objects in it. Tree identification is carried out as the first step toward classification and identification. Tree identification is done through the Primary Component Analysis method which is one of the dimensionality reductions. The mosaic of the input image is shown in Fig. 9.

1) *Decorrelation*: Decorrelation stretch enhances the visual perception of the image so that the objects can be differentiated due to enhanced color separation of the image (See Fig. 10).

2) *Principal component analysis*: Principal Component Analysis (PCA) is a technique used for dimensionality reduction. It directly decreases the number of feature variables by saving only the required variables. PCA is used to obtain a single-band image. First, compute the covariance matrix of the data and further calculate the eigenvalues and vectors for the computed covariance matrix. Select only the most significant feature vectors using the eigenvalues and vectors, and then convert the data onto those vectors to reduce dimensionality (See Fig. 11).



Fig. 9. Mosaic of the input image.

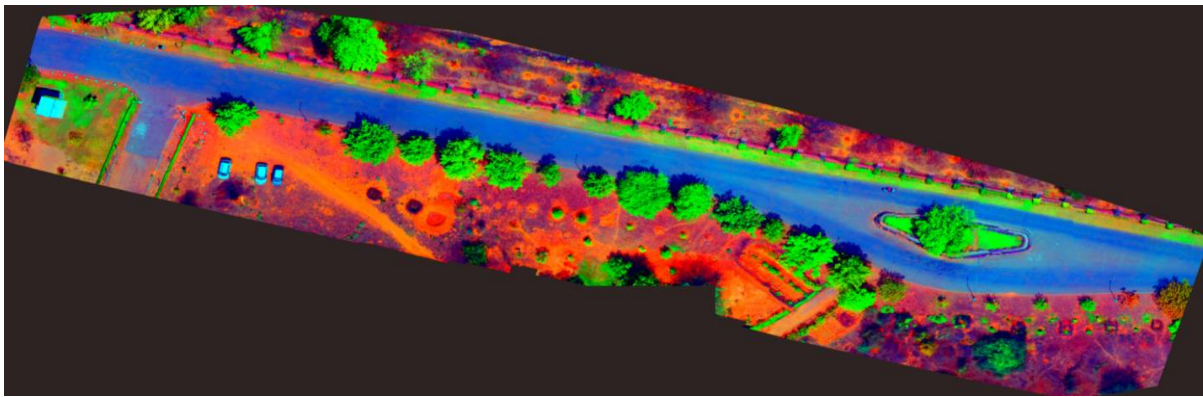


Fig. 10. Image after decorrelation stretch image.

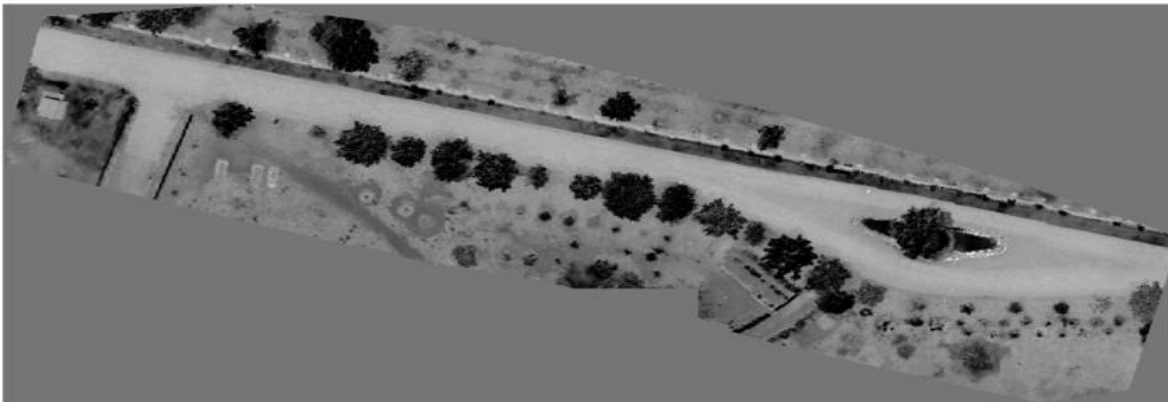


Fig. 11. Image after applying the PCA technique.

3) *Thresholding*: Thresholding is done by using Otsu's Method in our model. Otsu's thresholding approach iterates through all the threshold values and calculates a measure of pixel-level spread on each side [30] of the threshold, categorizing the pixels as foreground or background. The goal is to discover the threshold value at which the sum of foreground and background spreads is the smallest.

4) *Edge detection*: Edge detection is an image processing approach for detecting object boundaries within images. It works by sensing brightness discontinuities. It's utilized for data extraction and image segmentation. Edge detection in our model helps define the tree's outline and separate the tree from the background for further analysis. Noise in the form of unwanted background and shadow of the trees can be removed using the Gaussian Filter method. This helps in reducing false detection. The Gaussian filter is represented as equation (2).

$$H_{ij} = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{(i-(k+1))^2 + (j-(k+1))^2}{2\sigma^2}\right) \quad (2)$$
$$1 \leq i, j \leq (2k + 1)$$

Two methods namely Sobel and Canny can be applied to detect the edges. Canny has four different filters to detect edges in different directions in blurred images. The output of Sobel is taken as the input of canny. Canny gives an output of an image with thicker outer lines and thinner inner edges.

5) *Implementation and preprocessing*: Tree crowning is followed by Dataset Processing after the edge detection of the vegetation. Trees and plants are cropped from the resultant image of tree crowning and are stored in a folder for further processes. Multiple images are cropped from a single image using Matlab. Several trees are counted by the application of the PCA method and that helps in the training of the model. All the trees that are cropped and stored are now sorted. Identification of the species is done manually. The data set folders thus obtained are divided into the testing set and training set.

Keras is a Python based open-source deep learning framework that is used to implement the model. All the experiments were run on a Windows server with a GTX 1050Ti GPU (8 GB memory) and a 3.40 GHz i7-8750H CPU (16 GB memory).

6) *Training algorithm*: A feed-forward neural network is like a logistic regression algorithm, but the input values get more linearly separable with the addition of hidden layers. A gradient descent-based strategy was used to train this model by minimization of backpropagated error. The backpropagation algorithm is used to iteratively discover a local optimum of the loss function [31]. The error backpropagation through the output layer L to layer $l < L$ will be done via a recursive computation as equations (3, 4, 5).

$$\frac{\partial L}{\partial w_k^{(l)}} = \frac{\partial \eta^{(l)}}{\partial w_k^{(l)}} \frac{\partial h^{(l)}}{\partial \eta^{(l)}} \frac{\partial \eta^{(l+1)}}{\partial h^{(l)}} \frac{\partial L}{\partial \eta^{(l+1)}} \quad (3)$$

$$= - \frac{\partial \eta^{(l)}}{\partial w_k^{(l)}} \frac{\partial h^{(l)}}{\partial \eta^{(l)}} \frac{\partial \eta^{(l+1)}}{\partial h^{(l)}} \Delta^{(l+1)} \quad (4)$$

$$= - \frac{\eta^{(l)}}{w_k^{(l)}} \Delta^{(l)} \quad (5)$$

IV. RESULT ANALYSIS

A series of experiments were done to find the best accuracy and improve the quality of the result. The comparison between accuracy and epochs of the validation accuracy and testing accuracy has been measured. The proposed model results in 97.8% accuracy over the UAV dataset and 96% accuracy in the terrestrial dataset (See Fig. 12). The output image after applying the proposed CNN model is shown in Fig. 13. For the size of the dataset, this five-layered CNN is the best tradeoff between model capacity and accuracy.

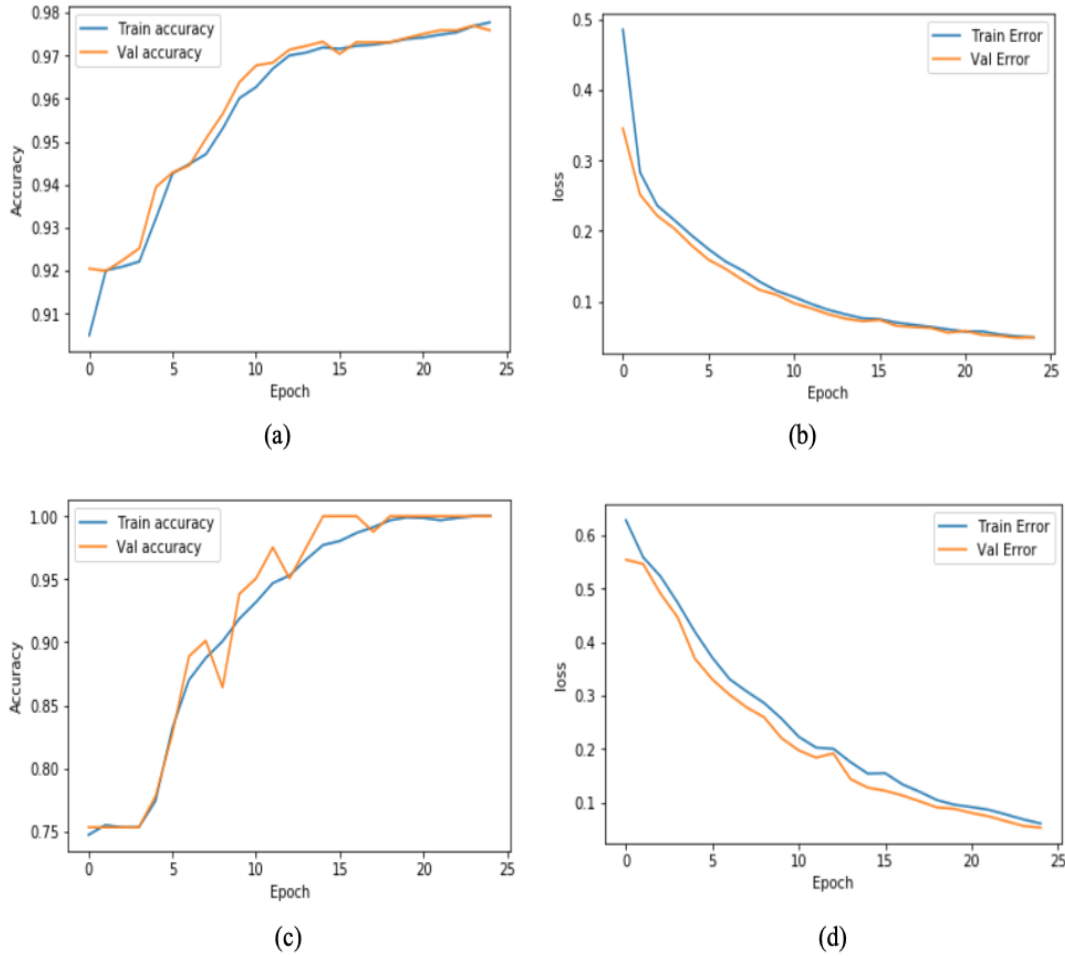


Fig. 12. Performance of the model.

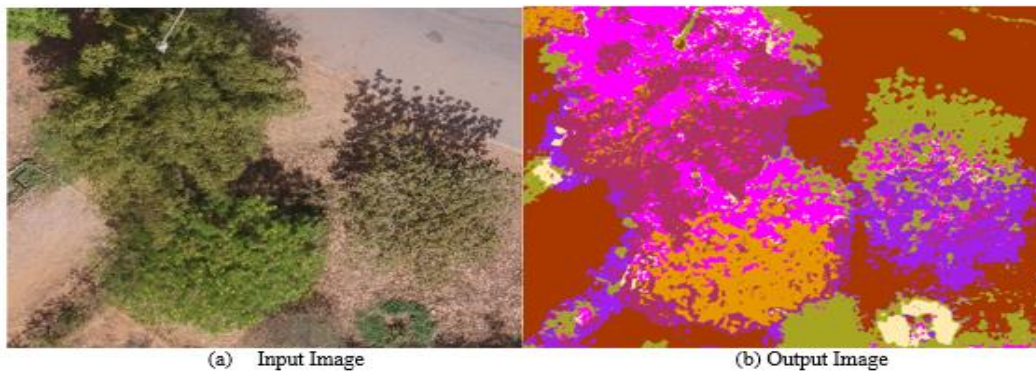


Fig. 13. Resultant image after applying the proposed CNN model.

V. CONCLUSION

The first mobile device acquired the NU108 dataset containing 300 images of nine plant species and the NU101 UAV data set which paves a path for further plant identification studies on UAV imagery. The proposed model achieves an accuracy of 97.8%, proving that deep learning is an encouraging technology for classifying plants in the natural environment. The NU101 database will be expanded in the future to include more plant species at various stages of their life cycles, as well as more thorough annotations. The drone imagery will be widened and work on more sophisticated systems for higher accuracy over a larger area will be done in future research work. The future scope of the present work will be utilizing the advanced DL models with additional datasets.

ACKNOWLEDGMENT

Jayadev Gyani would like to thank Deanship of Scientific Research at Majmaah University for supporting this work under Project No. R-2023-364.

REFERENCES

- [1] P. Bonnet, H. Goëau, S.T. Hang, M. Lasseck, M. Šulc, V. Malécot, P. Jauzein, P.-C. Melet, C. You, and A. Joly, "Plant identification: experts vs. machines in the era of deep learning," *Multimedia Tools and Applications for Environmental & Biodiversity Informatics*, Springer, pp. 131–149, 2018.
- [2] A. Begue, V. Kowlessur, U. Singh, F. Mahomoodally, and S. Pudaruth, "Automatic recognition of medicinal plants using machine learning techniques," *International Journal of Advanced Computer Science and Applications*, vol. 8(4), pp. 166–175, 2017. <https://dx.doi.org/10.14569/IJACSA.2017.080424>.
- [3] S.H. Lee, C.S. Chan, S.J. Mayo, and P. Remagnino, "How deep learning extracts and learns leaf features for plant classification," *Pattern Recognition*, vol. 71, pp. 1–13, 2017. <https://doi.org/10.1016/j.patcog.2017.05.015>.
- [4] A. Krizhevsky, I. Sutskever, and G.E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing*, vol. 25, MIT Press, Cambridge, MA, 2012.
- [5] J. Stallkamp, M. Schlipsing, J. Salmen, and C. Igel, "Man vs. computer: Bench-marking machine learning algorithms for traffic sign recognition," *Neural Networks*, vol. 32, pp. 323–332, August 2012. <https://doi.org/10.1016/j.neunet.2012.02.016>.
- [6] W. Chan, N. Jaitly, Q. Le, and O. Vinyals, "Listen, attend and spell: A neural network for large vocabulary conversational speech recognition," *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP, IEEE*, pp. 4960–4964, 2016. <https://doi.org/10.1109/ICASSP.2016.7472621>.
- [7] J. Wang, Y. Ma, L. Zhang, R.X. Gao, and D. Wu, "Deep learning for smart manufacturing: Methods and applications," *J. Manuf. Syst.*, vol. 48, pp. 144–156, 2018. <https://doi.org/10.1016/j.jmsy.2018.01.003>.
- [8] D. Wu, C. Jennings, J. Terpenney, R.X. Gao, and S. Kumara, "A comparative study on machine learning algorithms for smart manufacturing: tool wear prediction using random forests," *J. Manuf. Sci. Eng.* 139 (7), July 2017. <https://doi.org/10.1115/1.4036350>.
- [9] P. Barré, B.C. Stöver, K.F. Müller, and V. Steinhage, "LeafNet: A computer vision system for automatic plant species identification," *Ecological Informatics*, vol. 40, pp. 50–56, July 2017. <https://doi.org/10.1016/j.ecoinf.2017.05.005>.
- [10] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning. *Nature* 521 (7553), pp. 436–444, 2015. <https://doi.org/10.1038/nature14539>.
- [11] O. Söderkvist, "Computer Vision Classification of Leaves from Swedish Trees, 2001.
- [12] H. Fu, Z. Chi, J. Chang, and C. Fu, "Extraction of leaf vein features based on artificial neural network—Studies on the living plant identification I," *Chinese Bulletin of Botany*, vol. 21, pp. 429–436, 2003.
- [13] Y. Li, Q. Zhu, Y. Cao, and C. Wang, "A leaf vein extraction method based on snakes technique," *IEEE International Conference on Neural Networks and Brain (ICNN&B '05)*, pp. 885–888, 2005. <https://doi.org/10.1109/ICNNB.2005.1614763>.
- [14] P. He, and L. Huang Lin, "Feature extraction and recognition of plant leaf," *Journal of Agricultural Mechanization Research*. 6, pp. 168–170, 2008.
- [15] M. E. Nilsback, and A. Zisserman, "Delving deeper into the whorl of flower segmentation," *Image and Vision Computing*, 28(6), pp. 1049–1062, 2010. <https://doi.org/10.1016/j.imavis.2009.10.001>.
- [16] A.R. Backes, D. Casanova, and O.M. Bruno, "Plant leaf identification based on volumetric fractal dimension," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 23, issue 06, pp.1145–1160, 2009. <https://doi.org/10.1142/S0218001409007508>.
- [17] Y. Sun, Y. Liu, G. Wang, and H. Zhang, "Deep Learning for Plant Identification in Natural Environment. *Computational Intelligence and Neuroscience*, vol. 2017, 7361042, 1–6, May 2017. <https://doi.org/10.1155/2017/7361042>.
- [18] S.V. Militante, B.D. Gerardo, and N.V. Dionisio, "Plant leaf detection and disease recognition using deep learning," *IEEE Eurasia Conference on IOT, Communication and Engineering (ECICE)*, 19276490, December 2009. <https://doi.org/10.1109/ECICE47484.2019.8942686>.
- [19] L. Nian, and K. JiangMing, "Plant leaf identification based on the multi-feature fusion and deep belief networks method," *Journal of Beijing Forestry University*, 38(3), pp. 110–119, 2016.
- [20] M.A. Haq, "Planetoscope Nanosatellites Image Classification Using Machine Learning," *Computer Systems Science and Engineering*, vol. 42(3), pp. 1031–1046, 2022. <https://doi.org/10.32604/csse.2022.023221>.
- [21] C. Zhang, J. Liu, C. Liang, Q. Huang, and Q. Tian, "Image classification using Harr-like transformation of local features with coding residuals," *Signal Processing*, vol. 93, issue 8, pp. 2111–2118, 2013. <https://doi.org/10.1016/j.sigpro.2012.09.007>.
- [22] Y. Li, J. Nie, and X. Chao, "Do we really need deep CNN for plant diseases identification?," *Computers and Electronics in Agriculture*, vol. 178, 105803, Nov. 2020. <https://doi.org/10.1016/j.compag.2020.105803>.
- [23] A. Joly, H. Goëau, P. Bonnet, V. Bakić, J. Barbe, S. Selmi, I. Yahiaoui, J. Carrée, E. Mouysset, J.-F. Molino, N. Boujema, and D. Barthélémy, "Interactive plant identification based on social image data," *Ecological Informatics*, vol. 23, pp. 22–34, 2014. <https://doi.org/10.1016/j.ecoinf.2013.07.006>.
- [24] N. Kumar, P.N. Belhumeur, A. Biswas, D.W. Jacobs, W.J. Kress, I.C. Lopez, and J.V.B. Soares, "Leafsnap: a computer vision system for automatic plant species identification," *Proceedings of the Computer Vision-ECCV 2012*, pp. 502–516, 2012.
- [25] Microsoft Research. 2022. Flower Recognition, a Garage Project (Retired) - Microsoft Research. [online] Available at: <https://www.microsoft.com/en-us/research/project/flowerreco/> [Accessed 5 April 2022].
- [26] iNaturalist. 2022. iNaturalist. [online] Available at: <https://www.inaturalist.org/> [Accessed 5 April 2022].
- [27] Plantsnap - Identify Plants, Trees, Mushrooms With An App. 2022. PlantSnap - Plant Identifier App, #1 Mobile App for Plant Identification. [online] Available at: <https://www.plantsnap.com/> [Accessed 5 April 2022].
- [28] D. Popescu, and L. Ichim, "Image recognition in UAV application based on texture analysis," *16th International Conference on Advanced Concepts for Intelligent Vision Systems- ACIVS 2015*, vol. 9386, pp. 693–704, 2015, Springer, Cham.
- [29] V. Nair, and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," *27th International Conference on Machine Learning- ICML'10*, pp. 807–814, 2010.
- [30] F.A. Gougeon, "A crown-following approach to the automatic delineation of individual tree crowns in high spatial resolution aerial images," *Canadian Journal of Remote Sensing*. Vol 21(3), pp. 274–284, 1995. <https://doi.org/10.1080/07038992.1995.10874622>.
- [31] F. Farhadi, "Learning Activation Functions in Deep Neural Networks," Doctoral dissertation, École Polytechnique de Montréal, 2017.

- [32] M.A. Haq, "CNN Based Automated Weed Detection System Using UAV Imagery," *Computer Systems Science and Engineering*, vol 42(2), pp. 837-849, 2012. <https://doi.org/10.32604/csse.2022.023016>.
- [33] M.A. Haq, "Deep Learning Based Modeling of Groundwater Storage Change," *Computers Materials and Continua*, 70(3), pp. 4599–4617, 2021.
- [34] M.A. Haq, M. Alshehri, G. Rahaman, G., A. Ghosh, P. Baral, and C. Shekhar, "Snow and glacial feature identification using Hyperion dataset and machine learning algorithms," *Arabian Journal of Geosciences*, 14:15, pp. 1–21, 2021.
- [35] M.A. Haq, P. Baral, S. Yaragal, and B. Pradhan, "Bulk processing of multi-temporal modis data, statistical analyses and machine learning algorithms to understand climate variables in the indian himalayan region," *Sensors*, vol 21, issue 21, 2021. <https://doi.org/10.3390/s21217416>.
- [36] M.A. Haq, A. Ghosh, G. Rahaman, and P. Baral, "Artificial Neural Network-Based Modeling of Snow Properties Using Field Data and Hyperspectral Imagery," *Natural Resource Modeling*, vol 32, issue 4, pp. 1-36, July 2019.
- [37] M.A. Haq, A. Ahmed, I. Khan, J. Gyani, A. Mohamed, El-A Attia, P. Mangan and D. Pandi, "Analysis of environmental factors using AI and ML methods," *Scientific Reports* 12, 13267, August 2022. <https://doi.org/10.1038/s41598-022-16665-7>.
- [38] K. O'Shea and R. Nash, "An Introduction to Convolutional Neural Networks," 2015. <https://arxiv.org/abs/1511.08458>.

An IoT-based Framework for Detecting Heart Conditions using Machine Learning

Mona Alnaggar¹, Mohamed Handosa², T. Medhat³, M. Z. Rashad⁴

Robotics and Intelligent Machines Department-Faculty of Artificial Intelligence, Kafrelsheikh University, Egypt¹

Computer Science Department-Faculty of Computers and Information, Mansoura University, Egypt^{2,4}

Electrical Engineering Department-Faculty of Engineering, Kafrelsheikh University, Egypt³

Abstract—A lot of diseases may be preventable if they can be analyzed or predicted from patient historical and family data. Predicting diagnosis depends on the gathered clinical and physiological data of patients. The more collected clinical and medical healthcare data, the more knowledge the medical support system may support. Hence, real monitoring clinical and healthcare data for patients is the trend of this decade based on Internet of Things technologies (IoT). IoT models facilitate human life by easily collecting clinical data remotely for recognizing diseases that are easily treatable if it is diagnosed early. This paper proposes a framework consisting of two models: (i) heart attack detection model (HADM); (ii) Electrocardiosignal ECG heartbeat multiclass-classification model (ECG-HMCM). Gridsearch is used to the hyperparameters optimization for different machine learning (ML) techniques. The used dataset in HADM consists of 1190 patients and 14 features. As the foundation of diagnosing cardiovascular disease is arrhythmia detection hence, we propose an ECG heartbeat multi-class classification model using MIT-BIH Arrhythmia and PTB Diagnostic ECG signals dataset which contains five categories with 109446 samples. K Nearest Neighbor (KNN) technique is applied to build ECG-HMCM in addition to the using of Gridsearch algorithm for hyperparameter optimization aiming to improve the accuracy of classification which achieved 97.5%. The proposed framework aims to facilitate human life by easily collecting clinical data remotely. The outcomes of the experiments show that the suggested framework works well in a practical setting.

Keywords—Medical healthcare; medical support system; real monitoring; Internet of Things (IoT); ECG signals, Gridsearch; hyperparameter optimization

I. INTRODUCTION

Heart disease and other cardiovascular diseases (CVDs) are among the prominent reasons of mortality in the globe[1]. Corresponding to the World Health Organization (WHO), around 31% of fatalities globally are caused by cardiac disorders[2]. Furthermore, almost a quarter of a million Americans lose their lives each year to heart attacks. Heart disease is difficult for medical professionals and specialists to diagnose since it takes knowledge and experience to do so. Cardiac disorders, which are roughly categorized as various sorts of faulty heart problems, are responsible for about 1 in every 4 fatalities. However, diagnosing CVDs requires human analysis of data from several clinical tests, which takes time.

To give medical professionals faster analysis by lowering the time it takes to get a diagnosis and improving outcomes,

new methods for automating the identification of such anomalies in human heart diseases should be created. Electronic health records (EHRs) [3] are frequently used to find insightful data patterns that enhanced Machine Learning(ML) algorithms' ability to anticipate. ML makes a substantial contribution to the resolution of problems like forecasts in numerous fields, such as healthcare. Doctors and healthcare workers always face different epidemics. They suffer from their infection by these epidemics during their work. Hence, researchers and scientists must support them with new innovative ideas, tools, smart devices, and medical systems aiming to eliminate the increased infection and protect medical staff and patients from these epidemics. Applying ML and DL algorithms for prediction and diagnosing supports many medical branches such as oncology[4], surgery[5], Dentistry [6], Cardiology [7] Fetal Heart Diagnosis [8], diseases classification [9], and remote respiration rate monitoring [10, 11]. To aid in everyday medical inspections, a multifunctional, portable health monitoring device was developed and put into use[12].

Even though most feature-based ML approaches employ Heart rate variability (HRV) analysis to diagnose Electrocardiosignal (ECG), their robustness cannot be assured. Most crucial HRV characteristics are always influenced by unrehearsed variations, breathing, medication interactions, age, and gender. As an outcome, analyzing HRV should not serve as the major pillar for evaluating heart disorders. Most current ECG classification techniques offer accurate classification outcomes that discriminate between CHF and NSR patients. Building an effective automated framework that can correctly differentiate CHF, ARR, and NSR scenarios and operate in real-time with minimal hardware complexity is still quite challenging.

To the extent of our knowledge, all previous works focus on either heart attack detection or ECG classification. This work combines the two objectives in an IoT based framework with a notification system. Additionally, the highest accuracy achieved for heart attack detection is 92.28%. This accuracy is still to be enhanced. This work tries to enhance the accuracy of heart attack detection based on the combination of two different models for building the proposed framework. Therefore, improving heart disease detection models using ML algorithms is essential for spotting heart illness early on. However, using ML technique requires figuring out how the heart failure dataset's characteristics relate to one another. The main contribution of this paper is to:

- Propose a multi-class classification model for heart attack detection.
- Enhance the accuracy of heart attack detection model using ML techniques.
- Propose a classification model for multi-class classification of ECG readings that represent the heartbeat.
- Optimize the finding results of ECG heartbeat signals classification.
- Propose an IoT framework with a notification system to provide suitable alerts or assist in the medical decision support systems and real monitoring for patients.

The rest of the paper is designed to contain the related work in Section II. The material and methods of the proposed models will be presented in Section III. Section IV will discuss implementation and methods. Section V will show the achieved results with discussions. Finally, the conclusion of our proposed paper and the future work will be discussed in Section VI.

II. RELATED WORK

Currently, transforming traditional healthcare services depends on applying models of ML on the clinical and healthcare data. Applying ML techniques in the medicine and healthcare area have proven great results in different specializations such as using medical images for body organs recognition, tumor detection, reconstruction of medical images, lung nodules detection and diseases classification, to name a few[13]. The doctor may monitor the patient's health, wellness, and health using an IoT medical gadget. The history of physiological medical tools and procedures is simple to find. The patient's physiological indicators should be timely monitored to prevent any health issues before they arise[14]. Sensors are adopted with real monitoring systems. This adaptation concerned the user's daily life, which is very simple to enable the collected data to be simply recorded over long hours. Through this means, collected data can be recorded over long hours to produce a huge amount of data. These data need processing and transformations to be more useful when applying Artificial intelligence algorithms. This is for increasing accuracy, decreasing the error rate, or improving computational efficiency to support suitable decisions. The IoT Monitoring system consists of three sections, the network of used medical sensor devices, IoT cloud, and graphical user interface (GUI). Hence, this section will mention some related work in (i) heart attack detection, (ii) ECG heartbeat signal classification.

A. Heart Attack Detection

One of the important techniques of DL is the Convolutional neural network (CNN) which is employed for building applications of image processing. There are many other techniques of DL have common uses[15]. Diagnosing diseases is a very attractive research point. It is the hope for patients to live a more easily life, especially, if it is prevented from being infected with diseases according to analyzing and predicting of patient's data. Using UCI repository and CNN for possible

heart disease prediction[16, 17] supported in the medical decision for patients. In this study [18], a distance-based ensemble for the KNN (k Nearest Neighbor) approach is introduced, and the results of its use in the diagnosis of heart disease are shown. There are two ensemble formations in use. One employs three distances, the previous five. For all the formations and versions, the evaluation was with the Cleveland dataset of UCI heart disease, this ensemble had an accuracy of close to 85%. The dataset for Cleveland heart disease was employed in the suggested study, and data mining algorithms, including regression and classification were applied. Random Forest (RF) and Decision Tree (DT) algorithms are used. Three ML methods are utilized in the proposed implementation: RF, DT, and Hybrid Model (Hybrid of random forest and decision tree). According to experimental findings, the achieved accuracy based on hybrid model was 88.7%. A hybrid model of DT and RF was utilized to forecast heart disease using the user's input response in the interface[19]. One of the most widely utilized ML algorithms, KNN, is frequently used for data categorization. Analyzing each patient's health metrics can also help forecast heart disease. The accuracy of the parameter comparison approach can be improved by combining KNN. The accuracy of the 13 definite parameters obtained by this study, which employs UCI ML dataset, was 86%[20]. Verma et al., for instance, introduced a hybrid model for the prediction of heart illness using particle swarm optimization (PSO) and ML classifiers, namely K-nearest neighbor and multi-layer perceptron (MLP), which reached a 90.28% accuracy [21]. Five ML algorithms were used by Alotaibi et al.[22] (LR, RF, Naive Bayes, DT, SVM). Rapid Miner is utilized for implementation, and it produces good results when compared to prior studies. 1013 patient records from the Cleveland dataset (UCI) were employed. Shah et al.[23]'s efficient model applies KNN, DT, RF, and Naive Bayes while utilizing just 14 key features from the Cleveland dataset from UC Irvine. The best outcome is produced by KNN among these algorithms. Another cardiac disease detection model was created by Jindal et al.[24]. To build a model, RF, LR, and KNN algorithms are applied. Of RD and LR, KNN provides the best accuracy.

B. ECG Heartbeat Signal Classification

The gold specification for noninvasive diagnosis of several forms of cardiac problems is the ECG. When identifying CVDs, the ECG signal is quite important. We can learn about the heartbeat from the ECG readings. Cardiac arrhythmia can be found using ECGs. ECG is a non-invasive diagnostic that measures the electrical activity of the heart muscles to evaluate heart function. Since it gives cardiologists all the information they need regarding heart problems, ECG serves as an effective tool for detecting a variety of cardiac illnesses. Yet, a critical factor in increasing treatment options and slowing the course of CHF is early diagnosed of the condition. The arrhythmia is a significant cardiac condition that contributes to several instances of sudden death (ARR). ARR stands for aberrant heart rhythm, which is a result of erratic heart rate. The provision of quality care necessitates accurate patient diagnosis as well as the identification of appropriate treatments while avoiding incorrect diagnoses. Early CVD identification also lowers costs and decreases CVD mortality. A classification method, which is crucial in clinical research, may be used to do

the task effectively and inexpensively with data mining approaches.

ECG data is made up of a collection of electrical impulses from the heart that are collected, and it may be used to spot several different cardiac disorders including arrhythmias, coronary artery disease, and cardiomyopathies. One of the most crucial techniques for assessing the general condition of the heart is heart rate variability (HRV). It shows how flexible the cardiac system is to changes in internal and external stimuli. The fluctuation in the interval between successive heartbeats is called a HRV time series (RR intervals) [25]. Congestive heart failure (CHF) is a dangerous cardiac condition that significantly raises death rates across the world. With CHF, the heart can no longer pump blood across the body efficiently enough to meet tissue demands for oxygen and metabolism. CHF affects more than 26 million individuals worldwide, with 3.6 million new cases added each year[26].

Robustness, efficiency and accuracy of applying ML/ DL techniques in healthcare encourage researchers to keep going to improve and enhance the used methodologies in medical systems, such as the proposed DL Modified Neural Network (DLMNN) to support in medication, particularly in heart diseases diagnosing by using a wearable IoT sensor device, then saving these normal and abnormal signals on cloud [27]. The information gained on this job can be used to the myocardial infarction (MI) classification challenge, according to a recommended strategy. On the MIT-BIH and PTB diagnostics datasets from PhysionNet, the proposed methodology is assessed. The findings show that the proposed technique can predict arrhythmia classification with average accuracies of 93.4% and 95.9%, respectively[28]. This paper [29] suggested an enhanced ResNet-18 model for ECG heartbeat signals classification of based on a convolutional neural network (CNN) method through suitable model training and parameter tuning. The model's distinct residual structure allows for the deepening of CNN layered structure to improve classification performance. The MIT-BIH arrhythmia database application findings show that the suggested model attains greater accuracy (96.50%). The suggested technique[30] is tested on the benchmark MIT-BIH arrhythmia database before being prototyped on a commercially accessible ARM-based embedded device. Moreover, the prototype is assessed using two schemes, namely class and customized schemes, and both schemes indicated greater overall accuracy for state-of-the-art CVDs diagnosis than the current works, at 96.29% and 96.08%, respectively.

III. MATERIAL AND METHODS

In this section, the steps of the proposed framework will be discussed as a preview architecture to facilitate decision making for doctors and remotely support patients with medical decisions. Our proposed architecture aims to achieve a more compatible framework to follow up the patient using IoT components. The framework phases can be listed as the following:

Collecting Data phase: depending on the selected IoT sensor for gathering the patient's medical, then start the phase of gathering data.

Data Preparation and cleaning: data processing, this step is responsible for raw data changing to be more useful.

Model selection: according to the issue and collected data, the algorithm should be implemented.

Train the model: using Artificial Intelligence Algorithms to enhance the accuracy of the prediction or decision making.

Model Evaluation: evaluating the proposed model and comparing it with others.

Notification system: when using the proposed system, and abnormality of signals appear then the family and doctors receive alerts about the case to take a right decision.

Even though most feature-based ML approaches employ HRV analysis to diagnose ECGs, their robustness cannot be guaranteed. Most crucial HRV characteristics are always influenced by unrehearsed variations, breathing, medication interactions, age, and gender. As a result, analyzing HRV should not serve as the major pillar for evaluating heart disorders. By creating novel diagnostic procedures for heart disorders, these difficulties can be eliminated.

The proposed framework consists of several phases such as cleaning data, storage of data, analysis data, implementing models, and notification system. Fig. 1 shows the basis of the proposed framework which oversees gathering physical data from the human body's surface and transmitting it wirelessly to the IoT cloud. Then, the signals are put through a variety of steps to increase their quality and make them suitable for wireless transmission, such as amplification, filtering, etc. Typically, a smart terminal (for example: a smartphone or computer) is required to receive the data before transmitting it wirelessly to the IoT cloud. After making data processing, artificial intelligence algorithms applied to predict if there will be any issue. If it exists, the monitoring system sends alarm to the connected computer or smart phone. As the foundation of diagnosing cardiovascular illness is arrhythmia detection, we propose this framework which consists of two different models: (i) Heart attack detection and (ii) ECG heartbeat signals classification.

A. The Heart Attack Detection Model (HADM)

A heart attack, also identified as a myocardial infarction, happens when a part of the heart muscle is not receiving enough blood, according to the center of disease control, and prevention (CDC). Hence, with the help of the dataset's variables, we will create a model to forecast the chance of a heart attack. ML techniques can be highly helpful in the early detection and treatment of patients with cardiovascular disease or who are at high cardiovascular risk beneficial (due to the presence of one or more risk factors, including hypertension [31], diabetes, hyperlipidemia [32], or previously existing illness).

Some of the preparation steps were executed such as dropping duplicates, Outlier Detection, Features encoding such as Binary features encoding and One-hot encoding were applied on the used dataset. Dividing dataset into training and testing datasets is executed for building and evaluating the HADM.

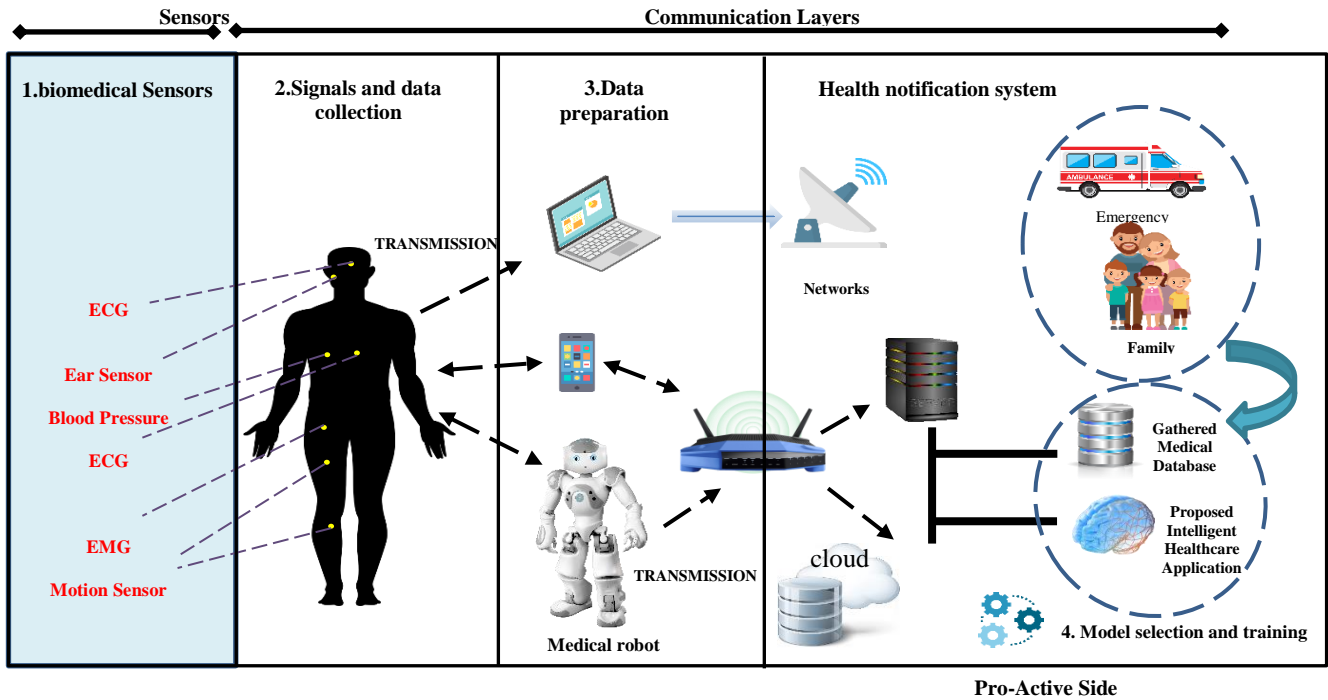


Fig. 1. Proposed framework phases to enhance the performance of the medical system based on the IoT sensors.

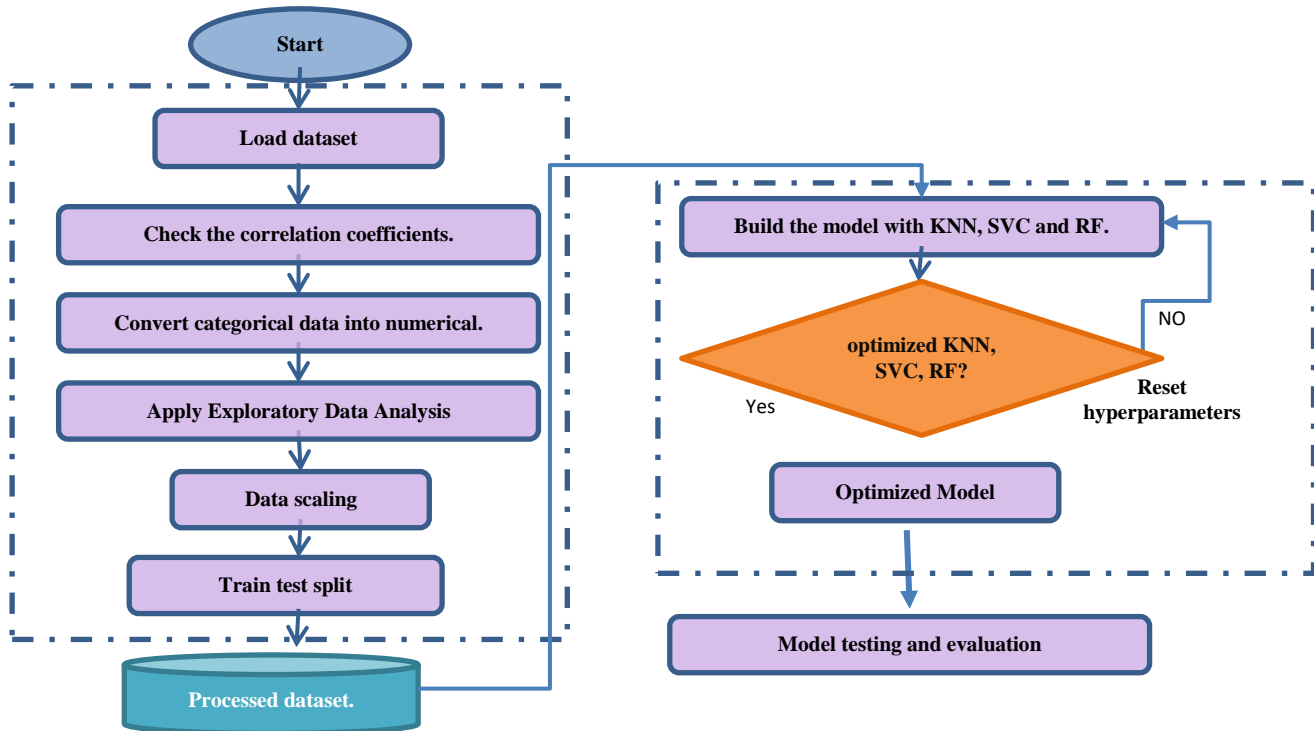


Fig. 2. Flowchart of the proposed Heart Attack Detection Model (HADM).

The proposed model depends on using the K Nearest Neighbors algorithm, support vector machine classifier (SVM) and random forest (RF) to determine whether cardiac disease is present or not. In Fig. 2, the HADM flowchart will be shown. Gridsearch technique is used for hyperparameters optimization. These parameters might be a continuous variable, a categorical variable, or an integer with values spanning from the lower to

upper bounds. As hyperparameters remain constant throughout training, the model's accuracy is increased while training time and memory requirements are concurrently decreased. Different models employ a variety of hyperparameters depending on the issue description. There are no hyperparameters that are optimal for all models.

B. ECG Heartbeat Signals Multi-class Classification Model (ECG- HMCM)

A rigorous and consistent examination by cardiologists is required, which is difficult and time-consuming, to give an effective and precise diagnosis of ARR and CHF. For the proper detection of cardiac disorders, a completely automated diagnosis system is urgently required. The creation of diagnostic systems can help cardiologists diagnose ECG recordings quickly and accurately while also cutting down on the time and money needed for clinical interpretation. The heartbeat classification dataset from the MITBIH is used in this suggested model. Up sampling is done such that there is the same number of instances in each of the five classes, each of which has a different number of samples. The KNN model is trained on the data once it has been divided into training and testing set. The test data is run through the trained model. Individual successfully and erroneously categorized signals are examined in terms of correlation to the class mean as well as plotted to examine explainability. Here, we looked at whether such affordable and straightforward algorithms may be useful enough for clinical usage and pave the way to better services. The model architecture will be shown in Fig. 3. The steps of applying the proposed model will be summarized in the next algorithm.

The proposed algorithm for ECG heartbeat multi-class classification model
<i>Step 1) load the data set,</i>
<i>Step 2) define classes,</i>
<i>Step 3) assign a data frame to each class,</i>
<i>Step 4) extract shape of the training and test sets, Upsample the minority classes.</i>
<i>Step 5) split original training data into a training and test subset (test_size =0.2)</i>
<i>Step 6) apply KNN algorithm on the dataset.</i>
<i>Step 7) attributes are ranked according to their value.</i>
<i>Step 8) apply (Grid search) for hyperparameters optimization.</i>
<i>Step 9) estimate the classifier's accuracy, which measures the capability of the classifier to classify heartbeat ECG signals correctly.</i>

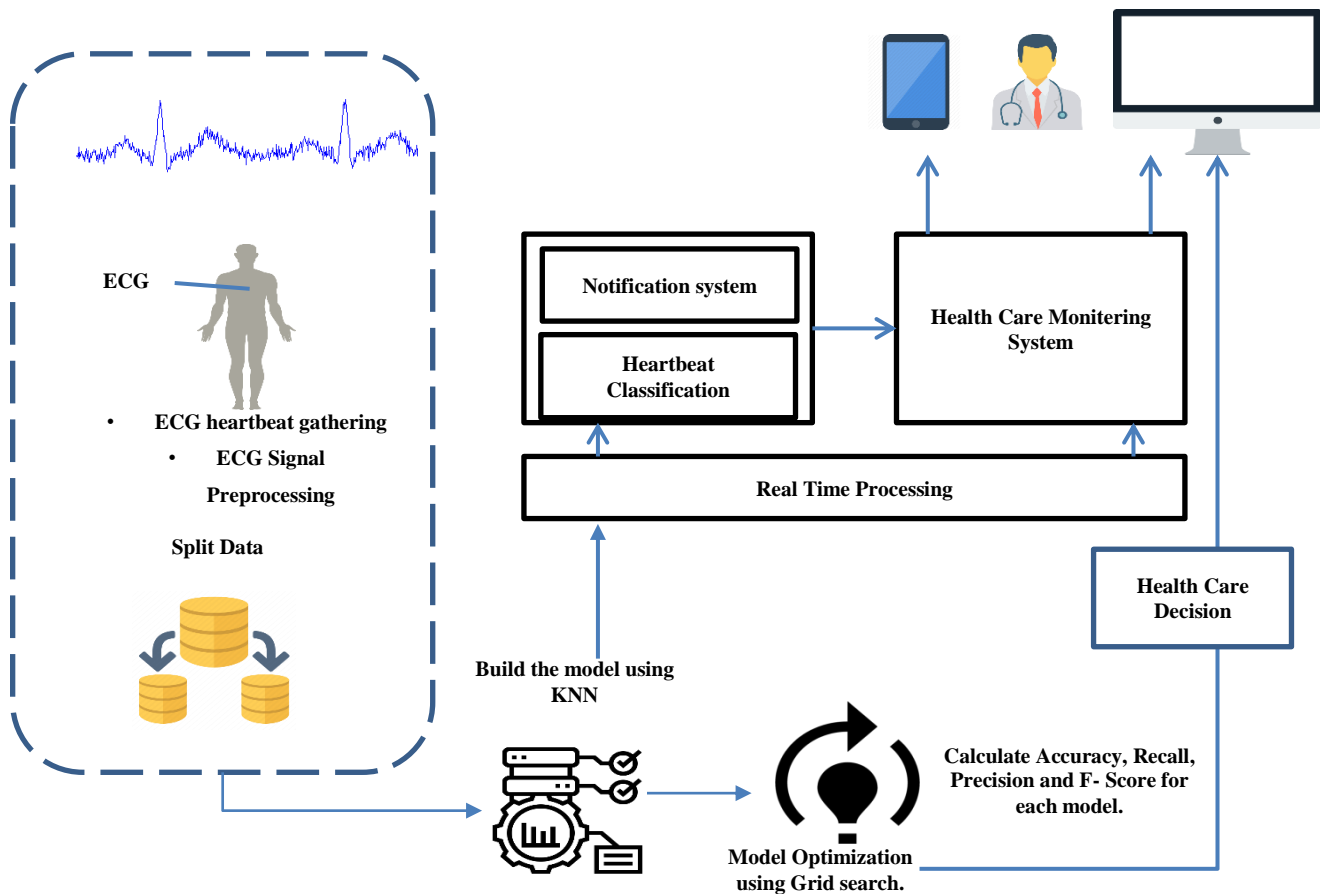


Fig. 3. ECG heartbeat multi-class classification model architecture.

IV. IMPLEMENTATION AND METHODS

In this section, the experiments of the proposed framework will be discussed which are divided into (i) heart attack detection, and (ii) ECG heartbeat signals classification, in addition to the description of the used datasets for both models.

To support the implementation of our suggested modal, our experiments were implemented in Python notebook version (6.4.8) and several open-source libraries. This system makes use of the packages Matplotlib, NumPy, sklearn, and Keras. It was carried out a laptop with a 2.30GHz, 11th-generation Intel Core (TM) i7-11800H processor, and 16.0 GB (15.8 GB usable).

A. Heart Attack Detection Model (HADDM)

The heart is a vital organ in humans. It sends blood to all the organs in our body. If it malfunctions, the body's other organs will stop working and the person will pass away in a matter of seconds. Several different things can lead to heart failure. These components have been split up by scientists into two categories: risk variables that cannot be changed and risk factors that can be. Historical Family data [33], gender, and age are risk aspects that cannot be changed. Excessive cholesterol, smoking [34], inactivity, and blood pressure are risk factors.

The heart attack detection model depends on ML algorithms, namely, KNN, SVM and RF. The KNN algorithm is distinguished by its straightforward design and practical use. KNN was selected because it can compete with the most accurate models and offers incredibly exact predictions. How accurate the forecasts are determined by the distance. KNN technique can be used in applications that demand great accuracy as a result. The SVM algorithm was also selected to participate in our classification model because of its mathematical base in statistical learning theory which provides a systematic solution to ML challenges. A portion of the training input is used to create the solution via SVM. Classification, regression, novelty detection, and feature reduction problems have all seen widespread usage of SVM.

Random forest classifier is also used to build our model. A method for lowering the variance of an estimated prediction function is RF. Regression and classification are also possible uses for RF. A random forest collects class votes from each tree before classifying using a majority vote when used for classification. The predictions from each tree at the target point x are simply summed when used for regression. When used for classification tasks, RF gets a class vote for each tree and then classifies the data using the results of the majority vote. The predictions from all trees at a target point x are merely averaged when RF is used to regression problems. The values of the parameters that are most appropriate for a job rely on those parameters, which are understood and employed as tuning parameters. Then, Gridsearch is separately applied to each technique to get optimized hyperparameters. To assess the intensity of the key elements that influence the prognosis of heart disease, a generalized algorithm was developed and optimized. Cleveland, Hungary, Switzerland, Long Beach and Stalog (CHSLBS) datasets were used to assess the models. The selected ML algorithms were used to build the proposed models for classification. Also, we will use grid search for

tuning models' hyperparameters, additionally we will evaluate their performance.

1) *The used dataset:* The original source for this dataset was the [35]. The dataset variables are mentioned in [35]. By combining multiple datasets that were previously available separately, one dataset was created. This dataset, which includes five heart datasets including eleven features in public, is the greatest heart disease dataset currently available for research objectives. The five datasets utilized for its curation are: Cleveland: 303 observations; Hungarian: 294 observations; Switzerland: 123 observations; Long Beach VA: 200 observations; and 270 observations in the Stalog (Heart) (CHSLBS) datasets. The total number of observations is 1190, but we found 272 duplicated rows. Hence, the preprocessing step is to drop this duplication to finally get 918 rows. Analyzing CHSLBS shows that there are 508 with heart disease and 410 normal cases. This leads us to have a balanced dataset. It was found that the presence of HeartDisease is substantially linked with all numerical characteristics. Hence, we will first get dummies for categorical characteristics. In our experiment, some dataset analysis has been extracted such as maximum, minimum, average, and median to determine whether the data indicate a skewness distribution. Each of these traits is looked at individually because they all have different units and meanings.

The used dataset [35] contains different attributes. Sex, FastingBs and Exercise angina are binary data. Chest Pain type, Resting ECG and ST_slope are categorical data, and the continuous set consists of Age, Resting blood pressure, cholesterol, MaxHR and old peak. Some of preparations steps were executed such as dropping duplicates, Outlier Detection, Features encoding such as Binary features encoding and One-hot encoding were applied on the used dataset. 4.58% of instances that are outliers in resting blood pressure (restingBP) were removed. Since characteristics like cholesterol and resting blood pressure are exceptional situations that do not accurately reflect the state of the general population and are not suitable for training our model of predictions, preprocessing is necessary.

Exploratory Data Analysis EDA [36] seeks to condense descriptions to make them easier to process with the cognitive capacity at hand; and tries to make the description more effective by looking behind the previously mentioned surfaces. The used dataset needs some steps of preprocessing and exploratory data analysis. First, we investigate the correlations between the patient's numerical characteristics and the target column. It was found that the presence of cardiac disease is closely associated with all numerical characteristics. Then, the categorical data is handled to be numerical.

The distribution analysis is a graphical representation of a dataset's distribution that displays the frequency of data points within various intervals. The histogram for distribution analysis of the features is a common graphing technique used to include both continuous and discrete data that was collected on an interval scale. It is widely used to conveniently illustrate the main ideas of data distribution. The conclusion after

applying EDA on the used dataset shows the following information: the risk of suffering CVDs increases with age. Fig. 4 shows the histogram for distribution analysis of some features such as age, oldpeak, resting Blood Pressure, and cholesterol.

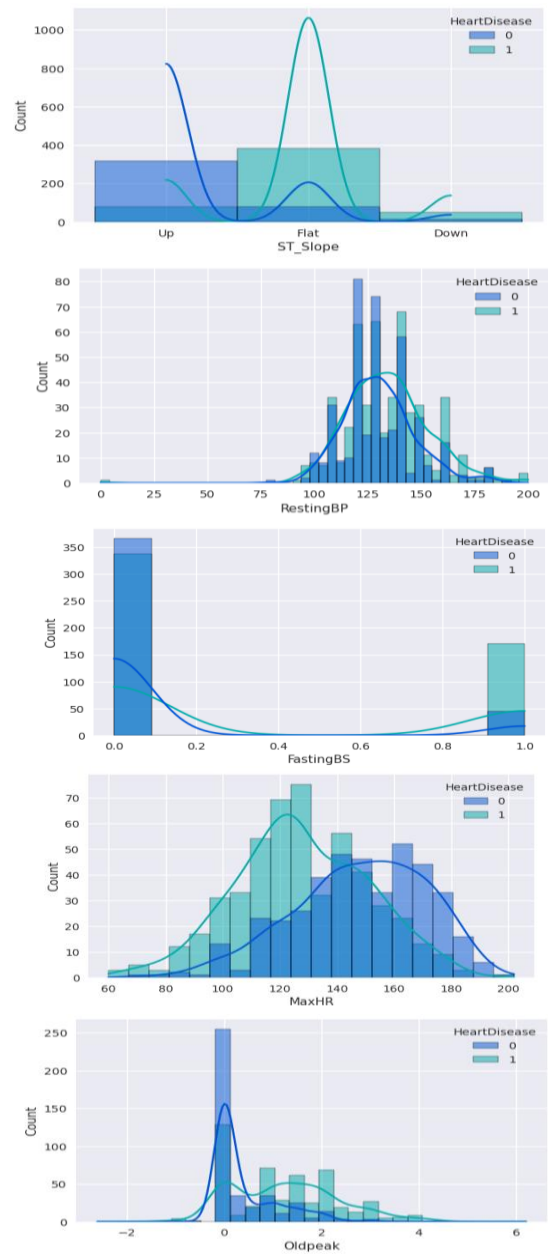
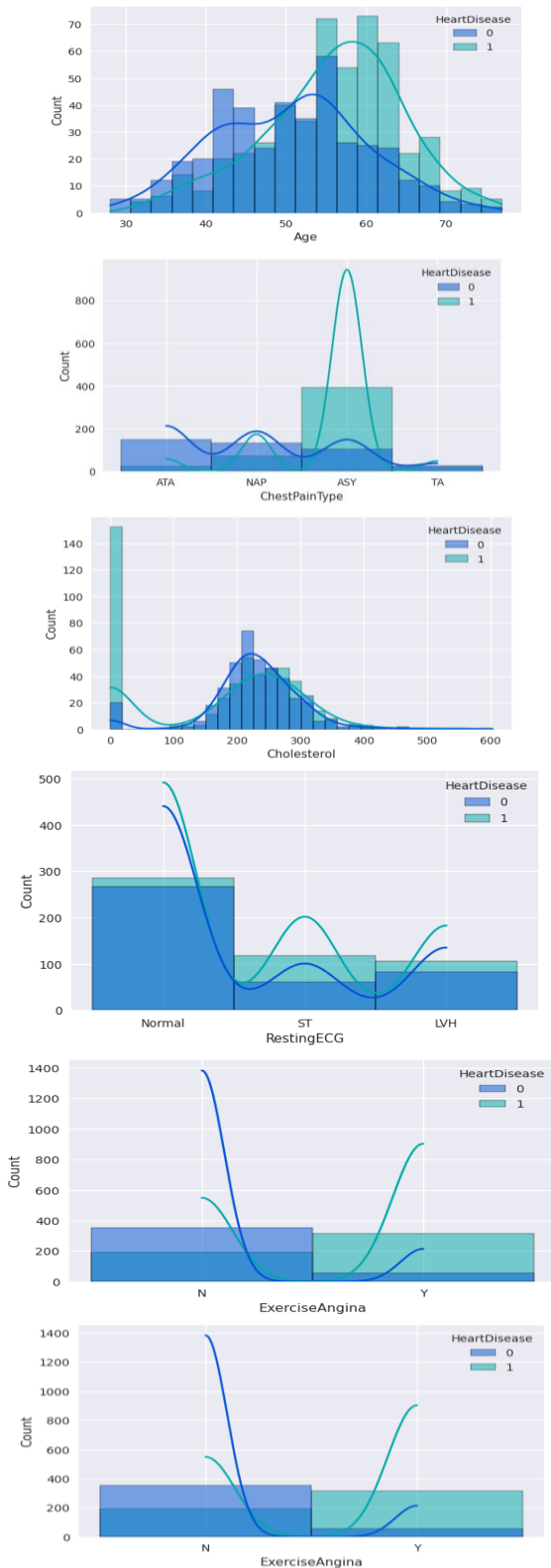


Fig. 4. Distribution analysis of some dataset features.

Apparently, the men have a major risk than women of suffering CVDs. Most cases of CVDs present absence of chest pain or the usual anginal equivalents followed by the cases that present chest pain non-anginal. We also see in this histogram a slight increase in CVDs cases when increasing the resting blood pressure. Fig. 5 will show the histogram for distribution analysis according to Age and oldpeak. Fig. 6 shows the correlation between dataset features.

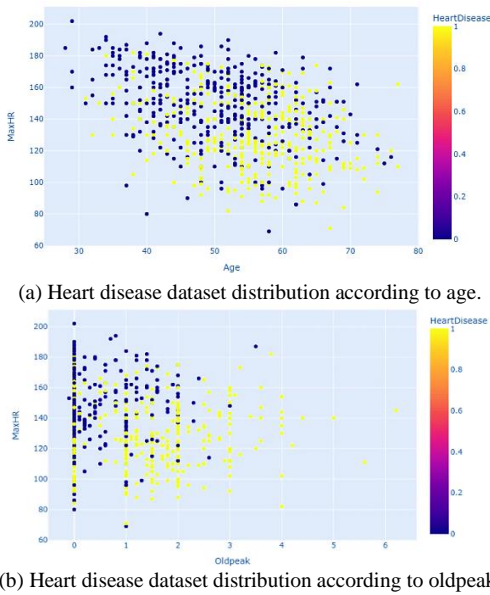


Fig. 5. The histogram for distribution analysis according to Age, oldpeak.

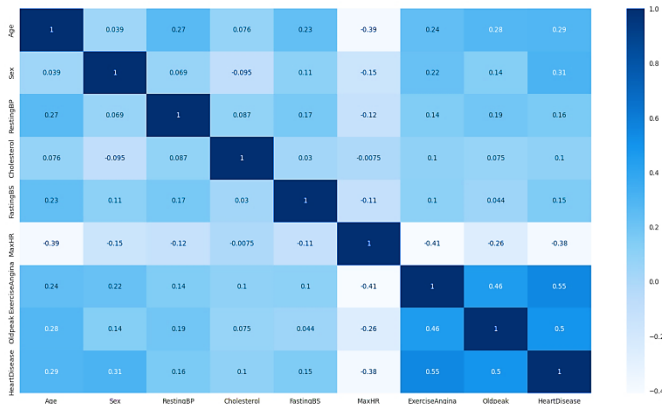


Fig. 6. The correlation between dataset variables.

B. ECG Heartbeat Signal Multi-class Classification

A default KNN model is trained on the data once it has been divided into a training and testing set. The test data is run through the trained model. Individual successfully and erroneously categorized signals are examined in terms of correlation to the class mean as well as plotted to examine explainability. Gridsearch algorithm is used to get the optimized k-parameter, then the model is retrained on the original test set by achieving an enhancement of accuracy.

1) *The used datasets:* Two collections of the signals of heartbeat that add up to this dataset are derived from two well-known datasets in classification of heartbeat, the MIT-BIH Arrhythmia Dataset [37] and The PTB diagnostic ECG database [38]. The MITBIH heartbeat dataset categorization is the subject of this experiment. The dataset contains five classes with these labels (i) normal beat ('N'): 0, (ii) supraventricular ectopic beats ('S'): 1, (iii) ventricular ectopic beats ('V'): 2, (iv) fusion Beats ('F'): 3, (v) unknown Beats ('Q'): 4. Upsampling is done such that there are the same number of instances in each of the five classes, each of which has a various number of

samples. Both sets provide a large enough number of samples to support deep neural network training. This dataset has been used to research the classification of heartbeats and to investigate some of the transfer learning possibilities using deep neural network architecture. The signals correspond to the ECG types of heartbeats for both the normal case and cases affected by different arrhythmias and myocardial infarction. These signals are preprocessed and divided into segments, each of which represents a heartbeat. Arrhythmia Dataset consists of five categories which are labeled to be ['N': 0, 'S': 1, 'V': 2, 'F': 3, 'Q': 4] with 109446 samples, the sampling frequency is 125HZ and the source is from Data Source: Physionet's MIT-BIH Arrhythmia Dataset. The PTB Diagnostic ECG dataset contains two categories with 14552 samples and the source of data is from Physionet's PTB Diagnostic dataset.

C. Evaluation Matrix

This section presents the results of the proposed models and evaluates their performance using AUC. The AUC has been chosen according to the providing equations[39]. Eq. (1) - (5) described the evaluation way of the proposed models. Precision, recall or sensitivity, f1-score, and other model performance indicators were also assessed, in addition to accuracy. To evaluate the model's performance, the true positive, true negative, false positive, and false negative values must first be established. When the model correctly identified positive and negative predictions, they are referred to as true positive and negative values. False positives and negatives, on the other hand, occur when the model incorrectly divides positive and negative predictions into positive and negative categories.

Accuracy:

$$Accuracy = \frac{True\ Negative + True\ Positive}{Total\ Data} \quad (1)$$

Precision:

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \quad (2)$$

Recall:

$$Recall\ (Sensitivity) = \frac{True\ Positive}{True\ Positive + False\ Negative} \quad (3)$$

Specificity:

$$Specificity = \frac{True\ Negative}{True\ Negative + False\ Positive} \quad (4)$$

F1 score:

$$F1 = \frac{2 \times Precision \times recall}{Precision + recall} \quad (5)$$

D. Model Optimization

Hyperparameters are peripheral parameters[40] that are not a model component and cannot be predicted from the dataset. For classification challenges, a user must manage the hyperparameters setting procedure, which varies for various algorithms and datasets, to obtain high accuracy. Finding a tuple that suggests an ultimate model and minimizes the loss function is the goal of optimization [41]. A Gridsearch algorithm was used to optimize the hyperparameters for the

experimental classification model when developing HADM using KNN, SVC and RF classification techniques. Table I lists the top hyperparameters for each used technique.

TABLE I. THE OPTIMIZED HYPERPARAMETERS FOR THE USED TECHNIQUES

Technique	Best hyperparameters
KNN	'n_neighbors': 9, 'p': 2, 'weights': 'distance'
SVC	'C': 1.0, 'gamma': 'auto', 'kernel': 'sigmoid'
RF	'criterion': 'gini', 'max_features': 30, 'n_estimators': 700,

V. RESULTS AND DISCUSSIONS

This evaluation section consists of (i) heart attack detection model, (ii) ECG heartbeat signal multi-class classification to be proposed models that can early support with an appropriate decision for saving patients.

A. Heart Attack Detection Results

The heart attack detection model depends on the use of KNN, SVC and RF classifiers achieved the best accuracy compared with another previous research when applied on the greatest heart disease dataset which consists of five datasets with 1190 observations formed (CHSLBS) datasets. During the execution of the experiments, Gridsearch is applied to optimize the hyperparameters for achieving higher accuracy. Area under the receiver operating characteristic curve (AUC), accuracy, sensitivity, and specificity were among the results that were reported. AUC curves will be shown in Fig. 7 for KNN, SVC and RF. Table II shows the summarized results for building the model of Heart Attack Detection.

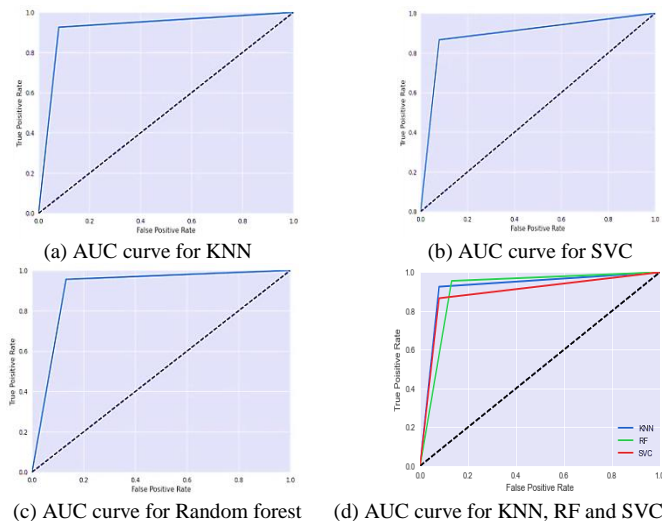


Fig. 7. The dataset’s AUC curve for test data using the proposed models.

TABLE II. THE EVALUATION OF HEART ATTACK DETECTION MODEL USING THE EVALUATION MATRIX

Technique	K-Nearest Neighbors	C-Support Vector	Random Forest
Accuracy	92.307692	89.510490	90.909091
Precision	91.176471	90.625000	86.486486
Recall	92.537313	86.567164	95.522388
f1- score	92.310711	89.491892	90.915316

We must find a balance between recall score and accuracy score; in the case of our focus, the Random Forest classifier had the greatest recall score, but also the poorest precision score. Hence, the KNN classifier is the best algorithm in this experiment. This study demonstrates that the RF classifier outperforms other classification algorithms when employing Gridsearch for hyperparameters optimization [40] and evaluating depending on the recall evaluation matrix parameter. KNN achieved higher accuracy compared with random forest and support vector classifier algorithms. Table III shows comparison between the result of our proposed result and other used techniques.

TABLE III. COMPARISON OF ACHIEVED RESULTS AND PREVIOUS RESEARCH

References	Year	The used technique	Accuracy
Saqlain, et al. [42]	2019	SVM and MCC	81.91%
Beunza, et al. [43]	2019	NB, BN, RF, and MP	85.48%
Kavitha, et al.[19]	2020	Hybrid DT and RF	88%
Kavitha, et al.[44]	2021	KNN with SBS feature selection	90%
Pawlovsky, et al. [18]	2022	KNN	85%
Anderies, et al. [20]	2022	KNN	86%
Chauhan, et al. [21]	2023	PSO, KNN	90.28%
The proposed model	2023	KNN, SVC and RF	92.31%, 89.51%, 90.91%

B. ECG Heartbeat Monitoring Results

In this model, KNN is used for classifying the ECG heartbeat for patient monitoring as a proposed solution to assist in the decision-making support health system. The achieved results were optimized according to using the grid search optimizer. The model is instantiated and performed depending on the use of a grid search for setting the number of neighbors (the k parameter value) fitting five folds for each of five candidates, for a total of 25 fits, for the first 40,000 shuffled training data. In Fig. 8, the preview of 5 ECG beats from the five categories will be shown when using the training dataset.

The finding results have been enhanced 0.5 % after using the grid search hyperparameter optimization and 97.5% Overall classification accuracy. Fig. 9 will preview the signals for the correct prediction according to our proposed model when applied on the test set.

This shows how the suggested approach might help cardiologists improve the precision of ECG diagnosis in real-time clinical situations.

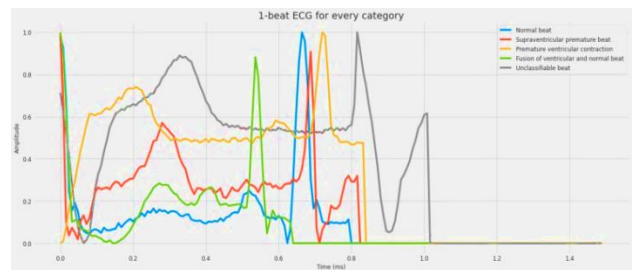


Fig. 8. Previewing 1 beat ECG from each category.

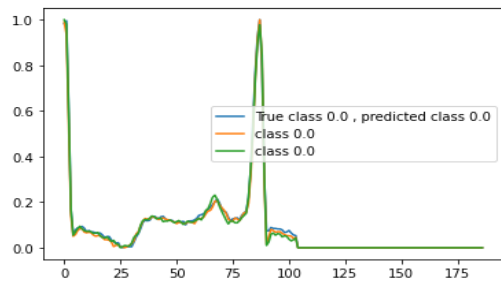


Fig. 9. Sample of correct predicted signal.

The proposed model has some advantages over previous works. The model combines heart rate detection and ECG classification in one IoT based framework with a notification system. The model also enhanced the accuracy of heart attack detection over most previous works. It is anticipated that the created method would make it much easier for doctors to diagnose heart issues in a more practical way. Above all, the study has significantly improved the computation of strength ratings, which are excellent indicators of the prognosis of heart conditions depending on the proposed framework.

VI. CONCLUSIONS AND FUTURE WORK

The leading cause of death is heart disease. Heart disease mortality will decrease because of the identification of key risk factors, the creation of decision support systems, effective control methods, and health education initiatives. If CVDs are detected and treated early, many lives might be saved. Cardiologists cannot manually examine the vast quantity of collected data for a patient to develop a timely medication plan. One of the most widely utilized ML algorithms, KNN, SVM and RF are frequently used for data categorization in our proposed model. The five datasets utilized for its curation are: Cleveland: 303 observations; Hungarian: 294 observations; Switzerland: 123 observations; Long Beach VA: 200 observations; and 270 observations in the Stalog (Heart) (CHSLBS) datasets. The total number of observations is 1190 and the achieved accuracy were 92.3%, 89.51% and 90.91% when applying KNN, SVM classifier and RF recursively with hyperparameters optimization using Gridsearch technique. Analyzing each patient's health metrics can also help to forecast heart disease. Hence, KNN is used for heart attack detection using UCI dataset, in addition to using KNN to build an ECG heartbeat multi class classification model. Two heartbeat signals were examined and evaluated in this study. The Arrhythmia and PTB diagnostic ECG datasets, both of which have 125 Hz sampling rates, each comprise 109446 samples in five categories and 14552 samples in two categories. Using two ECG datasets, the KNN methodology with the hyperparameters optimizer employing Gridsearch algorithm was applied, and higher accuracy has been achieved when compared with previous recent research in this field was attained. The results of this model have been optimized using grid search for hyperparameters optimization. The major goals of the proposed framework are to provide a method that can effectively identify heart attacks and help professionals to choose the best decision for saving human life. In future work, other different ML techniques can be used. By adding more data and testing with more important or statistically produced

data, such as numeric data augmentation, the applicable procedure may be enhanced. The writers see this as a future work that may be improved.

REFERENCES

- [1] Centers for Disease Control and Prevention (CDC). Deaths: leading causes. Available: <https://www.cdc.gov/nchs/fastats/leading-causes-of-death>.
- [2] S. Kaptoge, L. Pennells, D. De Bacquer, M. T. Cooney, M. Kavousi, G. Stevens, et al., "World Health Organization cardiovascular disease risk charts: revised models to estimate risk in 21 global regions," *The Lancet Global Health*, vol. 7, pp. e1332-e1345, 2019.
- [3] M. R. Cowie, J. I. Blomster, L. H. Curtis, S. Duclaux, I. Ford, F. Fritz, et al., "Electronic health records to facilitate clinical research," *Clinical Research in Cardiology*, vol. 106, pp. 1-9, 2017.
- [4] H. Shimizu and K. I. Nakayama, "Artificial intelligence in oncology," *Cancer science*, vol. 111, pp. 1452-1460, 2020.
- [5] P. Mascagni, A. Vardazaryan, D. Alapatt, T. Urade, T. Emre, C. Fiorillo, et al., "Artificial intelligence for surgical safety: automatic assessment of the critical view of safety in laparoscopic cholecystectomy using deep learning," *Annals of surgery*, vol. 275, pp. 955-961, 2022.
- [6] T. Shan, F. Tay, and L. Gu, "Application of artificial intelligence in dentistry," *Journal of dental research*, vol. 100, pp. 232-244, 2021.
- [7] D. Itchhaporia, "Artificial intelligence in cardiology," *Trends in cardiovascular medicine*, 2020.
- [8] S. Liu, Y. Sun, and N. Luo, "Doppler Ultrasound Imaging Combined with Fetal Heart Detection in Predicting Fetal Distress in Pregnancy-Induced Hypertension under the Guidance of Artificial Intelligence Algorithm," *Journal of Healthcare Engineering*, vol. 2021, 2021.
- [9] M. Alnaggar, M. Handosa, T. Medhat, and M. Z. Rashad, "Thyroid Disease Multi-class Classification based on Optimized Gradient Boosting Model," *Egyptian Journal of Artificial Intelligence*, 2023.
- [10] A. I. Siam, N. A. El-Bahnasawy, G. M. El-Banby, A. Abou Elazm, and F. E. Abd El-Samie, "Efficient video-based breathing pattern and respiration rate monitoring for remote health monitoring," *JOSA A*, vol. 37, pp. C118-C124, 2020.
- [11] M. Alnaggar, A. I. Siam, M. Handosa, T. Medhat, and M. Rashad, "Video-based Real-Time Monitoring for Heart Rate and Respiration Rate," *Expert Systems with Applications*, p. 120135, 2023.
- [12] A. I. Siam, M. A. El-Affendi, A. Abou Elazm, G. M. El-Banby, N. A. El-Bahnasawy, F. E. Abd El-Samie, et al., "Portable and Real-Time IoT-Based Healthcare Monitoring System for Daily Medical Applications," *IEEE Transactions on Computational Social Systems*, 2022.
- [13] A. Qayyum, J. Qadir, M. Bilal, and A. Al-Fuqaha, "Secure and robust machine learning for healthcare: A survey," *IEEE Reviews in Biomedical Engineering*, vol. 14, pp. 156-180, 2020.
- [14] R. Karthick, R. Ramkumar, M. Akram, and M. V. Kumar, "Overcome the challenges in bio-medical instruments using IOT-A review," *Materials Today: Proceedings*, vol. 45, pp. 1614-1619, 2021.
- [15] S. Serte, A. Serener, and F. Al-Turjman, "Deep learning in medical imaging: A brief review," *Transactions on Emerging Telecommunications Technologies*, vol. 33, p. e4080, 2022.
- [16] A. Mehmood, M. Iqbal, Z. Mehmood, A. Irtaza, M. Nawaz, T. Nazir, et al., "Prediction of heart disease using deep convolutional neural networks," *Arabian Journal for Science and Engineering*, vol. 46, pp. 3409-3422, 2021.
- [17] M. M. Eisa and M. H. Alnaggar, "Hybrid Rough-Genetic Classification Model for IoT Heart Disease Monitoring System," in *Digital Transformation Technology*, ed: Springer, 2022, pp. 437-451.
- [18] A. P. Pawlovsky, "An ensemble based on distances for a kNN method for heart disease diagnosis," in 2018 international conference on electronics, information, and communication (ICEIC), 2018, pp. 1-4.
- [19] R. Indrakumari, T. Poongodi, and S. R. Jena, "Heart disease prediction using exploratory data analysis," *Procedia Computer Science*, vol. 173, pp. 130-139, 2020.
- [20] A. Anderies, J. A. R. W. Tchin, P. H. Putro, Y. P. Darmawan, and A. A. S. Gunawan, "Prediction of Heart Disease UCI Dataset Using Machine

- Learning Algorithms," Engineering, Mathematics and Computer Science (EMACS) Journal, vol. 4, pp. 87-93, 2022.
- [21] A. Chauhan, A. Jain, P. Sharma, and V. Deep, "Heart disease prediction using evolutionary rule learning," in 2018 4th International conference on computational intelligence & communication technology (CICT), 2018, pp. 1-4.
- [22] F. S. Alotaibi, "Implementation of machine learning model to predict heart failure disease," International Journal of Advanced Computer Science and Applications, vol. 10, 2019.
- [23] D. Shah, S. Patel, and S. K. Bharti, "Heart disease prediction using machine learning techniques," SN Computer Science, vol. 1, pp. 1-6, 2020.
- [24] H. Jindal, S. Agrawal, R. Khera, R. Jain, and P. Nagrath, "Heart disease prediction using machine learning algorithms," in IOP conference series: materials science and engineering, 2021, p. 012072.
- [25] H. ChuDuc, K. NguyenPhan, and D. NguyenViet, "A review of heart rate variability and its applications," APCBEE procedia, vol. 7, pp. 80-85, 2013.
- [26] P. Ponikowski, A. A. Voors, S. D. Anker, H. Bueno, J. G. Cleland, A. J. Coats, et al., "2016 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure," *Kardiologia Polska (Polish Heart Journal)*, vol. 74, pp. 1037-1147, 2016.
- [27] S. S. Sarmah, "An efficient IoT-based patient monitoring and heart disease prediction system using deep learning modified neural network," *Ieee access*, vol. 8, pp. 135784-135797, 2020.
- [28] M. Kachuee, S. Fazeli, and M. Sarrafzadeh, "Ecg heartbeat classification: A deep transferable representation," in 2018 IEEE international conference on healthcare informatics (ICHI), 2018, pp. 443-444.
- [29] E. Jing, H. Zhang, Z. Li, Y. Liu, Z. Ji, and I. Ganchev, "ECG heartbeat classification based on an improved ResNet-18 model," *Computational and Mathematical Methods in Medicine*, vol. 2021, 2021.
- [30] S. Raj and K. C. Ray, "A personalized arrhythmia monitoring platform," *Scientific reports*, vol. 8, pp. 1-11, 2018.
- [31] I. Shiue, "Are urinary polyaromatic hydrocarbons associated with adult hypertension, heart attack, and cancer? USA NHANES, 2011-2012," *Environmental Science and Pollution Research*, vol. 22, pp. 16962-16968, 2015.
- [32] J. Stewart, T. McCallin, J. Martinez, S. Chacko, and S. Yusuf, "Hyperlipidemia," *Pediatrics in review*, vol. 41, pp. 393-402, 2020.
- [33] K.-T. Khaw and E. Barrett-Connor, "Family history of heart attack: a modifiable risk factor?," *Circulation*, vol. 74, pp. 239-244, 1986.
- [34] D. Cook, S. Pocock, A. Shaper, and S. Kussick, "Giving up smoking and the risk of heart attacks: a report from the British Regional Heart Study," *The Lancet*, vol. 328, pp. 1376-1380, 1986.
- [35] Heart Failure Prediction Dataset. Retrieved [Date Retrieved] from [Online]. Available: <https://www.kaggle.com/fedesoriano/heart-failure-prediction>.
- [36] G. Wang, B. Zhao, B. Wu, C. Zhang, and W. Liu, "Intelligent prediction of slope stability based on visual exploratory data analysis of 77 in situ cases," *International Journal of Mining Science and Technology*, vol. 33, pp. 47-59, 2023.
- [37] MIT-BIH Arrhythmia Database [Online]. Available: <https://www.physionet.org/content/mitdb/1.0.0/>.
- [38] PTB Diagnostic ECG Database [Online]. Available: <https://www.physionet.org/content/ptbdb/1.0.0/>.
- [39] M. J. Zaki, W. Meira Jr, and W. Meira, *Data mining and analysis: fundamental concepts and algorithms*: Cambridge University Press, 2014.
- [40] H. Alibrahim and S. A. Ludwig, "Hyperparameter optimization: comparing genetic algorithm against grid search and bayesian optimization," in 2021 IEEE Congress on Evolutionary Computation (CEC), 2021, pp. 1551-1559.
- [41] I. Priyadarshini and C. Cotton, "A novel LSTM-CNN-grid search-based deep neural network for sentiment analysis," *The Journal of Supercomputing*, vol. 77, pp. 13911-13932, 2021.
- [42] 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.
- [43] J.-J. Beunza, E. Puertas, E. García-Ovejero, G. Villalba, E. Condes, G. Koleva, et al., "Comparison of machine learning algorithms for clinical event prediction (risk of coronary heart disease)," *Journal of biomedical informatics*, vol. 97, p. 103257, 2019.
- [44] M. Kavitha, G. Gnaneswar, R. Dinesh, Y. R. Sai, and R. S. Suraj, "Heart disease prediction using hybrid machine learning model," in 2021 6th international conference on inventive computation technologies (ICICT), 2021, pp. 1329-1333.

Solar Energy Forecasting Based on Complex Valued Auto-encoder and Recurrent Neural Network

Aymen Rhouma¹, Yahia Said²

Department of Computer Science-College of Science and Humanities, Afif, Shaqra University, Saudi Arabia¹
Remote Sensing Unit, College of Engineering, Northern Border University, Arar, Saudi Arabia²
Laboratory of Electronics and Microelectronics (LR99ES30), University of Monastir, Tunisia²

Abstract—Renewable energy is becoming a trusted power source. Energy forecasting is an important research field, which is used to provide information about the future power generation of renewable energy plants. Energy forecasting helps to safely manage the power grid by minimizing the operational cost of energy production. Recent advances in energy forecasting based on deep learning techniques have shown great success but the achieved results still too far from the target results. Ordinary deep learning models have been used for time series processing. In this paper, a complex-valued autoencoder was coupled with an LSTM neural network for solar energy forecasting. The complex-valued autoencoder was used to process the time series with the advantage of processing more complex data with more input arguments. The energy value was used as a real value and the weather condition was considered as the imaginary value. Taking into account the weather condition helps to better predict power generation. The proposed approach was evaluated on the Fingrid open data dataset. The mean absolute error (MAE), root-mean-square error (RMSE) and mean absolute percentage error (MAPE) was used to evaluate the performance of the proposed method. A comparison study was performed to prove the efficiency of the proposed approach. Reported results have shown the efficiency of the proposed approach.

Keywords—Solar energy forecasting; artificial intelligence; complex-valued autoencoder; long-short term memory; deep learning

I. INTRODUCTION

Modern cities have elevated the power demand and fuel-based energy sources are polluting the globe. Renewable energy has been considered as a solution to provide power needs while reducing pollution and power production costs. Sun, wind, and water are natural clean power sources that can be considered trusted sources, but their power production was based on weather conditions. Recently, the energy produced by clean resources was integrated into the power grid. The main concern of those sources is that the generated power cannot be controlled and varies with weather conditions. Integrating clean power into the grid is a very challenging task because of the variability of the generated energy amount.

Solar power is a clean energy source that converts daylight to electricity. The main issue of this type of energy generator is that the amount of generated power is unpredictable. Generated solar power must be consumed at the same time of production because of the volatility of this energy. However, overload power causes many problems such as voltage regulation and reverse power flow. Besides, low power can cause

discontinuity and can affect all devices connected to the grid. Subsequently, for the safe integration of the solar power generator into the grid, it is important to predict the grid power need and integrate the solar power at its stable generation period to benefit from it the maximum possible. A high-performance energy forecasting system can help to solve the problem. Mainly there are two energy forecasting systems, physical models and artificial intelligence models. Physical models are based only on whether conditions generate a forecast based on wind speed or solar power curves based on light availability. Those models were not very accurate because many other conditions must be taken into account such as solar panels' conditions and temperature. However, artificial intelligence models have been efficiently and flexibly used for energy forecasting. Those models do not need to explicitly scale the physical process of power generation. In effect, it builds a relationship between the input and output through processing data and reducing the error with the target values.

Energy forecasting has been widely studied to manage energy production costs by predicting clean power generation. Many averaging techniques were deployed for energy forecasting such as Autoregressive Integrated Moving Average (ARIMA) [1], simple moving average (SMA) [2], and weighted moving average (WMA) [3]. The averaging techniques have achieved low performance and the predicted results were different from the expected. Methods such as simple exponential smoothing (SES) [4] were used for forecasting but the achieved results were not convincing.

Recently, the deep learning technique [5] has achieved great success in many applications. Deep learning techniques are based on deep neural networks. Convolutional neural networks [6], recurrent neural networks [7], and auto-encoder [8] were the most used deep learning models. The mentioned deep learning models have boosted the performance of many applications such as object detection [9, 31], indoor object detection [10, 29], face mask detection [11], pedestrian detection [12], traffic signs detection [30], time series forecasting [13], and many others. The outcomes demonstrated the effectiveness of deep learning models for solving many problems and can be used to solve more problems. The ability of automatically learning features without any handcrafted engineer features has been very effective. Besides, such models mimic the biological brain's decision-making process. Considered to solve renewable energy forecasting using deep learning models. Energy forecasting is based on analyzing historical time-series records. Similar to economic time series

forecasting [14], renewable energy forecasting needs deep learning models with memory such as recurrent neural networks. For better performance, we proposed to use an autoencoder to denoise and better represent the input data. Then, a long-short-term memory network was used to process the temporal information of the data for power prediction. The proposed autoencoder is a complex-valued autoencoder with real and imaginary outputs. The energy value was used as real input and the weather condition was used as an imaginary value. Combining the energy value with the weather condition helps to learn more relevant features and generate more trusted predictions.

The proposed method's key benefit is that it may be customized for various forecasting categories based on the horizon range such as short, middle, and long-range forecasting. The proposed method can apply for data forecasting for hours, days, or months. This advantage makes the model useful for many cases and not designed for a specific case.

The main contribution of this paper is the following: (1) proposing a solar power forecasting system; (2) proposing the use of a complex-valued autoencoder for data denoising and features clustering; (3) combining the complex-valued autoencoder with an LSTM neural network to process time-series data for energy forecasting; (4) validate the proposed approach on the open solar dataset.

The paper is organized as follows: Section II presents related works. The proposed approach is described and detailed in Section III. In Section IV, the experiment and results are presented and discussed. The paper was concluded in Section V.

II. RELATED WORKS

The application of renewable energy forecasting for grid management and cost reduction makes it a significant study area. Many works have been proposed for achieving trusted predictions. For a complete overview of existing works, the readers can refer to the survey presented in [15]. Energy forecasting techniques can be categorized based on the base model. The first category is based on averaging techniques and the second category is based on deep learning techniques. In this paper, we have focused on deep learning techniques.

Abedinia et al. [16] proposed a solar energy forecasting method based on neural networks combined with the metaheuristic algorithm. The metaheuristic algorithm was used to optimize the forecasting parameters. The neural network was composed of an input layer, 3 hidden layers, and output layers to generate a prediction. The Levenberg-Marquardt algorithm [17] was used to train the suggested neural network. The Levenberg-Marquardt presents a combination between the gradient descent algorithm [18] and the Newton technique [19]. After training the neural networks, the optimized weights were optimized using the metaheuristic algorithm based on the improved shark smell optimization (ISSO) technique. The proposed method was evaluated in a real solar energy production case.

In [20], a photovoltaic power forecasting model is based on the combination of the stationary wavelet transform, an LSTM

network, and a deep neural network. First, the stationary wavelet transformation was used to fix the high fluctuation and the non-stationary behavior of the input data. The stationary wavelet transformation allows the transformation of the original photovoltaic power data to wavelets which makes the data processible and manageable. Second, an ensemble of four LSTM networks was used to extract temporal features of the input data. The extracted features were reconstructed using the inverse stationary wavelet transformation. Third, statistical features were extracted from the input data by calculating many features such as the Mean, Standard Deviation, Variance, Skewness, and Kurtosis. The statistical features, the reconstructed data from the LSTM network ensemble, and the temperature data were combined and used to train the deep neural network to generate power prediction. The proposed method was evaluated for horizon forecasting but it can be adapted for multi-step forecasting based on the resolution of the available input data.

A solar energy forecasting method was proposed in [21] based on combining the LSTM network with a convolutional neural network. The LSTM network was first used for temporal features extraction and the convolutional neural network was then used for the extraction of the spatial features. Experimental results have shown that extracting temporal features and then extracting spatial features is more effective than extracting spatial features and then extracting temporal features. The proposed model was compared to existing models based on a single network either Convolutional Neural Network or LSTM network and presented superior performance.

Lin et al. [22] proposed the use of a temporal convolutional neural network for solar energy forecasting. The temporal convolutional neural network is an improved convolutional neural network used for temporal data processing. It is composed of a combination of regular convolution layers, dilated convolution layers [23], and residual connection [24]. A temporal convolutional neural network has a hierarchical architecture with several hidden layers that have the same size as the input layer. This type of convolutional neural network was designed for autoregressive prediction with a long memory. The proposed method proved its efficiency compared to regular neural networks such as multi-layer feedforward neural network, LSTM neural network, and GRU neural network.

The majority of the aforementioned techniques are intended for processing historical data for energy power forecasting. Generally, solar energy production is not stationary and not predictable and it is challenging to process it. In this work, we propose to use the weather condition as additional input data to generate more trusted predictions.

III. PROPOSED METHOD

We outline the suggested strategy for solar energy forecasting in this section, which combines a complex-valued autoencoder and an LSTM neural network. In the first subsection, the complex-valued autoencoder will be detailed. In the second subsection, the background of the LSTM neural network will be presented and detailed. The last subsection will

be reserved for explaining the proposed method for solar energy forecasting.

A. Complex Valued Autoencoder

Complex valued autoencoders are the same as autoencoders but with an input, weights, and a mapping function in the complex space. A complex space variable has two parts, one of which is imaginary, and a real part. Assuming a variable x , real part a , and the imaginary b . An example of complex variable can be represented as Eq. (1).

$$x = a + jb \quad (1)$$

Variables in the complex space have more information. A complex-valued autoencoder can be used to process more complex data without raising the computational complexity of the model. Generally, an autoencoder consists of an input layer, a hidden layer, and an output layer. It is used for data reconstruction by regenerating the input data in the output with a minimum error. A complex-valued autoencoder takes as input a real value Re and an imaginary value Im . A combination function is used; C was used to both values to generate the input of the autoencoder. Fig. 1 presented the structure of a complex-valued autoencoder. The autoencoder has two main processes, the encoding process, and the decoding process. The encoding process compresses the data through the hidden layer. The decoding process explodes the data size to reconstruct the input data while presenting more reliable features.

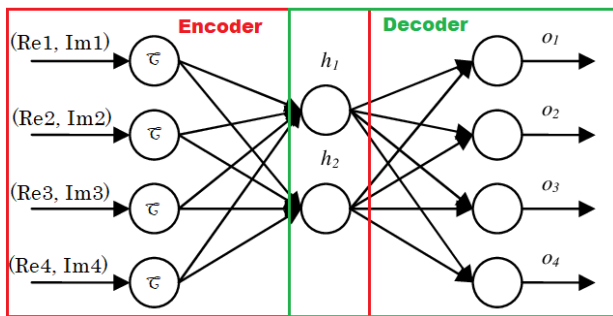


Fig. 1. Complex-valued autoencoder.

Generally speaking, the encoding process is a mapping function that takes the input X and generates an output Y with a size less than the input. The mapping function can be computed as Eq. (2).

$$Y = f(X) = E(wX + b_e) \quad (2)$$

Where w is the weights matrix and the bias vector $b_e \in \mathbb{C}^n$ of the neuron connections. E is a nonlinear activation function. All variables are presented in the complex space. The decoding process is an inverse mapping function that takes Y as input and reconstructs the input data X' . The decode function can be computed as Eq. (3).

$$X' = f'(y) = D(w'Y + b_d) \quad (3)$$

For the encoding process, the radial basis function (RBF) was proposed as a non-linear activation function. The RBF is a very useful function for time-series prediction and

approximations. The RBF activation function with complex variables can be computed as Eq. (4).

$$\phi_k(x_i) = \exp\left(\frac{-j\|x_i - c_h\|^2}{2\sigma_k^2}\right) \quad (4)$$

Where x_i is the hidden layer's h^{th} neuron's center is designated as c_h and the σ_k^2 is the square of the variance of this neuron. $\|x_i - c_h\|$ define the Euclidian distance between the input data and the center of the hidden layer.

For the output layer, the sigmoid function was proposed as an activation function. The sigmoid function is a bounded function between zero and one. Therefore, it is very useful for probability prediction. Besides, the sigmoid function is differentiable, and the slope can be found between every two points. The sigmoid activation function can be computed as Eq. (5).

$$\phi_k(x_i) = \frac{1}{1+e^{jx_i}} \quad (5)$$

In summary, the proposed complex-valued autoencoder has three layers, an input layer, a hidden layer, and an output layer. For the hidden layer input, the RBF activation function was used. For the output layer, the sigmoid function was used. The input values were concatenated through the C function to provide the complex representation of the input. All weights, activations, and biases were presented in complex space.

B. LSTM Neural Network

The LSTM neural network is a recurrent neural network variant with better temporal data processing. LSTM is very useful for long-term memory which is very useful for time series forecasting. In addition, the LSTM allows avoiding the gradient vanishing problem through processing longer input sequences. It is composed of an input gate, an output gate and a forget gate. As the same as recurrent neural networks, LSTM takes into account the previous time step data value. The input gate controls as to which features are important and should be let through the network. The output gate decides which features that will be pushed to the next layer. The forget gate decides which features to reject. The architecture of an LSTM unit is presented in Fig. 2. An LSTM unit can add and remove features to the current state $S^{(t)}$. Adding features can be performed through the input state and removing features can be performed through the forget gate. The output of the unit is filtered using an activation function (\tanh).

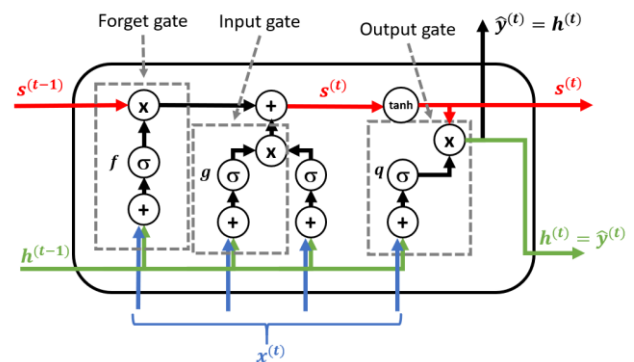


Fig. 2. LSTM neural network unit.

C. Proposed Approach for Solar Energy Forecasting

Solar energy forecasting is based on historical time series processing with consideration to other factors such as weather conditions and the solar panels' condition. Considering that the solar panel produces the maximum possible power, it remains the weather condition to focus on for a trusted prediction on power generation.

The primary goal of the suggested approach was to merge the time-series through the complex-valued autoencoder and use the reconstructed single to train the LSTM neural network for solar energy forecasting. The LSTM neural network learns temporal features according to weather conditions and predicts solar energy production.

The proposed LSTM neural network is composed of three levels of LSTM units. All levels have the same dimension as the input. The proposed LSTM neural network is presented in Fig. 3.

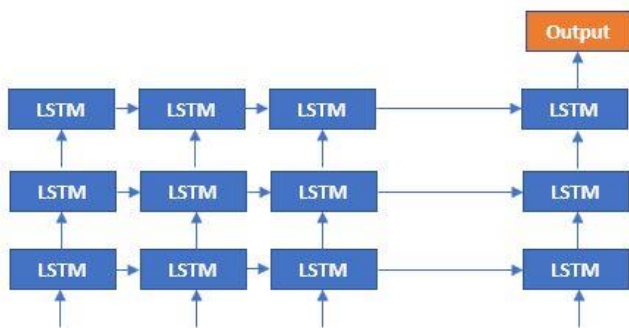


Fig. 3. Proposed LSTM neural network for solar energy forecasting.

In this work, we proposed the combination of a complex-valued autoencoder and an LSTM neural network. The complex-valued autoencoder takes as input the historical time series and the weather conditions. The input data were reconstructed in the output while more features were extracted. The output of the autoencoder was pushed to the LSTM neural network. First, the complex-valued autoencoder was trained separately and then the LSTM neural network. Then, the pretrained complex-valued autoencoder was connected to the LSTM neural network and both networks were trained jointly and considered as a single network. The flowchart of the proposed approach is presented in Fig. 4.

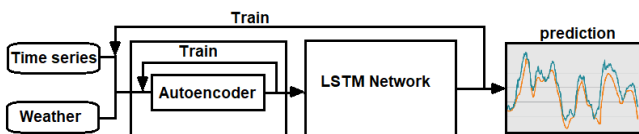


Fig. 4. Proposed approach for solar energy forecasting.

IV. EXPERIMENTS AND RESULTS

All experiments of this work were performed on a desktop with a Linux operating system equipped with an Intel i7 CPU and an Nvidia GTX960 GPU. All models were developed based on the keras deep learning framework. The training data was collected from the Fingrid open data [25]. The collected data was updated hourly. The historical time series and the weather conditions were estimated from a solar power facility

installed in Finland. The facility has a total capacity of 1 megawatt per hour. The data was downloaded in the CSV format. The training data was normalized between zero and one based on the min-max normalization technique. Data normalization allows the comparison of the obtained performances without taking into account the capacity of the solar power facility. The data were filtered by eliminating power values at night and on cloudy days. 60% of the data was used for training, 10% for validation, and 30% for testing after dividing it into three sets.

The proposed complex-valued autoencoder was trained using the gradient descent algorithm with a learning rate of 0.002 and a weight decay of 0.002. It was trained and then its layers were frozen and used for the training of the LSTM neural network. The LSTM neural network was trained using gradient descent with momentum with a learning rate of 0.001 and a weight decay of 0.005. There was a 0.9 fixed momentum.

For the evaluation of the proposed approach, three error metrics were proposed. The mean absolute error (MAE) [26], root-mean-square error (RMSE) [26] and mean absolute percentage error (MAPE) [27] were proposed as evaluation metrics. The MAE can be computed as Eq. (6).

$$MAE(y, x) = \frac{1}{N} \sum_{n=1}^N |y_n - x_n| \quad (6)$$

Where N is the number of samples of the time series. y is the predicted time series and x is the measured time series. The RMSE can be computed as Eq. (7).

$$RMSE(y, x) = \sqrt{\frac{1}{N} \sum_{n=1}^N (y_n - x_n)^2} \quad (7)$$

The MAPE can be computed as Eq. (8).

$$MAPE(y, x) = \frac{1}{N} \sum_{n=1}^N \left| \frac{y_n - x_n}{y_n} \right| \quad (8)$$

The choice of the evaluation metrics was not arbitrary because the comparison between those errors can provide more information on the forecasting system. For example, if the RMSE is higher than the MAE with a big margin then the forecasting system has a big deviation to the measured power. If the RMSE is approximately equal to the MAE then the forecasting system has a small deviation to the measured power. The MAPE is used to measure the robustness of the proposed approach for energy forecasting. A low MAPE means that high performance was achieved.

To avoid an overfitting problem, an early stopping condition was applied. The training stops if the RMSE does not change for 50 consecutive training iterations. The proposed approach was used to predict a day ahead forecast horizon of 24h to 48h. The proposed approach has an alternative training process by pretraining the complex-valued autoencoder then using its pretrained weights in the training of the LSTM neural network. While training the LSTM neural network, the complex-valued autoencoder was used for data reconstruction and its weights are not updated. The LSTM neural network was initialized using samples of two previous time steps.

Calculating the proposed error metrics was done in order to evaluate the proposed approach. Table I presents the obtained

error values. The results demonstrated the effectiveness of the suggested strategy. The obtained RMSE is 0.167 which is higher than the MAE with a margin of 0.074; it means that the forecasting approach has a good deviation from the measured power. The achieved MAPE of 0.028 proves that the proposed approach can generate trusted predictions with a low error.

TABLE I. OBTAINED ERRORS FOR THE PROPOSED METHOD

Error metric	Complex valued autoencoder + LSTM neural network
MAE	0.093
RMSE	0.167
MAPE	0.028

The proposed approach has achieved good results based on the obtained error values. A comparison against state-of-the-art models is presented in Table II. Among all methods, the proposed method has a better prediction effect with a better deviation to the measured power. Compared to the state-of-the-art methods, the proposed method has the lowest MAPE which means that it has the best prediction accuracy and can be considered a reliable forecasting system. More accurate prediction means that the integration cost of solar energy in the grid is reduced and fewer problems can be encountered.

TABLE II. COMPARISON AGAINST STATE-OF-THE-ART METHODS FOR SOLAR ENERGY FORECASTING

Method	MAE	RMSE	MAPE
Hybrid neural network [16]	0.17	0.32	0.067
LSTM-CNN [21]	0.124	0.184	-
TCNN [22]	0.221	0.621	0.042
RNN [28]	0.485	0.84	0.03
Proposed	0.093	0.167	0.028

Reliable forecasting values allow the control of solar power injection in the grid safely and that reduces pollution and using fuel power stations. Fig. 5 presents a comparison between state-of-the-art methods and the proposed method in 18 days of January 2023. As shown in Fig. 5, the proposed method (com-Auto-LSTM) has the lowest error compared to the target values. From day 6 to day 13 the energy production is zero and this may be caused by snowfall. Since the proposed method considers weather conditions the predicted power was very close to the target.

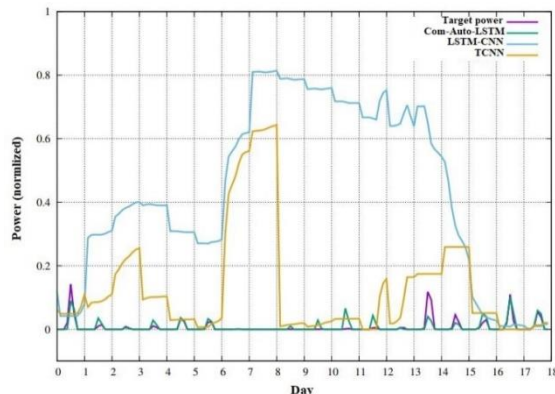


Fig. 5. Comparison against state-of-the-art methods and the proposed method in 18 days of January 2023.

Based on the obtained results, the proposed approach has proved its efficiency. The proposed approach takes advantage of features extraction and combination of the complex-valued autoencoder and the temporal features extraction of the LSTM neural network. The forecasting power of the LSTM neural networks has been improved by combining the historical time series of solar energy production with weather conditions (temperature, humidity, and cloud variation). Through the ability of the proposed complex-valued autoencoder to process complex data with multiple inputs, the input signal of the LSTM neural network was reconstructed by combining the different features. Overall, the proposed method has shown its efficiency for solar power forecasting through the ability to combine historical time series with weather conditions.

V. CONCLUSIONS

Solar power is considered clean energy that can be used to reduce pollution and power costs. However, integrating a solar power generator into the grid is very challenging due to many problems such as high voltage injection and power stability. Energy forecasting is considered as a solution for the safe integration of solar power energy into the grid. In this study, we suggested a neural network-based method for forecasting solar energy that combines two neural network models. First, a complex-valued autoencoder was used to combine the historical time series of power generation with weather conditions. Second, the reconstructed signal was used to train an LSTM neural network. The proposed LSTM neural network was composed of 3 levels of LSTM units with the same size as the input data. The proposed approach was tested and trained using the Fingrid open data. The obtained findings have demonstrated the effectiveness of the suggested strategy, which has a low MAPE when compared to cutting-edge techniques. Besides, the proposed method has a good deviation from the measured power. A comparison against existing models on 18 days of January 2023 has shown that the proposed method has the lowest prediction error. The evaluation has demonstrated the suggested method's high performance, which is attributable to the usage of a combination of an LSTM neural network and a complex-valued autoencoder.

ACKNOWLEDGMENT

The authors extend their appreciation to the deanship of scientific research at Shaqra University for funding this research work through the project number (SU-ANN-202258).

REFERENCES

- [1] Y. Chaoqing, S. Liu, and Z. Fang, "Comparison of China's primary energy consumption forecasting by using ARIMA (the autoregressive integrated moving average) model and GM (1, 1) model," *Energy* 100, 2016, pp. 384-390.
- [2] F. R. Johnston, J. E. Boyland, M. Meadows, and E. Shale, "Some properties of a simple moving average when applied to forecasting a time series," *Journal of the Operational Research Society* 50, no. 12, 1999, pp. 1267-1271.
- [3] J. M. Lucas, and M. S. Saccucci, "Exponentially weighted moving average control schemes: properties and enhancements," *Technometrics* 32, no. 1, 1990, pp. 1-12.
- [4] O. Eva, and O. Ostertag, "Forecasting using simple exponential smoothing method," *Acta Electrotechnica et Informatica* 12, no. 3, 2012, p. 62.

- [5] G. Ian, Y. Bengio, A. Courville, and Y. Bengio, "Deep learning," Vol. 1, no. 2. Cambridge: MIT press, 2016.
- [6] A. Saad, T. A. Mohammed, and S. Al-Zawi, "Understanding of a convolutional neural network." In 2017 International Conference on Engineering and Technology (ICET), pp. 1-6. IEEE, 2017.
- [7] S. Martin, R. Schlüter, and H. Ney, "LSTM neural networks for language modeling," In Thirteenth annual conference of the international speech communication association. 2012.
- [8] W. Yasi, H. Yao, and S. Zhao, "Auto-encoder based dimensionality reduction," *Neurocomputing*, 184, 2016, pp. 232-242.
- [9] A. Riadh, Y. Said, and M. Atri, "A Convolutional Neural Network to Perform Object Detection and Identification in Visual Large-Scale Data," *Big Data*, 9(1), 2021, pp. 41-52.
- [10] A. Mouna, R. Ayachi, E. Pissaloux, Y. Said, and M. Atri. "Indoor objects detection and recognition for an ICT mobility assistance of visually impaired people," *Multimedia Tools and Applications*, 79, no. 41, 2020, pp. 31645-31662.
- [11] Y. Said, "Pynq-YOLO-Net: An Embedded Quantized Convolutional Neural Network for Face Mask Detection in COVID-19 Pandemic Era," *International Journal of Advanced Computer Science and Applications(IJACSA)*, 11(9), 2020, pp.100-106.
- [12] A. Riadh, Y. Said, and A. B. Abdelaali, "Pedestrian Detection Based on Light-Weighted Separable Convolution for Advanced Driver Assistance Systems," *Neural Processing Letters*, 52, no. 3, 2020, pp. 2655-2668.
- [13] S. Alaa, and M. Kotb, "Unsupervised pre-training of a Deep LSTM-based Stacked Autoencoder for Multivariate time Series forecasting problems," *Scientific Reports*, 9, no. 1, 2019, pp. 1-16.
- [14] S. Sima, N. Tavakoli, and A. S. Namin, "The performance of LSTM and BiLSTM in forecasting time series," In 2019 IEEE International Conference on Big Data (Big Data), pp. 3285-3292. IEEE, 2019.
- [15] N. V. Anantha, and P. Karatampati, "Survey on renewable energy forecasting using different techniques," In 2019 2nd International Conference on Power and Embedded Drive Control (ICPEDC), pp. 349-354. IEEE, 2019.
- [16] A. Oveis, N. Amjady, and N. Ghadimi, "Solar energy forecasting based on hybrid neural network and improved metaheuristic algorithm," *Computational Intelligence*, 34, no. 1, 2018, pp. 241-260.
- [17] S. A. Abolfazl, M. B. Tavakoli, and A. Hoseinabadi, "Modified Levenberg-Marquardt method for neural networks training," *World Acad Sci Eng Technol*, 6, no. 1, 2005, pp. 46-48.
- [18] D. Simon, J. Lee, H. Li, L. Wang, and X. Zhai, "Gradient descent finds global minima of deep neural networks," In International Conference on Machine Learning, pp. 1675-1685. PMLR, 2019.
- [19] S. Rudy, and L. C. K. Hui, "Use of a quasi-Newton method in a feedforward neural network construction algorithm," *IEEE Transactions on Neural Networks*, 6, no. 1, 1995, pp. 273-277.
- [20] O. Juan, A. Newaz, and M. O. Faruque, "Forecasting of PV plant output using hybrid wavelet-based LSTM-DNN structure model," *IET Renewable Power Generation*, 13, no. 7, 2019, pp. 1087-1095.
- [21] K. Wang, X. Qi, and H. Liu, "Photovoltaic power forecasting based LSTM-Convolutional Network," *Energy*, 189, 2019, p.116225.
- [22] L. Yang, I. Koprinska, and M. Rana, "Temporal Convolutional Neural Networks for Solar Power Forecasting," In 2020 International Joint Conference on Neural Networks (IJCNN), pp. 1-8. IEEE, 2020.
- [23] Z. Xiaohu, Y. Zou, and W. Shi, "Dilated convolution neural network with LeakyReLU for environmental sound classification," In 2017 22nd International Conference on Digital Signal Processing (DSP), pp. 1-5. IEEE, 2017.
- [24] H. Kaiming, X. Zhang, S. Ren, and Ji. Sun, "Deep residual learning for image recognition," In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770-778. 2016.
- [25] Fingrid Solar power generation Search and download data. Available at: <https://data.fingrid.fi/en/dataset/solar-power-generation-forecast-updated-every-hour/resource/8b6b8bff-0181-48e1-aa86-1af97f81ce5a> Last accessed: 20/02/2023.
- [26] C. Tianfeng, and R. R. Draxler, "Root mean square error (RMSE) or mean absolute error (MAE)?—Arguments against avoiding RMSE in the literature," *Geoscientific model development*, 7, no. 3, 2014, pp. 1247-1250.
- [27] D. Myttenaere, A. B. Golden, B. Le Grand, and F. Rossi, "Mean absolute percentage error for regression models," *Neurocomputing*, 192, 2016, pp. 38-48.
- [28] L., Gangqiang, H. Wang, S. Zhang, J. Xin, and H. Liu, "Recurrent neural networks based photovoltaic power forecasting approach," *Energies*, 12, no. 13, 2019, pp. 2538.
- [29] A. Mouna, R. Ayachi, Y. Said, and M. Atri, "An evaluation of EfficientDet for object detection used for indoor robots assistance navigation." *Journal of Real-Time Image Processing*, 19, no. 3, 2022, pp. 651-661.
- [30] A. Riadh, A. Mouna, Y. Said, and A. B. Abdelali, "Real-time implementation of traffic signs detection and identification application on graphics processing units," *International Journal of Pattern Recognition and Artificial Intelligence*, 35, no. 07, 2021: 2150024.
- [31] A. B. Atitallah, Y. Said, M. A. B. Atitallah, M. Albekairi, K. Kaaniche, T. M. Alanazi, S. Boubaker, and M. Atri, "Embedded implementation of an obstacle detection system for blind and visually impaired persons' assistance navigation," *Computers and Electrical Engineering*, no. 108, 2023:108714.

Classification with K-Nearest Neighbors Algorithm: Comparative Analysis between the Manual and Automatic Methods for K-Selection

Tsvetelina Mladenova¹, Irena Valova²

Department of Computer Systems and Technologies, University of Ruse, Ruse, 7017, Bulgaria

Abstract—Machine learning and the algorithms it uses have been the subject of many and varied studies with the development of artificial intelligence in recent years. One of the popular and widely used classification algorithms is the nearest neighbors' algorithm and in particular k nearest neighbors. This algorithm has three important steps: calculation of distances; selection of the number of neighbors; and the classification itself. The choice of the value for the k parameter determines the number of neighbors and is important and has a significant impact on the degree of efficiency of the created model. This article describes a study of the influence of the way the k parameter is chosen - manually or automatically. Data sets, used for the study, are selected to be as close as possible in their features to the data generated and used by small businesses - heterogeneous, unbalanced, with relatively small volumes and small training sets. From the obtained results, it can be concluded that the automatic determination of the value of k can give results close to the optimal ones. Deviations are observed in the accuracy rate and the behavior of well-known KNN modifications with increasing neighborhood size for some of the training data sets tested, but one cannot expect that the same model's parameter values (e.g. for k) will be optimally applicable on all data sets.

Keywords—Machine Learning; KNN; classification

I. INTRODUCTION

K-Nearest Neighbors (KNN) is a simple and easy-to-use supervised machine learning (ML) algorithm that can be used for solving classification problems in different domains - education, healthcare, livestock and crop production, administration, production, transport, etc. [1, 2, 3]. KNN is an extension of the idea of the nearest neighbor method introduced by E. Fix and J. Hodges in 1951 [4, 5, 6], which is based on the calculation of the distance between an unlabeled sample and the nearest sample - neighbor, from the training set.

This extension was proposed by Cover and Hart [7] who added the parameter k representing the number of nearest neighbors to be considered for classification. For k=1, it can be said that the nearest neighbor algorithm, NN, is considered, and for k >1, KNN. The KNN algorithm is a relatively simple example of a non-parameterized classifier, easy to understand and implement [8, 9, 10]. The main stages of the KNN algorithm are three, (Fig. 1).

- 1) Calculate the distances from the unlabeled sample to each sample in the training set.
- 2) Processing the calculated distances and selecting the k neighbors [11] that will form the neighborhood.
- 3) Determining the class to which the unlabeled sample belongs - the classification stage.

The main challenges specific to the nearest-neighbor method can be divided into the following groups:

- Determination of the k parameter - the size of the neighborhood depends on it, which is decisive for the classification. It has been proven that the parameter is sensitive and at the same time important for the degree of efficiency of the model.
- Choosing the neighbors - the calculation of distances between samples is traditionally carried out using the Euclidean distance. Choosing the neighbors that will shape the neighborhood is no less important a step than choosing its size.
- Classification rule – the classification stage is the last stage in which the decision is made as to which class to classify the unlabeled sample. An inappropriate classification rule could mean an incorrect classification of the unlabeled sample.

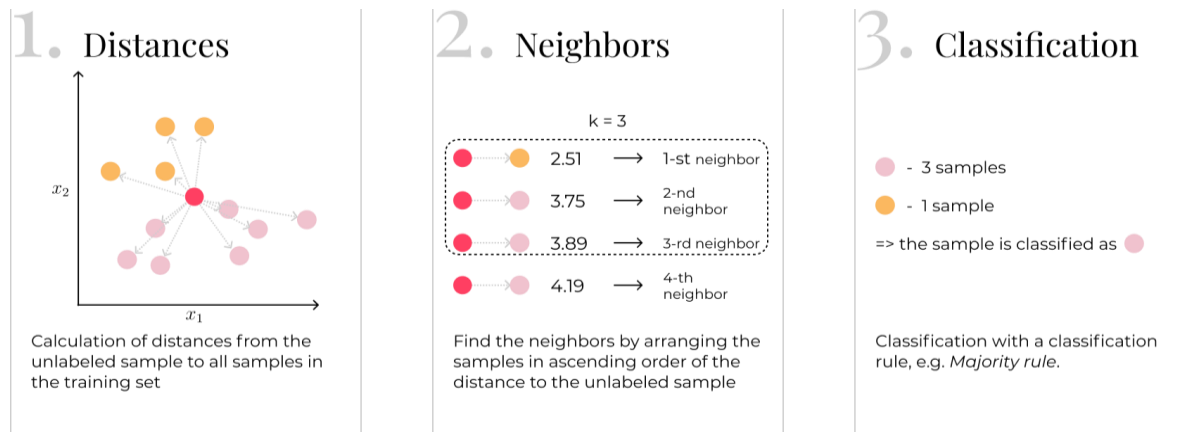


Fig. 1. Steps in classification with the method of nearest neighbors.

II. METHODS FOR CALCULATING THE K VALUE

Determining the value on which the neighborhood size will depend is a problem addressed by many scientific studies. The parameter is particularly sensitive, since at a value larger than necessary, the model's efficiency decreases. A smaller value, respectively, will make the model inaccurate. Additionally, a smaller value will mean that noise in the data will have a greater impact on the classification, and a larger value will mean that additional computing resources are required.

There are different approaches to choosing k . One approach uses a pre-definition of a single value, regardless of the particularities of the training set (type, size, subject area, etc.). The other approach is to determine the value of each set. It is used more, and with it, the value is most often found by the elbow method - the training set is trained with a series of k values and the one is found where the effectiveness of the model begins to decrease.

A third method is the *m-fold cross-validation* [12]. In this approach, the training set is divided into m disjoint sets. The cross-validation method is then applied to each set. Finally, the k value of the subset that gave the best results is taken as the k value of the entire training set.

A disadvantage of this approach is that the selection of the k value does not consider the distribution of the data [13]. Such a finding of the optimal value of k is called "manual search" and requires in-depth analysis by a person who understands and knows machine learning. In addition, such methods often require training the model several times, which means additional time for preparation and calculations.

One of the more automated methods for determining k is the dynamic one, where various approaches are applied to analyze the training set, feature distribution, class information, anomalies, etc., to find the best value.

It is important to note that no matter how the value for k is calculated, it does not guarantee that the model will have high efficiency or that this value will not change. Adding just one object or sample to the training set, or altering any step of the algorithm, may result in an inefficient neighborhood and an inapplicable value of k .

In [14] and [15] an approach is used where the value of k is determined automatically and can vary depending on the size of the set. According to the authors, this approach is suitable for reducing the set size, lowering the classification time, and is a suitable replacement for the traditional KNN for working with big data. The main idea on which the algorithm is built is to find local neighborhoods consisting of samples belonging to the same class. Thus, one can find the number of samples in each local neighborhood, as well as the similarities between the most distant sample in the neighborhood and its center.

By taking these samples as well as the centers of each neighborhood, the size of the training set is reduced while preserving its distribution. Each new, unlabeled sample is compared to the new, reduced set. In this way, there is no need to explicitly define a value for k . It should be noted that the authors propose the text classification algorithm where the classes from which the training set is composed are sufficient for the classification to be accurate.

In [16], [17], the number of neighbors on which the classification will depend is determined by the number of possible classes (1). Through experimental studies, it is proven that this determination of neighbors is appropriate when the boundaries between classes in the data are not clear enough and are face-to-face, so-called overlapping data.

$$k = c + 1 \quad (1)$$

Where:

c – the number of unique classes.

An approach with a dynamic selection of a value for k is also considered in [18], where the number of neighbors depends on the distribution of samples, relative to the classes in the training set. A Chinese text classification method is proposed, and it is proved through an experimental study that the modification achieves good results in classifying documents having classes with few samples.

Another approach used is equation (2).

$$k = \sqrt{N} \quad (2)$$

Where:

N – the number of samples in the training set.

A common practice of ML researchers is to apply the following rule – when the number of unique classes in the training set is even, the value of k is odd to avoid equality between neighboring classes. This is not a guaranteed approach, as samples from only one class or an equal number of samples from two or more classes can fall into the neighborhood, but it is an additional step to improve the model's performance.

The automated way of obtaining the neighborhood size has proven effective when the choice of a value for k needs to be made quickly and without additional steps before training the model. The results obtained with an automatically calculated value for k are comparable to the results obtained using some of the manual methods.

The main advantage of this approach is that it provides an additional step of automation and makes the nearest-neighbor method more accessible to users who do not understand ML. This is a solution that can be implemented in an algorithm aimed at automating learning processes and providing understandable results.

III. NEIGHBORHOOD SIZE AND NEIGHBORS' SELECTION

The neighborhood size, in the nearest neighbor method, is determined by the chosen value for the k parameter. If the value is too small, noise in the data may have too much influence on the classification. Too large a value means more computation time and resources. The exact value of k is difficult to determine and depends on the characteristics of the training set. Taking into account the fact that the sought-after solution has to handle most small datasets, the possibility that the value of k can be further restricted should also be taken into account.

In Fig. 2, Fig. 3, and Fig. 4 can be seen how increasing the number of neighbors leads to a change in the accuracy of the model. The figures show the decision boundaries for different values of k .

At $k=5$, Fig. 2, the overfitting of the model is noticeable, i.e. the model tries to classify as many "single" cases as possible, making it unstable and unreliable.

The increase in the number of neighbors, for example, $k=15$, Fig. 3, results in a normalization of the classification, although areas where it can be said that there is overfitting are still observed.

Fig. 2, Fig. 3, and Fig. 4 visualize a case where the data in the training set is unbalanced and the boundaries between classes are not clearly defined. For the set in the figures, the traditional nearest neighbor method achieves a classification accuracy of 73% at $k=5$, 73% at $k=15$, and 76% at $k=20$. The increase in accuracy rate is a result of expanding the neighborhood. When the training set is large enough to allow this, this is not a problem, despite the additional computational resources required for the larger neighborhood. However, when the set is on the order of 25 samples, $k=20$ will mean that in practice the entire set will be used for training.

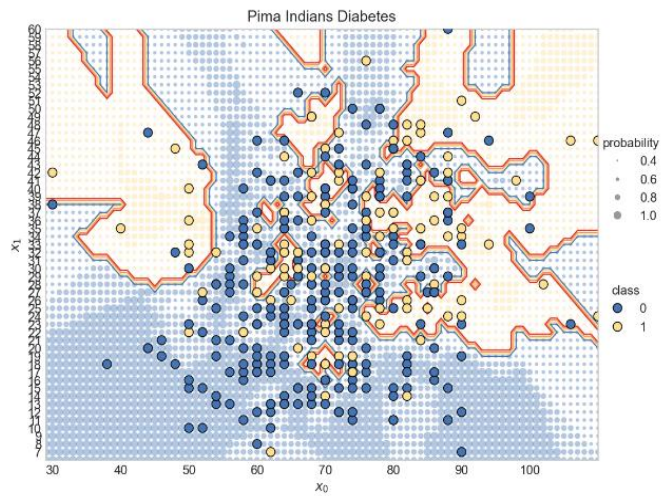


Fig. 2. Decision boundaries for $k=5$.

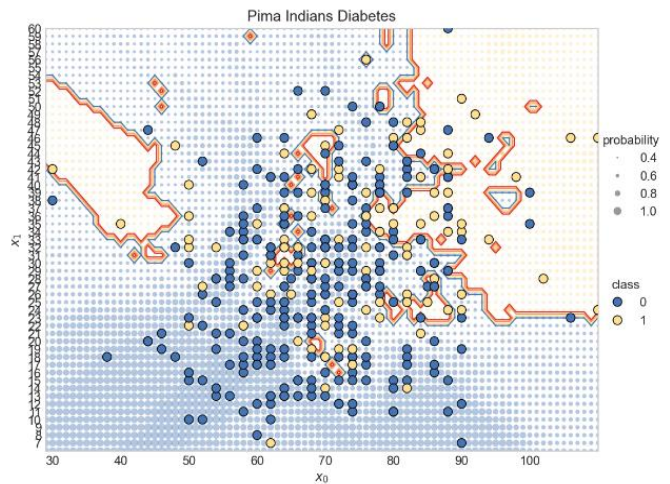


Fig. 3. Decision boundaries for $k=15$.

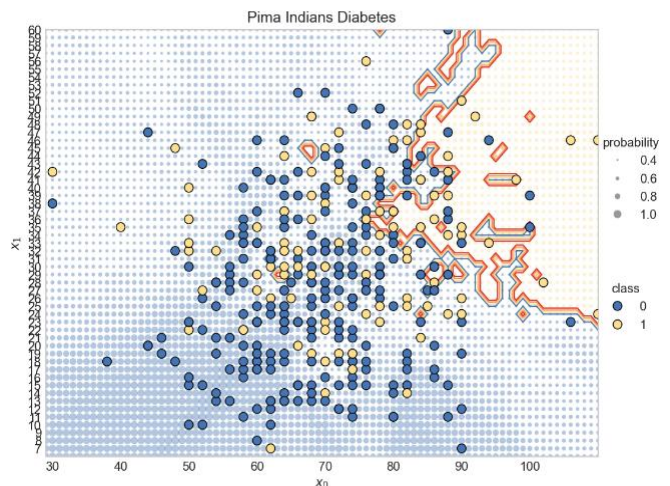


Fig. 4. Decision boundaries for $k=20$.

IV. METHODS FOR CALCULATING THE DISTANCES BETWEEN SAMPLES AND FORMING THE NEIGHBORHOOD

The calculation of the nearest neighbors consists of using a function to calculate the distance between the unlabeled sample and all other samples from the training set. There is no single metric for distance measurement that is applicable in all cases [19], although attempts for finding such a metric are not lacking [20, 21, 22].

According to [20], the neighborhood must meet two criteria: 1) the neighbors are close to the unlabeled sample and 2) the neighbors are symmetrically located around it. The idea of nearest neighbors takes into account only the first criterion [23]. Therefore, the neighborhood may not be symmetrically located if the data in the set of neighbors is inhomogeneous.

A. Nearest Centroid Neighborhood (NCN)

In [23] a new definition of the term "neighborhood" is proposed, which does not require user-defined parameters. The proposed algorithm is called Nearest Centroids or Nearest Centroid Neighborhood (NCN). The basic idea can be described as follows: let be a point whose neighbors are to be found and added to the set of points. The neighbors are such that they fulfill the following conditions: 1) they are as close as possible to and 2) the center between the neighbor and the point is as close as possible [24].

B. Distance Weighted K-Nearest Neighbor (WKNN)

In 1975, S. Dudani proposed a modification of the nearest neighbor algorithm [25]. The author argues that it is logical that the distances between individual neighbors should be proved in the form of a "weight" that varies depending on the distance from the unlabeled sample to its neighbors.

In [26], a modification of S. Dudani's algorithm (DWKNN) is proposed for calculating neighbor weights. The proposed algorithm reduces the weights of the nearest neighbors, except for the first and k neighbors. The purpose of weight reduction is to limit the influence of anomalies and improve classification accuracy. If the size of the training set is too large compared to the number of neighbors that are taken into account, then the presented algorithm and the majority rule achieve close results. The author claims that his proposed algorithm achieves better results on small and medium-sized training sets.

Choosing an optimal value for the parameter is difficult when using the majority rule due to the nature of data variation and the probability of classification error. The variation may be because the classification is largely due to the number of neighbors and the number of classes. Additionally, as, after a certain value depending on the size of the training set, increases, the probability of error may increase under certain circumstances.

Similar difficulties in choosing an optimal value for k do not exist when using WKNN. Therefore, it can be argued that with the use of the proposed algorithm, the selection of an optimal value for k can be made without worrying about increasing the probability of error. Several experiments proving the truth of the statement have been conducted. Doudani offers two more weighting functions – inverse weight and rank weight.

C. Pseudo Nearest Neighbor Rule (PNNR)

In [27], the proposed algorithm is based on two others – the weight calculation algorithm, WKNN, and the Local Mean Learning (LM) algorithm [28]. Conventionally, the algorithm is called Uniform. A variation of the algorithm called UWKNN is often used, which differs in that one is added to the neighbor's rank.

D. Dual K-Nearest Neighbor

In [29], a weighting function is proposed, aiming to reduce the sensitivity of the parameter by using a combination of the distance of the samples and their rank in the neighborhood.

E. Fuzzy K-Nearest Neighbor (FKNN)

The Fuzzy KNN algorithm [30] is based on the fuzzy set theory first proposed by Zadeh in 1965. [31]. The main idea of the algorithm is the calculation of the degree of belonging of each element of the training set to the classes. The degree of belonging is taken into account when classifying the unlabeled sample.

V. IMPLEMENTATION OF THE EXPERIMENTS

Often, programming languages such as Python and R are used to train machine learning models. Over the years, they have proven to be some of the most suitable solutions for artificial intelligence and ML. Some of the most developed libraries are written and adapted specifically for these programming languages.

The scikit-learn library contains many predefined algorithms for ML. In addition to the basic algorithms, the library also allows the application of some modification or set of additional parameters that give some flexibility. In some cases, however, despite the availability of parameters to control a given model, there is a need for the so-called "custom changes". In such cases, the library cannot always be used.

The current study does not use a predefined library. All source code for the classification and evaluation algorithm is written in the Python programming language.

A. Datasets for the Experimental Study

The data sets used for the experimental study are selected to match the data that a small business would have – small training sets, heterogeneous data, data with anomalies, and unbalanced data. A more unusual category of data has also been added to the sets used – those that have the $n \gg p$ problem, i.e. the number of attributes is many times greater than the number of samples. Although rare, such data do occur. A typical example is medical data, where the characteristics describing a sample are much more than the samples under study.

The structure of the used datasets is presented and described in Table I. The sets are arranged in increasing order of the number of samples. Data on the balance between the number of representatives of each class are taken from the sources of the sets. The percentage of abnormalities in each set was calculated using the Z-score method (3).

$$z = \frac{x - \mu}{\sigma} \quad (3)$$

Where:

x – sample of the set

μ – average value

σ – the standard deviation

The Z-score formula finds the number of standard deviations from the mean. It is considered an anomaly if the value of the sample obtained by (3) is above three or below three. For the purposes of the experimental study, an "anomaly set" will mean a set containing more than 10% anomalies.

TABLE I. TRAINING SETS USED IN THE EXPERIMENTAL STUDY

Dataset	Attributes	Samples	Classes	No balanced	Anomaly
Lenses ¹	4	24	3	Yes	100%
Lung Cancer ²	56	32	3	No	78.12% (25)
Soybean (small) ³	35	47	4	Yes	0%
SRBCT ⁴	2308	83	4	No	0%
Cryotherapy ⁵	6	90	2	No	3.33% (3)
Beavers ⁶	3	114	2	Yes	0.87% (1)
Iris ⁷	4	150	3	No	0.66% (1)
Hepatitis ⁸	19	155	2	Yes	92.9% (144)
Wine ⁹	13	178	3	No	5.62% (10)
Glass ¹⁰	9	214	6	Yes	9.35% (20)
Thyroid ¹¹	5	215	3	Yes	8.83% (19)
Stars ¹²	4	240	6	No	2.92% (7)
Algerian Forest Fires ¹³	11	243	2	No	7.41% (18)
Ecoli ¹⁴	8	335	8	Yes	0%
Ionosphere ¹⁵	34	351	2	Yes	0%
Breast Cancer Wisconsin ¹⁶	30	569	2	Yes	13% (74)
Absenteeism ¹⁷	19	740	28	Yes	77.84% (576)
Pima Indians Diabetes ¹⁸	8	768	2	No	10.41% (80)

¹ <https://archive.ics.uci.edu/ml/datasets/lenses>

² <https://archive.ics.uci.edu/ml/datasets/lung+cancer>

³ [https://archive.ics.uci.edu/ml/datasets/soybean+\(small\)](https://archive.ics.uci.edu/ml/datasets/soybean+(small))

⁴ <https://rdrr.io/cran/plsgenomics/man/SRBCT.html>

⁵ <https://archive.ics.uci.edu/ml/datasets/Cryotherapy+Dataset+>

⁶ <https://stat.ethz.ch/R-manual/R-devel/library/datasets/html/beavers.html>

⁷ <https://archive.ics.uci.edu/ml/datasets/iris>

⁸ <https://archive.ics.uci.edu/ml/datasets/hepatitis>

⁹ <https://archive.ics.uci.edu/ml/datasets/wine>

¹⁰ <https://archive.ics.uci.edu/ml/datasets/glass+identification>

¹¹ <https://archive.ics.uci.edu/ml/datasets/thyroid+disease>

¹² <https://www.kaggle.com/brsdincer/star-type-classification>

¹³

<https://archive.ics.uci.edu/ml/datasets/Algerian+Forest+Fires+Dataset++>

¹⁴ <https://archive.ics.uci.edu/ml/datasets/ecoli>

¹⁵ <https://archive.ics.uci.edu/ml/datasets/ionosphere>

¹⁶

<https://archive.ics.uci.edu/ml/datasets/breast+cancer+wisconsin+%28original%29>

¹⁷ <https://archive.ics.uci.edu/ml/datasets/Absenteeism+at+work>

¹⁸ <https://www.kaggle.com/uciml/pima-indians-diabetes-database>

The Z-score formula finds the number of standard deviations from the mean. It is considered an anomaly if the value of the sample obtained by (3) is above three or below three. For the purposes of the experimental study, an "anomaly set" will mean a set containing more than 10% anomalies.

B. Data Description

The data used in this experimental study varied in volume, the number of classes, the percentage of anomalies, and the imbalance. Some training sets are of greater interest for research because they contain a larger number of anomalies or are unbalanced. They are Lenses, Lung Cancer, Soybean, SRBCT, Beavers, Hepatitis, Glass, Thyroid, Ecoli, Ionosphere, Breast Cancer Wisconsin, and Absenteeism.

The *Lenses* training set contains 24 samples with 4 attributes each, divided into 3 classes. It is unbalanced and all samples are considered an anomaly. The data is used to classify patients who need contact lens fitting. The potential business application is in optics, where patients are examined and the need for contact lens fitting is determined.

The Lung Cancer set contains information about people who have characteristics of lung cancer patients. The classification task is to correctly diagnose new patients, based on the data of 32 patients (samples), 56 indicators (attributes), and 3 possible diagnoses (classes). Anomalies account for 78% of the data. A potential business application of such a dataset is in laboratories examining patient samples.

Soybean contains information on soybean diseases. It contains 47 samples with 35 attributes classified into 4 classes. It is defined as unbalanced, without anomalies. A set containing information on cereals, legumes, and similar crops can find application in any farm growing them.

SRBCT contains data from 83 patients and 2308 attributes (genes) classified into 4 classes. The data is intended to aid in the correct classification of various childhood cancers. The set is balanced, has no anomalies, and is considered an $n \gg p$ problem. It can be used in laboratories, research, and scientific centers.

Cryotherapy is a set containing data for 90 samples with 6 attributes. Unbalanced and free of anomalies.

Iris is a well-known and researched set for the classification of flowers of the Iris species. It is used as the basis for the study of many machine learning algorithms. It contains 150 samples having 4 attributes and samples are divided into 3 categories. It is balanced and has no anomalies. The application of such a set is in the field of the flower business - a flower shop can recognize the types of the received goods, and greenhouses can more accurately classify the flowers grown.

Hepatitis contains information on 155 samples, each described by 19 attributes and classified into 2 categories. The set is not balanced and has a high percentage of anomalies – 92.9%. Contains information for patients diagnosed with hepatitis. The classification problem it solves is to answer the question of whether the patient will live. It can be used in doctor's offices.

Wine is a dataset much like Iris - well-known, used, and studied. Contains data for 178 samples, 13 attributes, and 3 categories. The data is balanced and the anomaly rate is not high. A similar set can be used by winemakers.

Glass contains data for 214 samples, 9 attributes, and 6 classes. The set is unbalanced, but there are no anomalies. The classification problem is the recognition of types of glass, specifically glass shards found at crime scenes. It can be used in laboratories and glass manufacturers.

Thyroid described 215 patients with 5 attributes and divided them into 3 classes. The data is unbalanced and there are no anomalies. They are used to classify the action of the thyroid gland and therefore such a set can be used in doctors' offices and laboratories.

Stars dataset aims to categorize different celestial bodies into 6 categories, according to 4 attributes. There are 240 samples examined, and the set is considered balanced and contains no anomalies. The application could be in business organizations involved in space exploration.

Algerian Forest Fires contains information on 243 fires in Algeria. Each fire is classified into 2 categories and described with 11 characteristics. The set is balanced and has no anomalies. It can be used in applications for early fire warning, by conservation organizations, fire departments, for research purposes, etc.

Ecoli is a set that contains information about 335 studied samples, described in 8 categories and characterized by 8 attributes. There are no anomalies detected, but the data is unbalanced. It is used to detect specific proteins. The business application of such a set is in laboratories, doctor's offices, and scientific and research centers.

Ionosphere contains information on 351 signals passing through the Earth's ionosphere. The attributes describing them are 34, and the classes classifying them are 2. There are no detected anomalies, but the data is unbalanced. The data can be used in any organization involved in space and earth research.

Breast Cancer Wisconsin contains information on 569 patients described by 30 attributes and classified into 2 categories. The set is unbalanced and the anomaly rate is 13%. The data is used to detect malignant tumors in patients. The application can be in doctors' offices and laboratories.

Absenteeism contains information about 740 employees in different companies. Each employee is described with 19 indicators and 28 different degrees of possibility for the

employee to be absent from work are defined. The high number of possible classes into which a sample can be classified causes the set to have a high rate of anomalies and to be unbalanced. Training a model with such data makes it possible to use it anywhere regardless of the subject area of the business.

Pima Indians Diabetes is a set that has 768 samples classified into two categories and described by 8 attributes. Contains data for patients who have indicators similar to those of patients with diabetes. The data is balanced, but anomalies are present. The application of such a set could be in doctor's offices.

All experimental studies were done with m-fold cross-validation, where m=10. According to [32, 33, 34, 35], to avoid the chance of random results, each model should be trained more than once and the average of the training to be accepted, for this purpose many authors suggest the number of training iterations to be ten, this is the accepted number in the current paper. The results of each iteration were recorded and the average value of the model's accuracy was obtained. For each model training, all available attributes were used without further attribute processing.

For the k-parameter studies, all training sets were trained with the traditional neighborhood method in combination with the traditional nearest neighborhood method and the weighted neighborhood method. The automatic calculation of the k parameter is done in two ways: the training set size method, described with an equation (2), and the number of classes method - equation (1) (Table II).

In the training set method the number of the classes is taken into consideration. For every m_i training iteration around 70% of the data is used and around 30% is used for testing of the model, meaning that the number of the samples is not a constant.

Table II shows the results from the experimental study. The second column of the table shows the number of samples with which the k value is determined, and the following columns shows the highest achieved result for every trained model.

Every neighborhood method is experimented, and for every method, three types of k determination are done – a manual and two automatic methods. In every cell, the accuracy percentage is noted and in brackets is the number of the neighbors (k value) that has achieved this accuracy score.

TABLE II. HIGHEST PERCENT ACCURACY OBTAINED WITH MANUAL AND AUTOMATED K VALUE SELECTION

Dataset	N	KNN		WKNN		DWKNN		UWKNN		Dual		Uniform		Inverse	
		Manual method	$k = \sqrt{N}$	Manual method	$k = \sqrt{N}$	Manual method	$k = \sqrt{N}$	Manual method	$k = \sqrt{N}$	Manual method	$k = \sqrt{N}$	Manual method	$k = \sqrt{N}$	Manual method	$k = \sqrt{N}$
			$k = c + 1$		$k = c + 1$		$k = c + 1$		$k = c + 1$		$k = c + 1$		$k = c + 1$		$k = c + 1$

Dataset	N	KNN		WKNN		DWKNN		UWKNN		Dual		Uniform		Inverse	
Lenses	ir 21	0.85% (2)	0.750% (4)	0.8% (10)	0.750% (4)	0.833% (6)	0.783% (4)	0.817% (2)	0.700% (4)	0.817% (2)	0.733% (4)	0.8% (1)	0.717% (4)	0.783% (1)	0.717% (4)
		0.683% (4)	0.767% (4)		0.883% (4)		0.733% (4)		0.750% (4)		0.733% (4)		0.750% (4)		
Lung Cancer	ir 28	0.6% (8)	0.533% (5)	0.567% (9)	0.533% (5)	0.608% (10)	0.533% (5)	0.633% (5)	0.600% (5)	0.558% (6)	0.483% (5)	0.575% (3)	0.492% (5)	0.6% (7)	0.467% (5)
		0.442% (4)	0.450% (4)		0.433% (4)		0.517% (4)		0.517% (4)		0.467% (4)		0.483% (4)		
Soybean Small	ir 42	1.0% (3)	0.980% (6)	1.0% (4)	1.000% (6)	1.0% (1)	1.000% (6)	1.0% (3)	0.980% (6)	1.0% (6)	0.980% (6)	0.98% (1)	0.98% (6)	1.0% (4)	0.980% (6)
		0.980% (5)	1.000% (5)		1.000% (5)		0.980% (5)		0.98% (5)		0.98% (5)		0.980% (5)		
SRBCT	ir 74	0.943% (3)	0.893% (8)	0.965% (16)	0.942% (8)	0.975% (18)	0.975% (8)	0.976% (16)	0.939% (8)	0.953% (28)	0.914% (8)	0.976% (23)	0.918% (8)	0.951% (3)	0.940% (8)
		0.868% (5)	0.918% (5)		0.928% (5)		0.954% (5)		0.926% (5)		0.942% (5)		0.917% (5)		
Cryotherapy	ir 81	0.933% (2)	0.722% (9)	0.933% (4)	0.833% (9)	0.944% (9)	0.922% (9)	0.933% (2)	0.822% (9)	0.933% (4)	0.922% (9)	0.911% (3)	0.844% (9)	0.944% (7)	0.933% (9)
		0.844% (3)	0.933% (3)		0.933% (3)		0.856% (3)		0.933% (3)		0.933% (3)		0.933% (3)		0.911% (3)
Beavers	ir 102	0.949% (9)	0.947% (10)	0.949% (13)	0.948% (10)	0.949% (11)	0.947% (10)	0.949% (28)	0.948% (10)	0.949% (10)	0.939% (10)	0.948% (16)	0.948% (10)	0.949% (7)	0.930% (10)
		0.948% (3)	0.939% (3)		0.922% (3)		0.948% (3)		0.931% (3)		0.939% (3)		0.947% (3)		
Iris	ir 135	0.973% (12)	0.973% (11)	0.973% (18)	0.960% (11)	0.973% (16)	0.960% (11)	0.967% (7)	0.973% (11)	0.967% (22)	0.960% (11)	0.967% (8)	0.960% (11)	0.98% (13)	0.960% (11)
		0.960% (4)	0.960% (4)		0.960% (4)		0.960% (4)		0.953% (4)		0.960% (4)		0.960% (4)		
Hepatitis	ir 139	0.795% (26)	0.776% (11)	0.799% (20)	0.740% (11)	0.795% (30)	0.734% (11)	0.8% (24)	0.767% (11)	0.737% (27)	0.703% (11)	0.779% (29)	0.742% (11)	0.795% (23)	0.761% (11)
		0.715% (3)	0.669% (3)		0.671% (3)		0.703% (3)		0.665% (3)		0.698% (3)		0.715% (3)		
Wine	ir 160	0.795% (26)	0.686% (12)	0.786% (3)	0.707% (12)	0.78% (4)	0.720% (12)	0.775% (2)	0.719% (12)	0.781% (22)	0.752% (12)	0.781% (22)	0.775% (12)	0.782% (18)	0.771% (12)
		0.714% (4)	0.754% (4)		0.729% (4)		0.752% (4)		0.753% (4)		0.74% (4)		0.73% (4)		
Glass	ir 192	0.743% (1)	0.608% (13)	0.734% (1)	0.668% (13)	0.739% (4)	0.682% (13)	0.727% (2)	0.697% (13)	0.762% (8)	0.748% (13)	0.739% (3)	0.687% (13)	0.742% (1)	0.667% (13)
		0.649% (7)	0.687% (7)		0.719% (7)		0.691% (7)		0.725% (7)		0.692% (7)		0.660% (7)		

Dataset	N	KNN		WKNN		DWKNN		UWKNN		Dual		Uniform		Inverse	
Thyroid	193	0.949% (1)	0.888% (13)	0.949% (2)	0.930% (13)	0.953% (6)	0.940% (13)	0.954% (1)	0.921% (13)	0.958% (8)	0.939% (13)	0.958% (2)	0.930% (13)	0.954% (1)	0.926% (13)
			0.926% (4)				0.958% (4)				0.931% (4)				0.935% (4)
Stars	216	0.725% (1)	0.621% (14)	0.742% (2)	0.638% (14)	0.725% (2)	0.654% (14)	0.725% (1)	0.654% (14)	0.737% (6)	0.721% (14)	0.717% (9)	0.687% (14)	0.712% (3)	0.654% (14)
			0.604% (7)				0.683% (7)				0.667% (7)				0.654% (7)
Algerian Forest Fires	218	0.942% (4)	0.901% (14)	0.931% (6)	0.893% (14)	0.93% (1)	0.926% (14)	0.935% (5)	0.889% (14)	0.947% (1)	0.927% (14)	0.938% (2)	0.917% (14)	0.938% (1)	0.914% (14)
			0.930% (3)				0.931% (3)				0.934% (3)				0.909% (3)
Ecoli	301	0.425% (18)	0.424% (17)	0.425% (15)	0.424% (17)	0.424% (1)	0.424% (17)	0.425% (3)	0.424% (17)	0.425% (2)	0.424% (17)	0.425% (6)	0.425% (17)	0.425% (3)	0.424% (17)
			0.424% (9)				0.424% (9)				0.424% (9)				0.424% (9)
Ionosphere	315	0.863% (1)	0.838% (17)	0.872% (9)	0.852% (17)	0.872% (1)	0.846% (17)	0.872% (2)	0.826% (17)	0.875% (15)	0.869% (17)	0.866% (3)	0.849% (17)	0.863% (2)	0.832% (17)
			0.853% (3)				0.857% (3)				0.866% (3)				0.826% (3)
Breast Cancer Wisconsin	512	0.937% (10)	0.924% (22)	0.94% (22)	0.94% (22)	0.94% (22)	0.937% (22)	0.938% (7)	0.932% (22)	0.931% (27)	0.924% (22)	0.94% (25)	0.933% (22)	0.935% (17)	0.933% (22)
			0.930% (3)				0.917% (3)				0.921% (3)				0.930% (3)
Absenteeism	666	0.345% (5)	0.284% (25)	0.376% (7)	0.324% (25)	0.369% (11)	0.331% (25)	0.369% (26)	0.353% (25)	0.362% (28)	0.345% (25)	0.376% (23)	0.365% (25)	0.372% (29)	0.357% (25)
			0.276% (28)				0.315% (29)				0.334% (29)				0.350% (29)
Pima Indians Diabetes	691	0.755% (13)	0.751% (26)	0.75% (28)	0.742% (26)	0.749% (17)	0.741% (26)	0.75% (24)	0.733% (26)	0.726% (21)	0.712% (26)	0.738% (29)	0.720% (26)	0.749% (19)	0.725% (26)
			0.699% (3)				0.685% (3)				0.669% (3)				0.702% (3)

VI. CONCLUSIONS

From the obtained results, it can be concluded that the automatic determination of the k value can give results close to the optimal, but in some cases, the difference in the percentage of accuracy is about 10-15% lower, compared to manual methods.

For example, with Lenses, the obtained classification accuracy with the manual method is 85%, while with the automated selection method – 68%. Training with a manually found k value of the Lung Cancer set with the DWKNN

method can reach 60% classification accuracy. The same set and method (DWKNN), but with an automatically selected k value, reaches only 43% accuracy. Considering the nature of the training set, this percentage is unsatisfactory.

There are quite a few cases where the automatic selection of the neighborhood size gives close to optimal results. In some cases, the difference between the obtained values is less than a percentage, and in addition, the number of neighbors used is less than with the manual method. The use of a relatively small, but not too small, size of the neighborhood is one of the set criteria for accepting the algorithm as optimal (item 3.2).

In the Thyroid set trained with the Dual method, manual selection can classify with 95.8% accuracy at neighborhood size $k=8$. Using the automatic neighborhood of size $k=4$, i.e. half, achieves 94.4% accuracy. The difference between the two methods is negligibly small and it can be argued that in such cases the automatic selection of the k parameter is appropriate, efficient, and optimal in terms of the required resources for classification.

Known modifications of weighted nearest neighbors (WKNN, DWKNN, UWKNN, Dual, Uniform, and Inverse) do not behave stably and deviations in accuracy rate are observed with the increasing neighborhood for some training sets.

The performance of a model depends on many factors regarding the training set – size, number of unique classes, data balance, missing data anomalies, etc. Therefore, it cannot be expected that some model parameters, e.g. $k=5$, will be optimally applicable to all sets.

REFERENCES

- [1] Stoyanov, I. S., Iliev, T. B., Mihaylov, G. Y., Evstatiev, B. I., & Sokolov, S. A. (2018, October). Analysis of the cybersecurity threats in smart grid University of telecommunications and post, Sofia, Bulgaria. In 2018 IEEE 24th International Symposium for Design and Technology in Electronic Packaging(SIITME) (pp. 90-93). IEEE.
- [2] Mladenova, T., & Valova, I. (2022, September). Fake news detection from Bulgarian Facebook pages. In AIP Conference Proceedings (Vol. 2449, No. 1, p. 040013). AIP Publishing LLC.
- [3] Evstatiev, B. (2013). Evaluation of thermal diffusivity of soil near the surface: methods and results. Bulgarian journal of agricultural science, Agricultural academy, 19(3), 467-471.
- [4] Fix, E. and J. Hodges, "An important contribution to nonparametric discriminant analysis and density estimation," *Int. Stat. Rev.*, vol. 3, no. 57, pp. 233–238, 1951.
- [5] Fix, E. and J. L. Hodges, "Nonparametric discrimination: Consistency properties," Randolph Field, Texas, Proj., pp. 21–49, 1951.
- [6] Fix, E. and J. L. Hodges Jr, "Discriminatory analysis-nonparametric discrimination: Small sample performance," 1952.
- [7] Cover, T. and P. Hart, "Nearest neighbor pattern classification," *IEEE Trans. Inf. theory*, vol. 13, no. 1, pp. 21–27, 1967.
- [8] Qin, Z., A. T. Wang, C. Zhang, and S. Zhang, "Cost-sensitive classification with k-nearest neighbors," in *International Conference on Knowledge Science, Engineering and Management*, 2013, pp. 112–131.
- [9] Zhang, S., X. Li, M. Zong, X. Zhu, and R. Wang, "Efficient kNN classification with different numbers of nearest neighbors," *IEEE Trans. neural networks Learn. Syst.*, vol. 29, no. 5, pp. 1774–1785, 2017.
- [10] Mladenova, T., & Valova, I. (2021, June). Analysis of the KNN classifier distance metrics for Bulgarian fake news detection. In 2021 3rd International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA) (pp. 1-4). IEEE.
- [11] Zhang, S., D. Cheng, Z. Deng, M. Zong, and X. Deng, "A novel kNN algorithm with data-driven k parameter computation," *Pattern Recognit. Lett.*, vol. 109, pp. 44–54, 2018.
- [12] Meng, L., "Efficient M-fold Cross-validation Algorithm for KNearest Neighbors," 2010.
- [13] Qin, Y., S. Zhang, X. Zhu, J. Zhang, and C. Zhang, "Semi-parametric optimization for missing data imputation," *Appl. Intell.*, vol. 27, no. 1, pp. 79–88, 2007.
- [14] Guo, G., H. Wang, D. Bell, Y. Bi, and K. Greer, "An kNN model-based approach and its application in text categorization," in *International Conference on Intelligent Text Processing and Computational Linguistics*, 2004, pp. 559–570.
- [15] Guo, G., H. Wang, D. Bell, Y. Bi, and K. Greer, "KNN model-based approach in classification," in *OTM Confederated International Conferences*, 2003, pp. 986–996.
- [16] Li, B., Y. W. Chen, and Y. Q. Chen, "The nearest neighbor algorithm of local probability centers," *IEEE Trans. Syst. Man, Cybern. Part B*, vol. 38, no. 1, pp. 141–154, 2008.
- [17] Gates, G., "The reduced nearest neighbor rule (corresp.)," in *IEEE transactions on information theory*, 1972, vol. 18, no. 3, pp. 431–433.
- [18] Li, B. L., Yu, S. W., & Lu, Q. (2003). An improved k-Nearest neighbor algorithm for text categorization. In *International Conference on Computer Processing of Oriental Languages [ICCPOL]*, 2003.
- [19] Zhang, S., "Challenges in KNN Classification," *IEEE Trans. Knowl. Data Eng.*, 2021.
- [20] Hastie, T. and R. Tibshirani, "Discriminant adaptive nearest neighbor classification and regression," in *Advances in neural information processing systems*, 1996, pp. 409–415.
- [21] Domeniconi, C., J. Peng, and D. Gunopulos, "Locally adaptive metric nearest-neighbor classification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 9, pp. 1281–1285, 2002.
- [22] Mladenova, T., & Valova, I. (2022, October). Comparative analysis between the traditional K-Nearest Neighbor and Modifications with Weight-Calculation. In 2022 International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT) (pp. 961-965). IEEE.
- [23] Chaudhuri, B. B., "A new definition of neighborhood of a point in multi-dimensional space," in *Pattern Recognition Letters*, 1996, vol. 17, no. 1, pp. 11–17.
- [24] Mladenova, T., & Valova, I. (2022, June). Comparative Analysis Between the Traditional and Nearest Centroid Methods in the K-Nearest Neighbor Algorithm. In 2022 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA) (pp. 1-4). IEEE.
- [25] Dudani, S. A., "The distance-weighted k-nearest-neighbor rule," in *IEEE Transactions on Systems, Man, and Cybernetics*, 1976, no. 4, pp. 325–327.
- [26] Gou, J., L. Du, Y. Zhang, T. Xiong, and others, "A new distance-weighted k-nearest neighbor classifier," *J. Inf. Comput. Sci.*, vol. 9, no. 6, pp. 1429–1436, 2012.
- [27] Zeng, Y., Y. Yang, and L. Zhao, "Pseudo nearest neighbor rule for pattern classification," in *Expert Systems with Applications*, 2009, vol. 36, no. 2, pp. 3587–3595.
- [28] Mitani, Y. and Y. Hamamoto, "A local mean-based nonparametric classifier," in *Pattern Recognition Letters*, 2006, vol. 27, no. 10, pp. 1151–1159.
- [29] Gou, J., T. Xiong, and Y. Kuang, "A Novel Weighted Voting for K-Nearest Neighbor Rule," *J. Comput.*, vol. 6, no. 5, pp. 833–840, 2011.
- [30] Keller, J. M., M. R. Gray, and J. A. Givens, "A fuzzy k-nearest neighbor algorithm," in *IEEE transactions on systems, man, and cybernetics*, 1985, no. 4, pp. 580–585.
- [31] Zadeh, L. A., "Fuzzy sets," in *Fuzzy sets, fuzzy logic, and fuzzy systems: selected papers by Lotfi A Zadeh*, World Scientific, 1996, pp. 394–432.
- [32] Gu, Q., L. Zhu, and Z. Cai, "Evaluation measures of the classification performance of imbalanced data sets," in *International symposium on intelligence computation and applications*, 2009, pp. 461–471.
- [33] Hockenmaier, J., "Introduction to Machine Learning, Lecture 9: Evaluation." University of Illinois, 2015. [Online]. Available: <http://courses.engr.illinois.edu/cs446>
- [34] Hossin, M. and M. N. Sulaiman, "A review on evaluation metrics for data classification evaluations," *Int. J. data Min. & Knowl. Manag. Process*, vol. 5, no. 2, p. 1, 2015.
- [35] Watt, J., R. Borhani, and A. Katsaggelos, *Machine learning refined: foundations, algorithms, and applications*. Cambridge University Press, 2020.

The Impact of Design-level Class Decomposition on the Software Maintainability

Bayu Priyambadha¹, Tetsuro Katayama²

Interdisciplinary Graduate School of Agriculture and Engineering, University of Miyazaki, Miyazaki, Japan^{1,2}
Faculty of Computer Science, Universitas Brawijaya, Malang, Jawa Timur, Indonesia¹

Abstract—The quality of the software's internal structure tends to decay due to the adaptation to environmental changes. Therefore, it is beneficial to maintain the internal structure of the software to benefit future phases of the software life cycle. A common correlation exists between decaying internal structures and problems like software smell and maintenance costs. Refactoring is a process to maintain the internal structure of software artifacts based on the smell. Decomposition of classes is one of the most common refactoring actions based on Blob smell performed at the source code level. Moving the class decomposition process to the design artifact seems to affect the quality and maintainability of the source code positively. Therefore, studying the impact of design-level class decomposition on source code quality and software maintainability is essential to ascertain the benefits of implementing design-level class decomposition. The metrics-based evaluation shows that the design-level class decomposition positively impacts the source code quality and maintainability with the rank biserial value is 0.69.

Keywords—Design level refactoring; class decomposition; class diagram decomposition; software quality; software internal quality; software maintainability

I. INTRODUCTION

Internal structure quality maintenance is an essential task in every phase of the software development process. Due to software changes, software's internal structure or design tends to decay or decrease quality [1]–[3]. Some software is changed without using a modern software engineering approach. As a result, it may never be well structured and optimized for understandability. A software developer often changes the software artifact for the short-term goal. The changes are made mostly not followed by a comprehensive analysis of the existing artifact structure. Sometimes, the changes in one unit of software artifact require adjustment at the other unit, so the structure is always maintained. It is not uncommon that the decaying condition of the software structure will have other impacts, such as the immersing of smell on the software artifacts.

Finding and solving the smell problems in the software artifact is a software research field that continues growing until now [2], [4]–[7]. Many approaches are immersed in detecting the smell in the source code artifact [2]. Also, several researchers have defined how to solve every type of smell in source code without altering the outside behavior, known as Refactoring. The smell term was also started to understand at the level of design phases (software developing process) [4]–[6]. Design smell is defined using the knowledge that lies in the

design artifact, for example, the class diagram or the other architectural diagram that expresses the software architecture. But it is still a lack of refactoring process at the design-level artifact using the terms of design smell due to the abstract information in the design artifact [1], [8].

A previous study was conducted to do the refactoring process at the design artifact. The studies focused on the Blob smell that was identified in the class based on the data from the class diagram. Class decomposition is the process of refactoring to solve the Blob smell in the class diagram. Using the threshold-based agglomerative hierarchical clustering algorithm that is based on the class diagram metrics the class decomposition shows promising results [8]. The class diagram metrics, in this case, are *syn* and *sem*, representing syntactic and semantic metrics, respectively. Those metrics measure the relationship between class elements. Then, the approach was enhanced by using the evaluation process to solve the misplaced element and unusable class [9]. The research has shown promising results and is worth continuing to the other pathway solution than the Blob smell.

Before it is continued to the other pathway refactoring solution in the design-level artifact, it is better to know the impact of existing approaches to software maintainability. The existing refactoring process at the source code level has already been proven to have a good impact on software maintainability [10], [11].

This research examines the impact of design-level class decomposition on software maintenance. The decomposition process will be carried out using class diagrams. Once the decomposition recommendations are generated, it will be implemented into the source code. After the code has been implemented (class decomposition), several software metrics are used to measure the quality of a piece of source code after its implementation. Measurement results are compared, before and after decomposition, to determine whether there are any differences or effects on software maintainability.

The following section will describe the literature study (Section II) and the whole method of the design-level class decomposition. The class decomposition process on the class diagram is described in Section III. Section IV describes the current experiment scenario to know the impact of the design-level class decomposition on software maintenance. Finally, Sections V and VI represent the result of the experiment, the analysis using the statistical approach, and the overall conclusion of this experiment.

II. LITERATURE STUDY

Maintaining the software's internal structure also benefits the future phase of the software life cycle. The software life cycle is not only one cycle and finish. Mostly it will continue cycling as long as the user and environment need the software. During the cycle, the software experiences change due to user needs and environmental changes. The changes must be applied to the software to prevent the existing software in the specific environment. The software changes require costs we must pay [12], [13]. Therefore, the software engineer must maintain the internal structure to make changes easier and not costly. The bad structure makes the artifact difficult to understand, change, and maintain.

Every effort has been made to maintain the software's internal structure's quality, starting from the source code. Therefore, shifting from the source code to the design level of Refactoring is considered worth doing to increase the quality awareness of the design artifact as early as possible. But, refactoring activity at a higher level of abstraction has a specific problem [1], [8]. The design artifact contains less information than the implementation artifact. Therefore, excavating or mining the design artifacts and analyzing their in-depth information is necessary.

Generally, design artifacts are only written (text-based) with information that is contained within them. However, information may sometimes contain hidden meanings that require further analysis to be understood. The use of natural language processing (NLP) or semantic analysis (SA) is one approach that can provide functionality for understanding the meaning of information [14].

In contrast, source code-level information clearly provides complete information about the source code profile [1] for example, the number of operands or operators in the source code can be measured to determine the complexity of the source code by the software engineer. In addition, by reviewing the internal source code, the developer will be able to understand the relationship between attributes and methods. They can review the assigning value statement to determine the relationship between the method and attribute.

On the side of design, a review and assessment of the quality of the artifact design can be carried out by utilization of NLP and SA. Furthermore, the refactoring activity, such as class decomposition, using the design artifact is also possible using the NLP and SA analysis [1], [8].

Shifting the refactoring activity to the design artifact is expected to contribute and positively impact software maintenance [11], [15]. There are two fundamental theories of the research of design-level software refactoring. First, the theories of Software Evolution (specifically in software refactoring) [2] and the second is Model-Driven Software Engineering (MDSE) [16]. Software refactoring preserves the quality of the software's internal structure in relation to the software evolution, specifically in the case of software maintainability. On the other hand, MDSE provides the concept that software development is oriented on a model. Therefore, the model acts as the core of action and the guidance of the implementation phase. In other words, the model is the bridge between requirement analysis and implementation. The previous research on design-level Refactoring aims to combine the theories (Refactoring and MDSE) to propose a better approach for refactoring for better software quality [1]. It also has the aim to increase the awareness of internal quality as soon as possible. The other reason for the shift to the design phase is to gain the benefit of MDSE. Fig. 1 shows the thinking schema of the approach according to the justification or rationale. Furthermore, the impact of the Refactoring on the design phase is needed to be investigated.

MDSE uses a software model as the primary artifact of software development [16]. Compared to the implementation artifact (source code), the software model is closer to the problem domain. The model transformation is the main process of the MDSE since the MDSE aims to generate the source code from the models. On the other hand, there is another approach to the development of software called Code-centric Development (CcD). A comparison study between MDSE and CcD has been done for over a decade. From the review paper by Domingo et al., many researchers have been evaluating the benefit of the MDSE [17]. Some works said that MDSE decreases development time (up to 89%) relative to Code-centric Development (CcD). The other works suggest that the MDSE is suitable for academic exercise. Furthermore, the other works assert that MDSE is also suitable for inexperienced developers. Finally, Domingo et al., based on their review of the MDSE, conclude that the MDSE is suitable for academic exercise and inexperienced developers. It would be beneficial to move refactoring activities to the design artifact utilizing the benefits of MDSE.

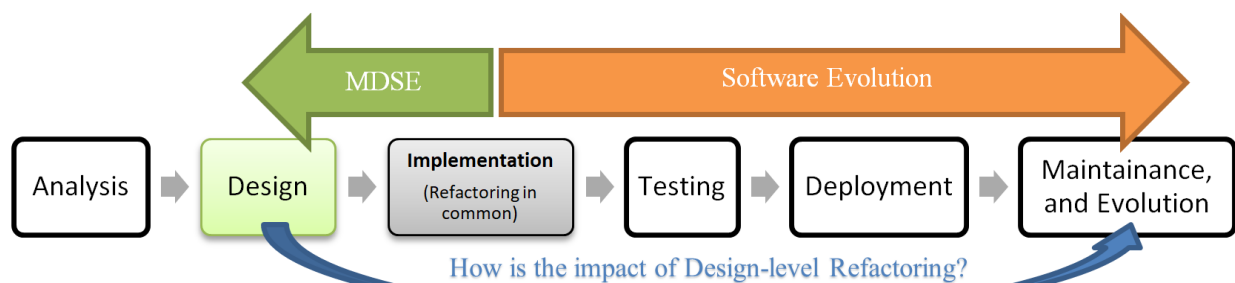


Fig. 1. Thinking schema of design-level class decomposition.

On the other side of view, the impact of the design-level refactoring process on maintenance is questionable and needs study. There are several methods to measure software maintainability. Software maintainability, the ease with which a software system can be understood and modified to accommodate bug fixes, new features, and improvements in general, plays an important role in software quality. Since it is difficult to measure software maintainability without measuring the actual maintenance process, researchers and practitioners often use product metrics as indicators. Some practitioners sometimes face difficulty in finding suitable metrics, specifically software maintainability metrics, according to the practitioner's scope or subjective. The paper of Saraiva et al. has aimed to propose Object-Oriented Software Maintainability (OOSM) metrics categorizations. It aims to make it easy to find suitable maintainability metrics [18]. Maintainability is related to external quality attributes such as analyzability, changeability, stability, and testability. All external attributes are expressed as maintainability characteristics that are "easy to adapt." To achieve this maintainability objective, most researchers measure from the internal quality attributes at least the system's size, complexity, coupling, and cohesion [18].

The other metric that points to software maintainability is named Maintainability Index (MI) [19]. MI is a software metric that is used to measure the level of software, whether it is easy or difficult to maintain in the future. The MI is calculated by considering the Line of Code (LOC), Cyclomatic Complexity, and Halstead Volume (HV). MI is measured using the information of software source code.

III. DESIGN-LEVEL CLASS DECOMPOSITION

This section describes how the refactoring activity can perform in the design-level artifact. This research uses the class diagram as the main object in the refactoring process. The refactoring activity on the class diagram utilizes the class information extracted from the class diagram. One of the challenges in this research is using the existing information to do the refactoring activity. Fig. 2 explains the proposed design-level class decomposition approach to solve Blob smell in class (class diagram).

A. Design-Level Information Extraction

The first task in this approach is to collect or extract information from the class diagram. The class diagram is a notation-based diagram that is expressed as an image showing the software's architecture in case of a class arrangement, and it is a static diagram. The information in the class diagram is important to collect and analyze to support the process. The information extraction needs a specific strategy to make it easy to implement.

The first step in this process is converting the class diagram to an XML formatted file. The conversion aims to transform the notation based to text-based information. Then, the information (text-based) is extracted from the XML syntax to get important data for smell detection (the next process). The extraction uses a specific search algorithm called Tree-based keyword search [20].

The candidate data (classes) for smell detection consists of seven pieces of information. The data are the number of attributes, number of methods, number of relations between methods and attributes, number of relations between methods and attributes, number of relations between attributes and attributes, the capacity of the relationships within the class, and the degree of cohesion [21]. The NLP and SA determine the relationship between the class elements. In addition, the degree of cohesion is calculated based on the relationship between the class elements [22].

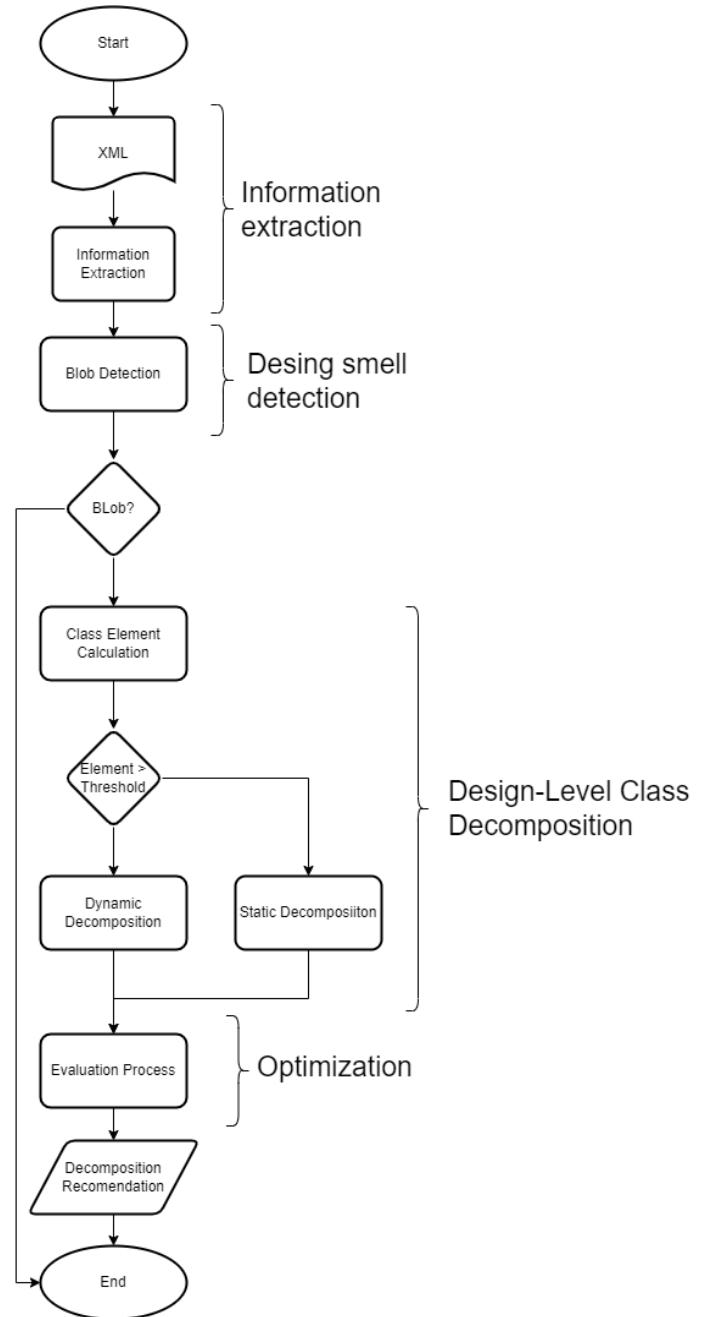


Fig. 2. The design-level class decomposition (design-level refactoring).

B. Design Smell Detection

The classification method will be used to detect the smell. The experiment used three classifiers: the j48, the Multi-Layer Perceptron, and the Naïve Bayes. The experiment uses Weka as machine learning software to solve data mining problems and run using a basic configuration. Classifiers are used to demonstrate that the dataset can be distinguished from bad smells (Blob and Feature Envy) [21]. The experiment result shows an average accuracy of 80.67% for the Blob smell. The accuracy value means that the information data can be utilized for the Blob detection process in a class diagram.

The information (seven pieces of information from the class diagram) contained characteristics that led to the identification of the Blob smell [21]. Hence, the Blob smell is being addressed in the Refactoring.

C. Design-Level Class Decomposition

Once the Blob smell is detected, Refactoring must be performed to solve the smell. Based on the experiences, the Blob smell is solved using the class decomposition or extraction. Mostly the research about class decomposition is done at the level of source code [2], [23]–[30]. Then based on the information in the class diagram [21], the class decomposition is done.

The design-level class decomposition in this research uses threshold-based agglomerative hierarchical clustering. It is divided into static and dynamic threshold hierarchical clustering. Static thresholds differ from dynamic thresholds in defining the threshold before decomposing the cluster. In the static approach, the threshold value is defined at the beginning of the decomposition process (the threshold is defined only once). According to the dynamic approach, the threshold is calculated at every stage of the decomposition process. In this study, Hamdi's algorithm is used, but it is implemented at the level of design [29]. Shifting objects to class diagrams requires defining new metrics for clustering. Syntax (*syn*) and semantic (*sem*) aspects of the class element's label are considered in the similarity matrix to do the clustering process (class decomposition) [8]. The two aspects (syntax and semantics) are considered due to the nature of class diagram information, which is more abstract than source code information. To determine the relationship between class elements, it has to determine the closeness meaning of the label name between elements. This seems to be the essential approach in this process.

The process shows the promising result of decomposition (based on the Silhouette value). But, there are still shortcomings to solve. Some elements still have a negative Silhouette value in the decomposition result. Negative Silhouette values indicate that the current element is far from other cluster elements or has the wrong placement. Additionally, the negative Silhouette elements are considered to be the least desirable.

The results also indicate that some clusters are considered unimplementable due to the possibility that they may produce objects that cannot collaborate with each other. A cluster with only one element, particularly if the element has a private modifier, is considered useless. Evaluating the moving

mechanism of the negative element is considered important as an optimization process. The next process is the evaluation process to the result of this process.

D. Optimization of Class Decomposition Result

The result of the decomposition from the previous process has to be optimized to solve the negative element and unusable cluster. The elements are evaluated by considering the value of Silhouette ($s(i)$) and class usability ($CUability$). $CUability$ value calculated based on the existence of the public method inner the cluster (value one if exists and 0 if not). Then to evaluate the cluster and elements, the following formula is used [9].

$$Eval = a.s(i) + b.CUability \quad (1)$$

Where a and b are the weight of every factor to adjust during the experiment. Threshold-based agglomerative hierarchical clustering experiment has been optimized by adding an evaluation process. During the evaluation process, a specific element with a negative Silhouettes value in each cluster is intended to be moved to a better cluster.

In comparison to the previous approach, the evaluation process increases the average Silhouettes of the cluster by using a higher or equal to 0.7. There has been an average increment of Silhouettes of about 40% [9]. Based on the results of the previous approach, the evaluation process is also able to solve the unusable cluster.

Finally, the whole process produces a set of clusters that represent the classes as the result of the decomposition. In this step, the result is the recommendation to be implemented at the source code level. Then, the impact of the decomposition implementation on the source code level will be described and analyzed in the following section.

IV. EXPERIMENT SCENARIOS

Understanding the impact of the refactoring result on the software maintainability is essential. The impact of class decomposition recommendation result (design-level class decomposition) on the quality of source code is the way to know how is the performance of the design-level class decomposition approach.

A. Overview of Scenario

The decomposition of the class on the source code level based on the recommendation from class decomposition on the design level aims to compare the quality. The source code quality before and after implementing the class decomposition recommendation will lead us to conclude how the design-level decomposition will impact the source code. The quality of code is measured using the source code quality metrics. Fig. 3 shows how the experiment is held to study the impact of design-level class decomposition. The following five internal quality attributes are associated with software maintainability: size, complexity, coupling, cohesion, and constraints associated with software architecture [18]. Four of the internal quality attributes are expressed in the 18 metrics. Table I shows the list of the 18 metrics used in this experiment to compare the before and after decomposition process. The purpose of this experiment is not only to compare the quality of source code

based on the 18 metrics but also to compare the MI [19] as the final proof of this research.

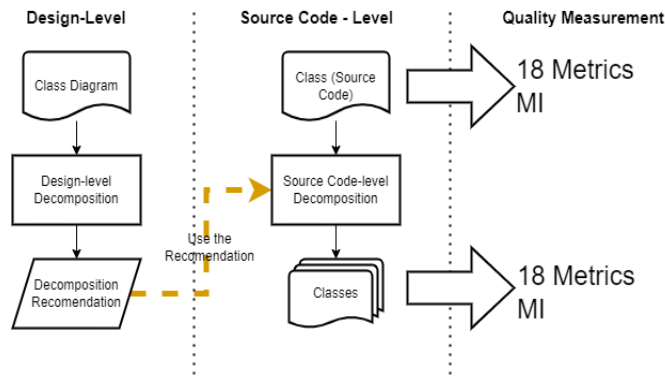


Fig. 3. The experiment scenario.

TABLE I. THE LIST OF THE 18 METRICS

No.	Name	Description
1.	CBO	Coupling Between Object Classes, the number of coupled classes (Coupling)
2.	RFC	Response For a Class, the number of methods that can be potentially invoked in response to a public message received by an object of a particular class (Complexity)
3.	SRFC	Simple Response For a Class, the number of methods that can be potentially invoked in response to a public message received by an object of a particular class (Complexity)
4.	DIT	Depth of Inheritance Tree, the position of the class in the inheritance tree (Complexity)
5.	NOC	Number of Children, the number of direct subclasses of a class (Coupling)
6.	WMC	Weighted Method Count, The weighted sum of all class' methods and represents the McCabe complexity of a class (Complexity)
7.	LOC	Line Of Code (Complexity, Size)
8.	CMLOC	Class-Method Lines of Code, Total number of all nonempty, non-commented lines of methods inside a class (Complexity, Size)
9.	NOF	Number Of Fields, the number of attributes in class (Complexity, Size)
10.	NOSF	Number of Static Fields, the number of static attributes (Complexity, Size)
11.	NOM	Number of Methods (Complexity, Size)
12.	NOSM	Number of Static Methods (Complexity, Size)
13.	NORM	Number of Overridden Methods (Complexity)
14.	LCOM	Lack of Cohesion of Methods, measure how methods of a class are related to each other (Cohesion)
15.	LCAM	Lack of Cohesion Among Methods (1-CAM), CAM metric is the measure of cohesion based on parameter types of methods (LCAM = 1-CAM) (Cohesion)
16.	LTCC	Lack of Tight Class Cohesion, The Lack of Tight Class Cohesion metric measures the lack of cohesion between the public methods of a class (Cohesion)
17.	ATFD	Access to Foreign Data, the number of classes in which the attributes are directly or indirectly reachable from the investigated class (Coupling)
18.	SI	Specialization Index measures subclasses override their ancestor's classes (Complexity)

All the measurement results are collected to be analyzed in the following step. For the result of measurement using the 18 metrics, the data will be recapped and show the trend of comparison before and after decomposition.

Lastly, for measuring MI, statistical analysis is needed to determine the impact of the decomposition recommendation on the source code quality.

B. Experiment Data

TABLE II. MI CLASSIFICATION

MI Value	Classification
>85	Highly maintainable
>65 and ≤85	Moderate maintainable
≤65	Difficult to maintain

This experiment used two study cases, jHotDraw and AgroUML source code. There are 67 classes identified as Blob classes using jDeodorant in both applications. But, after measuring the MI, not all classes are considered problematic in maintenance. The classification of MI value refers to Table II, which explains how the value is classified based on maintainability [19] There are only 33 classes that have moderate and difficult to maintain. Therefore, only 33 classes are used as the object in this experiment. The acquisition of data is shown in Fig. 4.

The Blob classes classified as highly maintainable are not used in this experiment because it assumed not to be included in problematic classes in the maintainability manner.

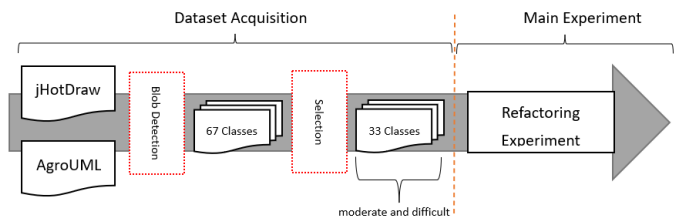


Fig. 4. Data acquisition.

But, it has to solve from the other perspective manner.

C. Tools

There are three tools used in this experiment, the jDeodorant plugin for Eclipse IDE [31], the CodeMR (Code Magnetic Resonance), and the prototype application that implements the MI measurement [19].

The jDeodorant is used in the data acquisition process to select the classes that contain a Blob smell. The CodeMR is the application that has the ability to measure the quality of source code based on the 18 metrics measurement. CodeMR is a static analysis tool for source code. And the last is a custom application that can show the value of the Maintainability index of source code. CodeMR and the custom application used on the before and after decomposition process. The result of measurement is recapped and analyzed to know how the impact of the usage design-level class decomposition recommendation on the source code quality.

V. EXPERIMENT RESULT AND DISCUSSION

The class decomposition recommendation is implemented on the source code to get the benefit of it. The result is measured using 18 metrics and MI to know how difference before and after the decomposition process.

A. Measurement using the 18 Metrics

The source code decomposition result is measured using the 18 metrics for the first result. The 18 metrics are grouped into four metrics based on the metrics type. Every value of each metric is calculated by averaging the values of particular metrics in every case study. Then, it differentiated before and after decomposition. The groups are coupling, complexity, cohesion, and size metric. Fig. 5 shows the result of the measurement that is described on the line graph. Another reason for grouping metrics is that each type of metric has a different range of values, so separating each type into groups will clarify trends for each type of metric.

Fig. 5(a) shows the trend of measurement in the type of coupling metric. There are three metrics in the category coupling metric, CBO, NOC, and ATFD. In this result, CBO shows a decrement value from before to after decomposition.

The other metrics, NOC and ATFD, do not show decrement due to the value equality between before and after decomposition.

Fig. 5(b) shows the group of complexity metrics consisting of ten metrics. The metrics are RFC, SRFC, DIT, WMC, SI, NOF, NOSF, NOM, NOSM, and NORM. Those metrics measure the complexity of source code from several sides. For the ten metrics, the graph shows the trend that the values decrease after decomposition. Two metrics show the same value before and after decomposition. The metrics are SI and NORM that has a value of 0 before and after decomposition.

Fig. 5(c) shows the cohesion metric, consisting of three metrics: LCOM, LCAM, and LTCC. All metrics show the measurement of a lack of cohesion in the class. Higher values show a higher lack of cohesion in the source code. The value of those metrics decreases before and after the decomposition process.

Fig. 5(d) shows the size metric, which measures the size of the source code. It seems to be the same trend as the other type of metrics. The value of LOC and CMLOC, before and after decomposition, decreases due to the result of decomposition implemented to the source code.

All metric types show the same trend that, after decomposition, tend to be lower value of metrics. The 18 metrics show the same meaning of the value: the lower value means the better condition of the source code. Implementing class decomposition on the source code seems to make the source code better quality measured by the 18 metrics.

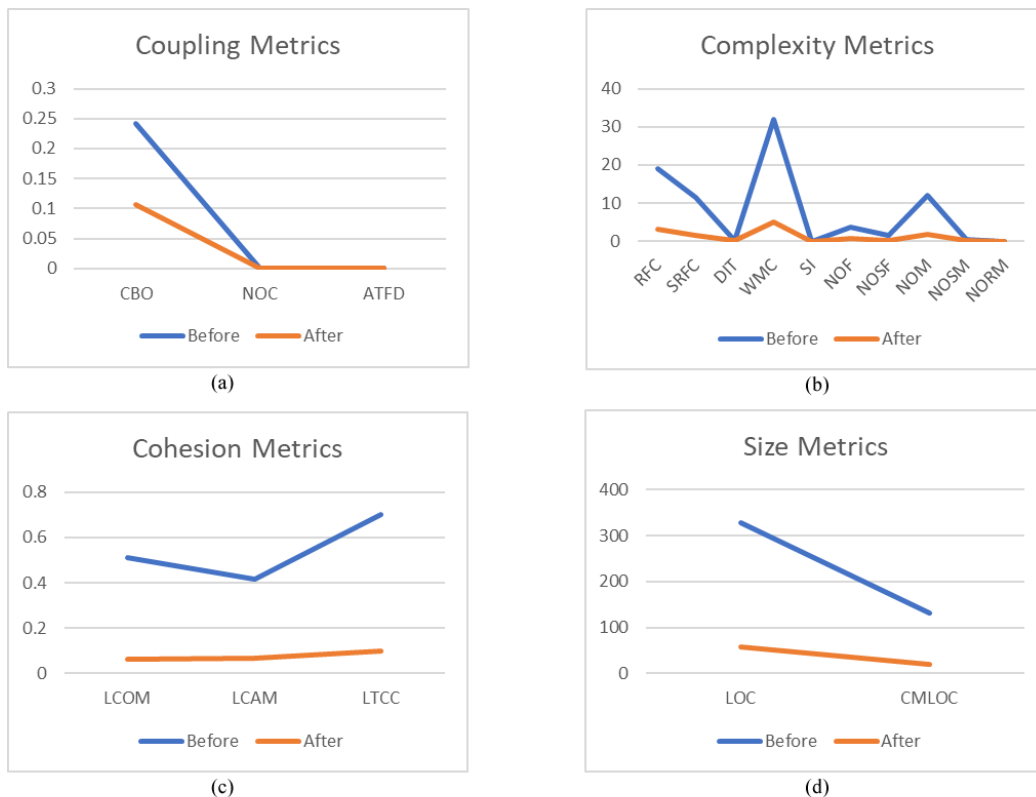


Fig. 5. Result of measurement using the 18 metrics of a) Coupling metrics, b) Complexity metrics, c) Cohesion metrics, and d) Size metrics.

B. Statistical Analysis of the Maintainability Index (MI)

The Maintainability Index (MI) measurement is done using the prototype application. The measurement is applied before and after the class decomposition process based on the design-level recommendation. The MI value for the after-decomposition is calculated by averaging each class's value aims to represent one value. Table III shows the result of the measurement of MI compared before and after the decomposition process.

In this experiment, the differences between before and after are worth calculating to ensure the differences after decomposition. The differences calculation uses the statistical approach, in this case, Wilcoxon signed-rank. Analysis differentiation aims to make sure that there is a difference between before and after decomposition. Differentiation can be used as a sign that the decomposition process causes an impact on the source code in case of maintainability.

Besides the differences, how strong the effect of the usage of design-level decomposition recommendation in the source code level decomposition to the value of MI is also important to know. The Wilcoxon signed rank is able to inform both the differences and the effect size of the approach.

Based on the result of the Wilcoxon signed rank, there are several interpretations based on the test result. Fig. 6 shows the first indicator by the p-value of the result. The significant value of differential analysis is lower than the 0.05 p-value. The current result shows that the p-value is 6.02e-04, lower than 0.05. So, based on the p-value, the result concluded that the MI before and after the decomposition process is significantly different.

The second indicator is the median value from the plot in Fig. 6. Even though the median value cannot act as the main indicator of differentiation and it shows how the differences in spreading data differ. Based on the result, the median values before and after decomposition differ in favor. The median of after decomposition data is increased by 25.03 to the before.

This research aims to know how the impact of the utilization of design-level class decomposition recommendations on the source code level. The p-value and median value only show that the data before and after is different. It does not show how the impact of the design-level class decomposition on the source code quality simultaneously. The other value could be used to know how strong the impact design-level decomposition is rank biserial, as shown in Fig. 6. The rank biserial is used to examine the relationship between dichotomous (binary) nominal data and ordinal (ranked) data. Before running the statistical analysis, the data measurement of MI is calculated to find the data rank based on the differences in the value of MI before and after decomposition. It is one of the Wilcoxon sign rank method's requirements before shown in the plot as shown in Fig. 6. The rank biserial value shown in Fig. 6 is 0.69. Therefore, the higher value is better. Based on Funder's interpretation [32], 0.69 can be interpreted as very

large. In other words, the use of design-level decomposition recommendation on the source code level decomposition gives a very large, positive, and significant effect on the MI.

TABLE III. MI BEFORE AND AFTER DECOMPOSITION PROCESS

Differentiation of MI			
No.	Class Name	Before	After
1.	ArgoEventPump	80.78	99.02
2.	ArgoFontChooser	82.76	100.13
3.	ArgoParser	84.66	116.53
4.	DetailsPane	80.38	97.84
5.	DrawApplet	75.41	120.47
6.	DrawApplication	68.84	122.26
7.	ExplorerPopup	62.42	91.33
8.	FindDialog	66.81	70.96
9.	GenericArgoMenuBar	62.19	89.18
10.	GraphLayout	72.14	72.43
11.	Import	69.4	81.42
12.	MyTokenizer	84.63	98.55
13.	NotationSettings	80.38	107.75
14.	PathItemPlacement	80.93	98.25
15.	PerspectiveConfigurator	63.85	46.43
16.	PerspectiveManager	66.65	88.60
17.	ProfileConfigurationParser	80.97	83.60
18.	ProfileUML	70.41	35.21
19.	ProjectBrowser	57.37	105.78
20.	SettingsTabProfile	53.51	62.48
21.	StandardDrawingView	71.48	110.70
22.	TabConstraints	73.45	104.59
23.	TabStyle	80.92	124.74
24.	TargetManager	71.62	82.97
25.	ToDoList	81.37	125.75
26.	TodoParser	84.07	92.41
27.	UMLActivityDiagram	69.27	101.92
28.	UMLAddDialog	83.8	42.25
29.	UMLDeploymentDiagram	76.53	113.31
30.	UMLStateDiagram	65.81	104.69
31.	UMLUseCaseDiagram	84.26	101.36
32.	UserDefinedProfile	67.57	98.48
33.	WizOperName	77.7	39.25

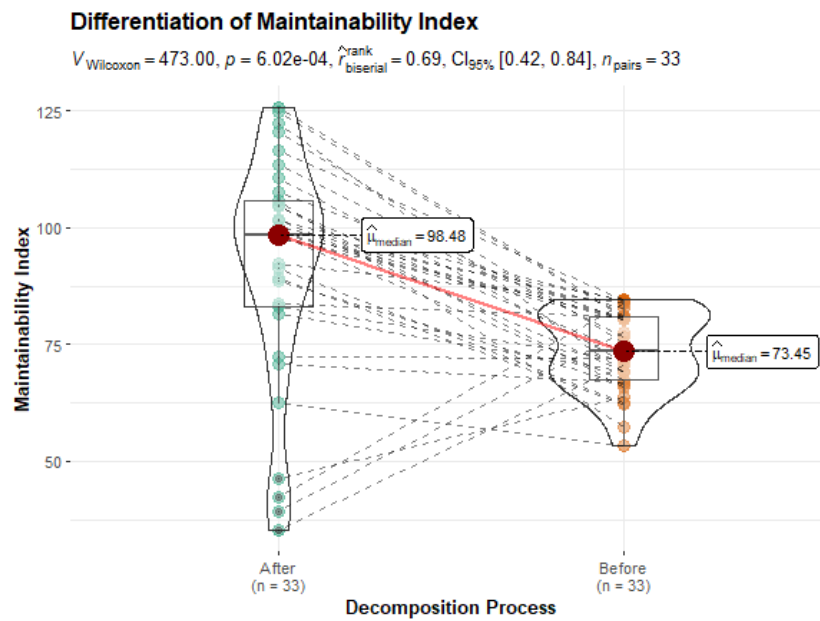


Fig. 6. Differentiation of MI.

C. Research Limitation

This section explains all things that have a possibility to threaten the validity of the experiment result. There are two main limitations to this experiment. First is the dataset used and the manual process conducted in the experiment.

A limited amount of data in the experiment would be the main limitation of this research. The study case is taken from only the two application resources (jHotDraw and jDraw). The dataset collected from both applications is only 33 instant data. It is sufficient for the scope of this experiment, but it is considered better to add the amount of data in the future to increase the result validity.

The process of class decomposition is done automatically. But, the implementation of decomposition recommendations to the source code is done manually based on the location of class elements.

VI. CONCLUSION

The quality measurement before and after the decomposition process on the source code is by using two approaches. First, the source code was measured using the 18 metrics representing coupling, complexity, cohesion, and size. Those groups of metrics are the type of metrics that are related to software maintainability based on the existing references. There is a trend in all metrics types that after decomposition, the metrics tend to have a lower value than before decomposition. In all 18 metrics, a lower value represents a better condition of source code. By implementing design-level class decomposition on the source code, the source code seems to be of better quality as measured by the 18 metrics.

The second quality measurement uses MI as one specific metric to measure the software maintainability of the source code. The measurement result differentiated before and after the decomposition process. The Wilcoxon signed-rank analysis

was applied to the result of measurement to get a deep analysis of the result. A p-value less than 0.05 indicates significant differential analysis in the first test. According to the current results, the p-value is 6.02e-04, which is less than 0.05. Therefore, it is concluded that the MI before and after decomposition is significantly different. The second indicator is the median value from the plot. Regardless of the fact that the median value cannot be used as the primary indicator of differentiation, it does at least indicate how the spread of data differs.

According to the results, the median value before and after decomposition differs in favor of decomposition. After decomposition, the median has increased by 25.03 compared to the before data.

The final test is rank biserial. There is a rank biserial value of 0.69, which can be interpreted as being very large. As a result, using design-level recommendations on source code decomposition has a very large, positive, and significant effect on the MI.

The 18 metrics and MI analysis show the same favorable result. The use of design-level class decomposition recommendation is able to increase the source code quality significantly based on the analysis result.

The shifting refactoring process to the design artifact is still challenging in the future. This is because so many code smell types could detect and refactor from the design artifact. This research only focuses on the Blob smell on the design artifact, only defining the pathway solution based on the existing Blob smell in the class diagram. The research will continue to the other pathway solution than the Blob smell.

This research uses the data collected from the existing open-source application. The limitation on the number of data might be lacking in the meter of data validity. Increasing the number of data is a plan that has been recorded to be carried

out in the future. The complete software documentation will be interesting data to analyze for future research.

REFERENCES

- [1] B. Priyambadha and T. Katayama, "Enhancement of Design Level Class Decomposition using Evaluation Process," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 8, pp. 130–139, 2022, doi: 10.14569/IJACSA.2022.0130816.
- [2] M. Fowler et al., *Refactoring Improving the Design of Existing Code Second Edition*, Second Ed. United State of America: Pearson Education - Wesley, 2019.
- [3] I. Sommerville, *Software Engineering*, 9th ed. Harlow, England: Addison-Wesley Professional, 2010.
- [4] K. Alkharabsheh, Y. Crespo, E. Manso, and J. A. Taboada, "Software Design Smell Detection: a systematic mapping study," *Software Quality Journal*, vol. 27, no. 3, pp. 1069–1148, 2019, doi: 10.1007/s11219-018-9424-8.
- [5] B. K. Sidhu, K. Singh, and N. Sharma, "A Catalogue of Model Smells and Refactoring Operations for Object-Oriented Software," *Proceedings of the International Conference on Inventive Communication and Computational Technologies, ICICCT 2018*, pp. 313–319, 2018, doi: 10.1109/ICICCT.2018.8473027.
- [6] B. Kaur Sidhu, "Model Smells In Uml Class Diagrams," *International Journal of Enhanced Research in Management & Computer Applications*, vol. 5, pp. 2319–7471, 2016, Accessed: Apr. 10, 2019. [Online]. Available: <https://pdfs.semanticscholar.org/bced/a3ff00a0b577007d17abfb6bfd406058def6.pdf>
- [7] R. C. Martin, *Clean Architecture: A Craftsman's Guide to Software Structure and Design*. in Robert C. Martin Series. Boston, MA: Prentice Hall, 2017.
- [8] B. Priyambadha and T. Katayama, "Design Level Class Decomposition using the Threshold-based Hierarchical Agglomerative Clustering," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 3, pp. 57–64, 2022, doi: 10.14569/IJACSA.2022.0130310.
- [9] B. Priyambadha and T. Katayama, "Enhancement of Design Level Class Decomposition using Evaluation Process," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 8, 2022, doi: 10.14569/IJACSA.2022.0130816.
- [10] G. Szöke, G. Antal, C. Nagy, R. Ferenc, and T. Gyimóthy, "Empirical study on refactoring large-scale industrial systems and its effects on maintainability," *Journal of Systems and Software*, vol. 129, pp. 107–126, 2017, doi: 10.1016/j.jss.2016.08.071.
- [11] S. Kaur, A. Kaur, and G. Dhiman, "Deep analysis of quality of primary studies on assessing the impact of refactoring on software quality," *Mater Today Proc*, no. xxxx, 2021, doi: 10.1016/j.matpr.2020.11.217.
- [12] A. Yamashita and L. Moonen, "Exploring the impact of inter-smell relations on software maintainability: An empirical study," in *Proceedings - International Conference on Software Engineering*, 2013, doi: 10.1109/ICSE.2013.6606614.
- [13] F. Palomba, D. Di Nucci, A. Panichella, R. Oliveto, and A. De Lucia, "On the diffusion of test smells in automatically generated test code: An empirical study," in *Proceedings - 9th International Workshop on Search-Based Software Testing, SBST 2016*, 2016, doi: 10.1145/2897010.2897016.
- [14] B. Priyambadha, T. Katayama, Y. Kita, H. Yamaba, K. Aburada, and N. Okazaki, "Utilizing the similarity meaning of label in class cohesion calculation," *Journal of Robotics, Networking and Artificial Life*, vol. 7, no. 4, pp. 270–274, 2021, doi: 10.2991/jrnal.k.201215.013.
- [15] S. Kaur and P. Singh, "How does object-oriented code refactoring influence software quality? Research landscape and challenges," *Journal of Systems and Software*, vol. 157, p. 110394, Nov. 2019, doi: 10.1016/j.jss.2019.110394.
- [16] M. Brambilla, Jordi. Cabot, and Manuel. Wimmer, *Model-driven software engineering in practice*. Morgan & Claypool, 2012.
- [17] Á. Domingo, J. Echeverría, Ó. Pastor, and C. Cetina, "Evaluating the Benefits of Model-Driven Development," *Advanced Information Systems Engineering*, no. June 2021, pp. 353–367, 2020, doi: 10.1007/978-3-030-49435-3_22.
- [18] J. D. A. G. Saraiva, M. S. De França, S. C. B. Soares, F. J. C. L. Filho, and R. M. C. R. De Souza, "Classifying metrics for assessing Object-Oriented Software Maintainability: A family of metrics' catalogs," *Journal of Systems and Software*, vol. 103, pp. 85–101, 2015, doi: 10.1016/j.jss.2015.01.014.
- [19] D. Coleman, D. Ash, B. Lowther, and P. Oman, "Using Metrics to Evaluate Software System Maintainability," *IEEE Computer*, vol. 27, no. 8, pp. 44–49, 1994.
- [20] B. Priyambadha and T. Katayama, "Tree-based keyword search algorithm over the visual paradigm's class diagram xml to abstracting class information," *2020 IEEE 9th Global Conference on Consumer Electronics, GCCE 2020*, pp. 280–284, 2020, doi: 10.1109/GCCE50665.2020.9291865.
- [21] B. Priyambadha, T. Katayama, Y. Kita, H. Yamaba, K. Aburada, and N. Okazaki, "The Seven Information Features of Class for Blob and Feature Envy Smell Detection in a Class Diagram," *The 2021 International Conference on Artificial Life and Robotics (ICAROB2021)*, pp. 348–351, 2021.
- [22] B. Priyambadha, T. Katayama, Y. Kita, K. Aburada, H. Yamaba, and N. Okazaki, "The Measurement of Class Cohesion using Semantic Approach," *Proceedings of International Conference on Artificial Life and Robotics*, vol. 25, pp. 759–762, 2020, doi: 10.5954/icarob.2020.os14-4.
- [23] G. Bavota, A. De Lucia, A. Marcus, and R. Oliveto, "A two-step technique for extract class refactoring," *ASE'10 - Proceedings of the IEEE/ACM International Conference on Automated Software Engineering*, pp. 151–154, 2010, doi: 10.1145/1858996.1859024.
- [24] G. Bavota, A. De Lucia, and R. Oliveto, "Identifying Extract Class refactoring opportunities using structural and semantic cohesion measures," *Journal of Systems and Software*, vol. 84, no. 3, pp. 397–414, Mar. 2011, doi: 10.1016/j.jss.2010.11.918.
- [25] G. Bavota, "Using structural and semantic information to support software refactoring," *Proceedings - International Conference on Software Engineering*, pp. 1479–1482, 2012, doi: 10.1109/ICSE.2012.6227057.
- [26] G. Bavota, A. de Lucia, A. Marcus, and R. Oliveto, "Automating extract class refactoring: an improved method and its evaluation," *Empir Softw Eng*, vol. 19, no. 6, pp. 1617–1664, 2014, doi: 10.1007/s10664-013-9256-x.
- [27] M. Fokaefs, N. Tsantalis, E. Stroulia, and A. Chatzigeorgiou, "Identification and application of Extract Class refactorings in object-oriented systems," *Journal of Systems and Software*, vol. 85, no. 10, pp. 2241–2260, 2012, doi: 10.1016/j.jss.2012.04.013.
- [28] M. Fokaefs, N. Tsantalis, A. Chatzigeorgiou, and J. Sander, "Decomposing object-oriented class modules using an agglomerative clustering technique," *IEEE International Conference on Software Maintenance, ICSM*, pp. 93–101, 2009, doi: 10.1109/ICSM.2009.5306332.
- [29] M. Hamdi, R. Pethe, A. S. Chetty, and D. K. Kim, "Threshold-driven class decomposition," *Proceedings - International Computer Software and Applications Conference*, vol. 1, pp. 884–887, 2019, doi: 10.1109/COMPSAC.2019.00130.
- [30] Y. Wang, H. Yu, Z. Zhu, W. Zhang, and Y. Zhao, "Automatic Software Refactoring via Weighted Clustering in Method-Level Networks," *IEEE Transactions on Software Engineering*, vol. 44, no. 3, pp. 202–236, 2018, doi: 10.1109/TSE.2017.2679752.
- [31] M. Fokaefs, N. Tsantalis, E. Stroulia, and A. Chatzigeorgiou, "JDeodorant: identification and application of extract class refactorings," in *2011 33rd International Conference on Software Engineering (ICSE)*, 2011, pp. 1037–1039. doi: 10.1145/1985793.1985989.
- [32] D. C. Funder and D. J. Ozer, "Evaluating Effect Size in Psychological Research: Sense and Nonsense," *Adv Methods Pract Psychol Sci*, vol. 2, no. 2, pp. 156–168, Jun. 2019, doi: 10.1177/2515245919847202.

Listening to the Voice of People with Vision Impairment

Revealing Outdoor Mobility Traits for Improved Assistive Technologies

Abeer Malkawi¹, Azrina Kamaruddin², Alfian Abdul Halin³, Novia Admodisastro⁴

Faculty of Computer Science and Information Technology, University Putra Malaysia (UPM), Serdang, Malaysia

Abstract—Extensive research developed assistive technologies (ATs) to improve mobility for people with vision impairment (PVI). However, a limited number of PVI rely on ATs for mobility. One of the factors contributing to the limited reliability and low acceptance of ATs is the developers' failure to consider PVI mobility traits from the target group's perspective. Many developers and researchers proposed solutions based on their knowledge and experiences, where PVI have been excluded from several studies except for limited involvement in testing phases. Accordingly, this study aims to bridge this gap by providing comprehensive information on PVI's behaviors, challenges, and requirements for safe and independent outdoor mobility. Therefore, a total of 15 participants with vision impairment were involved in semi-structured interviews and two observation sessions. One key finding highlights the need for AT that complements the conventional cane and overcomes its limitations, not substituting the cane. Moreover, the proposed AT should address instant mobility and future needs simultaneously. Overall, the study contributes to providing comprehensive knowledge on PVI safe and independent mobility traits, which assist AT developers to explore the potential barriers and facilitators of the adoption of ATs among PVI and leads to develop effective and reliable ATs that meet their needs. For future work, the researchers will develop an AT that complements the conventional cane, supports instant mobility, and enhances cognitive map formation.

Keywords—People with vision impairment; assistive technology; outdoor mobility; behaviors; challenges; requirements

I. INTRODUCTION

Outdoor mobility is one of the significant challenges for people with vision impairment (PVI), especially in an unfamiliar environment. Globally, PVI still rely on the white cane and guide dog for primary mobility [1]. The white cane gained widespread attention due to its robustness, object detection effectiveness, simplicity, low cost, lightweight, and portability. Nevertheless, the limitation of the cane is the inability to provide full safe mobility. For example, the cane lacks the capacity to detect chest or head-level obstacles [1], small obstacles and various terrain types [2]. Besides, cane techniques require a long training period to be well-understood [1]. The provision of complete safe mobility remains challenging when using a guided dog with limitations in dog cost, speed and path directions [3].

Based on the limitations of conventional mobility aids (white cane and guide dog), high-performance technologies have dominated interest from various disciplines to develop

ATs to support mobility for PVI [4], [5]. Although several proposed systems reported high performance and achieved their goals, only a few ATs have been accepted and considered reliable by the end-users [6], [7]. A significant contributing factor to the low acceptance of proposed ATs is the lack of understanding of the target group's requirements. Several researchers [8], [9] developed ATs based on their knowledge and experience without engaging the target group in any design phase. However, this exclusion of PVI from the research has resulted in a poor understanding of the PVI needs [10] and negatively impact the acceptance and success of the proposed AT [11]. The challenges in recruiting people with disabilities are the reason for this avoidance, as [12] stated. These challenges include logistic problems, participants' safety, and study duration.

In contrast, involving the target group in the design phases of ATs provides the developers with a comprehension of PVI's preferences, lifestyles, challenges, and requirements [3], [13], [7]. This knowledge plays a vital role to develop more effective ATs. For this reason, the study aims to answer two primary research questions: 1) what are the prevailing mobility behaviors and challenges encountered by PVI? and 2) What are the ATs' characteristics that potentially can address these challenges and fulfill PVI ambitions?

To achieve this goal, a qualitative study was conducted involving 15 participants from the PVI community. Three sessions were performed, face-to-face interviews, observation in a familiar environment, and observation in an unfamiliar environment. Participants shared their mobility practice, challenges, and narrated personal stories and ambitions for the future ATs. The study approach of involving PVI participants with three different instruments adds to the study's novelty, where the study provided insights into PVI's lived experiences that are often missing in other studies.

This study mainly highlights extensive illustration of mobility behaviors applied by the PVI to achieve safe outdoor mobility. Additionally, the significant challenges associated with outdoor mobility from the PVI perspective. Moreover, provides recommendations for ATs developers about the features of novel ATs that may satisfy the requirements and expectations of PVI to enhance outdoor safe mobility.

The subsequent aspects of this article are organized as follows: Section II presents a literature review of studies on PVI's mobility traits, challenges, and features inherent in existing solutions. The methodology employed is highlighted

in Section III whereas Section IV presents the findings. Next, the discussion and recommendation of the study are outlined in Section V and a brief conclusion was presented in Section VI.

II. RELATED WORK

A substantial amount of literature exists on mobility for PVI from target groups' perspectives. The main focus in the published studies includes the PVI behaviors for outdoor mobility, the challenges faced, requirements, and expectations of the solution that may improve their safe mobility. Although most past studies reported one or two mobility aspects, this study attempts to aggregate the whole mobility context and reveal new details expressed by PVI expressed via verbal and non-verbal behaviors.

The researchers in [7] conducted a User-Centered Design (UCD) study to specify the PVI requirements of an AT for outdoor navigation in the cultural environment. The study that involved PVI and sighted participants determined several functionality, design, and operational features of the ATs. Some findings revealed the need for GPS, obstacle avoidance, network aspects, video streaming, and vocal feedback. The researchers developed an initial prototype AT based on the findings. However, they reported that a lot of work is needed to fulfill the PVI ambitions especially from the design and functionality aspects. One study limitation is even though the study applied interviews and questionnaire instruments; PVI was involved in the questionnaire only, and the interview sessions were with stakeholders. Additionally, the findings revealed the ATs features that support mobility but it lacks the behaviors and challenges that the target group face.

The study of [14] aimed to highlighting the navigation behaviors and challenges for PVI. A total of 30 participants were interviewed through phone calls. Based on the key findings, the researchers reported that mastering orientation and mobility (O&M) skills were enough for navigation. Additionally, the white cane seeks for obstacles whereas the guide dog avoids obstructions. In terms of AT, the study found that there was no appropriate navigation technology for all PVI, rather it was based on individual needs and preferences. Factors involving high cost and complexities contributed to the absence of navigation technologies. One study limitation is the avoidance of ATs as supportive aids for PVI mobility.

In another related work, PVI-related environmental information for safe and independent commutes was investigated [15]. Semi-structured interviews were performed with 18 PVI participants. The main mobility behaviors reported in the study were asking people for help, using smell and auditory senses, assigning points of reference, and using sidewalk edges. In contrast, orientation and obstacles were demonstrated as the major challenges experienced during daily activities. Also, crossing the street was a significant risk faced by the PVI. The main obstacles impacting PVI's safety were potholes, sidewalk surface, windows and doors, pedestrians, and vendors. The public also had negative attitudes towards PVI. The study provided major challenges that PVI faces; however, it lacks the ATs features which can address these challenges and support safe mobility for PVI.

The study of [16] investigated how navigational technologies improved mobility for PVI. This qualitative research interviewed 23 participants with VI relying on the white cane, while seven of them also used a guided dog. All the participants used their smartphones to access navigation technology such as Google Maps, Apple Maps, Siri, and video calls. The key study finding suggested that safe mobility for PVI could be made more effective by combining conventional mobility aid represented by O&M skills with navigation technologies. Such aids detect the front area, and the navigation technology conveys information about the turn-by-turn route. Although the study demonstrated the importance of ATs in supporting mobility for PVI, it focused on the apps only.

Although navigation is an essential part of mobility, navigation technologies failed to consider the limitations of using conventional aids for detection, such as distinguishing small obstacles and obstacles higher than the ground level. Additionally, several researchers argued using a smartphone on-journey because it impacts on user's safety [11]. Besides, at least one hand should be free during the navigation [17] and users are concerned about battery drain for the whole trip [18].

Despite much research on mobility by involving a group of participants from the target community, several studies relied solely on interviews and excluded the real interactions with the environment. Only a few studies aggregated the mobility behaviors, challenges, requirements, and solution ambitions from PVI's viewpoints. In contrast, this study applied different instruments for data gathering; face-to-face interviews and two real-world observation sessions in familiar and unfamiliar environment, thus enabling the collection of verbal and non-verbal reactions. This study also combined the mobility context (behaviors, challenges, ambitions, and suggestions of the solution) from a diverse group of PVI.

III. METHODOLOGY

This study employed qualitative methodologies for data collection and analysis. Ethical approval with reference number (JKEUPM-2019-463) was obtained from the Ethical Committee of the University of Putra Malaysia (UPM).

A. Participants

A total of 15 adult participants with vision impairment comprising eight females and seven males were recruited. The mean age of the participants is 41 years and the standard deviation (SD) is 16.3. Among the participants, 11 were totally blind (B1) while the remaining four were partially blind (B2). Furthermore, six participants were congenitally blind, nine experienced gradual blindness occurring later in life, and none of the participants had other disabilities.

All participants relied on the white cane for mobility, 12 of them passed the O&M training course and the other three were in O&M training when this study was conducted. The manager of the Research and Development Department in the Malaysia Association for Blind (MAB) assisted in the recruitment process. The participants' demographic information is shown in Table I.

TABLE I. PARTICIPANT DEMOGRAPHIC INFORMATION

Participant	Gender	Age Range	Blindness Level	Blindness Age	Activity ^a
P1	F	41-60	B1	Adulthood	Int, Ob1, Ob2
P2	F	26-40	B1	Congenitally	Int, Ob1, Ob2
P3	F	26-40	B1	Childhood	Int, Ob1, Ob2
P4	M	26-40	B1	Adulthood	Int, Ob1, Ob2
P5	F	Above 60	B1	Adulthood	Int, Ob1, Ob2
P6	M	26-40	B2	Congenitally	Int, Ob1, Ob2
P7	M	Above 60	B1	Congenitally	Int, Ob1, Ob2
P8	F	26-40	B1	Congenitally	Int, Ob1, Ob2
P9	M	18-25	B2	Adulthood	Int, Ob1, Ob2
P10	F	26-40	B1	Childhood	Int, Ob1, Ob2
P11	F	26-40	B1	Congenitally	Int, Ob1, Ob2
P12	M	26-40	B1	Congenitally	Int, Ob1, Ob2
P13	M	18-25	B1	Adulthood	Int
P14	F	41-60	B2	Adulthood	Int
P15	M	41-60	B1	Adulthood	Int
Overall	8 F 7 M	Mean: 41 SD: 16.3	11 B1 4 B2	6 Congenitally 9 Later	

^a. Activity: Int: interview. Ob1: observation session in a familiar environment. Ob2: observation session in an unfamiliar environment

B. Instruments

Data were collected via three sessions; a semi-structured interview, Session A and two outdoor observation sessions, Session B in a familiar environment and Session C in an unfamiliar environment. All participants were involved in the face-to-face semi-structured interview. Specifically, 12 participants engaged in the two observation sessions. As advised by the O&M instructor, three participants were excluded from the observation sessions as they lacked the ability to navigate independently and they were involved in O&M training when this study was conducted.

1) *Session A – semi-structured interview*: To elicit the mobility context for PVI, face-to-face individual interviews were conducted with the 15 participants taking between 35 to 60 minutes to be completed. The interview questions were structured into three categories; demographic information and visual impairment history, outdoor mobility behaviors and challenges, and knowledge/experience on mobility ATs. All sessions were video and audio recorded, and all verbal and non-verbal cues, such as emotional signs were collected.

2) *Session B – observation I*: The first observation session was in a familiar environment and the purposes were to verify that participants could navigate independently without being exposed to any risk in the unfamiliar environment. In addition, the session was performed to verify the interview responses and identify how the PVI interact with the elements of the familiar sites. These events are expected to reveal the mobility behaviors, especially the emotional indications. Other reasons for this session were to uncover the mobility challenges with

the existence of the cognitive map and to compare the differences in mobility behaviors between the familiar and unfamiliar environments.

Each participant strolled with the white cane for about 10 minutes in the area surrounding MAB Brickfields; Kuala Lumpur, MAB buildings, and MAB courtyard. The site contains a flat area, pavements with tactile blocks, and several obstacles like trees, poles, staircases with handrails, streets, and traffic lights. The think-aloud protocol was applied, thereafter, the participants were asked to express their knowledge about environmental elements surrounding them. All the sessions were video recorded. Fig. 1 presents a sample of the familiar observation sessions.

3) *Session C – observation II*: The second observation session was in an unfamiliar environment. For safety purpose, a sighted individual walked close to the participant to intervene if required. Moreover, medical insurance was provided for each participant following the recommendation of the ethics committee.

The purposes of the unfamiliar observation session were to uncover PVI mobility behaviors and challenges during navigations, verify what the participants expressed in the interview, and identify the compatibility with environmental reactions [12]. In addition, this session was designed to gather the non-verbal cues, clues, and reactions that an individual with VI practices under real-world situations.

Each participant navigated in an unfamiliar environment for 15 to 20 minutes. The site used for the navigation exercise was in the UPM campus, besides the Sultan Salahuddin building and along a path of 250 to 300 meters as shown in Fig. 2(a) and

2(b) show unfamiliar observation site. This area was selected since all the participants were unfamiliar with the environment and controllable in terms of pedestrians, traffic, and obstacles. During the observation sessions, the think-aloud protocol was applied and the participants were asked to express their thought about everything they feel surrounding them. The sessions were also video recorded.



a) Area between MAB buildings b) Brickfields area surrounding to MAB

Fig. 1. Familiar observation sessions.

C. Data Analysis

All audio and video files were transcribed and thematic analysis was applied to analyze the data. Thematic analysis was performed because of its flexibility and provision of rich, detailed, and well-structured results [19], [20]. The analysis yielded a total of 554 quotations, which were combined into 26, which were then filtered and combined into four themes including, 1) mobility behaviors, 2) mobility challenges, 3) mobility aids, and 4) AT-required features. Data management and analysis were carried out using the NVivo 12.0 software [21].

IV. STUDY FINDINGS

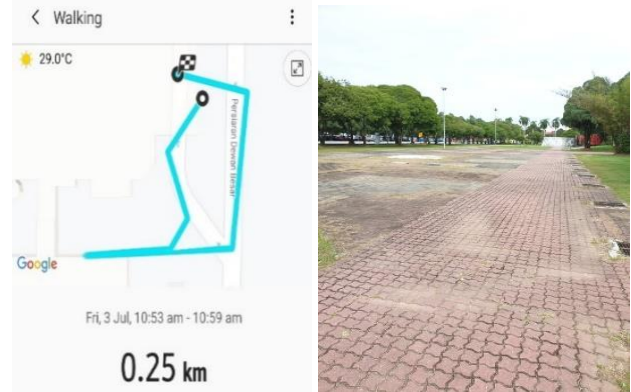
Several studies demonstrated the mobility behaviors, challenges, and solutions PVI desired to enhance their outdoor mobility. However, this study demonstrated further insights into mobility traits that were expected and unexpected. This section describes the outcomes from the perspectives of PVI mobility behaviors and challenges, as well as their ambitions and preferences of the future mobility AT to support their safe mobility.

A. Mobility Behaviors and Techniques

Spatial awareness is the PVI primary behavior in a familiar environment. In contrast, different techniques are performed to compensate for awareness in an unfamiliar environment.

1) *Cognitive map*: A cognitive map is a brain presentation of a spatial location [22]. It plays a main role in controlling the brain's navigation system [3]. Applying the cognitive map is the primary behavior associated with PVI mobility in a familiar environment, where PVI can visualize the surrounding area [23]. During the observation sessions in the familiar environment, all participants demonstrated high abilities to use

the cognitive map for real-time awareness. For instance, they recognized the accurate location of different objects like poles, and staircases. By generating relationships between objects' positions, the participants were able to consider objects such as doors, drains, and floor texture changes as landmarks. Fig. 3(a), presents a participant indicating a door as a landmark. All participants explained their mobility abilities and independence in a familiar environment. For example, participant P13 mentioned:



a) Unfamiliar Observation Path b) Sample of Unfamiliar Observation Site

Fig. 2. Sample from the unfamiliar observation site.

"Outside with a white cane and without any help, needs to feel free ... familiar place is easy."

Moreover, a few participants do not use a white cane in familiar areas. For instance, participant P9 emphasized:

"For a familiar place, I do not use the cane so much because I already have the cognitive map."

The cognitive map reflected on the participants' mobility with confidence, comfort, pride, and independence. Participant P12 reported that he engages in social services to assist blind people in familiar places. Additionally, for a question about the number of visits needed to familiarize a place, nine participants responded that, after three to four visits with a companion, they familiarize the place, but it is difficult to familiarize complex structure areas.

2) *Techniques*: This study collected about 22 techniques performed by the PVI to navigate safely in both familiar and unfamiliar environments. However, all participants expressed their disappointment, fear, and anxiety when visiting an unfamiliar environment due to bad experiences. The primary technique used by all participants when visiting an unfamiliar place is accompanying a sighted individual. Most participants (n = 10) narrated their bad experiences in unfamiliar environments which involved falling, getting lost, traffic, and immoral attitudes of strangers. According to participant P1:

"I don't like to fall... what I'll do if I'm lost? I will ask people, but you must be very careful asking people because people can even take you for a ride."

Participant P15:

"If unfamiliar (place), I think everybody does the same, go to an unfamiliar area with a person."

The second-most used technique when visiting an unfamiliar place is *information collecting* as posited by 13 participants. Participants obtained relevant information about the site before embarking on the trip. The information includes the distance from a specific location to the destination, landmarks in route and the area, useable public transportation, and area accessibility for the PVI. Participant P15 declared as follows:

"Normally I will ask my colleague and relatives. How does this place look like? So, they will tell me I try to imagine how it is."

As well participant P11 mentioned:

"I study the place before going. I get to know if I can take public transport and the accessibility of the place."

Different sources of information were used including sighted individuals and software applications (e.g., Waze and Apple/Google Maps). Before the unfamiliar observation sessions, several participants asked about the site features. Furthermore, participants engaged in slow walking and wide floor sweeping of the cane in an attempt to carefully recognize the area surrounding their bodies. Also, they kept talking to the researcher to feel comfortable and safe.

The participants protect their face and head using their freehand as additional safety techniques for familiar obstacles above the ground. For the staircases, they relied on the handrail and check each stair's height by the cane and their foot before they step which is applied also for pavements and curbs. They relied on tactile paving, trailing along a wall, a sidewalk border, or a road's edge if it is available to keep walking in a straight line unless it is not practicable [3]. Fig. 3(b) presents a participant trailing a sidewalk edge. To cross the road, they rely on hearing sense. The participants evaluated the potholes and drains by checking the depth, diameter, and drain covered by the cane and then avoids it by jumping above the small ones, or moving to the left or right. For an obstacle, they touch it with the cane in an attempt to recognize its features before addressing the issue. In public places, they ask sighted people for help.

3) *Alerting the awareness:* During the unfamiliar observation sessions, when the participants contacted a barrier via the cane, they stopped or slowed down to the maximum, swept the cane on the floor widely, and carefully searched for any change or obstacle. When the individual recognized the situation, they performed an effective and correct action to tackle it. This experience demonstrated the necessity of providing PVI with spatial information before they face a new situation. Hence, no need to inform individuals about what to do or how to address the challenge. However, individuals flinched, stumbled, and were prone to fall when their body contacted with an obstacle before the cane. Moreover, most participants expressed their emotions according to the situation, such as anxiety, comfortable, fear, stress, and pride. P6 mentioned:

"No, please no holes. Holes, especially holes, it is very difficult."

Accordingly, AT developers should avoid informing the user.

What to do (e.g., "turn left, go right"), rather informing them about the obstacle within an appropriate time should be emphasized. Table II summarizes the mobility behaviors and techniques that PVI perform in the familiar and unfamiliar environments to protect their safety.

B. Mobility Challenges

Several challenges were experienced by the PVI during mobility. This research demonstrates the eight most critical challenges reported by the participants encountered during the interview and observation sessions. All participants explicitly asserted that low awareness of *public behavior* is a severe defect that makes mobility an extreme challenge for PVI. They clarified that sighted people stop on the tactile paving for chatting, park their motorbike on it, and compete to walk on the tactile blocks. Participant P2 expressed:



a) A participant indicates a landmark in a familiar area



b) A participant trailing the sidewalk border in an unfamiliar area

Fig. 3. Sample of mobility behaviors.

"Normal people don't care. We need a minimum level of awareness. I'm not suffering from my blind status; I'm suffering the attitude of people every day and it is stressful."

Additionally, the presence of different *terrain types* was another significant challenge. All participants reported that potholes, drains, and stairs form a hazard that could directly impact their safety. The participants shared either their personal experiences or friends' stories about falling in potholes and drains. Fig. 4 presents participants while they were addressing different terrain types in the unfamiliar site. Participant P8 stated:

"The worst thing is falling down in the drain, fall in a drain or a hole."

Participants' anxiety and fears were expressed when approaching various terrain types even in the familiar environment, especially potholes that could suddenly appear as a result of weather or constructions. Participants mentioned several other challenges including traffic and crossing the road. Participant P12 expressed,

"Crossing the road is a big problem ..., crossing the road for the blind is very dangerous."

The participants lose their direction when they find themselves in a wide area. Crowded places, on the other hand, make them feel uncomfortable and embarrassed, especially if they contact pedestrians with their canes or bodies. Similar feelings were expressed when they approached inaccessible pavements that do not follow the height and width standardization or without tactile paving.

Participant P7 sadly mentioned:

"I fell before... I didn't know the curb was so high."

TABLE II. SUMMARY OF MOBILITY BEHAVIORS AND TECHNIQUES PRACTICED DAILY BY THE PVI

No	Behaviors	Familiar	Unfamiliar	Actions	Emotions
1.	Utilize the conventional cane	√	√	Contacting the space in front of them to recognize the situation	Confidence
2.	Utilize the cognitive map	√		Use environmental objects for real-time awareness	Confidence, pride, anxiety from any changes
3.	Landmarks	√	√	Use all senses, improve the cognitive map	Happiness
4.	Accompanying		√	Accompany a sighted individual	Sadness, embarrassed, confidence
5.	Go with a group	√	√	Group visit to a place	Happiness, confidence
6.	Information collecting		√	Ask relatives, use apps.	Confidence, anxiety
7.	Slow walking		√	Slow down and be careful	Anxiety, careful
8.	Wide floor-sweeping of the cane		√	To recognize the area surrounding their bodies	Anxiety, careful
9.	Talking to the companion	√	√	Keep chatting with the companion individual	Confidence
10.	Recognize a change	√	√	Stop and scan the cane on the floor widely	Fears, Anxiety
11.	Face obstacles on the head level	√		Raise the freehand to the head or face for protecting	Anxiety
12.	Utilize curb and pavements	√	√	Utilize foot to determine the height and cane for the width	Careful
13.	Pavement ramp	√	√	Looking for, to utilize it	Confidence
14.	Utilize tactile paving	√	√	Looking for tactile and walk on it	Confidence, anxiety from any obstacle on it
15.	Ascending/descending staircases	√	√	Use foot for the height, rely on handrailing if exist unless use the cane	Anxiety, careful
16.	Walk in a straight line	√	√	Tactile paving, keep the cane trail on the sidewalk edge, wall, or road boundary	Happiness
17.	Cross the road	√	√	Hearing sense, raise the cane upper than head level.	Fears, stress
18.	Presence of pothole or drain	√	√	Use cane to check the depth, diameter, and drain's cover then avoid	Fears, anxiety, careful
19.	Contact obstacle	√	√	Touch by the cane or hand to recognize then avoid	Careful, anxiety
20.	Feel lost		√	Ask people, or applications	Fears, stress, sadness
21.	Stop and raise the cane to public	√	√	Wait for someone to offer assistance	Embarrassed, anxiety
22.	Alert awareness about a new situation	√	√	Stop or slow down, seek for the situation by the cane, recognize the case, tackle it.	Fears, anxiety, careful



Fig. 4. Participants addressing various terrain types in an unfamiliar area.

Participants were also exposed to obstacles above the ground level such as signboards, open windows, and tree branches

Another critical challenge expressed by participants was public transportation without audio sign. The challenges faced by the PVI are summarized in Table III.

C. Mobility Aids

This section clarifies the participants' opinions and reliability of the white cane and mobility assistive technologies.

TABLE III. SUMMARY OF MOBILITY CHALLENGES EXPERIENCED BY THE PVI

No.	Challenge	Familiar	Unfamiliar	Actions	Emotions
1.	Public Attitude	√	√	Stop on the tactile paving for chatting. Park motorbike on it. Walk on the tactile blocks. Ignore PVI	Sadness, anger, stress
2.	Terrain types	√	√	Presence of Pothole, drain, stairs Lack of standard stairs, curbs and pavements Lack of handrailing of staircase	Fears, sadness, anxiety
3.	Traffic	√	√	Cross the road	Fears, anxiety
4.	Wide area	√	√	Lost direction, ask people	anxiety
5.	Strangers	√	√	Ask strangers for help,	Fears, anxiety
6.	Inaccessible pavement	√	√	Pavement without tactile block. Not standard height and width Obstacles, rubbish, motorbike	Sadness, anxiety
7.	Obstacles above the ground level	√	√	Strike upper body parts like chest, face, and head. Signboards, tree branches, open windows, hanging obstacles.	Fears, sadness, anxiety
8.	Public transportation	√	√	Visual signs, gap between the train and its platform	Fears, sadness, anxiety

Participant P15 rated safe mobility using the white cane to range from 70 to 75%. This is not surprising as several researchers verified the limitations of the white cane [24], [3].

2) *Assistive technologies*: In terms of the reliability of the ATs for PVI, all participants responded that they do not rely on any existing AT from either electronic devices or apps. Several researchers [1], [25] reported that PVI trust the white cane because of its cost-effectiveness and provision of better performance compared to smart canes.

None of the 15 participants utilized any mobile electronic device in their daily life. They justified that absence of usage of ATs by expensive, have low performance, are heavier compared to the white cane, and cause anxiety about the battery draining.

Regarding the use of mobile apps, 13 participants owned smartphones and use social media but none of them relied on

1) *White cane*: The 15 participants in this study emphasized that they rely only on the white cane for mobility. Furthermore, they affirmed that white cane enables them to avoid obstacles, feel safe, and inform the public that they have a vision impairment. Participant P2 said:

"If there is no cane, I feel something is missing; I feel I am lost."

However, 11 participants argued that the white cane also has its limitations and it does not provide 100% safe mobility. Some of the limitations that participants shared include lacking the ability to recognize the obstacle, detecting small terrain types, and detecting obstacles above the ground level. Therefore, the participants expressed a desire to have additional aid to overcome the limitations. Participant P7 remarked,

"It is quite safe. I mean, if you're in a familiar place, it's quite safe but in an unfamiliar place, maybe a bit risky."

apps during mobility. Nevertheless, six participants employed Maps apps, Waze [26], Lazarillo [27], and Moovit [28] to collect data before the trip.

Although some of the participants do not rely on mobile Apps, they opined that apps could satisfy some of their desired features. For instance, App can function as complementary with the white cane. Since the users already own a smartphone, the App can be installed and eliminates the need to purchase a new device, which making App an affordable option. Additionally, apps/smartphones do not attract unwanted attention and could be used for multimodal feedback. Apps are usually free or may incur a low cost compared to gadget ATs. However, the participants expressed their reasons for not using Apps during navigation which included the hazard of holding a smartphone when walking, the importance of having a free hand for safe mobility, and the need for an internet connection.

D. Perception of PVI for New AT Features

The appropriate mobility AT for PVI depend on the user's personality and activities [29]. However, the information gathered in this study included the most obvious participants' preferences according to AT features that may address their requirements and gain acceptance from the perspective of design, feedback, and distance to reach the obstacle.

Several designs were selected by the participants in terms of the AT *features*. All participants preferred a solution that is less obvious and does not attract pedestrian attention. This is consistent with the findings of [30], [10] that emphasized the rejection of any design that advertises the disability and attracts unwanted attention. Participant P1 reported;

"I do not want something obvious. It shouldn't make it as an alien."

Eight participants selected the smart cane as the preferred design, whereas two participants preferred wearable devices, such as eyeglasses or a jacket in addition with the conventional cane. Two participants selected a mobile application solution, but again accompanied by the conventional cane. In contrast, the remaining three participants reported that they do not trust any type of AT and preferred to stick with the conventional white cane, indicating a resistance to change.

According to the *feedback*, the study emphasized that it should be conveyed in real-time and precise in terms of the obstacle name and distance (e.g., pothole, 2 meters). The majority of respondents (n = 14) stated that the speech message is the best feedback type because it provides a good perception and enables them to recognize their environment. Additionally, three participants mentioned the importance of vibration especially in a noisy environment. Each feedback style has its advantages and disadvantages. Audio provides rich interaction spatial information that lets the user visualize the environment [31]. Which can improve the cognitive map also [23] but it might be missed in a noisy environment [14]. In contrast, haptic feedback provides privacy, however, the spatial information is limited [32]. Thus, this study agrees with some earlier reports discussing the functionality of multimodality feedback, (i.e., speech message and vibration). Multimodality provides better interaction modes and improves the proposed solution durability, flexibility, and accessibility for a wider range of acceptance [31], [33].

Information was also gathered in this study about the required *coverage distance* to reach the obstacle. Surprisingly, most participants were less informed about the distance units and unable to recognize how the distance is measured in meters following the usage of their arms to specify a given distance. Seven participants mentioned that it should neither be too close nor too far. They added that they needed enough time to prepare themselves for the new situation without losing it. The responses were between 0.5 to 4 meters with a mean of 2.25 meters, which approximately validates the coverage distance reported in various studies ranging from 2 to 4 meters, [2], [34]. The participants' preferences for the main features of the future AT are summarized in Table IV.

V. DISCUSSION AND RECOMMENDATIONS

This study was conducted to provide comprehensive information about mobility context that developers need to consider when developing new mobility ATs for the PVI. A qualitative study involved 15 participants with vision impairment, collected information about outdoor mobility, behaviors, challenges, and the preferred AT features.

An interesting finding demonstrated that PVI who either had solid mobility experience were able to safely address various mobility situations via the white cane only. However, the mobility hazards faced by the PVI were primarily from a lack of spatial information about the environment in front of them. For instance, the cane may miss between its up-down or left-right movement when approaching potholes, short stairs, as well as obstacles above the ground level such as tree branches. These events reflect the increased risk of PVI to various hazards resulting from a lack of spatial information and awareness.

Another significant findings is that PVI were unable to navigate without a cane, similar outcomes have been reported in previous studies [29], [35], [1]. Based on these findings, this study highlights the necessity of an AT that works as a complementary to the cane and address its limitations. A previous study also presented a similar complementary function between the conventional cane with wayfinding apps [19].

In terms of mobility behaviors, the study highlights the role of the cognitive map in performing safe and independent mobility in a familiar environment. The superiority of employing the cognitive map was evident when comparing mobility tasks in familiar and unfamiliar environments. This study encourages AT developers to prioritize the improvement of the cognitive map as a mobility aid, besides the most current research that solely focuses on addressing instant mobility.

In contrast, PVI primarily relied on accompanying a sighted individual when they visit an unfamiliar environment. However, they also depended on various techniques such as gathering information about the route and the area they intent to visit and trailing along a wall or a sidewalk border, which were also reported in a previous study [36]. Likewise, the PVI stopped when they feel a change of the ground surface until they recognize the situation, and they walk slowly and sweep the cane widely. The study suggests that these mobility behaviors need to be considered by the AT developers when designing future ATs. For instance, users could be directed to the nearest wall or sidewalk edge for it to be utilized for straight-line walking.

An unexpected challenge emerging from this study was the frequent suffering of PVI from the attitude of a large number of sighted individuals who compete with them to utilize infrastructures specially designed for PVI, including tactile blocks. Additionally, other challenges were reported by participants including, the presence of different types of terrain, crossing the road, navigating in wide areas, and obstacles above the ground level, inaccessible pavements in terms of height, width, and lack of tactile paving, and using public transportation. This indicates the need for ATs to tackle these

barriers, especially from the aspects of detection performance, usability, and cost.

In terms of AT design preferences, again the significance of the cane was obvious, where most participants preferred utilizing the conventional cane with another assistance (i.e., app or wearable device). This leads to recommend the developers to adopt the complementary function between AT

and the conventional cane, thereby eliminating the idea of substituting the cane with another AT.

For the feedback, 93.3% of participants prefer speech messages, however audio could be missed in a noisy environment. Therefore, the study suggests providing multimodal feedback (haptic and audio), which improves flexibility, usability and enhances the acceptance of the solution.

TABLE IV. SUMMARY OF THE PREFERENCES OF FUTURE AT FEATURES

Feature	Preference	Advantage	Disadvantage
Design	Smart cane	Still utilize a cane	Heavyweight compared to the white cane, costly, anxiety for the battery draining
	Mobile app	No more device, installed in their smartphones	Using smartphone during the navigation
	Not obvious, small size	Do not attract public attention, can be hanged, keep one hand free	
	Affordable	Able to purchase and try even they do not trust technology	
	Lightweight	Comfortable as the white cane	
Feedback	Precise amount	Simple, specific, and direct information	
	Speech message	Rich information to recognize the situation, improve the cognitive map	Loss in a noisy environment, block surrounding sounds
	Vibration	Can be used in noisy areas, provides privacy	Limited information, confusing
Coverage Distance	Around 2.5 meters	Enough time to prepare themselves for the new situation	

According to the coverage distance to convey the feedback, the findings showed that 2 to 3 meters is sufficient distance for the users to recognize the change and prepare themselves for the new situation.

VI. CONCLUSION AND FUTURE WORK

The purpose of the current research is to provide comprehension of the outdoor mobility behaviors, challenges, and the target group's perceptions of the solution. The research confirms previous findings and contributes to our understanding of the new mobility context.

The study answered the research questions by elaborating on the behaviors and challenges of PVI mobility, in addition to clarifying the ATs features that fulfill the PVI expectations. The findings demonstrated that PVI are able to navigate safely when they are informed about the environmental situation within a sufficient time. Also, the white cane is a primary mobility tool for PVI and any other assistant could be used as a complementary. Hence, it is suggested that the designers develop mobility ATs that work with the white cane and not substitute it. The results also emerge lack of spatial information and lack of public awareness as significant factors that increase mobility difficulty. The study reported perception of the new AT requirements. The results revealed the preference of the cane design, multimodal feedback, and two meters coverage distance to reach the obstacle. For future work, the authors intend to propose a prototype of AT to support outdoor mobility systems and meet the participants' requirements and perceptions based on the findings in this study.

ACKNOWLEDGMENT

We thank all members and instructors of MAB – KL and Ipoh branches for their assistance which facilitated this research. Also, we thank all participants who dedicated their time and effort to contribute to this study. Finally, we would like to thank the Human-Computer Interaction Research Group, Faculty Comp Science & Information Tech, University Putra Malaysia.

REFERENCES

- [1] M. A. Hersh and A. R. García Ramírez, "Evaluation of the electronic long cane: improving mobility in urban environments," *Behav. Inf. Technol.*, vol. 37, no. 12, pp. 1203–1223, 2018.
- [2] K. Yang et al., "Unifying Terrain Awareness for the Visually Impaired through Real-Time Semantic Segmentation," *Sensors*, vol. 18, no. 1506, pp. 1–32, 2018.
- [3] E. Pissaloux and R. Velázquez, *Mobility of Visually Impaired People*. Gewerbestrasse, Switzerland: Springer International Publishing, 2017.
- [4] M. M. Islam, M. S. Sadi, K. Z. Zamli, and M. M. Ahmed, "Developing Walking Assistants for Visually Impaired People: A Review," *IEEE Sens. J.*, vol. 19, no. 8, pp. 2814–2828, 2019.
- [5] R. Tapu, B. Mocanu, and T. Zaharia, "Wearable assistive devices for visually impaired: A state of the art survey," *Pattern Recognit. Lett.*, no. xxxx, pp. 1–16, 2018.
- [6] J. Guerreiro, D. Sato, D. Ahmetovic, E. Ohn-bar, K. M. Kitani, and C. Asakawa, "Virtual navigation for blind people: Transferring route knowledge to the real-world," *J. Hum. Comput. Stud.*, vol. 135, no. 102369, 2020.
- [7] C. Ntakolia, G. Dimas, and D. K. Iakovidis, "User - centered system design for assisted navigation of visually impaired individuals in outdoor cultural environments," *Univ. Access Inf. Soc.*, no. 0123456789, 2020.

- [8] A. Chai, B. Chun, L. B. Theng, L. Deverell, A. A. L. Mahmud, and C. McCarthy, "An Autonomous LiDAR Based Ground Plane Hazards Detector for the Visually Impaired," *IEEE-EMBS Conf. Biomed. Eng. Sci.*, pp. 346–351, 2018.
- [9] M. Cornacchia, B. Kakillioglu, Y. Zheng, and S. Velipasalar, "Deep Learning-Based Obstacle Detection and Classification with Portable Uncalibrated Patterned Light," *IEEE Sens. J.*, vol. 18, no. 20, pp. 8416–8425, 2018.
- [10] P. Chanana, R. Paul, M. Balakrishnan, and P. Rao, "Assistive technology solutions for aiding travel of pedestrians with visual impairment," *J. Rehabil. Assist. Technol. Eng.*, vol. 4, pp. 1–16, 2017.
- [11] A. Khan and S. Khushro, "An insight into smartphone-based assistive solutions for visually impaired and blind people: issues, challenges and opportunities." Springer Berlin Heidelberg, 2020.
- [12] J. Lazar, J. H. Feng, and H. Hochheiser, *Research Methods in Human-Computer Interaction*, 2nd Ed. Cambridge, MA, USA: Elsevier, 2017.
- [13] R. Jafri and M. M. Khan, "User-centered design of a depth data based obstacle detection and avoidance system for the visually impaired," *Human-centric Comput. Inf. Sci.*, vol. 8, no. 1, 2018.
- [14] S. Varghese Jacob and I. S. MacKenzie, "Comparison of feedback modes for the visually impaired: Vibration vs. audio," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 10907 LNCS, pp. 420–432, 2018.
- [15] J. A. Rey-Galindo, L. Rizo-Corona, E. L. González-Muñoz, and C. Aceves-González, "Environmental information for people with visual impairment in Mexico - or what they need and how they use it," *Appl. Ergon.*, vol. 85, no. 103079, 2020.
- [16] V. Kameswaran, A. J. Fiannaca, M. Kneisel, A. Karlson, E. Cutrell, and M. R. Morris, "Understanding In-Situ Use of Commonly Available Navigation Technologies by People with Visual Impairments," in *The 22nd International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '20)*, 2020, pp. 1–12.
- [17] M. Hersh and M. Johnson, *Assistive Technology for Visually Impaired and Blind People*. London: Springer, 2008.
- [18] J. Xu, S. M. Billah, A. Balasubramanian, and R. Shilkrot, "DarkReader: Bridging the gap between perception and reality of power consumption in smartphones for blind users," *ASSETS 2019 - 21st Int. ACM SIGACCESS Conf. Comput. Access.*, pp. 96–104, 2019.
- [19] V. Braun and V. Clarke, "Using thematic analysis in psychology," *Qual. Res. Psychol.*, vol. 3, no. 2, pp. 77–101, 2006.
- [20] L. S. Nowell, J. M. Norris, D. E. White, and N. J. Moules, "Thematic Analysis: Striving to Meet the Trustworthiness Criteria," *Int. J. Qual. Methods*, vol. 16, no. 1, pp. 1–13, 2017.
- [21] "NVIVO." [Online]. Available: <https://www.qsrinternational.com/nvivo-qualitative-data-analysis-software/home>. [Accessed: 10-Apr-2021].
- [22] E. C. Tolman, "Cognitive maps in rats and men," *Image Environ. Cogn. Mapp. Spat. Behav.*, pp. 27–50, 1948.
- [23] N. Banovic, R. L. Franz, K. N. Truong, J. Mankoff, and A. K. Dey, "Uncovering information needs for independent spatial learning for users who are visually impaired," in *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility*, 2013, pp. 1–8.
- [24] J. K. Tan, T. Ishimine, and S. Arimasu, "Walk environment analysis using my vision: Toward a navigation system providing visual assistance," *Int. J. Innov. Comput. Inf. Control*, vol. 15, no. 3, pp. 861–871, 2019.
- [25] A. D. P. Santos, F. O. Medola, M. J. Cinelli, A. R. Garcia Ramirez, and F. E. Sandnes, "Are electronic white canes better than traditional canes? A comparative study with blind and blindfolded participants," *Univers. Access Inf. Soc.*, pp. 93–103, 2021.
- [26] "Waze App." [Online]. Available: <https://www.waze.com/>. [Accessed: 28-Jul-2021].
- [27] "Lazarillo App." [Online]. Available: <https://lazarillo.app/>. [Accessed: 28-Jul-2021].
- [28] "Moovit App." [Online]. Available: <https://moovit.com>. [Accessed: 28-Jul-2021].
- [29] M. A. Williams, A. Hurst, and S. K. Kane, "' Pray Before You Step Out ': Describing Personal and Situational Blind Navigation Behaviors," in *15th International ACM SIGACCESS Conference on Computers and Accessibility*, 2013, pp. 1–8.
- [30] D. J. Calder, "Ecological solutions for the blind," *4th IEEE Int. Conf. Digit. Ecosyst. Technol. - Conf. Proc. IEEE-DEST 2010, DEST 2010*, pp. 625–630, 2010.
- [31] B. Kuriakose, R. Shrestha, and F. E. Sandnes, "Tools and Technologies for Blind and Visually Impaired Navigation Support: A Review," *IETE Tech. Rev. (Institution Electron. Telecommun. Eng. India)*, vol. 39, no. 1, pp. 3–18, 2020.
- [32] B. Kuriakose, R. Shrestha, and F. E. Sandnes, "Multimodal navigation systems for users with visual impairments—a review and analysis," *Multimodal Technol. Interact.*, vol. 4, no. 4, pp. 1–19, 2020.
- [33] D. Bohus and E. Horvitz, "Facilitating multiparty dialog with gaze, gesture, and speech," *Int. Conf. Multimodal Interfaces Work. Mach. Learn. Multimodal Interact. ICMI-MLMI 2010*, no. September, pp. 1–8, 2010.
- [34] W. M. Elmannai and K. M. Elleithy, "A Highly Accurate and Reliable Data Fusion Framework for Guiding the Visually Impaired," *IEEE Access*, vol. 6, pp. 33029–33054, 2018.
- [35] W. P. Uzan G, Hanse P-C, Seck M, "Solid: a model of the principles, processes and information required to ensure mobility for all in public transport systems," in *Proceedings 19th triennial congress of the IEA*, 2015.
- [36] M. A. Williams, C. Galbraith, S. K. Kane, and A. Hurst, "' Just Let the Cane Hit It ': How the Blind and Sighted See Navigation Differently," in *16th international ACM SIGACCESS conference on Computers & accessibility*, 2014, pp. 217–224.

A Deep Learning based Approach for Recognition of Arabic Sign Language Letters

Boutaina Hdioud¹, Mohammed El Haj Tirari²

Department of Computer Science, ENSIAS, Mohammed V University, Rabat, Morocco¹

Department of Statistic, INSEA, Rabat, Morocco²

Abstract—No one can deny that the deaf-mute community has communication problems in daily life. Advances in artificial intelligence over the past few years have broken through this communication barrier. The principal objective of this work is creating an Arabic Sign Language Recognition system (ArSLR) for recognizing Arabic letters. The ArSLR system is developed using our image pre-processing method to extract the exact position of the hand and we proposed architecture of the Deep Convolutional Neural Network (CNN) using depth data. The goal is to make it easier for people who have hearing problems to interact with normal people. Based on user input, our method will detect and recognize hand-sign letters of the Arabic alphabet automatically. The suggested model is anticipated to deliver encouraging results in the recognition of Arabic sign language with an accuracy score of 97,07%. We conducted a comparison study in order to evaluate proposed system, the obtained results demonstrated that this method is able to recognize static signs with greater accuracy than the accuracy obtained by similar other studies on the same dataset used.

Keywords—Deep learning; hand landmark model; convolutional neural network; Arabic sign language recognition

I. INTRODUCTION

According to the statistical information issued by the World Health Organization in 2021, people with hearing disabilities and deaf persons represent more than 5% of people on the planet [1]. Additionally, it is predicted that in the following 30 years, this number will double. As a result, governments and researchers therefore place a high priority on helping those individuals to participate in their societies. Nowadays, these individuals are the primary users of sign language recognition typically used to communicate both inside and outside of their community.

In this context, we can understand the importance of sign language as a form of non-verbal communication that based on signs and gestures. So, Sign Language Recognition systems (SLR) have been given attention recently. Despite the fact that movement expressions in sign language are the most structured [2], the difficulty of sign language recognition lies in the way that it is expressed. There are two main ways to represent words in sign language; these two ways can be considered complementary to one another: The first way uses particular body movements (e.g. human hands and arms to convey meanings) [3] and facial expressions (e.g. eyebrow raising and mouth shaping) [4]-[6], whereas the second way uses a fingerspelling approach (e.g. orientation, hand posing, and trajectory) [7]. Unfortunately, there is no universal sign language. It's typically different from one country to another

and from one language to another, including the level of the alphabet, where each letter has its own conventional sign. But, with the recent advance in the domain of sign recognition, the researchers have prompted the creation of robust sign language recognition systems for various sign languages, for instance, Brazilian Sign Language [8], British Sign Language [9], Chinese Sign Language [10] and the American Sign Language [11]. However, Asian, English, and Latin sign languages have all been the subject of substantial research, whereas Arabic has received relatively little attention. This can be due to the Arabic language's complexity or the absence of a widely used dataset for the Arabic sign language available to academics. As a result, researchers were forced to build their own datasets, which is a tedious task.

It is crucial to create systems that can translate sign languages into text or speech to help people, who are not deaf or mute, to communicate. Typically, to create any sign language recognition system we can use an image-based approach or a sensor-based technique [12]. Specifically, sensor-based systems work by connecting a glove to a number of sensors to read the gesture, which the system can then recognize; such solutions suffer from a lack of usability. On the other hand, image-based solutions have alleviated this issue and offered a solution in which signs are translated using the cameras. Ideally, those systems can reduce the need for human intervention in the exchange of information between normal people and deaf [13]. Image-based solutions relying on image processing can be carried out in two stages: detection and classification. Each image is first pre-processed during the detection phase, after which the regions of interest are found. The classification process can then be carried out using the results of the preceding procedure. During the recognition stage, each segmented hand sign has a set of features that are extracted in order to perform in the recognition process. Therefore, it is possible to understand the differences between the different signs by using these features as a guide.

Despite extensive research in this field, and as indicated in [14], each study has limitations in terms of the image processing, cost, and sign classification. In general, the system performance depend on the accuracy of the image pre-processing stage, which separates the hand region from the full image, therefore, the main focus of this work is on the recognition of Arabic Sign Language (ArSL) alphabets based on fingerspelling [15]. In order to improve the images of hand gestures accuracy, we proposed to implement a novel ArSL recognition system based on deep convolutional neural network, which incorporates a novel pre-processing stage that

can detect sign gestures by feeding it images of hand gestures performed in varied lighting and orientation, this stage can also alleviate in a small way the problem of strong similarities within the sign language alphabet that present highly similar visual properties, for example, the letter pairs DELL/DHELL, RAA/ZAY, Ayn/Ghayn, ...

Using a dataset of 28 classes of different images of Arabic signs, the objective of this research is to demonstrate how the pre-processing images can be aided in the representation of gestures in feature extraction and how proposed architecture can offer improved accuracy in comparison to other existing methodologies. This work has the following contributions:

- Create a new database of 5000 RGB images and combining them with existing Arabic Sign Language ArSL dataset to input into proposed model and use them as training data.
- Integrate proposed pre-processing stage to enhance the representation gestures.
- Training and comparison of others' works as well as the interpretation of static Arabic sign letters using proposed architecture.
- Analysis of performance of the proposed model in terms of Arabic sign language interpretation.

The rest of the paper is divided into the following sections. An overview of the connected works is given in Section II. The proposed methodology for recognizing hand signs is presented in Section III. The experimental results are described in Section IV. Finally, conclusions and perspectives are presented in Section V.

II. RELATED WORK

Recently, several research papers focused on ArSLR systems. For recognizing the ArSL with gestures, there are three partitions of the method: recognizing the Arabic alphabet, recognizing individual words, and recognizing entire sentences. In this research we focused on the recognizing of the Arabic alphabet. In this context, we provide a summary of some previously employed methods in this field. Automatic Sign Language Recognition is a field to replace sign language interpreters. There have been many studies performed in this way, and several technologies have been built, such as pre-processing to extract hand gesture, feature extraction to reduce an input data into relevant features, and classification to identify the class of each hand sign. In [16] the authors have summarized the work carried out previously, which focuses on making a difference between dynamic sign language and isolated, static, alphabetic non-Arabic and Arabic sign languages, as well as classification methods used in recognition and that relied on traditional machine learning methods or deep learning methods.

An automatic system for recognizing numbers from 0 to 10 and Arabic letters was implemented in [17]. The authors used a dataset with 7869 images overall. Seven layers composed the suggested model, which was trained repeatedly using various training-testing configurations. Finally, the authors offered a comparison with various strategies based on

KNN and SVM, demonstrating the benefit of the suggested model. An automated method for translating Arabic sign language was suggested by the authors in [18]. This system relies on the building of two datasets for Arabic alphabet gestures. In order to extract Arabic sign gestures from images or videos, this system suggests a hand coverage-based manual detection approach. It then uses a range of statistical classifiers, compares the results, and produces a more accurate classification.

With the use of Microsoft Kinect, Hisham et al. [19] introduced a dynamic Arabic sign language recognition. Two machine learning methods, Bayesian Network and Decision Tree are used for recognition, and the Ada-Boosting methodology is then used to improve system recognition. In [20], the authors used deep transfer learning for developing a robust recognition system for Arabic sign language. To prevent overfitting and improve overall system performance, they used data-augmentation. For the target recognition task, several network architectures have been studied. The Arabic sign language (ArSL2018) public dataset was used in this experiment. Other researchers, such as Saleh et al. [21], enhanced the accuracy of Arabic sign language hand gesture recognition through the use of deep CNN and transfer learning. A novel ArSLR system was proposed in [22] to recognize and classify Arabic alphabetical letters, which uses microscopic images along with an unsupervised deep learning algorithm built on a deep belief network (DBN). In [23], authors developed a system that automatically recognizes 28 letters in Arabic Sign Language using a CNN model with a grayscale image as input. In [24], the authors applied ontology to the sign language domain in order to address some sign language problems. They used simple static signs composed of Arabic alphabets. An architecture of CNN was trained and evaluated using a dataset that was collected and a pre-made dataset for Arabic sign language. A new framework for the automatic recognition of Arabic sign language was proposed by Duwairi et al. [25] and is based on transfer learning applied via popular deep learning models (AlexNet, VGGNet, and Inception Net) for image processing. They proposed using VGGNet architecture, which performed better than previous trained models for automatically recognizing Arabic alphabets in sign language. The authors in [26] suggested a system that can be used by the impaired people. The proposed system converts hand gestures in Arabic sign language into vocalized speech. They used Deep Convolutional Network to extract features from the data gathered by the sensor devices and they employed DG5-V hand gloves to capture the hand movements in the dataset. Finally, they applied CNN method for classification. Furthermore, recent advances in sign language have produced a number of models that are suitable for a variety of tasks; however neither of them actually possesses the necessary generalization capability.

III. PROPOSED METHODOLOGY

Using two-dimensional images, we seek to recognize the hand gestures of the ArSL and convert them into alphabet in LSA language. The main objective of this research is to build a deep CNN capable of accurately recognizing the ArSL alphabets. An overall view of the representation of the system is presented in Fig. 1.

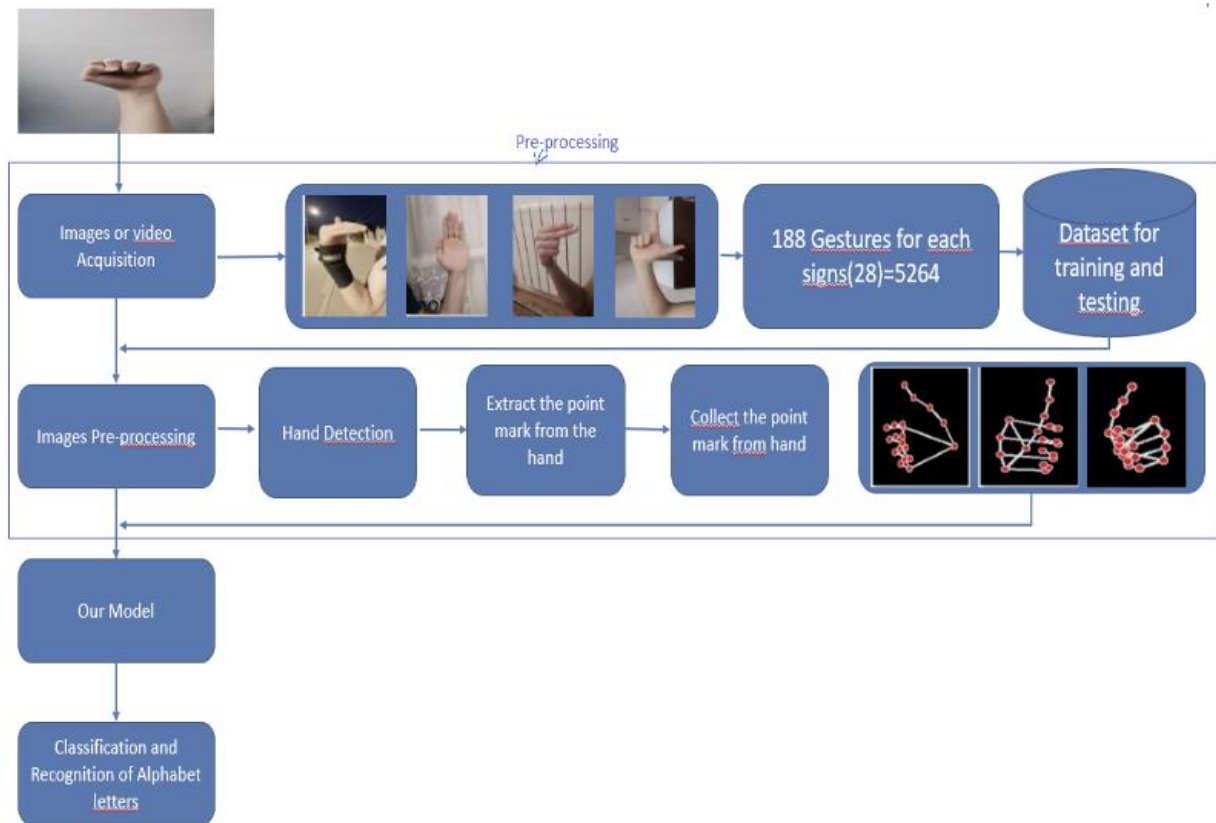


Fig. 1. Overview of the proposed method.

The general architecture of the proposed system comprises two stages: image pre-processing and a proposed model for recognizing Arabic sign language. In the following section we will cover different stages.

A. Proposed Pre-processing

After the Arabic sign language images are captured by the camera, they are subjected to a number of main processes during the pre-processing stage. The data are images depth that contain a hand. In order to extract the exact position of the hand we used the mediapipe framework as a solution to detect the hand [27]. The latter has a number of techniques for detection and tracking. Mediapipe Hand is one of its techniques. It consists of two different models working together, the first is namely Palm Detection Model which uses the whole image, produces an oriented hand bounding box made up of rigid objects (e.g. palms and fists) this model has the ability to detect occluded and self-occluded hand, and the second is namely Hand Landmark Model which produces high-fidelity 3D hand keypoints using the cropped image area defined by the palm detector. After the localization phase we cut the image by transforming it into black images that contain just the hand landmarks with a line connecting them after we resize the images to 240×240 (Fig. 2.). Finally, we divided our dataset into three parts- training, validation and test for training our model. Fig. 2 shows the input image Fig. 2(a) and the outcomes of the image pre-processing with edge detection in Fig. 2(b).

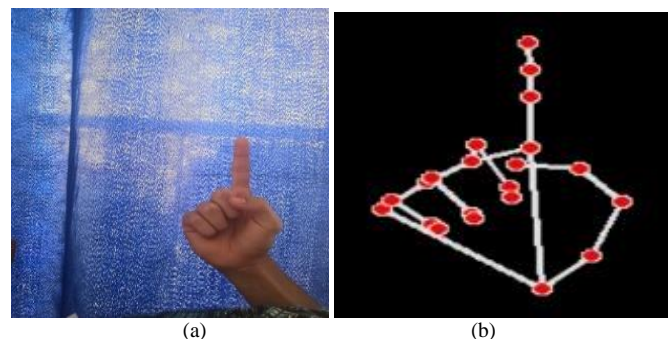


Fig. 2. Input image (a) Before preprocessing and Sample of characters extracted (b) After preprocessing.

B. Proposed CNN Architecture

The convolutional neural network CNN model we used in this study is composed of several layers. Fig. 3. shows the proposed architecture of the CNN, which consists of an input layer that accepts images with a size of $240 \times 240 \times 3$ pixels, which corresponds to the size of the images that the system accepts.

The feature extraction section consists of three convolutional layers (Conv2d, Conv2d_1, Conv2d_2) in which each convolution filter has dimensions of 3 by 3. There are thus 16 filters for Conv2d, 32 filters for Conv2d_1, 64 filters for Conv2d_2.

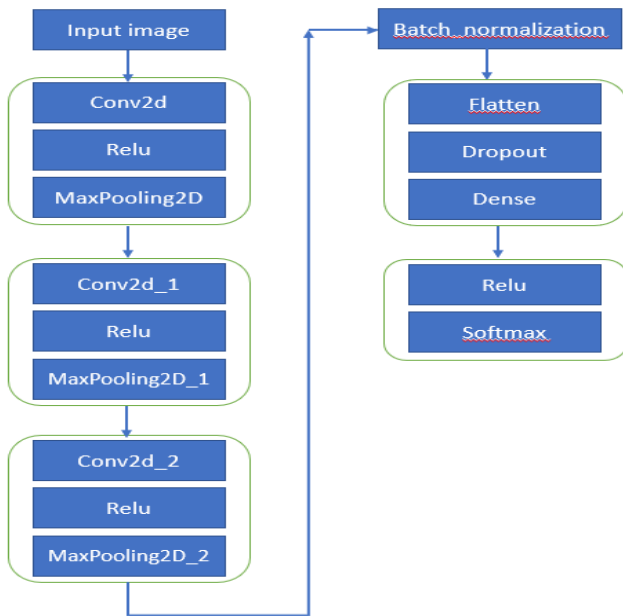


Fig. 3. Proposed CNN network.

Note here that each convolution operation is followed by an activation layer which uses the corrected linear unit function (ReLU). After each activation of the neurons, we will use a Maxpooling layer with a size of 3x3 in an attempt to reduce the size of the images without modifying their important characteristics. This will help to significantly reduce the computational power required by the model. Next, we will use a flatten layer followed by a dropout layer to disable 25% of the neurons. This type of layer helps to reduce overlearning by randomly disabling neurons in the network. Finally, we will use an activation layer using the RELU function followed by a dense classification layer using the softmax activation function. An overview of the parameters in each layer and the overall parameters in the proposed network are shown in Table I.

TABLE I. SUMMARY OF THE PROPOSED CNN NETWORK

Network Layer	Output Shape	Parameters
conv2d(Conv2D)	(None, 240, 240, 16)	448
Max_pooling2d(MaxPooling2D)	(None, 80, 80, 16)	0
conv2d_1(Conv2D)	(None, 80, 80, 32)	4640
max_pooling2d_1(MaxPooling2D)	(None,26,26,32)	0
conv2d_2(Conv2D)	(None, 26, 26, 64)	18496
max_pooling2d_2(MaxPooling2D)	(None, 8, 8, 64)	0
batch_normalization(BatchNormalization)	(None, 8, 8, 64)	256
flatten(Flatten)	(None, 4096)	0
dropout(Dropout)	(None, 4096)	0
dense(Dense)	(None, 256)	1048832
Dense_1(Dense)	(None, 28)	7196
Total parameters	1,079,868	

IV. EXPERIMENT AND RESULTS

A. Dataset

To evaluate proposed system, we create a combination of the Arabic Sign Language ArSL dataset [28] and our dataset created by professional camera from different signers and different luminosity intensities.

The Arabic Sign Language ArSL dataset is mainly composed of two folders, the first one containing 1160 images each having a size of 416x416 pixels, and the second one containing 1160 text files describing respectively the content of each image in the first folder. The second dataset was created in such a way as to have 28 classes, each containing about 135 images describing an Arabic letter. In order to combine the two datasets, a modification to the format of the Arabic Sign Language ArSL dataset was necessary. We therefore grouped the images of the same letter under the same folder in order to get rid of the need for text files and to unify the format of both datasets. This made it easier for us to combine the two datasets later on. Note here that the Arabic Sign Language ArSL dataset had a problem with the non-existence of the images of the letter "Noun" in their appropriate class which required a manual correction of this by moving them to their original class.

B. Results

As previously mentioned, the dataset used for training contained more than 5000 images, distributed in a unified format on 28 different classes of Arabic Sign Language gestures. Three distinct sets of data were created from the dataset, with 60% of the data used for training, 20% for testing, and 20% for validation. The proposed technique would then involve the following steps: We started by pre-processing stage which we outcome black images that contain just the hand landmarks with a line connecting. After that, we inputted the data into proposed model, where the proposed model is iterated through several epochs and at the end of each iteration, accuracy values are provided. Finally, after running the last training epoch, the final accuracy is shown, and the training model is finished. The training and validation accuracy can be seen in Fig. 4(a) and training and validation loss in Fig. 4(b).

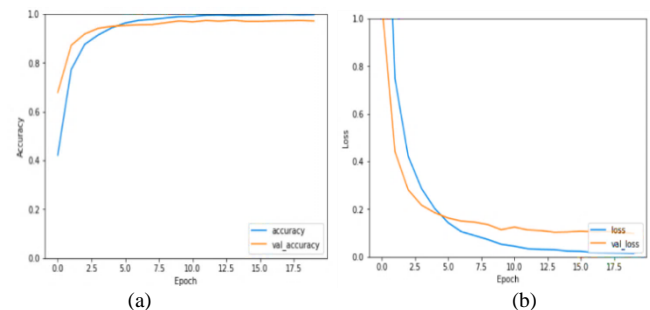


Fig. 4. (a) Training accuracy vs. validation accuracy, (b). Training loss vs. validation loss.

The architecture of CNN was trained for 20 epochs with a batch size of 32. At the 10th epoch, the model was able to validate with an accuracy of 97%, while the 20th epoch saw the highest accuracy of 97.07%. In Fig. 4(a), it can also be

noted how the model training progressed, which demonstrates also how the model stays ahead in every iteration of training and how the loss decreased (see Fig. 4(b)). According to the obtained result we observe no sign of the overfitting.

The confusion matrix shows how well the system performs in terms of correctly and incorrectly classifying developed data. The CM obtained by the CNN model using our methodology is shown in the following Fig. 5. The results obtained show that almost all gestures are correctly classified by our model.

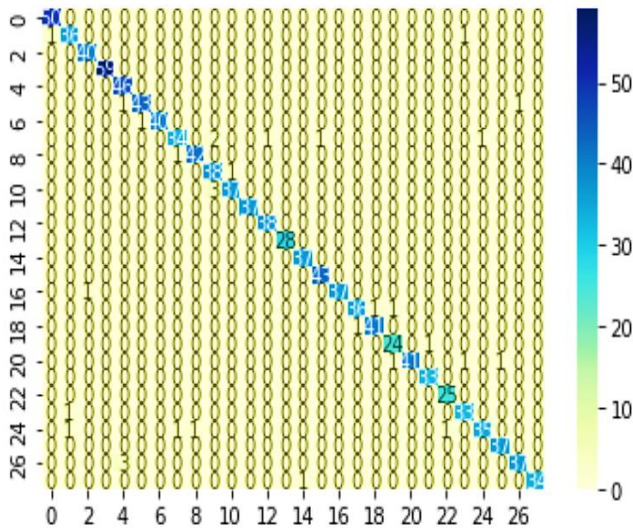


Fig. 5. Confusion matrix.

The classification report, as evident in Fig. 6, confirms what the confusion matrix states. Thus, it can be seen that the rate is actually low, and the majority of the class is properly categorized. The accuracy for the letters ("THA", "KHAA", "SEEN", "SAD", "ZAY") is high and ranges between 95 and 100%, whereas the accuracy for other gestures, such as the letters ("FAA", "JEEM", "RAA", "DEEL", "MEEM"), ranges between 88% and 94%. The principal reason for this difference is that each letter requires a different finger position. This confirms that accurate results depend greatly on how gestures are represented when features are extracted.

According to the obtained result, we can observe that the proposed method allows us to reduce the total number of incorrectly recognized signs to about 32. The misclassification rate is related as previously cited, to the fact that some letters have strong similarities in gestures (RAA/DELL, DELL/DHELL, RAA/ZAY, etc.), which has probably forced proposed model to extract similar features, rendering their classification process more difficult. The following figure (Fig. 7) displays an extract of misclassified images.

C. Comparative Study

To show the performance of the proposed system, we have compared the results obtained from recent literature [18],[29] and [30] with our method. We tested all methods on the same dataset used in this work. Table II reports the results of the three methods under consideration.

	precision	recall	f1-score	support
ALIF	0.98	1.00	0.99	50
BAA	0.95	0.95	0.95	38
TA	0.98	1.00	0.99	40
THA	1.00	1.00	1.00	59
JEEM	0.92	1.00	0.96	46
HAA	0.98	0.96	0.97	45
KHAA	1.00	0.98	0.99	41
DELL	0.94	0.87	0.91	39
DHELL	0.98	0.98	0.98	43
RAA	0.88	0.97	0.93	39
ZAY	0.97	0.93	0.95	40
SEEN	1.00	1.00	1.00	37
SHEEN	0.97	1.00	0.99	38
SAD	1.00	1.00	1.00	28
DAD	0.97	1.00	0.99	37
TAA	0.98	1.00	0.99	43
DHAA	1.00	0.97	0.99	38
AYN	0.97	0.95	0.96	38
GHAYN	0.98	0.98	0.98	42
FAA	0.92	0.96	0.94	25
QAAF	1.00	0.93	0.96	44
KAAP	0.97	1.00	0.99	33
LAAM	0.96	1.00	0.98	25
MEEM	0.94	0.97	0.96	34
NOON	0.97	0.90	0.93	39
HA	0.97	1.00	0.99	37
WAW	0.97	0.93	0.95	40
YA	1.00	0.97	0.99	35
accuracy			0.97	1093
macro avg	0.97	0.97	0.97	1093
weighted avg	0.97	0.97	0.97	1093

Fig. 6. Results for our approach.

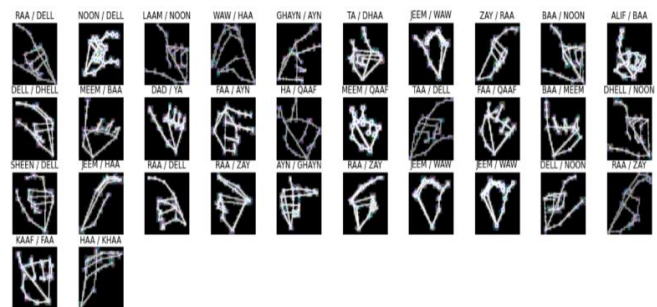


Fig. 7. An extract of the incorrectly classified images.

TABLE II. COMPARISON OF THE PROPOSED SYSTEM WITH OTHER METHODS

Methods	Accuracy
HAYANI et al. [18]	79.27%
KAMRUZZAMAN et al. [29]	83.91%
Ghazanfar Latif et al. [30]	90.61%
proposed model	97.07%

As compared with the results of the models listed above. The best result was reached with our suggested architecture, which has proven effective in the recognition of sign language due to a validation accuracy achieved of 97.70%.

V. CONCLUSION

In this paper, we used deep learning techniques to perform Arabic sign language recognition based on the hands of deaf people in real time in order to translate the signals into alphabets and improve communication between deaf and non-deaf people.

We started by collecting data and creating our own dataset, then applying a preprocessing that consists in cropping the hand from the image, and extracting its shape, then we proposed four CNN models with different architectures of which three architectures are proposed in three different scientific papers and the last one is proposed by ourselves, and applied them afterwards on our datasets with the aim of making a comparative study in order to choose the best model. The comparative study allowed us to conclude that our architecture is the best performing, where we obtained a score of 97.07%.

In the future work, we would like to extend the dataset for example, increase the variety of images in terms of noise, orientation, etc. and implement new techniques to convert hand movements into written text, and to use NLP (Natural Language Processing) techniques in order to process the obtained texts and present them with a comprehensible and adequate way.

REFERENCES

- [1] W. H. Organization, "Deafness and hearing loss", 2021. <https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss>.
- [2] S. Agrawal, A. Jalal, R. Tripathi, "A survey on manual and non-manual sign language recognition for isolated and continuous sign", *International Journal of Applied Pattern Recognition*, vol. 3, no. 2, pp. 99-134, 2016, doi: 10.1504/IJAPR.2016.079048.
- [3] B. Demircioglu, G. Bulbul, and H. Kose, "Turkish Sign Language recognition with Leap Motion", 24th Signal Processing and Communication Application Conference (SIU), no. 2015, pp. 589-592, 2016.
- [4] E. G. Mounq, C. C. Wooi, M. M. Sufian, C. K. On, and J. A. Dargham, "Ensemble-based face expression recognition approach for image sentiment analysis", *Int. J. Electr. Comput. Eng.*, vol. 12, no. 3, pp. 2588-2600, Jun. 2022.
- [5] B. Hdioud and M. E. H. Tirari, "Facial expression recognition of masked faces using deep learning", *IAES Int. J. Artif. Intell.*, vol. 12, no. 2, pp. 921-930, Jun. 2023.
- [6] S. P. Das, A. K. Talukdar, and K. K. Sarma, "Sign Language Recognition Using Facial Expression", *Procedia Comput. Sci.*, vol. 58, pp. 210-216, Jan. 2015.
- [7] D. Waters, R. Campbell, C. M. Capek, B. Woll, A. S. David, P. K. McGuire, M. J. Brammer, and M. MacSweeney, "Fingerspelling, signed language, text and picture processing in deaf native signers: The role of the mid-fusiform gyrus", *Neuroimage*, vol. 35, no. 3, pp. 1287-1302, 2007.
- [8] E. P. da Silva, K. M. O. Kumada, and P. D. P. Costa, "Analysis of Facial Expressions in Brazilian Sign Language (Libras)", *Eur. Sci. Journal*, ESJ, vol. 17, no. 22, pp. 1-1, Jul. 2021.
- [9] KumarKrishan, "DEAF-BSL: Deep IEarning Framework for British Sign Language recognition", *Trans. Asian Low-Resource Lang. Inf. Process.*, vol. 21, no. 5, pp. 1-14, Aug. 2022.
- [10] X. Jiang, S. C. Satapathy, L. Yang, S. H. Wang, and Y. D. Zhang, "A Survey on Artificial Intelligence in Chinese Sign Language Recognition", *Arab. J. Sci. Eng.*, vol. 45, no. 12, pp. 9859-9894, Dec. 2020.
- [11] K. Bantupalli and Y. Xie, "American Sign Language Recognition using Deep Learning and Computer Vision", *Proc. - 2018 IEEE Int. Conf. Big Data, Big Data 2018*, pp. 4896-4899, Jan. 2019.
- [12] K. Assaleh and M. Al-Rousan, "Recognition of Arabic sign language alphabet using polynomial classifiers", *EURASIP J. Adv. Signal Process.*, vol. 2005, no. 13, pp. 2136-2145, Jan. 2005.
- [13] M. Mohandes, J. Liu, and M. Deriche, "A survey of image-based Arabic sign language recognition", 2014 IEEE 11th Int. Multi-Conference Syst. Signals Devices, 2014.
- [14] Suharjo, R. Anderson, F. Wiryana, M. C. Ariesta, and G. P. Kusuma, "Sign Language Recognition Application Systems for Deaf-Mute People: A Review Based on Input-Process-Output", *Procedia Comput. Sci.*, vol. 116, pp. 441-448, 2017, doi: 10.1016/J.PROCS.2017.10.028.
- [15] O. Al-Jarrah and F. A. Al-Omari, "IMPROVING GESTURE RECOGNITION IN THE ARABIC SIGN LANGUAGE USING TEXTURE ANALYSIS", vol. 21, no. 1, pp. 11-33, Jan. 2007, <http://dx.doi.org/10.1080/08839510600938524>.
- [16] M. H. Ismail, S. A. Dawwd, and F. H. Ali, "Arabic Sign Language Recognition : A Review", *Int. J. Adv. Comput. Electron. Eng.*, vol. 6, no. June, pp. 1-12, 2021.
- [17] S. Hayani, M. Benaddy, O. El Meslouhi, and M. Kardouchi, "Arab Sign language Recognition with Convolutional Neural Networks", *Proc. 2019 Int. Conf. Comput. Sci. Renew. Energies, ICCSRE 2019*, Jul. 2019, doi: 10.1109/ICCSRE.2019.8807586.
- [18] A. MAhmed, R. Abo Alez, M. Taha, and G. Tharwat, "Automatic Translation of Arabic Sign to Arabic Text (ATASAT) System", pp. 109-122, 2016.
- [19] B. Hisham and A. Hamouda, "Supervised learning classifiers for Arabic gestures recognition using Kinect V2", *SN Appl. Sci.*, vol. 1, no. 7, pp. 1-21, 2019, doi: 10.1007/s42452-019-0771-2.
- [20] A. I. Shahin and S. Almotairi, "Automated Arabic Sign Language Recognition System Based on Deep Transfer Learning", *IJCSNS Int. J. Comput. Sci. Netw. Secur.*, vol. 19, no. 10, p. 144, 2019.
- [21] Y. Saleh and G. F. Issa, "Arabic sign language recognition through deep neural networks fine-tuning", *Int. J. online Biomed. Eng.*, vol. 16, no. 5, pp. 71-83, 2020, doi: 10.3991/IJOE.V16I05.13087.
- [22] A. Hasasneh, "Arabic sign language characters recognition based on a deep learning approach and a simple linear classifier", *Jordanian J. Comput. Inf. Technol.*, vol. 6, no. 3, pp. 281-290, 2020, doi: 10.5455/jicit.71-1587943974.
- [23] A. Althagafi, G. Althobaiti, T. Alsubait, and T. Alqurashi, "ASLR : Arabic Sign Language Recognition Using Convolutional Neural Networks", vol. 20, no. 7, pp. 124-129, 2020.
- [24] E. K. Elsayed and D. R. Fathy, "Sign language semantic translation system using ontology and deep learning", *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 1, pp. 141-147, 2020, doi: 10.14569/ijacsa.2020.0110118.
- [25] R. M. Duwairi and Z. A. Halloush, "Automatic recognition of Arabic alphabets sign language using deep learning", *Int. J. Electr. Comput. Eng.*, vol. 12, no. 3, pp. 2996-3004, 2022.
- [26] El Rwelli, R. Shahin, O. R. and Taloba, A. I. (2021). Gesture based Arabic Sign Language Recognition for Impaired People based on Convolution Neural Network," *Int. J. Adv. Comput. Sci. Appl.*, 12(12), 574-582, doi: 10.14569/IJACSA.2021.0121273.
- [27] F. Zhang et al., "MediaPipe Hands: On-device Real-time Hand Tracking", 2020, [Online]. Available: <http://arxiv.org/abs/2006.10214>.
- [28] S. BELMADOU, "Arabic Sign Language ArSL dataset | Kaggle", <https://www.kaggle.com/datasets/sabrielmadoui/arabic-sign-language-augmented-dataset/code>.
- [29] M. M. Kamruzzaman, "Arabic Sign Language Recognition and Generating Arabic Speech Using Convolutional Neural Network," *Wirel. Commun. Mob. Comput.*, vol. 2020, 2020, doi: 10.1155/2020/3685614.
- [30] G. Latif, N. Mohammad, R. AlKhalaf, R. AlKhalaf, J. Alghazo, and M. A. Khan", "An automatic arabic sign language recognition system based on deep CNN: An assistive system for the deaf and hard of hearing", *Int. J. Comput. Digit. Syst.*, vol. 90, no. 4, pp. 715-724, 2020, doi: 10.12785/ijcds/090418.

Improving QoS in Internet of Vehicles Integrating Swarm Intelligence Guided Topology Adaptive Routing and Service Differentiated Flow Control

Tanuja Kayarga¹, Ananda Kumar S^{2*}
SCOPE, Vellore Institute of Technology, Vellore, India^{1,2}

Abstract—Internet of Vehicles (IoV) is an evolution of vehicular adhoc network with concepts of internet of things (IOT). Each vehicle in IOV is an intelligent object with various capabilities like sensors, computation, storage, control etc. Vehicles can connect to any other entity in the network using various services like DSRC, C2C-CC etc. Ensuring QoS for vehicle to everything (V2X) communication is a major challenge in IoV. This work applies an integration of hybrid metaheuristics guided routing and service differentiated flow control to ensure QoS in Internet of Vehicles. Clustering based network topology is adopted with clustering based on hybrid metaheuristics integrating particle swarm optimization with firefly algorithm. Over the established clusters routing decision is done using swarm intelligence. Packet flows in the network are service differentiated and flow control is done at cluster heads to reduce the congestion in the network. High congestion in routes is mitigated with back up path satisfying the QoS constraints. Due to optimization in clustering, routing and data forwarding process, the proposed solution is able to achieve higher QoS. Through simulation analysis, the proposed solution is able to achieve 2% higher packet delivery ratio and 9.67% lower end to end packet latency compared to existing works.

Keywords—Particle swarm optimization (PSO); QoS; bio inspired technique

I. INTRODUCTION

Internet of Vehicles is seen as a most promising solution to address the challenges of modern transportation like improving road safety, mitigating traffic congestion, reduce fuel consumption etc[1]. Each vehicle is an intelligent object with sensing, computing, storing, control and communication facilities. Vehicles can connect to any entity within vehicular adhoc network (VANET) like Road Side Units (RSU), other vehicles and entities outside network like application servers, cloud etc. The connectivity is enabled using various mechanisms like dedicated short range communication (DSRC), car to car communication consortium (C2C-CC) etc. IoV is different from traditional IoT networks as it can generate information volume thousand times more than traditional IoT networks. IoV need to support messages of different categories like emergency messages, real time cooperative control messages, infotainment message etc with different delay and reliability constraints. Supporting wide variety of IoV applications with guaranteed quality of service (QoS) is a challenge in IoV networks. QoS requirements are in terms of packet delivery ratio, delay and jitter. Due to vehicle movement, network topology is dynamic and connectivity

between vehicles is on constant change. This affects the stability of the routes and in turn affects all the QoS parameters. Providing guaranteed QoS in highly dynamic IoV network is a very challenging multi constraint optimization problem [2]. The optimal solution to this problem can be found through hybrid metaheuristics.

In this work, an integrated solution combining hybrid metaheuristics for topology adaptive routing and service differentiated flow control is proposed to provide guaranteed QoS in IoV networks. The network is partitioned to clusters with efficient cluster head selection using hybrid metaheuristics based multi criteria optimization. Clustering is adaptive for a hybrid IoV with or without RSU availability. A QoS guaranteed routing using Ant colony optimization is proposed on the clustered topology. In addition it, packet flows in the network is differentiated and flow control is done at cluster heads to reduce the congestion and facilitate differential QoS for different services. In short, the proposed work is able to provide differential QoS guarantee for different services in IoV through three factors of topology control, topology adaptive routing and service differentiated flow control. Compared to earlier works addressing any one particular area of topology, routing or data forwarding optimization the proposed solution integrates optimization in all three areas of topology, routing and data forwarding. This integration has provided added advantage in achieving higher QoS in proposed solution.

The paper is organized as follows. Section II presents the existing works in VANET for ensuring higher QoS and their shortcomings. Section III presents the research gap. Section IV presents the proposed solution integrating clustering, routing and data forwarding optimizations to ensure higher QoS. Section V presents the novelties in the proposed solution. Section VI presents the results of the proposed solution and its comparison to existing works. Section VII presents the conclusion and scope of future work.

II. RELATED WORK

Authors in [3] proposed a QoS routing protocol for VANET. The protocol is designed specifically for high wayscenario. The network is clustered with cluster head selected based on the link expiration time of the neighbors. Artificial Bee Colony (ABC) algorithm is used to find the optimal path with higher QoS. This work enforces QoS requirements both during clustering and routing. The performance of solution was tested against different traffic

scenarios and the work was found to perform well for highway scenario where traffic density is normal. The performance degrades for higher urban traffic density. Authors in [4] proposed a QoS based clustering algorithm for mobile adhoc networks. The algorithm tries to create stable clusters and maintain stability during communications and link failure at the same time considers QoS requirements. Mobility metrics like velocity and distance are added to QoS functions to select stable clusters. The relay nodes for routing are selected using ant colony algorithm with requirement of satisfying mobility and stability constraints. A cheating prevention mechanism is also proposed in this work for fair and reliable relay selection. Direction of movement, load processed by MPR and packet flow categories are not considered in this work. A passive clustering aided routing protocol is proposed in [5] to improve the routing performance for one way multi-lane highway scenario. A multi metric distributed election strategy using metrics such as node degree, expected transmission count, and link lifetime is proposed for cluster head selection. The election strategy tries to maximize the stability of cluster structure. A weighted routing protocol based on metrics of expected transmission time, link life time and node degree is used for selecting the best relay cluster head. The work could not accommodate dynamic variation in speed and density. Authors in [6] proposed a QoS multicast routing protocol for VANET based on bee communication principle. The proposed protocol is able to find and maintain robust route between the source and multicast group members. The path is selected to maximize the average bandwidth and packet delivery ratio at same time minimizing the end to end delay and normalized overhead load. Failure of link cause due to vehicle dynamics is not considered for multicast tree reorganization in this work. A situation aware multi constrained routing protocol is proposed in [7] for VANET combining situational awareness and ant colony algorithm. From among multiple feasible routes between source and target vehicle, a best route satisfying multiple QoS constraints and reliability metrics is found. The reliability risks in the best route selected is mitigated using situation awareness and ant colony optimization by proactively predicting link failure and replacing them with backup links. The work predicts only movement based failures, but could have been extended for congestion based failures. Authors in [8] proposed a clustering algorithm for IoV network with nature inspired Moth flame optimizer. The cluster heads are selected in such a way that average of distance between cluster members to cluster head is minimal in the network. By minimizing the distance, the transmission energy is also reduced in the network. Clustering did not accommodate any QoS requirements in this work. A new stability based clustering algorithm is proposed for VANET is [9]. The algorithm has two stages: setup and maintenance. In setup stage, close proximity nodes are organized to clusters with selection of cluster head. In maintenance stage, a backup cluster head is selected and it becomes primary to maintain stability and reliability of cluster. The stability of cluster is measured in terms of velocity differences between the vehicles in the cluster and this difference is important criteria for clustering in this work. QoS requirements are not considered as criteria for clustering in this work. Authors in [10] proposed a hybrid of harmony search algorithm (HSA) and PSO algorithm

for cluster head selection in sensor networks. The cluster heads are selected to minimize the energy consumption in the network. With high search efficiency of HSA and dynamic capability of PSO, efficient clustering of network is done resulting in increased life time of the network. Authors in [11] used multi objective particle swarm optimization to optimize the number of clusters in adhoc network. Degree of nodes, transmission power, energy consumption of nodes and velocity of nodes and are the parameters considered for optimization. The goal of cluster head selection is to reduce the overall energy consumption in the network. Authors in [12] proposed a clustering algorithm for VANET using grey wolf optimization. Social behavior and hunting mechanism of grey wolfs are replicated to create efficient clusters. Stability of clusters is the optimization criteria for the fitness function in grey wolf optimization algorithm. Authors in [13] proposed two bio inspired evolutionary algorithms for clustering in VANET – comprehensive learning PSO and Multi objective PSO. Maximizing the number of cluster members without affecting the stability due to velocity of the vehicles is the optimization criteria used in both the evolutionary algorithms. Authors in [14] proposed a clustering algorithm using ant colony optimization for VANET. The variables considered for clustering are distance of neighboring nodes from cluster head, speed of vehicles, direction of cluster head and cluster nodes within a cluster. The algorithm tries to find the minimal number of clusters so that packet routing cost is minimized. Dragon fly optimization algorithm is used for clustering VANET in [15]. The approach also adapted the transmission rate dynamically based on the traffic density and mobility pattern. Dragon fly fitness function for cluster selection is based on speed, velocity and direction of the vehicles. A routing algorithm integrating ant colony optimization with Dynamic MANET (DYMO) on demand protocol is proposed for VANET in [16]. The path selection is based on two parameters of hop distance and path reliability. Path reliability is measured in terms of expected number of transmissions. Authors in [17] proposed an efficient clustering for flying adhoc networks. A variant of K-means density clustering algorithm is used for selection of cluster heads. Transmission power of nodes is adjusted based on their operational requirements. The approach works only for low mobility of nodes. Authors in [18] proposed a novel path planning algorithm for reliable and efficient routing in VANET. Congestion is managed by distributing load over alternate paths and splitting the larger packets to small packets. Routing is based on multiple QoS constraints of packet delivery ratio and energy. Authors in [19] proposed a novel, secure and reliable multi constrained QoS aware routing algorithm for VANETs. Ant colony optimization is used to compute feasible routes ensuring differential reliability for different traffic types. Authors in [20] proposed a novel moving zone based routing protocol. The moving zone is formed by clustering the vehicles with similar movement patterns and trajectories. A coordinator node is selected for each moving zone to assist in message dissemination. Authors in [21] used ant colony optimization for selection of routing paths with higher QoS in VANET. The factors considered for QoS improvement in this work were packet delivery ratio, delay and link connectivity. Routing to improve QoS based on vehicle orientation is proposed in [22]. The salient part of this

work is that it translated multi constraint QoS objectives to single constraint objectives using a cosine similarity scalarization model. Authors in [23] proposed an angle based clustering algorithm for VANET. The most stable vehicles that can act as cluster head are selected based on the angular position and direction of the vehicles. Each vehicle builds its neighborhood information using an angular technique. Clusters are created based on this neighborhood information. Authors in [24] used the concept of location aided routing to propose a reactive routing protocol for VANET. The protocol used multi objective particle swarm optimization for searching the route. Authors in [25] surveyed different QoS routing protocols in VANET. The study observed QoS cannot be guaranteed in VANET only with routing and it needs much more mechanisms like flow control, class differentiation etc. The solution proposed by us is in this direction of integrating flow control and class differentiation along with routing and topology control for guaranteed QoS in VANET.

III. RESEARCH GAP

- Following are some of important research gaps identified from the survey
- Most of the existing solutions consider ensuring QoS in IoV as topology control and routing problem without consideration of different services and their QoS needs. Better QoS management strategy must also consider service differentiation and flow control along with topology and routing management.
- Most of existing multi objective optimization based clustering works use limited parameters like velocity, direction etc, without consideration for their resource capability. Multi objective optimization based clustering methods must also include resource metrics affecting QoS, so that the QoS can be improved in the network.
- In this work, we address the above two open issues and design a swarm intelligence guided topology and routing management with multi objective QoS guarantee and integrating it with service differentiated flow control for better QoS management in IoV.

IV. PROPOSED SOLUTION

The proposed solution for improving QoS in IoV has following important functionalities

A. Multi Objective QoS Constrained Clustering

Multi objective QoS constrained clustering is implemented using particle swarm optimization (PSO). Each particle is the array of length N with each element in the array is the ID of the cluster head node. Each particle covers whole of the network and particle does not have any duplicate ID. The multi objective particle swarm optimization algorithm proposed in [11] is adapted. While the approach [11] is designed with goal of clustering for energy optimization, in this work, clustering is done for multi objective QoS optimization. A new multi constraint QoS objective function is introduced. The function considers following parameters

- Average Stability of cluster (S)

- Average effective hop count between clusters (h_c)
- Average Degree difference (D_d)
- Average uncovered vehicles on speed/direction variations. (U_c)

In [9], the objective values are measured for each cluster head in the particle and they are then summed up to provide the total objective value for the particle. Deviating from it, this work models selects the cluster heads as a multi objective optimization problem using hybrid metaheuristics. Compared to using single optimization algorithm, use of hybrid metaheuristics solves the problem of local minima. Combining algorithms with contrasting features of exploration and exploitation results in optimal solution avoiding the local minima. In this work Particle swarm optimization (PSO) with exploration capability is combined with firefly algorithm with exploitation capability to select the optimal cluster heads.

PSO is a swarm intelligence algorithm (Kennedy et al 1995) simulating the social behavior of swarm of organisms. This method is popular for solving optimization problems due to its simplicity, flexibility and versatility. Organisms move randomly with different velocities and use these velocities to update their individual position. Each candidate solution is a 'particle'. Each particle tries to attain its best velocity based on its own local best (p_{best}) value and its neighbor's global best (g_{best}). Each particle's next position depends on the current position, current velocity, distance from current position to p_{best} , distance from current position to g_{best} . The movement of particle in its search space depends on its velocity. For a particle X, its current position X_i and current velocity V_i is updated as

$$X_i(t+1) = X_i(t) + V_i(t+1) \quad (1)$$

$$V_i(t+1) = wV_i(t) + c_1r_1(p_{besti}(t) - X_i(t)) + c_2r_2(g_{besti}(t) - X_i(t)) \quad (2)$$

In the above equations, t is the iterative value. c_1 and c_2 are acceleration coefficients, r_1 and r_2 are random numbers, w is the inertia weight. The iteration is repeated till termination condition is met.

Firefly is swarm intelligence algorithm (Yang et al 2008) based on the behavior of fireflies in naturally occurring environment. Fireflies exhibit unique light flashes for various purposes like mating, warning about potential danger etc. Fireflies operation is guided by two parameters: light intensity and level of attractiveness. Light intensity (I) is inversely proportional to the distance between the emitting and observing firefly (r). It is given as

$$I = \frac{1}{r^2} \quad (3)$$

Level of attractiveness is proportional to the light intensity. It is calculated as

$$\beta(r) = \beta_0 e^{-\gamma r^2} \quad (4)$$

β_0 is the attractiveness at $r=0$ and γ is the light absorption coefficient. Euclidean distance formula is used for calculating r .

The movement of firefly (FX_i) governed by attraction from another firefly (FX_j) is calculated as

$$FX_i = FX_i + \beta_0 e^{-\gamma r^2} (FX_j - FX_i) + \alpha \epsilon_i \quad (5)$$

In the above equation α is the randomization parameter and ϵ_i is a random number.

The hybrid metaheuristics algorithm for clustering first invokes firefly algorithm to select the cluster heads with random cluster heads as input. PSO algorithm is then started with the cluster heads selected by firefly as input and the output of the PSO is the optimal cluster heads.

The fitness function for selection of optimal cluster heads is framed as

$$F_f = w_1 S + w_2 h_c + w_3 \frac{1}{d_d} + w_4 \frac{1}{u_c} \quad (6)$$

Firefly algorithm is started with K fireflies each firefly has N bit string with M (number of desired cluster head) out it is marked as 1 and rest marked as 0. Value of 1 represents, the corresponding node is cluster head. The initial N bit string is formed randomly. Firefly algorithm is run with fitness function defined in Eq. (6) on each iteration and the firefly with maximum value for F_f is selected as best and movement of other firefly is done based on best firefly. Once the maximum iteration is reached or error has reached minimum value firefly algorithm is stopped and N bit string of the best firefly is taken as the seed. From this seed, K particles are initialized with N bit string by toggling P random positions. The PSO algorithm is then started with this solution formed from firefly. At each iteration, the fitness function defined in Eq. (6) is calculated and the particle is maximum value for F_f is selected as globally best solution. Once the maximum iteration is reached or error has reached minimum value PSO algorithm is stopped and N bit string of the globally best particle is taken. The positions where 1 is marked in N bit string are taken and the corresponding nodes are made as cluster heads.

B. QoS based Routing

The proposed QoS based routing uses both two different routing modes: vehicle to cluster head and cluster head to cluster head. Vehicle to cluster head is through geographic routing. Cluster head to cluster head routing is using ant colony optimization algorithm.

The source vehicle which needs to send packet to destination vehicle sends routing request message to the source cluster head. The routing request message also carries the service identifier. Source cluster head initiates a group of forward ant in search of the destination vehicle cluster head. The service identifier is carried in the forward ant. When forward ant reaches any cluster head i, it records the cluster head ID and transmission delay, it makes a decision to choose the next cluster head based on the global pheromone (gp) and local pheromone (lp) stored at cluster head i. The probability of selecting the next cluster head j is given as

$$p_{ij} = \frac{gp_{ij}^\alpha lp_{ij}^\beta}{\sum_{m=1}^K gp_{im}^\alpha lp_{im}^\beta} \quad (7)$$

Where K is the number of next hop cluster heads to i. α, β are the weight values. Local pheromone (lp) and global pheromone (gp) is calculated in terms of QoS requirements as

$$lp_{ij} = w_1 C_p(e_{ij}) + w_2 pdr_p(e_{ij}) + w_3 d_p(e_{ij}) + w_4 j_p(e_{ij}) + w_5 L_p(e_{ij}) \quad (8)$$

Where w_x is the weight associated with each factor such a way that

$$\sum_{x=1}^5 w_x = 1 \quad (9)$$

e_{ij} is the link between cluster head i and j. C_p is the connection probability, pdr_p is the predicted packet delivery ratio, d_p is the predicted delay, j_p is the predicted jitter and L_p is the predicted link expiration time. For two cluster heads u and v positioned at (x_u, y_u) and (x_v, y_v) , moving with velocity v_u and v_v in direction of θ_u and θ_v , the link expiration time is calculated as

$$L(u, v) = \frac{\sqrt{d^2(a^2+c^2)-(ad-bc)^2-(ab+cd)}}{a^2+c^2} \quad (10)$$

Where

$$\begin{aligned} a &= v_u \cos \theta_u - v_v \cos \theta_v \\ b &= x_u - x_v \\ c &= v_u \sin \theta_u - v_v \sin \theta_v \\ d &= y_u - y_v \end{aligned}$$

In each link, the past values of delivery ratio, connectivity probability, delay and jitter are collected. Based on the previous history, prediction over new interval is done applying exponential moving average as

$$v(t+1) = \alpha * x + (1 - \alpha)v(t) \quad (11)$$

Where x is the current observed value and $v(t+1)$ is the predicted value. α is a constant.

When forward ant arrives at destination cluster head, it checks the delay value in the forward ant against the delay threshold set for the service. If the delay in the forward ant is less the delay threshold, the path passed by the ant is selected as a probable candidate route. The forward ants which arrive through candidate route are converted to backward ant. The backward ant traverses back to the source cluster head. As it traverses, its updates the global pheromone value as

$$\begin{aligned} gp_{ij} &= \\ (1 - \delta)gp_{ij} + \delta * \\ \text{best local phenome value over the path} \end{aligned} \quad (12)$$

Once all backward ants arrive at source cluster head, for each backward ant, a QoS fitness score fs is calculated for the route provided by the backward ant. The route having the highest value for fs is the path used for routing between source and destination cluster head. QoS fitness score fs is calculated as

$$fs(P_x) = w_1 C_p(P_x) + w_2 pdr_p(P_x) + w_3 d_p(P_x) + w_4 j_p(P_x) + w_5 L_p(P_x) \quad (13)$$

Where P_x is the path traversed by the backward ant.

The backward ant calculates the C_p of entire path as minimum value of connection probability of all links in the path. It calculates the pdr_p of entire path as minimum value of packet deliver ratio of all links in the path. It calculates the d_p of entire path as maximum value of delay of all the links in the path. It calculates the j_p of entire path as maximum value of jitter of all the links in the path. It calculates the L_p of entire path as minimum value of link expiration time of all links in the path.

C. Service Differentiation and Flow Control

The flows or sessions in the network are categorized based on the services to different priorities. At each of the cluster head, a differential flow control is done. Cluster head maintains multiple queues, each corresponding to a priority for queuing the incoming packets. Different from allocating processing slots based on priority alone, in this work, we use estimated traffic demand in each priorities, priority of the packet and estimated congestion on immediate forwarding links as the three parameters for allocating processing slots to the queues. The processing slots are grouped into time frames. The traffic demand for each priority queue is estimated as

$$DQ = \min(MA_i + D_i, T \times \Delta) \quad (14)$$

The exponential moving average of incoming traffic to queue MA is calculated with D as total packets in queue and physical transmission rate as T as

$$MA_i = \begin{cases} \alpha \cdot T + (1 - \alpha)MA_{i-1} & \text{if } T \neq 0 \\ (1 - \alpha)MA_{i-1}, & \text{otherwise} \end{cases} \quad (15)$$

From the estimated traffic demand the slots n_s for queue x is allocated out of N total slots as given below:

$$n_{S_x} = \frac{FD_x}{\sum_{x=1}^n FD_x} * \frac{P_x}{\sum_{x=1}^n P_x} * N \quad (16)$$

Where P is the priority of queue. The total slots N is not fixed and it should be adapted to congestion on immediate forwarding links. We model congestion in terms of predicted round trip time(RTT). RTT prediction is done as a probability mass function of past delay distribution. It is calculated as:

$$RTT = \begin{cases} \sum_{i=0}^{\infty} f_i(a) \cdot f_i(b), & x = 0 \\ \sum_{i=0}^{\infty} f_i(a) \cdot f_{2x+i}(b) + \sum_{i=0}^{\infty} f_i(b) \cdot f_{2x+i}(a), & x > 0 \end{cases} \quad (17)$$

The forward direction for packet travel is denoted as a. The backward direction for packet travel is denoted as b. The probability mass function of delay in direction v is denoted as f(v).

From the predicted RTT, the value of N is calculated as:

$$N = \frac{k}{RTT} \quad (18)$$

Where, k is the total number of messages that can be forwarded when RTT is at a best lower bound value. This value is configured by network administrator.

V. NOVELTY IN PROPOSED SOLUTION

Following are the novelties in the proposed solution

- Ensuring QoS in IoV is solved through an integrated solution combining QoS based topology control, QoS based routing and service differentiated flow control.
- Clustering using swarm intelligence considers multiple QoS metrics in addition to usual parameters of velocity, direction, etc.
- Multi class multi QoS optimization based ant colony optimization is done to find effective routes for different service flows.
- Adaptive service differentiated flow control based on network congestion feedback is proposed.

VI. RESULTS

NS2 simulator is used for measuring the effectiveness of the proposed solution. The vehicle traces are generated using SUMO and NS2 extension code implements the proposed solution on these traces.

The simulation was conducted against configuration parameters in Table I.

TABLE I. SIMULATION CONFIGURATION

Sl no.	Simulation Configuration	
	Parameter	Value
1	Road length	4Km
2	Road length	4Km
3	Road topology	Highway
4	Number of lanes	4
5	Vehicle density	25 to 200
	Vehicle speed	30 m/s
7	Transmission range	250 m
8	MAC protocol	IEEE 802.11 p
9	Data packet size	1000 bytes
10	Data rate	5 packet/second
11	Simulation time	1000 seconds

Four different traffic types are used for simulation. The traffic types and their priorities are given in Table II.

TABLE II. TRAFFIC CLASSES

Sl no.	Simulation Configuration	
	Traffic class	Priority
1	Bulk transfer	2
2	Audio	3
3	Video	3
4	General applications	5

The performance of the proposed solution is compared against CBQoS-Vanet [3] and SAMQ [7] in terms of following metrics:

- Packet delivery ratio
- Packet dropped ratio
- Normalized overhead load
- Average end to delay
- Throughput

Varying the number of vehicles, packet delivery ratio is measured and the results are given in Fig. 1.

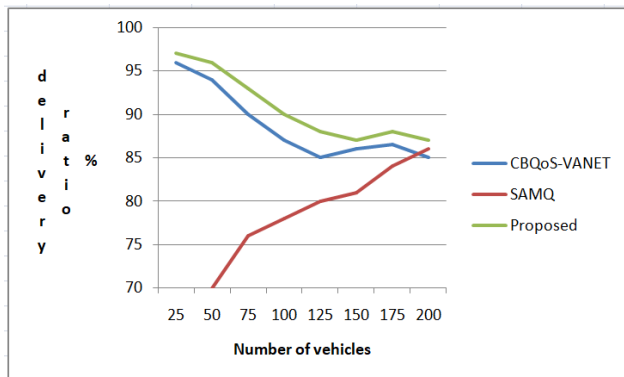


Fig. 1. Packet delivery ratio.

From the Fig. 1, it can be seen that in both CBQoS-VANET and Proposed, as the vehicle density increases, the delivery ratio drops. But the delivery ratio in Proposed is much higher compared to SAMQ and CBQoS-VANET. The increase in delivery ratio in the proposed when compared to CBQoS-VANET is due to implementation of flow control in the proposed. On an average packet delivery in the proposed solution is 2.07 % more compared to CBQoS-VANET and 12.88% more compared to SAMQ.

The results of packet dropped ratio for different vehicle density is given in Fig. 2.

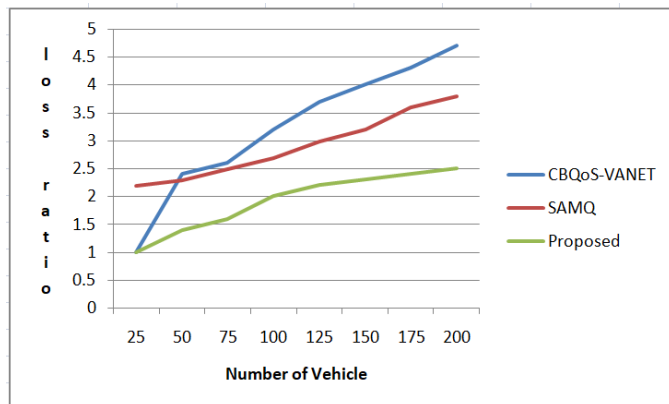


Fig. 2. Loss ratio.

From the Fig. 2, it can be seen that although the loss rate is within 2 to 6% in all the three protocols, best performance is achieved in proposed with average loss rate 70% lower compared to CBQoS-VANET and 53.15% lower compared to SAMQ. Congestion is controlled based on RTT feedback in the

proposed solution. Due to this, the loss rate is lower in the proposed solution.

Varying the vehicle density, normalized overhead is measured and the result is given in Fig. 3.

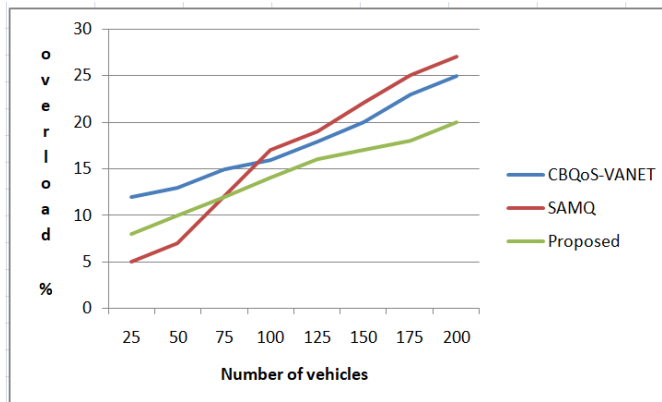


Fig. 3. Network overhead.

From the Fig. 3, the normalized overload in the proposed solution is on average is 16.52% lower compared to SAMQ and 23.47% lower compared to CBQoS-VANET. The overhead increases with higher vehicle density. The overhead is higher in proposed compared to SAMQ in lower vehicle density is due to clustering and routing finding process, but SAMQ overhead increased for higher vehicle density.

Varying the vehicle density, the average end to end delay for packet delivery is measured and the result is given in Fig. 4.

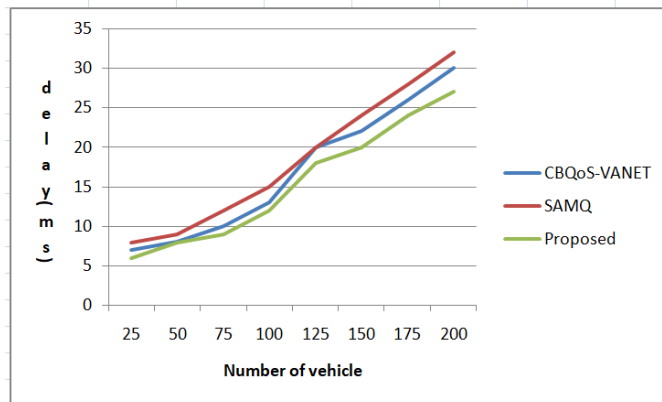


Fig. 4. Delay.

As seen from Fig. 4, delay increases with higher vehicle density. The average end to end delay in proposed solution is 9.67% lower compared to CBQoS-VANET and 19.35% lower compared to SAMQ. Service differentiation and flow control has reduced the average end to end delay in the proposed solution.

Varying the vehicle density, throughput is measured and the result is given in Fig. 5.

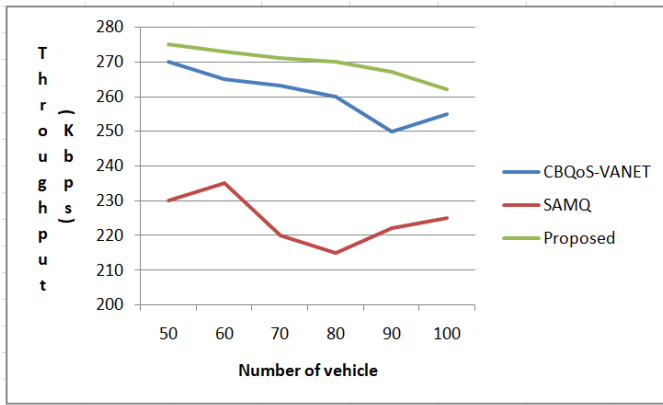


Fig. 5. Throughput.

As seen from Fig. 5, throughput reduces with higher vehicle density. The average throughput in proposed solution is 3.51% higher compared to CBQoS-VANET and 16.85% higher compared to SAMQ. The throughput has increased in proposed due to consideration of QoS constraints both during clustering and routing.

Varying the speed for 100 vehicles, the packet delivery ratio is measured and the results are given in Fig. 6.

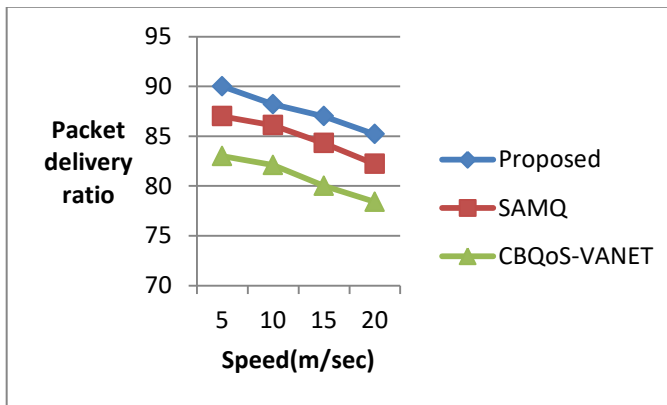


Fig. 6. Packet delivery ratio vs speed.

The packet delivery ratio is higher in proposed even as the speed increases. The average packet delivery in proposed solution is at least 3% higher compared to SAMQ and 7.6% higher compared to CBQoS-VANET. The higher packet delivery ratio in proposed solution is due to selection of more stable cluster head and effective routing path for forwarding the packet. In addition, the service differentiation based source rate control limits the packets during congestion and this helped to improve the packet delivery ratio even more in proposed solution.

Varying the speed for 100 vehicles, the packet delay is measured and the results are given in Fig. 7.

The delay is lower in proposed solution compared to existing works. It is on average 13% lower compared to SAMQ and 9.4% lower compared to CBQoS-VANET. The delay has reduced due to reduction of congestion and retransmissions due to congestion in the proposed solution.

This was achieved using service differentiated control of source rate of packets.

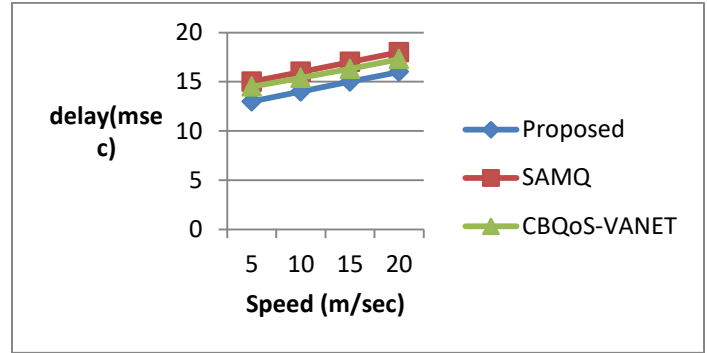


Fig. 7. Packet delay vs speed.

The optimization performance of the proposed hybrid metaheuristics is measured in terms of best value and worst value for the fitness function for six different network configurations (vehicle density). The performance is compared against PSO, Firefly and genetic algorithm and results are given in Tables III and IV.

TABLE III. BEST VALUES OF FITNESS FUNCTION

No of vehicles	Fitness Function			
	PSO+Firefly	PSO	Firefly	GA
50	4.72E+04	3.12E+08	1.86E+09	3.40E+06
60	2.62E+02	2.68E+10	7.11E+10	3.39E+06
70	3.04E+02	4.06E+04	3.79E+05	4.25E+02
80	4.00E+02	3.55E+03	1.09E+04	4.33E+02
90	6.06E+02	6.34E+02	6.44E+02	6.21E+02
100	7.00E+02	9.30E+02	1.56E+03	7.02E+02

TABLE IV. WORST VALUES FOR FITNESS FUNCTION

No of vehicles	Fitness Function			
	PSO+Firefly	PSO	Firefly	GA
50	1.62E+07	5.57E+08	5.27E+09	5.47E+07
60	1.55E+04	4.18E+10	1.14E+11	3.63E+06
70	1.01E+03	5.94E+04	1.60E+08	5.12E+04
80	5.80E+02	6.44E+03	3.30E+04	6.44E+02
90	6.17E+02	6.39E+02	6.53E+02	6.30E+02
100	7.01E+02	1.04E+03	2.01E+03	7.03E+02

In terms of best values and worst values, the proposed PSO+Firefly combination has better values compared to PSO, Firefly and GA algorithm. This is due to combination of both exploration and exploitation capability in the PSO+Firefly compared to their usage in separation.

A. Limitations

The proposed solution has following limitations:

- The solution was tested only for highway scenarios where directional mobility is not dynamic. The

applicability of solution for city scenarios with dynamic directional mobility is yet to be tested.

- The solution made assumption on ratio of traffic distributions and data generators. Testing for realistic scenarios like peak hour periods, etc. was not considered in this work.

VII. CONCLUSION

An integrated approach combining swarm intelligence guided clustering/routing with service differentiated flow control is proposed in this work. Clustering is done using multi optimization hybrid metaheuristics and routing paths are found using ant colony optimization. In both clustering and routing swarm intelligence algorithms, objective functions are designed with QoS constraints. The proposed solution is able to achieve 2% more packet delivery ratio, 9.67% lower end to end delay, 3.51% more throughput compared to existing solution. Extending the solution for city scenarios and realistic traffic distributions is in scope of future work.

REFERENCES

- [1] Kaiwartya, O., Abdullah, A., Cao, Y., Altameem, A., Prasad, M., Lin, C. and Liu, X., (2016). Internet of Vehicles: Motivation, Layered Architecture, Network Model, Challenges, and Future Aspects. *IEEE Access*, 2016, 4, pp.5356-5373
- [2] Rontala Subramaniam, Prabhakar & Thangavelu, Arunkumar & Chitra, Venugopal. (2011). QoS for highly dynamic Vehicular ad hoc network optimality. 2011 11th International Conference on ITS Telecommunications, ITST 2011. 405-411. 10.1109/ITST.2011.6060091.
- [3] Abderrahmane Lakas, Mohamed El Amine Fekair, Ahmed Korichi, Nasreddine Lagraa, "A Multiconstrained QoS-Compliant Routing Scheme for Highway-Based Vehicular Networks", *Wireless Communications and Mobile Computing*, vol. 2019, Article ID 4521859, 18 pages, 2019.
- [4] O. A. Wahab, H. Otrok, and A. Mourad, "VANET QoS-OLSR: QoS-based clustering protocol for Vehicular Ad hoc Networks," *Computer Communications*, vol. 36, no. 13, pp. 1422–1435, 2013.
- [5] S.-S. Wang and Y.-S. Lin, "PassCAR: A passive clustering aided routing protocol for vehicular ad hoc networks," *Computer Communications*, vol. 36, no. 2, pp. 170–179, 2013.
- [6] S. Bitam, A. Mellouk, and S. Fowler, "MQBV: Multicast quality of service swarm bee routing for vehicular ad hoc networks," *Wireless Communications and Mobile Computing*, vol. 15, no. 9, pp. 1391–1404, 2015.
- [7] M. H. Eiza, T. Owens, Q. Ni, and Q. Shi, "Situation-aware QoS routing algorithm for vehicular ad hoc networks," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 12, pp. 5520–5535, 2015
- [8] Khan, Muhammad & Aadil, Farhan & Maqsood, Muazzam & Bukhari, Syed & Hussain, Maqbool & Nam, Yunyoung. (2018). Moth Flame Clustering Algorithm for Internet of Vehicle (MFCA-IoV). *IEEE Access*. PP. 1-1. 10.1109/ACCESS.2018.2886420.
- [9] A. Ahizoune and A. Hafid, "A new stability based clustering algorithm (SBCA) for VANETs," in *Proc. IEEE 37th Conf. Local Comput. Netw. Workshops (LCN Workshops)*, Oct. 2012, pp. 843–847
- [10] T. Shankar, S. Shanmugavel, and A. Rajesh, "Hybrid HSA and PSO algorithm for energy efficient cluster head selection in wireless sensor networks," *Swarm Evol. Comput.*, vol. 30, pp. 1–10, Oct. 2016.
- [11] H. Ali, W. Shahzad, and F. A. Khan, "Energy-efficient clustering in mobile ad-hoc networks using multi-objective particle swarm optimization," *Appl. Soft Comput.*, vol. 12, no. 7, pp. 1913–1928, 2012.
- [12] M. Fahad et al., "Grey wolf optimization based clustering algorithm for vehicular ad-hoc networks," *Comput. Electr. Eng.*, vol. 70, pp. 853–870, Aug. 2018
- [13] M. Fahad, F. Aadil, S. Ejaz, and A. Ali, "Implementation of evolutionary algorithms in vehicular ad-hoc network for cluster optimization," in *Proc. IEEE Conf. Intell. Syst. Conf. (IntelliSys)*, Sep. 2017, pp. 137–141.
- [14] F. Aadil, K. B. Bajwa, S. Khan, N. M. Chaudary, and A. Akram, "CACONET: Ant colony optimization (ACO) based clustering algorithm for VANET," *PLoS ONE*, vol. 11, no. 5, p. e0154080, 2016
- [15] F. Aadil, W. Ahsan, Z. U. Rehman, P. A. Shah, S. Rho, and I. Mehmood, "Clustering algorithm for Internet of vehicles (IoV) based on dragonfly optimizer (CAVDO)," *J. Supercomput.*, vol. 74, no. 9, pp. 4542–4567, 2018.
- [16] A. M. Oranj, R. M. Alguliev, F. Yusifov, and S. Jamali, "Routing algorithm for vehicular ad hoc network based on dynamic ant colony optimization," *Int. J. Electron. Elect. Eng.*, vol. 4, no. 1, pp. 79–83, Feb. 2016.
- [17] F. Aadil, A. Raza, M. F. Khan, M. Maqsood, I. Mehmood, and S. Rho, "Energy aware cluster-based routing in flying ad-hoc networks," *Sensors*, vol. 18, no. 5, p. 1413, 2018.
- [18] F. I. Shaikh and H. A. Hingoliwala, "Path planning based QoS routing in VANET," 2017 International Conference on Big Data, IoT and Data Science (BIG), Pune, 2017, pp. 37-43, doi: 10.1109/BIG.2017.8336570
- [19] Mahmoud Hashem Eiza, Thomas Owens and Qiang Ni, "Secure and Robust Multi-Constrained QoS Aware Routing Algorithm for VANETs", *IEEE Transactions on Dependable and Secure Computing*, vol. 13, no. 1, 2016.
- [20] An Lin, Jian Kang, Anna Squicciarini, Yingjie Wu, Sashi Gurung and Ozan Tonguz, "MoZo: A Moving Zone Based Routing Protocol Using Pure V2V Communication in VANETs", *IEEE Transactions on Mobile Computing*, 2016.
- [21] G. Li, L. Boukhatem and J. Wu, "Adaptive Quality-of-Service-Based Routing for Vehicular Ad Hoc Networks With Ant Colony Optimization", *IEEE Transactions on Vehicular Technology*, vol. 66, no. 4, pp. 3249-3264, April 2017.
- [22] Ankur Nahar, Debasis Das, and Sajal K. Das. 2020. OBQR: Orientation-Based Source QoS Routing in VANETs. *Proceedings of the 23rd International ACM Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems*. Association for Computing Machinery, New York, NY, USA, 199–206.
- [23] Hadded, Mohamed, Paul Muhlethaler, Anis Laouiti, and Leila Azzouz Saidane. "A novel angle-based clustering algorithm for vehicular ad hoc networks." In *Vehicular Ad-Hoc Networks for Smart Cities*, pp. 27--38. Springer, Singapore, 2017
- [24] Mustafa Qasim AL-Shammari and Ravie Chandren Muniyandi, "Optimised Tail-based Routing for VANETs using Multi-Objective Particle Swarm Optimisation with Angle Searching" *International Journal of Advanced Computer Science and Applications (IJACSA)*, 11(6), 2020. <http://dx.doi.org/10.14569/IJACSA.2020.0110629>
- [25] Oche, Michael & Bello Tambawal, Abubakar & Chemebe, Christopher & Md. Noor, Rafidah & Distefano, Salvatore. (2018). VANETs QoS-based routing protocols based on multi-constrained ability to support ITS infotainment services. *Wireless Networks*. 10.1007/s11276-018-1860-7.

A Comparative Performance Evaluation of Routing Protocols for Mobile Ad-hoc Networks

Baidaa Hamza Khudayer¹, Lial Raja Alzabin², Mohammed Anbar³, Ragad M Tawafak⁴, Tat-Chee Wan⁵, Abir AlSideiri⁶, Sohail Iqbal Malik⁷, Taief Alaa Al-Amiedy⁸

National Advanced IPv6 Centre, Universiti Sains Malaysia, 11800 USM, Penang, Malaysia^{1, 3, 5, 8}
Information Technology Department, AlBuraimi University College, Buraimi, Oman^{1, 4, 6, 7}
AlZahra College for Women, Muscat, Oman²

Abstract—Mobile Ad Hoc Network (MANET) is a group of wireless mobile nodes that can connect with each other over a number of hops without the need for centralized management or pre-existing infrastructure. MANET has been used in several commercial areas such as intelligent shipping systems, ad hoc gaming, and clever agriculture, and non-commercial areas such as army applications, disaster rescue, and wildlife observing domains. One of the main challenges in MANET is routing mobility management which affects the performance of MANET seriously. The routing protocols have been functionally classified into proactive routing protocols, reactive routing protocols, and hybrid routing protocols. The objective of this paper is to create observations about the advantages and disadvantages of these protocols. Thus, the aim of this paper is to conduct a comparative analysis of the three groups of MANET routing protocols by comparing their features and methods in terms of routing overhead, scalability, delay, and other factors. It was shown that the proactive protocols guarantee the availability of the routes. However, it suffers from scalability and overhead. Whereas, reactive protocols initiate route discovery only when data needs to be sent. However, reactive protocols introduce an undesirable delay due to route establishment, which affects the network performance. Hybrid protocols, attempt to utilize the beneficial features of both reactive and proactive protocols, hybrid protocols are suitable for large networks and keep up-to-date information, but they increase operational complexity. It was concluded that MANET needs enhancement with regard to routing in order to meet the required performance.

Keywords—MANET; routing protocols; proactive protocols; reactive protocols; hybrid protocols; Ad Hoc Networks

I. INTRODUCTION

The routing process in MANET is responsible for discovering, establishing, and maintaining a route between two mobile nodes. Routing of packets can be performed using either a single-hop or a multi-hop paradigm [1]. In the single-hop paradigm, the destination node is assumed to be inside the communication range of the source node. Thus, the source node can connect with its destination directly. Within the multi-hop model, the source node can interact with its destination via intermediate nodes while the destination is outside the source node's communication range. MANET is regarded as a multi-hop network where mobile nodes in the network collaboratively help in forwarding the data or control packets between the source node and its destination. The

mobile nodes are involved in the discovery of routes, and once found, the intermediate mobile nodes on the routes would have key roles in maintaining the routes. There are some difficulties in establishing a route between source and destination nodes through intermediate nodes including low bandwidth, limited coverage and connectivity due to limited transmission range, higher error rate, high possibility of interference, power consumption, no centralized mechanism for routing, and frequent network topology changes due to mobility. Mobility makes routing a more complex task in MANET. Routing protocols should be capable of managing routing in MANET efficiently; therefore, it is important to investigate the advantages and disadvantages of the different protocols for MANET to identify the performance evaluation of each routing protocol as the applications of MANET are strongly dependent upon the underlying routing protocol which must be reliable and robust to accommodate frequent disruptions in the communication between mobile node pairs due to node mobility, interference, and lack of infrastructure. These routing protocols can be categorized functionally and structurally based on their routing processes and structures; therefore, the main goal of this work is to perform a comparison between the routing protocol categories with respect to the common parametric evaluation metrics.

The remainder of this paper is structured as follows: Section II describes the general issues and challenges in MANET, routing structure and protocols in MANET are covered in Section III, and Section IV briefly describes the functional classification of routing protocols which includes table-driven, on-demand, and hybrid routing protocols. The structural classification of routing protocols is covered in Section V. The routing protocols are discussed in Section VI with their limitations. The discussion and studies are presented in Section VII. Finally, the conclusion and future works are discussed in Section VIII.

II. GENERAL ISSUES AND CHALLENGES IN MANET

Routing and mobility management are the two key issues with MANET. Routing becomes more challenging due to mobility in MANET, which generally consists of a group of decentralized mobile nodes, that move randomly and frequently causing topology changes [2], [3], [4]. The following are the subsections that summarize the major challenges and issues in MANET.

A. Routing Traffic

Every node in MANET works as a source, a destination, or a router to send data to other nodes. Therefore, mobile nodes are equipped with a discovery function for their environment where a node can forward a message directly to nodes within range or to other unreachable nodes through the intermediate node(s) [5]. The main mechanism used to raise the whole network capacity and performance is multi-hopping. Thus, a node can send data to a specific destination on behalf of another node [6]. This means that even if a source is outside of the destination's radio range, the destination can still receive data from it. Packets travel through numerous wireless nodes to reach their destination. According to Minhas et al. [7], the multi-hopping mechanism helps in conserving energy resource conservation, interference reduction, and increasing the network throughput.

The network can stay operational by constructing new routes by flooding to deliver the data using a multi-hopping mechanism. In the flooding procedure, control packets are moving infinitely in the entire network. As a result, the flooding procedure consumes too much energy from the network resources when it is used for data transmission. Thus, controlling flooding is one of the major challenges to such networks.

B. Nodes Mobility Management

In normal conditions, any two neighboring nodes can exchange data between them (see Fig. 1). Nevertheless, their connection will vanish if any of them leaves the transmission range of the other. Thus, in MANET with high mobility nodes, the probability of a link breaking between any two network neighbors is considerable. This is another significant obstacle to such networks. Due to MANET's dynamic nature, the network topology frequently changes, which therefore results in frequent connection failures [8], [9], [10]. The network must create new routes, just as in the case of broken links, to assure data transmission [11], [12]. A dynamic routing system is required to maintain routes between a source and its destination because of the frequent topology changes.

Therefore, the reliability and success of MANET depend on the effectiveness of the routing protocol and the attribute and usefulness of the collected data [13].

C. Scalability and No Fixed Boundaries

MANET is subject to several challenges such as scalability and no fixed boundaries [14]. MANET is naturally dynamic where mobile nodes arrive and exit arbitrarily without control from a base station (BS) or other central points. Furthermore, as nodes in MANET join and leave arbitrarily, the number of nodes and the size of the network can grow erratically which introduces a heavy burden on the routing mechanism. Consequently, scalability becomes a major issue in MANET [15].

D. Node Density

The density of nodes in regions such as a national or urban park, where high density is presented, compared to highways where the density is varied from high to low depending on rush hour times, should be considered [16] Modeling the mobile

nodes and communication links is one of the problems in MANET. Such modeling can provide valuable information regarding the pattern or behaviors of the wireless transmission under different situations as wireless transmissions in a MANET functioning on a flat open environment can be different from such transmissions in an ad hoc network of nodes placed on a building [17].

The scatter or the distribution of nodes in a geographical area affects the efficiency of routing, especially when there are a lot of middle nodes between the source and the destination. In Fig. 2, where S and D denote the source node and the destination node correspondingly, the light gray area shows the potential flooding and the dark area shows the potential intermediate nodes involved in routing.

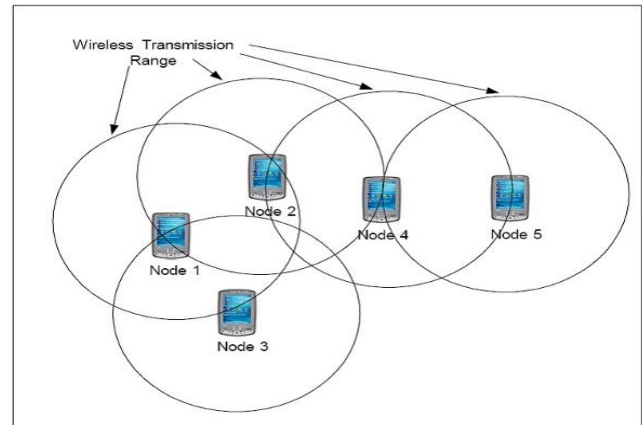


Fig. 1. Communications between adjacent mobile nodes.

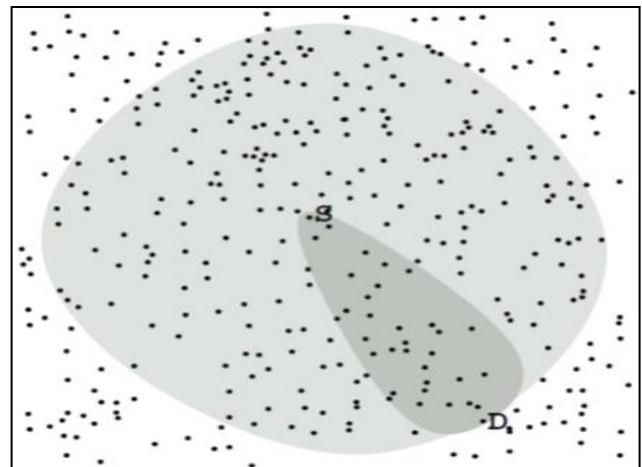


Fig. 2. Big number of nodes between the source and destination.

E. Security Concerns

Besides the technical problems mentioned above, In MANET, where trust relationships must be established, security is a significant problem [18], [19]. It is crucial to note that using several hops can cause a problem because it enables unauthorized individuals to intercept data illegally. In addition, there is intentional electronic interference or unintentional interference occurring while many nodes share the same air interface domain. The major challenges and issues in MANET are shortened in Table I.

TABLE I. MAJOR CHALLENGES AND ISSUES IN MANET

Issue	Description
Routing Traffic	The flooding procedure used in the discovery of new routes consumes too much energy from the network resources
Nodes Mobility Management	Nodes mobility causes frequent link breakages
Scalability and No Fixed Boundaries	As nodes in MANET join and leave arbitrarily, the size of the network can grow erratically which introduces a heavy burden on the routing mechanism
Node Density	The scatter or the distribution of nodes in a geographical area affects the efficiency of routing
Security Concerns	The use of multiple hops can be problematic since it makes it easier for an unauthorized person to intercept data.

III. ROUTING STRUCTURE AND PROTOCOLS IN MANET

The routing process in MANET is responsible for discovering, establishing, and maintaining a route between two mobile nodes. Routing of packets can be performed using either a single-hop or a multi-hop paradigm. In a single-hop paradigm, the destination node is assumed to be within the communication range of the source node. Thus, the source can communicate with its destination directly. Within the multi-hop paradigm, the source node can communicate with its destination through intermediate nodes as the destination is out of the communication range of the source node [20]. MANET is considered a multi-hop network where mobile nodes in the network collaboratively help in forwarding the data or control packets between the source node and its destination. The mobile nodes are involved in the discovery of routes, and once found, the intermediate mobile nodes on the routes would have key roles in maintaining the routes. Therefore, routing protocols should be capable of managing routing in MANET efficiently. There are some difficulties in establishing a route between source and destination nodes through intermediate nodes including low bandwidth, limited coverage and connectivity due to limited transmission range, higher error rate, high possibility of interference, power consumption, no centralized mechanism for routing, and frequent network topology changes due to mobility.

A lot of protocols have been developed for routing in MANET. These routing protocols can be classified functionally and structurally according to their routing processes and structures.

IV. FUNCTIONAL CLASSIFICATION OF ROUTING PROTOCOLS

According to the methods that are used in discovering and maintaining routes, routing protocols in MANET are categorized into three groups; table-driven (proactive) routing protocols, on-demand (reactive) routing protocols, and hybrid routing protocols [4], [14], [21], [22].

A. Table Driven Routing Protocols

Tables-driven protocols also called proactive protocols are developed depending on link state and distance vector routing techniques that are traditionally used on the Internet. The main

characteristic of this type of protocol is that they are proactive in the sense that every mobile node maintains an updated routing table to any other node in the network. Therefore, each node should periodically communicate routing information with other nodes in order to maintain its routing table up-to-date on whether the routes are used or not [23], [24]. The frequency of updating the routing tables is crucial. Even though it can reflect the state of the network accurately, and the routing process would be robust to the dynamic changes in the network; however, the bandwidth usage for exchanging routing information will be high. This would leave not much bandwidth for delivering data packets, which affects throughput at the destination nodes considerably. Furthermore, it causes Broadcast Storm Problems (BSP) [25], [26], as the network will be flooded with routing information updates. Hence, the bandwidth for sending data packets will be reduced significantly; especially, in MANET with high node density. On the other hand, as table-driven protocols ensure that routes to destinations are always available, this would reduce the delay in sending data packets once required. In reaction to network topology changes, each proactive protocol reacts differently according to its routing structure, the size of the routing table, and the frequency of routing information updates.

B. On-demand Routing Protocols

On-demand routing protocols also called reactive routing protocols were developed to improve scalability and overhead problems presented by table-driven routing protocols. The aim is to save bandwidth by reducing the number of control messages sent across the network. Therefore, a route to a destination is only looked up when the higher protocol levels demand it, compared with the periodic search for routes and updating them as with proactive protocols. Subsequently, the routing overhead is decreased significantly, which makes it more suitable for mobile network environments [15]. There are two main processes in reactive routing; which are route detection and route maintenance. When a Source node (S) needs to forward data, it first searches its routing table to examine whether it has a route to the desired Destination (D). If there is no route found, a route detection procedure is generated in order to discover a route to the destination. In route detection, the source node floods the network by broadcasting Route Request (RREQ) packets as shown in Fig. 3 [27]. When the destination or an intermediate node that has an active path to the destination receives the RREQ packet, it broadcasts or unicasts a Route Reply (RREP) back to the source node.

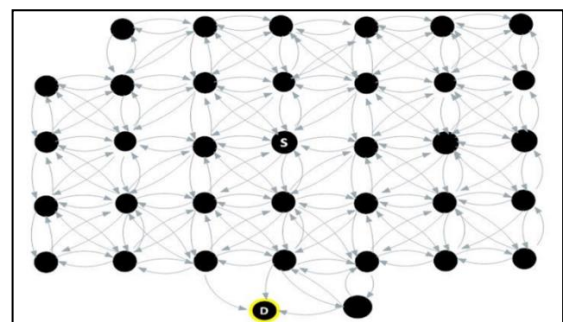


Fig. 3. Route detection in conventional on-demand routing protocol.

The route maintenance process starts when the route that is used currently to transport data is disconnected. The node that detects the route failure may repair the route using its local repair process, or otherwise, forward a Route Error (RERR) packet to the source node which will initiate a new route discovery attempt. The main difference between proactive and reactive routing methods in MANET is revealed in Table II [28].

TABLE II. DIFFERENCE BETWEEN PROACTIVE AND REACTIVE ROUTING METHODS IN MANET

Parameter	Table Driven	On-Demand
Route availability	Constantly available	Calculated when needed
Periodic updates	Always required	Not required
Handling mobility	Updates happen at regular intervals	Use localized route discovery
Control traffic generated	Usually higher than on-demand	Growth with mobility of active routes
Storage requirements	Higher than on-demand	Depends on the number of routes maintained or needed
Scalability	Typically, up to 100 nodes	Frequently higher than table-driven

Reactive routing protocols can be classified into two classes; which are hop-by-hop and source-based routing protocols. Source-based routing methods convey the whole path to the destination, while, hop-by-hop routing protocols hold only the destination and next hop addresses in their data packets header.

C. Hybrid Routing Protocols

Hybrid routing is a combination of distance-vector routing and link-state routing. Thus, hybrid routing protocols share the properties and useful features of both reactive and proactive protocols. These protocols are developed to increase the scalability and improve routing in MANET by determining the optimal routes to a destination and reporting network topology when there is a change only [29]. In cases where connectivity to nearby nodes should be maintained, reactive routing is used, while proactive routing can be used if routes to remote nodes are required. This minimizes the periodic propagation of routing information and may provide accurate and reliable routes for transmitting data packets to their intended destination. Moreover, these protocols are able to reduce the number of rebroadcasting nodes in the network using different hierarchical strategies [30]. These strategies enable the nodes to organize themselves to provide effective routing where only selected nodes are used to perform route discovery. Nevertheless, the disadvantage of these protocols is that their efficiency depends on the number of nodes activated in the network. In addition, the gradient of traffic volume plays an important role in reacting to traffic demand. Compared to reactive or proactive protocols, hybrid routing protocols are naturally more complex and require a high computation level to investigate their performance in large MANET.

V. STRUCTURAL CLASSIFICATION OF ROUTING PROTOCOLS

Based on their routing structures, routing protocols in MANET can be classified into three categories; flat, hierarchical, and geographic position routing protocols [31]. Every protocol in these categories performs routing whether proactively or reactively or both. For example, flat routing protocols can be reactive such as AODV and DSR, or proactive such as DSDV and OLSR. In hierarchical routing protocols such as ZRP, nodes are grouped into zones (cluster-based) or trees that would help in limiting the flooding area during the route discovery process. In hierarchical routing, the group leader is responsible for routing management within its group which can reduce the global exchange of routing information (overhead) and the size of routing tables [32]. In addition, hierarchical routing protocols scale better than flat routing protocols in large MANET. Nonetheless, these protocols cause high overhead in highly dynamic MANET due to the frequent reconstruction of zones and cluster head election [33]. Geographic position routing protocols such as ZHLS require that each mobile node must be equipped with GPS in order to acquire their location information when needed. In geographic position routing protocols, data are sent to all mobile nodes in a particular region using geographical information and routing. Hence, the propagation of routing information to the whole network is obviated. The use of geographical information makes those protocols adjust themselves to topology changes quickly. However, high overhead is introduced due to the mapping of address to location procedure.

VI. LITERATURE REVIEW

This section discussed the previously mentioned routing methods along with their limitations in terms of the route discovery process.

Some of the table-driven routing protocols, like Optimized Link State Routing (OLSR) [34], [35], Mobility based OLSR (Mob-OLSR) [36], [37], and Fisheye State Routing (FSR) [38], [39], are developed based on link-state routing algorithm where nodes maintain link-state cost to their neighbouring nodes [40]. Other routing protocols in this category, such as Destination Sequenced Distance Vector (DSDV) [41] and Wireless Routing Protocol (WRP) [42] developed based on distance vector routing where the shortest paths to a destination are checked and maintained periodically by every node.

DSDV routing protocol [41], which is a table-driven routing mechanism based on the Bellman-Ford algorithm [43], was developed to overcome the routing loop problems based on the sequence number of each route stored in the routing table that is announced by the destination. Hence, the data packets are routed through the route with the most recent sequence number. DSDV requires a consistent update of the routing tables [44] which utilizes some of the bandwidth even when the network is not used, which leads to fast depletion of battery power. DSDV is not appropriate for a very dynamic or large-scale MANET [45] as it needs a new sequence number whenever the network topology changes.

Similar to DSDV, WRP [42] was developed to diminish route loops and confirm reliable message exchange based on

the Bellman-Ford algorithm. WRP preserves an up-to-date view of the network by using a set of tables. Maintaining multiple tables requires a significant amount of memory and greater processing power. As WRP uses hello messages to ensure connectivity with neighbours, in highly dynamic MANET, the control overhead involved in updating tables is high and more bandwidth and energy are consumed. Therefore, WRP is not suitable for large MANET since it suffers from limited scalability issues.

FSR [38], [39] is a link state-based routing protocol that controls the overhead by sending out information about mobile nodes that are within its range only. In FSR, a node maintains the link state for every destination in the network by periodically broadcasting update messages to its neighboring nodes. In addition, route updates related to closer nodes are propagated more frequently. FSR provides good packet delivery when mobility in a MANET is low. However, in highly dynamic MANET where the network topology changes repeatedly, FSR presents inaccurate routing information to the destination which makes it not suitable for large MANET.

OLSR [35], [46] is a proactive link-state routing protocol, which discovers and propagates information using Topology Control (TC) and hello messages. OLSR is a shortest-path first-based algorithm. OSPF (Optimize Shortest Path First) floods the topology data using a reliable algorithm that is not suitable for MANET nature. Accordingly, OLSR is considered an unreliable protocol for a highly dynamic MANET. Also, it does not sense the quality of the route; it just assumes that the route is active if some of the hello packets have been received properly. Furthermore, OLSR uses many network resources i.e., bandwidth and energy that are limited in MANET. It is the same for the enhanced version of OLSR where a new technique for node mobility measurement was proposed by [36].

The common on-demand routing protocols in practice are Dynamic Source Routing (DSR) [47], [48], Ad-hoc On-Demand Distance Vector (AODV) [49], [50], Dynamic MANET On-demand (DYMO) [51], [52], Location-Aided Routing (LAR) [53], and Temporally-ordered routing algorithm (TORA) [54].

DSR [47] is an on-demand reactive routing protocol that uses source routing rather than depending on the routing table information at each intermediate node. DSR has two main mechanisms; which are route discovery and route maintenance. Route discovery is initiated when a node requests a route to a specific destination. Route maintenance is triggered when a link between two nodes that are involved in the active route breaks down. DSR, like other on-demand routing protocols, floods the network with RREQ packets during the route discovery process. In determining the route to a destination, the addresses of the intermediate nodes between the source and destination are accumulated during the route discovery process where each node caches the route information. The learned route is used to transmit data packets that contain the address of each node along the path to the destination. DSR controls the bandwidth consumed by control packets, eliminating the periodic update messages required in proactive routing. Load balancing is achieved by using multiple routes which can

increase robustness as well DSR is beacon-less. Moreover, although DSR performs well in networks with low mobility, its performance degrades significantly in highly dynamic networks [55], [56], [57]. Furthermore, its route maintenance strategy does not locally repair a broken route. If routes in the cache are stale, it can cause incompatibility when the route is reconstructed. Also, the delay in establishing a connection is higher compared to that of table-driven protocols.

AODV [48], [58] is a hop-by-hop reactive routing protocol that broadcasts discovery packets only when needed. AODV applies destination sequence numbers to find the latest route to the destination, which also helps in avoiding the infinite loops problem. In addition to that, the delay of the connection establishment in AODV is lower. Overheads and contention are reduced since AODV maintains only active routes. However, having an old source sequence number in the intermediate nodes can lead to unreliable routes. Also, heavy control overhead can be caused by multiple RREP packets in response to a single RREQ and due to the use of periodic "HELLO" packets route maintenance. Furthermore, AODV uses periodic beaconing to keep routing tables updated, which results in unnecessary bandwidth consumption. Moreover, AODV shows better performance in terms of throughput and delay in small-size MANET with no node mobility; and in dense networks with minimum mobility [59]. However, the quality of its performance decreases as the node mobility increases. The mobility-aware approach was added to AODV [60] to improve the management of high mobility in MANET by avoiding the frequent link breakages associated with using unstable paths that contain high mobile nodes. This added feature has shown some enhancement compared to AODV [61].

DYMO (also known as AODVv2) [51] is designed for dynamic environments such as MANET where network topology changes frequently. DYMO shares many benefits of the operational structure of DSR and AODV. DYMO outperforms AODV and DSR protocols as it uses accumulative routing which reduces RREQs noticeably [62]. DYMO was improved by considering the energy and traffic parameters of the network and showed better performance compared to the original DYMO and AODV routing protocols [63].

LAR [53] is an on-demand routing protocol that uses geographical location information to limit the propagation of RREQ packets to a certain number of nodes rather than flooding the network, which in turn reduces the routing overhead considerably. The location of nodes is detected using Global Positioning System (GPS) information and defined in an area that is called a "Request Zone". Only nodes in this zone are required to forward RREQ packets. This can help in avoiding the broadcast storm [20]. However, connection and tracking problems may appear with the use of LAR. It is because when a source node has to find a route to a destination, it should first get the coordinates of the destination from an external location service. LAR has been enhanced further to improve the link ability during routing [64], and to control the overhead found in LAR [65].

TORA [54] is an adaptive routing protocol designed to restrict the propagation of control messages in the highly

dynamic context of mobile computing. In TORA, each node has to explicitly start a query when it needs to forward data to a specific destination. TORA tries to figure out what is known as a Directed Acyclic Graph (DAG) which is rooted at the destination. Even though TORA performs well in dense networks, it does not scale by any means. Several evaluation studies showed that DSR and AODV outperform TORA [66], [67]. It was enhanced to provide better packet delivery, and acceptable routing overhead and packet latency [66].

Location Update Routing Protocol (SLURP) [68], Zone Routing Protocol (ZRP) [40],[69] and Zone-based Hierarchical Link State (ZHL) [70],[71] routing protocols are among the common hybrid routing protocols served the base for the development of several protocols proposed later.

SLURP [68] uses GPS instead of a cluster head to manage the node location and coordinate the transmission of data packets. It utilizes the identification (ID) of the node and zone ID of the destination to perform routing. Therefore, SLURP shares the same advantages mentioned above. Moreover, it limits the need for flooding as the nodes within the zone maintain location information with each other. Thus, nodes know how to find an efficient route to destinations when required. SLURP limits the overhead of maintaining routing information further. This is achieved by restricting route discovery to the home region or specific zone assigned to each node in the network. The home region is determined by a static mapping function that is known to all nodes in the network. The drawback of SLURP is that it depends on a predefined static zone map.

ZRP [40], [69] was developed to speed up data delivery and reduce processing overhead. In ZRP, mobile nodes are clustered into zones where communications among nodes are performed according to their locations in the zone. ZRP maintains robust network connectivity within the routing zones using the reactive routing technique. Also, it reactively discovers remote routes faster. Nonetheless, ZRP behaves as a proactive protocol if the routing zone is too large. On the other hand, if the routing zone is too small, ZRP performs as a reactive protocol. Thus, it is important to set the value of the zone radius according to the density of nodes in the network.

In ZHLS [70], the nodes are divided into non-overlapping zones where each node is associated with two identifiers which are a node ID and a zone ID that is calculated using GPS. Traffic bottleneck is avoided in ZHLS as it does not require a cluster head to coordinate data transmission. Therefore, there is no processing overhead necessary for electing the cluster and restructuring the zone in case of a single point of failure. Hence, the communication overhead is reduced significantly, compared to the flooding method in reactive protocols. Furthermore, ZHLS can adapt to the changes in network topology faster as it only requires the node ID and the zone ID of the destination when routing data packets. However, in ZHLS, for a node to function, it should have a static zone map, which is not practicable in networks like MANET where the geographical boundary is dynamic.

VII. DISCUSSION AND STUDIES

This section discusses the advantages and disadvantages of the routing methods mentioned earlier and presents a comparison between different types of MANET routing protocols in terms of the common parametric evaluation metrics.

Proactive routing protocols have been evaluated theoretically and through simulation [23],[36],[44],[72],[73],[74],[75] and it was found that the main advantage of the proactive routing protocol is to ensure the availability of routes whenever needed with a minimum delay of data delivery, as every node should maintain routing information to every node in the network. The main disadvantages of this type of routing are the continuous discovery of routes and the broadcast of routing information which introduces high overhead and consumes high energy and bandwidth. Therefore, table-driven routing protocols are not appropriate for large and highly dynamic MANET since every node is required to maintain entries in the routing table about all nodes in the network. Because of the nature of MANET, a routing protocol designed for such networks should improve the scalability and decrease the routing overhead by restraining route computations to situations when a route is needed.

The evaluation of reactive routing protocols [17], [66],[76],[77] showed that reactive routing introduces lower overhead as loop free since routes are only constructed when required, which is a privilege of this type of the protocol compared to proactive routing. The disadvantages of reactive routing are that, due to the initial route discovery process, there is a critical delay between the time a source node requests a route for data transmission and the time when the actual transmission takes place. The source node must wait until a route is found then it can start transmitting its data. Rapid changes in a MANET topology due to mobility may break active routes and cause subsequent route discoveries which can substantially impact the network's performance. Additionally, the flooding technique utilized throughout the route discovery phase can result in a broadcast storm.

The performance evaluation of hybrid routing protocols [2],[29],[71],[78],[79] showed that, in comparison with reactive protocols, hybrid protocols can reduce the average length of the routes in terms of the number of nodes and the physical length of the route as well. It was found that the overhead cost in hybrid routing is tolerable in most of the evaluation scenarios. However, even though that hybrid protocols are suitable for large networks and make available up-to-date information, they increase operational complexity.

Table III [80] presents a methodological comparison between different categories of MANET routing protocols with respect to the corresponding parametric evaluation.

TABLE III. COMPARISON BETWEEN ROUTING PROTOCOLS CATEGORIES

Parameter	Reactive Routing Protocols	Proactive Routing Protocols	Hybrid Routing Protocols
Routing Structure	Flat	Flat, Hierarchical	Flat Hierarchical, Geographic

Routing Method	On-demand	Table driven	Both
Routing Overhead	Low	High	Medium
Delay	High-as a result of flooding	Low as a result of routing tables	Low inside the zone; high outside the zone
Scalability	Inappropriate for large-size MANET	Low	Suitable for large-size MANET
Route Availability	Required when needed	Always available in routing tables	Both
Periodic Updates	Not required	Required for updating on topology changes	Required inside the zone only
Storage	Low-depending on the routes number	The high-routing table can be large	High inside the zone
Mobility Management	Route maintenance	Periodic updates	Both

VIII. CONCLUSION AND FUTURE WORKS

This paper discusses that data packet routing is the main concern to improve the performance of MANET where nodes move arbitrarily with no central administration. This presents a heavy burden on the routing protocol in use. In regards to a functional classification of routing protocols discovering and maintaining routes are considered, the protocols have been classified into table-driven (proactive) routing protocols, on-demand (reactive) routing protocols, and hybrid routing protocols. It was shown that in proactive protocols, nodes should maintain a routing table for forwarding packets to any other node in the network. This enforces periodic exchange of information between mobile nodes to keep their routing table up-to-date. In this type of protocol, scalability and overhead become serious issues. Whereas, reactive protocols initiate route discovery only when there is data to send. However, it was recognized that such a process introduces an undesirable delay between the request for data transmission and the actual transmission of data before route establishment, which affects the network performance. Moreover, the flooding procedure used in the route discovery process can cause broadcast storms. Hybrid protocols, however, attempt to utilize the beneficial features of both reactive and proactive protocols to tackle these problems. Nonetheless, it was concluded that while hybrid protocols are suitable for large networks and keep up-to-date information, they increase operational complexity.

In this paper, in connection with the routing methods mentioned earlier, the routing protocols are also classified based on the structure of the network into flat, hierarchical, and geographical position routing protocols, along with a discussion on their performance in terms of some common evaluation metrics. It was concluded that MANET needs enhancement with regard to routing in order to satisfy the service quality requirements of the user applications with desirable performance. The insights gathered from this study will be useful to researchers, network designers, and professionals that work in this area as they design and optimize future MANETs. Future research should include recommendations for selecting the best routing protocol for various scenarios and conduct a comparative analysis of additional routing protocols, including their advantages and disadvantages. Furthermore, future research could be the

investigation on how routing protocols in MANETs can be improved using machine learning and artificial intelligence techniques.

REFERENCES

- [1] B. U. I. Khan et al., "Exploring Manet Security Aspects: Analysis of Attacks and Node Misbehaviour Issues," *Malaysian J. Comput. Sci.*, vol. 35, no. 4, pp. 307–338, 2022.
- [2] D. Bhatia and D. P. Sharma, "A comparative analysis of proactive, reactive and hybrid routing protocols over open source network simulator in mobile ad hoc network," *Int. J. Appl. Eng. Res.*, vol. 11, no. 6, pp. 3885–3896, 2016.
- [3] V. Polara and J. M. Rathod, "A study of mobile ad hoc network and its performance optimization algorithm," in *Computer Networks, Big Data and IoT: Proceedings of ICCBI 2020*, 2021, pp. 131–142.
- [4] U. Srilakshmi, N. Veeraiha, Y. Alotaibi, S. A. Alghamdi, O. I. Khalaf, and B. V. Subbayamma, "An improved hybrid secure multipath routing protocol for MANET," *IEEE Access*, vol. 9, pp. 163043–163053, 2021.
- [5] B. Victor, "Descriptions of Mobile Ad-hoc Networks," 2016.
- [6] M. K. Suthar and A. K. Vyas, "Design and Development of Hybrid Routing Protocol Algorithm with Attack Detection & Protection Mechanism for Mobile Ad-hoc Network Application." Gujarat Technological University Ahmedabad, 2022.
- [7] Q.-A. Minhas, H. Mahmood, and H. Malik, "Incentive driven cooperation to avoid packet loss in multihop ad hoc networks," in 2012 International Conference on Emerging Technologies, 2012, pp. 1–6.
- [8] A. Goncalves, C. Silva, and P. Morreale, "Design of a mobile ad hoc network communication app for disaster recovery," in 2014 28th International Conference on Advanced Information Networking and Applications Workshops, 2014, pp. 121–126.
- [9] S. Dalal et al., "An adaptive traffic routing approach toward load balancing and congestion control in Cloud-MANET ad hoc networks," *Soft Comput.*, vol. 26, no. 11, pp. 5377–5388, 2022.
- [10] B. H. Khudayer, M. M. Kadhum, and T.-C. Wan, "Topology-Aware Mechanism to Improve Routing in Mobile Ad Hoc Networks," in *Advances in Machine Learning and Signal Processing: Proceedings of MALSIP 2015*, 2016, pp. 13–24.
- [11] V. G. Menon and P. M. Joe Prathap, "Routing in highly dynamic ad hoc networks: issues and challenges," *Int. J. Comput. Sci. Eng.*, vol. 8, no. 4, pp. 112–116, 2016.
- [12] B. H. Khudayer, M. M. Kadhum, and W. T. Chee, "Multi-disjoint Routes Mechanism for Source Routing," *Inf. Sci. Appl.*, vol. 339, no. 23, pp. 189–204, 2015.
- [13] V. K. Quy, N. T. Ban, V. H. Nam, D. M. Tuan, and N. D. Han, "Survey of Recent Routing Metrics and Protocols for Mobile Ad-Hoc Networks," *J. Commun.*, vol. 14, no. 2, pp. 110–120, 2019.
- [14] A. Sethi, S. Suthar, V. Yadav, and A. Kumar, "A survey of QoS multicast protocols for MANETs," *J. Netw. Commun. Emerg. Technol.*, vol. 6, no. 3, pp. 77–81, 2016.
- [15] L. R. Raju and C. Reddy, "A survey on routing protocols and QoS in mobile ad hoc networks (MANETs)," *Indian J. Sci. Technol.*, vol. 10, no. 8, 2017.
- [16] K. Konte, *Mobile ad hoc networks in transportation data collection and dissemination*. Rowan University, 2019.
- [17] D. Ramphull, A. Mungur, S. Armoogum, and S. Pudaruth, "A review of mobile ad hoc NETWORK (MANET) Protocols and their Applications," in 2021 5th international conference on intelligent computing and control systems (ICICCS), 2021, pp. 204–211.
- [18] N. Sivapriya and R. Mohandas, "Analysis on Essential Challenges and Attacks on MANET Security Appraisal," *J. Algebr. Stat.*, vol. 13, no. 3, pp. 2578–2589, 2022.
- [19] T. A. Alamiydy, M. F. R. Anbar, B. Belaton, A. H. Kabla, and B. H. Khudayer, "Ensemble feature selection approach for detecting denial of service attacks in RPL networks," in *Advances in Cyber Security: Third International Conference, ACeS 2021, Penang, Malaysia, August 24–25, 2021, Revised Selected Papers 3*, 2021, pp. 340–360.

- [20] S. Paul, Introduction to MANET and Clustering in MANET. Anchor Academic Publishing, 2016.
- [21] M. S. Avhankar, D. J. A. Pawar, S. Majalekar, and S. Kedari, "Mobile Ad Hoc Network Routing Protocols Using OPNET Simulator," *Int. J. Recent Innov. Trends Comput. Commun.*, vol. 10, no. 1, pp. 1–7, 2022.
- [22] Y. H. Jazyah, "Enhancing the Performance of Wireless Routing Protocols of MANET using AI," *J. Comput. Sci.*, vol. 17, no. 10, pp. 953–959, 2021.
- [23] V. Pondwal and H. Saini, "A Comprehensive Survey on Routing Schemes for High Speed Networks.," *IUP J. Telecommun.*, vol. 8, no. 4, 2016.
- [24] V. Kumar and S. Singla, "Performance Analysis of Optimized ACO-AOMDV Routing Protocol with AODV and AOMDV in MANET," in *Advances in Computing and Data Sciences: 6th International Conference, ICACDS 2022, Kurmool, India, April 22–23, 2022, Revised Selected Papers, Part I, 2022*, pp. 415–425.
- [25] B. Bachir, H. Ahmed, and G. Zouhair, "Topology Lightening and Forwarders Aggregation for Routing Based Network Flooding," *Int. J. Comput. Sci. Inf. Secur.*, vol. 14, no. 12, 2016.
- [26] N. Akhtar, M. A. Khan, A. Ullah, and M. Y. Javed, "Congestion avoidance for smart devices by caching information in MANETS and IoT," *IEEE Access*, vol. 7, pp. 71459–71471, 2019.
- [27] B. H. Khudayer, M. M. Kadhum, and T.-C. Wan, "Feasible Progressive Source Routing-Based Strategy for Mobile Ad-hoc Networks," *Adv. Sci. Lett.*, vol. 22, no. 10, pp. 2628–2632, 2016.
- [28] N. S. Yadav and R. P. Yadav, "Performance comparison and analysis of table-driven and on-demand routing protocols for mobile ad-hoc networks," *Int. J. Electron. Commun. Eng.*, vol. 2, no. 12, pp. 2809–2817, 2008.
- [29] G. A. Walikar and R. C. Biradar, "A survey on hybrid routing mechanisms in mobile ad hoc networks," *J. Netw. Comput. Appl.*, vol. 77, pp. 48–63, 2017.
- [30] R. Kochher and R. Mehta, "Performance analysis of reactive AODV and DSR with Hybrid GRP Routing Protocols under IEEE 802.11 g MANET," in *2016 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*, 2016, pp. 1912–1916.
- [31] D. E. M. Ahmed and O. O. Khalifa, "A comprehensive classification of MANETs routing protocols," 2017.
- [32] S. Sisodia and S. Raghwanishi, "Performance evaluation of a table driven and on-demand routing protocol in energy constraint MANETs," in *2013 International Conference on Computer Communication and Informatics*, 2013, pp. 1–8.
- [33] D. E. M. Ahmed, "Enhanced Scheme for Video Transmission over Mobile Ad hoc Networks based on Optimized Link State Routing Protocol." Sudan University of Science and Technology, 2020.
- [34] T. Clausen and P. Jacquet, "RFC3626: Optimized link state routing protocol (OLSR)." RFC Editor, 2003.
- [35] A. Guillen-Perez, A.-M. Montoya, J.-C. Sanchez-Aarnoutse, and M.-D. Cano, "A comparative performance evaluation of routing protocols for flying Ad-Hoc networks in real conditions," *Appl. Sci.*, vol. 11, no. 10, p. 4363, 2021.
- [36] A. Ouacha, N. Lakki, and A. Habbani, "OLSR protocol enhancement through mobility integration," in *2013 10th IEEE INTERNATIONAL CONFERENCE ON NETWORKING, SENSING AND CONTROL (ICNSC)*, 2013, pp. 17–22.
- [37] O. Barki, Z. Guennoun, and A. Addaim, "Improving the selection of MPRs in OLSR protocol: a survey of methods and techniques.," *Int. J. Electr. Comput. Eng.*, vol. 10, no. 1, 2020.
- [38] M. Gerla, "Fisheye state routing protocol (FSR) for ad hoc networks," Draft. txt, 2002.
- [39] A. H. Mohsin, "Optimize Routing Protocol Overheads in MANETs: Challenges and Solutions: A Review Paper," *Wirel. Pers. Commun.*, vol. 126, no. 4, pp. 2871–2910, 2022.
- [40] E. Kurode, N. Vora, S. Patil, and V. Attar, "MANET routing protocols with emphasis on zone routing protocol—an overview," in *2021 IEEE Region 10 Symposium (TENSYP)*, 2021, pp. 1–6.
- [41] K. L. Arega, G. Raga, and R. Bareto, "Survey on performance analysis of AODV, DSR and DSDV in MANET," *Comput. Eng. Intell. Syst.*, vol. 11, no. 3, pp. 23–32, 2020.
- [42] K. Prabha, "Performance assessment and comparison of efficient ad hoc reactive and proactive network routing protocols," *SN Comput. Sci.*, vol. 1, pp. 1–7, 2020.
- [43] D. Sinwar, N. Sharma, S. K. Maakar, and S. Kumar, "Analysis and comparison of ant colony optimization algorithm with DSDV, AODV, and AOMDV based on shortest path in MANET," *J. Inf. Optim. Sci.*, vol. 41, no. 2, pp. 621–632, 2020.
- [44] N. Aggarwal, T. S. Chohan, K. Singh, R. Vohra, and S. Bahel, "Relative analysis of AODV & DSDV routing protocols for MANET based on NS2," in *2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT)*, 2016, pp. 3500–3503.
- [45] M. A. Shyaa, H. M. Ibrahim, A. Meri, M. Dauwed, and A. Flayh, "The impact of node density over routing protocols in manet by using NS-3," *J. Adv. Res. Dyn. Control Syst.*, vol. 11, no. 5 Special Issue, pp. 184–191, 2019.
- [46] A. Kurniawan, P. Kristalina, and M. Z. S. Hadi, "Performance analysis of routing protocols AODV, OLSR and DSDV on MANET using NS3," in *2020 International Electronics Symposium (IES)*, 2020, pp. 199–206.
- [47] D. Johnson, Y. Hu, and D. Maltz, "The dynamic source routing protocol (DSR) for mobile ad hoc networks for IPv4," 2007.
- [48] B. H. Khudayer, M. Anbar, S. M. Hanshi, and T.-C. Wan, "Efficient route discovery and link failure detection mechanisms for source routing protocol in mobile ad-hoc networks," *IEEE Access*, vol. 8, pp. 24019–24032, 2020.
- [49] C. Perkins, E. Belding-Royer, and S. Das, "RFC3561: Ad hoc on-demand distance vector (AODV) routing." RFC editor, 2003.
- [50] N. I. S. Ramli, S. I. Hisham, N. S. N. Ismail, and M. Ramalingam, "Performance comparison between AODV and DSR in mobile ad-hoc network (MANET)," in *2021 International Conference on Software Engineering & Computer Systems and 4th International Conference on Computational Science and Information Management (ICSECS-ICOCOSIM)*, 2021, pp. 217–221.
- [51] S. Ratliff, J. Dowdell, and C. Perkins, "Dynamic MANET on-demand (AODVv2) routing," *Internet Draft*.(2013), 1-60, 2013.
- [52] S. M. Alkahtani and F. Alturki, "Performance Evaluation of Different Mobile Ad-hoc Network Routing Protocols in Difficult Situations," *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 1, 2021.
- [53] B. K. Tripathy, S. K. Jena, P. Bera, and S. Das, "An adaptive secure and efficient routing protocol for mobile ad hoc networks," *Wirel. Pers. Commun.*, vol. 114, pp. 1339–1370, 2020.
- [54] P. M. Thakrar, V. Singh, and K. Kotecha, "Improved route selection algorithm based on TORA over mobile adhoc network," *J. Discret. Math. Sci. Cryptogr.*, vol. 23, no. 2, pp. 617–629, 2020.
- [55] B. H. Khudayer and M. M. Kadhum, "Reliability of dynamic source routing in heterogeneous scalable mobile ad hoc networks," in *2014 IEEE International Conference on Communication, Networks and Satellite (COMNETSAT)*, 2014, pp. 71–79.
- [56] B. ÖZYURT, İ. DOĞRU, and M. AKCAYOL, "Analyzing the Route Cache Timeout Parameter of DSR Protocol in Mobile Ad Hoc Networks," *Int. J. Comput. Commun. Eng.*, vol. 6, no. 1, 2017.
- [57] B. H. KHUDAYER et al., "An optimizing rebroadcast mechanism for minimizing the control overhead in mobile ad-hoc networks," *J. Theor. Appl. Inf. Technol.*, vol. 100, no. 8, 2022.
- [58] I. D. Chakeres and C. E. Perkins, "Dynamic MANET on-demand routing protocol," *ETF Internet Draft*. Draft. txt, 2008.
- [59] K. M. Omran, "The routing control in mobile ad hoc network using intelligent optimization algorithms," in *2020 International Conference on Electrical, Communication, and Computer Engineering (ICECCE)*, 2020, pp. 1–6.
- [60] Y. Khamayseh, O. M. Darwish, and S. A. Wedian, "Ma-aodv: Mobility aware routing protocols for mobile ad hoc networks," in *2009 Fourth International Conference on Systems and Networks Communications*, 2009, pp. 25–29.

- [61] A. Hinds, M. Ngulube, S. Zhu, and H. Al-Aqrabi, "A review of routing protocols for mobile ad-hoc networks (manet)," *Int. J. Inf. Educ. Technol.*, vol. 3, no. 1, p. 1, 2013.
- [62] S. Hasdemir, S. Yilmaz, and S. Sen, "A novel multi-featured metric for adaptive routing in mobile ad hoc networks," *Appl. Intell.*, vol. 49, pp. 2823–2841, 2019.
- [63] M. C. Aravind, C. P. Sangeetha, and C. D. Suriyakala, "Enhanced dynamic MANET on-demand (En-DYMO) routing protocol for mobile adhoc networks," in *2015 Global Conference on Communication Technologies (GCCT)*, 2015, pp. 544–549.
- [64] R. Aboki, E. Shaghghi, P. Akhlaghi, and R. M. Noor, "Predictive location aided routing in mobile ad hoc network," in *2013 IEEE 11th Malaysia International Conference on Communications (MICC)*, 2013, pp. 57–61.
- [65] A. Veerasamy and S. S. R. Madane, "An improved opportunistic location aided routing (iolar) in mobile ad hoc networks," in *Second International Conference on Current Trends In Engineering and Technology-ICCTET 2014*, 2014, pp. 367–370.
- [66] J. L. E. K. Fendji and S. D. Samo, "Energy and performance evaluation of reactive, proactive, and hybrid routing protocols in wireless mesh network," *arXiv Prepr. arXiv1903.06875*, 2019.
- [67] K. Thamizhmaran and A. Charles, "Comparison of On-Demand Routing Protocol for MANET using Simulation," *i-Manager's J. Commun. Eng. Syst.*, vol. 11, no. 1, p. 13, 2022.
- [68] T. K. Saini and S. C. Sharma, "Prominent unicast routing protocols for Mobile Ad hoc Networks: Criterion, classification, and key attributes," *Ad Hoc Networks*, vol. 89, pp. 58–77, 2019.
- [69] R. Shanthi and T. Padma, "A zone routing protocol incorporated with sleep scheduling for MANETs," *J. Ambient Intell. Humaniz. Comput.*, vol. 12, pp. 4181–4191, 2021.
- [70] M. Sharma, M. Singh, K. Walia, and K. Kaur, "Comprehensive Study of Routing Protocols in Adhoc Network: MANET," in *2019 IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, 2019, pp. 792–798.
- [71] A. M. Abdullah, "Mobile Ad hoc Networks: A Survey of Existing Mobility Models and Routing Protocols," 2021.
- [72] D. Kaur and N. Kumar, "Comparative analysis of AODV, OLSR, TORA, DSR and DSDV routing protocols in mobile ad-hoc networks," *Int. J. Comput. Netw. Inf. Secur.*, vol. 5, no. 3, p. 39, 2013.
- [73] S. R. Inamdar and R. M. Yadahalli, "Paradigm shift in routing approaches for high speed MANET applications," *Eur. J. Eng. Technol. Res.*, vol. 2, no. 1, pp. 59–64, 2017.
- [74] B. Memić, A. Čolaković, A. H. Džubur, and E. Avdagić-Golub, "Efficiency-complexity evaluation methods of routing algorithms in mobile ad hoc networks," *Sci. Eng. Technol.*, vol. 1, no. 2, pp. 24–31, 2021.
- [75] S. Kumar and D. Sharma, "Performance Evaluation of Routing Protocols in MANETs with Variation in Pause Time," in *Optical and Wireless Technologies: Proceedings of OWT 2020, 2022*, pp. 105–113.
- [76] K. Rampurkar, N. Lavande, S. Shilgire, and S. N. Mane, "Study of routing overhead and its protocols," *Int. J. Adv. Eng. Manag.*, vol. 2, no. 2, pp. 52–55, 2017.
- [77] L. Enciso, P. Quezada, J. M. Fernandez, B. Figueroa, and V. Espinoza, "Analysis of Performance of the Routing Protocols Ad Hoc using Random Waypoint Mobility Model Applied to an Urban Environment," in *WEBIST (1)*, 2016, pp. 208–213.
- [78] J. Huang, X. Fan, X. Xiang, M. Wan, Z. Zhuo, and Y. Yang, "A clustering routing protocol for mobile ad hoc networks," *Math. Probl. Eng.*, vol. 2016, 2016.
- [79] A. K. Yadav, "Comparative study of routing protocols in MANET," *Orient. J. Comput. Sci. Technol.*, vol. 10, no. 1, pp. 174–179, 2017.
- [80] R. Kaur and M. K. Rai, "A novel review on routing protocols in MANETs," *Undergrad. Acad. Res. J.*, vol. 1, no. 1, pp. 103–108, 2012.

Deep Learning for Combined Water Quality Testing and Crop Recommendation

Tahani Alkhudaydi, Maram Qasem Albalawi, Jamelah Sanad Alanazi, Wejdan Al-Anazi, Rahaf Mansour Alfarshouti
Faculty of Computers and Information Technology
University of Tabuk, Tabuk, Saudi Arabia

Abstract—The field of agriculture and its specifics has been gaining more attention nowadays due to the limited present resources and the continuously increasing need for food. In fact, agriculture has benefited greatly from the advancements of artificial intelligence, namely, Machine Learning (ML). In order to make the most of a crop field, one must initially plan on what crop is best for planting in this particular field, and whether it will provide the necessary yield. Additionally, it's very important to constantly monitor the quality of soil and water for irrigation of the selected crop. In this paper, we are going to rely on Machine Learning and data analysis to decide the type of crop that we will use, and the quality of soil and water. To do so, certain parameters must be taken into consideration. For choosing the crop, parameters such as sun exposure, humidity, soil pH, and soil moisture will be taken into consideration. On the other hand, water pH, electric conductivity, content of minerals such as chloride, calcium, and magnesium are among the parameters taken into consideration for water quality classification. After acquiring datasets for crop and water potability, we implemented a deep learning model in order to predict these two features. Upon training, our neural network model achieved 97% accuracy for crop recommendation and 96% for water quality prediction. However, the SVM model achieves 96% for crop recommendation and 92% for water quality prediction.

Keywords—Deep learning; irrigation; artificial intelligence; soil moisture

I. INTRODUCTION

As humans became more advanced, they learned that plants do not only provide necessary food for both humans and animals, but it also plays a very important role in the fields of medicine, energy production, and the wellbeing of the entire planet. Agriculture is one of the most essential needs for humans to maintain a sustainable livelihood. Without agriculture, humans would not be getting the sufficient nutrients in their meals, and livestock will not have food to eat which means soon cows and sheep will no longer survive [1]. Thus, it is important to play close attention to crops and to constantly monitor their needs. One of the important needs is water, and not just any water. Irrigation water must have certain qualities in order to be used for planting the crops. So, a water quality assessment or classification system is necessary. In addition, it is essential to know which crops grow best in which environmental conditions, which means a crop prediction system becomes necessary as well.

In many fields such as aquaculture, livestock production, and food industry, water is a critical raw material. For this reason, not any kind of water can be used in any field. To illustrate, not all water is good enough for drinking, or watering plants, etc. [2]. Chatterjee described four different water qualities which are palatable water, infected water, potable water, and contaminated water. From these four types, only palatable and potable water are useable. The water can be classified based on several parameters that have different effects on the water quality.

Generally speaking, the water parameters can be divided into three categories: physical parameters, biological parameters, and chemical parameters as shown in Fig. 1. The physical parameters include total suspended solids (TSS), temperature, electric conductivity (EC), and turbidity. The biological parameters consider whether the water contains any microorganisms. On the other hand, the chemical parameters include Sulfate, pH, heavy metals, and total nitrogen [3].

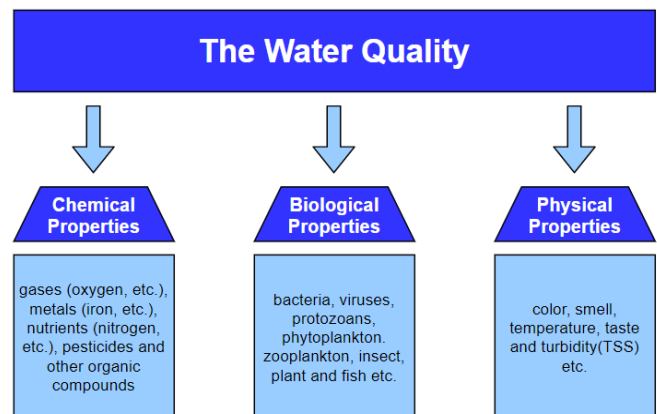


Fig. 1. Water quality parameters.

Predicting the crop is critical for sustainable intensification and efficient use of natural resources [4]. There are many factors that play a role and influence the yield production of crops including environmental conditions and management [5]. In Fig. 2, some of the factors are soil conditions such as pH and moisture play a role in determining the yield of a crop, in addition to weather conditions such as humidity, temperature, and rainfall. Furthermore, factors such as the genotype of the plant, the implemented water irrigation systems, and pesticide control also contribute greatly to how much yield a plant will produce [6].

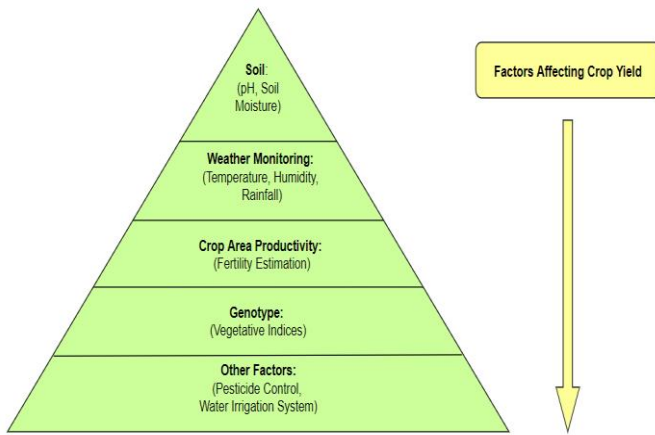


Fig. 2. Factors that influence the crop.

In order to achieve the objectives of the study, we first have to mention the problems that we are trying to address. As mentioned previously, two of the major issues or challenges in the proper maintenance of an agricultural field are the quality of irrigation water that is reaching the plants and that the plants need for proper growth, as well as the appropriate crop type since crops don't react and grow the same as each other in different climates and conditions. Thus, we propose a model that can attempt to solve the water quality analysis and crop recommendation problems at the same time. The contributions that we offer in our study can be summed as follows:

- We developed a system that can predict whether the available water is suitable for irrigation or not, and at the same time can perform crop recommendation based on crop prediction.
- Our model can provide reliable and accurate recommendations for farmers.
- The model that we propose is an inexpensive solution to solve the common problems that farmers face such as low crop.

In summary, this paper aims to address the challenges faced by farmers in selecting the appropriate crop type and evaluating the quality of irrigation water. We propose a model that utilizes deep learning and machine learning algorithms to predict the suitability of available water for irrigation and recommend the best crop type based on crop prediction. Our model offers a cost-effective solution to the common problems faced by farmers, providing reliable and accurate recommendations. The paper also includes a literature review on water quality studies and crop recommendation datasets, as well as a detailed description of the dataset used in our model. Additionally, we discuss the factors that influence crop growth and the contributions of our study.

II. LITERATURE REVIEW

Due to the importance of correctly selecting a crop type, and to continuously evaluate the quality of irrigation water, there's a huge pool of studies that revolve around these two topics.

A. Water Quality Studies

El Bilali et al. [7] designed and implemented a total of 8 Machine Learning models to statistically predict the irrigation water quality (IWQ) parameters that make water suitable for irrigation. Ten irrigation water quality parameters, namely Kelly ration (KR), sodium absorption ratio (SAR), adjusted SARA, Sodium percentage ($\text{Na}^+\%$), exchangeable sodium percentage (ESP), residual sodium carbonate (RSC), total dissolved solids (TDS), chloride (Cl^-), permeability index (PI), and magnesium absorption ratio (MAR) were measured by analysis of 264 samples that were gathered from the Bouregreg watershed, as well as 29 samples from Cherrate and 35 samples from Nfifikh watersheds in Morocco. As for the Machine Learning algorithms, the chosen ones were: multiple linear regression, artificial neural network (ANN), decision tree (DT), random forest (RF), K-nearest neighbor (KNN), support vector machine/regression (SVM/R), stochastic gradient descent (SGD), and adaptive boosting. Upon testing, the results show that all of the 8 models except for SVR and KNN, are capable of predicting only 8 IWQ parameters through the use of electrical conductivity and pH as input variables in Bouregreg watershed surface. To further confirm the results, the six validated ML models were generalized to the Cherrate and Nfifikh watersheds. The results of this generalization attempt revealed that the previous models can be generalized for three parameters in Cherrate watershed and 4 in Nfifikh watershed. Some of these models were not able to statistically predict the MAR and the PI, possibly because of the poor relationship between the EC and pH input variables and these two parameters. Additionally, it was revealed that the adaptive boosting model achieves better performance in comparison with other models in Bouregreg watershed. Thus, it is possible to confirm that ML models can help farmers to better manage the irrigation water quality through extensive analysis.

Ali Mokhtar et al. [8] purposed to study the quality index of the irrigation water of Bahr El-Baqr drain. The authors gathered data from the analysis of 105 water samples of 1L each at a depth of 1m collected from Bahr El-Baqr during the month of July in 2020, ten features were taken into consideration, including pH level, electrical conductivity, sodium concentration, potassium concentration, Ca^{2+} , Mg^{2+} , chloride, carbonate, bicarbonate and sulfate composition. These data were cleaned and modified in order to reach the best score and the most accurate prediction possible. Multiple regressions including principal component regression, stepwise regression, partial least squares regression and ordinary least squares regression, in addition to Machine Learning methods including random forest, extreme gradient boosting, and support vector machine were applied in order to go through the features and determine the features most responsible of identifying the quality index. After applying the root mean square error, the best performance was for the stepwise regression with the values of 0.21% and 0.03%. While after applying the scatter index, all models gave values less than 0.1% except RSC.

B. Crop Recommendation Studies

Bandara et al. [9] proposed a system that can be used for predicting what type of crop should be planted in a certain area within Sri Lanka with the help of artificial intelligence. The

system in fact is a recommendation system based on the collection of multiple environmental factors that directly impact the growth and yield of a certain crop. These factors are collected via sensors, namely temperature and humidity sensor, soil moisture sensor, pH sensor, and sunlight sensor, and are then communicated through Arduino microcontroller to the database for storage and analysis. The study relied on datasets collected by the authorities, as well agricultural books and websites. The collected data is preprocessed before being used by two algorithms which are the support vector machine (SVM) and naïve Bayes (NB) algorithms to perform a prediction based on the input data. The parameters are used to generate a PLU code related to each crop as an output. The added value of this study is that it offers an option for feedback by the users, from which the system will self-train accordingly as a response. Initially the model generated a 92% accuracy which can be enhanced to 95% upon constant use and feedback from farmers.

Priyadarshini et al. [10] purposed to help farmers finding the best crop to be grown in their lands by creating a recommendation system. The authors gathered data from government website and from Kaggle including the yield dataset which includes 16 major crops, the cost of cultivation dataset indicating the cost of each crop, the modal price of crops dataset which gives the market prices of crops among two months, the soil nutrient content dataset which includes five features which are order state, nitrogen content, phosphorous content, potassium content, and average pH. And the rainfall temperature dataset which includes crops, min and max rainfall, min and max temperature, and pH values. These data were cleaned and modified so that the null numbers are replaced by -1 in order not to affect the prediction process. Seven different methods were applied in order to go through the features in order to achieve accuracy and precision, these methods are the linear regression, the neural network, decision tree, K nearest neighbor, K nearest neighbor with cross validation, naïve Bayes and support vector machine. The best performance was for the neural network, with an accuracy of 89.88%.

Shilpa Mangesh Pande [11] and his colleagues in this paper proposed a prediction system in order to help farmers choose the most profitable crop with maximum yield and the best time for using fertilizers in order to reach the best results possible. The authors collected historical data for Maharashtra and Karnataka from different sources among which are indianwaterportal.com, data.gov.in and kaggle.com, while taking six features into consideration which are region, soil type, crop type, area, season, and year. These data were cleaned and modified so that the unavailable values are substituted with the mean values. Five different methods were applied in order to go through the features which are the support vector machine, artificial neural network, K nearest neighbor algorithms, random forest, and multivariate linear regression algorithms. The best performance was for the random forest algorithm with an accuracy of 95%.

Jadhav et al. [12] purposed to deal with the difficulties faced by farmers and find the best solutions and crops for farmers to grow in order to reach the best results possible. The authors gathered data from Kaggle while taking seven features

into consideration; the features selected are the ratio of nitrogen content in soil, the phosphorous ratio, the potassium ratio, the temperature, humidity, rainfall and pH value. While these data were cleaned and modified so that only important features are selected using bar charts, scatter plots, box plots etc., also a UI was built so that the farmer can enter his data in order to get immediate results and recommended crops to grow. Four different methods were applied in order to go through the features leading to the ones responsible of finding the most accurate and the best result of which crop to grow, these methods are random forest, decision tree, logistic regression and XGBoost. The best performance was for the random forest algorithm, reaching an accuracy of 98.9%.

In conclusion, the literature review has highlighted the importance of water quality assessment and crop prediction in agriculture. However, the studies have gaps such as the lack of a comprehensive model for both water quality assessment and crop recommendation, and underutilization of deep learning and machine learning algorithms. Our proposed model addresses these gaps by using these algorithms to predict water quality and recommend crop types. Additionally, our model provides a cost-effective solution to common farmer problems with reliable and accurate recommendations. The limitations across the four references on crop recommendation systems include the lack of information on datasets used, need for further validation in real-world scenarios, generalizability of the systems, and lack of comparison with existing systems. Our study contributes to the literature by providing an efficient model for water quality assessment and crop recommendation.

III. METHODOLOGY

A. Deep Learning Model

There are in the least millions of neurons in a human brain, that interaction among each other and communicate information. The neuronal interactions usually occur via electrochemical signals. The parts of the neurons that are responsible for connecting them to others are known as a Synapse, and they are the location of the passage of electrochemical signals. Neural networks resemble the functioning of the human central nervous system [13], since deep neural networks comprise a huge number of processing units that are connected to each other [14]. Deep neural networks are an important concept of Machine Learning as they can process large amounts of data, and then use them to come up with a pattern from which it can learn [15]. In the majority of cases, neural networks are used for classification tasks because they have a robust and efficient capability of processing the datasets [16]. There are three basic layers in the artificial neural network, which are the input layer, hidden layers, and output layer (Fig. 3).

1) *Input layer*: In the input layer, each input is assigned a vector where the attributes are represented. However, an output must also be given in order to be able to evaluate the model based on its accuracy.

2) *Hidden layer*: The hidden layers consist of weights and thresholds that improve the attributes. Two main processes take place in the hidden layers which are the multiplication of

weights and attributes, followed by sigmoid function that is used for output generation.

3) *Output layer*: The output layer is the level on which comparisons between the resultant output and the actual output take place. Depending on the similarities or differences, the feedback is given to the hidden layers. Additionally, permutation and combination of the weights and attributes occur in the output layer, to ensure better accuracy.

B. Forward propagation

In the case of forward propagation, the data flow is unidirectional in the sense of the output direction. In addition, no feedback is available which makes the accuracy measurements difficult as a way of evaluating the model [17].

C. Back propagation

Back propagation is different from forward propagation since it is constantly training itself by looping the comparisons between the actual and the desired outputs. The comparison is then propagated to the error function which alters the weights of the hidden layers in order to significantly reduce the differences between the actual and the desired outputs.

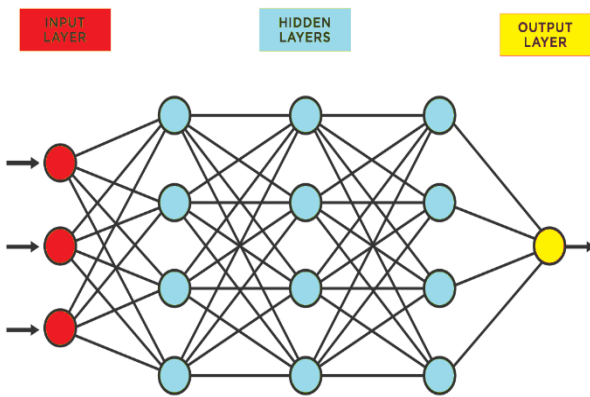


Fig. 3. Artificial neural network.

D. Activation Function

1) *Sigmoid function*: It can be used whenever the values are between 0 and 1. Where 0 represents the minimal probability of an event and 1 represents the maximum probability. Thus, for event predictions, the sigmoid function is very suitable. In addition, it is possible to find the derivative of the sigmoid function, being a curve between two points.

$$f(x) = \frac{1}{1+e^{-x}} \tag{1}$$

2) *SoftMax function*: It is similar to the sigmoid function, yet the output values in the SoftMax function are divided and can be summed up to a total of 1 [18]. Thus, it is like a probability distribution of the output values.

$$softmax(z_i) = \frac{exp(z_i)}{\sum_j exp(z_j)} \tag{2}$$

3) *Rectified linear unit (ReLU) function*: It is beginning to replace the sigmoid function. In this case, whenever the output

value is below zero, it will be rounded up so that the output is zero. The output and input values are considered equal when the input value is greater than zero.

$$RELU(x) = \begin{cases} 0 & \text{if } x < 0 \\ x & \text{if } x \geq 0 \end{cases} \tag{3}$$

E. Support Vector Machines

Support vector machines (SVM) is a supervised machine learning method that is commonly used for solving classification and regression problems. It was initially developed by Vapnik and Chervonenkis, and has its roots in Statistical Learning Theory. SVM is designed to learn structure from given data and can handle both continuous and categorical variables. The model represents different classes in a hyperplane within a multi-dimensional space, and its objective is to categorize a dataset into different classes by identifying the maximum marginal hyperplane (MMH) [19].

SVM employs kernel functions to transform input data into a desired form. For non-linear problems, the kernel trick technique is utilized in SVM with the aid of slack variables and additional dimensions, which transforms the data into a higher dimensional space. SVM utilizes several types of kernels, which are listed in Table I.

TABLE I. DESCRIPTION OF WATER QUALITY PARAMETERS

Kernel Type	Equation
Polynomial	$k(x, y) = (ax^T y + c)^2$
Linear	$k(x, y) = (ax^T y + c)$
Sigmoid	$k(x, y) = \tanh(ax^T y + c)$
Laplacian kernel	$k(x, y) = \exp\left(-\frac{\ x - y\ }{2\sigma}\right)$
Radial Basis Function (RBF)	$k(x, y) = \exp\left(-\frac{\ x - y\ ^2}{2\sigma^2}\right)$

F. Proposed System Workflow

The following figure describes the workflow of the overall system including the steps required in both of its subsystems.

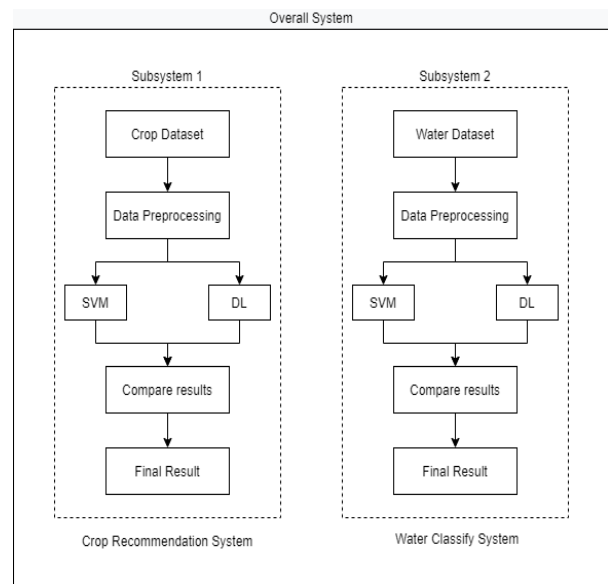


Fig. 4. The proposed workflow of our proposed prediction system.

Fig. 4 describes the general process performed in order to create our proposed prediction model. Our overall model is concerned with two separate tasks which are crop recommendation, and water quality classification. Thus, we decided to divide the system into two subsystems: subsystem one for crop recommendation, and subsystem two for water quality.

Mainly, similar processes take place in both of the subsystems, which slight differences. Initially, two different datasets are used for the different subsystems. In subsystem one, a crop dataset is acquired from Kaggle, where data selection and preprocessing take place. After that, a deep learning algorithm, namely neural network is implemented to see the results it can achieve, while SVM machine learning algorithm is also used for the same purpose. The objective is to determine for the case of crop recommendation, which perform better: neural network or SVM? The answer to this question can be obtained by comparing their results and thus determining whose result will be taken into consideration when recommending crops.

In subsystem two, a water quality assessment dataset is used and subjected to preprocessing, where it is later fed to a neural network and an SVM algorithm, in order to compare which one of them performs better in terms of classifying the quality of water which will be used for irrigating the crops.

G. Dataset Description and Preprocessing

1) Dataset Description

a) *Crop recommendation dataset:* For subsystem one, we use a dataset comprising the soil-specific attributes which are collected from online sources [20]. The crops considered in our model include 'rice', 'maize', 'chickpea', 'kidney beans', 'pigeon peas', 'moth beans', 'moonbeam', 'blackgram', 'lentil', 'pomegranate', 'banana', 'mango', 'grapes', 'watermelon', 'muskmelon', 'apple', 'orange', 'papaya', 'coconut', 'cotton', 'jute', 'coffee'.

Fig. 5 illustrates the number of instances or repetitions of each crop available in the training dataset. These crops differ in the climate that they need to grow and have good yields. For instance, the rice crop requires a lot of water, thus it is suitable for planting in areas with a lot of rainfall. The coffee crop is suitable to be planted in tropical regions. Jute for example is a crop that requires rainfall but also special soil conditions, thus it only grows in specific regions around India. Black gram crop conversely requires hot and humid climate to grow well and provide the best yield. Therefore, the different crop types require different climate conditions for their prosperity.

The five different attributes or parameters that were considered in the crop yield prediction dataset can be visualized in Fig. 6. For instance, the majority of data about the Phosphorous content fall between 45 and 60 units, whereas the Potassium values are always below 50 units. The pH value falls almost always between 6 and 7, humidity is above 80% most of the times, and the rainfall is between 40 to 120 units in most cases.

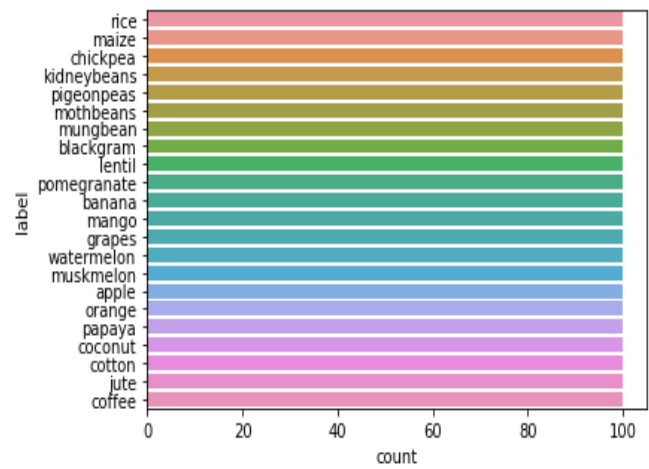


Fig. 5. Number of how many times each crop is present in the training dataset.

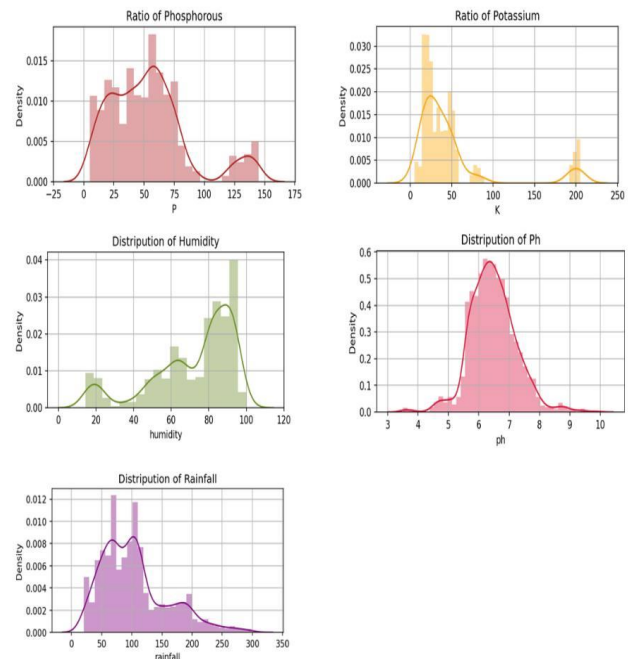


Fig. 6. Distribution of the crop prediction parameters.

b) *Water potability dataset:* The water quality dataset is acquired from Kaggle [21]. This dataset includes metrics for 3276 water bodies that differ in their water quality. This dataset is often used to determine the potability of the tested water. The parameters according to which the water is assessed are shown in Table II:

The nine different parameters that were considered in the water potability dataset are shown in Fig. 7. The ratio between the parameters being suitable or not is different and this difference in each of the features is used to predict the water quality. For instance, the difference between the values of turbidity being suitable for irrigation or unsuitable for irrigation determines if this parameter in particular classifies the tested water as suitable for irrigation or not.

TABLE II. DESCRIPTION OF WATER QUALITY PARAMETERS

Parameter	Description
<i>pH value</i>	representing the acid-base balance in the water, since this factor can be quite harmful to human or natural life if it were significantly off.
<i>Hardness</i>	representing the amount of calcium and magnesium salts present in water.
<i>Total dissolved solids</i>	representing the collection of organic and inorganic minerals that can be dissolved in water.
<i>Chloramines</i>	representing the measure of disinfectants that remain in the water after its treatment.
<i>Sulfate</i>	it is a mineral that can be found groundwater and its concentration varies depending on location.
<i>Conductivity</i>	representing the degree to which the water is capable of conducting electricity based on its minerals content.
<i>Organic carbon</i>	representing the organic matter that can be found in water as a result of decaying natural or synthetic matter.
<i>Trihalomethanes</i>	representing the chemicals that can be detected in water after chlorine treatment.
<i>Turbidity</i>	describing the light emitting properties of water depending on the amount of solid matter in its suspension.
<i>Potability</i>	describing the safety of water for human drinking.

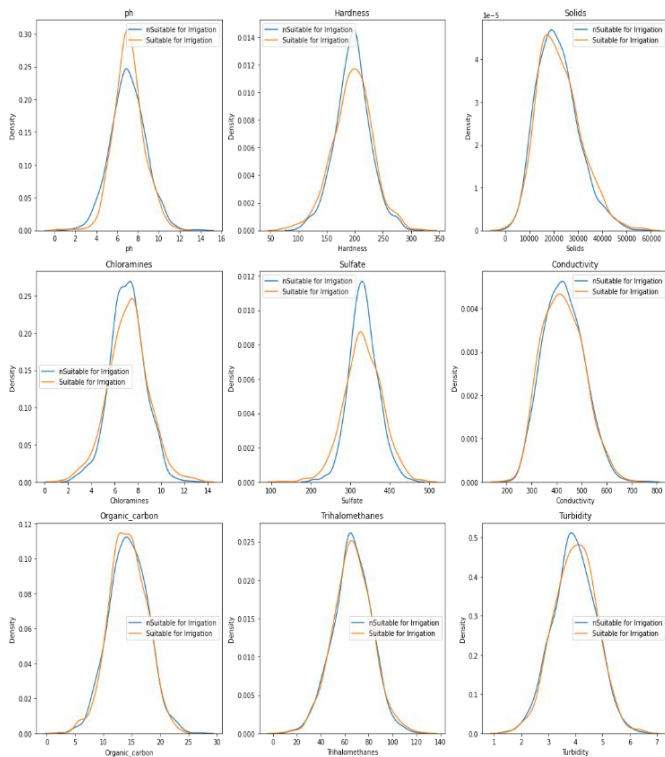


Fig. 7. Water quality parameter distribution.

2) *Data preprocessing*: The acquired data are not always clean, in fact they often include missing values, null values, and noise which make the dataset unfit to be used in ML or DL algorithms. Thus, data preprocessing is performed to clean the data and prepare it to be a suitable input for the algorithms. Preprocessing includes many tasks such as the removal of outliers or flawed data, as well as replacing the missing values if present. There are two techniques that can be performed to resolve the missing data, among which is deletion. Deletion

means the removal of the entire row where the missing data reside. In the case of deletion, this might lead to significant reduction in the size of the dataset if it was already small. The other technique includes filling the missing spaces with the average or mean values of the attributes.

a) *Crop recommendation system data preprocessing*: The crop recommendation dataset did not contain any empty values. The data was entered to Standard Scaler in order to standardize the features since the input from the dataset have very different characteristics range. Feature standardization takes place through subtracting the mean and then scaling to unit variance, where unit variance is achieved by dividing all the values by the standard deviation.

b) *Water data preprocessing*: The water quality classification dataset contained some null values, thus these null value rows were dropped from the data so that the performance of the algorithms is not negatively affected by them. Just like the crop recommendation dataset, the data was entered to StandardScaler.

H. Experimental Set Up

In order to achieve the best results, the ML and DL algorithms must be trained and tested under a variety of scenarios. In this study, we trained models on the data so that it will predict the crop that can be grown based on various given parameters such as the soil nutrients and environmental factors. We give different set of input parameters and based on them we train the data to predict the exact crop to be grown. We fit the data to the X, Y training values and make predictions on the X test data. We trained the model for 100 epochs. The model with the lowest loss is considered as the best model and that model is used for evaluation and testing.

The DL model's performance is evaluated in the python environment. TensorFlow is a free deep-learning framework tool it offered by google. It provides a library of various models for data preprocessing, classification, clustering, forecasting, visualization, etc. The collab in which our experiments were conducted contains many powerful features that help the developer and researchers in the development and research process.

IV. RESULTS

A. Correlation for Water

Seaborn heat map function (fig. 8) was used to determine the correlations between the different factors. Each two factors affect each other to a certain degree, which is referred to as correlation.

The correlation matrix illustrates that each feature is strongly correlated to itself only (+1 score), and not to other features. In fact, the water potability/quality factors don't show any string correlations between each other, except for the two parameters pH and hardness, where a weak correlation exists (0.08).

These correlation results imply that dimension reduction is not possible in the water quality data due to the lack of correlation between its parameters.

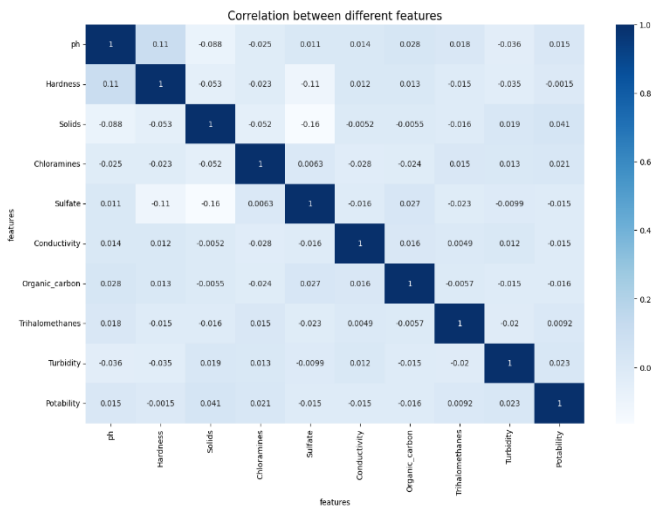


Fig. 8. Correlation between different features.

In the water quality classification dataset, all of the parameters are regarded as independent features, whereas the water potability parameter is the only dependent factor.

B. Evaluation Metrics

There are several metrics that can be used to evaluate the performance of the models such as precision, recall, f1 score, and accuracy [see (4), (5), (6) and (7)]. Recall is another term used for sensitivity, which resembles the true positive value, which is also the portion of the correctly classified inputs as positive among the entire inputs that should have been classified as positive. Precision is the portion of the true positive classifications over the entirety of the positive results. F-measure is the harmonic mean of the precision and recall and sums up the predictive performance of a model.

$$Recall = TP / (TP + FN) \tag{4}$$

$$Precision = TP / (TP + FP) \tag{5}$$

$$F\text{-Measure} = 2 \cdot Precision \cdot Recall / (Precision + Recall) \tag{6}$$

$$Accuracy = (TN + TP) / (TN + TP + FN + FP) \tag{7}$$

Where, True positive is designated by TP. True negative is designated by TN. False positive is designated by FP. False negative is designated by FN. Area under curve AUC is also a beneficial metric, where the values must be between 0 and 1, such that the higher the AUC value, the better the performance. If the model can discriminate between the instances of two classes perfectly, then AUC would be 1. Conversely, if the model fails to distinguish between any instances, the AUC would be 0.

C. System Results Evaluation

The evaluation metrics were obtained for the proposed water quality and crop recommendation for irrigation system during the training phase. The subsystem was evaluated based on accuracy, precision, f1 score, recall, and loss.

For the crop recommendation system, both the DL and SVM models achieved high levels of accuracy, precision, recall, and F1 score, with only slight variations between the

two models. As shown in Fig. 9, the DL model achieved a slightly higher accuracy of 0.975, compared to the SVM model's accuracy of 0.968. Both models achieved high precision scores of 0.97 and high recall scores of 0.97 and 0.98, respectively. The F1 score was also high for both models, at 0.97. These results suggest that both the DL and SVM models were effective in predicting crop recommendations, with the DL model performing slightly better than the SVM model in terms of accuracy.

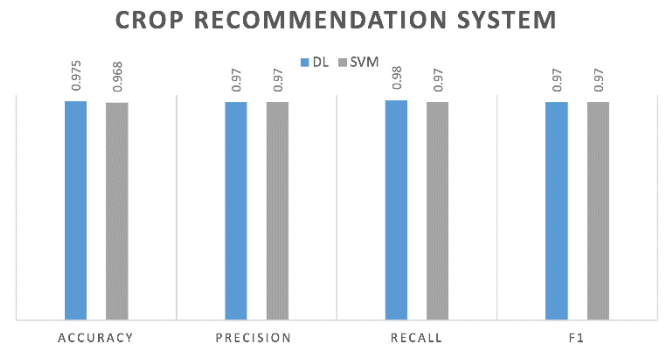


Fig. 9. Crop recommendation system.

As shown in Fig. 10, for the water classify system, the DL model achieved a higher level of accuracy (0.964) compared to the SVM model (0.927), as well as higher precision and recall scores (0.94 for both). The SVM model achieved a lower precision score of 0.914 and a lower recall score of 0.9. The F1 score for both models were relatively similar, with the DL model achieving 0.94 and the SVM model achieving 0.9. These results suggest that the DL model outperformed the SVM model in predicting water potability, with a higher accuracy, precision, and recall score. Overall, the DL model was more effective in classifying water samples as potable or non-potable compared to the SVM model.

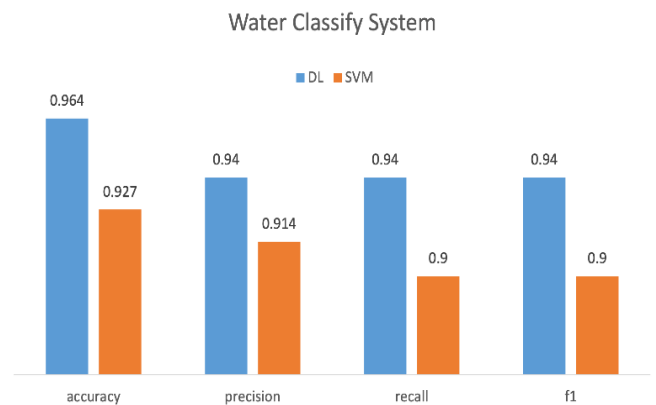


Fig. 10. Water classify system.

As a result, the SVM algorithm was able to achieve the accuracy less than neural network model by scoring 97% for crop recommendation for DL as shown in Fig. 11, and 96% for SVM. On the other hand, the water classify subsystem achieve 96% accuracy for DL as shown in Fig. 12, and 92% for SVM.

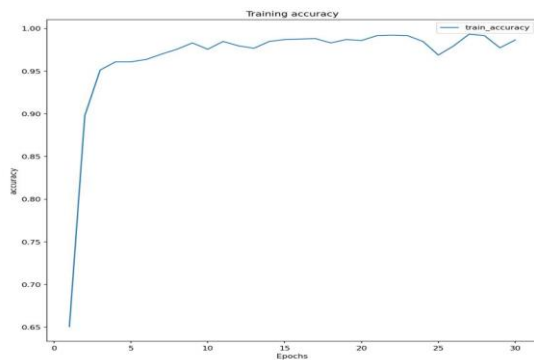


Fig. 11. Training accuracy for crop recommendation using DL.

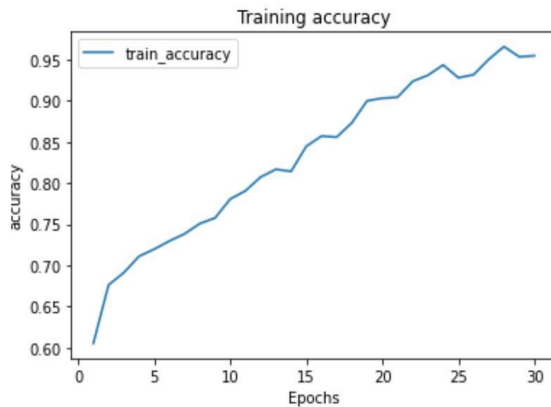


Fig. 12. Training accuracy for water classify using DL.

Compared to the related works mentioned in the literature review, our work stands out for providing a solution for both crop recommendation and water potability prediction, while some of the studies focused on only one of these tasks. While our work and the related studies all utilize datasets for crop recommendation systems, our approach stands out due to the unique nature of our dataset. Our dataset consists of soil-specific attributes that were collected from online sources, providing a comprehensive and informative dataset for crop recommendations. In contrast, the datasets used in the related studies are either unspecified or contain a limited number of crops. For example, [10] only contains 16 crops. Furthermore, our dataset includes 23 crops, which is a more extensive and diverse selection compared to some of the other studies that do not mention the number or types of crops considered. Therefore, our dataset is more comprehensive and suitable for accurate crop recommendations.

Additionally, our work used a variety of parameters for prediction, such as sun exposure, humidity, soil pH, and soil moisture, as well as water pH, electric conductivity, and content of minerals such as chloride, calcium, and magnesium. Some of the references also used similar parameters, but the methods varied, including decision trees, random forests, and Naive Bayes classifiers. Results from the four references show that accuracy ranged from 90% to 96.7% using different machine learning algorithms. However, the proposed system in our work achieved a higher accuracy for both crop recommendation and water quality prediction, demonstrating the effectiveness of the proposed model.

V. CONCLUSION

Two of the major issues or challenges in the proper maintenance of an agricultural field are the quality of irrigation water that is reaching the plants and that the plants need for proper growth, as well as the appropriate crop type since crops don't react and grow the same as each other in different climates and conditions.

In conclusion, the study proposed a binary model based on deep learning to address the challenges of water quality analysis and crop recommendation. The model was divided into two subsystems that relied on data collected from separate sources for training and testing. The performance of the model was evaluated using various metrics, including a confusion matrix, accuracy, recall, and precision. The neural network achieved high accuracy rates of 97% and 96% for crop recommendation and water quality prediction, respectively, while SVM achieved 96% and 92% accuracy. The results suggested that the binary model had the potential to serve as an effective tool for addressing the complex issues of water quality analysis and crop recommendation simultaneously.

The crop recommendation dataset used in the study has some limitations, including missing or incomplete soil-specific attributes, reliance on online sources for data collection, and a limited number of crops. Meanwhile, the water potability dataset may be limited by a lack of representativeness in sampled water bodies and incomplete or missing data, which could affect the accuracy of water potability predictions.

In the future, we might add to our dataset a wider collection of crops that the system can choose from based on the selected parameters. Additionally, we can implement other algorithms in the future to check if the accuracy can be improved, or if tweaking the algorithms, a bit can add more efficiency. We can also integrate IoT systems in order to be able to collect more data from the field for both the water quality prediction and the crop recommendation features.

REFERENCES

- [1] R. Lal, "Soil science and the carbon civilization," *Soil Science Society of America Journal*, vol. 71, pp. 1425-1437, 2007.
- [2] H. D. Nguyen, T. Q. D. Nguyen, H. N. Thi, B. Q. Lap, and T. T. H. Phan, "The use of machine learning algorithms for evaluating water quality index: A survey and perspective," in *2022 International Conference on Multimedia Analysis and Pattern Recognition (MAPR)*, Phu Quoc, Vietnam, 2022, pp. 1-6.
- [3] N. Koralay and Ö. Kara, "Forestry activities and surface water quality in a watershed," *European Journal of Forest Engineering*, vol. 4, pp. 70-82, 2018.
- [4] B. Phalan, R. Green, and A. Balmford, "Closing yield gaps: Perils and possibilities for biodiversity conservation," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 369, p. 20120285, 2014.
- [5] D. Paudel, H. Boogaard, A. de Wit, S. Janssen, S. Osinga, C. Pylaniadis, *et al.*, "Machine learning for large-scale crop forecasting," *Agricultural Systems*, vol. 187, p. 103016, 2021.
- [6] G. Pratyusha and A. Sinha, "Prediction of crop using deep learning techniques: A concise review," in *Recent advances in computer based systems, processes and applications*, A. Namburu and S. S. Barpanda, Eds. Boca Raton: CRC Press, 2020, pp. 145-159.
- [7] A. T. Ali El Bilali. (2020). Prediction of irrigation water quality parameters using machine learning. University of Hassan II, Casablanca, Faculty of Sciences and Techniques of Mohammedia, Morocco.

- [8] Mokhtar Ali, El-Ssawy Wessam, He Hongming, Al-Anasari Nadhir (2022). Using Machine Learning Models to Predict Hydroponically Grown Lettuce Yield. *Frontiers in Plant Science*.
- [9] P. Bandara, T. Weerasooriya, R. T.H, W. J. M. Nanayakkara, D. M.A.C, and P. M.G.P, "Crop recommendation system," *International Journal of Computer Applications*, vol. 175, pp. 22-25, 2020.
- [10] A. Priyadharshini & Chakraborty, Swapneel & Kumar, Aayush & Pooniwala, Omen. (2021). Intelligent Crop Recommendation System using Machine Learning. 843-848. 10.1109/ICCMC51019.2021.9418375.
- [11] Shafiulla Shariff, Shwetha R B, Ramya O G, Pushpa H, Pooja K R, 2022, Crop Recommendation using Machine Learning Techniques, *INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) ICEI – 2022 (Volume 10 – Issue 11)*,
- [12] A. Jadhav, N. Riswadkar, P. Jadhav, and Y. Gogawale, "Crop recommendation system using machine learning algorithms," *International Journal of Engineering Research & Technology* vol. 9, pp. 1227-1230, 2022.
- [13] S. Mishra, D. Mishra, and G. H. Santra, "Applications of machine learning techniques in agricultural crop production: A review paper," *Indian Journal of Science and Technology*, vol. 9, pp. 1-14, 2016.
- [14] V. V. Erokhin and V. E. Eliseeva, "Influence of neural network architecture on its performance," *Soft Measurements and Computing*, vol. 8, pp. 5-17, 2022.
- [15] M. I. C. Rachmatullah, J. Santoso, and K. Surendro, "A novel approach in determining neural networks architecture to classify data with large number of attributes," *IEEE Access*, vol. 8, pp. 204728-204743, 2020.
- [16] K. Sulaiman, L. Hakim Ismail, M. Adib Mohammad Razi, M. Shalahuddin Adnan, and R. Ghazali, "Water quality classification using an artificial neural network (ANN)," *IOP Conference Series: Materials Science and Engineering*, vol. 601, p. 012005, 2019.
- [17] T. R. Zhangirov, A. B. Krevchik, D. V. Zuev, A. A. Liss, A. V. Ekalo, and N. Y. Grigoryeva, "Forward propagation neural network weighting analysis as a model estimation method," in *2021 II International Conference on Neural Networks and Neurotechnologies (NeuroNT)*, Saint Petersburg, Russia, 2021, pp. 10-12.
- [18] P. Ramachandran, B. Zoph, and Q. V. Le, "Searching for activation functions," *arXiv preprint arXiv:1710.05941*, 2017.
- [19] K. Lata, "Analysis of SVM and RNN-LSTM on Crop Datasets," in *2020 5th International Conference on Computing, Communication and Security (ICCCS)*, 2020, pp. 1-5.
- [20] A. Ingle, "Crop Recommendation Dataset," Kaggle, 2021. [Online]. Available: <https://www.kaggle.com/datasets/atharvaingle/crop-recommendation-dataset>. [Accessed: Mar. 31, 2023].
- [21] A. Kadiwal, "Water Potability," Kaggle, 2021. [Online]. Available: <https://www.kaggle.com/datasets/adityakadiwal/water-potability>. [Accessed: Mar. 31, 2023].

Context Aware Automatic Subjective and Objective Question Generation using Fast Text to Text Transfer Learning

Arpit Agrawal¹, Pragya Shukla²

Institute of Engineering and Technology, DAVV Indore, India^{1,2}

Abstract—Online learning has gained a tremendous popularity in the last decade due to the facility to learn anytime, anything, anywhere from the ocean of web resources available. Especially the lockdown all over the world due to the Covid-19 pandemic has brought an enormous attention towards the online learning for value addition and skills development not only for the school/college students, but also to the working professionals. This massive growth in online learning has made the task of assessment very tedious and demands training, experience and resources. Automatic Question generation (AQG) techniques have been introduced to resolve this problem by deriving a question bank from the text documents. However, the performance of conventional AQG techniques is subject to the availability of large labelled training dataset. The requirement of deep linguistic knowledge for the generation of heuristic and hand-crafted rules to transform declarative sentence into interrogative sentence makes the problem further complicated. This paper presents a transfer learning-based text to text transformation model to generate the subjective and objective questions automatically from the text document. The proposed AQG model utilizes the Text-to-Text-Transfer-Transformer (T5) which reframes natural language processing tasks into a unified text-to-text-format and augments it with word sense disambiguation (WSD), ConceptNet and domain adaptation framework to improve the meaningfulness of the questions. Fast T5 library with beam-search decoding algorithm has been used here to reduce the model size and increase the speed of the model through quantization of the whole model by Open Neural Network Exchange (ONNX) framework. The keywords extraction in the proposed framework is performed using the Multipartite graphs to enhance the context awareness. The qualitative and quantitative performance of the proposed AQG model is evaluated through a comprehensive experimental analysis over the publicly available Squad dataset.

Keywords—Automatic question generation; Text-to-Text-transfer-transformer (T5); natural language processing; word sense disambiguation (WSD); domain adaptation; multipartite graphs; beam-search decoding

I. INTRODUCTION

Assessment has always been a very crucial tool in the educational ecosystem to identify the attainment of the learning outcomes and the curriculum gaps. Various evaluation techniques and assessment methods have been proposed by the researchers and academicians to exploit various aspects of students' learning. Question-answer technique has found to be the most effective way of evaluating the students' knowledge as it serves various purposes simultaneously like offering the

opportunity to practice the retrieval of information from memory, identifying the misconceptions, reinforced learning through iterative core concepts, engagement in learning activities, etc. However, deriving these questions manually is a huge task as it requires the thorough knowledge, training and depth of understanding. The challenge increases many folds when the questions need to be replaced periodically to ensure the validity and value which reduces after few rounds of usage. The emergence of e-learning has added a new dimension over the last decade through the available massive open online courses (MOOCs) and adaptive learning. These tools require a large pool of questions to ensure their effectiveness [1-3].

Considering the amount of labor, time and cost associated with the process of generating questions manually, automatic question generation (AQG) has been introduced recently to automate the process. It produces the questions automatically through the structured or unstructured knowledge source and hence enables the educators to invest the time in some other instructional activities. A well-structured and customized question bank can also be created by controlling the difficulty and cognitive level to make the testing adaptive as per the students' requirements. The type of the questions may be subjective or objective. Subjective questions can be used to assess the depth and diversity of understanding of the students, while the binary mode of assessment in objective questions may be used to evaluate the logical understanding in a faster and reliable manner. The most popular varieties of objective questions include the multiple-choice questions (MCQs), fill in the blank questions, true or false questions or matching questions. Although the subjective question-based assessment has remained the most trusted tool in the traditional education systems so as to assess the writing skills and critical reasoning of the students, the recent growth in the e-learning paradigm has attracted many universities towards the objective questions. This is due to the reason that in computer-assisted examination, accurate assessment of numerous students through subjective questions will be a very tedious task. Automatic generation of semantically meaningful and well-formed questions (subjective and objective) has the potential to greatly enhance the learning and assessment experience [4-6].

Traditionally, the process of AQG is classified in two categories, rule-based approach and deep learning approach. Rule-based approach derives the questions on the basis of the hand-crafted rules which are derived through a high level of manual interference. However, the limitation of this approach is that the rules derived for one domain may not be suitable for

other domains. On the other hand, deep learning approach utilizes the natural language processing (NLP) techniques like text abstraction, text summarization, machine translation, etc to generate the questions automatically. Many question answer scenarios and various datasets have been presented over the last decade with different size of context, different formats of answers and sizes of training datasets [7,8]. However, the performance of the AQG model is greatly subject to the understanding and reasoning about the relationship between the sentences in the contextual data. The most important requirement from a machine learning model applied for NLP task is to process the text in a way which is acquiescent to downstream the learning through a general-purpose knowledge framework which understands the text.

Recently, transfer learning has emerged as a very powerful technique in NLP which utilizes the knowledge retrieved from one task to the other related task and therefore reduces the necessity of fine-tuning dataset and improves the performance. This feature of transfer learning technique is also known as domain adaptation which is achieved through the word vector mapping between the similar words and similar vectors. The general-purpose knowledge and abilities in transfer learning is achieved through the pre-training of the entire model over a data rich task. This knowledge is further been transferred to the downstream tasks afterwards. Generally, supervised learning on a large labeled dataset is used to pre-train the transfer learning model for computer vision applications and unsupervised learning on unlabeled data is utilized for the same in NLP applications. The enormous amount of text data available on internet may be used in transfer learning to train the network. Various transfer learning models have been proposed for the NLP applications like GPT, ELMo, BERT, XLNET, ALBERT, RoBERTa, etc. [9-12]. All of these models have shown promise for doing particular tasks, but when asked to fulfill a more comprehensive set of requirements, they fell short. The performance of these models differs from one task to the next due to the fact that even the procedures, practices, and processes that they use are not consistent. Therefore, in order to comprehend transfer learning more thoroughly, a methodology that is both consistent and methodical is essential.

Various unifying approaches like language modeling, span extraction, casting of all text as question answering, etc have been proposed by the researchers in last few years for NLP tasks. But, Text-to-Text-Transfer-Transformer (T5) has evolved as the most powerful unified framework which treats every text processing problem as a “text-to-text” problem [13]. It takes text as input, process or transforms it and generates text as output. The unification has made it possible to apply the same model, objectives, training and decoding procedure to any NLP task at hand like text abstraction, document summarization, question-answer generation, text classification, sentiment analysis from text, etc. The encoder-decoder model in T5 is pretrained on a multi-purpose blend of supervised and unsupervised tasks. The encoded input is fed through a cross-attention layer to generate the autoregressive decoded output using the scalar embedding. The training algorithm in T5 is teacher forcing and hence requires an input sequence and target sequence compulsorily.

However, due to the sequential nature of T5 model, the text-to-text transformation is naturally slow. The model speed even reduces for the larger T5 models and makes the implementation really difficult with limited resources. Augmenting the transformer model with WSD to generate the objective questions in terms of multiple choice, fill-in-the-blank, true/false or pair matching questions enhances the complexity further. This paper presents a fast automatic question generation model for objective and subjective questions by using the fastT5 model which is capable of inferencing faster than the conventional transformer for a reduced model size. The proposed model is run on the onnx runtime which quantizes the whole process and gives the model as output in a single line of code. The flexibility to customize the whole model as per the application has been achieved in this work through the PyTorch Lightning library. It is a lightweight, scalable and high-performance deep learning framework which provides a flexible interface for PyTorch which can easily work on distributed hardware while keeping the models hardware skeptical. The keywords extraction in the proposed framework is performed using the Multipartite graphs to enhance the context awareness. The speed of finding and replacing the keywords in objective questions generation is further improved by using the FlashText library.

The major contribution of the proposed work lies within the application of FastT5 transfer learning model for the subjective and objective question generation automatically. The augmentation of word sense disambiguation (WSD) and domain adaptation framework with the proposed model has improved the meaningfulness of the questions. As the model is required to understand and reason about the relationship between the sentences in the story, the context awareness in the proposed work is enhanced using the multipartite graph-based keyword extraction and FlashText library and the flexibility is achieved through the PyTorch Lightning library. The performance of the proposed technique is evaluated thoroughly through an experimental analysis over SQuAD dataset. Specifically, “teacher forcing” methodology has been used here to train the model with a maximum likelihood objective regardless of the task. Due to the quantized model over onnxruntime, the proposed system is lightweight and faster with good BLEU score.

Rest of the paper is organized as follows: Section II deals with the literature survey through the analysis of related work. The mathematical framework of the text-to-text transformer is given in Section III. The proposed automatic subjective and objective question generation using the proposed fastT5 model is discussed in Section IV. Effectiveness of the proposed strategy is illustrated through the experimental analysis in Section V while Section VI concludes the paper.

II. RELATED WORK

The potential of AQG to change the complete paradigm of online education system, information retrieval systems and interactive support systems has attracted a lot of researchers towards it. Rus et al. [14] has defined the problem of AQG as “the task of automatically generating questions from various inputs such as raw text, database, or semantic representation”. The increasing availability of the digital information on

internet and the amazing advancement in the field of NLP has driven the pace of research in the area of AQG. Conventionally, the problem of AQG is classified in two categories rule-based approaches and neural network approaches (also called neural approaches). In rule-based approaches, human designed syntactic rules are used to transform the declarative sentences in text to the interrogative sentences. It typically applies some lexical transformation around the main verb in the sentence. Heilman and Smith [15] presented a trained logistic model to generate the questions from a paragraph by deriving the transformation rules to produce multiple declarative sentences. These sentences are further converted into questions by syntactic and lexical transformations. Dhole and Manning [16] proposed a collection of semantic and syntactic rules-based heuristics model and named it Syn-QG. They addressed the problem of rule-based approach of generating the questions on Blooms Taxonomy level 1 like “what” and “yes or no”. The proposed Syn-QG model utilizes the VerbNet to generate a set of semantically richer questions. However, this approach requires extensive knowledge in linguistics and well-designed transformation rules to convert the declarative sentences into questions.

Wang et al. [17] proposed a parser based AQG model for medical knowledge evaluation. The medical terms in the articles are extracted and the unstructured entries are classified into different fields using the parsers. They formed more than 100 templates to match the parsed data entries to the question template. However, this method required an extensive labor and the respective templates are domain specific and cannot be applied to other applications. Fabbri et al. [18] developed a context-answer pair based syntactic dataset and used it for the training of deep learning model. The accuracy and effectiveness of this model was dependent on the quality of the dataset and was not generalized for any application. The generated questions were also monotonous.

RNN Encoder-decoder based framework was proposed by Zhou et al. [19] for AQG using attention mechanism for generating natural language questions. The network in the encoder is bidirectional Gated Recurrent Unit (GRU) and left-to-right GRU in the decoder. A multi-perspective context matching algorithm has been proposed by Song et.al. [20] in the sequence-to-sequence LSTM encoder-decoder framework with the copy mechanism for the AQG. Two sets of hidden vectors were generated by encoding the context passage and answer pair through two bi-directional LSTMs. Yuan et al. [21] proposed a combination of supervised learning and reinforcement learning to train a model for AQG. The proposed framework has trained the model first with an objective function of minimizing the cross-entropy loss and then maximized the reward function using the policy gradient method.

Kim et al. [22] derived an answer masking based AQG framework to encode the answers and the sentences separately where the answer text is replaced special tokens. The stack of attention layers with a dot-product based alignment score is used as attention module which cooperates with the answer encoder. The output gate of the LSTM network utilized retrieval style word generator to predict the token. Recently the

potential of transfer learning has been exploited by Mitkov et. al. [23] to generate the multi-choice questions answering. They have showed that the unsupervised transfer learning can be helpful in NLP based applications through iterative self-labeling technique. The transferability of knowledge in encoder and decoder has been explored by See et.al. [24] through a thorough experimental analysis with source question dataset and target dataset. They have used a sequence-to-sequence pointer-generator network over a smaller sized dataset and evaluated the performance over semi-supervised learning.

Most of the prior works in the field of AQG has assumed the access to the sufficiently large dataset for the training and validation. The time taken in the training of the existing models and transformers is also very large. The techniques proposed in the traditional techniques were mostly domain specific and were either suitable for subjective question generation (long answers and short answers) or for objective question generation. Motivated by these facts, our work aims to provide a general purpose, fast and more meaningful AQG model for subjective and objective questions generation using transfer learning.

III. T5 TRANSFORMER MODEL DESCRIPTION

Transfer learning has gained a lot of attention over the last years due to its potential to utilize the knowledge learned from one task for the inferential study of another task and thereby reducing the necessity of fine tuning the dataset. It typically reuses the learned weights of a base network which is trained with a large dataset to the target network by changing the training objectives over a smaller dataset. Typically, the objective of transfer learning is to enhance the learning characteristics through the target conditional probability distribution $P(Y_T|X_T)$ in a target domain D_T utilizing the knowledge attained from the source domain D_S and Source task T_S . **Text-to-Text Transfer Transformer (T5)** is an advance model based on transfer learning converts every problem of NLP like question answering, text classification, translation, question generation, etc. to a text-to-text problem [25]. The model takes text as input and gets trained with the input so as to generate some target text. It enables the model to use the same model, hyperparameters, loss function, activation function, etc. for any diverse set of NLP application. The typical structure of the transformer model is shown in Fig. 1 which is based on an encoder-decoder model. The input sequence of symbols represented by $X=(x_1, x_2, \dots, x_n)$ is mapped to a sequence of continuous representation $Z=(z_1, z_2, \dots, z_n)$ at the encoder. Taking the sequence z , decoder generates an output sequence of symbols $Y=(y_1, y_2, \dots, y_n)$ with one element at a time. The model works in autoregressive style to generate the next symbol by utilizing the previously generated symbols.

The encoder comprises of a sequentially connected stack of N identical encoder layers where the output of one layer is the input to the next layer. The input token sequences fed to the

encoder first concerted into vectors of size d_{model} by the token embedding layer and the position embedding layer. Each block of encoder layer is further comprising of two components; a multi-head self-attention layer and a fully connected feed-forward layer. A simplified layer normalization is applied to

the input of each sublayer where the activation function is rescaled without adding bias. Residual skip connection method is applied afterwards to map each sub-component's input to the output. Dropout is applied within the feed-forward network, on the skip connection, on the attention weights, and at the input and output of the entire stack. Self-attention mechanism plays a vital role in transformer model which relates different positions of a single sequence to compute the representation of the sequence. It typically utilizes the weighted average of the rest of the sequences to replace each element of a sequence. The process of multi-head self-attention mechanism can be represented as

$$MultiHeadAttn(\tilde{Q}, \tilde{K}, \tilde{V}) = [\phi_1; \phi_2; \phi_3; \dots; \phi_h] W^O \quad (1)$$

and

$$\phi_i = Attention(\tilde{Q}W_i^Q, \tilde{K}W_i^K, \tilde{V}W_i^V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d}}\right)V \quad (2)$$

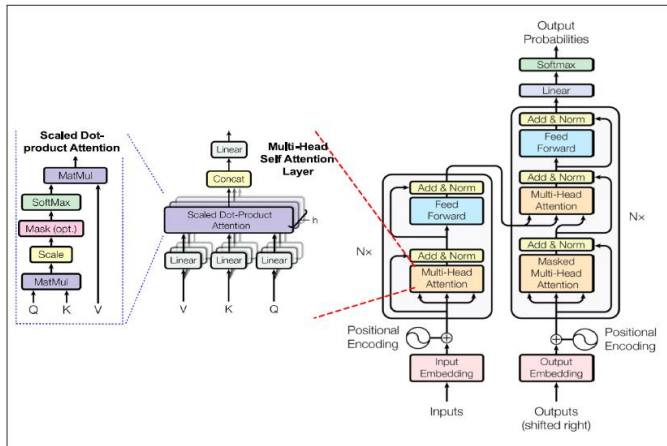


Fig. 1. Text-to-text-transformer model [25].

Similar to the encoder, the decoder structure also comprises of a stack of identical layers. However, each layer in decoder comprises of three sublayers. The third extra sub-layer in decoder is of the form of autoregressive or casual self-attention which performs the multi-head attention over the output of encoder stack and allows the model to attend to past outputs. The residual connections are employed around each sub-layer followed by layer normalization. The self-attention sub-layer is modified here to prevent the positions from attending to subsequent positions. The output of the final decoder block is given to the dense layer with 'softmax' output. The weights of this dense layer are further shared with the input embedding matrix. All attention mechanisms here are fragmented into independent "heads" whose outputs are concatenated before being further processed.

The computation complexity in case of convolutional neural network based like Extended Neural GPU, ByteNet and ConvS2S increases with the increase in the distance between the two arbitrary inputs or output positions [26-28]. It varies linearly for ConvS2S and logarithmically for ByteNet which

makes it extremely difficult to establish the learning among the dependencies between distant positions. This issue has been resolved in T5 transformer architecture by limiting the computation to a constant number by establishing a tradeoff with reduced resolution. Considering the fact that self-attention is an order independent operation, an explicit position is provided in the transformer architecture. Relative position embedding is used here instead of fixed embedding to provide the learned embedding with respect to the offset between the "key" and "query". A simplified form of position embedding is used in T5 model where the attention weights is computed by adding a scaler "embedding" to the corresponding logit. The model efficiency is further enhanced by sharing the position embedding parameters across all the layers. The major difference in the structure of T5 model with the conventional transformer model is that the layer normalization in T5 is placed outside the residual path. It also utilizes different position embedding scheme and removes the Layer norm bias as compared to the traditional model without affecting the performance considerably due to the orthogonal changes in the structure.

IV. PROPOSED METHODOLOGY

The proposed framework is designed for the automatic subjective and objective questions using the fast T5 Model as shown in Fig. 2.

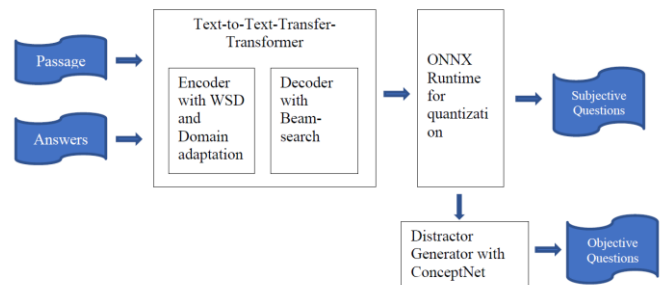


Fig. 2. Proposed model for subjective and objective question generation.

The methodology used for both the cases is discussed here in two parts. Part 1 will discuss the AQG for Subjective question generation and part 2 will discuss the objective questions generation.

Part 1. AQG for Subjective questions: It comprises of four phases of processing namely problem formulation, data preprocessing, baseline model training and domain adaptation

A. Problem Formulation

Considering w_k as the word in the input sentence $S = \{w_k\}_{k=1}^L$ with length L^S , the problem is formulated as the generation of natural question Q with an objective of maximization of the conditional probability of the predicted question sequence Q given the input sentence S as

$$P(Q|S, \theta) = \prod_{t=1}^{L^Q} P(q_t | S, \{q_\tau\}_{\tau=1}^{t-1}, \theta) \quad (3)$$

where L^Q and q_t denote the length of the output question Q and words within Q respectively, θ represents the set of parameters of the prediction model. It is assumed here that the answer of the generated question is a consecutive segment in S so as to generate factual questions.

B. Data Preprocessing

The dataset comprises of three fields (C, Q, A) where C is the prefix which is a string indicating the task to perform, Q and A are the question (target text: The target sequence) and answer (input text: The input text sequence) from the context respectively. As mentioned earlier, our source data set, Squad [29], is a general-purpose QA dataset containing questions generated from Wikipedia pages that cover various topics. Preprocessing is performed to derive the pair of input sequence S and the corresponding question from the context. The larger paragraph is first of all transformed to the abstractive form through text abstraction process. The sentence which contains the answer A is extracted from the short paragraph C. Sentence-question (S,Q) pair is generated through this process which can be utilized for the training of the transformer model.

C. Baseline Model

The baseline model used in the proposed work is T5 model which converts every NLP problem into a text-to-text problem. The architecture of the model is discussed in earlier section. However, the sequential manner of text-to-text transformation is naturally slow and makes the implementation in real time challenging. The performance deteriorates further for larger models as the decoder in T5 model is utilized repeatedly for inference. To overcome this challenge, the conventional T5 Transformer model in this paper is augmented with fast T5 library which makes the inference of the conventional transformer faster. The high inference speed is achieved by running it on onnx runtime which quantize the whole model to decrease its size. FastT5 library converts the pretrained model to Open Neural Network Exchange (onnx) for quantization and generates a model as output which can run in a single line of code through a cross-platform inference framework, onnx runtime. It is a training machine-learning accelerator which offers compatibility among different hardware, drivers and operating systems to optimize the overall performance. The resultant quantized model is lightweight models which offer almost the same accuracy with low latency (due to graph optimization) as compared to the conventional transformer models.

The encoder and decoder are exported to the onnx model separately to reduce the computation complexity and the memory requirement considerably. It is also observed in the conventional transformer model that a constant number of inputs are given to the models, but first level of decoder does not take the previous key values. In contrast, the past key values are provided to the other layers of decoder. This uneven computation issue is addressed in fastT5 model by creating two different decoders; one for the first step without past key values and other for the other steps with past key values. The augmentation of word sense disambiguation (WSD) [30] and domain adaptation framework with the proposed model has improved the meaningfulness of the questions. As the model is required to understand and reason about the relationship

between the sentences in the story, the context awareness in the proposed work is enhanced using the multipartite graph-based keyword extraction and FlashText library and the flexibility is achieved through the PyTorch Lightning library. Beam search algorithm has been used in this model in the decoder which tracks the n (number of beams) most likely hypotheses (based on word probability) at each timestep and finally chooses the hypothesis with the highest overall probability. Teacher forcing has been used in the model as the error propagation-based learning strategy during the training phase of the T5 model where the inference of a new token is depending upon the previous predicted tokens and current hidden state. This strategy modifies the training process by using the true tokens partially instead of always using the generated tokens.

D. Domain Adaptation

Domain adaptation is the process to learn a predictive function F_t which can be used to map the knowledge attained from the source domain D_S for task T_S to the target domain D_T for task T_T . The transfer of knowledge is done in such a way that domain distribution discrepancy between D_S and D_T is reduced to a minimum possible value. Supervised domain adaptation has been used in this work where the best model M_b is selected on the basis of best model evaluation parameters over the validation dataset in target domain for a given number of epochs. The best model is further fine-tuned for the target domain.

Part 2. AQG for Objective questions: Automatic generation of objective questions comprises of three phases of processing namely problem formulation, data preprocessing, and baseline model and distractor generation

E. Problem Formulation

Considering $C = \{w_t^c\}_{t=1}^{t=L_c}$ as the contextual passage, $Q = \{w_t^q\}_{t=1}^{t=L_q}$ as the question and $A = \{w_t^a\}_{t=1}^{t=L_a}$ as the correct answer with lengths L_c, L_q and L_a respectively. The problem here is to design a transfer learning model M to generate a distractor $D = \{w_t^d\}_{t=1}^{t=L_d}$ about the question. The determination of best possible distractors \bar{D} is done in such a way that the conditional likelihood is maximized given by $\bar{D} = \arg \max_D \log \Pr(D|C, Q, A)$

F. Data Preprocessing

The preprocessing stage for objective questions generation is similar to that discussed earlier for subjective questions. The dataset comprising of three fields (C, Q, A) but the difference is that the field A will be a word or a phrase. The length of A as compared to subjective question generation scenario is very small instead. The dataset will be transformed to Sentence-question (S,Q) pair which can be utilized for the training of the transformer model.

G. Baseline Model

The baseline model for objective question generation is FastT5 model where the encoder and decoder are exported to the onnx model separately to achieve the faster speed and low complexity. However, the model is augmented with distractor

generating algorithm to generate other options which are similar to the correct answer but are wrong answers and are used to befuddle the examinee. The keywords extraction in the proposed framework is performed using the multipartite graphs to enhance the context awareness. An objective question, especially a multiple-choice question or fill in the blank question, comprises of three important parts; a question stem, a correct answer and distractors. Due to the limited scope of varying stem or answer, a right set of distractors can greatly control the level and relevance of the automatically generated questions. For MCQs, this task has been performed in this paper through the ConceptNet which is a freely available semantic framework designed to enable the computer understand the meaning of the commonly used words. Word embeddings has been created by ConceptNet [30] by representing the word meanings as vectors. These word embeddings are free, multilingual, aligned across languages, and designed to avoid representing harmful stereotypes. It is a graph knowledge base $G \subseteq C \times R \times C$ where C and R represent the natural language concepts and commonsense relations respectively. ConceptNet contains 32 million triplets where each triplet instance in the respective graph (c_1, r, c_2) represents a commonsense knowledge such as '(Commitment, Leadsto, Success)'. The words and relations for which the created MCQs are referred as seed words and seed relations, respectively. In the generated graph, the directed edge from node c to node p with relation r , p is referred as a parent of node c and c as a child of node p with relation r . The siblings of a node, with respect to a specific relation, are defined as all the children of its parent node, except for the node itself. The selection of distractors for objective questions is done on the basis of hypothesis that the distractors should not share any common property with correct answer. It has been ensured by finding non-overlapping graph communities within words. To attain these leading nodes of non-overlapping community, a one hop expansion is performed in that community and repeated words are removed. The community for each seed word along with its leading nodes is identified to derive the objective questions. Depending on the existence of path between the respective leading node and seed word, distractors from the same community can be chosen using the same seed relation. These generated distractors can be used to derive multiple choice questions, fill-in-the-blanks or True-false questions.

V. EXPERIMENT RESULTS AND DISCUSSION

An experimental analysis has been performed to evaluate the performance of the proposed AQG model. Stanford Question Answering Dataset (SQUAD) has been used in this work which comprises of a rich set of 107785 questions collected from the crowd workers on 536 Wikipedia articles. It is generated through the question-answer pairs from the paragraphs of Wikipedia articles. The selection of Squad in this paper is to derive the questions from a larger spectrum of fields. The available SQUAD dataset consists of two sets: a training set and development set which has been divided further for training, testing and validation. The dataset is having three fields namely context, question and answer. The part of the dataset which is accessible has 490 articles which is divided randomly as given in Table I.

TABLE I. STATISTICS OF DATA SPLIT

Data Split	Articles	Passage-answer pairs	Avg. passage tokens	Avg. question tokens
Train	442	75668	154.32	12.26
Validation	24	10566	159.61	12.58
Test	24	11877	133.64	12.55

The dataset has been used here for the subjective and objective questions generation. The data is first processed using the Wordpiece tokenizer which is a data-driven sub-word level method and consists of 30522 tokens trained from Google. Context selection has been implemented in this work by truncating the input tokens as per the need. The maximum limit of number of tokens in the tokenizer has been chosen 512 tokens as maximum limit so as to attain the optimal training speed and number of truncated data. The limit for number of question tokens has been set to 96 here. The context and answer are given as input to the proposed fast T5 model and question is the output of the transformer.

The proposed T5 model and its training procedures have been implemented using the PyTorch v1.5.0, Pytorch-Lightning frameworks along with the Transformersv3.0.2 from Hugging Face which is a python library providing various transformer architectures. Cross entropy loss and teacher training has been used here for the efficient training performance whereas GPT2 has been used in the decoder. The encoder and decoder of the transformer are exported to the onnx model separately to reduce the computation complexity and the memory requirement considerably. The reasoning and understanding between the sentences in the story to attain the context awareness is enhanced using the multipartite graph-based keyword extraction and FlashText library and the flexibility is achieved through the PyTorch Lightning library. The decoder has utilized the Beam search algorithm in this model so as to reduce the risk of hidden high probability word sequences by keeping the most likely number beams of hypotheses at each time step and help in choosing the hypothesis that has the overall highest probability.

The model training and the evaluation of the proposed framework is performed with Google Cloud Compute Engine. One virtual machine with one 16 GB NVIDIA Tesla T4 GPU has also been used. The sum of token embeddings, segment embeddings and position embedding are set to 0.1. The dropout probability of the self-attention layer and all fully connected layers is also kept 0.1 similar to the original transformer.

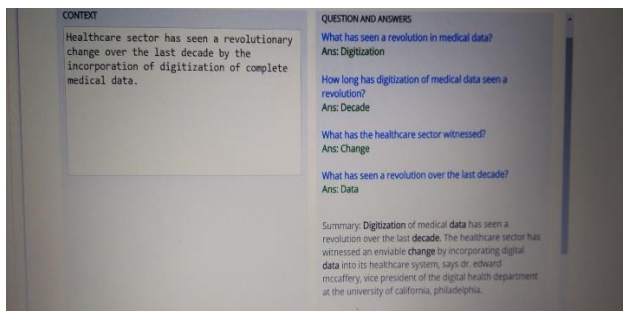
The initial learning rate is set to 5×10^{-5} for the training of the FastT5 model with Adam optimizer. The batch size for the model is kept 64 and is trained only with 2 epochs with equal number of iterations due to the GPU memory constraints. The quality of question generation has been evaluated by BLEU-4 score during the training process which is an automatic performance measuring parameter. It resembles to the 4-gram precision of the hypothesis against the corresponding reference. The performance of the proposed fastT5 model based AQG has also been compared with some conventional

models over BLEU-4 score with 2000 samples. The respective comparison has been shown in Table II.

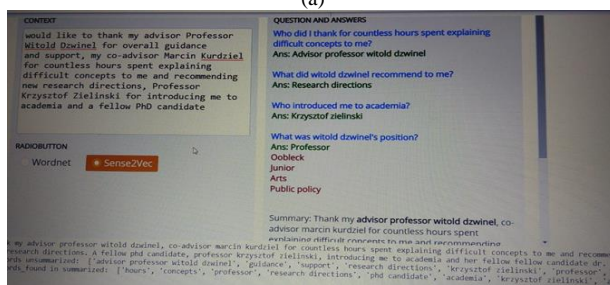
TABLE II. COMPARATIVE ANALYSIS OF THE PROPOSED WORK

Model	BLEU-4
NQG[31]	12.28
M2S+cp [32]	13.98
Ass2s [33]	16.20
S2S-a-at-mp-gsa [34]	16.38
Proposed model	20.28

A graphical user interface (GUI) has been designed in this work using Gradio to present a user-friendly interface to utilize the model as shown in Fig. 3(a) and 3(b). It creates customizable UI components quickly around your TensorFlow or PyTorch models. The GUI shows four fields, namely, context, RadioButton(corpus type), question and answer, and summary field. The user needs to enter the context in the context field and select the suitable corpus; the generated questions and the summary will be automatically displayed in the respective fields. The sense2vec is a powerful variation of word2vec which improves the performance of algorithms like syntactic dependency parsing while significantly reducing computational overhead for calculating the representations of word senses. It enables the model to be implemented for the specific and typical context as well.



(a)



(b)

Fig. 3. (a) Graphical user interface designed in gradio for AQG
(b) Graphical user interface designed in gradio for AQG.

VI. CONCLUSION

A challenging problem of automatic subjective and objective question generation with context awareness has been addressed in this work using unified text to text transfer learning. It presents a fastT5 based transformer model which reframes natural language processing tasks into a unified text-

to-text-format and augments it with word sense disambiguation (WSD), concept net and domain adaptation framework to improve the meaningfulness of the questions. The proposed T5 model and its training procedures have been implemented using the PyTorch v1.5.0, Pytorch-Lightning frameworks along with the Transformersv3.0.2 from Hugging Face which is a python library providing various transformer architectures. Beam-search decoding algorithm has been used here to reduce the model size and increase the speed of the model through quantization of the whole model by Open Neural Network Exchange (onnx) framework. The keywords extraction in the proposed framework is performed using the Multipartite graphs to enhance the context awareness. It can be used to derive multiple choice questions, fill-in-the-blanks or True-false questions. The qualitative and quantitative performance of the proposed AQG model is evaluated through a comprehensive experimental analysis over the publicly available SQuAD dataset. The research can further be extended in future with the more advance text transformers. The computation complexity can also be addressed in the future research work.

REFERENCES

- [1] Lane, H. C. and Vanlehn, K. "Teaching the tacit knowledge of programming to novices with natural language tutoring", Journal Computer Science Education, 15, pp. 183–201, 2005.
- [2] Graesser, A. C., Rus, V., D'Mello, S. K., and Jackson, G. T., "AutoTutor: Learning through natural language dialogue that adapts to the cognitive and affective states of the learner", In D. H. Robinson & G. Schraw (Eds.), Recent innovations in educational technology that facilitate student learning, Information Age Publishing, pp. 95-125, 2008.
- [3] Heath, T. and Bizer, C., "Linked Data: Evolving the Web into a Global Data Space", Morgan & Claypool Publishers, 2011
- [4] Ali, H., Chali, Y. and Hasan, S. A., "Automation of question generation from sentences. In Boyer, K. E. and Piwek, P. (eds.), Proceedings of the 3rd Workshop on Question Generation, held at ITS 2010, pp.58-67, 2010.
- [5] Chen, W., Aist, G., and Mostow, J., "Generating questions automatically from Informational text", In: Craig, S. D. & Dicheva, D. (eds.), Proceedings of the 2nd Workshop on Question Generation, held at AIED 2009, pp.17-24. 2009
- [6] Jouault, C., and Seta, K., "Building a Semantic Open Learning Space with Adaptive Question Generation Support", In Proceedings of the 21st International Conference on Computers in Education, 2013.
- [7] Amidei, J., Piwek, P., Willis, A., "Evaluation methodologies in automatic question generation", , In Proceedings of The 11th International Natural Language Generation Conference (pp. 307–317). Tilburg University: Association for Computational Linguistics, pp. 2013-2018, 2018.
- [8] Chen, G., Yang, J., Hauff, C., Houben, G.-J. , "Learningq: A large-scale dataset for educational question generation", In Twelfth International AAAI Conference on Web and SocialMedia, pp. 481–490, 2018.
- [9] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805, 2018.
- [10] Li Dong, Nan Yang, Wenhui Wang, Furu Wei, Xiaodong Liu, Yu Wang, Jianfeng Gao, Ming Zhou, and Hsiao-Wuen Hon. Unified language model pre-training for natural language understanding and generation. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett (eds.), Advances in Neural Information Processing Systems 32, pp. 13063–13075. 2019.
- [11] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. Roberta: A robustly optimized BERT pretraining approach. arXiv preprint arXiv:1907.11692, 2019.

- [12] Zhenzhong Lan, Mingda Chen, Sebastian Goodman, Kevin Gimpel, Piyush Sharma, and Radu Soricut. Albert: A lite BERT for self-supervised learning of language representations. arXiv preprint arXiv:1909.11942, 2019.
- [13] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J Liu. Exploring the limits of transfer learning with a unified text-to-text transformer. arXiv preprint arXiv:1910.10683, 2019.
- [14] Rus, V., Cai, Z. & Graesser, A., "Question Generation: Example of A Multi-year Evaluation Campaign", In: Rus, V. and A. Graesser (eds.), Online Proceedings of 1st Question Generation Workshop, NSF, Arlington, VA. 2008.
- [15] Michael Heilman and Noah A. Smith. "Good Question! Statistical Ranking for Question Generation". In: HLT-NAACL. 2010.
- [16] Kaustubh D. Dhole and Christopher D. Manning. "Syn-QG: Syntactic and Shallow Semantic Rules for Question Generation". In: ArXiv abs/2004.08694 (2020).
- [17] Weiming Wang, T. Hao, and W. Liu. "Automatic Question Generation for Learning Evaluation in Medicine". In: Advances in Web Based Learning – ICWL 2007 4823 (2007), pp. 242–251.
- [18] A. R. Fabbri et al. "Template-Based Question Generation from Retrieved Sentences for Improved Unsupervised Question Answering". In: ACL. 2020.
- [19] Qingyu Zhou et al. "Neural Question Generation from Text: A Preliminary Study". In: NLPCC. 2017.
- [20] Linfeng Song et al. "Leveraging Context Information for Natural Question Generation". In: NAACL-HLT. 2018.
- [21] Xingdi Yuan et al. "Machine Comprehension by Text-to-Text Neural Question Generation". In: Rep4NLP@ACL. 2017.
- [22] Yanghoon Kim et al. "Improving neural question generation using answer separation". In: Proceedings of the AAAI Conference on Artificial Intelligence 33 (2019), pp. 6602–6609.
- [23] Mitkov, R., Ha, L.A., Varga, A., Rello, L.: Semantic similarity of distractors in multiplechoice tests: extrinsic evaluation. In: Proceedings of the EACL 2009 Workshop on GEometrical Models of Natural Language Semantics, pp. 49–56 (2009).
- [24] Abigail See, Peter J. Liu, and Christopher D. Manning. 2017. Get to the point: Summarization with pointer generator networks. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 1073–1083.
- [25] Ashish Vaswani et al. "Attention is All you Need". In: Advances in Neural Information Processing Systems 30. Ed. by I. Guyon et al. Curran Associates, Inc., 2017, pp. 5998–6008.
- [26] Łukasz Kaiser and Samy Bengio. Can active memory replace attention? In Advances in Neural Information Processing Systems, (NIPS), 2016.
- [27] Nal Kalchbrenner, Lasse Espeholt, Karen Simonyan, Aaron van den Oord, Alex Graves, and Koray Kavukcuoglu. Neural machine translation in linear time. arXiv preprint arXiv:1610.10099v2, 2017.
- [28] Jonas Gehring, Michael Auli, David Grangier, Denis Yarats, and Yann N. Dauphin. Convolutional sequence to sequence learning. arXiv preprint arXiv:1705.03122v2, 2017.
- [29] Rajpurkar, P., Zhang, J., Lopyrev, K., Liang, P. (2016). SQuAD: 100,000+ Questions for Machine Comprehension of Text, In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing (pp. 2383–2392). Austin: Association for Computational Linguistics.
- [30] Robyn Speer and Catherine Havasi. Representing general relational knowledge in ConceptNet 5. In Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12), pp. 3679–3686, Istanbul, Turkey, May 2012. European Language Resources Association (ELRA).
- [31] Qingyu Zhou, Nan Yang, Furu Wei, Chuanqi Tan, Hangbo Bao, and Ming Zhou. 2017. Neural question generation from text: A preliminary study. In NLPCC
- [32] Linfeng Song, Zhiguo Wang, Wael Hamza, Yue Zhang, and Daniel Gildea. 2018. Leveraging context information for natural question generation. In Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers), pages 569–574.
- [33] Yanghoon Kim, Hwanhee Lee, Joongbo Shin, and Kyomin Jung. 2018. Improving neural question generation using answer separation. In AAAI.
- [34] Yao Zhao, Xiaochuan Ni, Yuanyuan Ding, and Qifa Ke. 2018b. Paragraph-level neural question generation with maxout pointer and gated self-attention networks. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, pages 3901–3910.

Multistage End-to-End Driver Drowsiness Alerting System

Sowmyashree P, Sangeetha J

Department of Computer Science and Engineering, M S Ramaiah Institute of Technology, MSR Nagar, Bangalore- 560 054, India

Abstract—Drowsiness in drivers is the major cause for these fatal road accidents. Hence detecting drowsiness in drivers and alerting them on time is very important to avoid accidents. Researchers have developed several techniques to detect drowsiness in driver and warn the driver. However, in the past there is no work done towards the end-to-end driver drowsiness alerting system. Therefore, in this proposed system, it will ensure that the driver is awake through its end-to-end multi-stage (i.e., three stage) alerting system. The proposed system, at first performs driver authentication. Next, it detects the driver's face and also checks whether he/she has consumed alcohol or not, in either case the car engine will not start, and a warning mail is sent. Then the system performs drowsiness detection. If the driver is found drowsy then a multi-stage alerting system (i.e., voice alert, seat vibration alert and physical alert) is performed to wake him/her. After the voice alert, the driver has to give his/her fingerprint as proof for not being drowsy. If the system fails to get a fingerprint it starts the vibration alert. Once again system asks for driver's fingerprint, without which the system starts physical alert through robot arm which is performed with three different frequencies (i.e., Low, Medium and High) and three questions are asked after each frequency to make sure the driver is alert. In the process, it creates a log file which contains the driver's drowsiness details, after analyzing which it gives rating to the driver and mail this rating to the concerned person. This rating can be used to choose the driver for a safe and comfortable journey. Thus, the system ensures that driver is alert and avoids road accidents.

Keywords—Driver drowsiness detection; internet of things; voice alert; seat-vibration alert; physical alert; driver rating; haar cascade classifier; eye aspect ratio

I. INTRODUCTION

The Internet of Things (IoT) indicates billions of appliances worldwide connected to the internet, gathering and distributing data. IoT is a technology that has steadily gained momentum and is now quietly changing our future. The world around us is becoming more intelligent and receptive because of the Internet of Things, which is merging the digital and physical worlds. IoT technologies have a good range of applications in agriculture, health sector, homes, cities, environment, industry and transport system [1]. In the contemporary world, transportation plays a significant role. People rely on automobiles as their primary source of transportation. In 2021, 79.1 million vehicles were produced globally, which is 1.3 percent more than the year 2020 [2]. Although the automobile has changed how people live and made doing everyday activities easier, it also has a connection to a variety of negative effects, such as road accidents.

As reported in a survey by Foundation for Traffic safety [3], over 96 percent of motorists consider drowsy driving to be highly or extremely risky. Only significantly fewer than 40% of respondents believed that fatigued drivers ran the chance of being pulled over by the police. Approximately 27% of drivers acknowledged that they had driven while feeling so sleepy that it was difficult, at least once in the previous 30 days to keep their eyes open, despite high rates of perceived risk and personal/social disapproval surrounding drowsy driving. According to [4], in India there were nearly 4.49 lakh of road accidents in the year of 2019. Our nation has faced 4,69,418 number of serious injuries and 1,51,417 number of deaths in the year 2019. Each year sleepy driving has caused around 100,000 road mishaps and nearly 1,500 deaths in the United States [5]. Therefore, being drowsy poses a severe risk to road safety, resulting in serious injuries, fatalities, and financial losses. Therefore, it is essential to detect drowsiness in drivers and alert them in time.

Many researchers have implemented numerous techniques to identify driver drowsiness and alert him/her. These researchers have considered different parameters to detect drowsiness. Some of them have used parameters like facial expressions [6], head position [7], blink rate [8], eye closure rate [9, 10] and yawning [11]. Many other systems use driver's biological conditions such as heartbeat rate [12, 13], pulse rate [14], galvanic skin response [15] and also features like steering wheel angle [16, 17], vehicle lane [18], speed variation, etc. Some researchers have combined these parameters for attaining better results [19, 20]. Most of the researchers have considered eye as the region of interest to find drowsiness. Eye Aspect Ratio (EAR) [21], Percentage of Eye Closure (PERCLOS) [9], Template Matching [9, 16] are some of the algorithms used for detecting drowsiness. EAR and PERCLOS gives good accuracy among these algorithms. Many authors have also used different Machine learning algorithms. Classifiers used for classifying drowsy and non-drowsy are SVM [6, 14, 18, 20], CNN [7, 8, 13]. CNN gives good accuracy, but it is computationally intensive.

The authors in [21] have developed a real-time system, implemented using mobile application which detects drowsiness using Eye Closure Ratio (ECR) and Eye Aspect Ratio (EAR). They use Dilib library to detect facial characteristics like eye and eyelids. Then by using these features EAR output is computed and a threshold is set for this output which classifies driver as drowsy or non-drowsy. When the driver is found drowsy, he is given an alert using mobile alarm. In [22], the researchers have created a fatigue detection system depending on grayscale image scanning and

PERCLOS. The suggested system comprises of three sections. In the first section, it determines where the driver's face is generally located in grayscale photos and next it examines the eye locations using a mini template. In second section, it creates a fatigue model applying the information from the previous part and PERCLOS. Depending on the driver's unique sleep model, in the third section, the system continually assesses the driver's condition. When the driver is found drowsy, the system warns him/her using an alarm sound. A drowsiness detection system which detects drowsiness by observing the eye movements is proposed in [23]. By combining data from numerous successive video frames with machine learning models' capacity to recognize various eye behaviors, a temporal aspect is added to the system. EAR value and blink classification techniques are used for detecting drowsiness in drivers. Here each video frame contains a calculation and storage of the EAR values. The time dimension was added to the ML model by concatenating a certain number of consecutive EAR values. Previously known blink patterns are used to assess if a user is sleepy or not. When the driver is found sleepy, the system sends a caution message, and a voice is played to warn the driver. The suggested methodology also includes a user feedback process to modify models based on particular user's feedback to produce even better outcomes.

A drowsiness detection system to examine the potential of smart watches and other wrist-worn wearables for the identification of driver drowsiness across a range of age groups is suggested in [24]. The authors have organized two simulators for this study – a low-level simulator and a high-level simulator. Study with high-level simulator was undertaken with two age groups to more thoroughly validate the methodology. Estimation of the heart rate signal and driving status were taken in the first step. The gathered information was prepared for feature extraction in the subsequent step. The final phase involved creating data sets by labeling the features in accordance with the subject's detected state during the simulator drive. A variety of machine learning methods were then used to classify the labeled data. The findings showed that drowsiness varies by age group i.e., younger drivers are more likely to become sleepy than older drivers. Authors have not discussed about in what way they will alert the driver once he/she is found drowsy. The study in [25] discusses a deep learning network-based Electroencephalogram (EEG) categorization system for drowsiness detection. The device EmotivEPOC+ headset is used to collect EEG signals, which is then preprocessed. Data preparation, signals interpretation, and data amplification are the three main focuses of the preprocessing stage. In data preparation, noises are removed from the data. Signal interpretation is based on research into Alpha-Theta waves from the occipital and temporal areas, which are used to assess alertness and drowsiness, respectively. By multiplying the vectors at random intervals, segments were extended to a predetermined length in the data amplification. Then the proposed model is analyzed by using different CNN networks. The system achieved an accuracy of 90.14%.

From all these approaches, it can be seen that the way of warning the driver when he is found drowsy is not specified in many of the works. Many researchers have used simple

buzzers or alarms to alert the driver and there is no way provided to ensure that the driver is awake after the buzzer or alarm goes off. It is important to have a good system to alert the driver, because many people have the habit of sleeping off after the alarm stops. In a real scenario, when a person is in drowsy state, simple alarms or buzzers will not help him/her to be alert. Hence, the proposed system introduces a multi-stage alerting system to wake the driver when he/she is found drowsy.

The proposed system starts with driver authentication by detecting the driver's face and also checks whether he/she has consumed alcohol or not, in either case the car engine will not start, and a warning mail is delivered to the concerned person. Then the system performs drowsiness detection. If the driver is found drowsy then a multi-stage alerting system (i.e., voice alert, seat vibration alert and physical alert) is performed to wake him/her. After the voice alert, the driver has to give his/her fingerprint as proof for not being drowsy. If the system fails to get a fingerprint, it starts the vibration alert. Then again driver has to give his/her fingerprint, without which the system starts physical alert through robot arm which is performed with three different frequencies (i.e., Low, Medium and High) and three questions are asked after each frequency to make sure the driver is awake. If the driver is found alert after any of these multistage alerts, the system continues with detecting drowsiness. In the process, it creates a log file containing the driver's drowsiness details, by analyzing which it gives rating to the driver and mail the rating to the concerned person. This rating can be used by passengers while choosing drivers for their journey. Thus, the system ensures that driver is alert and avoids road accidents to save lives. This paper is separated into III principle segments. In segment II we discuss the methodology used in the proposed system. Result of the proposed system is discussed in segment III. The conclusion is given in the end.

II. METHODOLOGY

The proposed system has two modules – Driver Authentication and Driver Drowsiness Detection and Alerting. Driver authentication module checks for authorized driver and alerts the owner in case the driver is not authorized. Driver drowsiness detection alerting module detects drowsiness in driver and alerts him until he is awake with its multistage alerts. The Fig. 1 represents the flow chart of the suggested system.

A. Driver Authentication

The proposed system starts with driver authentication. As soon as the driver enters the car the system does a face recognition using Haar Cascade Classifier. If the face matching fails, a warning message is delivered to the possessor of the vehicle and the vehicle's engine will not start. After successful face recognition, the system does an alcohol test on the driver using alcohol sensor. If he/she has been detected positive, then the vehicle's engine will not start, and a warning message is sent to the concerned person. Hazards due to drunk driving can be prevented by this alcohol test. If no alcohol content is detected, the system moves to drowsiness detection and alerting module.

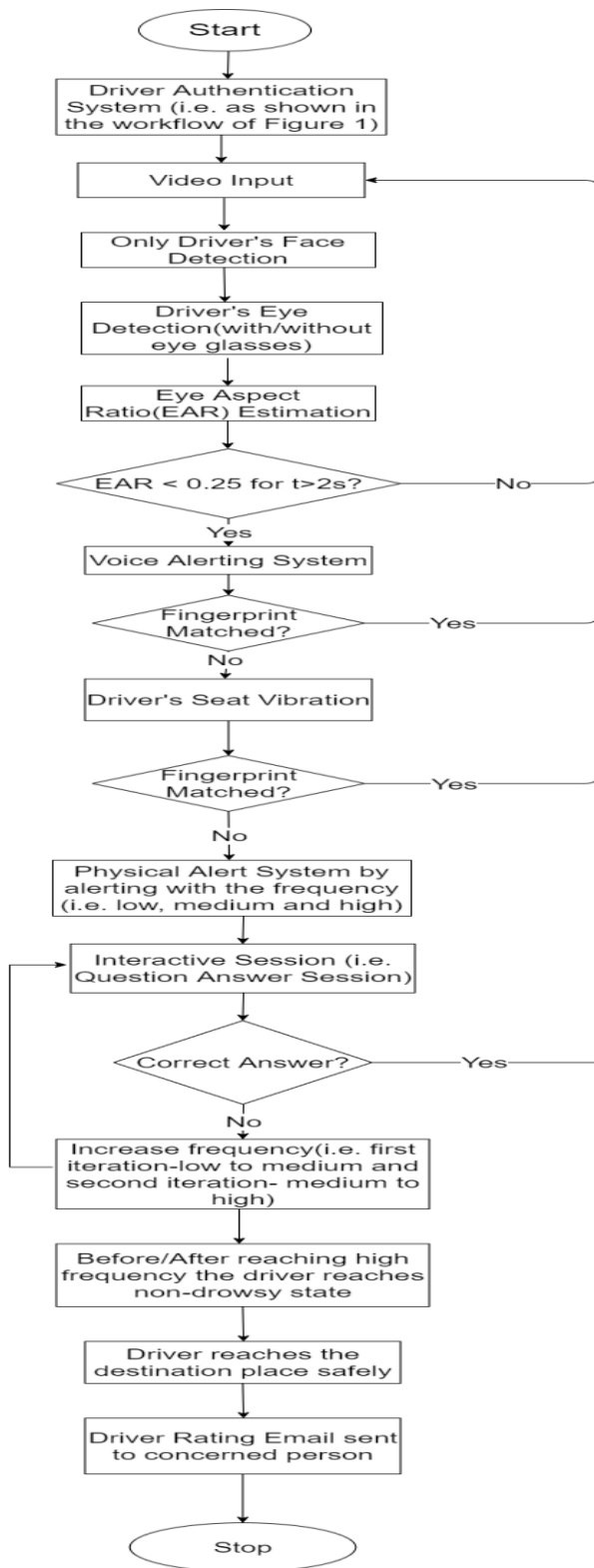


Fig. 1. Flowchart of the proposed system.

B. Driver Drowsiness Detection and Alerting

The system, as seen in Fig. 1 takes continuous video input of the vehicle driver's face and performs face detection and eye detection respectively. After the eye detection, the proposed

system performs EAR (Eye Aspect Ratio) estimation. If the estimated EAR is greater than threshold amount (i.e., 0.25) then the driver is not in sleepy state and hence the system again goes back to monitor the video input. If the estimated EAR is less than threshold amount for more than 2 seconds, then the driver is drowsy. Now, the system undergoes multi-stage alerting (i.e., voice alert, driver's seat vibration alert and physical alert).

1) *Voice alerting system*: When a sleepy driver is discovered, the voice alert is performed by telling the driver to be alert. Next the system asks for the fingerprint of the driver. If the system gets the fingerprint input, the driver is awake, and the system goes back to monitor the video input. If fingerprint input is not provided by the driver, then the system starts the second stage of alerting (i.e., Seat Vibration Alert).

2) *Driver's seat vibration alerting system*: In this stage, the system vibrates the driver's seat of the vehicle to awake him/her. After the car seat vibrates, the system again asks driver to provide his fingerprint, if it is given by the driver then he is awake, and the system goes back to video input monitoring. If the fingerprint is not obtained, then the system starts the third stage of alerting.

3) *Physically alerting the driver by using robot arm*: The proposed system physically alerts the driver by three measures of frequencies (i.e., low, medium and high frequency). After physically alerting the driver with low frequency, our proposed system asks the driver's name, father's name and place. If correct answer is provided by the driver, then the system continues to observe the video input of the driver (i.e., he/she is in non-drowsy state). If it receives wrong answer, then the system again starts physically alerting the driver with the medium frequency. Likewise, the physical alerting system alerts the driver with high frequency.

All the above stages of drowsy detection and alerting are recorded in the form of log file and using this file it gives rating to the driver as Excellent, Good, Average and Bad. This rating is then mailed to the concerned person. The system stops once the driver reaches the destination.

C. Haar Cascade Classifier

Haar cascade classifier suggested by Viola and Jones [26] is one among the few object identification algorithms. We use this algorithm in the proposed system for driver's face detection. For detecting a face, this technique uses four steps: Haar-like attribute, intrinsic image, AdaBoost learning, and The Attentional Cascade. The camera module captures driver's facial behaviour in the form of video. This video is divided into many image frames and these image frames are given as input to Haar Cascade Classifier.

1) *Haar-like Attribute*: Collecting Haar-like attributes on the input image frame is the first step in Haar Cascade Classifier. The crucial component of the Haar cascade classifier is the Haar-like attributes. These attributes make it simple to determine lines or borders and parts where there is a sudden change in the pixel brightness. These rectangular attributes are moved across the image, and the aggregate

number of pixels in the white and black parts is subtracted from one another. The darker portion has a pixel value of 1 and lighter portion has a pixel value of 0. We compute the aggregate of all the image pixels situated in the Haar attributes black and white areas respectively and then Haar value is calculated by taking the difference between them as shown in equation (1). If there is an edge in the image separating light pixels on the left from dark pixels on the right, the Haar value will be close to 1.

$$\text{Haar value} = (\text{Sum of dark pixels} / \text{Number of dark pixels}) - (\text{Sum of light pixels} / \text{Number of light pixels}) \quad (1)$$

The edge attribute is used to detect edges, is the difference between the sums of the pixels inside the two rectangular sections. The line feature determines the sum within two outer rectangles subtracted from the sum within a centre rectangle and is used mostly for line detection. The four-rectangle attribute calculates the difference between rectangle diagonal pairings.

2) *Intrinsic image*: Every pixel in an intrinsic image is calculated by using the primary image so that it is equivalent to the sum of all the pixels to its upper right and to its left. The amount at some point (m, n) is given by equation (2).

$$I(m, n) = \sum_{\substack{m' < m \\ n' < n}} i(m', n') \quad (2)$$

where $i(m, n)$ is the original image and $I(m, n)$ is the integral image. The value of the intrinsic image, $I(m, n)$, is derived by adding the preceding index values from left to right. Over the primary image, the integral image can be calculated in a single pass by applying the following equations (i.e., equation (3) and equation (4)).

$$s(m, n) = s(m, n - 1) + i(m, n) \quad (3)$$

$$I(m, n) = I(m - 1, n) + s(m, n) \quad (4)$$

The Intrinsic Image's endmost pixel in the bottom right corner will be equal to the aggregate of all the pixels in the Primary Image. For whatever attribute size, the Intrinsic Image demands only four continual value additions every time to calculate the Haar value. The total of $i(m, n)$ across the rectangle extended by A, B, C, and D is shown in equation (5).

$$\sum_{\substack{m_0 < m < m_1 \\ n_0 < n < n_1}} i(m, n) = I(D) + I(A) - I(B) - I(C) \quad (5)$$

Where $i(m, n)$ is the sum of pixel values of the rectangle ABCD and the values $I(A), I(B), I(C)$ and $I(D)$ are the pixel values at position A, B, C and D respectively.

3) *AdaBoost algorithm*: After finding Haar attributes in the input image, we have to make a classification function, which will classify the input image frames into two classifications (i.e., the input image that contains face and that does not contain face). To achieve this, we need a training data set which contains images with and without faces, and the Haar attribute set. In the selected Haar attributes, majority of them won't complement the face features nicely or will be unnecessary. Typically, a basic 24×24 image yields

approximately 1,60,000 attributes. Many of these Haar attributes will be irrelevant while performing face detection. So, in this case, a feature selection strategy is required to choose a subset of attributes from the vast set, which will not only choose attributes that perform better than the rest but will also eliminate the unnecessary ones. For this purpose, we use a boosting technique called AdaBoost which is an iterative machine learning process which will select subset of attributes and also train the classifier.

AdaBoost uses the combination of several weak classifiers generated using same set of training dataset which creates a strong classifier. Training data set contains many images with and without human face. At first, on the training set, a basic classifier is trained and predictions are made. Each sample is initially given the same weight. The weight will be decreased if the sample is correctly classified and increased otherwise. This will make the incorrectly classified sample stand out so that a new sample set may be created. Next, the weak classifier is created by training the new sample set and it makes predictions based on the training set, updates the weights, and so on. By overlapping these weak classifiers, the strong classifier will be created. At the end we will be left with most pertinent features needed for face detection. In this way, the number of features selected is reduced to 6000.

4) *The attentional cascade*: After selecting Haar attributes and training the classifier, now system detects face in the input image frame. The Attentional Cascade is an approach to building a cascade of classifiers that improves face detection performance while drastically cutting down on calculation time. AdaBoost-trained classifiers are used to build the stages of the cascade. Fig. 2 represents the architecture of the cascade classifier.

The Haar attributes selected after applying AdaBoost are run on the input image frames by using cascade classifier to detect the face. This is done with the presumption that not all the Haar attributes need to be run on every window. We can say that the face features are absent on a specific sub-window if one Haar attribute fails on that sub-window and hence we can move to next sub-window where there might be a facial feature present. The cascade classifier works as below:

1) The classifier divides the input image into smaller sections, or sub-windows.

2) We then make use of a cascade ordering of N detectors, each of which picks up a type of Haar attributes from the image frame it is fed with. After attribute extraction is finished, a confidence rating is given for each sub-window.

3) The sub-windows (or images) that have the high confidence rating are identified as faces and are submitted to the accumulator while the remaining images are rejected. The cascade then restores the succeeding frame or image, if any, and continues the process.

The process continues in the same manner and at the last stage driver's face is detected.

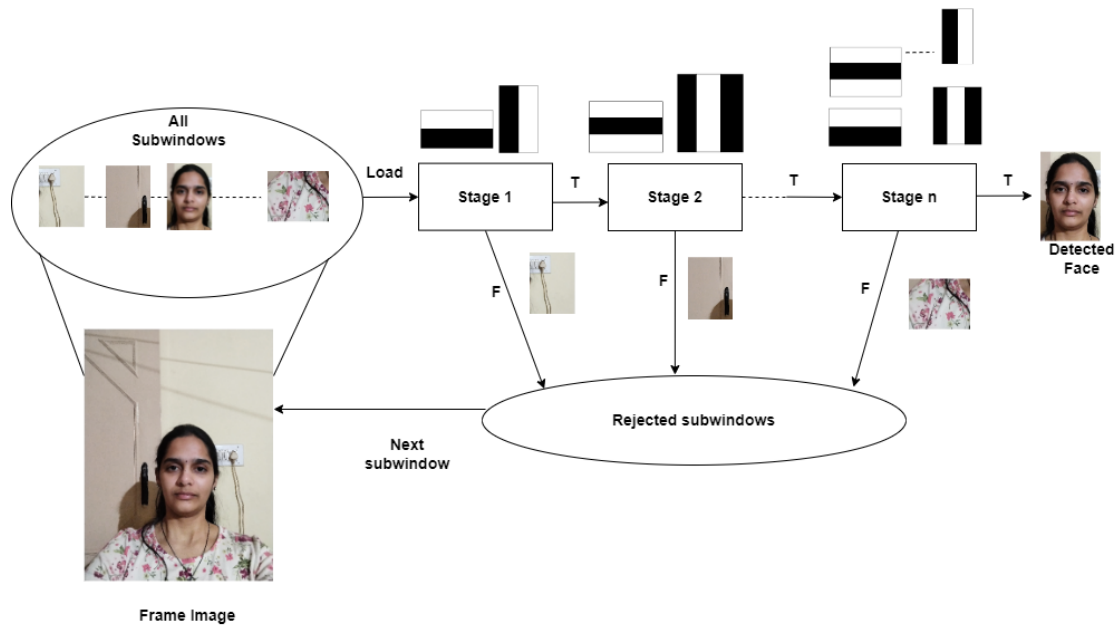


Fig. 2. Cascade classifier architecture.

D. Eye Aspect Ratio (EAR)

After the driver’s face is detected, face landmarks are detected using Dlib library. We use the Dlib library’s pre-trained model to initialize the face landmark detector. After detecting facial landmarks, the drowsiness is detected on the basis of eye blink rate. Eye Aspect Ratio (EAR) [27], detects eye blink. Suppose that the driver flickers his/her eyes repeatedly, it indicates that they are drowsy. In order to determine the eye blink frequency, it is important to precisely detect the eyes shape. Based on the eye landmarks detected previously using Dilib library, the EAR value is calculated and is used to assess the eye-opening state. For every video frame six landmark coordinates around the eyes are taken. Then equation (6) is applied to compute the eye aspect ratio.

$$EAR = \frac{||X2-X6||+||X3-X5||}{2||X1-X4||} \quad (6)$$

Where X1, X2, X3, X4, X5 and X6 are the 2D landmark positions. As indicated in Fig. 3(a), the X2, X3, X5, and X6 are utilized to compute height, whereas the X4 and X1 are utilized to compute eye width in meters (m). The EAR remains persistent when the eye is open, but as the eye gets closed, it rapidly goes to zero, as indicated in Fig. 3(b).

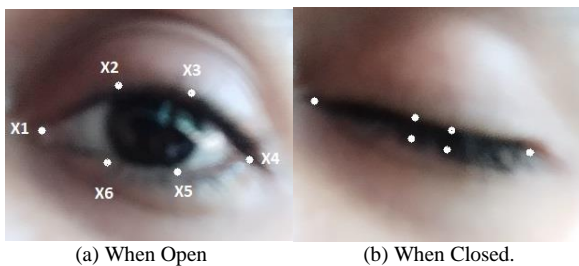


Fig. 3. Eye landmarks.

$$EAR = \begin{cases} y > 0; \text{eyes open} \\ y \cong 0; \text{eyes close} \end{cases} \quad (7)$$

Equation (7) illustrates the EAR value extent when the eyes are closed and open. When eyes are open, the EAR can be some integer value y that is larger than 0, but when they are closed, the EAR will be near to 0. A threshold amount of 0.25 is taken for deciding whether the person is drowsy or not, If the EAR value remains close to 0.25 or less for a time of 02 seconds or more, then the driver is considered to be sleepy [28].

III. RESULTS AND DISCUSSION

In this paper, we are proposing a Multistage End-to-End Driver Drowsiness Alerting System to detect drowsiness in drivers and alert them with multi-stage alerting system. The system detects drowsiness both in day and night condition, and also when the driver is wearing eye glasses. It is implemented using Raspberry Pi 3 Model B. Picamera has been utilized for recognizing and monitoring the driver’s face and Infrared (IR) Light Emitting Diode (LED) is used during night driving. The system starts with driver authentication by detecting driver’s face and recognizing it for an authorized driver. The Haar cascade classifier is applied for face recognition.

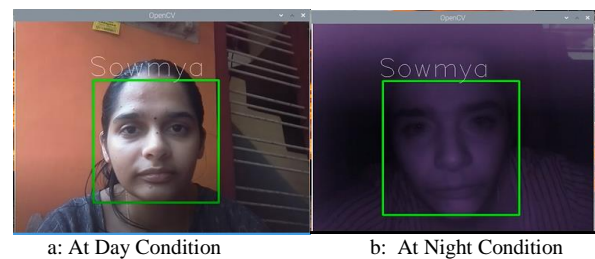


Fig. 4. Face authentication.

Fig. 4(a) displays the result of face authentication at daytime and Fig. 4(b) shows the outcome of face authentication at nighttime. Here the face is identified and recognized both at day and night conditions. A green box is

drawn around the identified face and his name is displayed on top of this box. The system then tests for alcohol consumption using alcohol sensor as shown in Fig. 5. If the driver is not authenticated or if he has consumed alcohol, then the system stops the vehicle engine, and an alert email is delivered to the vehicle owner. The alert mail is given in Fig. 6.

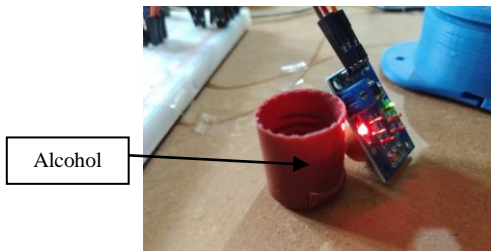


Fig. 5. Alcohol sensor.

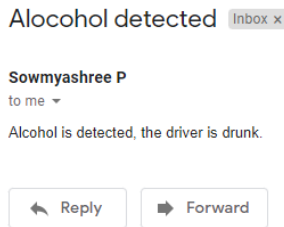
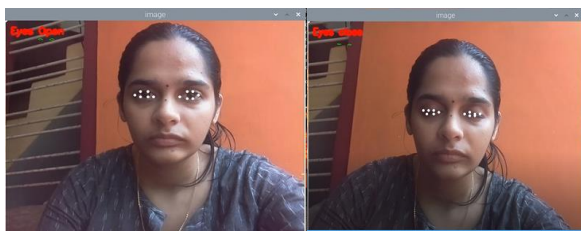


Fig. 6. Alcohol detection warning email.

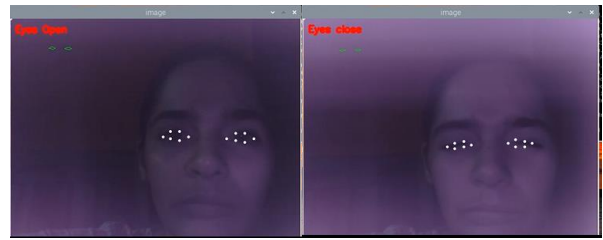
After driver authentication, the system starts detecting drowsiness. The system detects the shape of both left and right eye and estimates EAR to determine whether the eye is open or closed. A threshold value is taken to decide whether the eye is closed or open. If the EAR output remains less than threshold for greater than 2 seconds, the person is said to be drowsy. The system can detect drowsiness in both day and night condition and also when the driver is wearing eye glasses. The result of eye detection when the driver's eyes are open in day condition is shown in Fig. 7(a) and 7(b) show the result of eye detection when driver is drowsy.



a: Open Eye b: Drowsy Eye

Fig. 7. Eye detection at day condition.

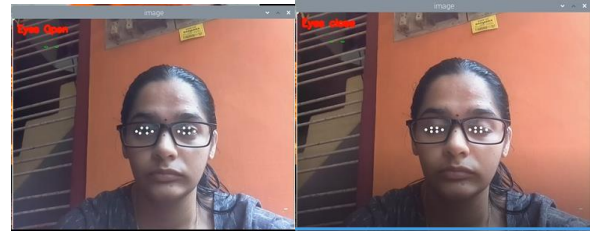
Fig. 8(a) and 8(b) show the result of eye detection when the driver's eyes are open, and eyes closed in night condition respectively.



a: Open Eye b: Drowsy Eye

Fig. 8. Eye detection at night condition.

The result of eye detection while driver is wearing eye glasses in day time is shown in Fig. 9.



a: Open Eye b: Drowsy Eye

Fig. 9. Eye detection with eye glasses.

After performing drowsiness detection, if the driver is found drowsy then a multistage alerting system (i.e., voice alert, seat vibration alert and physical alert) is performed to wake him/her. The multistage driver drowsiness alerting system consists of:

A. Voice Alerting System

The system continuously monitors the driver's eye for drowsiness. It detects drowsiness by applying EAR algorithm. If the driver is found drowsy, as seen in Fig. 10, then the system starts the first stage alert, i.e. voice alert.

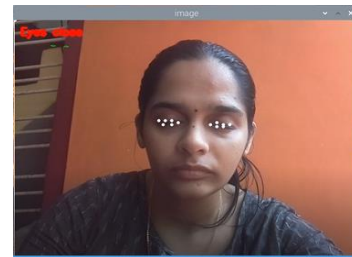


Fig. 10. Drowsy eyes.

Fig. 11 shows the voice alerting system. We use a Bluetooth speaker for providing voice alert to the driver. When the voice alert is initialized, a voice saying "wake up" is played thrice using the speaker, shown in Fig. 11(a). Fig. 11(b) shows respective output.



a: Bluetooth Speaker.

```
pi@raspberrypi: ~/Desktop
File Edit Tabs Help
pi@raspberrypi:~/Desktop $ python main.py
Initializing the system
Driver is authorised..., Testing for Alcohol
Alcohol Not Detected !
Initializing drowsiness detection
Drowsiness detected

Initializing voice alert
Voice alert completed
```

b:Output

Fig. 11. Voice alerting system.

Once the speaker stops playing the voice, the system asks the driver to give his/her fingerprint as a proof for not being drowsy. Fig. 12(a) shows the fingerprint sensor used in the system and Fig. 12(b) shows the respective output. If the system reads a fingerprint and it is matched with driver's fingerprint, then it indicates that the driver is alert. The system continuously observes the driver's eyes for detecting drowsiness. If the driver is still in drowsy state and the system doesn't receive fingerprint, then it starts the next stage alert (i.e., Seat Vibration Alert).



a: Fingerprint sensor

```
pi@raspberrypi: ~/Desktop
File Edit Tabs Help
pi@raspberrypi:~/Desktop $ python main.py
Initializing the system
Driver is authorised..., Testing for Alcohol
Alcohol Not Detected !
Initializing drowsiness detection
Drowsiness detected

Initializing voice alert
Voice alert completed
Waiting for finger...
Finger print matched

driver is alert
```

b: Output

Fig. 12. Fingerprint module.

B. Seat Vibration Alerting System

After voice alert, if the system does not receive the fingerprint, then the next stage of alerting system (i.e., seat vibration alert) is initialized. Here the driver's seat vibrates for one minute time to make him alert. Fig. 13(a) shows the DC motor we used to provide driver seat vibration alert and Fig. 13(b) shows the respective output. Once the seat vibration stops, the system asks for driver's fingerprint as shown in Fig. 14. If driver provides a fingerprint and it is matched, then the driver is alert and hence the system goes back to monitor driver's eye. Else, if no fingerprint is provided, the system starts next stage alert (i.e., Physical Alert).



a: DC Motor

```
pi@raspberrypi: ~/Desktop
File Edit Tabs Help
pi@raspberrypi:~/Desktop $ python main.py
Initializing the system
Driver is authorised..., Testing for Alcohol
Alcohol Not Detected !
Initializing drowsiness detection
Drowsiness detected

Initializing voice alert
Voice alert completed
Waiting for finger...
no finger print found

Initializing seat vibration alert
Seat Vibration alert completed
```

b: Output

Fig. 13. Seat vibration alerting system.

```
pi@raspberrypi: ~/Desktop
File Edit Tabs Help
pi@raspberrypi:~/Desktop $ python main.py
Initializing the system
Driver is authorised..., Testing for Alcohol
Alcohol Not Detected !
Initializing drowsiness detection
Drowsiness detected

Initializing voice alert
Voice alert completed
Waiting for finger...
no finger print found

Initializing seat vibration alert
Seat Vibration alert completed
Waiting for finger...
Finger print matched

driver is alert
```

Fig. 14. Output of fingerprint module after seat vibration alert.

C. Physical Alerting System

After seat vibration alert if the system does not receive the fingerprint, the system starts physical alert through robot arm which is performed with three different frequencies (i.e., Low, Medium and High Frequency). Fig. 15 shows the robot arm used to provide physical alert and respective output is shown in Fig. 16.

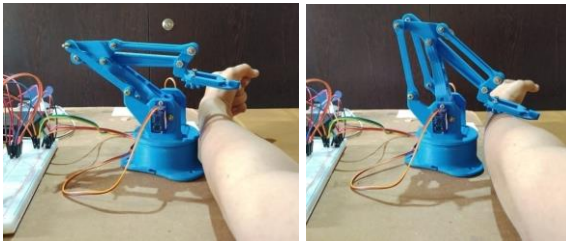


Fig. 15. Physical alert through robot arm.

```
pi@raspberrypi: ~/Desktop
File Edit Tabs Help
pi@raspberrypi:~/Desktop $ python main.py
Initializing the system
Driver is authorised..., Testing for Alcohol
Alcohol Not Detected !
Initializing drowsiness detection
Drowsiness detected

Initializing voice alert
Voice alert completed
Waiting for finger...
no finger print found

Initializing seat vibration alert
Seat Vibration alert completed
Waiting for finger...
no finger print found

Initializing physical alert
Initializing low Frequency physical alert
Low Frequency Robot arm alert completed
```

Fig. 16. Output of physical alerting system.

After low frequency physical alert, the system starts a question-and-answer session with the driver to make sure he is awake. The following are the questions asked:

- 1) What is your name?
- 2) What is your father's name?
- 3) Where are you from?

For all the above questions the driver has to answer correctly, even if one question is answered incorrectly then the system concludes the driver is drowsy and goes for next frequency physical alert (i.e., medium frequency). Fig. 17 shows the output of question-and-answer session.

```
pi@raspberrypi: ~/Desktop
File Edit Tabs Help
pi@raspberrypi:~/Desktop $ python main.py
Initializing the system
Driver is authorised..., Testing for Alcohol
Alcohol Not Detected !
Initializing drowsiness detection
Drowsiness detected

Initializing voice alert
Voice alert completed
Waiting for finger...
no finger print found

Initializing seat vibration alert
Seat Vibration alert completed
Waiting for finger...
no finger print found

Initializing physical alert
Initializing low Frequency physical alert
Low Frequency Robot arm alert completed

Initializing question answer session
What is your name ?
you said: Sowmya
What is your father's name?
you said: Sowmya
Where are you from?
you said: Bangalore

driver is drowsy
```

Fig. 17. Question and answer session.

After medium frequency physical alert, once again the question-and-answer session is started with the same questions. If the driver provides wrong answer again, the system starts high frequency physical alert. After these many alerts driver will definitely be alert. Fig. 18 shows the output of complete physical alert.

```
pi@raspberrypi: ~/Desktop
File Edit Tabs Help
Initializing physical alert
Initializing low frequency physical alert
Low Frequency Robot arm alert completed

Initializing question answer session
What is your name ?
you said: Sowmya
What is your father's name?
you said: Sowmya
Where are you from?
you said: Bangalore
driver is drowsy

Initializing medium frequency physical alert
Medium Frequency Robot arm alert completed

Initializing question answer session
What is your name ?
you said: Sowmya
What is your father's name?
you said: Keshava
Where are you from?
you said: Keshava
driver is drowsy

Initializing high frequency physical alert
High Frequency Robot arm alert completed

physical alert completed
```

Fig. 18. Complete output of physical alerting system.

We create a log file for each driver which contains his/her drowsiness and alerting details. Fig. 19 shows a part of log file.

```
Initializing the system 2022-07-29 15:59:13,135
Starting Driver Face Authentication 2022-07-29 15:59:18,148
Initializing the system 2022-07-29 16:03:33,145
Starting Driver Face Authentication 2022-07-29 16:03:38,151
Driver is authorised..., Driver name is Sowmya 2022-07-29 16:04:13,191
Testing for Alcohol 2022-07-29 16:04:13,192
Alcohol Not Detected ! 2022-07-29 16:04:13,192
Initializing drowsiness detection 2022-07-29 16:04:23,213
Drowsiness Detected 2022-07-29 16:04:44,988
Voice alert completed 2022-07-29 16:04:58,352
No fingerprint found, driver is still drowsy. Starting seat vibration 2022-07-29 16:05:03,293
Seat vibration alert completed 2022-07-29 16:05:18,310
No fingerprint found, driver is still drowsy. Starting physical alert 2022-07-29 16:05:23,273
Low frequency physical alert completed 2022-07-29 16:05:40,293
Sowmya is alert after low frequency physical alert 2022-07-29 16:06:03,132
physical alert completed 2022-07-29 16:06:08,140
Initializing drowsiness detection 2022-07-29 16:06:08,141
```

Fig. 19. Log file.

By analyzing this log file, we rate the drivers as Excellent, Good, Average and Bad. The driver who gets alert by voice alerting system is rated as Excellent and the driver who gets alert by seat vibration is rated as Good. The drivers who get alert after low frequency or medium frequency physical alert are rated Average. A Bad rating is given to the drivers who get alert after high frequency physical alert. This rating is then mailed to the concerned person. Fig. 20 shows the mail that has been sent to the concerned person containing the driver's name and his rating for the journey. This rating can help peoples to choose a good driver for a safe and enjoyable journey. Thus, our proposed system provides multistage end-to-end driver drowsiness detection and alerting functionality. This system helps us to avoid road accidents and save people's lives.



Fig. 20. Driver rating email.

IV. CONCLUSION AND FUTURE WORK

In the literature, various researchers have developed techniques to check drowsiness of the driver. However, in the past, there is no work done towards ensuring that driver is awake after the alerting system goes off. Hence, we suggest a driver drowsiness detection system which ensures that driver is alert through its multi-stage alerting system. The system starts with driver authentication by face recognition and alcohol testing. Then it detects drowsiness in driver, and if he/she is found drowsy, multi-stage alerting system is started. In the first stage a voice alert is given. After the voice alert, the driver has to give his/her fingerprint as a proof for not being drowsy. If the system fails to get a fingerprint it starts the vibration alert. Then again driver has to give his fingerprint, without which the system starts physical alert through robot arm which is performed with three different frequencies (i.e. Low, Medium and High) and three questions are asked after each frequency to make sure the driver is alert. We create a log file containing the driver's drowsiness details, by analyzing which we rate the driver and mail the rating to the concerned person. This rating can be used while hiring the driver for a safe and comfortable journey by avoiding road accidents. Thus, the system ensures that driver is alert and avoids road accidents. Future work can be to detect drowsiness when the driver is wearing sunglasses.

ACKNOWLEDGMENT

This research work is filed for patent at the Indian patent office with Patent Number – 202241076823.

REFERENCES

- [1] Alrehaili, Ahmed, Abdallah Namoun, and Ali Tufail. "A Comparative Analysis of Scalability Issues within Blockchain-based Solutions in the Internet of Things." *environments* 6: 7.
- [2] ACEA Driving mobility for Europe, <https://www.acea.auto/figure/world-motor-vehicle-production/>.
- [3] Traffic Safety Culture Index - AAA Foundation for Traffic, 2018, <https://aaafoundation.org/2018-traffic-safety-culture-index/>.
- [4] Ministry of Road Transport and Highway Government of India, <http://morth.nic.in/road-accidents-in-india>
- [5] Statista "Global No.1 business data platform" , <https://www.statista.com/statistics/200002/international-car-sales-since-1990/>.
- [6] Moujahid, Abdelmalik, FadiDornaika, Ignacio Arganda-Carreras, and Jorge Reta. "Efficient and compact face descriptor for driver drowsiness detection." *Expert Systems with Applications* 168 (2021): 114334.
- [7] Shen, Qi, Shengjie Zhao, Rongqing Zhang, and Bin Zhang. "Robust Two-Stream Multi-Features Network for Driver Drowsiness Detection." In *Proceedings of the 2020 2nd International Conference on Robotics, Intelligent Control and Artificial Intelligence*, pp. 271-277. 2020.
- [8] Deng, Wanghua, and Ruoxue Wu. "Real-time driver-drowsiness detection system using facial features." *Ieee Access* 7 (2019): 118727-118738.
- [9] Rostaminia, Soha, Addison Mayberry, Deepak Ganesan, Benjamin Marlin, and Jeremy Gummeson. "Ilid: low-power sensing of fatigue and drowsiness measures on a computational eyeglass." *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies* 1, no. 2 (2017): 1-26.
- [10] Kamarudin, Nora, NurAnidaJumadi, Ng Li Mun, C. K. Ng, Audrey HuangKahChing, W. M. H. W. Mahmud, MarliaMorsin, and Farhanahani Mahmud. "Implementation of Haar Cascade classifier and eye aspect ratio for driver drowsiness detection using raspberry pi." *Univ. J. Electr. Electron. Eng* 6, no. 5B (2019): 67-75.
- [11] Akrouf, Belhassen, and Walid Mahdi. "Yawning detection by the analysis of variational descriptor for monitoring driver drowsiness." In *2016 International Image Processing, Applications and Systems (IPAS)*, pp. 1-5. IEEE, 2016.
- [12] Fujiwara, Koichi, Erika Abe, Keisuke Kamata, Chikao Nakayama, Yoko Suzuki, Toshihiko Yamakawa, Toshihiro Hiraoka et al. "Heart rate variability-based driver drowsiness detection and its validation with EEG." *IEEE Transactions on Biomedical Engineering* 66, no. 6 (2018): 1769-1778.
- [13] Lee, Hyeonjeong, Jaewon Lee, and Miyoung Shin. "Using wearable ECG/PPG sensors for driver drowsiness detection based on distinguishable pattern of recurrence plots." *Electronics* 8, no. 2 (2019): 192.
- [14] Leng, Lee Boon, Lee Boon Giin, and Wan-Young Chung. "Wearable driver drowsiness detection system based on biomedical and motion sensors." In *2015 IEEE SENSORS*, pp. 1-4. IEEE, 2015.
- [15] Misbhauddin, Mohammed, AlReemAlMutlaq, AlaaAlmithn, Norah Alshukr, and Maryam Aleesa. "Real-time driver drowsiness detection using wearable technology." In *Proceedings of the 4th International Conference on Smart City Applications*, pp. 1-6. 2019.
- [16] Zhenhai, Gao, Le DinhDat, Hu Hongyu, Yu Ziwen, and Wu Xinyu. "Driver drowsiness detection based on time series analysis of steering wheel angular velocity." In *2017 9th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA)*, pp. 99-101. IEEE, 2017.
- [17] Li, Zuojin, Shengbo E. Li, Renjie Li, Bo Cheng, and Jinliang Shi. "Driver fatigue detection using approximate entropic of steering wheel angle from real driving data." *International Journal of Robotics and Automation* 32, no. 3 (2017).
- [18] Katyal, Yashika, SuhasAlur, and ShipraDwivedi. "Safe driving by detecting lane discipline and driver drowsiness." In *2014 IEEE International Conference on Advanced Communications, Control and Computing Technologies*, pp. 1008-1012. IEEE, 2014.
- [19] Choi, Minho, Gyogwon Koo, MinseokSeo, and Sang Woo Kim. "Wearable device-based system to monitor a driver's stress, fatigue, and drowsiness." *IEEE Transactions on Instrumentation and Measurement* 67, no. 3 (2017): 634-645.
- [20] Xing, Tianzhang, Qing Wang, Chase Q. Wu, Wei Xi, and Xiaojiang Chen. "dWatch: A Reliable and Low-Power Drowsiness Detection System for Drivers Based on Mobile Devices." *ACM Transactions on Sensor Networks (TOSN)* 16, no. 4 (2020): 1-22.
- [21] Mehta, Sukrit, SharadDadhich, SahilGumber, and ArpitaJadhav Bhatt. "Real-time driver drowsiness detection system using eye aspect ratio and eye closure ratio." In *Proceedings of international conference on sustainable computing in science, technology and management (SUSCOM)*, Amity University Rajasthan, Jaipur-India. 2019.
- [22] Yan, Jun-Juh, Hang-Hong Kuo, Ying-Fan Lin, and Teh-Lu Liao. "Real-time driver drowsiness detection system based on PERCLOS and grayscale image processing." In *2016 International Symposium on Computer, Consumer and Control (IS3C)*, pp. 243-246. IEEE, 2016.
- [23] Maior, CaioBezerraSouto, Márcio José das ChagasMoura, JoãoMateus Marques Santana, and Isis Didier Lins. "Real-time classification for autonomous drowsiness detection using eye aspect ratio." *Expert Systems with Applications* 158 (2020): 113505.
- [24] Kunding, Thomas, Phani Krishna Yalavarthi, Andreas Riener, Philipp Wintersberger, and Clemens Schartmüller. "Feasibility of smart

- wearables for driver drowsiness detection and its potential among different age groups." *International Journal of Pervasive Computing and Communications* (2020).
- [25] Chaabene, Siwar, BassemBouaziz, AmalBoudaya, Anita Hökelmann, AchrafAmmar, and LotfiChaari. "Convolutional neural network for drowsiness detection using EEG signals." *Sensors* 21, no. 5 (2021): 1734.
- [26] Viola, Paul, and Michael Jones. "Rapid object detection using a boosted cascade of simple features." In *Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001*, vol. 1, pp. I-I. Ieee, 2001.
- [27] Soukupova, Tereza, and Jan Cech. "Eye blink detection using facial landmarks." In *21st computer vision winter workshop, RimskeToplice, Slovenia*. 2016.
- [28] Hossain, Md Yousuf, and Fabian Parsia George. "IOT based real-time drowsy driving detection system for the prevention of road accidents." In *2018 International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS)*, vol. 3, pp. 190-195. IEEE, 2018.

Insights on Data Security Schemes and Authentication Adopted in Safeguarding Social Network

Nithya S¹, Rekha B²

Research Scholar, SJB Institute of Technology, Bangalore, India¹
SJB Institute of Technology, Bangalore, India²

Abstract—With the increased social network usage, there is a rising concern about potential security and privacy risks related to digital information data. Although there have been numerous studies in this area, a summary is necessary to understand the effectiveness of existing security approaches. The ultimate goal is to provide valuable insights into the effectiveness of existing security schemes in the social network ecosystem. Therefore, the proposed paper discusses the existing research that has been done on authentication and data security measures, including methodologies, issues, benefits, and drawbacks. The inquiry further contributes to highlighting existing research trends and identifying the gap. The paper concludes by stating its learning results that help to open possible insights into the effectiveness of existing security schemes in the social network. Furthermore, blockchain is witnessed with increased interest in distributed security over large data. The paper's outcome states that blockchain-based authentication systems possess better scope if subjected to amending their inherent shortcomings. The findings of this paper emphasize the importance of continuous innovation in data security to ensure the safety and privacy of user data in an ever-evolving digital landscape. This paper offers a foundational aspect for future research toward developing more secure, privacy-preserving solutions for social network users.

Keywords—Social network; security threat; authentication; blockchain

I. INTRODUCTION

With an ongoing demand and trend of sharing information, social network has been creating breakthrough innovations in this perspective [1]. It is not only used for sharing information but also extensively used for constructing business opportunities too [2]. A social network can be more generalized, while its implication can be multi-formed based on usage. It can be represented or rather classified into a consumer review network (e.g., TripAdvisor, Yelp, etc.), a network with content curation (e.g., Flipboard, Pinterest, etc.), a forum for discussion (e.g., Quora, Reddit, etc.), the network for media sharing (YouTube, Snapchat, Instagram, etc.), for the social purpose (LinkedIn, Twitter, Facebook, etc.) [3]. This classification of social network usage will eventually state that volumes of information are involved in each application that is stored, shared, and accessed seamlessly and concurrently [4]. Although most social network offers a simplified design that the account owner can customize, a security breach cannot be fully stopped [5]. As social networks are formed by highly

interconnected networks over the internet, recognizing and comprehending harmful behavior and differentiating them from regular behavior is quite challenging [6]. The attacker intrudes on the network by hiding their identity as a regular node whose malicious intention cannot be guessed initially. Using various alternative means and tools, it is always possible for an attacker to gain access to the user's private information in the social network [7]. Various studies have reported different variants of intrusive activities in the social network [8]. There are also accomplished studies and ongoing research toward strengthening social network security [9]. However, due to their publicly exposed contents, no tool or program has yet been identified or benchmarked as offering full-proof security in social networks. At present, the blockchain-based approach over large data, along with integrated encryption, is highly in demand and is increasingly adopted for securing the contents to avoid falling into the hands of an attacker [10]. Encryption is another preferred technique for securing the contents to offer more data integrity and privacy [11].

However, blockchain and encryption have potential flaws irrespective of their known beneficial features for security. One of the significant problems in blockchain implementation is achieving performance scalability [12]. It also suffers from limiting the number of transactions a network with the block can process. A different arena of problems also exists for an encryption-based solution. Usually, the strong encryption algorithm is characterized by a higher size of keys and is iterative in its operation [13]. Aside from that, the encryption technique is also reliant on a precise set of resources to be fully executed. These shortcomings eventually offer a possible hindrance to a robust authentication method in social networks.

Therefore, the proposed paper discusses existing security approaches in social networks, specifically emphasizing data security and authentication. The contributions of the paper are i) reviewing existing data security approaches exercised in social networks, ii) highlighting a few prominent works using encryption, privacy preservation, learning approaches, and other miscellaneous methodologies towards data security in social network, iii) identifying advantages and shortcomings of various methodologies, iv) highlighting research trends towards publications and v) exploring prominent research gap. The paper is organized as follows: Section II discusses security insights into social networks and discusses existing data security approaches in Section III. Section IV discusses the

current contribution of authentication. At the same time, research trends are highlighted in Section V. Discussion of the research gap is presented in Section VI. Section VII presents the findings and discussions. Section VIII concludes with highlights of its learning outcomes of the proposed review work in social network security.

II. SECURITY INSIGHTS IN SOCIAL NETWORK

In the current era of social networks, it is noted that most social media applications are publicly disclosed, where it is feasible for the attacker to aggregate the data stealthily without letting the user know [7]. The next level of attackers is more interested in illegitimately gaining access to genuine users' accounts. However, the degree of threat in the social network depends on their planned motive. Attackers deploy various alternative techniques to understand the user in social media, thereby initiating malicious activities. Various forms of intrusive activities performed by attackers are briefed as follows:

- **Data Breach:** Using multiple alternative approaches, an attacker can steal a user's credentials and gain illegitimate access to their account. This led to a potential breach of users' private information at the hand of the attacker.
- **Malware propagation:** An attacker can easily divert users to visit their sites using various counterfeited portals. Once the user visits such sites, they are prompted to do simple activities which lead to the activation of malicious codes, and thereby malicious malware starts propagating.
- **Data Theft:** If an attacker can access any business account on the social network, they can also exfiltrate sensitive information channeled to their account. Hence, both data and account eventually get compromised.
- **Impersonation of the brand:** An attacker can construct a counterfeited account of a specific brand where prospective customers can be tricked in various ways, either by maliciously draining their finances or stealing confidential data.
- **Phishing:** Such an attack calls for forwarding a malicious link in messages to the user in various ways. When the user clicks, it directs them to various security threats, including account hijacking. It could also lead to latent stealing of personal information stored in the device.
- **Social Engineering:** In this form of threat, the attacker convinces or tricks the user of their genuine and trustworthy profile. The users are prompted to forward either a financial asset or high-profile information at their wish that is maliciously forwarded to the attacker's account.

All of the above-mentioned invasive actions are common, but a standard classification also explains their variations. Apart from the above-mentioned activities, other forms of attacks in social network are cyberbullying too which are used to intimidate and harass the user by posting objectional

comments and spreading counterfeited rumors. Fig.1 highlights further classification in the form of traditional adversaries, modern adversaries, and targeted adversaries. All these adversaries work with different techniques; hence, there is no fair possibility of developing a common solution to stop all these adversaries. Work is being carried out towards modeling lethal threats in the social network [8], yet it has a restricted security feature.

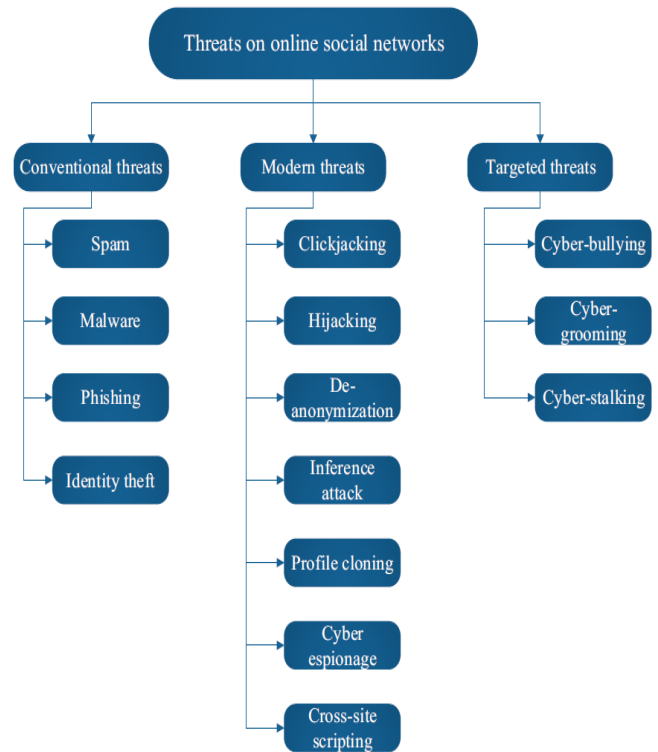


Fig. 1. Classification of adversaries in social networks [14].

Nevertheless, much research is being done to combat the risks spread through social networks. [8], there is also the evolution of smart solutions: cyberstalking, clickjacking, cyber grooming, cyberbullying detection, and phishing detection. From the commercial application viewpoint, the normal recommendation towards securing a social network is to adopt a strong credential, restrict location sharing, install threat detection software, understand and be aware of a third-party application, be vigilant towards sharing content, and review every new friend request [14]. However, from the research perspective, it is suggested to use a strong security protocol that uses potential data encryption, understanding the network connectivity, and vigilant towards usage of underlying threads in the device [15].

Unfortunately, the degree of attacks on social networks is consistently rising despite knowing the facts to be considered for security. It will eventually mean a potential tradeoff between actual security demand and existing effectiveness in security schemes. One of the most primitive intrusion points is via weak authentication in social networking applications [16][17]. Unfortunately, denser and more aggressive authentication protocol usage will also affect the actual motive of social networks, which is towards large-scale data sharing in

a multi-dimensional manner. Further social networking applications migrating to a cloud-based ecosystem offer some application-feature-based advantages but at the cost of security. Moreover, as one social network application is also connected to different applications, the level of threat propagation is quite excessive. Therefore, there is a potential demand to investigate better authentication with better data security. The next section discusses some of the contributions of existing schemes toward securing social networks.

III. DATA SECURITY APPROACHES IN SOCIAL NETWORK

The prior section noted various security threats in social network applications. There is also the evolution of various research techniques to circumvent such forms of threats due to different forms of threats. Due to a large chain of sophisticated networks, offering robust data security in the social network is challenging. Therefore, this section outlines the contribution of some of the identified essential techniques to offer data security.

A. Encryption-based Approaches

This approach is designed and implemented to encrypt the data propagated in social media. From the viewpoint of data, it is found that multimedia data is in circulation and shared in social networks along with the text. The recent work carried out by Ali, and Ali [18] has implemented an encryption strategy toward securing a color image by adopting a chaotic map. A non-linearity element is generated by amending the pixel values, which further results in a random sequence by diffusion. This operation is followed by mixing the encrypted image to generate a consistent distribution of randomness further. Huang et al. also investigated data encryption [19], where a re-encryption policy is applied based on the identity-based sharing of confidential data.

One of the significant advantages of this policy is that it only permits a re-encryption process for matched encrypted data, thereby offering better access control. The work by Qiu et al. [20] discussed a selective encryption system based on coding an embedded block. The investigation also uses optimized truncation to secure a selected part of bitstreams. The work by Zuo et al. [21] used homomorphic encryption to secure the graph operation in social networks associated with untrusted clouds. This investigation model offers effective data security and retains better privacy preservation. The beneficial point of adopting an encryption-based approach is its explicit resistivity towards a specific attack; however, its scope of applicability is limited to specific intruders and involves quite a sophisticated key computation that demands many computational resources.

B. Privacy Preservation-based Approaches

This type of security approach mainly concerns safeguarding all the necessary information that holds privacy details of the data and the user in the social network. A unique investigation formulated by Barni et al. [22] discussed security issues related to adopting biometrics in social media, specifically using iris. As a solution, the author has used a generative adversarial network to generate images with

eliminated biometric information in social networks. Another work by Chen et al. [23] used an integrated method of searchable encryption and blockchain to offer privacy. The presented mechanism is highly decentralized and uses public-key encryption to secure communication in a vehicular network. Li et al. [24] have provided a de-anonymization strategy for the heterogeneous social network in a different piece of work—the model claims to improve the detection system by using user profile and network structure information. Li et al. also presents a similar work form [25]. The investigation model assesses privacy factors associated with the behavioral attributes in social networks based on structural similarity.

The work carried out by Qu et al. [26] has addressed the solution to the gap between the utility of data and privacy customization. The investigation model customizes the level of protecting privacy by using the shortest distance between two nodes in the social network. The model has also adopted an improved version of the Laplacian method for noise modeling that is finally subjected to decoupling to prove its resistivity against collusion attack. Another unique work is carried out by Xu et al. [27] investigating the context behind the communication of anti-social elements using social networks. This investigation has presented a mechanism to retrieve data associated with privacy preservation which further performs a query on suspect communication. The method also securely implements classification and regression trees to resist privacy leakage. The work by Xu et al. [28] presented a selection technique of an optimal trajectory and adopted a heuristic method. The implemented technique also performs clustering operations for facilitating multiple forms of trajectories that finally assist in discovering the community. Assurance of privacy preservation was also discussed by Yin et al. [29] using deep learning approaches.

A hybrid scheme has been deployed using a Bayesian network, a federated learning approach, and a sparse differential gradient. This work aims to optimize the functional encryption operation to secure data sharing among social network multi-parties. A similar investigation was also carried out by Zhang et al. [30], where the data privacy factor is assessed by evaluating the sentence correlation using a convolution neural network and the firefly algorithm. This work aims to ensure secure privacy preservation for social users using large communication scenarios like Internet-of-Things (IoT). The investigation by Zhu et al. [31] presented a computational model for the propagation of privacy information for more in-depth identification of the malicious nature of social nodes. The benefit of adopting the privacy preservation model in securing social network communication is that it offers various techniques while offering higher coverage of security problems in the social network. However, the limiting factor associated with privacy preservation schemes of existing studies is related to their non-applicability in a different test environment.

C. Learning-based Approaches

Learning-based approaches are a part of artificial intelligence that can identify and solve complex problems facilitated by predictive outcomes. The algorithm written for

learning-based approaches offers an extensive capability of execution with a higher adaptation rate of intelligence. Adopting a learning approach facilitates evaluating the problems in social networks and further evolves with more accurate solutions.

The investigation carried out by Abbasi et al. [32] essentially focuses on implementing a learning approach toward identifying any drift factor associated with the concept of massive ranges of social data. The author uses ensemble learning for this purpose, where the idea is to perform an optimal classification of social data to identify the concept drift. The work carried out by Chen et al. [33] presented a model for evaluating multiple trusts for users of social networks considering discrete criteria as well as features, e.g., link, feedback, behavior, and profile. The investigation model implements multiple machine-learning approaches to evaluate its performance.

Another learning-based approach is implemented by Gao et al. [34] to confirm the presence of a Sybil attacker. The technique implements a convolution neural network for extracting low-end and high-end features by Long Short-Term Memory. Discussion towards the applicability of the deep learning approach to the social network security aspect has been carried out by Garg et al. [35]. The investigation has emphasized using Software Defined Networks (SDN) to improve social network security systems. The first module of the investigation uses a support vector machine with gradient descent for anomaly detection.

In contrast, the second module of the investigation implements SDN to ensure a better delivery system. Mei et al. [36] have presented a solution to inference attacks in social media by amending the existing deep learning approach. The work by Sansonetti et al. [37] has presented a technique for identifying the propagation of counterfeited news in social media. The learning-based technique is applied for this analysis in both online and offline modes.

Another learning-based approach is discussed by Shen et al. [38], where an encryption system and blockchain have been introduced to secure the classification operation. The presented investigation uses a support vector machine for performing training operations, while the scheme also assists in resisting any collusion of data involved in it. The adoption of an adversarial learning scheme has been presented by Zhang et al. [39], emphasizing privacy protection. The research helps defend against attacks that try to establish a connection between two nodes in graph embedding.

Consequently, the scheme minimizes the accuracy of the prediction of an attacker. The beneficial factor for adopting a learning-based approach is its effective modeling toward predictive computation considering complex security loopholes. However, a closer look into existing approaches exhibits the prevalence of using static use cases of adversaries, implying its applicability only to specific case studies. In

addition, the computational complexity linked to a higher number of repeats is still unaddressed by the current scheme.

D. Miscellaneous Approaches

Apart from the conventional security practices in social networks seen in prior sections, various off-beat mechanisms are introduced to address similar security problems. From the work presented by Song et al. [40], game theory has been increasingly adopted in modeling social networks. This article claims that game theory is used in behavioral analysis, community identification, and information dissemination to improve the security of social networks by improving access controls and formulating privacy policies. The adoption of game theory is seen in the work of Du et al. [41], where privacy protection is emphasized using the evolutionary game concept in the social network. The idea of this work is to investigate the selection strategy of a user toward privacy protection. The outcome of this investigation is analyzed concerning computational performance cost. Although blockchain is also reported to be used in securing social network communication and services, some studies enhance blockchain's usage towards more security. The work carried out by Fan et al. [42] has used blockchain to offer better non-repudiation in security services as well as to perform a better formulation of access policy using a secret sharing scheme.

Further information used for constructing access policy is hidden for enhanced security. The work carried out by Gao et al. [43] has developed a game-based framework to investigate the social reputation and its impact on controlling data access. A similar line of work is also carried out by Wang et al. [44] towards identifying and resisting counterfeited messages. Huang et al. also adopted a game-based framework [45] to construct an economic model. The core idea of this part of the implementation is to identify any alteration in the network concerning user income. The implementation also assists in developing a model for price decisions under a specific network condition.

The framework developed by Kong et al. [46] has constructed a security framework for strengthening the reputation system over the cloud ecosystem considering the use case of a large-scale healthcare system. The investigation has implemented a convolution neural network that categorizes textual data followed by applying a dynamic game model for constructing a strategy toward incentive allocation.

Su and Xu [47] have worked on allocating resources toward secure communication in social media using the game-based model. The primary agenda of this work is to carry out secure group-based communication for social networks followed by resource gain improvement. Sun et al. [48] have presented a key-based encryption strategy for securing social network data. Using a secure data-sharing scheme, the idea is to identify and protect against intrusion. Table I summarizes all the above data security approaches in the social network.

TABLE I. SUMMARY OF DATA SECURITY APPROACHES

Authors	Problems Addressed	Methodology Adopted	Advantage
Ali & Ali [18]	Image encryption	Chaotic Map	Efficient non-linearity in encryption
Huang et al.[19]	Secure Data Sharing and encryption	Identity-based Re-encryption	Better access control
Qiu et al. [20]	Data security	Selective encryption	Can secure text, image, and video file
Zuo et al. [21]	Data security	Homomorphic encryption	Ensure better data ownership
Barni et al. [22]	Privacy issues in biometrics	Generative adversarial network	Significantly control
Chen et al. [23]	Privacy in vehicular network	Searchable encryption	Satisfactory response time
Li et al. [24]	De-anonymization of user	Modeling using user profile and network structure	Maximize accuracy of detection
Li et al. [25]	Privacy measurement	Structural similarity model of behavior intimacy	Effectively reduces privacy leakage
Qu et al. [26]	Privacy preservation	Analytical model using the Laplacian method	Customization privacy
Xu et al. [27]	Sensing criminal communication in social network	Retrieval of data for privacy, classification & regression tree	Significantly less overhead
Xu et al. [28]	Discovering latent trajectory	Community discovery model using clustering	Offers higher accuracy
Yin et al. [29]	Data sharing in multi-party	Sparse differential gradient, functional encryption, federated learning	Enhance transmission efficiency
Zhang et al. [30]	Privacy preservation for social users in IoT	Convolution Neural Network, Firefly algorithm,	Optimize more usage of data in social network
Zhu et al. [31]	Privacy propagation	Empirical approach	Applicable for dynamic social network
Abbasi et al. [32]	Identification of concept drift	Ensemble learning	Satisfactory accuracy
Chen et al. [33]	Trust evaluation in social network	Multiple machine learning for feature selection	Better accuracy performance
Gao et al. [34]	Detection of Sybil attack	Long Short-Term Memory	Higher accuracy in detection
Garg et al. [35]	Anomaly detection	Support Vector Machine (Gradient Descent), SDN	Ensure secure flow routing
Mei et al. [36]	Inference attack in social network	Revised convolution neural network	Satisfactory accuracy performance
Sansonetti et al. [37]	Identification of counterfeited news in social media	Learning-based approach	Overall satisfactory accuracy performance
Zhang et al. [39]	Privacy preservation	Adversarial learning	Higher preservation performance
Song et al. [40]	Review of game theory	Reviewing existing approaches	Higher applicability
Du et al. [41]	Privacy protection	Evolutionary game	Offer consistency in privacy protection with increased network size.
Fan et al. [42]	Secure data sharing	Blockchain, secret sharing	Resistive against collusion attack
Gao et al. [43]	Impact on social reputation	Game-based framework	Improve the rate of cloud storage
Wang et al. [44]	Counterfeited Message	Game-based model	Increased probability of detection of malicious message
Huang et al. [45]	Issues in price decisions in cyber-physical system	Game-based model	Suitable for price adjustment
Kong et al. [46]	Data privacy for large network	Convolution neural network	Improve model reliability
Su and Xu [47]	Secure resource allocation	Coalition game model	Effective resource efficiency

IV. AUTHENTICATION APPROACHES IN SOCIAL NETWORK

Different authentication mechanisms have evolved with the increasing features of social network applications. However, such authentication mechanism differs strongly between commercially used applications and those reported in scientific papers. The commercially available applications use normal user identity-based credential mechanism that adopts user name and password. The authentication mechanism could be carried out in single execution and sometimes in multiple executions. However, almost all commercially available social network applications use a static form of security token to authenticate the user, where traces of authentications are stored

in mobile devices within its cache memory system. This is a highly vulnerable state for the user where their services and data are exposed to a potential threat. On the other hand, various recently developed protocols have been carried out to strengthen and evolve of authentication scheme.

The work by Alvarez et al. [49] has discussed different authentication system mechanisms for strengthening the associated privacy factors. The investigation suggests the usage of biometrics for this purpose. The majority of the social network application is investigated using a graphical concept. One such work by Jin et al. [50] emphasizes implementing a stochastic approach to authenticate such graphs. The technique

uses a supervised learning mechanism to identify such malicious activities in the social network. An authentication model discussed by Megouache et al. [51] has presented a unique scheme for privacy preservation in an environment with multi-clouds. The investigation uses an encryption approach toward data integrity as well as authentication. The work carried out by Park et al. [52] have presented distributed scheme of authentication considering multi-factor authentication based on trust score. Ruan et al. [53] have presented a work towards location privacy where replicated information is used to secure the user's privacy. At the same time, it controls all sorts of inference of activity track of use from the location server to keep it more secure. The work by Sinha et al. [54] used elliptical curve encryption and symmetric encryption to resist replay attacks and cryptanalysis attacks in the social network. The investigation model has implemented a key exchange mechanism to carry out authentication. Soni et al. [55] have presented a security scheme using fuzzy c-means algorithm. In contrast, it uses a series of encryption techniques (e.g., Rivest Shamir Algorithm (RSA), Advanced Encryption

Standard (AES), and Rivest Cipher 6 (RC6)) to cipher the data further. A recent investigation has also witnessed the adoption of blockchain to secure the trust factor in next-generation networks based on user behavior.

Tu et al. [56] carried out an investigation where a novel trust control modeling is based on user behavior. A unique authentication model is presented by Xu et al.[57] over the storage framework associated with the social network using blockchain. The authentication is provided by incorporating secure access control developed by integrating the Clark-Wilson model and blockchain technology. The adoption of homomorphic encryption was carried out by Zuo et al. [58], where a sub-graph matching mechanism was introduced to carry out authentication. The cloud carries out the query processing of the subgraph without any dependency on the secure and sensitive information of the user. The investigation model of authentication is claimed to offer data integrity too. Table II highlights the summarization of the existing authentication schemes.

TABLE II. SUMMARY OF AUTHENTICATION APPROACHES

Authors	Problems Addressed	Methodology Adopted	Advantage
Alvarez et al. [49]	Sensor-based authentication	Review work	Elaborated discussion of existing methods
Jin et al. [50]	Authentication of graph	Supervised learning, stochastic	84.4% of accuracy rate
Megouache et al. [51]	Authentication & Integrity	Encryption-based model	Stabilized system
Park et al.[52]	Cyber-security threat	Distributed authentication model	Reduced latency
Ruan et al. [53]	Privacy protection	Replicated identity construction	Lower communication and computation cost
Sinha et al. [54]	Replay attack on social network	Key exchange, Elliptical curve cryptography, symmetric encryption	Less processing time
Soni et al. [55]	Data security in the cloud	Fuzzy c-Means clustering, RSA, AES, RC6	Simplified technique
Tu et al. [56]	attacks in next-generation network	Blockchain-based trust model	Minimized network threats
Xu et al.[57]	Access Control on storage	Blockchain, Clark-Wilson Model	Offers data integrity
Zuo et al. [58]	Privacy preservation	Homomorphic encryption	Offers data privacy and integrity

V. RESEARCH TRENDS

To understand the existing research trend, data were collected from well-known publications, e.g., IEEE Xplore, Springer, ACM, Wiley, and Elsevier, ranging between the publication year of 2012-2022. It was noted that approximately 6538 conference papers and 1395 journals are being published towards discussing and evolving out of security models associated with social networks in IEEE alone [59]. After a complete evaluation of overall methodologies, it has been noted that there are different variants of techniques towards incorporating security, authentication-based approaches, blockchain technology, investigation towards data integrity, adoption of game theoretical framework, usage of different types of machine learning models, studies towards privacy preservation, and encryption-based approaches. All the studies mentioned above methodology are witnessed to address different forms of security problems arising in social media. From Fig. 2 it can be noted that more studies have been carried out to emphasize privacy preservation in social networks. The studies adopting blockchain-based and encryption-based mechanisms are also increasingly evolving. However, many of these approaches are very less than privacy preservation

approaches. Games and learning-based programs are still in the early stages of development.

Therefore, the inference obtained from the simplified analysis of the research trend of publication is as follows:

- The amount of research on protecting privacy on social networks is relatively higher. The scope of this outcome is that existing privacy preservation is carried out considering a set of adversaries that it can successfully resist [14],[16],[21]-[31],[39], [49]. Therefore, such a privacy preservation scheme is robust to act against specific attacks; however, the prime shortcoming is that there is a need to address dynamic attackers in the social network that are few to be reported. Similar problems are also applicable to data integrity schemes.
- From an implementation standpoint, research on authentication and encryption-based systems is closely related [11], [13], [18]-[20], [23]. There is an increasing trend toward using a symmetric key, elliptical curve cryptography, RSA, and many other approaches in public key encryption. Such encryption approaches introduce sophisticated techniques of private key

computation without considering encrypting default public keys. Apart from this, most authentication-based approaches using key-based mechanisms do not consider the device complexity used in social networks. Each social network application has a unique performance signature on different devices, which is not evaluated in existing research.

- Machine learning [16], [29],[32]-[35], [37], [39] and game theoretical models [40]-[47] are slowly gaining pace toward social network security. However, such approaches are yet to be addressed from their practical world implementation owing to their larger dependencies on additional information and data. The complex problems associated with them are yet to be researched.

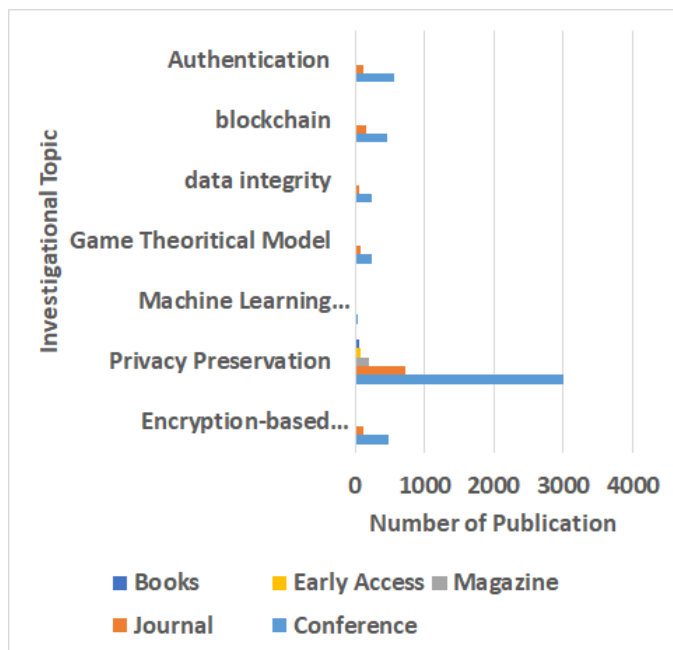


Fig. 2. Research trends of publication (2012-2022).

VI. GAP ANALYSIS

Undoubtedly, current research on secure social network applications has significantly contributed. With different classes of algorithms, various security solutions are available to resist potential threats. After reviewing various classes of existing schemes for securing communication in social networks, the following shortcomings have been witnessed in the form of a research gap:

- *Attack Specific Investigation:* There are variously reported adversaries present in the social network, e.g., cross-site scripting, clickjacking, SQL injection, whaling attack, malware propagation, spamming attack, etc., apart from various other conventional reported attacks. Different variants of methodologies presented to date have used only specific forms of attacks, and hence that solution is rendered inapplicable when exposed to a different set of attacks. The problem becomes more significant when the attacker introduces

malicious activities dynamically. Hence, there is a need to develop an adversarial model first which bears the maximum characteristics of existing attackers to prove the resiliency of security protocols.

- *Unbalance computational efficiency:* The primary stage of developing the social network is constructing a higher linked and inter-connected topology of the user (or nodes). As social networking application is run over user handheld device, hence, is predicted to support such resource constraint device. Applying a strong encryption algorithm will only negatively affect the system's performance, even if it can resist certain threats. Hence, introducing a simplified, lightweight encryption scheme can only ensure computational efficiency. Unfortunately, no reported work has presented evidence of consistency in the computational performance of security when confronted with large and heavy traffic on the social network.
- *Usage of Conventional Authentication Scheme:* From the viewpoint of practical deployment, existing social network applications still use static passwords. Adopting a key-management scheme induces maximized resource dependencies while using encryption sets introduces computational burden over the long run. Existing multi-factor schemes are executed without protecting the location where algorithms are executed and thereby introduce a tradeoff between security and computational demands.
- *Lack of integrated schemes:* As a social network is a large chain of nodes formed in a complex way, data leakage is also possible. Hence, the authentication mechanism should be studied alongside data privacy and integrity to offer maximum protection. Such integrated schemes are not yet found been introduced or benchmarked.

VII. FINDINGS AND DISCUSSION

There are mainly two core classes of research toward the direction of security in the existing system, i.e., data security and authentication approaches. More studies have been carried out on data security approaches compared to authentication approaches. The core approaches discussed in this paper are encryption-based, privacy preservation-based, machine learning-based, and game-based modeling to strengthen social network security.

Reviewing existing research trends showcase more concentrated work towards privacy preservation while very little research is towards learning and data integrity-based methods. It is, therefore, evident that current solutions do not provide comprehensive security services and can only provide data integrity, data privacy, or non-repudiation. A social network is a complex network exposed to multi-dimensional threats. Hence, it's predicted to offer maximum security services, which is not found reported in existing security schemes.

It has been also identified that the blockchain is one of the evolving solutions for security in the social network. However,

there are various pitfalls seen in implementation effectiveness. i) blockchain-based scheme demands a potential form of the secured and interconnected topology of nodes integrated with a service provider with resources to operate. Existing blockchain schemes don't address this fact, ii) adoption of blockchain approach is associated with complexity, especially if it's a large and heterogeneous network of complex user behavior; hence computational burden is inevitable in the blockchain. Hence, there is potential for improving these blockchain problems to harness their security strength. None of the existing solutions has discussed the essential type of content in the social network. Textual content is more extensively used than other forms, e.g., images, GIFs, video, and audio. Encryption algorithms are eventually a better alternative; however, achieving encryption performance with computation and service relaying performance in social networks is yet to be seen. Therefore, there is a need for an investigation that would emphasize securing the text contents from dynamic attackers in the social network.

VIII. CONCLUSION AND FUTURE WORK

This paper has presented a compact discussion of securing communication in the social network. It is noted that various classes of methodologies are being adopted towards improving the security aspect with claimed benefits; however, the paper has identified a shortcoming associated with it. Hence, based on the complete review, it can be said that multifold findings state that there is still a large open scope for improving the security aspect of the social network. Exploring the existing research literature followed by methodologies, challenges, benefits, and drawbacks has highlighted the need for continuous innovation and improvement to protect user data effectively. The paper also emphasizes the growing interest in blockchain technology as a promising distributed security and authentication solution. Ultimately, the findings of this paper underscore the importance of a collaborative and multidisciplinary approach to data security and authentication in social networks. In future work, the scope of this paper will be extended toward modeling computationally efficient and robust security approaches to address dynamic security and privacy issue in social networking applications.

REFERENCES

- [1] M. Burcher, *Social Network Analysis, and Law Enforcement Applications for Intelligence Analysis*, Springer International Publishing, ISBN: 9783030477714, 3030477711, 2020.
- [2] S. Alavi, V. Ahuja, *Managing Social Media Practices in the Digital Economy*, IGI Global, ISBN: 9781799821878, 1799821870, 2019.
- [3] D. Zahay, M. L. Roberts, J. Parker, D. I. Barker, M. Barker, *Social Media Marketing: A Strategic Approach*, Cengage Learning, ISBN: 9780357516287, 0357516281, 2022.
- [4] A. E. Hassani, A. Abraham, M. Panda, *Big Data Analytics-A Social Network Approach*, Taylor & Francis Group, ISBN: 9780367780777, 0367780771, 2021.
- [5] B. B. Gupta, S. R. Sahoo, *Online Social Networks Security-Principles, Algorithm, Applications, and Perspectives*, CRC Press, ISBN: 9781000347111, 1000347117, 2021.
- [6] R. Luttrell, *Social Media-How to Engage, Share, and Connect*, Rowman & Littlefield Publishers, ISBN: 9781538154434, 1538154439, 2021.
- [7] B.D. Deebak, Fadi Al-Turjman, *Security in IoT Social Networks*, Elsevier Science, ISBN: 9780128216033, 0128216034, 2020.

- [8] E. Etuh, F. S. Bakpo, E. Agozie H, "Social Media Networks Attacks and their Preventive Mechanisms: A Review," *ArXiv, Social and Information Networks*, 2022. DOI: <https://doi.org/10.48550/arXiv.2201.03330>.
- [9] R. Abid, M. Rizwan, P. Vesely, A. Basharat, U. Tariq, and A. R. Javed, "Social Networking Security during COVID-19: A Systematic Literature Review", *Hindawi-Wireless Communications and Mobile Computing*, Article ID 2975033, 2022, DOI: <https://doi.org/10.1155/2022/2975033>.
- [10] S. X. Wu, Z. Wu, S. Chen, G. Li and S. Zhang, "Community Detection in Blockchain Social Networks," in *Journal of Communications and Information Networks*, vol. 6, no. 1, pp. 59-71, March 2021, doi: 10.23919/JCIN.2021.9387705.
- [11] H. Qiu, M. Qiu, M. Liu, and Z. Ming, "Lightweight Selective Encryption for Social Data Protection Based on EBCOT Coding," in *IEEE Transactions on Computational Social Systems*, vol. 7, no. 1, pp. 205-214, Feb. 2020, doi: 10.1109/TCSS.2019.2952553.
- [12] A. Hafid, A. S. Hafid and M. Samih, "Scaling Blockchains: A Comprehensive Survey," in *IEEE Access*, vol. 8, pp. 125244-125262, 2020, doi: 10.1109/ACCESS.2020.3007251.
- [13] C. -I. Fan, Y. -F. Tseng, J. -J. Huang, S. -F. Chen and H. Kikuchi, "Multireceiver Predicate Encryption for Online Social Networks," in *IEEE Transactions on Signal and Information Processing over Networks*, vol. 3, no. 2, pp. 388-403, June 2017, doi: 10.1109/TSIPN.2017.2697580.
- [14] A.K. Jain, S.R. Sahoo, & J. Kaubiyal, "Online social networks security and privacy: comprehensive review and analysis," *SpringerOpen-Complex & Intelligent Systems* volume 7, pp.2157-2177, 2021. DOI:<https://doi.org/10.1007/s40747-021-00409-7>.
- [15] S. Szymoniak, "Security protocols analysis including various time parameters," *AIMS-Press-Mathematical Biosciences and Engineering*, Volume 18, Issue 2, pp.1136-1153, 2021. Doi: 10.3934/mbe.2021061.
- [16] T. Guo, F. Li, "Machine Learning-based Online Social Network Privacy Preservation," *ACM-Proceedings of the 2022 ACM on Asia Conference on Computer and Communications Security* May 2022 Pages 467-478 <https://doi.org/10.1145/3488932.3517405>.
- [17] X. Yu, R. Ge, F. Li, "Research on Blockchain-Based Identity Authentication Scheme in Social Networks," *ACM-Machine Learning for Cyber Security: Third International Conference, ML4CS 2020, Guangzhou, China, October 8-10, 2020, Proceedings, Part I* Oct 2020 Pages 558-565 https://doi.org/10.1007/978-3-030-62223-7_49.
- [18] T.S. Ali, R. Ali, "A novel color image encryption scheme based on a new dynamic compound chaotic map and S-box," *Springer-Multimedia Tools and Applications*, vol.81, pp.20585-20609, 2022, DOI: <https://doi.org/10.1007/s11042-022-12268-6>.
- [19] Q. Huang, Y. Yang, J. Fu, "PRECISE: Identity-based private data sharing with conditional proxy re-encryption in online social networks," *Elsevier-Future Generation Computer Systems*, vol.86, pp.1523-1533, 2018 DOI: <http://dx.doi.org/10.1016/j.future.2017.05.026>.
- [20] H. Qiu, M. Qiu, M. Liu, and Z. Ming, "Lightweight Selective Encryption for Social Data Protection Based on EBCOT Coding," in *IEEE Transactions on Computational Social Systems*, vol. 7, no. 1, pp. 205-214, Feb. 2020, doi: 10.1109/TCSS.2019.2952553.
- [21] X. Zuo, L. Li, S. Luo, H. Peng, Y. Yang, and L. Gong, "Privacy-Preserving Verifiable Graph Intersection Scheme With Cryptographic Accumulators in Social Networks," in *IEEE Internet of Things Journal*, vol. 8, no. 6, pp. 4590-4603, 15 March 15, 2021, doi: 10.1109/IJOT.2020.3028417.
- [22] M. Barni, R. D. Labati, A. Genovese, V. Piuri and F. Scotti, "Iris Deidentification With High Visual Realism for Privacy Protection on Websites and Social Networks," in *IEEE Access*, vol. 9, pp. 131995-132010, 2021, doi: 10.1109/ACCESS.2021.3114588.
- [23] B. Chen, L. Wu, H. Wang, L. Zhou, and D. He, "A Blockchain-Based Searchable Public-Key Encryption With Forward and Backward Privacy for Cloud-Assisted Vehicular Social Networks," in *IEEE Transactions on Vehicular Technology*, vol. 69, no. 6, pp. 5813-5825, June 2020, doi: 10.1109/TVT.2019.2959383.
- [24] H. Li, Q. Chen, H. Zhu, D. Ma, H. Wen, and X. S. Shen, "Privacy Leakage via De-Anonymization and Aggregation in Heterogeneous Social Networks," in *IEEE Transactions on Dependable and Secure*

- Computing, vol. 17, no. 2, pp. 350-362, 1 March-April 2020, doi: 10.1109/TDSC.2017.2754249.
- [25] X. Li, Y. Xin, C. Zhao, Y. Yang, S. Luo, and Y. Chen, "Using User Behavior to Measure Privacy on Online Social Networks," in IEEE Access, vol. 8, pp. 108387-108401, 2020, doi: 10.1109/ACCESS.2020.3000780.
- [26] Y. Qu, S. Yu, W. Zhou, S. Chen, and J. Wu, "Customizable Reliable Privacy-Preserving Data Sharing in Cyber-Physical Social Networks," in IEEE Transactions on Network Science and Engineering, vol. 8, no. 1, pp. 269-281, 1 Jan.-March 2021, doi: 10.1109/TNSE.2020.3036855.
- [27] J. Xu, A. Wang, J. Wu, C. Wang, R. Wang, and F. Zhou, "SPCSS: Social Network Based Privacy-Preserving Criminal Suspects Sensing," in IEEE Transactions on Computational Social Systems, vol. 7, no. 1, pp. 261-274, Feb. 2020, doi: 10.1109/TCSS.2019.2960857.
- [28] C. Xu, L. Zhu, Y. Liu, J. Guan, and S. Yu, "DP-LTOD: Differential Privacy Latent Trajectory Community Discovering Services over Location-Based Social Networks," in IEEE Transactions on Services Computing, vol. 14, no. 4, pp. 1068-1083, 1 July-Aug. 2021, doi: 10.1109/TSC.2018.2855740.
- [29] L. Yin, J. Feng, H. Xun, Z. Sun and X. Cheng, "A Privacy-Preserving Federated Learning for Multi-party Data Sharing in Social IoTs," in IEEE Transactions on Network Science and Engineering, vol. 8, no. 3, pp. 2706-2718, 1 July-Sept. 2021, doi: 10.1109/TNSE.2021.3074185.
- [30] P. Zhang, Y. Wang, N. Kumar, C. Jiang, and G. Shi, "A Security- and Privacy-Preserving Approach Based on Data Disturbance for Collaborative Edge Computing in Social IoT Systems," in IEEE Transactions on Computational Social Systems, vol. 9, no. 1, pp. 97-108, Feb. 2022, doi: 10.1109/TCSS.2021.3092746.
- [31] T. Zhu, J. Li, X. Hu, P. Xiong, and W. Zhou, "The Dynamic Privacy-Preserving Mechanisms for Online Dynamic Social Networks," in IEEE Transactions on Knowledge and Data Engineering, vol. 34, no. 6, pp. 2962-2974, 1 June 2022, doi: 10.1109/TKDE.2020.3015835.
- [32] A. Abbasi, A. R. Javed, C. Chakraborty, J. Nebhen, W. Zehra and Z. Jalil, "ElStream: An Ensemble Learning Approach for Concept Drift Detection in Dynamic Social Big Data Stream Learning," in IEEE Access, vol. 9, pp. 66408-66419, 2021, doi: 10.1109/ACCESS.2021.3076264.
- [33] X. Chen, Y. Yuan, L. Lu, and J. Yang, "A Multi-dimensional Trust Evaluation Framework for Online Social Networks Based on Machine Learning," in IEEE Access, vol. 7, pp. 175499-175513, 2019, doi: 10.1109/ACCESS.2019.2957779.
- [34] T. Gao, J. Yang, W. Peng, L. Jiang, Y. Sun and F. Li, "A Content-Based Method for Sybil Detection in Online Social Networks via Deep Learning," in IEEE Access, vol. 8, pp. 38753-38766, 2020, doi: 10.1109/ACCESS.2020.2975877.
- [35] S. Garg, K. Kaur, N. Kumar and J. J. P. C. Rodrigues, "Hybrid Deep-Learning-Based Anomaly Detection Scheme for Suspicious Flow Detection in SDN: A Social Multimedia Perspective," in IEEE Transactions on Multimedia, vol. 21, no. 3, pp. 566-578, March 2019, doi: 10.1109/TMM.2019.2893549.
- [36] B. Mei, Y. Xiao, R. Li, H. Li, X. Cheng, and Y. Sun, "Image and Attribute Based Convolutional Neural Network Inference Attacks in Social Networks," in IEEE Transactions on Network Science and Engineering, vol. 7, no. 2, pp. 869-879, 1 April-June 2020, doi: 10.1109/TNSE.2018.2797930.
- [37] G. Sansonetti, F. Gasparetti, G. D'aniello, and A. Micarelli, "Unreliable Users Detection in Social Media: Deep Learning Techniques for Automatic Detection," in IEEE Access, vol. 8, pp. 213154-213167, 2020, doi: 10.1109/ACCESS.2020.3040604.
- [38] M. Shen, J. Zhang, L. Zhu, K. Xu, and X. Tang, "Secure SVM Training Over Vertically-Partitioned Datasets Using Consortium Blockchain for Vehicular Social Networks," in IEEE Transactions on Vehicular Technology, vol. 69, no. 6, pp. 5773-5783, June 2020, doi: 10.1109/TVT.2019.2957425.
- [39] K. Zhang, Z. Tian, Z. Cai, and D. Seo, "Link-privacy preserving graph embedding data publication with adversarial learning," in Tsinghua Science and Technology, vol. 27, no. 2, pp. 244-256, April 2022, doi: 10.26599/TST.2021.9010015.
- [40] X. Song, W. Jiang, X. Liu, H. Lu, Z. Tian, and X. Du, "A survey of game theory as applied to social networks," in Tsinghua Science and Technology, vol. 25, no. 6, pp. 734-742, Dec. 2020, doi: 10.26599/TST.2020.9010005.
- [41] J. Du, C. Jiang, K. -C. Chen, Y. Ren, and H. V. Poor, "Community-Structured Evolutionary Game for Privacy Protection in Social Networks," in IEEE Transactions on Information Forensics and Security, vol. 13, no. 3, pp. 574-589, March 2018, doi: 10.1109/TIFS.2017.2758756.
- [42] K. Fan et al., "A Secure and Verifiable Data Sharing Scheme Based on Blockchain in Vehicular Social Networks," in IEEE Transactions on Vehicular Technology, vol. 69, no. 6, pp. 5826-5835, June 2020, doi: 10.1109/TVT.2020.2968094.
- [43] L. Gao, Z. Yan, and L. T. Yang, "Game Theoretical Analysis on Acceptance of a Cloud Data Access Control System Based on Reputation," in IEEE Transactions on Cloud Computing, vol. 8, no. 4, pp. 1003-1017, 1 Oct.-Dec. 2020, doi: 10.1109/TCC.2016.2632110.
- [44] X. Wang et al., "Game Theoretic Suppression of Forged Messages in Online Social Networks," in IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 51, no. 3, pp. 1601-1611, March 2021, doi: 10.1109/TSMC.2019.2899626.
- [45] M. Huang et al., "A Game-Based Economic Model for Price Decision Making in Cyber-Physical-Social Systems," in IEEE Access, vol. 7, pp. 111559-111576, 2019, doi: 10.1109/ACCESS.2019.2934515.
- [46] F. Kong, Y. Zhou, B. Xia, L. Pan, and L. Zhu, "A Security Reputation Model for IoT Health Data Using S-AlexNet and Dynamic Game Theory in Cloud Computing Environment," in IEEE Access, vol. 7, pp. 161822-161830, 2019, doi: 10.1109/ACCESS.2019.2950731.
- [47] Z. Su and Q. Xu, "Security-Aware Resource Allocation for Mobile Social Big Data: A Matching-Coalitional Game Solution," in IEEE Transactions on Big Data, vol. 7, no. 4, pp. 632-642, 1 October 2021, doi: 10.1109/TBDATA.2017.2700318.
- [48] J. Sun, H. Xiong, S. Zhang, X. Liu, J. Yuan, and R. H. Deng, "A Secure Flexible and Tampering-Resistant Data Sharing System for Vehicular Social Networks," in IEEE Transactions on Vehicular Technology, vol. 69, no. 11, pp. 12938-12950, Nov. 2020, doi: 10.1109/TVT.2020.3015916.
- [49] L. H. Álvarez, J. M. de Fuentes, L. G. Manzano, and L. H. Encinas, "Privacy-Preserving Sensor-Based Continuous Authentication and User Profiling: A Review," MDPI-Sensor, vol.21, No.92, 2021. DOI: <https://dx.doi.org/10.3390/s21010092>.
- [50] S. Jin, V. V. Phoha and R. Zafarani, "Graph-Based Identification and Authentication: A Stochastic Kronecker Approach," in IEEE Transactions on Knowledge and Data Engineering, vol. 34, no. 7, pp. 3282-3294, 1 July 2022, doi: 10.1109/TKDE.2020.3025989.
- [51] L. Megouache, A. Zitouni, and M. Djoudi, "Ensuring user authentication and data integrity in the multi-cloud environment," Springer-Human-Centric Computing & Information Sciences, vol.10, No.15, 2020.DOI: <https://doi.org/10.1186/s13673-020-00224-y>.
- [52] N-E Park, S-H Park, Y-S Oh, J-H Moon, and I-G Lee, "Distributed Authentication Model for Secure Network Connectivity in Network Separation Technology," MDPI-Sensors, vol.22, No.579, 2022. DOI:<https://doi.org/10.3390/s22020579>.
- [53] O. Ruan, L. Zhang, and Y. Zhang, "Location-sharing protocol for privacy protection in mobile online social networks," EURASIP Journal on Wireless Communications and Networking, 2021.DOI: <https://doi.org/10.1186/s13638-021-01999-z>.
- [54] V.K. Sinha, D. Anand, S. Kaur, P. Singh, and I. D. Noya, "Security Verification of Social Network Model Using Improved Three-Party Authenticated Key Exchange Protocol," MDPI-Symmetry, vol.14, No.1567, 2022. DOI: <https://doi.org/10.3390/sym14081567>.
- [55] D. Soni, D. Srivastava, A. Bhatt, A. Aggarwal, S. Kumar, and M.A. Shah, "An Empirical Client Cloud Environment to Secure Data Communication with Alert Protocol," Hindawi-Mathematical Problems in Engineering, Volume 2022, Article ID 4696649, 14 pages, 2022.
- [56] Z. Tu, H. Zhou, K. Li, H. Song, and Y. Yang, "A Blockchain-Enabled Trusted Protocol Based on Whole-Process User Behavior in 6G Network", Hindawi-Security and Communication Networks, Volume 2022, Article ID 8188977, 12 pages, 2022.

- [57] D. Xu, W. Wang, L. Zhu, J. Zhao, F. Wu, and J. Gao, "CL-BC: A Secure Data Storage Model for Social Networks," *Hindawi-Security and Communication Networks*, Volume 2022, Article ID 5428539, 13 pages, 2022.
- [58] X. Zuo, L. Li, H. Peng, S. Luo and Y. Yang, "Privacy-Preserving Subgraph Matching Scheme With Authentication in Social Networks," in *IEEE Transactions on Cloud Computing*, vol. 10, no. 3, pp. 2038-2049, 1 July-Sept. 2022, doi: 10.1109/TCC.2020.3012999.
- [59] https://ieeexplore.ieee.org/search/searchresult.jsp?queryText=security,%20social%20network&highlight=true&returnFacets=ALL&returnType=SEARCH&matchPubs=true&ranges=2012_2023_Year.

A Novel Framework for Semi-supervised Multiple-label Image Classification using Multi-stage CNN and Visual Attention Mechanism

Joseph James S*, Lakshmi C

Department of Computational Intelligence, Faculty of Engineering and Technology, SRM Institute of Science and Technology, Kattankulathur, Chennai, Tamil Nadu, India

Abstract—To train deep neural networks effectively, a lot of labeled data is typically needed. However, real-time applications make it difficult and expensive to acquire high-quality labels for the data because it takes skill and knowledge to accurately annotate multiple label images. In order to enhance classification performance, it is also crucial to extract image features from all potential objects of various sizes as well as the relationships between labels of numerous label images. The current approaches fall short in their ability to map the label dependencies and effectively classify the labels. They also perform poor to label the unlabeled images when small amount of labeled images available for classification. In order to solve these issues, we suggest a new framework for semi-supervised multiple object label classification using multi-stage Convolutional neural networks with visual attention (MSCNN) and GCN for label co-occurrence embedding (LCE) (MSCNN-LCE-MIC), which combines GCN and attention mechanism to concurrently capture local and global label dependencies throughout the entire image classification process. Four main modules make up MSCNN-LCE-MIC: (1) improved multi-label propagation method for labeling largely available unlabeled image; (2) a feature extraction module using multi-stage CNN with visual attention mechanism that focuses on the connections between labels and target regions to extract accurate features from each input image; (3) a label co-existence learning that applies GCN to discover the associations between different items to create embeddings of label co-occurrence; and (4) an integrated multi-modal fusion module. Numerous tests on MS-COCO and PASCAL VOC2007 show that MSCNN-LCE-MIC significantly improves classification efficiency on mAP 84.3% and 95.8% respectively when compared to the most recent existing methods.

Keywords—Semi-supervised; visual attention; multi-label; image classification; label propagation

I. INTRODUCTION

Multiple labels image classification has lately sparked a lot of interest in areas such as human attribution recognition, music mood categorization, and multi-object recognition. Multi-label image categorization, as compared to single-label image recognition, turns into a useful and difficult job that necessitates a deeper comprehension of the image objects because images from real world application always encompass numerous objects that contain extensive semantic information. Numerous of noteworthy works have been suggested to investigate the semantic connections between different labels

and accomplish successful classification of multiple label images. There are two major groups into which these strategies fit. In contrast to the former which focuses on learning the regional correlations between target labels and true labels, the latter uses a (GCN) [1] to acquire the overall label relationships among different objects. This multi-label image, as seen in Fig. 1, is typical in real life and primarily includes three objects: "person," "tennis ball," and "tennis racket." Each of these objects is annotated based on the target regions that have been highlighted in the image rather than other areas of it.

Wang et al. [2] and others [3, 4, 5] coupled recurrent neural network (RNN) and convolution neural network (CNN) to mutually describe the image label significance and label relationships, but they failed to take into account of the geographic setting of images. The main limitation of these techniques is that they ignore the intricate topology structure that exists between objects. This encourages study into methods for identifying and examining the label dependency in different approaches.

To directly model label dependencies, some methods built around RNN [2] or stochastic graph models [6, 7] have been suggested. The first approach views the multiple label problem of classification as a systematic reasoning problem that might have scaling issues due to its high computational complexity. Another work uses attention mechanisms to tacitly model the label correlations [8]. The global associations between labels that must be derived from information drawn from visuals other than the one being studied are still being ignored. Instead, they consider correlations between observed areas of an image, also referred to as regional associations. To address this issue, Zhu et al. [9] propose the Spatial Regularization Network (SRN), which learns an attention map for each label by using image-level supervisions and focused on the associated image region for every label. To update the complete network, however, they employ a subpar multi-step training workflow. Despite the fact that these methods [10] analyze label relationships with mechanisms of attention, they only take into account a small amount of local association among various target regions that show in a single image and ignore association of the labels distribution on every one of the training images.



Fig. 1. Multiple label images.

Fig. 1 shows that if two things are semantically associated in the real world, they are more probable to appear together in an image compared to when they are not. The ML-GCN architecture developed by Chen et al. [11] records the global label correlations on all training examples by employing GCN to generate label relationship embedding using the label statistics. It has been suggested that A-GCN [12] represent the label correlations in reference to ML-GCN by developing a method for label network construction. However, these two approaches use the dot product (DP) to finish the fusion of label co-occurrence embeddings and image features, severely impeding model convergence and multiple label image categorization performance improvement. In this work, we suggest a new multi-staged CNN with a visual attention mechanism to extract features and produce better feature representations from the training images. This method captures the local and global label dependency and speeds up model convergence. We use a more effective label propagation technique to label the unlabeled images and multi modal factorized bilinear pooling (MFB) [13] as an element of fusion. Then, we suggest an innovative semi-supervised multiple label classification of images model with an attention mechanism in order to successfully integrate visual characteristics and label co-occurrence embeddings (referred to as MSCNN-LCE-MIC). MSCNN-LCE-MIC is composed of four fundamental modules: cross-modal fusing with MFB; learning by label co-occurrence embedding with GCN; and feature extraction with multi-staged CNN, attention mechanisms, and improved label propagation techniques.

In the first module, we use CNN inspired by VGG-16 to learn the features of the images by creating a feature map for each label with multi-stage CNN, a visual attention mechanism and an improved label propagation technique to obtain the labels for unlabelled images. All labels are transformed into word vectors in the second module before being used as inputs to GCN to produce label co-occurrence embedding. The MFB component is finally integrated into our task in the module of cross-modal fusion, where it helps us effectively combine label embeddings and image characteristics to allow a complete classification system. Widespread tests on MS-COCO [14] and PASCAL VOC2007 [15] show that MSCNN-LCE-MIC significantly improves convergence efficiency and outperforms existing methods in terms of classification outcomes. The following are the paper's main contributions:

1) By combining the multi-staged CNN with a visual attention process for effective extraction of features and GCN to simultaneously detect regional label relationships in an

image and universal label relationships among different items over the data distribution, we provide an innovative complete capable of training multi-label image categorization framework, MSCNN-LCE-MIC.

2) To effectively extract features of each object from an image, which will improve the performance of the classification model, we suggest a multi-staged CNN with a visual attention mechanism.

3) We incorporate an improved label propagation technique to add labels to the training data's unlabeled images, and we also incorporate the fusion component MFB into MSCNN-LCE-MIC to effectively combine features of image and embeddings of label co-occurrence. As a result, model convergence is considerably accelerated and classification performance is improved.

4) The proposed method, MSCNN-LCE-MIC, consistently outperforms earlier competing approaches and is simple to apply end-to-end. We put our approach to the test using benchmark datasets for multiple label image identification.

The remaining portions of the piece of writing are structured as follows: In Section II, we will review recent works related to the topic. Section III will cover the suggested framework for semi-supervised multiple label image classification, along with a comprehensive explanation of each component. Section IV will discuss the experimental design, the datasets used, the architecture, and analysis of the results obtained. Section V, reviews the findings and concludes the whole work with the proposal to future research options.

II. RELATED WORKS

A. MultipleLabel Image Classification

Due to the emergence of large-scale datasets like Image Net [16], MSCOCO [14], and PASCAL VOC [15], as well as the quick advancement of deep convolutional networks [3, 17], the efficiency of image categorization has seen a quick improvement. Extending deep CNNs for categorization of images with multiple labels has received a lot of attention. Image categorization using a single label techniques have advanced significantly with the quick creation of CNN-based models. Experts employed binary methods to classify images that have multiple labels through developing a binary classification algorithm for each label. To recognize multi-label photos, [18] apply pre-built features of ImageNet. Chatfield et al.'s [19] model success is improved by using the target dataset to create task-specific image characteristics. A deep network that excels on particular items is VeryDeep [4]. These methods, however, deal every object in the image separately and ignore the connections between various elements. To jointly describe the label significance and relationships, Wang et al. suggest CNN-RNN [2].

The SRN created by Zhu et al. [9] places an emphasis upon the associated image area associated with every label and uses image level oversights to obtain an attention map for each label. They are unable to finish end-to-end training because they update the entire network using a subpar multi-step training workflow. The study [20] suggests employing

GCAM to record the label relationships between diverse image transforms. Graph convolution network (GCN) is used by MLGCN [11] and AGCN [12] to produce label co-occurrence embedding for multiple label picture categorizations. Unfortunately, they use dot product (DP) to finish the process of fusing label co-occurrence embeddings and image features, substantially impeding system convergence and more effectively classify images with multiple labels.

B. Learning with Structured Graph

By utilizing graph propagation and reasoning, it is possible to model the relationships between labels in a useful fashion as well. In order to improve image categorization, [21] used neural networks to analyze data with a graph layout and provide a paradigm using GNNs to discover more characteristic relationships. In order to investigate the relationships between various labels of multiple label zero-shot learning, [22] used information graphs. For the study of graph-structured data, [1] introduced a GCN method, which utilized layer wise propagation to encode both graph data and node attributes. According to the aforementioned GCN, Semantic embeddings and classification associations were merged by Wang et al. [23] to predict the effectiveness of the visual classification for each group. Moreover, [11] provided a framework built on GCN that modeled the relationships between labels for multi-label categorization using a predefined graph.

C. Cross-modal Fusion

Recent years have seen a rise in interest in the disciplines of visual question and answering (VQA) [24, 25] and multi-modal counterfeit news discovery [26], which attempts to successfully combine vectors from several modules. These fields combine visual and linguistic representation derived from open, sizable linguistic or visual databases. Nevertheless, current techniques are unable to produce expressive image-text features or to comprehend the complex connections between these characteristics. The majority of available solutions merely integrate cross-modal embeddings using linear models such element-wise addition. It has been suggested that MCB and MLB significantly lower the calculation costs associated with the VQA fusion procedure to address this issue. However, there are two significant drawbacks: MLB requires too much iteration before it converges, while MCB can only produce good results if it collects multi-dimensional feature vectors. Additionally, [13] proposed MFB, which successfully combines image features and text embeddings while also noticeably accelerating model convergence. MFB first transforms vectors of high dimension into vectors of low dimension before combining the matching position element of cross modal vectors. By incorporating MFB to be trained on the multiple label picture characteristics, F-GCN [27] successfully addresses the cross-modal fusion problem, but it disregards each image's attention process.

The problems in the existing methods are lack of efficient feature extraction methods to handle morphological similarity

issues and label dependency that exists between objects of an image.

The proposed work presents an innovative semi-supervised framework called MSCNN-LCE-MIC that acquires label relationship and their interdependence with GCNs in order to enhance the accuracy of multiple object label image categorization. The aforementioned structure learning methods served as inspiration for this framework. Our MSCNN-LCE-MIC explicitly stacks numerous GCN layers to build universal models for exploring the feature representations, in contrast with earlier structure learning methods. The main difference between MSCNN-LCE-MIC and rival approaches is that the proposed work uses an end-to-end methodology during the training stage to concurrently record localized label correlations within an image along with universal label correlations across multiple items in the data distribution. The label encoding and structured representation of graphs techniques used in MSCNN-LCE-MIC are more efficient at addressing the over-smoothing and scaling problems that are brought on by the substantial depth of GCN-based architecture.

III. PROPOSED WORK

The general structure of our suggested framework, multi-stage CNN with label co-occurrence embedding and attention mechanism (MSCNN-LCE-MIC) for semi-supervised classification of multiple label images, is given away in Fig. 2. This framework includes (1) enhanced label propagation to obtain labels for unlabelled images, (2) a multi-stage CNN with visual attention to mine the each input image features, (2) a module for learning label co-occurrence embedding that is built on the GCN, and (3) a module for fusing the two cross modal vectors described above successfully. We describe our plan's workflow in more detail below.

A. Improved Label Propagation

The label propagation method (LPA) is a well-known clustering technique due to its popularity and freedom from parameter dependence. Label propagation is a technique for propagating labels from labeled to unlabeled picture data based on the association between the two object classes [28][29]. The similarity weights in conventional label propagation systems, on the other hand, may not be appropriate for subsequent label propagation because they distribute labels after a certain data graph generation process. This strategy offers benefits, but it also has some disadvantages. Another drawback of LPA is the unpredictability with which nodes are clustered, which increases instability and the creation of big communities. When nearby nodes are chosen at random from a list of fixed constant hops, a collection of disconnected nodes results. We suggest an LPA-based solution to overcome the problem of random community allocation and build better clusters. These variations improve the quality of communities that are found by taking advantage of node attribute values and link strength.

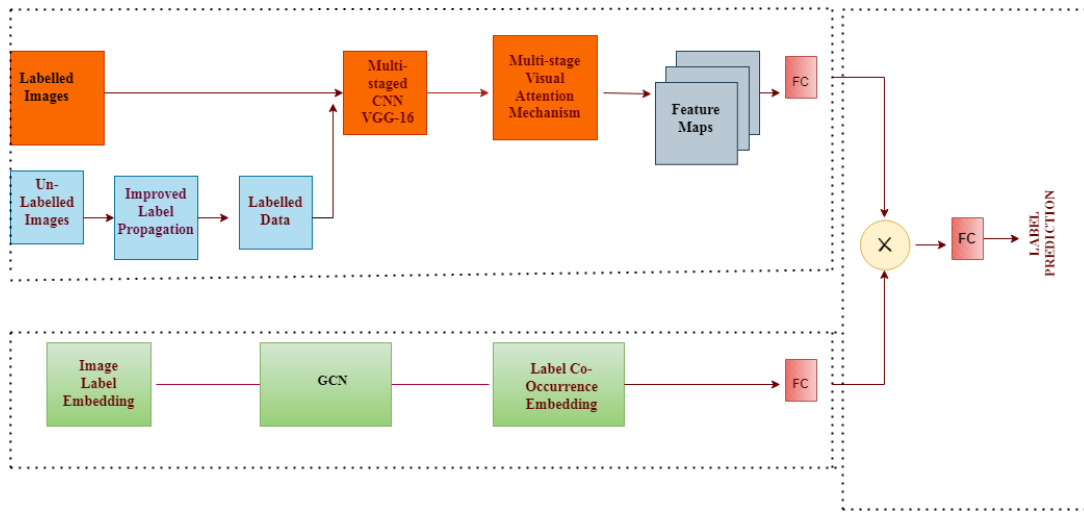


Fig. 2. Overall flow diagram of proposed method.

Using the feature data, a weighted graph is first created. The issue of random selection between nearest neighbors is then addressed via enhanced label propagation with dynamic connection strength metrics. The improved label propagation identifies the linked neighbors in the graph by using the heat kernel dispersion approach and its corresponding heat equation. The community arrangement of the node-attributed network is then returned by the program. In this form, the weighted common neighborhood (WCN) is employed as the link strength measure to award a community label (Murata & Moriyasu, 2007). The WCN value for each community label is then added together for the neighborhood. The system then chooses the community tag to apply to the selected node that has the highest total. The equation below describes how to measure the WCN connection strength.

$$Strength_{A,B} = \sum_{C \in \Gamma(A) \cap \Gamma(B)} w(A,C) + w(C,B) \quad (1)$$

The shared neighbors of nodes A and B are shown by the pair (A) (B). The edge weights that link the nodes A and B to their neighbor C are $w(A, C)$ and $w(C, B)$, respectively.

B. Multi-stage CNN

The projected model is based on an 18-layer CNN inspired by VGGNet [30]. However, as shown in Fig. 3, the proposed design classifies data by considering both high level and mid-level features. The second, third, and fourth network segments are used to extract these feature maps. The data from the first block is not directly used because it only contains minimal object label recognition filters. It is clear that the initial block's attributes made little of an impact. At this point, the networks can only understand and learn about simple textures, which is insufficient to define the essential traits required to differentiate objects.

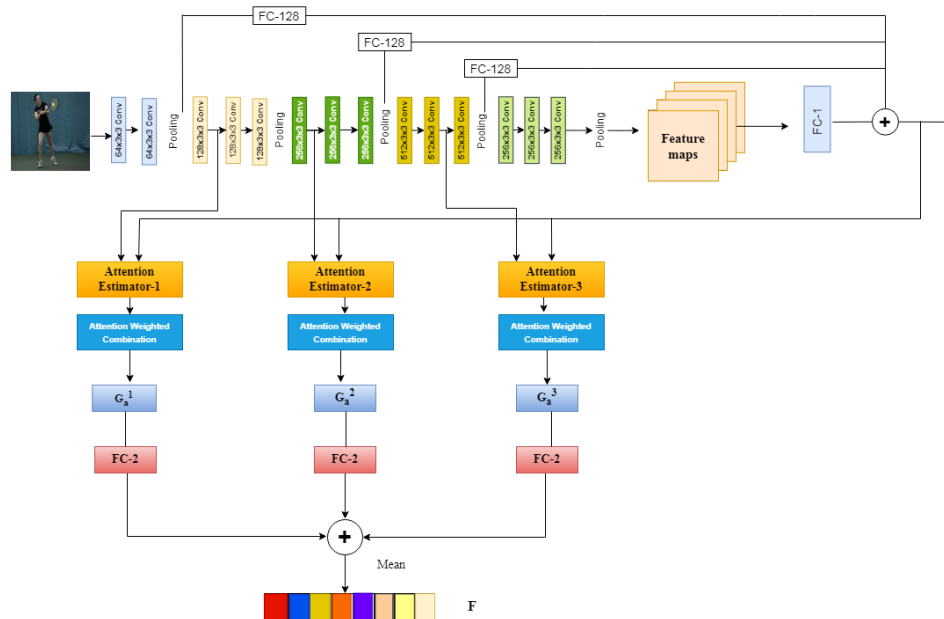


Fig. 3. Proposed multi-stage CNN with attention mechanism.

Furthermore, although mid-level layers in the second and third blocks produce local features, high-level layers in the fourth and fifth blocks greatly contribute to tiny object recognition via global characteristics. In this study, we propose MS-CNNs that combine local and global features in a predetermined way to produce the global feature vector g . Our networks, in contrast to conventional CNNs, generate a large number of essential characteristics that make the system robust to fluctuations in image quality and object occlusion problems. Also, a few pointless feature extraction filters are present in the same block, which is consistent with the discovery made regarding mid-level features in the prior section is given in eq.2.

$$(F_1^1, F_2^1, \dots)^T \oplus (F_1^2, F_2^2, \dots)^T = (F_1^1, F_2^1, \dots, F_1^2, F_2^2, \dots)^T \quad (2)$$

where $(F_1^1, F_2^1, \dots)^T$ and $(F_1^2, F_2^2, \dots)^T$ feature vectors that came from different network tiers. If the input is the result of Convolutional layers, vectorization is required before employing this operator.

C. Attention Mechanism

Attention estimators have been inserted after layers 7, 10, and 13. The attention estimator receives the layer 7 output and generates a "attention mask" consisting of integers between 0 and 1, then multiplies by output of the layer 7 to produce " g_a1 ". Following layers 10 and 13, the attention estimators go through a similar process, producing, respectively, g_a2 and g_a3 . There was often another fully connected layer (FC) after the FC layer with the number "16," but it has been eliminated, leaving only the fully connected layer that comes after the dense layer at the network end to make the label classification. Instead, the three attention estimators' inputs are now handled by a new fully connected layer. The process of visual attention method for feature extraction is described as follows,

Step 1: The compatibility score C is calculated with use of the regional feature vector l and the universal feature vector g . The compatibility score is meant to be high when the local characteristics-defined image patch contains components of the dominating picture category. For instance, if the image contains a multiple objects, we assume that the global feature vector g adequately describes all possible object features. The patch that most closely resembles a particular object is also expected to produce local traits l that, when paired with g , will result in a high compatibility score.

$$C_i^s = \langle l_i^s, g \rangle, i \in \{1, \dots, n\} \quad (3)$$

The local feature vector l and the global feature vector g are simply added together. Be aware that while the local features l will change based on the convolutional layer (layers 7, 10, and 13), the global feature vector g will remain constant. Projecting g into l 's lower-dimensional space will make sense if l and g are not the same dimensions.

Step 2: Calculate the Attention Weights a_i from the Compatibility Scores C shown in eq(4). The outcome is referred to as "a", after that the compatibility scores C are compressed into the range of (0, 1) using a softmax.

$$a_i^s = \frac{\exp(C_i^s)}{\sum_j^n \exp(C_j^s)}, i \in \{1, \dots, n\} \quad (4)$$

Step 3: Determine Each Layer's Final Attention Mechanism Output using equation 5.

$$g_a^s = \sum_{i=1}^n a_i^s l_i^s \quad (5)$$

Here, we calculate the attention mechanism's (g_a) final output for a certain layer using a weighted combination of the l for that particular layer (s). The attention weights "a" that we recently found are the weights that we employ.

Step 4: Create a categorization forecast based on final outcome of the attention module. We now want to select a classification using the attention outputs g_a that we just gathered for layers 7, 10, and 13. To acquire intermediate predictions, feed each attention output into a separate, completely connected layer. The final projections are then calculated by averaging these intermediate guesses.

D. Label Dependency Learning with GCN

According to [1] and [11], the label co-existence learning (LCL) module uses GCN of two-layers, with each layer taking the output of the layer before it and producing a new graph representation. Fig.4 shows the LCL module with GCN. The semantic encoding vectors $X = \{X_i\}_{i=1}^C$ and the equivalent association graph G are fed into the LCL component for the initial GCN layer. The graph representation and node feature may be easily combined by the LCL module in the convolution. The LCL module, which serves as the central part of the proposed MSCNN-LCE-MIC, intends to be trained on a set of classifier score $W \in R^{s \times d}$ to recalibrate early values for every label received from the image feature embedding module. In order to satisfy the requirements of the broadcast method, it is important to address two crucial challenges, namely word embedding and graph representation (5).

1) *Word embedding*: To retrieve the word embeddings for each of the labels across these multiple label image datasets, we use the 300-dim GloVe [31] algorithm learned on the Wikipedia dataset. We demonstrate that it is evenly efficient when using the suggested LCL method as demonstrated in Sections 4(c). To acquire the global label interdependence on the training set, we build a GCN with two layers. We use GloVe [31][33] to create 300 dimensional vectors for each of the objects in order to create the label embeddings matrix $Z \in R^{C \times 300}$, following MLCGN [32].

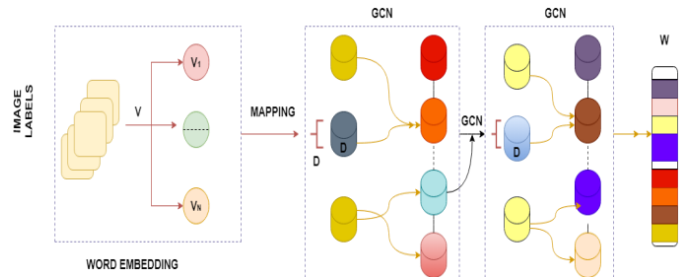


Fig. 4. Label Co-occurrence embedding network with GCN.

2) *Graph representation*: It goes without saying that the correlation matrix A affects how the node representations propagate. Instead of starting from scratch, researchers have suggested a number of methods for building the predetermined correlation matrix (explicit graph representations). For instance, WordNet [34] was used by Lee et al. [22] to develop their structured knowledge network. String matching was employed by Gao et al. [35] to plot concepts to nodes available in concept Net [36]. To express relationships between labels, these methods rely solely on semantic embedding and do not explicitly include information about interdependencies. Instead, we concentrate on the label co-existence matrix based on the training images, which is then used to represent the organized network composed of label associations by combining the label relationship data from neighboring nodes into a singular association matrix.

The following is the production procedure:

1) The first component of the label co-existence matrix P for the training image, we count the times at which pairs labels (L_i, L_j) first emerge.

$$P = \{p_i\}_{i=1,j=0}^c \in \mathfrak{R}^{c \times c} \quad (6)$$

2) The initial label correlation matrix P' can be constructed using the matrix R^{c x c}.

$$P' = \{p_i / N_i\}_{i=1,j=0}^c \in \mathfrak{R}^{c \times c} \quad (7)$$

where N_i is the object label's ith number. As a result, the graph constructed from such an asymmetric matrix P' is directed.

$$M_{i,j} = \begin{cases} p_{i,j} < \phi \\ 0, & p_{i,j} \geq \phi \\ \lambda^* \frac{p_{i,j}}{\sum_{j=0}^c p_j + \theta} \end{cases} \quad (8)$$

In order to safeguard label partnerships' specifics as stated in eq. (8), this work presents a nonlinear method for preparing matrix P' which can reduce noise (7), where φ is the noise filtration cutoff, 1^{e-6}, and M is the label association matrix employed in every GCN layer.

E. Multi-Label Classification Loss

Both the MS-COCO and the PASCAL VOC-2007 datasets exhibit the issue of class imbalance, which usually manifests as an inequity in the amount of negative and positive data. The suggested weighted cross entropy loss is taken into account as the multiple label categorization loss employed in our MSCNN-LCE-MIC to tackle this issue, and is described as:

$$Loss(b_i, l_i) = -wt_p \sum_{l_i=1} \log(\sigma(b_i)) - wt_n \sum_{l_i=0} \log(1 - \sigma(b_i)) \quad (9)$$

Where, $wt_p = \frac{|po| + |Ne| + |1|}{|Po| + |1|}$ and $wt_n = \frac{|po| + |Ne| + |1|}{|Ne| + |1|}$, |Po| and |Ne| are The total amount of both positive and negative image labels in a batch, b_i is belief of each class label, σ is the sigmoid function.

Formally, for all i ∈ [1,C], we will fuse F and W_i to obtain Yⁱ, or the ith component of the predicted label Y ∈ R^C, where W_i indicates the ith row vector of W. This is done provided the Ith image feature vector F. First, using two f_C layers, W_i and F will be transformed into the corresponding m dimensional vectors M1 and M2. Additionally, M1 and M2 are cross-modal vectors that will be multiplied element-wise into an m-dimensional vector M1 M2 to enhance the interaction between these two embeddings. We further convert M1 ⊗ M2 into a m / g dimensional vector M via group sum-pooling to decrease parameter inflation and over-fitting, the elements in each group are represented by the letter g, to hasten convergence. The ith constituent of Y is then obtained by creating a f_C layer. As a result, we are able to produce the full anticipated labels Y following fusion of C number of times. In order to create an end-to-end classification model, we employ the multiple label loss function by updating the Loss given in eq. (9) between predicted labels Y and the actual labels Y' ∈ {0, 1 }^C of Ith image.

IV. EXPERIMENTAL SETUP

A. Specifications of Implementation

PyTorch is used to perform all experiments. Each input image is resized to 448x448 before passing it to feature extraction component. Using the GloVe [37] model, each object is transformed to become a 300-dimensional vector that includes words in the label association embedding module. We set g = 2 to perform the group sum-pooling procedure and m = 358 to perform the fusion of the vectors of features and label co-existence embedded data with reference to the FGCN. A batch size of 32 is used when updating our network during the training procedure. The proposed model uses stochastic gradient descent (SGD) with 0.9 momentum, 10⁻⁴ weight decay and initial learning rate of 0.001.

1) *Datasets*: We conducted in-depth tests to confirm the effectiveness of MSCNN-LCE-MIC on MS-COCO [14] and PASCAL-VOC2007 [15]. To separate the datasets, we use the similar settings of MLGCN and FGCN. For further information, consult the references [5, 24].

2) *Metrics for evaluation*: We employ the following assessment metrics in accordance with mainstream techniques [9, 11]: (P-C) precision per class, (R-C) recall per-class, (mAP) mean average precision, F1 per class (F1-C) and F1 overall (F1-O). For fair comparisons, we also catalog the investigational findings on the top-3 classes of the categorization scores.

B. Investigational Results and Discussions

The convergence effectiveness and classification outcomes of MSCNN-LCE-MIC are compared to those of cutting-edge image classification methods.

1) *Convergence efficiency*: We track how the number of training epochs changed the mAP on the examination set in this section. We execute this experiment using the same parameters as ML-GCN, including SGD, batch size, learning rate, datasets, loss function, etc., to allow for fair comparisons. MSCNN-LCE-MIC has, on MS-COCO and PASCAL VOC2007, converged at the 25th and 23th epochs, respectively, and it produces superior mAP of 84.3% and 95.8%. Instead, MLGCN has not yet converged, and its mAP values are 45.8% and 75.7% lower than MSCNN-LCE-MIC at this time. Moreover, it will take the MLGCN roughly 200 epochs (more than 10 times as long as MSCNN-LCE-MIC) to complete its learning process. These findings show that multi-stage CNN, visual attention and cross module fusion significantly quickens model convergence and enhances performance of classification model.

2) *Results comparisons with the existing models*

a) *Evaluation results on MS-COCO*: The most recent existing techniques are compared to MSCNN-LCE-MIC, including FGCN [27], AGCN [12], MLGCN [11], ResNet-101 [3], SRN [9], Order Free RNN [38], RNN with Attention [8], and CNN+RNN [2]. Fig. 5(a) and 5(b) shows the graphical representations of results obtained on MS-COCO dataset using proposed method. MSCNN-LCE-MIC nearly outperforms other options on all metrics which are shown in Table I and Table II. MSCNN-LCE-MIC significantly enhances the classification outcomes by 7% mAP in comparison to ResNet-101 baseline [3]. This occurrence shows that understanding the label dependency to produce more precise label features is greatly influenced by GCN. Also, as compared to those DP-based approaches, cross fusion module effectively completes the modal fusion to improve the efficiency of MSCNN-LCE-MIC. Additionally, based on the discrepancy between the earlier FGCN [27] and MLGCN [11], the multi-stage CNN with attention mechanism does in fact assist in extracting more precise image features, improving the effectiveness of categorization outcomes by 1.3% mAP.

b) *Evaluation results on VOC2007*: The state-of-the-art approaches are compared to MSCNN-LCE-MIC, including CNN+RNN [2], MLGCN [5], AGCN [6], ResNet-101 [8], Attention Reinforce [16], RNN with Attention [14], Very Deep [13], and FGCN [24]. The AP values and mAP values for each category are listed in Table III. With the exception of somewhat poorer performance on "bike", "bird" and "person" MSCNN-LCE-MIC outperforms other contenders overall in every category. The main cause is that the majority of earlier techniques ignored the local and global label dependencies, allowing them to solely focus on one or a few objects while ignoring the distribution of global labels. Our MSCNN-LCE-MIC, in contrast to existing approaches, considers both the regional label relationships in addition to the universal label relationships among diverse objects (multi-stage CNN with attention mechanism) within an image. MSCNN-LCE-MIC outperforms other methods for the remaining 17 objects. Although it appears that there is a trade-off between other

approaches and our MSCNN-LCE-MIC in this phenomenon, we think that this global viewpoint is essential for multi-label picture categorization. The MSCNN-LCE-MIC produces a superior mAP value of 1.8% when compared to the existing method FGCN [27], which shows that both the multi-stage CNN with attention mechanism and the cross fusion module are involved and help to provide more accurate classification results. In Table IV of Section IV C, we also discuss how these modules have an impact.

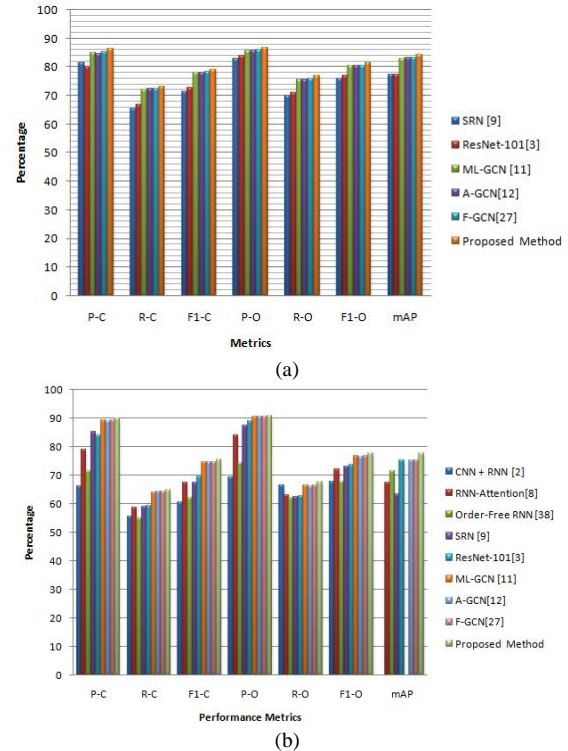


Fig. 5. Performance comparison of proposed method on MS-COCO dataset (a) Overall labels (b) top-3 class.

TABLE I. ACCURACY(%) OBTAINED ON MS-COCO IMAGE DATASET

Method	P-C	R-C	F1-C	P-O	R-O	F1-O	mAP
[9]	81.6	65.4	71.2	82.7	69.9	75.8	77.1
[3]	80.2	66.7	72.8	83.9	70.8	76.8	77.3
[11]	85.1	72.0	78.0	85.8	75.4	80.3	83.0
[12]	84.7	72.3	78.0	85.6	75.5	80.3	83.1
[27]	85.4	72.4	78.3	86.0	75.7	80.5	83.2
MSCNN-LCE-MIC	86.3	73.0	79.1	86.8	76.9	81.5	84.3

TABLE II. TOP-3 ACCURACY (%) ON MS-COCO DATASET

Method	P-C	R-C	F1-C	P-O	R-O	F1-O	mAP
[2]	66.0	55.6	60.4	69.2	66.4	67.8	-
[8]	79.1	58.7	67.4	84.0	63.0	72.0	67.4
[38]	71.6	54.8	62.1	74.2	62.2	67.7	71.5
[9]	85.2	58.8	67.4	87.4	62.5	72.9	63.2
[3]	84.1	59.4	69.7	89.1	62.8	73.6	75.2
[11]	89.2	64.1	74.6	90.5	66.5	76.7	-
[12]	89.0	64.2	74.6	90.5	66.3	76.6	75.2
[27]	89.3	64.3	74.7	90.5	66.6	76.7	75.2
MSCNN-LCE-MIC	89.9	65.0	75.4	91.0	67.7	77.6	77.8

TABLE III. CLASSIFICATION ACCURACY RESULTS (%) ON VOC 2007 DATASET

Method	plane	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	motor	person	plant	sheep	sofa	train	tv	mAP
CNN+RNN [2]	96.7	83.1	94.1	92.8	61.2	82.1	89.1	94.2	64.2	83.6	70.0	92.4	91.7	84.2	93.7	59.8	93.2	75.3	99.7	78.6	84.0
Very Deep [4]	98.9	95.0	96.8	95.4	69.7	90.4	93.5	96.0	74.2	86.6	87.8	96.0	96.3	93.1	97.2	70.0	92.1	80.3	98.1	87.0	89.7
ResNet-101 [3]	99.5	97.7	97.8	96.4	65.7	91.8	96.1	97.6	74.2	80.9	85.0	98.4	96.5	95.9	98.4	70.1	88.3	80.2	98.9	89.2	89.9
RNN+Attention[8]	98.6	97.4	96.3	96.2	75.2	92.4	96.5	97.1	76.5	92.0	87.7	96.8	97.5	93.8	98.5	81.6	93.7	82.8	98.6	89.3	91.9
Attention-reinforce[39]	98.6	97.1	97.1	95.5	75.6	92.8	96.8	97.3	78.3	92.2	87.6	96.9	96.5	93.6	98.5	81.6	93.1	83.2	98.5	89.3	92.0
MLGCN [11]	99.5	98.5	96.8	98.1	80.8	94.6	97.2	98.2	82.3	95.7	86.4	98.2	98.4	96.7	99.0	84.7	96.7	84.3	98.9	93.7	94.0
AGCN[12]	99.4	98.5	98.6	98.0	80.8	94.7	97.2	98.2	82.4	95.5	86.4	98.2	98.4	96.7	98.9	84.8	96.6	84.4	98.9	93.7	94.0
FGCN[27]	99.5	98.5	98.7	98.2	80.9	94.8	97.3	98.3	82.5	95.7	86.6	98.2	98.4	96.7	99.0	84.8	96.7	84.4	98.9	93.7	94.0
MSCNN-LCE-MIC	99.7	98.5	98.8	98.3	84.7	96.6	97.6	98.8	83.9	96.5	87.5	98.6	98.9	97.4	99.1	85.7	97.1	85.1	99.2	94.6	95.8

C. Ablation Studies

We carry out ablation study to examine how important elements and parameter settings affect MSCNN-LCE-MIC.

1) *Classification performance of MSCNN-LCE-MIC with/without different modules:* In this section, we examine how five crucial modules—namely, Multi-stage CNN, the visual attention method, the GCN, label propagation and the cross modal fusion affect our proposed model. Cross-modal fusion won't work if our suggested paradigm is used without GCN. Table IV displays the mAP results for MS-COCO and PASCAL VOC2007 with different module combinations. As can be seen, MSCNN-LCE-MIC performs best when all four of these modules are used at once. The evaluation findings show that any one module can enhance our model's mAP output. By incorporating the label dependencies between objects in global level, GCN especially improves MS-COCO and PASCAL VOC2007 mAP by 5.9% and 4.1%, respectively. In addition, the visual attention method boosts the mAP value by 0.3% on these two datasets while taking into account the label dependencies within an image. Finally, MFB keeps improving the classification outcomes with respective mAP improvements of 0.4% and 0.5%. These outcomes attest to the potency of our strategy, which notably benefits from both local and global level label dependencies, also the cross-modal fusion to boost the accuracy of classification.

2) *GCN with different layers:* The change in performance is shown in Table V and Fig. 6 after designing two (1024-2048), three (1024-1024-2048), and four (1024-1024-1024-2048) GCN layers in this section. With MS-COCO, MSCNN-LCE-MIC achieves its best performance with two GCN layers (Fig. 6(a)), after which its performance on mAP would degrade as GCN layers are added. Similar to Fig. 6(b), MSCNN-LCE-MIC produced the best outcome (mAP) on PASCAL VOC2007 using a 2-layer GCN. The main reason is because adding more layers of GCN would have a major negative impact on the model's functionality by making it so that the output features of nodes are no longer recognizable during the propagation process. We therefore employ two

GCN layers in order to acquire the label-occurrence embeddings.

3) *Number of units g in group sum pooling:* By using group sum-pooling, we reduce each **m** dimensional vector to a manageable small **m / g** dimensional vector. Changing **g** from 1 to 64 allows us to track how performance changes and the outcomes are exposed in Table VI. As can be shown, **g = 2** improves the outcome on MS-COCO even though the improvement in mAP on PASCAL VOC2007 is less pronounced. We think that **g = 2** is more appropriate for minimizing the dimension and convey the semantic significance of the top phrase, despite the fact that various values of **g** have a comparable effect.

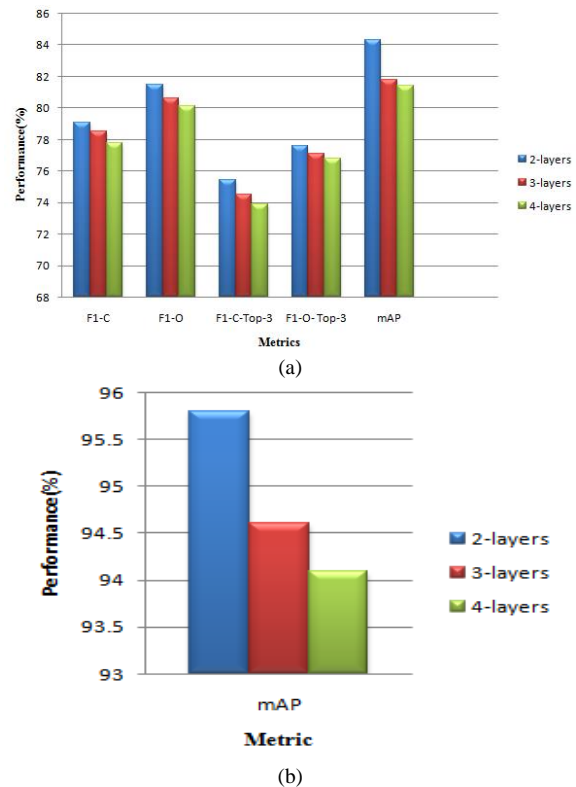


Fig. 6. Change in performance of proposed method with different GCN layers.

4) *Performance of improved label propagation:* The PASCAL VOC-2007 is a sizable multiple label standard dataset gathered for a number of computer vision tasks, including captioning, recognition and segmentation. The test collection contains 77,980 images, 24640 number of objects. There are about 2.4 object labels per picture, and there are 20 different classes in which the objects are divided. The 5K samples are used for testing, while the 20K samples are divided into 5 batches and used as training data. From each batch, we took 4k labeled samples at random and used the remaining 16K examples as unlabeled data. For each batch, this process is repeated five times. The end accuracy is the average of these trials. 32 mini-batch sizes were employed. Table VII and Table VIII display the outcomes of enhanced label propagation using various epochs. The proposed label propagation method achieves 3% less error than the existing label propagation method.

TABLE IV. CLASSIFICATION PERFORMANCE WITH / WITHOUT DIFFERENT MODULES

Improved label propagation	Multi-stage CNN	Visual Attention	GCN	MFB	mAP	
					MS-COCO	VOC-2007
Yes	Yes	Yes	Yes	Yes	84.3	95.8
Yes	Yes	Yes	No	No	83.5	94.4
Yes	No	No	No	No	77.4	90.6
No	No	Yes	Yes	Yes	83.7	94.6
No	Yes	No	Yes	Yes	84.0	94.6
No	Yes	Yes	No	No	77.2	90.2

TABLE V. PERFORMANCE COMPARISON ON GCN WITH VARIOUS LAYERS

Number of layers	MS-COCO				VOC-2007	
	F1-C	F1-O	F1-C-top-3	F1-O-top-3	mAP	mAP
2-layers	79.1	81.5	75.4	77.6	84.3	95.8
3-layers	78.5	80.6	74.5	77.1	81.8	94.6
4-layers	77.8	80.1	73.9	76.8	81.4	94.1

TABLE VI. MODEL PERFORMANCE WITH DIFFERENT G VALUE

Dimensions of g	mAP	
	MS-COCO	VOC-2007
1	83.8	95.6
2	84.3	95.8
4	84.1	95.6
8	84.0	95.4
16	83.7	95.2
32	83.5	95.1
64	83.4	94.9

TABLE VII. ERROR RATE COMPARISON ON CIFAR-10

Dataset	CIFAR-10			
	1000	2000	3000	4000
Labeled images				
LPA [40]	22.02±0.88	15.66±0.35	-	12.69±0.29
Improved LA(ours)	18.24±0.50	11.78±0.65	10.80±0.30	9.40±0.50

TABLE VIII. ERROR RATE COMPARISON ON PASCAL VOC-2007

Dataset	PASCAL VOC-2007			
	Labeled images	2000	4000	6000
LPA[40]		36.72±0.45	27.64±0.55	20.46±75
Improved LPA (ours)		30.33±0.25	20.25±0.56	17.70±0.44

V. CONCLUSION AND FUTURE WORK

This paper argues that most of the images in large scale datasets are unlabeled; the labels have dependency and morphological similarity issues. The existing methods using traditional convolution technique may fail to extract features efficiently and classify labels accurately due to large unlabeled data. We construct MSCNN-LCE-MIC, which combines multi-stage CNN with visual attention and GCN to concurrently collect global and regional level dependencies. A feature extraction component with multi-stage CNN and visual attention technique helps to create the most precise features of each image by concentrating on the relationships among labels and regions of target which solves the problem of morphological similarity issues exists between objects. MSCNN-LCE-MIC primarily consists of four key modules: improved label propagation technique to find labels of large volumes of unlabeled images available in real time applications; learning tool for label co-occurrence with GCN; and multi-stage CNN with attention mechanism. Comprehensive tests on MSCOCO and VOC2007 show that MSCNN-LCE-MIC significantly improves the effectiveness of our suggested framework and yields superior classification outcomes than the existing methods in the field. Experimental results demonstrate that the MSCNN-LCE-MIC method achieves higher mAP of 3.6% on MS-COCO dataset and 1.8% mAP on PASCALVOC-2007 dataset.

In the future, instead of using label propagation to train a neural network model, a generative adversarial network model could be used to get more labeled images from unlabeled image data and also reduce the time complexity due fusion of multiple components.

REFERENCES

- [1] Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks. In: 5th international conference on learning representations, Toulon, France, April 24-26, 2017.
- [2] Wang J, Yang Y, Mao J, Huang Z, Huang C, Xu W. CNN-RNN: A unified framework for multi-label image classification. In: 2016 IEEE conference on computer vision and pattern recognition, Las Vegas, NV, USA, June 27-30, 2016. p. 2285-94.
- [3] Kaiming He, Xiangyu Zhang, ShaoqingRen, and Jian Sun. Deep residual learning for image recognition. In CVPR, pages 770-778, 2016.
- [4] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. In ICLR, pages 1-8, 2015.
- [5] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and ZbigniewWojna. Rethinking the inception architecture for computer vision. In CVPR, pages 2818-2826, 2016.
- [6] Qiang Li, MaoyingQiao, Wei Bian, and Dacheng Tao. Conditional graphical lasso for multi-label image classification. In CVPR, pages 2977-2986, 2016.
- [7] Xin Li, Feipeng Zhao, and YuhongGuo. Multi-label image classification with a probabilistic label enhancement model. In UAI, pages 1-10, 2014.

- [8] Zhouxia Wang, Tianshui Chen, Guanbin Li, RuijiaXu, and Liang Lin. Multi-label image recognition by recurrently discovering attentional regions. In ICCV, pages 464–472, 2017.
- [9] Zhu F, Li H, Ouyang W, Yu N, Wang X. Learning spatial regularization with image-level supervisions for multi-label image classification. In: 2017 IEEE conference on computer vision and pattern recognition, Honolulu, HI, USA, July 21-26, 2017. p. 2027–36.
- [10] Mnih V, Heess N, Graves A, Kavukcuoglu K. Recurrent models of visual attention. In: Advances in neural information processing systems 27: annual conference on neural information processing systems, Montreal, Quebec, Canada, December 8-13, 2014. p. 2204–12.
- [11] Chen Z, Wei X, Wang P, Guo Y. Multi-label image recognition with graph convolutional networks. In: IEEE conference on computer vision and pattern recognition, Long Beach, CA, USA, June 16-20, 2019. p. 5177–86.
- [12] Li Q, Peng X, Qiao Y, Peng Q. Learning category correlations for multi-label image recognition with graph networks. 2019, CoRR abs/1909.13005.
- [13] Yu Z, Yu J, Xiang C, Fan J, Tao D. Beyond bilinear: Generalized multimodal factorized high-order pooling for visual question answering. IEEE Trans Neural Netw Learn Syst 2018;29(12):5947–59.
- [14] Lin T, Maire M, Belongie SJ, Hays J, Perona P, Ramanan D, Dollár P, Zitnick CL. Microsoft COCO: Common objects in context. In: Computer vision -ECCV 2014 - 13th European conference, Zurich, Switzerland, September 6-12, 2014. p. 740–55.
- [15] Everingham M, Gool LV, Williams CKI, Winn JM, Zisserman A. The pascal visual object classes (VOC) challenge. Int J Comput Vis 2010;88(2):303–38.
- [16] JiaDeng,Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. ImageNet: A large-scale hierarchical image database. In CVPR, pages 248–255, 2009.
- [17] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In CVPR, pages 7132–7141, 2018.
- [18] Razavian AS, Azizpour H, Sullivan J, Carlsson S. CNN features off-the-shelf: An astounding baseline for recognition. In: IEEE conference on computer vision and pattern recognition, Columbus, OH, USA, June 23-28, 2014. p. 512–9.
- [19] Chatfield K, Simonyan K, Vedaldi A, Zisserman A. Return of the devil in the details: Delving deep into convolutional nets. In: British machine vision conference, Nottingham, UK, September 1-5, 2014.
- [20] Wang Y, Xie Y, Liu Y, Fan L. G-CAM: Graph convolution network based class activation mapping for multi-label image recognition. In: ICMR '21: International conference on multimedia retrieval, Taipei, Taiwan, August 21-24, 2021. p. 322–30.
- [21] K. Marino, R. Salakhutdinov, and A. Gupta, “The more you know: Using knowledge graphs for image classification,” in Proc. Conf. Comput. Vision Pattern Recognit., 2016, pp. 20–28.
- [22] C.-W. Lee, W. Fang, C.-K. Yeh, and Y.-C. Frank Wang, “Multi-label zeroshot learning with structured knowledge graphs,” in Proc. IEEE Conf. Comput. Vision Pattern Recognit., 2018, pp. 1576–1585.
- [23] X. Wang, Y. Ye, and A. Gupta, “Zero-shot recognition via semantic embeddings and knowledge graphs,” in Proc. IEEE Conf. Comput. Vision Pattern Recognit., 2018, pp. 6857–6866.
- [24] Malinowski M, Fritz M. A multi-world approach to question answering about real-world scenes based on uncertain input. In: Advances in neural information processing systems 27: Annual conference on neural information processing systems, Montreal, Quebec, Canada, December 8-13, 2014. p. 1682–90.
- [25] Chen J, Zhang S, Zeng J, Zou F, Li Y-F, Liu T, Lu P. Multi-level, multi-modal interactions for visual question answering over text in images. World Wide Web 2021;1–17.
- [26] Zeng J, Zhang Y, Ma X. Fake news detection for epidemic emergencies via deep correlations between text and images. Sustainable Cities and Society 2021;66:102652.
- [27] Wang Y, Xie Y, Liu Y, Zhou K, Li X. Fast graph convolution network based multi-label image recognition via cross-modal fusion. In: The 29th ACM international conference on information and knowledge management, Virtual Event, Ireland, October 19-23, 2020. p. 1575–84.
- [28] I. Aviles-Rivero, N. Papadakis, R. Li, S. M. Alsaleh, R. T. Tan, and C.-B. Schonlieb, “When labelled data hurts: Deep semi-supervised classification with the graph 1-Laplacian,” 2019, arXiv:1906.08635. [Online]. Available: <http://arxiv.org/abs/1906.08635>.
- [29] AhmetIscen, GiorgosTolias, YannisAvrithis, Ondrej Chum, “Label Propagation for Deep Semi-supervised Learning”, Computer Vision and Pattern Recognition, IEEE Xplore, 2019.
- [30] T.-Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and S. Belongie, “Feature pyramid networks for object detection”, in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jul. 2017, pp. 936–944.
- [31] J. Pennington, R. Socher, and C. Manning, “Glove: Global vectors for word representation,” in Proc. Conf. Empirical Methods Natural Lang. Process., 2014, pp. 1532–1543.
- [32] B. Chen, J. Li, X. Guo, and G. Lu, “Dualhexnet: Dual asymmetric feature learning for Thoracic disease classification in chest X-Rays,” Biomed. Signal Process. Control, vol. 53, p. 101554, 2019.
- [33] H. Sak, A. Senior, and F. Beaufays, “Long short-term memory based recurrent neural network architectures for large vocabulary speech recognition,” 2014, arXiv:1402.1128.
- [34] G. A. Miller, “Wordnet: A lexical database for english,” Commun. The ACM, vol. 38, no. 11, pp. 39–41, 1995.
- [35] J. Gao, T. Zhang, and C. Xu, “I know the relationships: Zero-shot action recognition via two-stream graph convolutional networks and knowledge graphs,” in Proc. AAAI Conf. Artif. Intell., 2019, vol. 33, pp. 8303–8311.
- [36] H. Liu and P. Singh, “Conceptnet—a practical commonsense reasoning tool-kit,” BT Technol. J., vol. 22, no. 4, pp. 211–226, 2004.
- [37] Pennington J, Socher R, Manning CD. Glove: Global vectors for word representation. In: Proceedings of the 2014 conference on empirical methods in natural language processing, Doha, Qatar, a Meeting of SIGDAT, a special interest group of the ACL, October 25-29, 2014. p. 1532–43.
- [38] Chen S, Chen Y, Yeh C, Wang YF. Order-free RNN With visual attention for multi-label classification. In: Proceedings of the thirty-second AAAI conference on artificial intelligence, the 30th innovative applications of artificial intelligence, and the 8th AAAI symposium on educational advances in artificial intelligence, New Orleans, Louisiana, USA, February 2-7, 2018. p. 6714–21.
- [39] Chen T, Wang Z, Li G, Lin L. Recurrent attentional reinforcement learning for multi-label image recognition. In: Proceedings of the thirty-second AAAI conference on artificial intelligence, the 30th innovative applications of artificial intelligence, and the 8th AAAI symposium on educational advances in artificial intelligence, New Orleans, Louisiana, USA, February 2-7, 2018. p. 6730–7.
- [40] Iscen A, Tolias G, Avrithis Y and Chum O. Label Propagation for Deep Semi-supervised Learning, Computer Vision and Pattern Recognition, 2019, pp-5070-5079. <https://doi.org/10.48550/arXiv.1904.04717>.

Comparison of Predictive Machine Learning Models to Predict the Level of Adaptability of Students in Online Education

Orlando Iparraguirre-Villanueva¹ , Carmen Torres-Ceclén² , Andrés Epifanía-Huerta³ , Gloria Castro-Leon⁴ ,
Melquiades Melgarejo-Graciano⁵ , Joselyn Zapata-Paulini⁶ , Michael Cabanillas-Carbonell⁷ 

Facultad de Ingeniería y Negocios, Universidad Privada Norbert Wiener, Lima, Perú¹

Facultad de Ingeniería, Universidad Católica los Ángeles de Chimbote, Perú²

Facultad de Ingeniería, Universidad Tecnológica del Perú, Chimbote, Perú³

Facultad de Ingeniería y Gestión, Universidad Nacional Tecnológica de Lima Sur, Lima, Perú⁴

Facultad de Ciencias Empresariales, Universidad Científica del Sur, Lima, Perú⁵

Escuela de Posgrado, Universidad Continental, Lima, Perú⁶

Facultad de Ingeniería, Universidad Privada del Norte, Lima, Perú⁷

Abstract—With the onset of the COVID-19 pandemic, online education has become one of the most important options available to students around the world. Although online education has been widely accepted in recent years, the sudden shift from face-to-face education has resulted in several obstacles for students. This paper, aims to predict the level of adaptability that students have towards online education by using predictive machine learning (ML) models such as Random Forest (RF), K-Nearest-Neighbor (KNN), Support vector machine (SVM), Logistic Regression (LR) and XGBClassifier (XGB). The dataset used in this paper was obtained from Kaggle, which is composed of a population of 1205 high school to college students. Various stages in data analysis have been performed, including data understanding and cleaning, exploratory analysis, training, testing, and validation. Multiple parameters, such as accuracy, specificity, sensitivity, F1 count and precision, have been used to evaluate the performance of each model. The results have shown that all five models can provide optimal results in terms of prediction. For example, the RF and XGB models presented the best performance with an accuracy rate of 92%, outperforming the other models. In consequence, it is suggested to use these two models RF and XGB for prediction of students' adaptability level in online education due to their higher prediction efficiency. Also, KNN, SVM and LR models, achieved a performance of 85%, 76%, 67%, respectively. In conclusion, the results show that the RF and XGB models have a clear advantage in achieving higher prediction accuracy. These results are in line with other similar works that used ML techniques to predict adaptability levels.

Keywords—Machine learning; adaptability; students; online education; prediction; models

I. INTRODUCTION

In recent years, online education has experienced an unprecedented boom. The COVID-19 pandemic further accelerated this process, forcing millions of students and educators to adapt to a fully digital educational environment [1]. In response to the educational crisis caused by the COVID-19 pandemic, UNESCO has established partnerships with various international organizations, such as the International

Labour Organization (OIT), the United Nations High Commissioner for Refugees (UNHCR), the United Nations Children's Fund (UNICEF) and the World Health Organization (WHO) [2]. Companies such as Microsoft, Google, Facebook, and Coursera [3] also participate. With the aim of addressing the problems of connectivity, content through digital tools and seeking distance learning solutions for children, youth, and adults through innovation. Today, online education is not only a viable alternative to face-to-face education, but has become an indispensable tool for many students, professionals, and individuals around the world. This type of education is offered through online platforms and tools, such as websites, mobile applications, discussion forums, videoconferencing, among others [4], [5]. Since the advent of the COVID-19 pandemic, online education has been an indispensable alternative for students around the world [6]. However, the sudden transition from face-to-face to online education has posed many challenges to students, especially in terms of their ability to adapt to this new educational environment [7]. Adaptability is a fundamental skill that allows us to face new and unfamiliar situations with flexibility and creativity [8]. In the case of online education, adaptability refers to the ability of students to adjust to a new educational model [9], which requires different skills and tools than those used in face-to-face education [10]. However, not all students have the same ability to adapt to online education and some may have difficulty staying motivated and engaged [11]. Early identification of these students can be key to providing them with appropriate support and improving their online academic success [12]. For example, in Latin America and the Caribbean, Brazil has the highest rate of online students with 49%, Mexico with 15%, Colombia with 11%, Argentina with 8%, Chile with 6% and Peru with 5% [13] as shown in Fig. 1.



Fig. 1. Representation of online education in Latin America and the Caribbean.

Online education has revolutionized the world of teaching and learning, providing new opportunities to access education from anywhere in the world [14], [15]. In recent years, ML has started to play a very important role in online education, enhancing the learning experience and personalizing education for each student [16], [17]. ML is a discipline within the field of artificial intelligence (AI) that provides the ability for computers to learn and improve their performance on a specific task without being explicitly programmed. In the field of online education, ML is used to analyze student data and create personalized learning models [18], [19]. This work aims to predict the level of adaptability of students in online education for which we use the following models: RF, KNN, SVM, LR and XGB of ML. Furthermore, this work aims to provide a more in-depth understanding of how ML is transforming online education and how it can be used to improve education in the future. In addition, it will examine the challenges and limitations associated with the use of ML in online education. The predictive capacity of the models is also sought, for which the results of the trained models are compared.

This document is structured as follows: Section II presents the related works. Section III explains the materials and methods used, as well as the process of data collection and understanding. Section IV presents the results obtained after model processing and training. Section V discusses the analysis and discussion of the results, as well as the performance of the models. Finally, Section VI presents the conclusions drawn from the study and proposes future work.

II. RELATED WORKS

Predicting the level of student adoptability in online education has become an important issue. In this context, researchers and academics have carried out work related to ML

models for the analysis of distance education and student adaptability, these methods involve the use of algorithms that allow the analysis of large volumes of data and the extraction of significant patterns. For example, in [20] they developed a study on the role of student adaptability in virtual education during the COVID-19 pandemic, in this work they used a population of 1548 students. The findings showed that adaptability is a crucial personal resource that can help students in their online learning process. Similarly, in [21] The authors conducted an investigation to estimate the level of student adaptability in virtual education using ML techniques, where they explored the gap between virtual and face-to-face education from a Bangladeshi context. They found that 8.3% of the students adapted with high ease to online education, they also identified factors that make it difficult for them to adapt, such as Internet connectivity, lack of knowledge in the use of digital tools, problems of access to electricity, among others. Also, in [22] they presented the results of a survey conducted in June 2022 to beginning teachers, analyzing factors such as technology, Internet accessibility, teachers' competencies, pedagogical knowledge and learning opportunities. The results showed that for teachers to adapt easily to online education, they must have certain digital competencies, such as the adoption of information and communication technologies. Similarly, in the following work [23] they proposed an adaptive system to predict an educational path by applying different data mining algorithms to extract relevant features and build a personalized model. The results showed that ML algorithms are efficient in their performance and accuracy. Similarly, [24] made use of ML algorithms to predict student performance in online education. For which they used deep neural network algorithms and SVM and decision tree (DT) classifiers. As a result, they concluded that the SVM classifier is the one that obtains the best results in its accuracy and performance. Furthermore, in [25] examined the prospects, advantages, and problems of artificial intelligence (AI) in online education. They used ML algorithms to classify the level of adaptability of the students. As a result, an accuracy level of 93% was obtained, RF models and neural networks proved to be the most efficient. Similarly, the authors in [26] developed a study to determine the influence of IA on students' social adaptability. They used a population of 1328 students. They used ML algorithm to determine the influence on student adaptability, using variables such as: teacher-student, parents and society, age, connectivity, technological competencies, among others. The results showed that IA has a positive influence on adaptability in online education. Neural network classifiers are contributing significantly to learning prediction. For example, in the work [27] they developed a paper that seeks to associate learning styles with distance education student behavior, in which they applied linear regression models. The results of the study indicated that no significant relationship was found between learning styles and student performance in distance education. Furthermore, in the article [28], a study was conducted on the performance of students in online education using the deep learning short- and long-term memory (LSTM) technique, in which a total of 22437 students participated. As a result, an accuracy of 0.8457, precision of 0.8224 and F1-Score of 0.7943 were obtained using the LSTM technique.

III. METHODOLOGY

The purpose of this section is to present the methodology used in this work, which is divided into two parts: A) description of the ML models (RF, KNN, SVM, LR and XGB); B) the development of the case study. Within the case development process, the following stages are followed: 1) data set collection; 2) data set processing and exploratory data analysis; 3) training and validation of the ML models.

A. Description of the ML Models

ML is composed of a set of algorithms for regression, classification, reduction, clustering, among other types of algorithms. However, for this work that aims to predict the adaptability of online students, five classification algorithms have been chosen, such as: RF, KNN, SVM, LR and XGBC.

1) *Random forest*: The supervised learning (SL) algorithm known as RF model, according to [29], is used for classification and regression. This model is made up of a series of decision trees, where each is trained with a random subset of data and cast its vote for the outcome, as described in [30], as depicted in Fig. 2. This reduces overfitting and increases the accuracy of the model. It also allows the identification of the most important features for prediction.

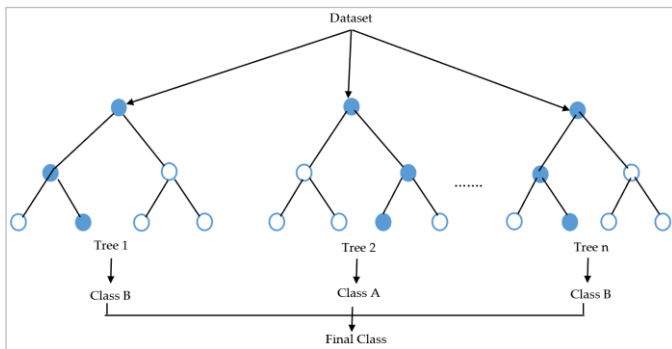


Fig. 2. RF model classification process.

2) *K-Nearest neighbor*: The K-NN model is an SL algorithm used for classification and regression [31]. In the context of regression, the k nearest neighbor's method consists of estimating the value of a data point from the values of the k nearest neighbors to that point. K-NN is simple, efficient, and easy to understand, but it can be sensitive to outliers and does not work well with high dimensionality data [32]. The formulation for calculating the vector distance is represented in equation (1).

$$d(q, p) = \sqrt{\sum_{i=1}^n (q_i - p_i)^2} \quad (1)$$

where q_n represents the results of one observation and p_n also represents the results of another observation.

3) *Support vector machine*: The SVM model is an SL algorithm used for classification and regression [33]. This hyperplane is found by maximizing the distance between the

closest points of each class, called support vectors [34]. SVM is effective in high dimensionality spaces and can handle nonlinear data using kernel functions [35]. However, it can be sensitive to hyperparameter selection and requires a large amount of data processing resources.

4) *Logistic regression*: The LR model is an SL algorithm used for binary classification. It models the probability that a data point belongs to the positive or negative class using the logistic function [36]. The model is trained using the likelihood function and fit using optimization methods such as gradient descent [37]. The LR is simple and easy to understand but can have problems with nonlinear data and can be sensitive to outliers [38]. The model is represented in equation (2) and (3).

$$PY = 1|X = \frac{1}{1 + \exp w_{o+} + \sum_{i=1}^n w_i x_i} \quad (2)$$

and

$$PY = 0|X = \frac{\exp w_{o+} + \sum_{i=1}^n w_i x_i}{1 + \exp w_{o+} + \sum_{i=1}^n w_i x_i} \quad (3)$$

Where, $PY|X$ represents the parametric distribution, Y is the described value, and $X = x_1, \dots, x_n$ is a vector with continuous values.

5) *eXtreme Gradient Boosting*: XGB is a decision tree-based ML algorithm used for regression and classification [39]. XGB improves the model at each iteration by adding sequential trees that focus on the most difficult to predict data points [40]. It uses a combination of gradient boosting and regularization to avoid overfitting and improve model accuracy [41]. XGB is scalable and can be used on large and complex datasets.

B. Case Study

1) *Understanding and collecting the data set*: In this section, the main task is to analyze and understand the dataset on online student adaptability to extract relevant information that can help improve the online learning experience. The first step is to understand the context of online education and how students' adaptability affects their experience. The dataset used in this work was obtained from Kaggle, this dataset is composed of a population of 1205 students between high school and college. The dataset is composed of 14 attributes such as: gender, age, educational level, type of educational institution, technology student (yes/no), location, internet quality, economic condition, type of internet, type of device to connect to the internet, duration of classes, knowledge of LMS, class schedule and level of adaptability, as shown in Fig. 3.

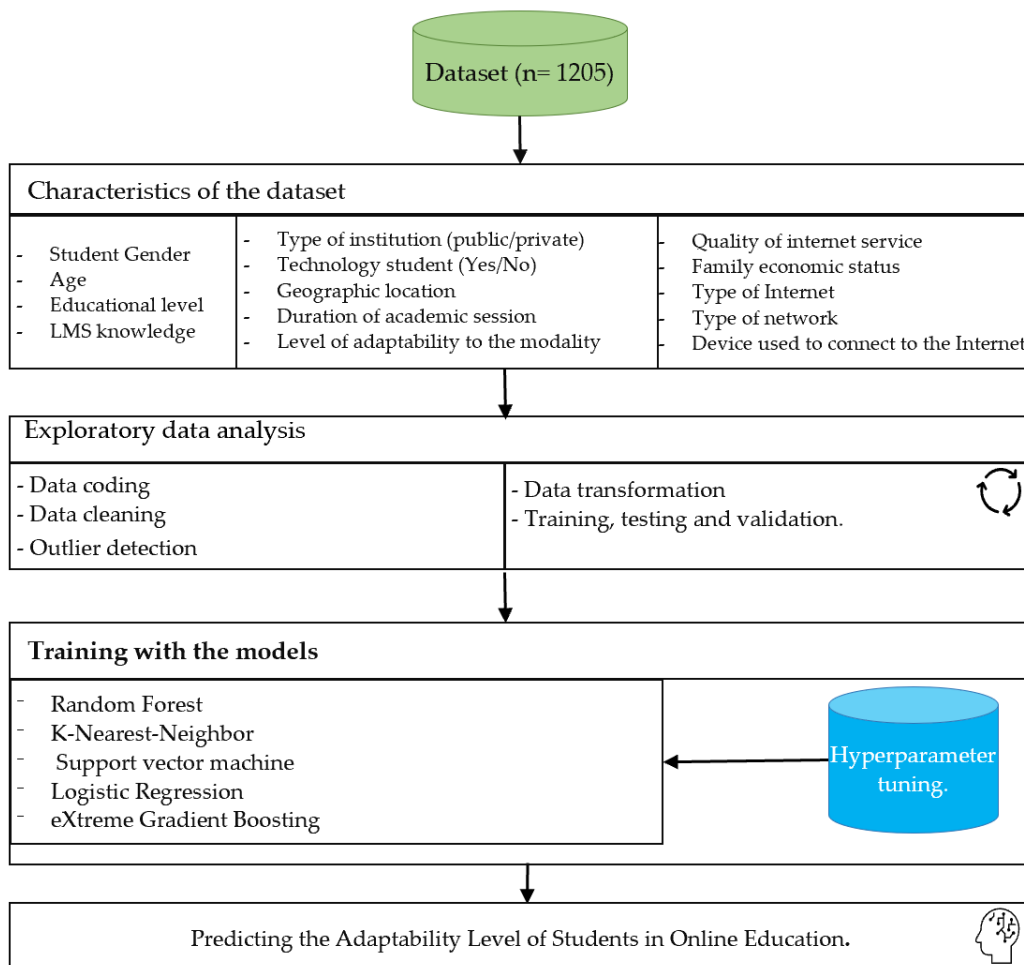


Fig. 3. The process shows the extraction, selection of attributes, analysis, training with ML models and training results.

Generally, when the data set is small, as is the case, and to solve this limitation, Synthetic Minority Over-sampling Technique (SMOTE) is used in this paper to deal with class imbalance in the data. SMOTE solves the problem by creating synthetic instances by interpolating the minority class in the dataset in order to increase the size of the minority class in the dataset. In this context, predicting the adaptability of online learners directly contributes to educators and educational program managers by anticipating students' needs and difficulties and providing immediate answers and supports to improve their online learning experience. In addition, ML methods directly help designers of online educational programs to improve the quality of their educational offerings and make them more accessible and effective for students.

2) *Data set processing*: The processing of the dataset involves performing a set of tasks, such as: data cleaning, transforming, and analyzing the collected data. As a first step we have to: import the libraries for loading the dataset, then we generally explore the data. In it, we can find the data types for each attribute, the content and description of the attributes, as shown in Table I. As a second step we proceed to eliminate any duplicate, missing or erroneous information. Also, outliers

and inconsistencies in the data are checked to ensure that they are accurate and reliable.

Next, the statistical values of the data set are checked for example, the number of students, unique values, frequency, and highest values as presented in Table II.

Performing a simplified analysis of the data set by constructing a general distribution of the characteristics, the following interpretation is reached. The gender variable is balanced, since the same number of males and females is applied to obtain a relevant result. Most of the students correspond to the age of 11-25 years; this is the age at which one speaks with confidence of sustainable adaptation to online learning. The values provided by the level of education indicate that most of the students have only a school education. This data correlates with a given timetable, in which many students are between the ages of 7 and 20. Existing educational services are provided by private institutions. This is due to the development of the educational market. It should also be noted that most of the educational programs are sold by telephone through the 4G network, which shows that people receive education at any place convenient for them.

Next, we analyze the number of students according to their level of adaptation in online education.

TABLE I. GENERAL ANALYSIS OF THE DATA SET

	Student Gender	Age	Educational level	Institution Type	Financial Condition	Internet Type	Device
0	Male	21	University	Non-public	Middle class	Wi-Fi	Tab
1	woman	20	University	Non-public	Middle class	Mobile Data	Mobile
2	woman	16	College	Non-public	Middle class	Wi-Fi	Mobile
3	woman	12	School	Non-public	Middle class	Mobile Data	Mobile
4	woman	15	School	Non-public	Lower class	Wi-Fi	Mobile
...
1201	woman	16	College	Non-public	Middle class	Wi-Fi	Mobile
1202	woman	16	College	Non-public	Middle class	Wi-Fi	Mobile
1203	Male	12	School	Non-public	Middle class	Mobile Data	Mobile
1204	woman	17	College	Non-public	Middle class	Wi-Fi	Mobile
1205	woman	11	College	Non-public	Lower class	Mobile Data	Mobile

TABLE II. DATA SET STATISTICS

	Count	unique	top	freq
Student's gender	1205	2	Male	663
Age range	1205	6	21-25	374
Student's educational level	1205	3	School	530
Type of educational institution	1205	2	private	823
Student's IT knowledge	1205	2	No	901
Student's geographic location	1205	2	Yes	935
Power system interruption	1205	2	Low	1004
Student's economic status	1205	3	Mid	878
Type of mobile / fixed Internet	1205	2	Mobile Data	695
Type of connectivity	1205	3	4G	775
Session duration	1205	3	1-3	840
LMS knowledge	1205	2	No	995
Device used	1205	3	Mobile	1013
Degree of student adaptability	1205	3	Moderate	625

According to the graphical representation presented in Fig. 4, it is observed that men tend to adjust more easily to new knowledge, while the rate of maladaptation is similar for both genders. In addition, it is evident that the optimal age for successful adaptation is between 21 and 25 years, while the worst adaptation occurs in people older than 26 years and in the age range of 16 to 20 years. This information suggests that age may play an important role in the ability to adapt to online education. The lower capacity to adapt to new knowledge can be attributed to various factors, both social and physiological. Importantly, those who belong to the middle class tend to take better advantage of online educational materials. In addition, a greater adaptive capacity has been observed among those students who reside in organized communities or environments, which may be related to economic and social factors. It is also important to consider that the level of adaptation to new knowledge is also influenced by the quality of the available Internet service.

Once the exploratory analysis of the data set has been completed, the next step is to perform the processing of the data set. As a starting point, it consists of dividing the data set in two, 80% will be used for training and 20% for testing. With

`Sklearn.Model_selection.train_test_split()`, it is possible to perform the splitting process by simply specifying the size of the test part for training. As in the exploratory analysis of the data set, nulls or missing values have been discovered. Therefore, it is necessary to remove them from the dataset in order to handle them, estimate those values and then add a mean value to the dataset, for which we use libraries such as: `Simple-Imputer ()` and `impute.fit_transform(x_train, x_test)`. We then proceed to normalize the data. There is no doubt that data normalization is a very important factor, since it consists of reorganizing and scaling each of the features according to the unit variance, so that all features can be compared equally. The following libraries are used for feature normalization: `StandardScaler()`, `tandardScaler.fit_transform(x_train_Impute, x_test_Impute)` and `data_train_normalized()`. Since the data set contains a relatively small number of observations, the SMOTE technique will be used to avoid over-fitting the data set. The objective of this technique is to compare the accuracy of the original data set with the accuracy of the sampled data set. This comparison involves comparing their accuracy. By using the SMOTE technique, class imbalance can be balanced, since class imbalance can lead to asymmetry and overfitting in training.

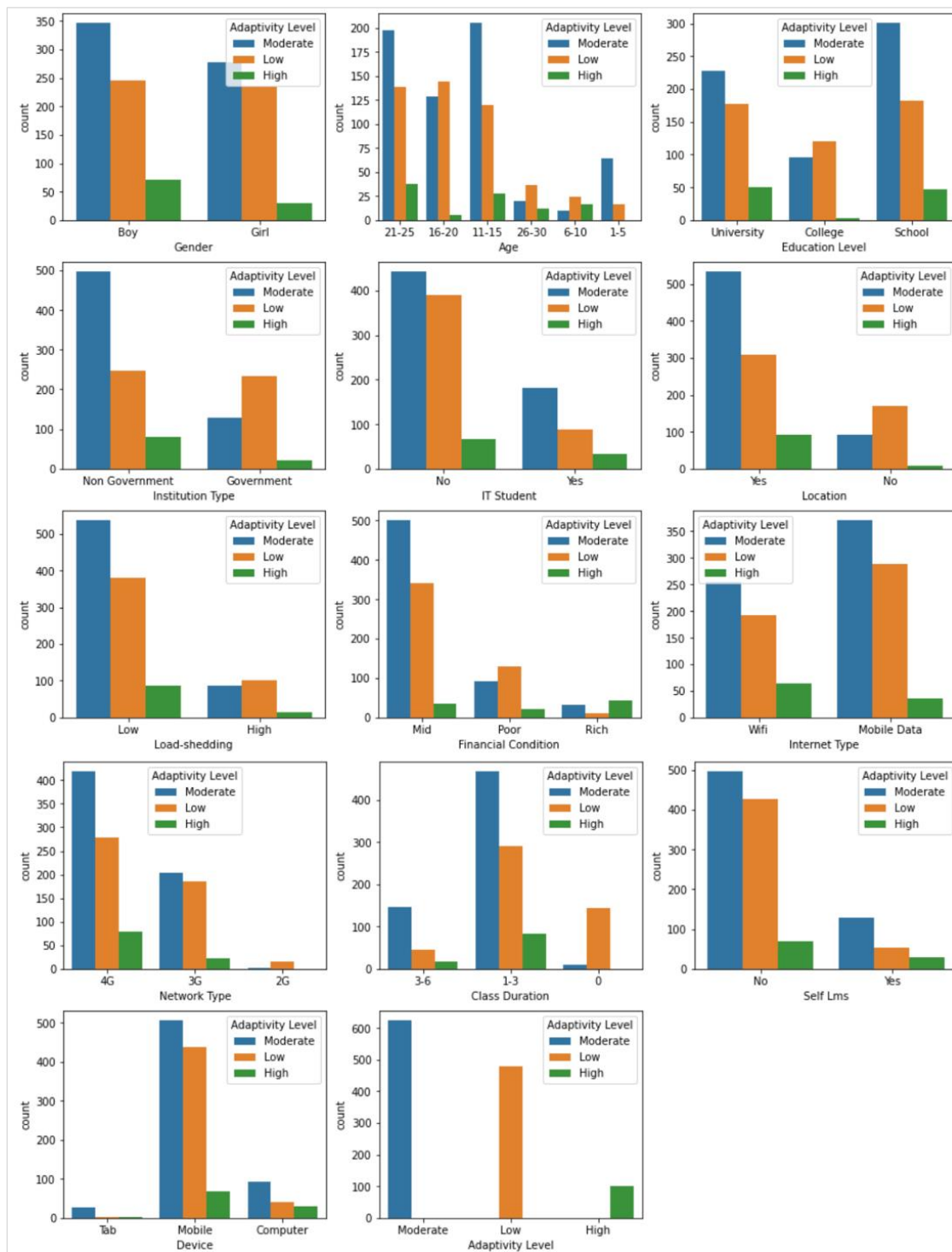


Fig. 4. Analysis of the level of adaptability.

3) *Training and data validation:* Dataset training is a process, in which ML models are fit to a dataset. During this process, the models are fitted to minimize the margin of error between the model predictions and the actual values of the data used in training the model. The purpose of this process is to achieve a model that can generalize well to unseen data, i.e., that can make accurate predictions on new data. For each

model is fitted with SMOTE technique, it means that, one uses the original data set, and the other uses the SMOTE balanced data set. For the parameter selection process, the cross-validation technique was used, which consists of training in two parts, first selecting one part of the data (training) and evaluating it with another part (test set), this process is repeated several times for different combinations. The result is

obtained by promising the results of each experiment, which gives us a more accurate assessment of the predictive ability of the model. Cross-validation is a very useful technique because it allows using all available data to train and evaluate the model, which can increase accuracy and reduce the risk of over-fitting. In addition, cross-validation allows us to compare different ML models and select the best model for our data set. To evaluate the performance of ML models, several measures are employed, which are defined as: a) True Positives (TP), in this work are classified as positive samples which are those students who have moderate level of adaptability to online education; b) True Negatives, in this work are classified as negative samples those students who do not have a level of adaptability to online education; c) False Positives, are the samples that have been misclassified as positive and d) False Negatives, are the samples that have been misclassified as negative, as seen in Fig. 5.

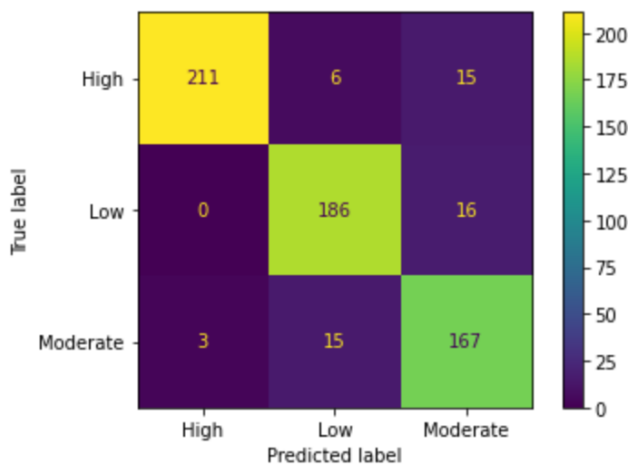


Fig. 5. Variable confusion matrix.

Fig. 5 shows the confusion matrix in general. For example, it can be seen that 211 samples are classified as positive (students with high level of adaptability), 186 samples are classified as positive (have high internet connectivity) and 167 samples are also classified as positive (students with moderate level of adaptability). After training the models, it is important to evaluate their performance to make sure it is accurate. This process follows the following steps: testing the models, in this work 20% of the dataset is used; evaluation metrics are used to assess the performance of the models, accuracy, recall, F1-Score and support, as shown in the following section in Table III.

IV. RESULTS

Training of various ML models, including RF, K-NN, SVM, LR and XGB, was carried out using a specific Kaggle dataset. Subsequently, a learning algorithm was developed to train the models. Subsequently, the performance of each model was evaluated using unobserved data. To carry out this evaluation, various metrics such as accuracy, precision, recall, ROC curve and F1 score were used. The results of these evaluations are presented in Table III.

TABLE III. RESULTS OF TRAINED MODELS

Random Forest				
	Precision (%)	Recall (%)	f1-score (%)	support
0	91	98	94	214
1	93	91	92	207
2	91	85	88	198
accuracy			92	619
macro avg	92	91	91	619
weighted avg	92	92	92	619
KNN classifier				
	Precision (%)	Recall (%)	f1-score (%)	support
0	86	100	93	214
1	86	82	84	207
2	86	73	78	198
accuracy			85	619
macro avg	85	85	85	619
weighted avg	85	85	85	619
Support vector machine				
	Precision (%)	Recall (%)	f1-score (%)	support
0	83	86	85	214
1	83	67	74	207
2	65	76	70	198
accuracy			76	619
macro avg	77	76	76	619
weighted avg	77	76	76	619
Logistic Regression				
	Precision (%)	Recall (%)	f1-score (%)	support
0	82	70	76	214
1	71	63	66	207
2	54	69	61	198
accuracy			67	619
macro avg	69	67	68	619
weighted avg	69	67	68	619
xGBClassifier				
	Precision (%)	Recall (%)	f1-score (%)	support
0	91	100	95	214
1	92	91	92	207
2	93	84	88	198
accuracy			92	619
macro avg	92	92	92	619
weighted avg	92	92	92	619

In this study, a standardized procedure has been carried out to calculate the true positive (TP) and false positive (FP) rate of all trained models. To illustrate this process, in the case of the RF model, the TP rate is obtained by dividing the number of true positives by the total number of positive cases. The FP rate, on the other hand, is obtained by dividing the number of false positives by the total number of negative cases. The same formula is applied for the other models (K-NN, SVM, LR, XGB). These rates are important metrics to evaluate the performance of a classification model and is applied to adjust the classification threshold of a model. For example, if you want to increase the rate of true positives, you can decrease the classification threshold of the model so that it predicts more samples as positive. However, this could also increase the FP rate. Therefore, it is important to find an appropriate balance between TP and FP. In this case, the RF, K-NN, SVM, LR and XGB models were trained, and the following performances were obtained: 91%, 85%, 76%, 67% and 92%, respectively.

Analyzing Table III, the RF and XGBClassifier models achieved the best rates in the f1-score, recall and accuracy metrics, in this order are analyzed. RF, 92%, 92%, 92%; XGBClassifier, 92%, 92% and 92%, although it is true that in this case, we analyze it at the average level. Therefore, it can be affirmed that the two models present optimal performances to predict the level of adaptability in online education; the second model to obtain better metrics is KNN with the following rates: 85% in F1-score, 85% in recall and 85% in precision; it is followed by the SVM model with 76% of F1-Score, 76% in recall and 77% of precision. Finally, the LR model achieved 68% in F1-Score, 67% in recall and 69% in precision. To reach a better interpretation of the results, it was decided to analyze the importance of each characteristic of the data set. For this purpose, the feature_importances technique was used to determine which features are the most important for the model. The calculation is based on how much each feature contributes to the model in terms of impurity reduction at the node of the tree where it is located. Therefore, the feature that obtains the highest score is considered the most important for the model. This is shown in Table IV.

TABLE IV. IMPORTANT CHARACTERISTICS OF THE DATA SET

#	Features	Feature importances
7	Financial Condition	16.35 %
1	Age	13.72 %
10	Class Duration	12.69 %
0	Gender	8.56 %
9	Network Type	7.54 %
3	Institution Type	6.08 %
2	Education Level	6.08 %
5	Location	5.59 %
8	Internet Type	5.54 %
12	Device	4.96 %
11	Self Lms	4.82 %
4	IT Student	4.37 %
6	Load-shedding	3.64 %

V. DISCUSSION

In the last three years, online education has experienced tremendous growth and has become a popular form of learning due to the COVID-19 pandemic. However, not all students have the same ability to adapt to this type of education. Some students may have more difficulty adapting to online education due to the lack of face-to-face interaction with the professor and their classmates, the need to be more autonomous and organized in their own study time. Predicting the level of adaptability of students in online education is an important factor for online educators, as it allows them to identify and support students who may have difficulties adapting to this type of education. ML methods are a useful tool for predicting the level of student adaptability in online education. Table IV shows that the variable economic condition, which represents the most important factor for student adaptability in online education, is followed in importance by the following variables: age, duration of sessions, type of network used, type of institution, level of education, geographic location, type of Internet, among other variables. ML models such as RF, KNN, LR, SVM, and XGB were used to analyze the data and rank

which of the models is optimal for predicting the level of student adaptability in online education. These ML models provided valuable information about which variables are the most important for students' adaptability in online education and how they are related to each other, as shown in Table IV. For example, Economic condition is the most important variable for adaptability, which represents 16.35% in importance, which is related to the results obtained in the study [20], where they concluded that student's financial condition plays a very important role in student's adaptability in online education. Similarly, the results of this work are related to [21], where they explored the gaps that exist in online and face-to-face education, and concluded that the variable of economic condition, age and Internet connection are determining factors for adaptability to online education. The results of the training with the predictive models showed that the RF (92%) and XGB (92%) model achieved the best rates for predicting the level of student adaptability in online education, which is related to the results obtained in the work [24], where they used ML classifiers and deep neural networks, and concluded that the SVM classifiers and the Decision tree, achieved the best results in accuracy and performance. AI has contributed significantly in the academic field, so much so that, unlike this work, in [23] and [25] examined perspectives, advantages and problems of AI in online education, for which they used different ML algorithms, and concluded that RF is the something-rhythm that achieved the best metrics, with an accuracy of 93%, therefore, it is the best suited for this type of tasks, this result is superior to that achieved in this work. The results will depend on the nature of the research and the volume of the data set. In the same line, in [26] they used IA to determine the influence on the adaptability of students in online education, using variables such as: teacher-student, parents and society, age, connectivity, among others. The results showed that AI has a positive influence on the adaptability of students in online education. Unlike the ML models used in this work to predict the level of adaptability of students in online education, in [28], they used the deep learning LSTM technique to analyze the performance of students in online education, where they obtained that deep learning techniques are a very good option to predict the level of adaptability, reaching an accuracy of 84. Predicting the adaptability level of students in online education using ML methods is a valuable tool for online educators. It can provide important information about which variables are most important for student adaptability in online education and how they relate to each other. In addition, it can be used to develop personalized support strategies for students who may have difficulty adapting to this type of education.

VI. CONCLUSIONS

After performing training and comparison of ML (RF, KNN, VSM, LR and XGB) predictive models for predicting the Adaptability Level of students in online education, using the Kaggle dataset consisting of 14 attributes, it reaches the following conclusions.

It was determined that the RF and XGBClassifier models obtained the best results in accuracy and performance. Therefore, for predicting the level of adaptability of students in online education, they are the best predictors, which suggests that these models can be useful for improving the effectiveness

of online education. Without detracting from the other models, they also obtained excellent performance results. The models used in this study demonstrated high prediction accuracy.

It was also found that the variables used in the set of variables are important for the accuracy of the predictions. Table IV shows the relevance of each of the variables for predicting the level of adaptability of students in online education, the most important being economic condition, age, duration of sessions, gender, type of network, type of institution, directly influencing the level of adaptability of students.

Finally, predictive ML models can be a valuable tool for predicting the level of student adaptability in online education. The use of ML models, such as RF, KNN, VSM, LR, and XGB, and the careful selection of variables can significantly improve the accuracy of predictions. These findings contribute in a significant way for online education and can help educators improve the effectiveness of their online educational programs. In the future, a possible development that would complement the use of the models would be the development of work to assess the academic performance of students in online education.

REFERENCES

- [1] R. E. Baticulon et al., "Barriers to Online Learning in the Time of COVID-19: A National Survey of Medical Students in the Philippines," *Med Sci Educ*, vol. 31, no. 2, pp. 615–626, Apr. 2021, doi: 10.1007/S40670-021-01231-Z.
- [2] UNESCO, "List of Geoparks & Regional Networks," 2023. <https://en.unesco.org/global-geoparks/list> (accessed Mar. 08, 2023).
- [3] U. Nations, "Human Development Report 2021-22," *Human Development Reports*, 2022, Accessed: Mar. 08, 2023. [Online]. Available: <https://hdr.undp.org/content/human-development-report-2021-22>
- [4] M. M. H. Suzan, N. A. Samrin, A. A. Biswas, and M. A. Pramanik, "Students' Adaptability Level Prediction in Online Education using Machine Learning Approaches," 2021 12th International Conference on Computing Communication and Networking Technologies, ICCCNT 2021, 2021, doi: 10.1109/ICCCNT51525.2021.9579741.
- [5] A. Jennifer G., M. George Thomas, and R. Vijay Solomon, "Does Virtual Titration Experiment Meet Students' Expectation? Inside Out from Indian Context," *J Chem Educ*, vol. 99, no. 3, pp. 1280–1286, Mar. 2022, doi: 10.1021/ACS.JCHEMED.1C01034.
- [6] A. Besser, G. L. Flett, and V. Zeigler-Hill, "Adaptability to a Sudden Transition to Online Learning During the COVID-19 Pandemic: Understanding the Challenges for Students," *Scholarsh Teach Learn Psychol*, vol. 8, no. 2, pp. 85–105, 2022, doi: 10.1037/STL0000198.
- [7] L. Yazel, C. Bishop, and H. Britt, "Adapting During the Ever-Changing Pandemic Environment: A One-Year Examination of How Health Education Specialists Remain Adaptable," *Am J Health Educ*, vol. 53, no. 3, pp. 142–148, 2022, doi: 10.1080/19325037.2022.2048746.
- [8] Y. Zhang, G. Zhao, and B. Zhou, "Does learning longer improve student achievement? Evidence from online education of graduating students in a high school during COVID-19 period," *China Economic Review*, vol. 70, p. 101691, Dec. 2021, doi: 10.1016/J.CHIECO.2021.101691.
- [9] C. Guo and B. Wan, "The digital divide in online learning in China during the COVID-19 pandemic," *Technol Soc*, vol. 71, p. 102122, Nov. 2022, doi: 10.1016/J.TECHSOC.2022.102122.
- [10] E. M. Azila-Gbettor, M. K. Abiemo, and S. N. Glate, "University support and online learning engagement during the Covid-19 period: The role of student vitality," *Heliyon*, vol. 9, no. 1, p. e12832, Jan. 2023, doi: 10.1016/J.HELIYON.2023.E12832.
- [11] F. Li et al., "Perceptions towards online learning among medical students during the COVID-19 pandemic," *Heliyon*, vol. 9, no. 2, p. e13119, Feb. 2023, doi: 10.1016/J.HELIYON.2023.E13119.
- [12] C. Tarchi, E. W. Brante, M. Jokar, and E. Manzari, "Pre-service teachers' conceptions of online learning in emergency distance education: How is it defined and what self-regulated learning skills are associated with it?," *Teach Teach Educ*, vol. 113, p. 103669, May 2022, doi: 10.1016/J.TATE.2022.103669.
- [13] M. Bower, "eLearning Statistics: 2023 Data, Trends & Predictions," *Design of Technology-Enhanced Learning*, pp. 261–304, Aug. 2023, doi: 10.1108/978-1-78714-182-720171011.
- [14] Y. Ma, R. Zhang, and Y. Zhang, "Visualization analysis of junior school students' pubertal timing and social adaptability using data mining approaches," *Heliyon*, vol. 8, no. 8, p. e10443, Aug. 2022, doi: 10.1016/J.HELIYON.2022.E10443.
- [15] M. Gamboa-Ramos, R. Gómez-Noa, O. Iparraguirre-Villanueva, M. Cabanillas-Carbonell, and J. L. H. Salazar, "Mobile Application with Augmented Reality to Improve Learning in Science and Technology," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 10, pp. 487–492, 2021, doi: 10.14569/IJACSA.2021.0121055.
- [16] A. A. Biswas, A. Majumder, M. J. Mia, I. Nowrin, and N. A. Ritu, "Predicting the enrollment and dropout of students in the post-graduation degree using machine learning classifier," *International Journal of Innovative Technology and Exploring Engineering*, vol. 8, no. 11, pp. 3083–3088, Sep. 2019, doi: 10.35940/IJITEE.K2435.0981119.
- [17] A. Qazi, N. Hasan, C. M. Owusu-Ansah, G. Hardaker, S. K. Dey, and K. Haruna, "SentiTAM: Sentiments centered integrated framework for mobile learning adaptability in higher education," *Heliyon*, vol. 9, no. 1, p. e12705, Jan. 2023, doi: 10.1016/J.HELIYON.2022.E12705.
- [18] A. Morgan, R. Sibson, and D. Jackson, "Digital demand and digital deficit: conceptualising digital literacy and gauging proficiency among higher education students," *Journal of Higher Education Policy and Management*, vol. 44, no. 3, pp. 258–275, 2022, doi: 10.1080/1360080X.2022.2030275.
- [19] J. C. Bricout, "Making computer-mediated education responsive to the accommodation needs of students with disabilities," *J Soc Work Educ*, vol. 37, no. 2, pp. 267–281, 2001, doi: 10.1080/10437797.2001.10779053.
- [20] A. J. Martin, R. J. Collie, and R. P. Nagy, "Adaptability and High School Students' Online Learning During COVID-19: A Job Demands-Resources Perspective," *Front Psychol*, vol. 12, p. 3181, Aug. 2021, doi: 10.3389/FPSYG.2021.702163/BIBTEX.
- [21] M. M. H. Suzan, N. A. Samrin, A. A. Biswas, and M. A. Pramanik, "Students' Adaptability Level Prediction in Online Education using Machine Learning Approaches," 2021 12th International Conference on Computing Communication and Networking Technologies, ICCCNT 2021, 2021, doi: 10.1109/ICCCNT51525.2021.9579741.
- [22] J. König, D. J. Jäger-Biela, and N. Glutsch, "Adapting to online teaching during COVID-19 school closure: teacher education and teacher competence effects among early career teachers in Germany," <https://doi.org/10.1080/02619768.2020.1809650>, vol. 43, no. 4, pp. 608–622, Aug. 2020, doi: 10.1080/02619768.2020.1809650.
- [23] M. Ezz and A. Elshenawy, "Adaptive recommendation system using machine learning algorithms for predicting student's best academic program," *Educ Inf Technol (Dordr)*, vol. 25, no. 4, pp. 2733–2746, Jul. 2020, doi: 10.1007/S10639-019-10049-7/METRICS.
- [24] X. Ma, Y. Yang, and Z. Zhou, "Using machine learning algorithm to predict student pass rates in online education," *ACM International Conference Proceeding Series*, pp. 156–161, Apr. 2018, doi: 10.1145/3220162.3220188.
- [25] R. G. Tiwari, A. Misra, V. Kukreja, A. K. Jain, and N. Ujjwal, "Education 4.0: Classification of Student Adaptability Level in E-Education," 2022 10th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions), ICRITO 2022, 2022, doi: 10.1109/ICRITO56286.2022.9964851.
- [26] C. Xie et al., "Influence of Artificial Intelligence in Education on Adolescents' Social Adaptability: A Machine Learning Study," *International Journal of Environmental Research and Public Health*

- 2022, Vol. 19, Page 7890, vol. 19, no. 13, p. 7890, Jun. 2022, doi: 10.3390/IJERPH19137890.
- [27] R. D. Costa, G. F. Souza, R. A. M. Valentim, and T. B. Castro, "The theory of learning styles applied to distance learning," *Cogn Syst Res*, vol. 64, pp. 134–145, Dec. 2020, doi: 10.1016/J.COGSYS.2020.08.004.
- [28] H. Waheed, S. U. Hassan, R. Nawaz, N. R. Aljohani, G. Chen, and D. Gasevic, "Early prediction of learners at risk in self-paced education: A neural network approach," *Expert Syst Appl*, vol. 213, p. 118868, Mar. 2023, doi: 10.1016/J.ESWA.2022.118868.
- [29] P. Palimkar, R. N. Shaw, and A. Ghosh, "Machine Learning Technique to Prognosis Diabetes Disease: Random Forest Classifier Approach," *Lecture Notes in Networks and Systems*, vol. 218, pp. 219–244, 2022, doi: 10.1007/978-981-16-2164-2_19/COVER.
- [30] O. Iparraguirre-Villanueva, A. Epifanía-Huerta, C. Torres-Ceclén, J. Ruiz-Alvarado, M. Cabanillas-Carbonell, and N. Wiener, "Breast Cancer Prediction using Machine Learning Models," *IJACSA International Journal of Advanced Computer Science and Applications*, vol. 14, no. 2, p. 2023, Accessed: Mar. 14, 2023. [Online]. Available: www.ijacsa.thesai.org
- [31] X. Zhang, H. Xiao, R. Gao, H. Zhang, and Y. Wang, "K-nearest neighbors rule combining prototype selection and local feature weighting for classification," *Knowl Based Syst*, vol. 243, p. 108451, May 2022, doi: 10.1016/J.KNOSYS.2022.108451.
- [32] L. Sun, J. Zhang, W. Ding, and J. Xu, "Feature reduction for imbalanced data classification using similarity-based feature clustering with adaptive weighted K-nearest neighbors," *Inf Sci (N Y)*, vol. 593, pp. 591–613, May 2022, doi: 10.1016/J.INS.2022.02.004.
- [33] M. Tanveer, T. Rajani, R. Rastogi, Y. H. Shao, and M. A. Ganaie, "Comprehensive review on twin support vector machines," *Ann Oper Res*, pp. 1–46, Mar. 2022, doi: 10.1007/S10479-022-04575-W/METRICS.
- [34] A. B. YILMAZ, Y. S. TASPINAR, and M. Koklu, "Classification of Malicious Android Applications Using Naive Bayes and Support Vector Machine Algorithms," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 10, no. 2, pp. 269–274, May 2022, Accessed: Mar. 14, 2023. [Online]. Available: <https://www.ijisae.org/index.php/IJISAE/article/view/2010>
- [35] Joseph Redshaw, D. S. J. Ting, Alex Brown, J. D. Hirst, and Thomas Gärtner, "Krein support vector machine classification of antimicrobial peptides," *Digital Discovery*, 2023, doi: 10.1039/D3DD00004D.
- [36] Y. Hu, Y. Fan, Y. Song, and M. Li, "A general robust low-rank multinomial logistic regression for corrupted matrix data classification," *Applied Intelligence*, pp. 1–17, Feb. 2023, doi: 10.1007/S10489-022-04424-0/METRICS.
- [37] M. Zivkovic et al., "Training Logistic Regression Model by Hybridized Multi-verse Optimizer for Spam Email Classification," pp. 507–520, 2023, doi: 10.1007/978-981-19-6634-7_35.
- [38] S. R. Darawsheh, A. S. Al-Shaar, F. A. Haziemeh, and M. T. Alshurideh, "Classification Thyroid Disease Using Multinomial Logistic Regressions (LR)," pp. 645–659, 2023, doi: 10.1007/978-3-031-12382-5_34.
- [39] H. Nguyen and N. D. Hoang, "Computer vision-based classification of concrete spall severity using metaheuristic-optimized Extreme Gradient Boosting Machine and Deep Convolutional Neural Network," *Autom Constr*, vol. 140, p. 104371, Aug. 2022, doi: 10.1016/J.AUTCON.2022.104371.
- [40] T. Thenmozhi and R. Helen, "Feature Selection Using Extreme Gradient Boosting Bayesian Optimization to upgrade the Classification Performance of Motor Imagery signals for BCI," *J Neurosci Methods*, vol. 366, p. 109425, Jan. 2022, doi: 10.1016/J.JNEUMETH.2021.109425.
- [41] A. Ramón, A. M. Torres, J. Milara, J. Cascón, P. Blasco, and J. Mateo, "eXtreme Gradient Boosting-based method to classify patients with COVID-19," *Journal of Investigative Medicine*, vol. 70, no. 7, pp. 1472–1480, Oct. 2022, doi: 10.1136/JIM-2021-002278.

Analyzing WhisperGate and BlackCat Malware: Methodology and Threat Perspective

Mathew Nicho¹, Rajesh Yadav², Digvijay Singh³

Research and Innovation Centre, Rabdan Academy

Abu Dhabi, United Arab Emirates¹

Department of Computer Science and Engineering, GITAM Univeraity

Visakhapatnam, India²

Department of Computer Science and Engineering, BML Munjal University

Gurgaon, India³

Abstract—The increasing use of powerful evasive ransomware malware in cyber warfare and targeted attacks is a persistent and growing challenge for nations, corporations, and small and medium-sized enterprises. This threat is evidenced by the emergence of the WhisperGate malware in cyber warfare, which targets organizations in Ukraine to render targeted devices inoperable, and the BlackCat malware, which targets large organizations by encrypting files. This paper outlines a practical approach to malware analysis using WhisperGate and BlackCat malware as samples. It subjects them to heuristic-based analysis techniques, including a combination of static, dynamic, hybrid, and memory analysis. Specifically, 12 tools and techniques were selected and deployed to reveal the malware’s innovative stealth and evasion capabilities. This methodology shows what techniques can be applied to analyze critical malware and differentiate samples that are variations of known threats. The paper presents currently available tools and their underlying approaches to performing automated dynamic analysis on potentially malicious software. The study thus demonstrates a practical approach to carrying out malware analysis to understand cybercriminals’ behavior, techniques, and tactics.

Keywords—Malware analysis; WhisperGate; BlackCat; malware sample; ransomware

I. INTRODUCTION

The geopolitical events in Ukraine at the start of 2022 were preceded by the devastating cyber warfare operation highlighted by WhisperGate malware (A malware that corrupts a system’s master boot record, displays a fake ransomware note, and encrypts files based on certain file extensions). WhisperGate is considered dangerous because it can launch cyber-attacks and compromise sensitive information against hardened targets. Since it was deployed in cyber warfare against Ukraine, it could exploit unknown vulnerabilities in a target’s security systems and cause significant harm. The destructive capabilities of WhisperGate make it a threat to individual, organizational, and national security. At the end of 2021, a sophisticated malware called BlackCat also known as “AlphaV,” emerged, targeting U.S. organizations and their affiliates in Europe, the Philippines, and other locations. While WhisperGate masquerades as ransomware targeting nation-states (in this case, Ukraine), BlackCat has emerged as deadly ransomware targeting U.S. and European retail, construction,

and transportation organizations. BlackCat appeared as an innovative ransomware-as-a-service (RaaS) group leveraging the Rust programming language and offering affiliates 80% to 90% of ransom payments [1]. Affiliates included Germany’s tank storage and terminal firm Oiltanking and energy firm Mabatnaft, Belgian energy firm Sea-Invest, and Dutch oil and gas firm Evos. These attacks underlined the growing vulnerability of critical infrastructure companies to malicious hackers [2]. Both pieces of malware were challenging to defend against due to their elusive and evasive nature, which intrigued cybersecurity analysts worldwide. Since the top three cyberattacks that organizations are most concerned about are ransomware, social engineering, and malicious insider activities [3], WhisperGate and BlackCat were ideal candidates for our practical malware analysis approach due to their stealth and evasive capabilities and the destructive consequences they can cause.

Multiple malware classes, such as worms, viruses, spyware, Trojan horses, rootkits, ransomware, keyloggers, and adware, are designed with specific functionalities namely data exfiltration, data encryption, and data destruction. Despite the widespread use of antimalware software, the number of malware infections continues to grow. Malware, especially zero-day malware, can evade antimalware solutions and even infect them with its built-in defensive mechanisms. Along with WhisperGate, malware deployed against Ukraine included HermeticWiper, IsaacWiper, HermeticWizard, and CaddyWiper. Once inside the initial network, it leverages that access to compromise user and administrator accounts in the active directory of Windows’ server and configures malicious group policy objects through Windows’ task scheduler [4].

This paper provides a practical approach to performing malware analysis using integrated tools and techniques to assess WhisperGate and BlackCat. The Microsoft Threat Intelligence Centre disclosed that WhisperGate, categorized as a wiper, targeted several organizations in Ukraine and was tracked as DEV-0586 with a design similar to ransomware but lacking a recovery mechanism [5]. BlackCat, which belongs to a sophisticated ransomware as a service (RaaS) family, extorts money from targeted institutions instead.

The structure of the remainder of this paper is as follows: Section II discusses extant literature on ransomware in general, the identified ransomware, and an evaluation of malware analysis techniques. Section III outlines the methodology, and Section IV focuses on the experimental analysis. Section V discusses the results, and Section VI concludes the research with suggestions on future research.

II. LITERATURE REVIEW

The section provides an overview of ransomware, discusses two specific types of malware (WhisperGate and BlackCat), and then evaluates malware analysis methodologies to focus on the appropriate technique(s).

A. Ransomware

Ransomware is considered one of the most threatening types of malware. Its attacks increased by 151% in 2012, averaging 270 cyberattacks per organization, with each successful breach resulting in a cost of \$3.6 million for the affected company [3]. Cyber-attacks predominantly occur through ransomware, social engineering, and malicious insider activity [3]. In particular, ransomware leverages social engineering methods to gain unauthorized access to the victim's network. Once an infection is spread, the user is extorted and asked for a monetary payment against the locked access [6], but there is no guarantee that they will regain access to their locked files after paying the ransom. Threat actors often receive the payment but still retain the data. These cybercriminals often request payment in cryptocurrency, as it is untraceable and allows them to evade responsibility [7]. Malwares leverage the Trojan by disguising themselves as legitimate software and download the malicious components, which negatively impact the system and tend to infect files and target other systems [8]. While commercial solutions are available, these are not 100% secure, because hackers use more sophisticated techniques to follow the evolution and bypass the protection techniques [9]. WhisperGate is classified as a wiper, i.e., it disguises itself as ransomware but instead aims to cause mass destruction by wiping out hard drives at targeted organizations [10].

Removing the ransomware or restoring the infected devices is ineffective, as the ransomware uses asymmetric cryptography [11], which makes it robust. The encryption makes it so that the victim is unable to access the data without first decrypting it using a key [12]. Threat actors usually ask for a ransom in exchange for the decryption key and target organizations that handle large amounts of sensitive data. The victim is faced with inaccessibility and damage to their data and often pays the ransom demand. Since most of the victims are threatened with their data and sensitive information being exposed [13].

Among the five types of ransomware—locker, crypto leakware, scareware, and pseudo-ransomware [14], WhisperGate comes under the pseudo-ransomware category, while BlackCat comes under leakware category. Also known as doxware, leakware presents a high-risk level because it is well-targeted to institutions such as banks or those that work with confidential and critical data. This ransomware does not destroy the data but threatens to release them into the public

domain. Furthermore, since the context can damage the institution's image, an even greater emphasis can be placed on the quick payment of a ransom. Accordingly, BlackCat operates as a RaaS option that permits earning a percentage of the ransom payment to all the persons who have low technical knowledge about how to create ransomware but are members of this network. It is only necessary that those members spread the ransomware as far as possible while the RaaS vendor can focus on how to make this malicious software cause even more damage.

B. WhisperGate

Unlike traditional ransomware campaigns where the motive is clear, the BlackCat campaign is believed to be pseudo, with its intention being to cause the destruction of infected systems, as evidenced by the Stage 4 wiper that overwrites data on the victim's system, making decryption impossible [15]. The malware that was explicitly launched against various Ukrainian organizations in geopolitically motivated attacks was first analyzed by the Microsoft Threat Intelligence Center and detected on January 13, 2022 [16]. The BlackCat ransomware campaign targeted Ukraine in 2021 prior to its physical invasion, but it was detected and neutralized before causing any severe damage [17]. Russian cyber operations have targeted Ukraine with destabilization efforts for years through attacks on critical infrastructure, influence operations, website defacement, and attacks against banks and military networks [16]. WhisperGate, while masquerading as ransomware, corrupts a system's master boot record, displays a fake ransomware note, and then encrypts files based on specific file extensions. While a ransomware message is displayed during the attack, the targeted data is destroyed and is not recoverable even if a ransom is paid [18]. The multi-stage infection chain downloads a payload that wipes the master boot record (MBR). Then, it downloads a malicious Dynamic-link library (DLL) file hosted on a discord server (a platform where people can interact with each other in real time), which drops and executes another wiper payload that destroys files on the infected machines [19]. The malware, which is designed to look like ransomware, is intended to render the targeted devices inoperable rather than to obtain a ransom, as it does not have an inbuilt recovery code. Using social engineering methods in an Advanced Persistent Threat (APT) campaign, the attackers might have used stolen credentials and likely had access to the victim's network for months before the attack [19]. The malware can also extend to extranet networks. The recommendations from the US Cybersecurity and Infrastructure Security Agency that organizations with ties to Ukraine should carefully consider how to isolate and monitor those connections to protect themselves from potential collateral damage are echoed.

Following the detection of WhisperGate, HermeticWiper, another similar malware masquerading as ransomware used against organizations in Ukraine, was discovered on February 23, 2022. The malware targets Windows devices, manipulating the master boot record and resulting in subsequent boot failure [18]. Since both these pieces of malware (WhisperGate and HermeticWiper) are similar, the WhisperGate malware was selected as an example to study. The diamond model of intrusion analysis (DMIA) illustrates (Fig. 1) the four

dimensions of the malware attack: the adversary profile, the affected infrastructure, the deployed capabilities, and the target. Specifically, the adversary deploys a capability over a specific infrastructure against a victim [20].

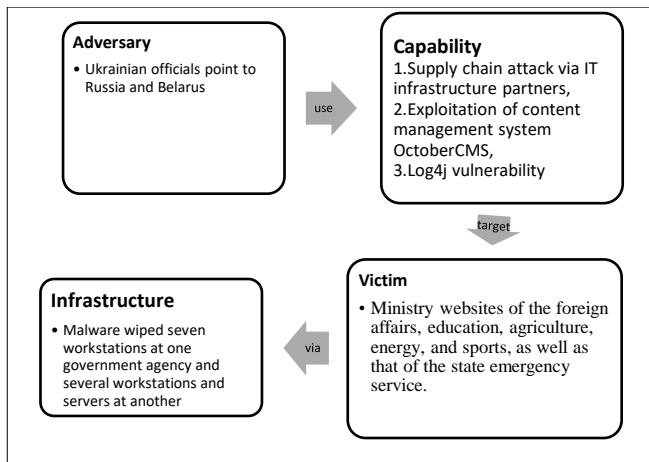


Fig. 1. DMIA model of WhisperGate malware attack.

C. BlackCat

BlackCat is a sophisticated and innovative ransomware family that surfaced in mid-November 2021. It operates as a RaaS business model, and it gained notoriety for soliciting affiliates in known cybercrime forums and offering them to leverage the ransomware and keep 80%–90% of the ransom payment [21]. BlackCat made headlines as one of the first ransomware families written in the Rust programming language, which is used to evade detection by conventional security solutions that may struggle to analyze and parse binaries written in Rust [5].

The rise of cybercrimes has been fueled by the anonymity and non-reversibility of cryptocurrencies, particularly Bitcoin, which makes ransomware payments simple for victims and risk-free for ransomware operators. The trend towards using cryptocurrencies such as Monero, which offers improved security, privacy, and anonymity, is growing, as Monero transactions cannot be traced back to a specific user or address, and the transaction history is kept private. Nonetheless, Bitcoin remains the most popular payment method for ransomware [14]. Among the 31 Ransomware listed by Unit 42, BlackCat has only the seventh largest number of victims listed on their leak site. However, while Lockbit 2.0 ransomware has a list of 50 victims over a period of six months, BlackCat has had an impressive record of 12 victims in just one month since its emergence tanner [21], which makes it a suitable candidate for analysis. In some cases, BlackCat operators use triple extortion by threatening to perform a Distributed Denial of Service (DDoS) attack on the victims' infrastructure: if the ransom is not paid, leak the information along with the data encryption [21]. The DMIA model illustrates the attack process (Fig. 2).

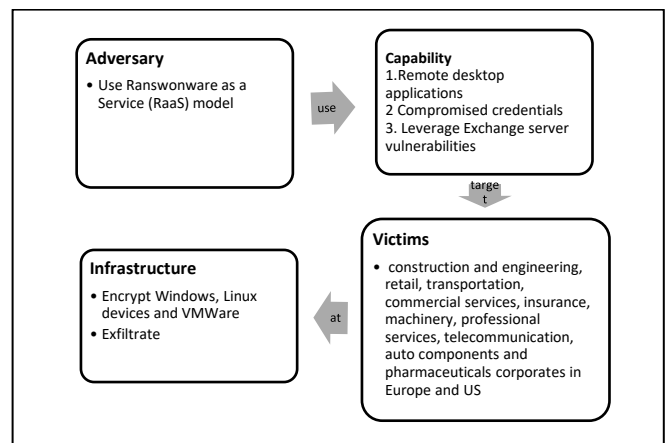


Fig. 2. DMIA model of a BlackCat malware attack.

D. Malware Analysis

Malware analysis applies program analysis and network analysis techniques to understand the behavior and evolution of malicious samples over time [22] and estimate the level of threat and harm a file can cause. Additionally, this kind of analysis helps identify a malicious file's purpose, origin, process execution, file monitoring, and hidden indicators [23].

Malware analysis from a heuristic-based detection (also known as anomaly or behavior based) consists of four types, namely, static, dynamic, hybrid and memory analysis [24]. Being heuristic-based, the proposed research focus on the static, dynamic, hybrid and memory analysis. First, static analysis was used to examine malware samples without the file's execution to extract necessary information from a suspicious file, which assisted us in classifying and identifying its execution. This information is usually gathered using static analysis tools, which examine the sample code more effectively [25]. Static analysis assists in the discovery of the binary code, which contains very useful information about the malicious behavior of a program in the form of op-code sequences, functions, and parameters. However, this method alone may not suffice for a zero-day malware (WhisperGate and BlackCat) before its discovery because new pieces of malware are created daily. The signature-based detection approach followed by the static analysis method requires frequent updates of the virus signature database, which is the method's main disadvantage [26].

Dynamic analysis is deployed, since hackers use various techniques, such as code obfuscation, dynamic code loading, encryption, and packing, to evade static analysis (including signature-based antivirus tools). Furthermore, dynamic analysis can help understand the analyzed file, thus improving detection capabilities [27]. In dynamic malware analysis, the suspicious files are executed and monitored in a controlled environment

[24]. Dynamic analysis includes function call monitoring, network simulation, and registry and file changes. Interactive behavioral properties are observed and analyzed after the simulation of malware. When malware is executed in a dynamic environment, it changes its behaviors. Therefore, static features can be extracted easily and correctly. Hence, the extraction of static features in a dynamic environment detects malware efficiently. The accuracy of dynamic malware analysis alone may instead not be efficient due to the malware's intelligent behaviors [28].

While both static and dynamic analysis techniques are effective on their own, in specific situations, an integrated technique combining the relative merits of each is more efficient. Hybrid analysis that combines both static and dynamic malware analysis is thus generally preferred [29]. Memory analysis, which is used in both malware analysis and malware forensics, involves both the acquisition and analysis phase, thus providing a more comprehensive view of the malware than static and dynamic analyses and an excellent way to analyze memory by preserving a system's contents [29]. Since malware can hide its code in the computer system effectively, it must execute its code in the memory to perform its tasks [24]. Therefore, based on the evaluation of static, dynamic, hybrid, and memory analysis, and having reviewed the efficiency and effectiveness of each of these approaches, the proposed research thus focusses on all four heuristic methods.

E. Methodologies Deployed in Malware Analysis

A survey of extant research presented relevant methods used for malware analysis, namely Eureka, disassembled code analyzer for malware (DCAM), malware analysis reverse engineering (MARE), and systematic approach to malware analysis (SAMA). Eureka, a framework allowing a static analysis of malware binaries, highlights the need to produce unpacked code. It provides Windows application programming interface (API) resolution to identify the system calls in the unpacked code [30]. Or-Meir et al. conducted an overview of existing dynamic analysis methods and provided a malware classification based on each category's behavior, mapping layouts, techniques, and flow comprising memory forensics using volatile tools [27]. Almashhadani et al. used the Lock family of crypto-ransomware as their case study for their comprehensive behavioral analysis (BA) of crypto-ransomware [31]. Their work assisted us in the malware analysis of BlackCat, as the latter showed similarities with crypto-ransomware. Ren et al. provided a three-level ransomware detection and prevention mechanism using virtual machines on Petya and NotPetya ransomware [32]. Similar to WhisperGate in terms of its behavior, NotPetya falls under the category of a wiper disguised as ransomware. Hence, its analysis assisted us in the analysis on WhisperGate.

DCAM is a static malware detection technique using code disassembly to recognize malware variants based on a common core signature with promising results on a set of malware [33]. MARE introduced a four-stage approach covering a structured analysis process that focuses on producing an objective outcome to detect malware followed by isolation and extraction phases, as shown by [34], who introduced the malware behavioral technique, malware reverse engineering,

and code analysis. The author in [35] proposed an automated analysis framework to analyze executable behaviors through a synergic combination of malware detection techniques, including using a virtual machine over a sandbox to enhance invisibility. SAMA provides detailed information on the working of malware, and its applicability over any type of malware makes it robust. It follows a four-stage approach, namely, an modified version of MARE, as shown by [36]. The authors pointed to the execution order provided by MARE and noted that code analysis must be executed along with behavioral analysis.

SAMA is a complete methodology for performing malware analysis, and malware analysts have used it to analyze the following malware threats: Stuxnet, Dark Comet, Poison Ivy, Locky, Careto, and Sofacy Carberp, including Flame and Red October, as shown by [36]. However, the authors did not explain how they used their tools for each step. Furthermore, through the use of SAMA, the authors have only partially discussed memory analysis. Additionally, the stage of packaging obfuscation is executed after the initial five steps of classification in SAMA. However, it would be more impactful to include obfuscation checking before any other step because the analysis could lead to incorrect results. Finally, a hybrid technique should be included as part of the methodology to perform an in-depth analysis, which is missing in SAMA. Accordingly, hybrid analysis was performed to obtain relevant information and fast results to assess WhisperGate and BlackCat malware. The proposed approach is illustrated against those presented in the literature in Table I. Specifically, the lab setup process and static, dynamic, code, and memory analysis is compared.

TABLE I. COMPARATIVE ANALYSIS OF ANALYSIS METHODOLOGIES

Research Methods	Lab Set-up	Static	Dynamic	Hybrid	Memory
DCAM	No	Yes	No	No	No
MARE	No	No	Partial	No	No
Vidarthi et al.	Partial	Yes	Yes	No	No
SAMA	Yes	Yes	Yes	No	Partial
Proposed approach	Yes	Yes	Yes	Yes	Yes

III. METHODOLOGY

Since the main objective of the paper is to illustrate a practical approach to carrying out malware analysis, the two sample pieces of malware were analyzed using the following integrated tools and the four analyses. The implemented methodology is shown in Fig. 3 and can be used as a guideline for future comprehensive malware analyses.

A. Lab Setup

Flare VM is an open-source collection of software installation scripts for Windows systems to easily setup and maintains a reverse engineering environment on a virtual machine. It was installed on Virtual Box hypervisor to analyze the encrypted malicious file downloaded from Malware Bazaar—a project from abuse.ch). Then, it was downloaded and installed on Windows 10 VM. A system snapshot was taken before each analysis to preserve the integrity of the

results and for future-reference analysis. The following sub sections discuss the rationale for choosing the four malware analysis techniques.

B. Applying Static Analysis

Static analysis examines the malicious sample by gathering maximum information without executing it by following a three-step phases approach: de-obfuscation, basic properties analysis (BPA), and advanced static analysis (ASA) (Fig. 4).

The cyber attacker uses obfuscation to intentionally disguise some attributes of the malware specimen by packing it. Hence, before proceeding with the analysis, the initial step should be to perform an obfuscation check and bring the sample to its unpacked state (BPA) In this respect, the file type and signature identification are crucial steps of static analysis to obtain useful information, such as the target operating system (OS) and the architecture of the suspicious file. Among Windows' basic executable files, the presence of portable executables in the form of .exe or .dll provides a glimpse of hexadecimal values and notes that are present in a file.

Malware hashing involves the generation of cryptographic hashes for a malicious file. Hashing algorithms are commonly used to generate hash values of the malicious files are MD5, SHA-256, and SHA-1. This step provides unique values that act as fingerprints for the malware samples. VirusTotal website allows for the flexibility to either upload a file or a URL or simply search the hash value of the sample, and it offers API-based support for detection and recognition by supplying the details of the previous records created by other researchers. String analysis extracts legible letters and words from the malware and focuses on critical information that can be fetched from strings, such as file names, IP addresses, registry keys, and URLs. However, an attacker may include fake strings to divert an analyst from disrupting their task, as strings provide an overview of what malware can do.

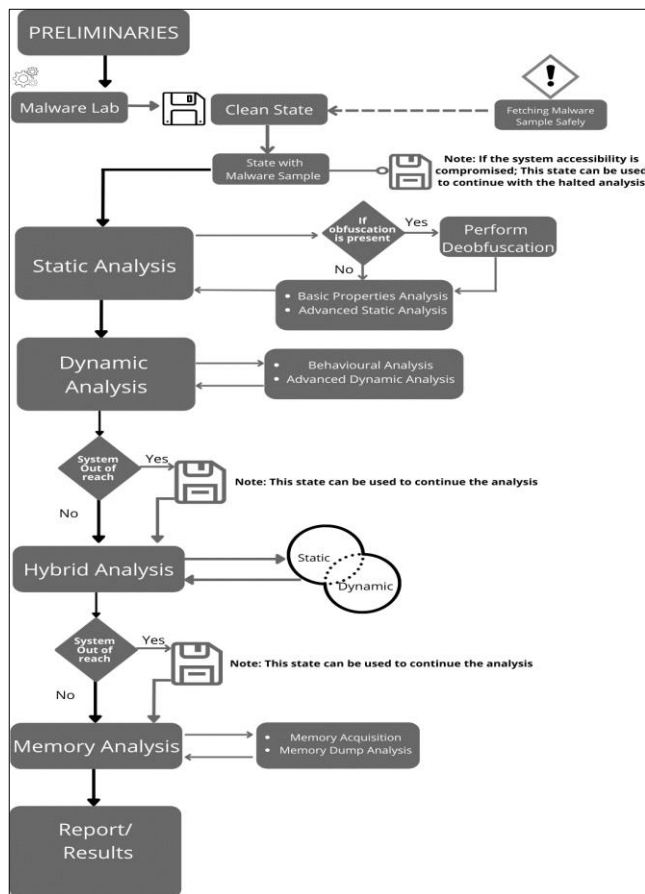


Fig. 3. Malware analysis methodology.

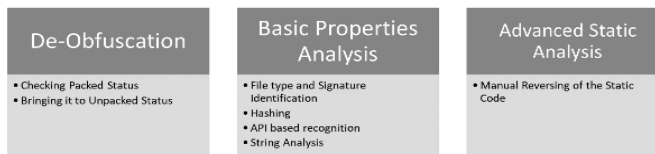


Fig. 4. The three-step phase of the static analysis.

The goal of advanced static analysis of the static code is to understand the malware code's design. This kind of analysis includes the analysis of the machine code by disassembling a file. Performing the static code analysis after BA is appropriate because it requires analyzing the processes and behavior of malware by comparing the two states, namely pre- and post-execution. Conversely, if static code analysis is performed before behavioral analysis, it might reduce the accuracy.

C. Applying Dynamic Analysis

Dynamic analysis implies the execution of the malware sample in a contained and safe environment (sandboxed) to understand the malware's functionality, which includes changes in registry and the files created by it. The objective is to cover two important phases of dynamic analysis, namely behavioral analysis (BA) and advanced dynamic analysis (ADA) (Fig. 5).

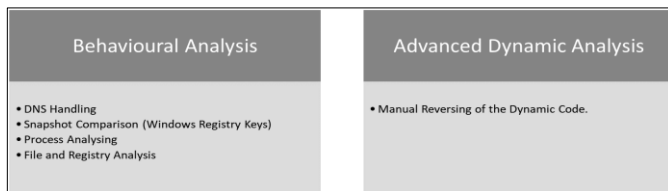


Fig. 5. The two step phase of dynamic malware analysis.

The objective of BA is to understand the suspicious behavior of malware through the interaction with the sample to gather maximum information. BA helps understand the changes in registry, network, and files. The methods in this step include domain name system (DNS) handling, snapshot comparison, process analysis, and registry analysis. DNS handling involves setting up a fake server to generate responses for the requests created by the suspicious file. When the malware is executed, it creates a DNS request so that it can perform the required malicious behavior. These requests are resolved by creating a fake server that fools the malware and generates the response. The snapshot comparison of Window's registry keys focuses on obtaining information about the changes in the number of registry keys and their values before and after the execution of the suspicious file. A registry key is an organizational unit that serves the purpose of an internal database and is used by the computer to store information related to configuration. In the process analysis stage, the malicious application is executed to elicit information regarding its behavior, namely the activities of the application and the details of threads, memory, handles, and the child processes created by it. Analysis of real-time registry, file system, and thread activity of the malicious file involves advanced monitoring of the applications using thread stacks, sessions created, and their activities. It also helps obtain information on the path the processes have traversed in the system, including the changes made. The objective of the ADA stage is to perform advanced analysis on the code by debugging the dynamic code by executing a file.

D. Applying Hybrid Analysis

Since hybrid malware analysis assist to obtain the benefits of both static and dynamic malware analysis [24], this increases their ability to detect malicious software correctly.

Furthermore, this analysis technique has all the strengths of static and hybrid analyses while overcoming the shortcoming they have when they are performed independently.

E. Applying Memory Analysis

The main objective of performing this stage is to gain information by monitoring memory changes. An analyst examines the memory dump to gain additional information on process execution and performs a restoration step to make a clean state for further analysis. Memory analysis is integrated with hybrid analysis when an analyst applies basic static analysis to the information gathered during interactive BA, namely the execution of malicious code to generate memory changes followed by dynamic analysis. This phase will be reverted to perform basic static analysis on that memory dump. The table below lists the tools used in malware analysis.

TABLE II. TOOLS AND TECHNIQUES USED IN MALWARE ANALYSIS

No.	Tools	Static	Dynamic	Hybrid	Memory
1	ExinfoPE	Y	X	X	X
2	Hex Editor	Y	X	X	X
3	PeStudio	Y	X	X	X
4	Virustotal	Y	X	X	X
5	Ghidra	Y	X	Y	X
6	ApateDNS	X	Y	X	X
7	Regshot	X	Y	X	X
8	Process Monitor	X	Y	X	X
9	IDA	X	X	Y	X
10	Procmon	X	X	Y	X
11	AccessData FDK	X	X	X	Y
12	Volatility	X	X	X	Y

IV. RESULTS AND ANALYSIS

This section presents a practical approach to analyzing malware using open-source and powerful tools (Table II). The WhisperGate malware is analyzed first, followed by the analysis of BlackCat malware. The preliminary lab setup was meticulously followed to ensure the status of the malware before and after each process (i.e., static, dynamic, hybrid and memory analysis). The following subsections highlight the experimental results obtained through the judicious use of appropriate tools.

A. Experimental Findings - WhisperGate

1) *Static analysis*: The static analysis followed a three-step approach, namely deobfuscation, BPA, and ASA. Exeinfo PE is a software that can be used to view executable file properties. When using the tool Exeinfo PE to deobfuscate and identify whether the malware was obfuscated (Fig. 6), it was found that the file was unpacked.

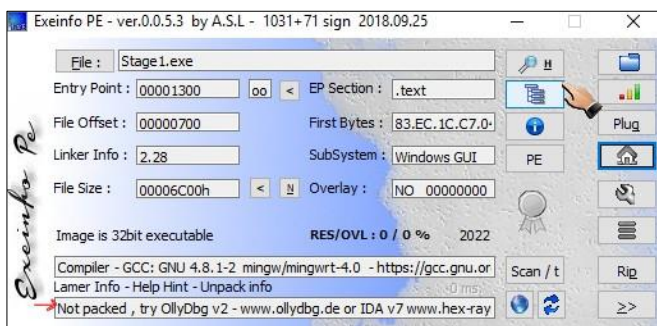


Fig. 6. Deobfuscation of WhisperGate using Exinfo PE.

In the File and Signature Identification step of BPA, the tools Hex Editor (Hxd) were used to obtain detailed information on the signature (Fig. 7 and 8), and PeStudio was used to identify the file type of the malware (Fig. 9). HxD is a tool that can inspect, compare, and verify files, disks, disk images, memory, and log files; patch errors, and repair disk structures. Used for malware detection, PeStudio analyzes the executable files and provides information about the file's properties, characteristics, and potential risks.

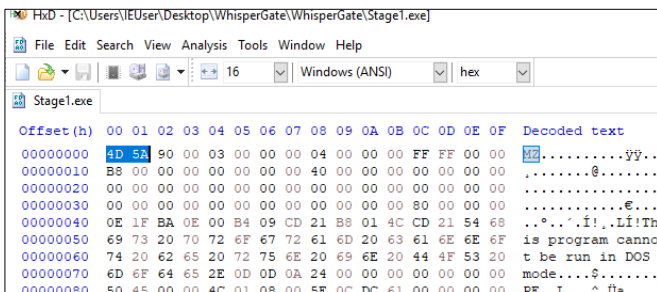


Fig. 7. File and signature identification of WhisperGate using Hxd.

The figure shows that the first two bytes contain 4D 5A, and the decoded text is MZ (which stands for Mark Zbikowski, a leading developer of DOS). Both these values are a crucial factor, which tells us that it is a portable executable. Another file signature that can be observed from the tool is the note that tells that "This program cannot be run in DOS mode." The decoded text states that "Your hard drive has been corrupted... In case you want to recover all hard drives of your organization, You, should pay us \$10k Dollars via bitcoin wallet ** and send message via tox ID ** with your organization name. We will contact you to give further instructions" (Fig. 8).

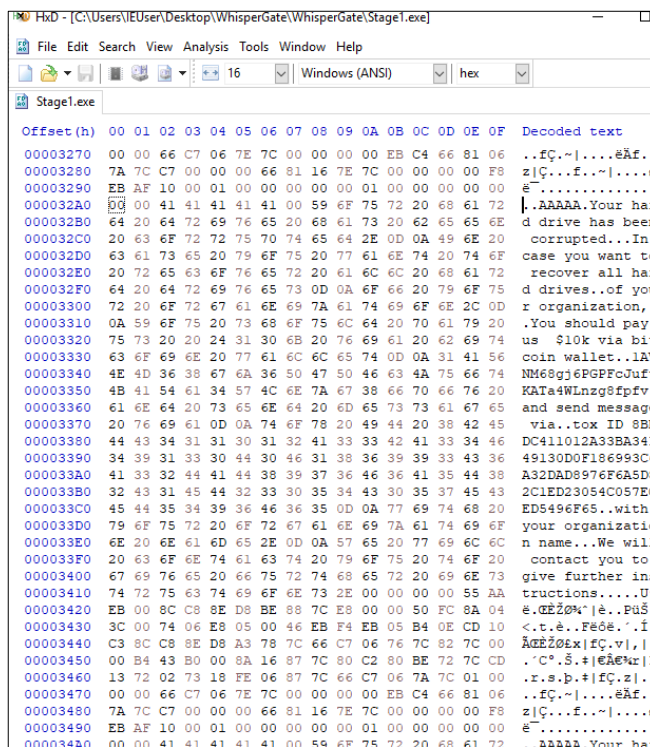


Fig. 8. Decoded text message of WhisperGate using Hxd.

After the signature identification, PeStudio was used to identify the correct type of file (Fig. 9). Based on the results, the file is built on a 32-bit CPU architecture with a file size of 27648 bytes.

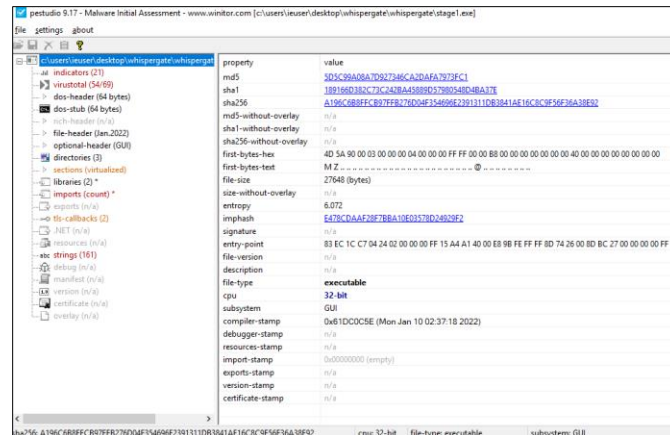


Fig. 9. File type identification of WhisperGate using PeStudio.

Upon hashing, PeStudio generated the hash values:

MD5 : 5D5C99A08A7D927346CA2DAFA7973FC1

SHA-1:

189166D382C73C242BA45889D57980548D4BA37E

SHA-256:

A196C6B8FFCB97FFB276D04F354696E2391311DB3841AE16C8C9F56F36A38E92

The API-based identification stage involves using the web-based tool VirusTotal. VirusTotal is an online portal where users can upload suspicious files. It uses antivirus engines and website scanners to detect various types of malware and malicious content. The identification and classification of the malware show a community score of 54 on a scale of 69, signifying malware detection by 54 security vendors out of 69. Details such as the history of the application, the compilation stamp, and the information of the target identified through the API are shown below in Fig. 10.

Fig. 10. API identification of WhisperGate from VirusTotal.

In the last step of BPA, namely string analysis, PeStudio was used to analyze the strings and retrieve useful information, as shown in Fig. 11.

encoding (2)	size (byte)	file offset	blacklist (4)	hex (23)	group (7)	index (16)
ascii	4	00333029	-	...	-	...
ascii	184	00333034	-	...	-	Your hard drive has been corrupted. You can use the recovery tool to recover all hard drive data.
ascii	184	00333039	-	...	-	Your hard drive has been corrupted. You can use the recovery tool to recover all hard drive data.
ascii	184	00333044	-	...	-	Your hard drive has been corrupted. You can use the recovery tool to recover all hard drive data.
ascii	184	00333049	-	...	-	Your hard drive has been corrupted. You can use the recovery tool to recover all hard drive data.
ascii	184	00333054	-	...	-	Your hard drive has been corrupted. You can use the recovery tool to recover all hard drive data.
ascii	184	00333059	-	...	-	Your hard drive has been corrupted. You can use the recovery tool to recover all hard drive data.
ascii	184	00333064	-	...	-	Your hard drive has been corrupted. You can use the recovery tool to recover all hard drive data.
ascii	184	00333069	-	...	-	Your hard drive has been corrupted. You can use the recovery tool to recover all hard drive data.
ascii	184	00333074	-	...	-	Your hard drive has been corrupted. You can use the recovery tool to recover all hard drive data.
ascii	184	00333079	-	...	-	Your hard drive has been corrupted. You can use the recovery tool to recover all hard drive data.
ascii	184	00333084	-	...	-	Your hard drive has been corrupted. You can use the recovery tool to recover all hard drive data.
ascii	184	00333089	-	...	-	Your hard drive has been corrupted. You can use the recovery tool to recover all hard drive data.
ascii	184	00333094	-	...	-	Your hard drive has been corrupted. You can use the recovery tool to recover all hard drive data.
ascii	184	00333099	-	...	-	Your hard drive has been corrupted. You can use the recovery tool to recover all hard drive data.
ascii	184	00333104	-	...	-	Your hard drive has been corrupted. You can use the recovery tool to recover all hard drive data.
ascii	184	00333109	-	...	-	Your hard drive has been corrupted. You can use the recovery tool to recover all hard drive data.
ascii	4	00333110	-	...	-	...

Fig. 11. String identification of WhisperGate using PeStudio.

A total of 161 strings were identified, four of which were blacklisted, and 16 carried the note “Your hard drive is corrupted,” indicating that malicious activity could be carried out using the sample. Advanced static analysis involves the single task of manually reversing the static code. The tool Ghidra (a free and open source reverse engineering tool) was used to disassemble the code and further examine the functions, which provided relevant information regarding the nature of the file, as shown in Fig. 12.

```

*(undefined4 *) (aparam_1 + iVar1) = 0;
*(undefined4 *) (stack0x00000000 + iVar1) = 0;
*(undefined4 *) ((int)sp_stack4 + iVar1) = 3;
*(undefined4 *) (stack0xffffffff + iVar1) = 0;
*(undefined4 *) (stack0xffffffff + iVar1) = 3;
*(undefined4 *) (stack0xffffffff + iVar1) = 0x10000000;
*(wchar_t **) (stack0xffffffff + iVar1) = L"\\\\.\\PhysicalDrive0";
*(undefined4 *) ((int)sp_stack24 + iVar1) = 0x403bcf;
pVar3 = CreateFileW(LPCWSTR *) (stack0xffffffff + iVar1), (DWORD *) (stack0xffffffff + iVar1),
    (DWORD *) (stack0xffffffff + iVar1),
    (LPSECURITY_ATTRIBUTES *) (stack0xffffffff + iVar1),
    (DWORD *) ((int)sp_stack4 + iVar1), (DWORD *) (stack0x00000000 + iVar1),
    (HANDLE *) (aparam_1 + iVar1));
*(HANDLE *) (stack0xffffffff + iVar1) = pVar3;
*(undefined4 *) ((int)sp_stack4 + iVar1) = 0;
*(undefined4 *) (stack0xffffffff + iVar1) = 0;
*(undefined4 *) (stack0xffffffff + iVar1) = 0x300;
*(undefined4 **) (stack0xffffffff + iVar1) = local_2020;
*(undefined4 *) ((int)sp_stack24 + iVar1) = 0x403bcf;
WriteFile((HANDLE *) (stack0xffffffff + iVar1), (LPCVOID *) (stack0xffffffff + iVar1),
    (DWORD *) (stack0xffffffff + iVar1), (LPOVERLAPPED *) (stack0xffffffff + iVar1),
    (LPOVERLAPPED *) ((int)sp_stack4 + iVar1));
*(HANDLE *) (stack0xffffffff + iVar1) = pVar3;
*(undefined4 *) ((int)sp_stack24 + iVar1) = 0x403bc9;
iVar4 = CloseHandle((HANDLE *) (stack0xffffffff + iVar1));
*(BOOL *) (stack0xffffffff + iVar1) = iVar4;
return 0;
    
```

Fig. 12. Code reverse of WhisperGate using ghidra.

The images taken from Microsoft documentation (Fig. 13 and 14) shows that the functions, CreateFileA (which opens a physical disk drive or a volume) and WriteFile (which writes data to the specified file or input/output (I/O) device) fall under the category of Data Access and Storage. Hence, it is possible to relate the general syntax with the disassembled code provided by Ghidra (Fig. 12), which contains the same two functions as the one from Microsoft documentation.

Fig. 13. The CreateFileA function that relates to the result in Fig. 12.

Fig. 14. The WriteFile function that relates to the result in Fig. 12.

From the disassembled code, useful information was extracted from the two functions. Their synchronization hints that a file is being opened and overwritten to execute a task. CreateFile accepts the parameter "Physical Drive," which is the name of the file being opened. The access mask used is "0xfffff0." WriteFile provides important details through the handle buffer. The handle returned by CreateFile is used by this function, while the buffer pvVar3 presents the variable Local 2020.

2) *Dynamic analysis:* The dynamic analysis followed two phases, namely BA and ADA. In BA, the malware was executed to interact with the sample to determine its behavior and intended purpose (see Fig. 5 for the four steps). In the DNS processing step, the ApatDNS tool (that aid analysts in DNS identification) was used to spoof DNS responses to DNA requests generated by the malware, as shown in Fig. 15.

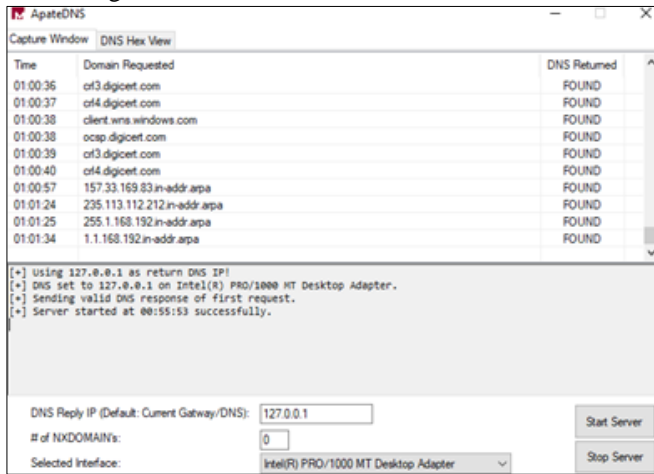


Fig. 15. DNS spoofing of WhisperGate using the tool ApatDNS

During the execution of the sample, the tool successfully captured a list of DNS requests along with the timestamp, indicating that the malware was attempting to connect to different IP addresses for malicious purposes. In the snapshot comparison step, the Regshot tool was used to take sequential snapshots for the pre- and post-execution states to monitor the changes to the registry and files. RegShot that is a tool for controlling changes in the Windows registry can compare the state of registry entries "before" and "after" system changes. The pre- and post-execution snapshots are shown in Fig. 16. The images indicate information regarding the keys, values, and related attributes in the snapshot.

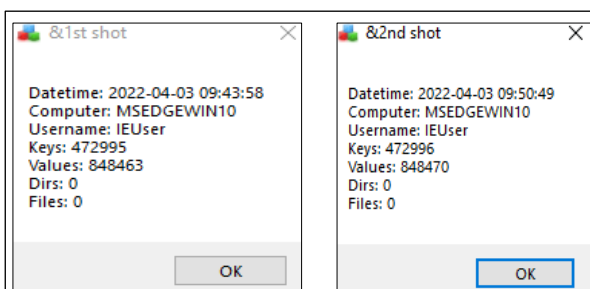


Fig. 16. The pre- and post-execution snapshots using regshot.

The Process Monitor tool was used in the Process Analysis step. Upon invoking the process name filter, the tool generated a list of sub-processes (Fig. 17). The Process Monitor is a troubleshooting and malware hunting monitoring tool for Windows that shows real-time file system, Registry and process/thread activity.

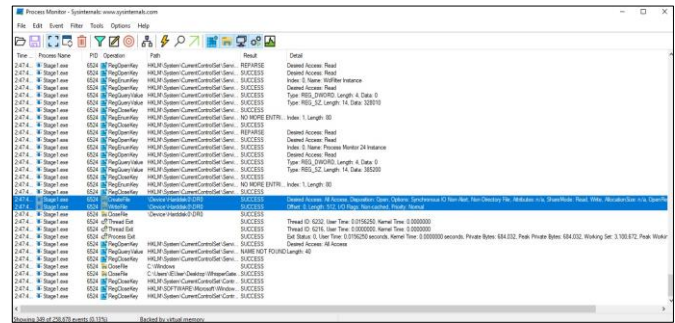


Fig. 17. Process monitoring of WhisperGate using process monitor.

The tool successfully displayed 258,678 events triggered by the malware sample. The main processes highlighted are, again, CreateFile and WriteFile (Fig. 17), and they provide useful information regarding the execution of malware. In particular, CreateFile offers the desired access for the malware to open a file, and WriteFile allows it to overwrite.

In the Registry and File Analysis step, the output generated by Regshot was used to compare the changes made in the registry values and the modifications in the system files, as shown in Fig. 18.

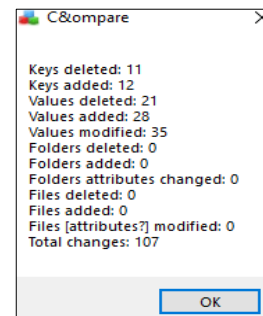


Fig. 18. Registry and file analysis of WhisperGate using regshot.

A total of 107 changes were observed, and the output file displays useful information, such as the addition and deletion of files and registry changes.

In the final ADA step, the IDA tool was used to obtain a debug view of the sample. The IDA tool creates maps of software's execution to display the binary instructions that are actually executed by the processor. The real-time import was used to obtain useful information, as shown in Fig. 19.


```

.ldata:0040A130 ; Segment type: Externs
.ldata:0040A130 ; _idata
.ldata:0040A130 __imp_CloseHandle dd offset kernel32_CloseHandle
.ldata:0040A130 ; DATA XREF: CloseHandle!r
.ldata:0040A134 __imp_CreateFile dd offset kernel32_CreateFileW
.ldata:0040A134 ; DATA XREF: CreateFile!w
.ldata:0040A138 __imp_DeleteCriticalSection dd offset unk_778B8C00
.ldata:0040A138 ; DATA XREF: DeleteCriticalSection!r
.ldata:0040A13C __imp_EnterCriticalSection dd offset unk_778BAFC0
.ldata:0040A13C ; DATA XREF: EnterCriticalSection!r
.ldata:0040A140 __imp_ExitProcess dd offset kernel32_ExitProcess
.ldata:0040A140 ; DATA XREF: ExitProcess!r
.ldata:0040A144 __imp_FindClose dd offset kernel32_FindClose
.ldata:0040A144 ; DATA XREF: FindClose!r
.ldata:0040A148 __imp_FindFirstFileA dd offset kernel32_FindFirstFileA
.ldata:0040A148 ; DATA XREF: FindFirstFile!r
.ldata:0040A14C __imp_FindNextFileA dd offset kernel32_FindNextFileA
.ldata:0040A14C ; DATA XREF: FindNextFile!r
.ldata:0040A150 __imp_FreeLibrary dd offset kernel32_FreeLibrary
.ldata:0040A150 ; DATA XREF: FreeLibrary!r
.ldata:0040A154 __imp_GetCommandLineA dd offset kernel32_GetCommandLineA
.ldata:0040A154 ; DATA XREF: GetCommandLine!r
.ldata:0040A158 __imp_GetLastError dd offset kernel32_GetLastError
.ldata:0040A158 ; DATA XREF: GetLastError!r
.ldata:0040A15C __imp_GetModuleHandleA dd offset kernel32_GetModuleHandleA
    
```

Fig. 19. The debug view of WhisperGate using the IDA tool.

Three vital imported files, CreateFile, CommandLine, and WriteFile, were observed. Out of the 53 imports, four imported an msvcrt library, and the remaining imported a KER-NEL32 library. After creating a breakout for the WriteFile, the file resumed at the GetCommandLine parser. The “stepinto” feature of the tool provides a view of the kernel-based library (Fig. 19).

3) *Hybrid analysis*: Hybrid malware analysis overcomes the shortcomings of individual malware analysis types, as relying on a single malware analysis method does not provide a comprehensive malware analysis report. In this stage, the information generated using the tools Ghidra and Procmon were integrated. Procmon is a utility for Microsoft Windows OS that captures and displays system and network activity. The static analysis provided an overview of the static code along with the basic properties, while dynamic analysis was applied simultaneously to strengthen the information gathered during the static analysis phase. As soon as the file was run, its behavior could be monitored and compare with the relevant information extracted from the static code analysis. This process highlights the importance of synchronization of both static and dynamic analyses (Fig. 20).

```

*(undefined4 *) (iParam_1 + iVar1) = 0;
*(undefined4 *) (stack0x00000000 + iVar1) = 0;
*(undefined4 *) ((int)sp_Stack4 + iVar1) = 3;
*(undefined4 *) (stack0xffffffff0 + iVar1) = 0;
*(undefined4 *) (stack0xffffffff4 + iVar1) = 3;
*(undefined4 *) (stack0xffffffff8 + iVar1) = PERM_ALL;
*(wchar_t **) (stack0xffffffffc + iVar1) = L"\\.\PhysicalDrive0";
*(undefined4 *) ((int)uStackY24 + iVar1) = 0x403bcf;
pVar3 = CreateFileW((LPCWSTR) (stack0xffffffffc + iVar1), (DWORD) (stack0xffffffff0 + iVar1),
    (DWORD) (stack0xffffffff4 + iVar1),
    (LPSECURITY_ATTRIBUTES) (stack0xffffffff8 + iVar1),
    (DWORD) ((int)sp_Stack4 + iVar1), (DWORD) (stack0x00000000 + iVar1),
    (HANDLE) (iParam_1 + iVar1));
*(HANDLE *) (stack0xffffffffc + iVar1) = pVar3;
*(undefined4 *) ((int)sp_Stack4 + iVar1) = 0;
*(undefined4 *) (stack0xffffffff8 + iVar1) = 0;
*(undefined4 *) (stack0xffffffff4 + iVar1) = 0x200;
    
```

Fig. 20. Result of ghidra illustrating the buffer size.

The buffer size in Hexadecimal code 0x200 (last line in Fig. 20), is revealed as 512 in decimal. The first 512 bytes are equal to the exact size of the MBR buffer. The buffer contains the string “Your hard drive has been corrupted.” It is possible that the sample made an effort to corrupt the MBR, but this hypothesis could not be confirmed without performing a hybrid analysis.

To add value to the information, the thread activity overwriting the device's hard disk with a length of 512 bytes was analyzed. The images shown below (Fig. 21 and 22) provide details such as operation, path, offset value, and the result status of the operation.

Process	Thread	Operation	Path	Result	Offset	Length	I/O Flags
C:\Users\IEUser\Desktop\WhisperGate\WhisperGate\Stage1.exe	6524	OpenControlSet.Serv...	HKLM\System\CurrentControlSet.Serv...	SUCCESS			
Stage1.exe	6524	RegEnumKey	HKLM\System\CurrentControlSet.Serv...	NO MORE ENTRI...	Index: 1	Length: 80	
Stage1.exe	6524	RegCloseKey	HKLM\System\CurrentControlSet.Serv...	SUCCESS			
Stage1.exe	6524	CreateFile	\Device\Harddisk0\DR0	SUCCESS	Desired Access: All Access, Disposition: Open, Options: Synchron...		
Stage1.exe	6524	WriteFile	\Device\Harddisk0\DR0	SUCCESS	Offset: 0	Length: 512	I/O Flags: Non-cached, Priority: Normal
Stage1.exe	6524	CloseFile	\Device\Harddisk0\DR0	SUCCESS			

Fig. 21. The thread activity showing the overwriting of the hard disk.

Event	Process	Stack
Date:	4/4/2022 2:23:03.916660 AM	
Thread:	6528	
Class:	File System	
Operation:	WriteFile	
Result:	SUCCESS	
Path:	\Device\Harddisk0\DR0	
Duration:	0.0386311	
Offset:	0	
Length:	512	
I/O Flags:	Non-cached	
Priority:	Normal	

Fig. 22. Event view of the result of the thread activity (see Fig. 21).

The thread confirms the successful execution of the operation WriteFile in overwriting 512 bytes of memory in the hard disk. By synchronizing the use of static code obtained under advanced static analysis and the behavioral characteristic observed under the dynamic analysis, critical information about the malware were successfully gathered. Specifically, the nature of the malware is to write the 512 bytes of the hard disk and corrupt the MBR.

4) *Memory analysis*: This phase involves a two-step approach: memory acquisition and memory dump analysis. First, a memory dump of the infected state was obtained, and then the analysis was completed by analyzing this memory dump. In the memory acquisition step, the tool AccessData FTK Imager was used to capture the memory dump of the infected state, as shown in Fig. 23. AccessData FTK Imager is a computer forensics software that can create copies, or forensic images of computer data without making changes to the original evidence.

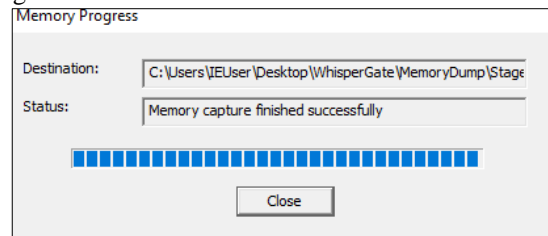


Fig. 23. Acquisition of the memory using the tool AccessData FTK imager.

Subsequently, the tool Volatility was used in the memory dump analysis step to analyze the memory dump of the infected state, which provided valuable information on the running processes (Fig. 24) and the mapping of physical offsets to virtual addresses (Fig. 25). Volatility is a command line memory analysis and forensics tool for extracting and

analyzing the volatile data that is temporarily stored in random access memory from memory dumps.

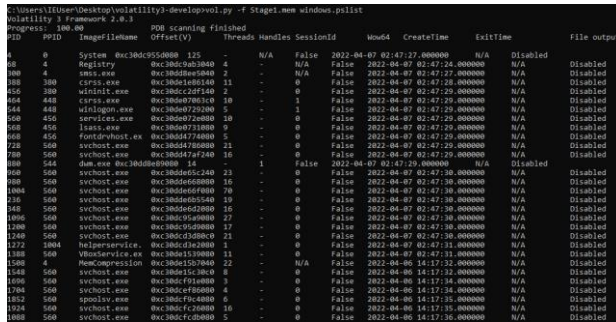


Fig. 24. View of the running processes using the tool volatility.

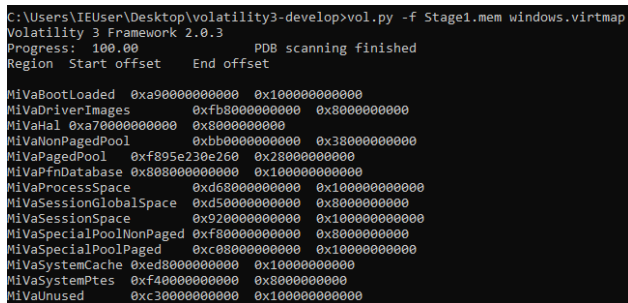


Fig. 25. View of the mapping of physical offsets to virtual addresses using the tool volatility.

While the images show a list of processes running, the Stage1.exe file is not visible. This indicates that the file has been executed with immediate effect to remove its traces. The virtual mapping provides an overview of the start and close offsets of different regions, such as BootLoaded DriverImages.

B. Experimental Findings - BlackCat

BlackCat malware can be analyzed in a similar manner. However, the analysis is not restricted to the tools which were used to analyze WhisperGate; rather, an analyst can use alternate tools as per the situation, but it should be ensured that the necessary parameters are evaluated according to the target. This section focuses on the findings resulting from the analysis of the BlackCat malware sample.

1) *Static analysis:* The static analysis followed a three-step approach: deobfuscation, BPA, and ASA. Deobfuscation using the tool ExeinfoPE revealed that the file was not packed (Fig. 26).

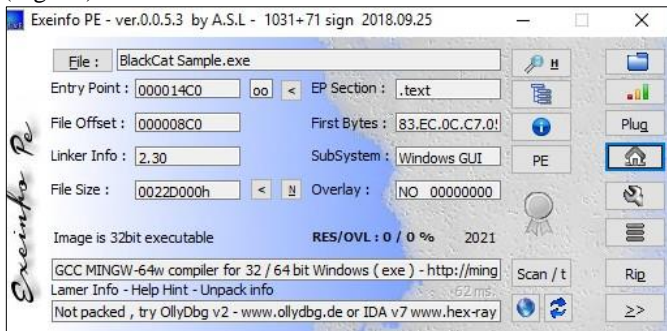


Fig. 26. Deobfuscation of BlackCat using exeinfo PE.

The tools HxD and Pestudnio were used in the BPA for file and signature analysis. The Hex view of the sample is provided in Fig. 27.

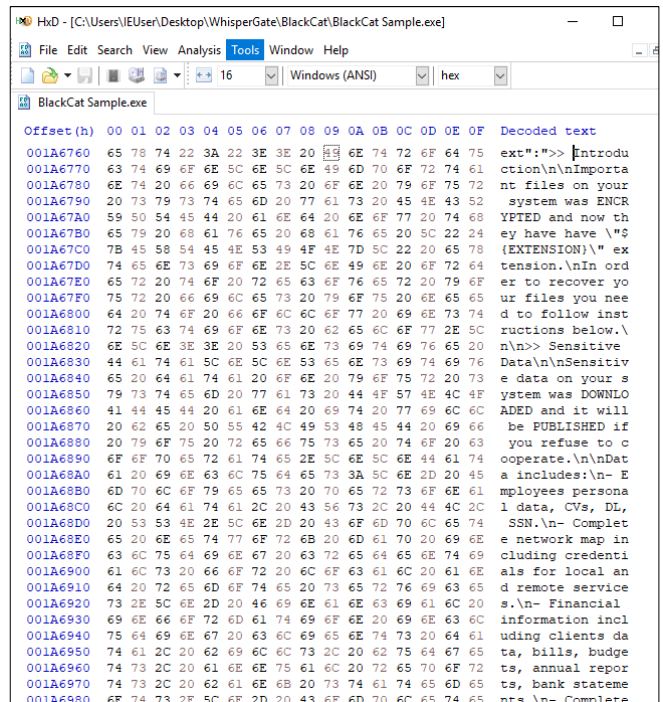


Fig. 27. Decoded text message of BlackCat using Hxd.

The findings revealed the malware to be a PE file with the following decoded text, as shown in Fig. 27: “Important files on your system were ENCRYPTED, and now they have, have. In order to recover your files, you need to follow the instructions below. Sensitive data on your system were downloaded, and they will be published if you refuse to cooperate. Data include: employees’ personal data, CVs, DL, SSN...Caution: do not modify files yourself. Do not use third party software to restore your data. You may damage your files; it will result in permanent data loss. Your data are strongly encrypted; you cannot decrypt it without a cipher key.”

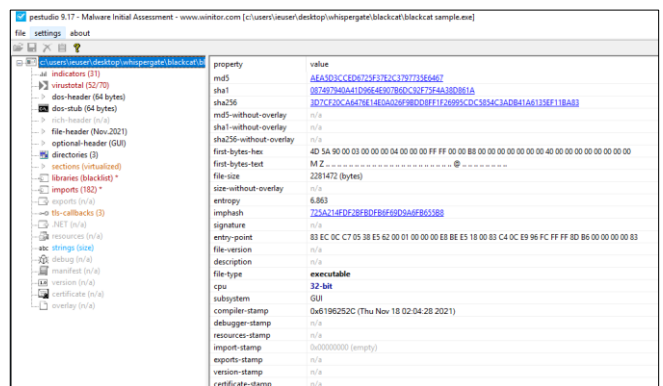


Fig. 28. PeStudio tool identifying the type of file.

After the signature identification, the tool PeStudio identified the type of file (Fig. 28). Based on the results, the file is built on a 32-bit CPU architecture with a file size of

2281472 bytes. In terms of hashing, the following values were obtained:

MD5 : AEA5D3CCED6725F37E2C3797735E6467

SHA-256: 087497940A41D96E4E907B6DC92F75F4 A38D 861°

SHA-1 : 3D7CF20CA6476E14E0A026F9BDD8FF1F26995C DC5854C3A

DB41A6135EF11BA83

The file was identified as a ransomware using Virustotal in the APU-based identification step, with a community score of 52/70.

String analysis through the tool PeStudio (Fig. 29) identified 13454 strings with 73 blacklisted. It was observed that the File-offset 0x0022C514 has a string value WriteFile. This is a critical finding that can assist in the advanced static analysis process.

encoding (2)	size (bytes)	file-offset	blacklist (73)	hint (208)	group (14)	value (13454)
ascii	7	0x0022C30E	x	utility	network	connect
ascii	4	0x0022C319	x	utility	network	send
ascii	19	0x0022C344	x	import	synchronization	GetOverlappedResult
ascii	25	0x0022C38E	x	import	synchronization	QueuePerformanceInformation
ascii	15	0x0022B830	x	import	storage	FindVolumeClose
ascii	30	0x0022C4E2	x	import	storage	WouldDisableWoodOffRedirection
ascii	14	0x0022B924	x	import	services	ControlService
ascii	16	0x0022B969	x	import	services	OpenProcessToken
ascii	22	0x0022C0DA	x	import	reckoning	GetTimeZonesInformation
ascii	16	0x0022C32C	x	import	network	NetApiBufferFree
ascii	13	0x0022C340	x	import	network	NetServerEnum
ascii	12	0x0022C350	x	import	network	NetShareEnum
ascii	10	0x0022C39C	x	import	network	WSACleanup
ascii	15	0x0022C36A	x	import	network	WSASetLastError
ascii	10	0x0022C35A	x	import	network	WSAStartup
ascii	11	0x0022C3D0	x	import	network	OpenSocket
ascii	12	0x0022C3E9	x	import	network	freeaddrinfo
ascii	11	0x0022C3F9	x	import	network	getaddrinfo
ascii	11	0x0022C308	x	import	network	socket
ascii	8	0x0022C31C	x	import	network	socket
ascii	10	0x0022C33A	x	import	network	socketopt
ascii	28	0x0022B94E	x	import	file	GetFileInformationByHandle
ascii	13	0x0022C302	x	import	file	MapFileToOffset
ascii	15	0x0022C488	x	import	file	UpdateViewOfFile
ascii	9	0x0022C314	x	import	file	WriteFile
ascii	24	0x0022B974	x	import	execution	CreateToolhelp32Snapshot
ascii	18				execution	GetCurrentProcess

Fig. 29. String analysis of BlackCat using PeStudio.

In the advanced static analysis step, the tool IDA revealed the presence of GetCommanLineW, indicating the intended behavior of the sample when it utilized the command line for a specific task (Fig. 30).

00000000...	FreeConsole	KERNEL32
00000000...	FreeEnvironmentStringsW	KERNEL32
00000000...	FreeLibrary	KERNEL32
00000000...	GetCommandLineW	KERNEL32
00000000...	GetComputerNameW	KERNEL32

Fig. 30. Advanced static analysis step results using the tool IDA.

2) *Dynamic analysis:* To analyze the malware behavior, the tool ApatеDNS was used to monitor the DNS requests generated by the malware. However, no legitimate responses were identified (Fig. 31).

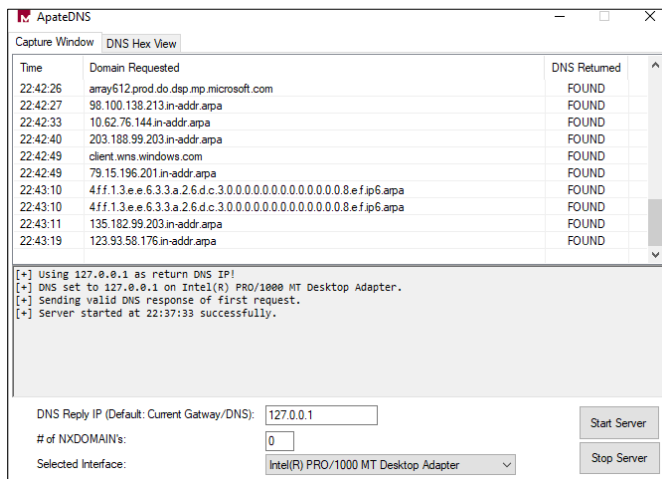


Fig. 31. DNS spoofing of BlackCat using the tool ApatеDNS.

The snapshot comparison tool Regshot was used to compare the snapshot of the registry before and after executing the executable (Fig. 32). The snapshots indicate the changes in the keys and values.

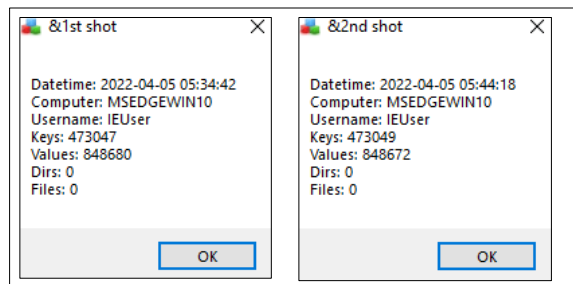


Fig. 32. The pre- and post-execution snapshots using regshot.

The tool Process Monitor was used for the process analysis, which revealed that 389633 processes were triggered on malware execution (Fig. 33). The CreateFile process was highlighted, but no such evidence of WriteFile was produced.

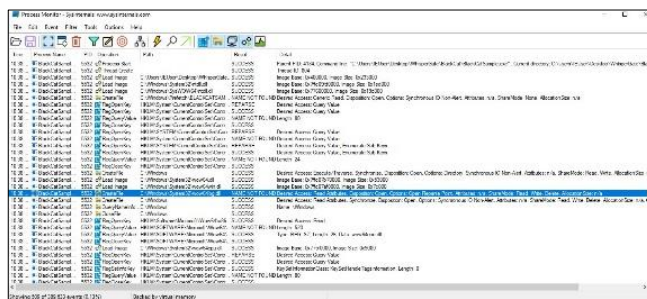


Fig. 33. The result from the tool process monitor with CreateFile highlighted.

In the subsequent step, the tool Regshot was used for registry and file analysis, which highlighted 138 changes (Fig. 34)



Fig. 34. Registry and file analysis using the tool regshot.

Using the ADA tool IDA, the snapshot revealed that the GetCommandLineW process imported a kernel-based library (Fig. 35).

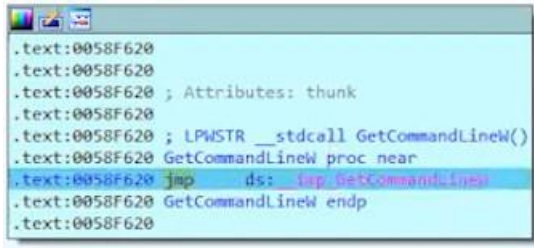


Fig. 35. Results of the snapshot using the tool IDA.

3) *Hybrid analysis:* In static analysis, information was gathered about the GetCommandLineW call. Through dynamic code analysis, the complete code was debugged to extract some useful keys: “h,” “p,” “e,” “-,” and “l.” By using the command prompt feature running ProcMon, it was surmised that the keys could be the instructions used in the command prompt, which could be executed to further examine the intended purpose (Fig. 36). Here, the command prompt executes the sample and passes a log file to a particular directory.

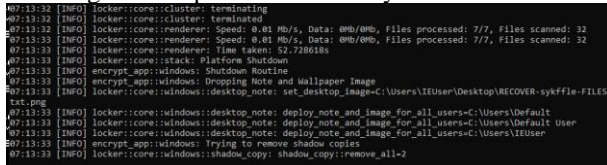


Fig. 36. Results from the command prompt running ProcMon.exe.

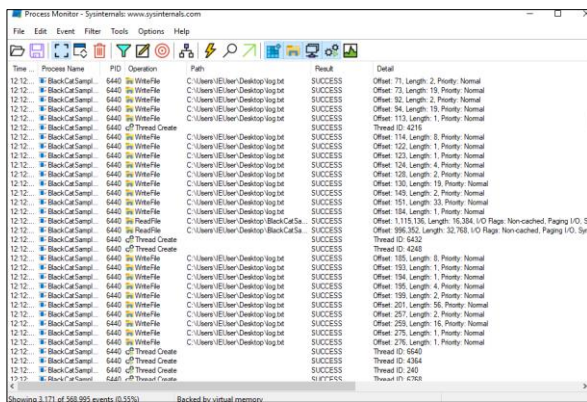


Fig. 37. Results from the tool process monitor.

When using the Process Monitor tool (Fig. 37), it was observed that the malware triggered almost 1.5 times (568,995) the number of processes executed by the same malware when compared to dynamic analysis (389,633). This time, the WriteFile operation was evident and confirmed based on the process executed through the command prompt. The directory set during the execution was successfully injected with the log file, thus corrupting the services.

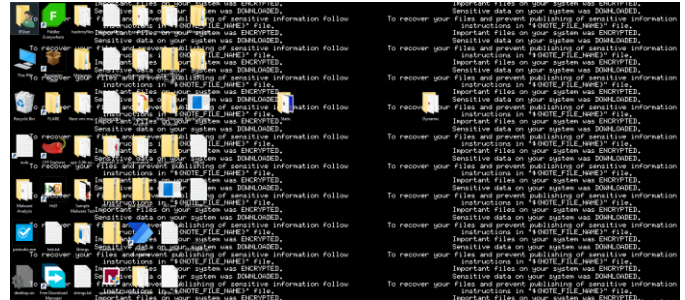


Fig. 38. The wallpaper image dropped through the execution of the malware using cmd.

Once the file was executed through the cmd, it took 53 seconds to corrupt the services, along with dropping a note and wallpaper image (Fig. 38). This demonstrated the speed and potency of the malware in infecting the system.

4) *Memory analysis:* The tool AccessData FTK was used to capture the memory dump of the infected state (Fig. 39) at the initial memory acquisition step.

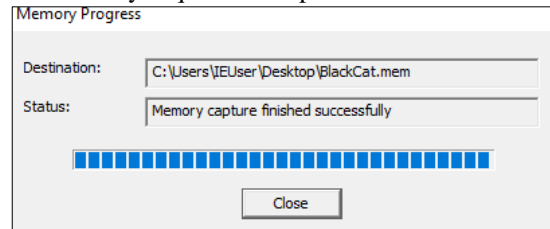


Fig. 39. Capturing the memory dump using the tool AccessData FTK.

Next, in the memory dump analysis step, the Processlist successfully showed the execution of the BlackCat sample in the infected memory dump (Fig. 40).

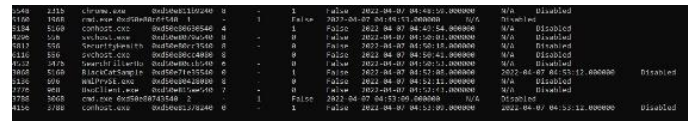


Fig. 40. Image showing the ProcessList upon execution.

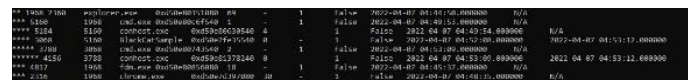


Fig. 41. Image showing the ProcessList details for the malware sample.

The Processtree provides details such as the execution time and the offset value for the malware sample (Fig. 41). The command line operation shows the request to memory accessibility at a particular offset value (Process ID 3788 in Fig. 42).

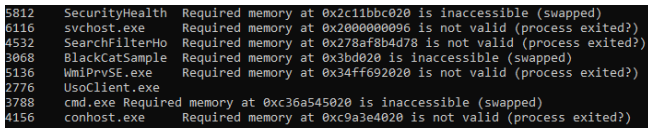


Fig. 42. Image showing the result of the memory accessibility (PID 3788)

V. DISCUSSION

The proposed methodology is applicable for analyzing different malicious files. The case study provides a demonstration of malware analysis on WhisperGate and BlackCat. It is advisable for an analyst to prepare a summary report based on the experimental results of the sample. This section presents a report of the results relative to the two candidate pieces of malware used in the case study.

C. WhisperGate

The analysis of WhisperGate shows a deobfuscated .exe file with 32-bit CPU architecture carrying threatening information in a static approach. While performing dynamic analysis through the registry modification, the impact of malware was also noticeable. Its nature was identified through hybrid analysis that used both static and dynamic processes in which the malware overwrote the MBR. Changes in the disk led by the malware sample were observed through the offset mapping to the bootloader and driver images (Table III).

TABLE III. SUMMARY REPORT OF WHISPERGATE MALWARE

Static	Dynamic	Hybrid	Memory
An unpacked .exe file with a CPU architecture of 32 bits and threatening strings (message) was identified. The API-generated score of 54/69 indicates the presence of characteristics typical of ransomware and wipers. This finding allowed for classifying the malware as a suspicious file with a monetary purpose.	Post-execution impact was visible in the registry modification. Along with a trigger of 2,58,678 events, the WriteFile event brought attention to the modification/overwriting of the file.	Static analysis indicated a buffer 0x200 equal to 512 in decimal, which is indeed the size of MBR. Dynamic analysis generated a thread to overwrite 512 bytes of memory in the hard disk. This finding confirmed the nature of malware corrupting the MBR through overwriting	Virtual mapping of the offsets related to the BootLoaded and DriverImages brought attention to the possible changes in the disk.

Fig. 43 indicates the impact of running the malware sample WhisperGate showing the output “Your hard drive has been corrupted”. The sample overwrites the MBR and displays a ransom note demanding \$10k via cryptocurrency (“You should pay us \$10k via bitcoin wallet”), thus validating the experimental analysis.

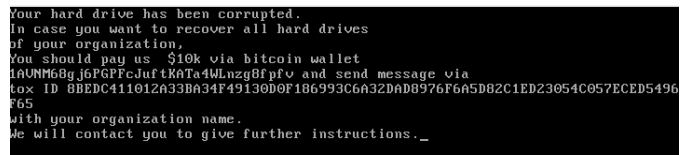


Fig. 43. Image showing the impact of executing WhisperGate.

D. BlackCat

Like WhisperGate, the static analysis of BlackCat showed a deobfuscated .exe file with a 32-bit CPU architecture carrying a threatening note. The API score indicates that the sample is ransomware with a monetary purpose. Through registry modification performed in a dynamic approach, the use of the command prompt by the sample was observed. During the hybrid approach, both key identifications used for executing the command prompt were performed by the malware along with a threatening message, indicating the malicious nature of the sample as ransomware. In memory analysis, the same result was revealed through the command line request for memory access (Table IV).

TABLE IV. SUMMARY REPORT OF BLACKCAT MALWARE

Static	Dynamic	Hybrid	Memory
An unpacked .exe file with a CPU architecture of 32 bits and threatening Strings (message) was identified. The API-generated score of 52/70 indicated the presence of characteristics typical of ransomware, and a Trojan classified the latter as a suspicious file with a monetary purpose.	Post-execution impact was visible in the registry modification. Along with a trigger of 3,89,633 events, a kernel-based library for the GetCommandLine function indicated the usage of the command prompt by the sample.	Static analysis helped identify keys that were then used for executing via the command prompt in the dynamic analysis. An injected log file encrypting files and a background image showing a threatening message confirmed the malicious nature of the ransomware.	Command line operation requesting memory accessibility indicated the suspicious nature of the sample.

Fig. 44 indicates the impact of running the malware sample BlackCat. The sample corrupted the directory and displayed a background image with a threatening note of the ransom, thus validating the results of the experimental analysis performed using static, dynamic, hybrid, and memory analysis.

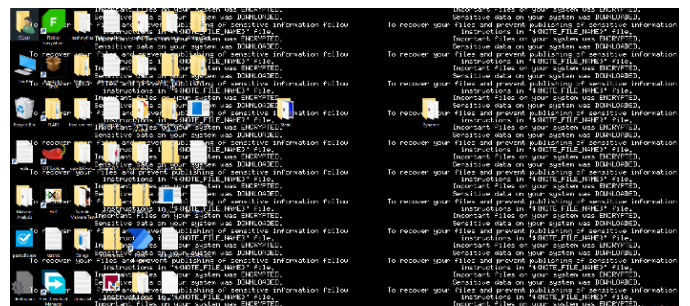


Fig. 44. Image showing the impact of executing BlackCat.

VI. CONCLUSION AND FUTURE WORK

The paper offers a comprehensive and practical approach to performing in-depth analyses of pseudo ransomware by illustrating two pieces of malware, namely WhisperGate (used in cyber warfare) and BlackCat (used to target critical organizations of target states). Twelve tools/techniques were selected, and a detailed description of the steps involved in the application, information extraction, analysis of results, and digital forensics is provided. The malware analysis was successfully executed through the use of static, dynamic, hybrid, and memory analysis and then validated. The detailed malware analysis using twelve tools revealed the embedded information and values in the malicious code for greater visibility and subsequent actions for information technology security personnel and forensic analysts. As malware attacks have rapidly risen with the appearance of innovative malware, the research demonstrated a successful methodology for analyzing potent malware through a comprehensive step-by-step approach. The work overcomes the limitations of relying on a single malware analysis technique thus providing a comprehensive approach to malware analysis.

WhisperGate came into the limelight at the beginning of 2022, when it was used to target multiple government and private organizations in Ukraine. The ransomware malware BlackCat was selected as a sample because it was reported to target European affiliations and U.S. organizations in late 2021. Out of the four malware analysis mentioned in the paper, hybrid analysis provided maximum information critical for the malware analyst to understand the extent of damage.

Three limitations have been observed in this study that can lead to further research. First, since only two pieces of malware (i.e., ransomware and pseudo ransomware) were observed, the experimented malware analysis methodology can be extended to diverse malware samples to validate the methodology. Secondly, since the study was limited to traditional malware analysis, appropriate machine learning methodologies can be deployed in future research to compare the findings with those obtained from traditional malware analysis. Thirdly, in this research, open-source tools were deployed for malware analysis that is already known to malicious hackers for circumventing the analysis process. Hence, future research can compare the results of open source tools with subscription based commercial tools.

REFERENCES

- [1] J. Greig, "BlackCat ransomware targeting US European retail, construction and transportation orgs." ZD Net. <https://www.zdnet.com/article/blackcat-ransomware-targeting-us-european-retail-construction-and-transportation-orgs/> (accessed February, 2023).
- [2] O. Analytica, "Cyberattacks on European energy cement new dynamic," *Emerald Expert Briefings*, no. oxan-db, 2022.
- [3] World Economic Forum, "Global Cybersecurity Outlook 2022." World Economic Forum. https://www3.weforum.org/docs/WEF_Global_Cybersecurity_Outlook_2022.pdf (accessed January, 2023).
- [4] Centre for Internet Security, "Breaking Down the BlackCat Ransomware Operation." Center for Internet Security. (accessed February, 2023).
- [5] Microsoft Security, "The Many Lives of BlackCat Ransomware." Microsoft Inc. <https://www.microsoft.com/en-us/security/blog/2022/06/13/the-many-lives-of-blackcat-ransomware/> (accessed February, 2023).
- [6] E. Berrueta, D. Morato, E. Magaña, and M. Izal, "A survey on detection techniques for cryptographic ransomware," *IEEE Access*, vol. 7, pp. 144925-144944, 2019.
- [7] A. Kapoor, A. Gupta, R. Gupta, S. Tanwar, G. Sharma, and I. E. Davidson, "Ransomware detection, avoidance, and mitigation scheme: a review and future directions," *Sustainability*, vol. 14, no. 1, p. 8, 2021.
- [8] M. Sikorski and A. Honig, *Practical malware analysis: the hands-on guide to dissecting malicious software*. William Pollock, 2012.
- [9] H. Madani, N. Ouerdi, and A. Azizi, "Ransomware: Analysis of Encrypted Files," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 1, pp. 213-217, 2023.
- [10] S. Alelyani and H. Kumar, "Overview of cyberattack on saudi organizations," *Journal of Information Security And Cybercrimes Research*, vol. 1, no. 1, pp. 42-50, 2018.
- [11] F. Mercaldo, V. Nardone, A. Santone, and C. A. Visaggio, "Ransomware steals your phone. formal methods rescue it," in *Formal Techniques for Distributed Objects, Components, and Systems: 36th IFIP WG 6.1 International Conference, FORTE 2016, Held as Part of the 11th International Federated Conference on Distributed Computing Techniques, DisCoTec 2016, Heraklion, Crete, Greece, June 6-9, 2016, Proceedings 36*, 2016: Springer, pp. 212-221.
- [12] U. Urooj, B. A. S. Al-rimy, A. Zainal, F. A. Ghaleb, and M. A. Rassam, "Ransomware detection using the dynamic analysis and machine learning: A survey and research directions," *Applied Sciences*, vol. 12, no. 1, p. 172, 2022.
- [13] P. Zavarsky and D. Lindskog, "Experimental analysis of ransomware on windows and android platforms: Evolution and characterization," *Procedia Computer Science*, vol. 94, pp. 465-472, 2016.
- [14] M. Horduna, S.-M. Lăzărescu, and E. Simion, "A note on machine learning applied in ransomware detection," *Cryptology ePrint Archive*, 2023.
- [15] Trellix, M. Kersten, and R. Samani, "Return of Pseudo Ransomware." Trellix. (accessed February, 2023).
- [16] O. S. Carlos, "Using cyber threat intelligence to support adversary understanding applied to the Russia-Ukraine conflict," *arXiv preprint arXiv:2205.03469*, 2022.
- [17] N. Kostyuk and E. Gartzke, "Why Cyber Dogs Have Yet to Bark Loudly in Russia's Invasion of Ukraine (Summer 2022)," *Texas National Security Review*, 2022.
- [18] Cybersecurity and Infrastructure Security Agency: uscert. "Update: Destructive Malware Targeting Organizations in Ukraine." Department of Homeland Security. <https://www.cisa.gov/uscert/ncas/alerts/aa22-057a> (accessed February, 2023).
- [19] N. Biasini *et al.* "Ukraine Campaign Delivers Defacement and Wipers, in Continued Escalation." Cisco Talos. <https://blog.talosintelligence.com/ukraine-campaign-delivers-defacement/> (accessed February, 2023).
- [20] S. Caltagirone, A. Pendergast, and C. Betz, "The diamond model of intrusion analysis," Center For Cyber Intelligence Analysis and Threat Research Hanover Md, 2013.
- [21] A. Tanner, A. Hinchliffe, and D. Santos, "Threat assessment: Blackcat ransomware." Palo Alto. <https://unit42.paloaltonetworks.com/blackcat-ransomware/> (accessed January, 2023).
- [22] D. Ucci, L. Aniello, and R. Baldoni, "Survey of machine learning techniques for malware analysis," *Computers & Security*, vol. 81, pp. 123-147, 2019.
- [23] M. Egele, T. Scholte, E. Kirda, and C. Kruegel, "A Survey on Automated Dynamic Malware Analysis Techniques and Tools ACM Computing Surveys," 2012.
- [24] R. Sihwail, K. Omar, and K. Z. Ariffin, "A survey on malware analysis techniques: Static, dynamic, hybrid and memory analysis," *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 8, no. 4-2, pp. 1662-1671, 2018.
- [25] E. Gandotra, D. Bansal, and S. Sofat, "Malware analysis and classification: A survey," *Journal of Information Security*, vol. 2014, 2014.
- [26] P. Shijo and A. Salim, "Integrated static and dynamic analysis for malware detection," *Procedia Computer Science*, vol. 46, pp. 804-811,

- 2015.
- [27] O. Or-Meir, N. Nissim, Y. Elovici, and L. Rokach, "Dynamic malware analysis in the modern era—A state of the art survey," *ACM Computing Surveys (CSUR)*, vol. 52, no. 5, pp. 1-48, 2019.
- [28] M. Ijaz, M. H. Durad, and M. Ismail, "Static and dynamic malware analysis using machine learning," in *2019 16th International bhurban conference on applied sciences and technology (IBCAST)*, 2019: IEEE, pp. 687-691.
- [29] R. Sihwail, K. Omar, K. A. Zainol Ariffin, and S. Al Afghani, "Malware detection approach based on artifacts in memory image and dynamic analysis," *Applied Sciences*, vol. 9, no. 18, p. 3680, 2019.
- [30] M. I. Sharif, V. Yegneswaran, H. Saidi, P. A. Porras, and W. Lee, "Eureka: A Framework for Enabling Static Malware Analysis," in *ESORICS*, 2008, vol. 8: Springer, pp. 481-500.
- [31] A. O. Almashhadani, M. Kaiiiali, S. Sezer, and P. O'Kane, "A multi-classifier network-based crypto ransomware detection system: A case study of locky ransomware," *IEEE access*, vol. 7, pp. 47053-47067, 2019.
- [32] A. Ren, C. Liang, I. Hyug, S. Broh, and N. Jhanjhi, "A three-level ransomware detection and prevention mechanism," *EAI Endorsed Transactions on Energy Web*, vol. 7, no. 26, 2020.
- [33] A. Sulaiman, K. Ramamoorthy, S. Mukkamala, and A. H. Sung, "Disassembled code analyzer for malware (DCAM)," in *IRI-2005 IEEE International Conference on Information Reuse and Integration, Conf, 2005.*, 2005: IEEE, pp. 398-403.
- [34] C. Q. Nguyen and J. E. Goldman, "Malware analysis reverse engineering (MARE) methodology & malware defense (MD) timeline," in *2010 Information Security Curriculum Development Conference*, 2010, pp. 8-14.
- [35] D. Vidyarthi, S. Choudhary, S. Rakshit, and C. S. Kumar, "Malware detection by static checking and dynamic analysis of executables," *International Journal of Information Security and Privacy (IJISP)*, vol. 11, no. 3, pp. 29-41, 2017.
- [36] J. Bermejo Higuera, C. Abad Aramburu, J.-R. Bermejo Higuera, M. A. Sicilia Urban, and J. A. Sicilia Montalvo, "Systematic approach to malware analysis (SAMA)," *Applied Sciences*, vol. 10, no. 4, p. 1360, 2020

A Cloud Native Framework for Real-time Pricing in e-Commerce

Archana Kumari¹, Mohan Kumar. S²

School of Engineering and Technology, CMR University, Bengaluru, India¹
Directorate of Research and Innovation, CMR University, Bengaluru, India²

Abstract—Real-time pricing is a form of 'dynamic pricing,' and it enables online sellers to adjust prices in real-time in response to variations in demand and competition to achieve higher revenue or improve customer satisfaction. As modern e-commerce implementations become more cloud-based, this paper proposes a cloud-native framework for a real-time pricing system. We take a requirement driven approach to come up with a modular architecture and a set of reusable components for real-time pricing. Following DSRM methodology, during the design phase, we identify and develop the theoretical foundations for key parts of the system, such as pricing models, competition and demand watchers, and other analytics components that fulfill the functional requirements of the system. At the stage of implementation, we describe how each of these components and the entire cloud application will be configured using an AWS cloud native implementation. As a framework, this work can support a variety of pricing models, demonstrating that multiple pricing models have been discussed. Other low-latency, reusable components described in this work provide the ability to react quickly to changes in demand and competition. We also provide a price-cache that decouples pricing model calculation from end-user price requests and keeps price query latency to a minimum. For a real-time system, where latency stands to be the most desired NFR, we validate the system for price-request latency (found to be a single digit of milliseconds) and market reaction latency (less than a second). Overall, our proposed framework provides a comprehensive solution for real-time pricing, which can be adapted to different business needs and can help online sellers optimize their pricing strategies.

Keywords—Real-time pricing; cloud-native design; system-design; pricing-framework; Amazon web services

I. INTRODUCTION

e-Commerce has become increasingly competitive; therefore, enterprises that intend to win must be able to offer their products and services at prices that are competitive with those of their rivals. Pricing is a crucial aspect of e-commerce, and proper pricing can help to attract and retain customers, increase sales and revenue, and optimize profits [1]. The term "dynamic pricing" refers to the practice of adjusting prices in response to fluctuations in supply, demand, and other factors in the market. Profits can be maximized, sales increased, customer happiness improved, and market share preserved by employing dynamic pricing [2]. Firms are looking to provide prices that are more personalized to each client by using a customer's location and purchase history, among other factors. Dynamic pricing is popular in industries other than e-commerce, including energy, transportation, and financial markets [3].

Real-time pricing is a form of dynamic pricing that involves adjusting prices in response to market factors as they take place in real time. Real-time pricing allows businesses to respond to market conditions in a near-instantaneous fashion, which can help them optimize their profits and minimize risk. By using real-time data, businesses can make more informed pricing decisions and ensure that they are not charging too much or too little for their goods or services [4].

Factors of Dynamic Pricing: Demand is the most important factor in dynamic pricing; when there is high demand, a company may be able to charge a higher price, and in the event of low demand, the company may need to lower the price in order to attract buyers. The amount of competition and what firms know about their prices can help them decide on pricing strategies like undercutting, matching, leading, or skimming. A company may use dynamic pricing to differentiate its prices based on customer segmentation. For example, a company may charge higher prices to customers who are more willing to pay, typically referred to as "myopic customers". Similarly, customer loyalty enables differentiated pricing, i.e., if a company has a loyal customer base, it may be able to charge a higher price for its products or services. Another crucial factor is the level of inventory, i.e., if a company has a limited supply of a product, it may be able to charge a higher price due to the scarcity of the product. Other pricing factors may include seasonality, time of day or week, weather, choice of payment methods, etc. [1] [2].

Cloud technology has undergone significant advancement in recent years, and there are several key developments like containers, serverless computing, AI/ML & IoT etc. The adoption of cloud computing has been growing rapidly in recent years. According to a report by IDC, global spending on cloud services is expected to reach \$1.3 trillion in 2025, up from \$706 billion in 2021 [5]. Cloud computing is increasingly important for businesses as they look to take advantage of cost savings, increased agility and flexibility, and improved scalability and reliability. A wide range of online businesses leverage cloud infrastructure to host stores, manage transactions, and offer integrations and extensions to customize and optimize their online presence [6]. Cloud native refers to a method of developing and deploying applications that fully utilize the cloud computing paradigm. Two popular cloud-native application development paradigms are: 1) Microservices architecture splits an e-commerce platform into discrete services that may be independently developed, deployed, and scaled [7]. 2) Serverless architecture leverages cloud services such as AWS Lambda to manage infrastructure

and run programs without servers. This decreases expenses and improves company agility [8].

A wide range of e-commerce businesses use cloud computing. Amazon's online business runs almost entirely on AWS [9], a platform that hosts and oversees its online store and other operations. Alibaba utilizes the cloud to power its operations on Alibaba Cloud, which handles and maintains all the company's business operations, including its e-commerce platform [10]. Apart from e-commerce businesses, there are cloud-based e-commerce platforms that host and administer online stores. Shopify [11] and Magento [12] are such popular e-commerce platform for establishing and managing online stores [11]. It leverages cloud infrastructure to host stores and manage transactions, and offers integrations and extensions to customize and optimize its online presence.

Real-time pricing has the potential to improve revenue and give businesses real-time control over prices. However, its implementation poses challenges such as real-time decision-making, managing large volumes of data, achieving scalability, and accommodating diverse product types. Cloud-native technology has become the preferred platform for contemporary e-commerce development, and designing a cloud-native application can help address many of these challenges. To increase adoption of real-time pricing by sellers, it is important to solve the challenge of real-time decision-making and develop a reusable framework that can be used for different e-commerce product types. The objective of this paper is to develop a framework for a real-time pricing system that helps online stores optimize revenues or other business specific pricing requirements. We will begin by identifying the requirements and specifications, using them to guide the design and build the necessary functional foundation. To implement the framework, we will use cloud-native technology, which offers cost-effectiveness, scalability, and quick time-to-market. In addition, we will follow the philosophy of "meeting customers where they are", by implementing the system on a popular cloud platform. Specifically, we have chosen AWS, which is currently the most widely used cloud solution.

The remaining sections of this paper are organized as follows: Section II surveys the existing literature, explores all the different types of work that has been carried out in the field of dynamic pricing. There we will also identify the lack of system design and implementation work in the existing literature, among other things as a research gap. The methodology and system requirements are discussed in Section III and Section IV, respectively. Section V presents the high-level design of the real-time pricing system, while Section VI focuses on the cloud-native implementation of the same. The results of this study are discussed in Section VII, and Section VIII provides the conclusion and potential future work.

II. LITERATURE REVIEW AND RESEARCH GAP

A. Background of Dynamic Pricing in e-Commerce

Dynamic pricing is a multi-disciplinary research area as it involves the integration of knowledge and techniques from multiple fields, including economics, mathematics, statistics, computer science, and behavioral psychology [13]. The primary goal of dynamic pricing is to adjust prices based on

various factors such as supply and demand, competition, and customer behavior, to remain competitive in the market [14]. Dynamic Pricing implementation may involve statistic, algorithms, and machine learning techniques to predict demand and adjust prices accordingly [15]. Dynamic pricing has several benefits, including increased sales, improved profit margins, and better customer satisfaction [16].

B. Algorithms for Dynamic Pricing

1) *Rule-based pricing*: This is dynamic pricing based on predefined rules and conditions that adjusts prices for goods and services accordingly. In this strategy prices are dynamically determined on factors such as demand, supply, competition, time of day, and other market conditions. [17].

2) *Competition pricing*: This refers to the practice of basing prices on those offered by competitors. It can be an effective strategy for a low-cost supplier entering a new market. However, blindly following competitive pricing can hinder a company's ability to capitalize on shifting customer perceptions of value and brand differentiation. [18].

3) *Linear programming approaches*: In order to maximize revenue, profit, or market share, among other business goals, linear programming can be used to optimize pricing decisions over time. Linear programming for dynamic pricing seeks to maximize a predetermined objective function by first developing a pricing model that factors in several factors, including but not limited to customer demand, product availability, and competitive pressure [19][20].

4) *Statistical and machine learning approaches*: (non) Linear regression can be used in dynamic pricing to model the relationship between the price of the product or service and the demand for it. One can collect data on past sales and prices and use it to train a linear regression model and possibly optimize sales revenues (or profit) based on a parametric model using (non) linear regression [21]. Machine learning algorithms like reinforcement learning (RL, which is typically used to maximize rewards in games) have the potential to solve dynamic pricing problems in various scenarios like varying demand and competitive settings. [22].

C. Cloud Native based Application Development

Cloud-native technology is becoming more popular as businesses seek ways to build and deploy software applications that can handle high traffic, frequent updates, and rapid scaling. Gannon et al. define cloud-native as the practice of developing and deploying highly scalable, resilient, and secure cloud-based software [23]. Cloud has popularized microservices that divides a larger application into smaller, more manageable services that can be built, deployed, and scaled independently. Kubernetes is an open-source container orchestration platform that automates the deployment, scaling, and management of containerized applications [24]. The current academic literature lists some interesting case studies conducted with cloud native technology. Using technologies such as microservice architecture, event-driven architecture, domain-driven design, and containerization, Pan et al. schedule CPU and GPU resources for their cloud-native online judge system [25]. A cloud-native smart operation management platform

architecture is proposed by Lang et al., which would centralize the technical architecture and data interaction of various operation subsystems [26].

D. Research Gaps

Although several studies have been conducted on pricing strategies for e-commerce, there is a lack of research that specifically addresses real-time pricing systems that can adapt to changing market conditions and customer types in real-time. There is also a lack of studies that examine the technical aspects of implementing real-time pricing systems, such as the selection of appropriate technology platforms, pricing algorithms and building reusable components.

III. METHOD

For this work, we use DSRM (Design Science Research Methodology), a research paradigm that emphasizes the generation and validation of prescriptive knowledge. Design science is a serious research approach for creating solutions to real-world engineering issues. As contrasted with how things are in the natural sciences, DSRM is concerned with how things ought to be, that is, with creating artifacts to accomplish things. The steps of the DSRM [27] process include problem identification and motivation, specification of the solution's objectives (requirements), design and implementation, demonstration, evaluation, and communication. Algorithms, user interfaces for computers and people, design approaches (including process models), and languages are among the kinds of artifacts for which design science research is often used. For this work, problem identification and motivation for a real-time pricing system have already been established (in the first section), for the next few sections, we are going to cover requirement, design, implementation, result analysis and will provide conclusion.

IV. REQUIREMENTS

A. Functional Requirement

FR1: A real-time pricing system should provide a suitable price, as per the business need, that optimizes revenue, customer experience or stakeholder's expectations.

FR2: Price should be able to be derived from one or more factors: market factors (demand, competition prices), inventory level, and possibly different customer segments.

B. Non-functional Requirement

NFR1: Configurability/Reusability - Depending on the product type and its current state in the product lifecycle, more than one pricing logic may be configured to run; i.e., the framework should have the ability to switch to different pricing logic.

NFR2: Response latency - Pricing response should be quick, and there should be very minimal deterioration in the price lookup.

NFR3: Reactive Latency - System should react to changes in factors listed in FR2 in an automated fashion within an acceptable time-frame.

NFR-4: Asses the aspects of availability, scalability, and infrastructure cost.

V. DESIGN

This section explains the theoretical design for a real-time pricing system that meets the functional requirement established in the previous section, as well as how it paves a way out for NFRs.

A. Typical Setup of Pricing-module in an e-commerce Application

In a typical e-commerce setup, the user interacts with the catalogue service, which sends them a list of products along with information like the price and the product's description. A product may have a static (fixed) price or be dynamically priced. When there is static pricing, a special database created just for this purpose provides the price data to the catalogue service. When there is dynamic pricing, the prices might need to be calculated on the fly. The catalogue service should contact the pricing service to learn the item's price. Fig. 1 provides an illustration of this variation. Even though dynamic pricing can automate the price generation process, a trivial implementation may result in longer "price-response" latencies and higher computational costs.

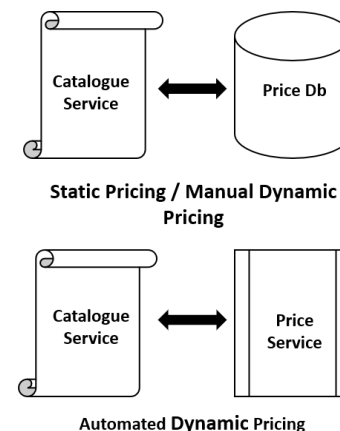


Fig. 1. Static vs dynamic pricing implementation.

B. Pricing Service - High Level Design

The functional requirements (FR1 and FR2) served as our inspiration for the high-level design of the pricing service that is presented below (Fig. 2). The pricing model is in charge of determining the prices. The model may be dependent on inventory, demand, and price competition. By depicting the components in Fig. 2 as distinct parts, we suggest that they can all function concurrently and prefetch/prebuild all the input required by the pricing model. The use of a pricing cache eliminates the need to invoke the pricing model, as it serves cached the pricing results produced by the model to the catalogue service.

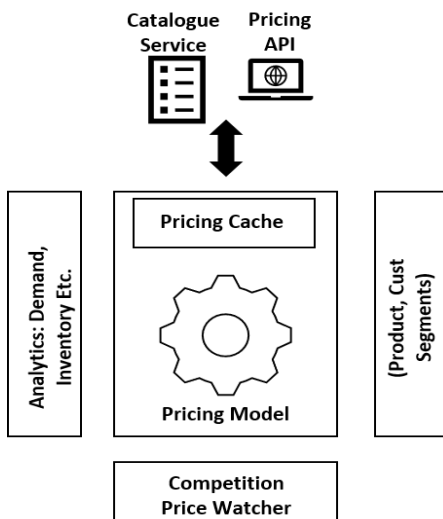


Fig. 2. HLD for real-time pricing system.

C. Component Design

1) *Competition watcher*: The competition watcher component has the responsibility of routinely monitoring and reporting changes in competition prices. The interval at which the “competition watcher” can detect a change in competition prices will determine how quickly the real-time pricing system responds to market prices. A trivial design for this component would be to poll the market for prices at regular intervals. Better reactive latency (FR3) would result from shorter polling intervals, but at the cost of higher computational costs. An ideal polling frequency would be determined by considering how frequently the competition changes their prices. Let us say a competitor increases the price of her product at a rate of λ , Poisson's theorem can be used to calculate the probability that the price (PCP_t) will change during the period t. (shown in Eq.(1)).

$$PCP_t = 1 - e^{-\lambda t} \quad (1)$$

PCP_t becomes an input configuration parameter for the system for a given competitor; let us call it the desired probability (to detect the competition price change, denoted by DP). This helps figure out the polling interval for which the probability of competition price change is configured.

$$Polling\ Interval, T = \frac{\ln(\frac{1}{1-DP})}{\lambda} \quad (2)$$

Initially rate λ is unknown and can be trivially calculated by polling at a small period during the initial setup phase. Once λ is known, Polling interval T can be found using Eq. 2, and should be used to poll the prices. Should competition later increase the frequency of price change, it would be seen by the competition watcher that, number of price changes in the interval T, is more than 1/DP, and that would trigger a re-calculation of λ . When there are fewer samples than 1/DP in the competition prices, this means the competition has decreased their price change rate, and a new polling interval can be calculated using Eq. (2) again.

2) *Demand watcher*: Unlike in competition, where price changes can occur all at once and must be monitored at short intervals, the demand can be calculated once, for a specific time period, such as the previous hour or day. There could be two potential indicators of demand: 1) the price request rate and 2) the item sale rate. However, our understanding is that, sale is also a function of price, and price is variable; thus, in this work, we only consider the first one, i.e., the rate of incoming requests, as an indicator of demand. The architecture influences demand estimation procedure; however, the two most common approaches are:

- *Transaction database query*: Transaction databases are mostly relational databases with transaction time as an index. Using an SQL query, it is trivial for such a database to get the number of transactions in a certain time frame.
- *Transaction stream accumulators*: Another approach is to stream the transaction as it occurs and store the summary in a database table. This method is useful when there is no database from which to obtain information.

3) *Other analytics*: The pricing models (many of it) need analytics data, for example sale-probability, customer segments, existing inventory (or projection of upcoming replenishment). Most of the analytics do not change very frequently and hence computing them once (say daily) would suffice.

D. Pricing Models

1) *Rule base pricing*: This style of pricing involves the dynamic adjustment of a product's price in accordance with a predetermined set of rules. Such a rule may be formed by a combination of supply and demand, time of day, day of the week, and supply and demand alone. One straightforward illustration of this is to raise the price of the product during periods of high demand and lower it during periods of low demand. The model can react instantly to changes in market factors by automatically changing prices in accordance with a predetermined set of rules. The pricing logic in this method is typically quite simple; the main challenge is retrieving underlying factors like supply and demand.

2) *Competition pricing*: The term "competition-based pricing" refers to a pricing approach in which a company's prices change over time in response to the prices offered by its rivals. This strategy can be used to stay competitive in a market by matching or beating the prices of other companies. The implementation for this model has just one dependency, i.e., access to competition prices.

3) *Linear regression*: This pricing strategy uses linear regression analysis to ascertain the connection between prices and demand for a product. This can assist in identifying a product's ideal price point in order to maximize sales or profits. The first step is to use historical sales and pricing data to train a linear regression model that can forecast demand based on various prices. Different linear regression methods,

such as simple linear regression, multiple linear regression, and sophisticated techniques, can be used to train the model. The product's demand is then predicted at various price points using the model, which is then used to determine the price that will generate the most revenue (or profits).

$$\text{Price} = b_0 + b_1 \cdot \text{Demand} + b_2 \cdot \text{CompPrice} + b_3 \cdot \text{Time} \quad (3)$$

Eq. (3) shows the outcome of linear regression analysis, is an expression that can be programmed into pricing model to calculate prices over the period of the time. In the equation, b_0 represents the baseline price of the product, b_1 represents the elasticity of demand, b_2 represents the effect of competition on price, and b_3 represents the effect of time of day on price.

4) *Linear programming*: Linear programming (LP) is an optimization method in which the objective function and all constraints are linear. Linear programming dynamic pricing is a pricing strategy that employs linear programming to calculate the best price for a product. The dynamic pricing problem can be expressed as a linear program, with the goal of maximizing revenue or profits while keeping prices, demand, and costs in check. The objective function and constraints are expressed as linear equations, with the solution being the set of prices that maximizes the objective function [19] (Eq. (4)) while satisfying all constraints (Eq. (5) to Eq. (9)).

$$Y_{(\text{revenue})} = \sum_{t=1}^{t=T} \sum_{k=1}^2 P_t Q_{tk} S_{tk} \quad (4)$$

Constraints:

$$\forall k \in K, RC_k \geq \sum Q_{tk} \quad (5)$$

$$T_{(\text{target sale})} \geq \sum_{t=1}^{t=T} \sum_{k=1}^2 Q_{tk} S_{tk} \quad (6)$$

$$1 = \sum_{k=1}^2 C_k \quad (7)$$

$$P_t := \{P_t \in \mathbb{R}^n \mid P_{\min} \leq P_t \leq P_{\max}\} \quad (8)$$

$$Q_{tk} := \{Q_{tk} \in \mathbb{Z}^{m \times n} \mid 0 \leq Q_{tk}\} \quad (9)$$

Where:

R is the total expected number of price requests across k customer segments.

K is the customer segments ($K = 2$ for myopic and strategic customers setup),

C_k is the ratio of expected request count in the customer category (k) and total requests.

Q_{tk} is number of price response by the DP engine with price P_t in customer segment k . (This is LP Variable).

P_t is the price point to offer, its range is P_{\min} and P_{\max} .

S_{tk} is the sale probability at price P_t , customer segment k .

E. Pricing Cache

Client price requests for a product many times more frequent than change in market factors hence the price calculated by model can be cached to avoid the model invocation time, which is not just time but also computationally expensive. Pricing cache decouples the

expensive pricing process from servicing the pricing. If cache is not used, price request latency is given by 10, i.e., sum of latencies of model evaluation and sum of all the analytic fetch.

$$Lat_{pr} = Lat_{\text{price-model}} + \sum Lat_{\text{analytics}} \quad (10)$$

By using the cache, the price-request-latency simply becomes the latency of cache lookup (11).

$$Lat_{pr} = Lat_{\text{cache}} \quad (11)$$

Please note that, the pricing cache is considered large enough to hold the calculated pricing results for all the items, and hence cache miss penalties are not considered.

F. Product Module

A product module is a component or module that manages and provides access to product information. This includes information such as the pricing model, minimum and maximum price settings, available inventory, price exploration factor, and any other relevant settings required for the product's pricing model to function. This data is typically stationary and varies infrequently (say, monthly or quarterly). Also note that the product module does not need to be exclusive to pricing system; it can be shared across other e-commerce services (such as catalogue, order etc).

VI. AWS CLOUD IMPLEMENTATION

We will go over the specifics of this work's cloud implementation in this section. Real-time pricing system implementation requires compute, a database, and the ability to react to events. Most cloud service providers can easily meet these requirements, but we have chosen to implement AWS for the proof-of-concept and because it is the most widely used cloud service. Python was used exclusively throughout the work to program everything.

A. A Quick Review of AWS Services

Amazon Elastic Compute Cloud (EC2) is an AWS web service that allows users to rent virtual computers to run their own computer programs. Elastic Container Service (ECS) is a fully managed container orchestration service that allows you to run and scale containerized applications on a cluster of EC2 instances using Docker and Kubernetes. Amazon Elastic Container Registry (ECR) is a fully managed Docker container registry hosted by Amazon Web Services (AWS). Docker images can be safely and saleably stored, managed, and deployed using ECR. EventBridge is a serverless event bus service that connects applications by automating event scheduling and routing from sources such as DynamoDB to destinations such as Lambda functions, Lambda is a serverless computing service that allows you to run code without the need for server provisioning or management. Lambda-powered applications can respond to events and run functions automatically in response to triggers such as changes to data in DynamoDB or an EventBridge-triggered event. Lambda does have some limitations, such as a maximum amount of memory and limited concurrency, which should be considered for scalability. DynamoDB is a fully managed NoSQL key-value and document database that scales to provide millisecond performance. It is an excellent choice for applications that require consistent, fast performance as well as the ability to

manage large amounts of data. Applications can use DynamoDB Streams to react to changes in DynamoDB tables and invoke AWS Lambda to handle the change [28].

B. Component Implementation

1) *Competition watcher*: The goal of the Competition Watcher software is to obtain competitor product pricing information by scraping competitor websites. Web-scraping tools are commonly used for this purpose. Several libraries and frameworks, such as “Beautiful Soup,” [29] “Scrapy,” and “Selenium,” are available for web scraping in the Python programming language. Under Amazon Web Services, there are two possibilities for designing and building this:

a) *Event-bridge-based periodic Lambda invocation*: EventBridge allows you to schedule a Lambda for a specific time (Say every 30 minutes). Lambda executes the web scraping business logic and updates the competition prices in DynamoDB. DynamoDB Stream can then be used to retrigger the pricing process by triggering a lambda function.

b) *Docker workflow with ECS, ECR (running on EC2)*: In this approach, web-scraping logic is executed in a Docker container powered by EC2, ECS, and ECR Services to retrieve competition prices, detect changes, and store in DynamoDB. A DynamoDB stream, like the previous option, can be used to trigger a lambda function for pricing. This process can monitor hundreds of products, making it more cost-effective for large e-commerce setups.

2) *Demand watcher*: The demand watcher watches the ongoing demand periodically, and if there is a change in the demand, it is supposed to trigger the repricing by invoking the lambda function. Demand Watcher implementation is dependent on E-Commerce architecture implementation. Assume that E-commerce architecture has a database table called REQUEST and a REQUEST_TIME index on it. Demand SQL expression is shown in Fig. 3.

```
SELECT count(*) from REQUEST
where REQUEST_TIME > Now()-POLL_WINDOW
```

Fig. 3. Demand calculation in a period using existing RDBMS DB.

In many use cases, the above approach would work, except when users want more real-time demand tracking. The alternative is to use streaming analytics, for this purpose, first, the incoming price requests can be posted to a Kinesis analytics stream. This stream has a consumer, a lambda function that is configured with batch triggers on Kinesis updates. The lambda analyses and records the demand information in real-time into a DynamoDB. Dynamodb stream then triggers the pricing function to recalculate the price for the new demand (Fig. 4).

3) *Customer segmentation and sale probability*: For the analytics that do not change in real-time, like sale-probability and customer-segments, in order to be cost-effective, the framework would compute these periodically (say, once a day). We use the price request stream and order stream, and register lambda triggers (with the largest possible batch

size/window, i.e., 5K records and 5 minutes). Both the lambdas add/update the entries in the Analytics table (Fig. 5). Their logic updates and aggregates the incoming stream in the format shown in Table I. This serves as the common source for segmentation and sale probability calculators. Customer segmentation is done by aggregating the analytics table by customer-id and the ratio of purchases to requests. Those with a higher purchase to request ratio are classified as myopic, while the rest are classified as strategic. Table II shows DynamoDB schema for customer segmentation table. Sale-probability is calculated as sale to request ratio for each segment and at each price points. Table III displays the schema for DynamoDB table for the same. Please Note, since DynamoDB is a KV store, probability data is stored in json format, and pricing-model is expected to deserialize that before using.

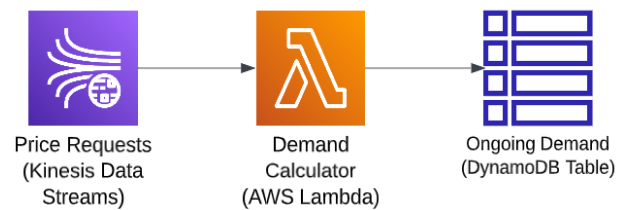


Fig. 4. Realtime demand-data accumulation.

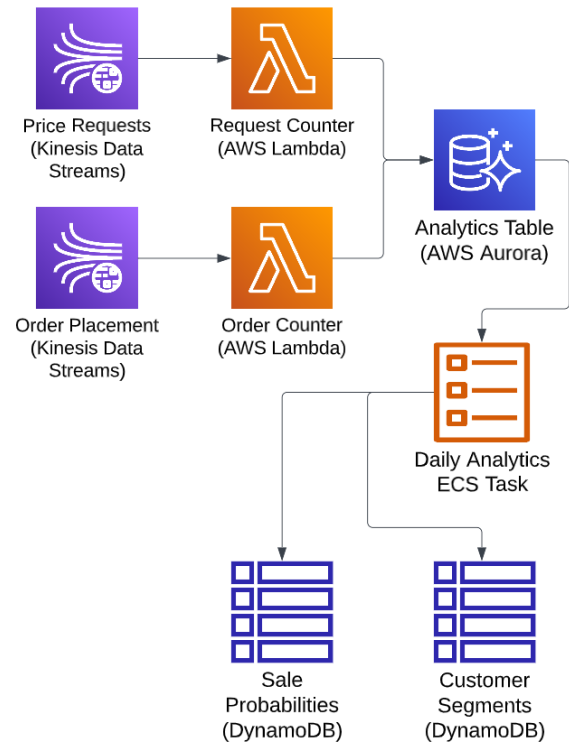


Fig. 5. Analytics: sale probability & customer segmentation.

TABLE I. ANALYTICS TABLE SCHEMA

Date	String	String	Float	Bool
Date	Cust-id	Product	Price	Purchased

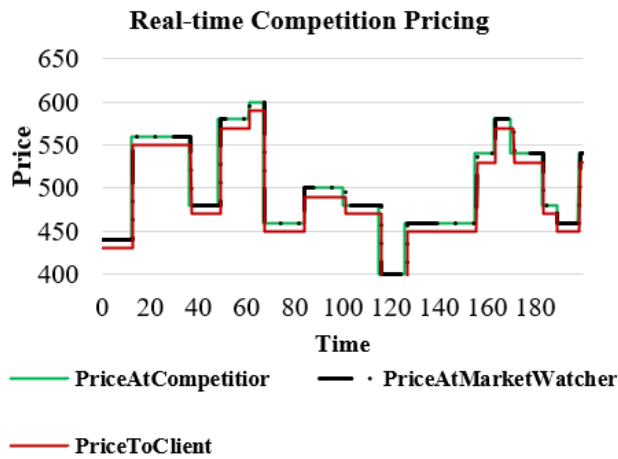


Fig. 8. Competition pricing tracking.

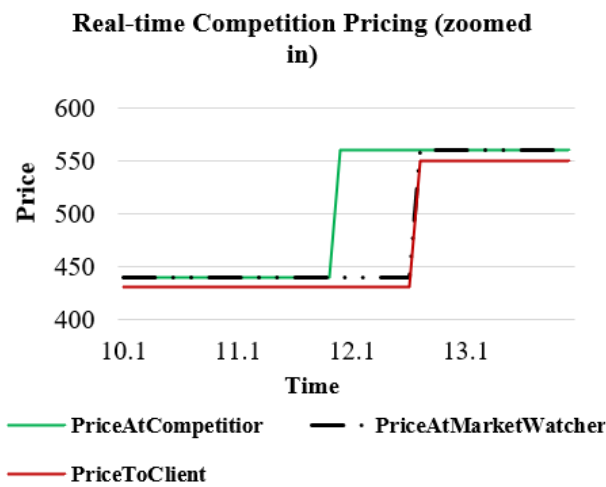


Fig. 9. Competition response latency.

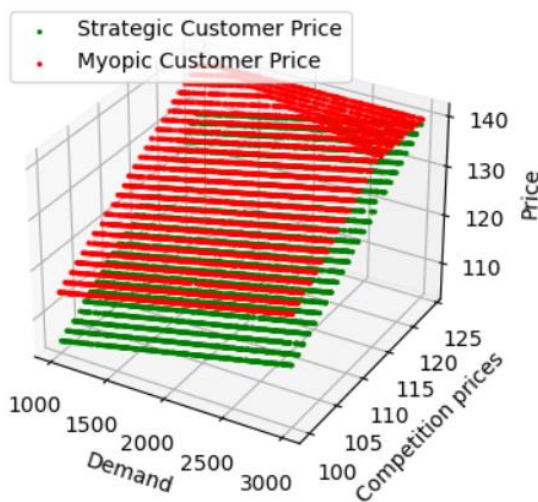


Fig. 10. Rule-based dynamic pricing.

Fig. 11 and 12 are the simulation outcomes for the linear programming-based pricing model. The solution for objective function (Eq. (4)) and constraint (Eq. (5) to Eq. (9)) can change the prices for changing demand or target sale-quantities and

maximize the revenue under the different circumstances. Fig. 11 shows optimal prices for different values of target sale-quantity, and Fig. 12 shows the same different demand values.

Fig. 13 and Fig. 14 explain the content and performance of the price cache. Fig. 13 is a snippet taken from AWS console depicting the content of price cache. Key shown in the Fig. 13 is the product-id, and the json value provides the price values for myopic and strategic customer.

It also has a default value, which is useful in situation when pricing-model is yet to provide any pricing value. Fig. 14 shows our typical distribution of price-request latency when using DynamoDB as a pricing cache. The latency is single-digit millisecond latency and comparable to static-pricing scenarios.

Linear Programming based Pricing for Target Sale Variations

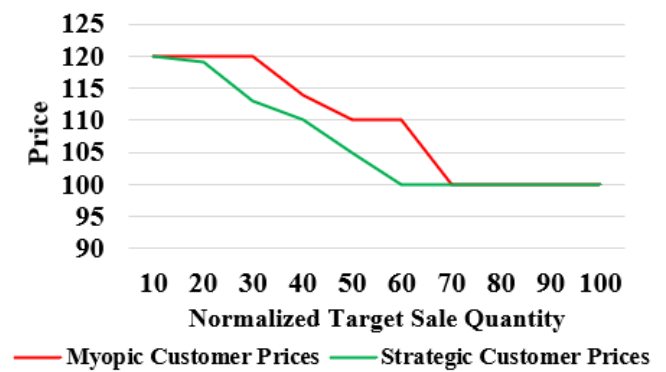


Fig. 11. Linear programming-based pricing (1).

Linear Programming based Pricing for Demand Variations

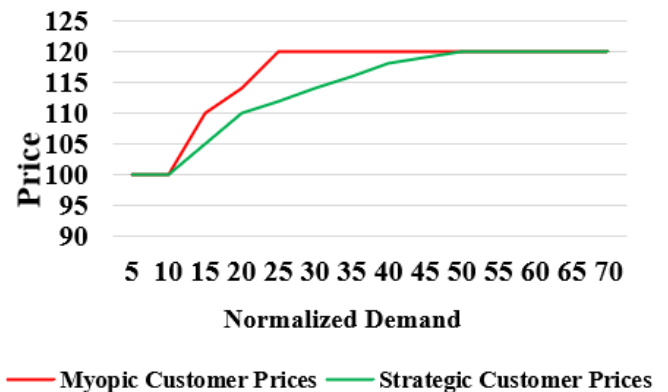


Fig. 12. Linear programming-based pricing (2).

key1	json
np374	{"default_price": 3500, "prices": [[[1, 3840]], [[1, 3835]]]}
np623	{"default_price": 7000, "prices": [[[1, 6330]], [[1, 6325]]]}

Fig. 13. Typical price cache values.

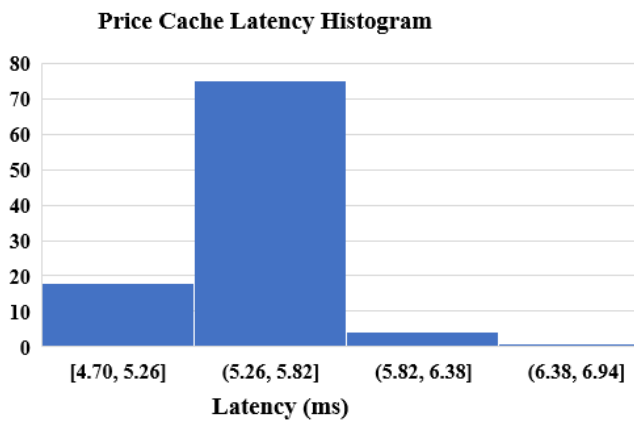


Fig. 14. Price cache latency distribution.

VIII. CONCLUSION

A. Summary and Discussion

Real-time pricing can help e-commerce owners; however, implementation requires careful handling of multiple requirements in order to perform dynamic pricing in real-time. Through this work, we defined functional and non-functional requirements and used them to guide our system design process. We created a pricing framework that can potentially provide the most suitable price for various products at different stages of their product lifecycle while taking a variety of market factors such as demand and competition into consideration (satisfying functional requirements FR1 and FR2).

The components (e.g., demand watcher, competition price fetcher, cache) are reusable, concurrent (designed with separation of concerns in mind), and allow plug-and-play of different pricing logic through the defined database schema (satisfies NFR1). By the virtue of design decision, price response latency (NFR2) equals to a cache-lookup which is similar to systems without dynamic pricing. The cache latency can further be improved (if required) by using DynamoDB-DAX [30]. It has been demonstrated in the result that, system response to any price change is within a single second, an acceptable latency in the context of e-commerce (NFR3). The cloud native building blocks (DynamoDB, lambda, etc.) inherently provide availability and scalability to our design and with the pay-as-you-go model for cost, cloud-native design satisfies the last one, NFR-4.

B. Research Contributions

Real-time e-commerce pricing is a relatively unexplored academic field. In the research gap analysis, we did not find any implementation-specific work, which emphasizes the timeliness of the pricing. As our first research contribution, we have been able to apply existing knowledge of 'dynamic pricing' to methodologically specify, design, and implement a real-time pricing system.

Other contribution of this research is the demonstration of cloud-native design principles for the implementation of real-time pricing. The framework provides, unlike anything else in the published literature, a low-cost (pay as you go) and scalable solution for real-time pricing by making use of cloud

implementation methodologies. By providing the ability to plug-and-play a pricing model for different e-commerce use-cases, the framework does not enforce 'one-size-fits-all', but rather promises flexibility while maintaining the reusability of building blocks.

C. Future Work

The primary goal of this work is to create a framework for a real-time pricing system, which has been demonstrated in the implementation and results section. However, because real-time pricing for e-commerce is going to evolve further, there are a few opportunities for future work:

- There is an opportunity to improve the analytics described by performing some finer optimizations. To give an example, advanced statistical techniques can be used to provide more accurate results for sale-probability estimation. Similarly, sophisticated forecasting techniques can be utilized to predict demand, which will help price the product more efficiently over the long run.
- Another potential area of research can be conducted on pricing models and their selection in the context of real-time pricing and cloud native. Models can be fast or time-consuming and may have different memory and compute requirements, leading to deployment choices such as serverless or containers.

REFERENCES

- [1] Y. Narahari, C. Raju, K. Ravikumar, and S. Shah, "Dynamic pricing models for electronic business," *Sadhana*, vol. 30, pp. 231–256, 2005.
- [2] D. Elreedy, A. F. Atiya and S. I. Shaheen, "Multi-Step Look-Ahead Optimization Methods for Dynamic Pricing With Demand Learning," in *IEEE Access*, vol. 9, pp. 88478–88497, 2021, doi: 10.1109/ACCESS.2021.3087577.
- [3] S. Christ, *Operationalizing Dynamic Pricing Models: Bayesian Demand Forecasting and Customer Choice Modeling for Low Cost Carriers*. 2011.
- [4] G. Mehra, "Real-time Pricing Affordable for Smaller Merchants," *Practical Ecommerce*, Jan. 22, 2017. <https://www.practicalecommerce.com/Real-time-Pricing-Affordable-for-Smaller-Merchants>.
- [5] "IDC Forecasts Worldwide," IDC: The premier global market intelligence company. <https://www.idc.com/getdoc.jsp?containerId=prUS48208321>.
- [6] Z. Mahmood, "Cloud computing for enterprise architectures: concepts, principles and approaches," in *Cloud computing for Enterprise architectures*, Springer, 2011, pp. 3–19.
- [7] M. Wu, X. Ding, and R. Hou, "Design and implementation of B2B E-commerce platform based on microservices architecture," in *Proceedings of the 2nd International Conference on Computer Science and Software Engineering*, 2019, pp. 30–34.
- [8] S. Athreya, S. Kurian, A. Dange, and S. Bhatsangave, "Implementation of Serverless E-Commerce Mobile Application," in *2022 2nd International Conference on Intelligent Technologies (CONIT)*, 2022, pp. 1–5.
- [9] "Cloud Computing Services - Amazon Web Services (AWS)," *Amazon Web Services, Inc.* <https://aws.amazon.com/>.
- [10] "Empower Your Business in USA & Canada with Alibaba Cloud's Cloud Products & Services," *Empower Your Business in USA & Canada with Alibaba Cloud's Cloud Products & Services.* <https://www.alibabacloud.com>.

- [11] "Start and grow your e-commerce business - 3-Day Free Trial," *Start and grow your e-commerce business - 3-Day Free Trial*. <https://www.shopify.com/?ref=mile-high-themes>.
- [12] "What Is Magento Ecommerce And Why Should You Use It?," *World's #1 POS for Magento*, Mar. 04, 2021. <https://www.magestore.com/blog/what-is-magento/>.
- [13] I. Yeoman, "The history of revenue and pricing management – 15 years and more," *Journal of Revenue and Pricing Management*, vol. 15, no. 3–4, pp. 185–196, Jun. 2016, doi: 10.1057/rpm.2016.36.
- [14] Kokkoris, I., & Lemus, C. (2022). *Research Handbook on the Law and Economics of Competition Enforcement*. E-CONTENT GENERIC VENDOR. <https://books.google.co.in/books?id=jqKCEAAAQBAJ>.
- [15] T. Wang *et al.*, "A framework for airfare price prediction: a machine learning approach," in *2019 IEEE 20th international conference on information reuse and integration for data science (IRI)*, 2019, pp. 200–207.
- [16] Gallego, G., & Topaloglu, H. (2019). *Revenue Management and Pricing Analytics*. Springer New York. <https://books.google.co.in/books?id=YFpDwAAQBAJ>.
- [17] S. Saharan, S. Bawa, and N. Kumar, "Dynamic pricing techniques for Intelligent Transportation System in smart cities: A systematic review," *Computer Communications*, vol. 150, pp. 603–625, 2020.
- [18] R. Phillips, *Pricing and Revenue Optimization*. Stanford University Press, 2005. [Online]. Available: <https://books.google.co.in/books?id=Xi17Xx9rD9wC>.
- [19] A. Kumari and B. R. K, "Design of a Real-Time Pricing System for E-commerce," *International Journal of Computer Theory and Engineering*, vol. 15, no. 1, pp. 46–53, 2023.
- [20] S. Kedia, S. Jain, and A. Sharma, "Price optimization in fashion e-commerce," *arXiv preprint arXiv:2007.05216*, 2020.
- [21] A. V. Den Boer, "Dynamic pricing and learning: historical origins, current research, and new directions," *Surveys in operations research and management science*, vol. 20, no. 1, pp. 1–18, 2015.
- [22] A. X. Carvalho and M. L. Puterman, "Dynamic pricing and reinforcement learning," in *Proceedings of the International Joint Conference on Neural Networks, 2003.*, 2003, vol. 4, pp. 2916–2921.
- [23] D. Gannon, R. Barga, and N. Sundaresan, "Cloud-native applications," *IEEE Cloud Computing*, vol. 4, no. 5, pp. 16–21, 2017.
- [24] J. Lee and Y. Kim, "A Design of MANO System for Cloud Native Infrastructure," *2021 International Conference on Information and Communication Technology Convergence (ICTC)*, Jeju Island, Korea, Republic of, 2021, pp. 1336–1339, doi: 10.1109/ICTC52510.2021.9620858.
- [25] G. -C. Pan, P. Liu and J. -J. Wu, "A Cloud-Native Online Judge System," *2022 IEEE 46th Annual Computers, Software, and Applications Conference (COMPSAC)*, Los Alamitos, CA, USA, 2022, pp. 1293–1298, doi: 10.1109/COMPSAC54236.2022.00204.
- [26] H. Lang, H. Tian, D. Li, Z. Niu and L. Wen, "Design of A Cloud Native-Based Integrated Management Platform for Smart Operation of Multi-Business Buildings," *2022 14th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC)*, Hangzhou, China, 2022, pp. 169–173, doi: 10.1109/IHMSC55436.2022.00047.
- [27] P. Offermann, O. Levina, M. Schönherr, and U. Bub, "Outline of a Design Science Research Process," in *Proceedings of the 4th International Conference on Design Science Research in Information Systems and Technology* (pp. 1–11). 2009. doi: 10.1145/1555619.1555629.
- [28] Cloud Products," *Amazon Web Services, Inc.* <https://aws.amazon.com/products/>.
- [29] L. Richardson, "Beautiful Soup: We called him Tortoise because he taught us.," *Beautiful Soup: We called him Tortoise because he taught us.* <https://www.crummy.com/software/BeautifulSoup/>.
- [30] "In-memory acceleration with DynamoDB Accelerator (DAX) - Amazon DynamoDB," *In-memory acceleration with DynamoDB Accelerator (DAX) - Amazon DynamoDB.* <https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/DAX.html>.

A Deep Learning Approach for Sentiment Classification of COVID-19 Vaccination Tweets

Haidi Said¹, BenBella S. Tawfik², Mohamed A. Makhoulouf³

Department of Information Systems-Faculty of Computers and Informatics, Suez Canal University, Ismailia, Egypt^{1,2,3}

Abstract—Now-a-days, social media platforms enable people to continuously express their opinions and thoughts about different topics. Monitoring and analyzing the sentiments of people is essential for governments and business organizations to better understand people's feelings and thoughts. The Coronavirus disease 2019 (COVID-19) has been one of the most trending topics on social media over the last two years. Consequently, one of the preventative measures to control and prevent the spread of the virus was vaccination. A dataset was formed by collecting tweets from Twitter for over a month from November 13th to December 31st, 2021. After data cleaning, the tweets were assigned a positive, negative, or neutral label using a natural language processing (NLP) sentiment analysis tool. This study aims to analyze people's public opinion towards the vaccination process against COVID-19. To fulfil this goal, an ensemble model based on deep learning (LSTM-2BiGRU) is proposed that combines long short-term memory (LSTM) and bidirectional gated recurrent unit (BiGRU). The performance of the proposed model is compared to five traditional machine learning models, two deep learning models in addition to state-of-the-art models. By comparing the results of the models used in this study, the results reveal that the proposed model outperforms all the machine and deep learning models employed in this work with a 92.46% accuracy score. This study also shows that the number of tweets that involve neutral, positive, and negative sentiments is 517496 (37%) tweets, 484258 (34%) tweets, and 409570 (29%) tweets, respectively. The findings indicate that the number of people carrying neutral sentiments towards COVID-19 immunization through vaccines is the highest among others.

Keywords—COVID-19 vaccination; sentiment analysis; Twitter; machine learning; deep learning; natural language processing (NLP)

I. INTRODUCTION

Recently, the massive amount of data generated by users on social media platforms and advances in computational power provide a strong impetus for sentiment analysis to develop and become the top research field under natural language processing (NLP) [1]. Sentiment analysis (SA), also known as sentiment or opinion mining, is the computational analysis of people's opinions and emotions about different topics [2]. This automatic analysis plays a significant role in extracting useful insights for decision-makers in various application domains. SA is a research field that aspires to understand and extract the sentiment of unstructured content. This content can be text, audio, image, or video. Following the technical definition of sentiment analysis introduced by [3], in this sentence, "Mark likes the camera of Samsung S10", here Mark acts as the opinion holder expressing his positive sentiment towards the

aspect camera of entity Samsung S10. SA uses NLP, and machine learning (ML) techniques to extract the subjective information of a document and classify it according to its opinion orientation or polarity which can be positive, negative, or neutral. Subjective information carries only sentiments about a particular topic or entity, while objective text contains facts with no sentiments [4]. For instance, "Star Wars is an awesome movie". This sentence has a sentiment (awesome); thus, it is a subjective text, not an objective.

The explosive growth of social media applications (e.g., Facebook, Instagram, Twitter) witnessed today opens the door for a continuous stream of opinions and thoughts [5]. This Internet-generated content enables people to communicate with each other by sharing their sentiments and how they feel in the form of opinions or reviews about any topic. This consequently leads to the generation of an enormous amount of unstructured data. Business organizations need to process and analyze this data to support them in their decision making by gaining deeper insights into user sentiments, which will accordingly improve customer satisfaction. Furthermore, the study of public opinion about a specific topic or issue is very important to governments for the study of human activity and behavior as well as crisis management. Sentiment analysis provides governments with the valuable information necessary to take true actions in time.

Monitoring and analyzing the opinions and sentiments for extracting valuable information from them manually is a challenging task, in terms of time and exerted effort, due to the boom of various internet-based applications and sites. So, automated sentiment analysis or opinion mining systems are required to overcome this problem. Several machine learning and natural language processing-based approaches have been proposed to analyze these sentiments. However, recently deep learning-based approaches have shown significant results and higher performance in the field of sentiment analysis [6].

COVID-19 has dramatically affected every aspect of our lives and was declared a worldwide pandemic by the World Health Organization (WHO) [7]. Vaccination against COVID-19 was introduced as an urgent solution to prevent the spread of the disease among people and reduce death rates [8]. Therefore, analyzing the sentiments of people towards COVID-19 vaccination is essential for governments and health ministries to understand the public mood of people regarding the vaccination process against that virus. Consequently, vaccine campaigns can be developed targeting the hesitant and anti-vaccine groups of people to raise their awareness about the pivotal role of vaccines and boosters in containing the pandemic.

A long short-term memory (LSTM) is an extension of recurrent neural network (RNN). LSTMs can handle long-term dependencies between time steps of sequential data [9]. A Bidirectional gated recurrent unit (BiGRU) is used to learn features in both directions to produce a more meaningful output. This study combines the strengths of LSTM and BiGRU models to enhance the accuracy of sentiment classification. To the best of our knowledge, several studies navigated the public views regarding the COVID-19 vaccines before November 2021 [10-13]. However, many important events related to COVID-19 vaccines have occurred after this date, like the emergence of the Omicron variant [14] as well as boosters. Therefore, to fill this gap, we analyzed the tweets related to COVID-19 vaccination during this period of time to fully understand the public mood and perspectives towards COVID-19 vaccination acceptance or hesitancy. The main contributions of this study are as follows:

- Creating and preparing a large dataset of 4,093,986 tweets related to COVID-19 vaccination and vaccines for the sentiment classification task.
- Visualizing and analyzing the sentiments of people about the topic of COVID-19 vaccination.
- Proposing a hybrid approach to determine people's thoughts and opinions towards COVID-19 vaccination. The collected dataset was labeled using Valence Aware Dictionary and sEntiment Reasoner (VADER) lexicon-based approach. Moreover, Conventional machine learning classifiers were trained on the annotated dataset and tested on unseen data to evaluate their performance.
- Improving the classification accuracy by introducing an ensemble model (LSTM-2BiGRU), combining long short-term memory and bidirectional gated recurrent unit.
- A performance comparison of the proposed model with traditional machine learning classifiers, two individual deep learning models (LSTM, BiGRU), and state-of-the-art models to identify the top performer model on our collected dataset.

The rest of the paper is organized as follows: Section II provides an overview of the related work to our study. The subsequent section describes the dataset and methods used in the study. The fourth section presents and discusses the results of the conducted experiments. Finally, the last section concludes and summarizes the paper.

II. RELATED WORK

COVID-19 has seriously affected the daily lives of people in different sectors of life [15]. As a result, several studies were conducted to analyze people's sentiments towards COVID-19 and its related aspects. For example, the study in [16] focused on analyzing and understanding the sentiments of the Canadian people towards social distancing related to COVID-19 using a hybrid approach. They used SentiStrength, a lexicon-based approach, with Support Vector Machine (SVM), a machine learning approach, for sentiment analysis. To perform this study, they collected Twitter data for one month from an open-

source publicly available IEEE website. They applied the SVM algorithm by splitting the dataset into 80% for training and 20% for testing the data. Their experiment resulted in 71% accuracy for sentiment classification. By excluding the neutral sentiments from data, the accuracy of the model rose to 81%. Additionally, the accuracy reached 87% by reducing the test data to 10%. The results showed that 40% of the Canadian people have neutral sentiments, followed by 35% expressed negative sentiments and only 25% have positive sentiments towards social distancing. One of the limitations they faced during the study was the insufficient number of training examples.

The authors in [17] compared the effectiveness of two machine learning models, Naïve Bayes (NB), and logistic regression (LR), on Twitter data of varying lengths collected from February to March of 2020. The collected data was cleaned and prepared for analysis using the R programming language and its related packages. NB achieved 91% accuracy with short tweets (less than 77 characters), while the accuracy obtained using the logistic regression model was 74% with short tweets. However, the performance of both models (Naïve Bayes and logistic regression) significantly decreased to 57% and 52%, respectively, in the case of longer tweets. For future work, the authors of this work recommended extending this study with more data and additional methods. This will consequently help policy makers and businesses in understanding the public sentiment and perspectives for making the right decisions in time.

The study in [18] performed a sentiment classification task on 3090 tweets collected from 23 March to 15 July 2020. The collected text was labeled as fear, sad, anger, and joy. Each sentiment was mapped from 0 to 3 (fear:0, sad:1, anger:2, joy:3). They applied a new deep learning model called Bidirectional Encoder Representations from Transformers (BERT). They compared it to three other classification algorithms: LR, SVM, and LSTM. For all the classifiers, they split the dataset into 85% for training and the remaining 15% was used for testing purposes. Their experiment showed that the BERT model outperformed the other models by having 89% accuracy, whereas the other three models scored 75%, 74.775%, and 65%, respectively.

In [19], logistic regression and LSTM learning models were used along with two different word embedding techniques: CountVectorizer and TfidfVectorizer to analyze and study people's attitude towards COVID-19. The dataset used in this study consists of tweets collected from 03/16/2020 to 04/14/2020. The author used only original tweet and label columns from the given columns in the dataset to investigate people's reactions towards COVID-19 during this period of time. The data was initially labeled as five classes (positive, extremely positive, negative, extremely negative, and neutral) but the author manually grouped them into three classes (positive, negative, and neutral) for more precise classification. Then in the featurization stage, the original tweets were pre-processed and vectorized using the previously mentioned two vectorization methods. After that, logistic regression and LSTM models were used to classify the sentiments of the given COVID-19 related tweets. The LSTM/TF-IDF model achieved the best result with an F1 score of 0.85.

The authors in [20] used the keyword “COVID-19” to search tweets on Weibo, one of the largest Chinese social media platforms, and collected 45,987 tweets. They used a lexicon-based approach for sentiment analysis by matching their dataset with words in the sentiment dictionary. BosonNLP sentiment dictionary was selected in this study as it supports the Chinese language and is constructed from millions of labeled data, including Weibo, news, forum data, and others. For future work, they suggest combining their approach with deep learning (DL) or random forest (RF) classifier to achieve more accurate sentiment prediction results. Kuar et al. devised an algorithm called Hybrid Heterogeneous Support Vector Machine (H-SVM) that performs sentiment classification on COVID-19 tweets in study [21]. They compared the performance of their proposed algorithm with RNN and SVM. The H-SVM model outperformed the other two models in terms of precision (86%), recall (69%), and F1 score (77%).

In [22], the authors proposed a feature extraction technique by concatenating bag-of-words (BoW) and term frequency – inverse document frequency (TF-IDF). They used five machine learning models, such as RF, SVM, decision tree (DT), XGBoost classifier, extra tree classifier (ETC), and LSTM to analyze sentiments of COVID-19 tweets. The results showed that ETC outperformed all other models using their proposed feature extraction technique with an accuracy score of 0.93. The authors stated that the poor performance of the LSTM deep learning model is due to the small dataset. The authors of paper [23] conducted sentiment analysis on 11,960 tweets in the UK related to COVID-19 using three different ensemble models. The dataset was labeled manually by three independent annotators. The stacking classifier (SC) achieved the highest F1 score with 83.5% in comparison with the voting classifier (VC) and bagging classifier (BC) with 83.3% and 83.2% scores, respectively.

Researchers in this study [24] applied sentiment analysis to tweets about COVID-19 different vaccines. They found that 48.49% were neutral, 33.96% were positive, and 17.55% had negative feelings about the coronavirus vaccines. The study used LSTM and bidirectional LSTM (BiLSTM) deep learning models to predict the sentiments of the tweets. According to the study, the resulting accuracy is 90.59% for LSTM and 90.83% for Bi-LSTM. The authors of paper [25] combined convolutional neural network (CNN) with BiLSTM to classify the polarity of COVID-19 tweets as positive, negative, or neutral. They used the hybrid CNN-BiLSTM model with two word embedding pre-trained models: FastText and Global Vectors for Word Representation (GloVe) to obtain higher accuracy levels. The authors conducted experiments on a dataset of 40000 tweets collected from Twitter. The CNN-BiLSTM model with FastText yielded 99.33% accuracy, which is higher than the accuracy score of CNN-BiLSTM with GloVe (97.55%).

The authors in [26] proposed a hybrid feature extraction approach for COVID-19 tweets sentiment classification. They combined TF-IDF with FastText and GloVe to increase the performance of classification. This study applied seven machine learning classifiers in addition to one deep learning model (CNN). It was observed that the best performance was achieved by SVM using the TF-IDF with FastText word

embedding proposed technique. This study [27] presented a BiLSTM model for public opinion analysis of COVID-19-related discussions on social media. They performed sentiment classification using four different scenarios on three datasets collected from Twitter and Reddit platforms. One of the limitations they encountered is that the BiLSTM method is time-consuming. Also, they recommended applying multilingual sentiment analysis to get a global understanding of COVID-19 sentiments.

A new hybrid deep learning method is proposed in [28] to find the general sentiment of people in eight countries around the world. Their method is based on the fusion of four deep learning and one machine learning model. Their proposed model shows superior performance over the other models used in the study with 85.8% accuracy. In [29], the authors performed sentiment analysis on Nepali COVID-19 tweets using an ensemble CNN model. They proposed three different CNN models with three feature extraction techniques such as FastText (ft), domain-specific (ds), and domain-agnostic (da). In addition, they combined the three CNN models to form an ensemble CNN model. Their model achieved an accuracy of 68.7% for sentiment classification of tweets in the Nepali language.

This research [30] used three machine learning classifiers to predict people’s awareness of COVID-19 preventative measures in Saudi Arabia. For this purpose, they prepared a dataset of Arabic tweets related to COVID-19 preventative procedures. They applied four feature extraction techniques (unigram, unigram TF-IDF, bigram, and bigram TF-IDF) with each classifier. Their results show that the SVM classifier with bigram TF-IDF achieved the highest accuracy of 85%. Furthermore, the people in the south region of Saudi Arabia had the highest level of awareness, where they reacted positively towards COVID-19-related precautionary measures. The sentiments of COVID-19 tweets were analyzed using the BERT model in research [31]. They created two datasets by collecting tweets from India and the whole world from January 20, 2020, to April 25, 2020. The accuracy of their proposed model is approximately 94%.

III. MATERIALS AND METHODS

This study introduced an ensemble model based on a hybrid approach to analyze the sentiments of people towards COVID-19 vaccination. We collected a large dataset containing tweets related specifically to the COVID-19 vaccines and vaccination process. Tweets were cleaned and pre-processed before feeding them to the machine and deep learning models. VADER, a lexicon-based approach was used to assign positive, negative, or neutral polarity to the collected tweets. Afterwards, the annotated dataset was split for training and testing purposes. The lexicon-based approach is integrated with machine learning and deep learning-based approaches for sentiment classification. Accuracy, precision, recall, and F1 score were used to analyze the performance of the models. Fig. 1 depicts the architecture of the proposed methodology.

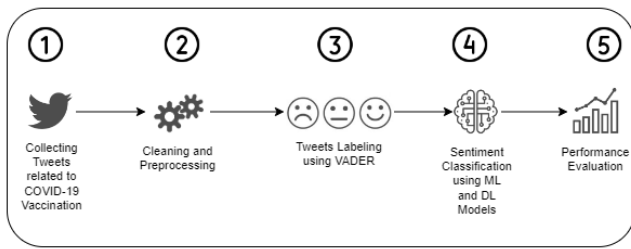


Fig. 1. Architecture of the proposed methodology.

A. Data Collection

To extract the COVID-19 vaccination-related tweets from Twitter, we applied for a Twitter developer account for the purpose of doing academic research. Accordingly, the Twitter developer team approved our request and enabled us to access Twitter’s application programming interface (API), which acts as an intermediary between the developer and the Twitter backend. Ultimately, the Twitter developer team provided us with the required credentials to authenticate access to the Twitter API. We used the Tweepy library in Python to interact with the Twitter API for collecting the required tweets. The data collection process was carried out on a daily basis from 13 November 2021 to 31 December 2021. Different hashtags were used to extract the tweets related to COVID-19 vaccines such as "#vaccinated", "#CovidVaccine", "#CovidVaccination", "#VaccinesWork", "#VaccineSideEffects", "#booster", "#AstraZeneca", "#Pfizer", and "#Moderna". In total, 4,093,986 tweets were extracted about vaccination against the coronavirus disease. The collected dataset shape is 4093986 by 5 columns, including tweet_id, text, favorite_count, created_at, and tweet_location. Table I shows a sample of some tweets from the collected dataset.

TABLE I. A SAMPLE OF TWEETS RELATED TO COVID-19 VACCINATION

Text
IM VACCINATED 🙏
I just got vaccinated on Tuesday and it hurts :(
@Dannyoconnor430 Get vaccinated to protect yourself.
So no need to panic ! Get vaccinated !! Protect yourself and others 😊 #omicron #covid #Valimai https://t.co/PAbdvkAGuA
@VishnuNDTV really a ridiculous rule for domestic passenger who r fully vaccinated.
Majority of mumps cases are among the vaccinated, CDC finds https://t.co/WuWbWy3AY via @nbcnews
@ScottATaylor waiting for the non vaccinated to get wise, so I can go out and have fun again
Saw a kid go up and thank the medical professional who vaccinated him today and my heart remembered the world has good in it
@FlutoShinzawa Good thing they are all double vaccinated.
@ABC Do what you want with your own body...But get vaccinated or lose your job

B. Data Cleaning and Preprocessing

After data collection and exploration, tweet texts were cleaned by removing noise or words that are irrelevant to our experiment. While extracting data from Twitter, all retweets were ignored by applying -filter: retweets to our search query.

In this study, we are interested only in analyzing the sentiments of tweets, so all the columns other than text were discarded. This led to having 4,093,986 rows and 1 column. In addition, duplicated tweets and rows with null values were removed. After these operations, the ultimate shape of the dataset became 1,411,324 rows by 1 column. Subsequently, irrelevant data elements to the sentiment analysis process were removed including hashtags (#), hyperlinks, mentions (@username), special characters, numbers, and punctuation marks. Table II shows some tweets before and after pre-processing operations. After labeling the tweets, we removed stop words and applied tokenization, case-folding, and stemming to avoid misleading the models.

TABLE II. A SAMPLE OF TWEETS BEFORE AND AFTER THE CLEANING PROCESS

Original raw tweets	Cleaned tweets
Please get vaccinated. https://t.co/YfZnFLRTSd	Please get vaccinated
Egypt bans unvaccinated people from public institutions#Egypt#vaccination#COVID19#coronavirus... https://t.co/0rvWPbaL8P	Egypt bans unvaccinated people from public institutions

C. Data Labeling using a Lexicon-based Approach

VADER is a lexicon-based approach used for text sentiment analysis. It depends on a dictionary comprising nearly 7517 features (words) with corresponding valence scores [32]. VADER sentiment analyzer returns a tuple of (neg, neu, pos, compound) for each sentence. A compound score is computed for each text, ranging from -1 (most negative feeling) to 1 (most positive feeling). It is a normalized score calculated by adding the scores of each word present in the sentence. If the compound score is greater than or equal to 0.05 then the tweet sentiment is considered positive and if it is smaller than or equal to -0.05 then the sentiment of the tweet is negative, otherwise the tweet is regarded as neutral. Some tweets with their polarity orientation are shown in Table III.

TABLE III. A SAMPLE OF TWEETS WITH THEIR CORRESPONDING SENTIMENTS

Tweets	Polarity
I just got vaccinated on Tuesday and it hurts	Negative
The city is starting to enforce the vaccine requirement to enter most indoor public spaces You MUST be vaccinated	Neutral
Good thing they are all double vaccinated	Positive

D. Sentiment Classification using ML and DL Models

This study used traditional machine learning in addition to deep learning models to classify the sentiments of tweets. LR, NB, DT, RF, and K-Nearest Neighbors (KNN) classifiers are implemented using scikit-learn library in Python. For our sentiment analysis task, two individual deep learning models, LSTM and BiGRU, were used in addition to our proposed ensemble model (LSTM-2BiGRU) to achieve higher levels of accuracy. The proposed model combines LSTM and BiGRU to obtain better results. In LSTM-2BiGRU, the output of the LSTM is fed to a stack of two sequential BiGRU layers for tweets sentiment classification.

E. Performance Evaluation

We used precision, recall, accuracy, and F1-Score to assess the performance of the sentiment classification models. These performance metrics are calculated using the following formulas:

$$\text{Precision} = \frac{TP}{TP+FP} \quad (1)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (2)$$

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

$$\text{F1-score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

Where TP, TN, FP, and FN represent true positive, true negative, false positive, and false negative instances respectively.

F. Architecture of the Proposed Model

The structure of the proposed model is composed of seven layers, as given in Fig. 2 below. The first layer of the proposed model is an embedding layer, where the number of unique words in the vocabulary is 5000. Each word is represented by a vector of 100 dimensions. A dropout layer with a 0.5 dropout rate follows the embedding layer. After the dropout layer, an LSTM layer is applied with 128 units. The output of the previous layer is fed to a stack of two BiGRU layers with 100 and 64 units for each layer, respectively. The output of the first BiGRU layer is processed by the subsequent BiGRU layer to improve the accuracy of the model. BiGRU traverses the input data two times through a forward and a backward pass. Lastly a dense layer is used with three neurons based on the number of target classes (positive, negative, or neutral) and a softmax activation function. Before the previous dense layer, a dropout layer is placed with a rate of 0.5. The dropout layers are employed to reduce the complexity of the model as well as its propensity for overfitting. The LSTM-2BiGRU model was trained using 21 epochs and a batch size of 128. Early stopping was used to prevent overfitting of the model. The hyperparameters used to train our proposed model are listed in Table IV below.

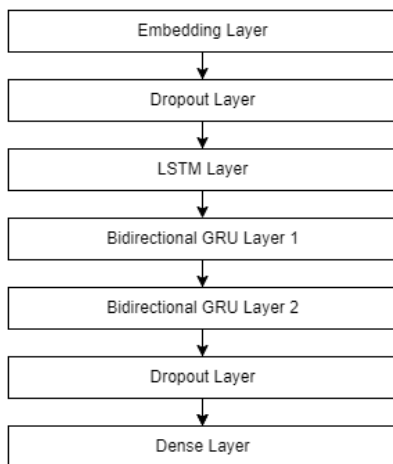


Fig. 2. Architecture of the proposed model.

TABLE IV. HYPERPARAMETERS USED FOR THE PROPOSED MODEL

Hyperparameter	Value
Optimizer	Adam
Loss	Categorical cross-entropy
Batch Size	128
Epoch	21
Dropout	0.5
Activation	Softmax

The pseudo-code below represents the steps taken to develop and evaluate the proposed model.

Algorithm 1: The pseudocode of the proposed model

- 1: **Import** Keras library for creating a deep learning model (M)
- 2: **Read** the collected dataset (D_i)
- 3: **Clean** the tweets from unwanted symbols like #, @, etc.
- 4: **Label** tweets using VADER tool
- 5: **If** the compound score >= 0.05
- 6: **Return** “Positive”
- 7: **Else If** compound score <= -0.05
- 8: **Return** “Negative”
- 9: **Else**
- 10: **Return** “Neutral”
- 11: **End If**
- 12: **Pre-process** cleaned tweets by removing stop words, applying case-folding, stemming, and tokenization
- 13: **Initialize** hyperparameters by setting input_dim=5000, output_dim=100, batch_size=128, epochs= 21
- 14: **For** no_epochs and batch_size **do**
- 15: **Train** the model (M)
- 16: **End For**
- 17: **Predict** the sentiment of unlabeled tweets using M
- 18: **Evaluate** the model and calculate the metrics accuracy, precision, recall, and F1-score
- 19: **Plot** the ROC curve of the model

IV. RESULTS AND DISCUSSIONS

This section demonstrates the results of sentiment analysis on COVID-19 vaccination tweets using machine learning and deep learning models. The collected dataset in this study is split into 85% for training the models and 15% for testing purpose on unseen and unlabeled data. The number of tweets in the training set is 1,199,625 while the testing set consists of about 211699 tweets. Fig. 3 shows a word cloud of the most frequent terms used in all the tweets. Fig. 4 to Fig. 6 present the key terms used in the positive, negative, and neutral tweets respectively.

TABLE V. SENTIMENT ANALYSIS RESULTS USING MACHINE LEARNING CLASSIFIERS

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Random Forest	87	87	87	87
Logistic Regression	87	87	87	87
Decision Tree	83	83	83	83
Naïve Bayes	80	81	80	80
KNN	57	68	56	55
Voting Classifier	88	88	88	88

TABLE VI. HYPERPARAMETERS SELECTED FOR MACHINE LEARNING MODELS

Algorithm	Hyperparameter
Logistic Regression	solver = 'saga'
Multinomial Naïve Bayes	default setting
Decision Tree	criterion = 'gini'
Random Forest	n_estimators = 100
KNN	n_neighbors = 3, weights = 'uniform'

C. Sentiment Analysis using Deep Learning Models

To attain higher levels of accuracy for our sentiment analysis task, two individual deep learning models are used including LSTM and BiGRU as well as our proposed ensemble model. The results of the deep learning models are provided in Table VII. The proposed LSTM-2BiGRU model shows a superior performance over the traditional machine learning and deep learning models used in the study. It achieved the highest accuracy score of 92.46%, followed by LSTM and BiGRU with accuracy scores of about 91%. It is important to mention that using an ensemble of LSTM and BiGRU improved the classification accuracy than using each model individually. Fig. 9 shows the receiver operating characteristic (ROC) curve for the proposed model. The micro-average value of the area under the curve (AUC) for the LSTM-2BiGRU model is 0.98.

TABLE VII. SENTIMENT ANALYSIS RESULTS USING DEEP LEARNING MODELS

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
LSTM	91.98	92.70	91.28	91.97
BiGRU	91.97	92.59	91.43	92.00
LSTM-2BiGRU	92.46	92.97	91.94	92.44

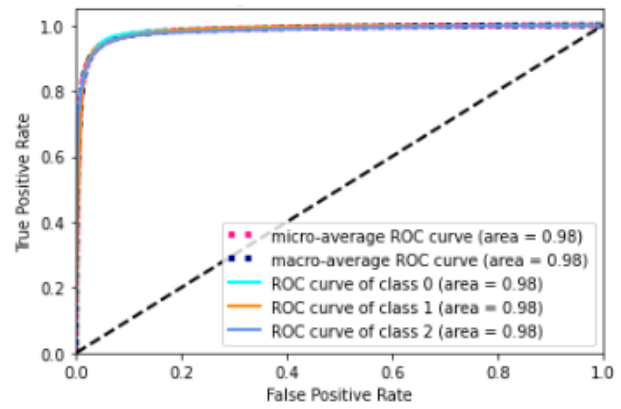


Fig. 9. ROC curve for the proposed model.

D. Comparison with State-of-the-Art Studies

To assess the performance of the proposed LSTM-2BiGRU model with respect to recent studies, a performance comparison was conducted on four cutting-edge studies related to sentiment analysis. These experiments were performed using the collected dataset in this study. In addition, all the steps involved in the proposed approach were applied to the models in the selected studies to guarantee a fair and accurate comparison. The study [33] introduced a stacked BiLSTM model to determine the polarity of deep fake tweets. The authors of paper [34] performed sentiment analysis of tweets related to COVID-19 vaccination using LSTM model. Likely, Reshi et al. presented an ensemble model called LSTM-GRNN for carrying out the same task in research [35]. In addition, paper [36] combined LSTM and GRU to analyze sentiments and detect emotions of tweets related to cryptocurrency. Table VIII compares the result of our proposed model to the findings of cutting-edge studies relevant to our task. The output results show that the proposed model outperforms other models in recent studies with 0.92 accuracy.

TABLE VIII. PERFORMANCE COMPARISON WITH STATE-OF-THE-ART STUDIES

Reference	Model	Accuracy (%)
[33]	Stacked Bi-LSTM	91.94
[34]	LSTM	91.67
[35]	LSTM-GRNN	90.99
[36]	LSTM-GRU	90.75
This study	LSTM-2BiGRU	92.46

V. CONCLUSION

This study performs sentiment analysis on tweets related to the COVID-19 vaccination. To analyze the sentiments of the tweets, this study uses a hybrid approach based on two techniques: natural language processing and machine learning. For conducting experiments, tweets were extracted from Twitter, and the dataset was annotated using VADER. Accordingly, the proportion of people having neither positive nor negative feelings towards the COVID-19 vaccines is the highest among others. Considering the machine learning side of the hybrid approach, the dataset was split into 85% for training and 15% for testing. The sentiments of the tweets were analyzed using five machine learning classifiers: LR, NB, DT, RF, and KNN. To obtain higher accuracy for sentiment classification, two individual deep learning models (LSTM, BiGRU) were used in addition to our proposed ensemble model LSTM-2BiGRU. The experimental results show that the LSTM-2BiGRU model performs significantly better than all the traditional machine learning and deep learning models used in this study. With a 92.46% accuracy score, the proposed model proves its superiority in predicting the sentiments of tweets that are related to the vaccination aspect of COVID-19. However, it is important to point out that BiGRU is slow and requires more time for training due to its complex nature. In future, we aim to perform multimodal sentiment analysis as social media users express their feelings and opinions in multimedia forms, such as videos or images, rather than text solely.

REFERENCES

- [1] R. Jagdale, V. Shirsath, and S. Deshmukh, "Sentiment Analysis on Product Reviews Using Machine Learning Techniques: Proceeding of CISC 2017," 2019, pp. 639-647.
- [2] B. Liu, "Sentiment analysis and opinion mining," Synthesis lectures on human language technologies, vol. 5, no. 1, pp. 1-167, 2012.
- [3] B. Liu and L. Zhang, "A Survey of Opinion Mining and Sentiment Analysis," in Mining Text Data, C. C. Aggarwal and C. Zhai Eds. Boston, MA: Springer US, 2012, pp. 415-463.
- [4] H. Kaur, V. Mangat, and Nidhi, "A survey of sentiment analysis techniques," in 2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), 10-11 Feb. 2017 2017, pp. 921-925, doi: 10.1109/I-SMAC.2017.8058315.
- [5] N. C. Dang, M. N. Moreno-García, and F. De la Prieta, "Sentiment Analysis Based on Deep Learning: A Comparative Study," Electronics, vol. 9, no. 3, 2020, doi: 10.3390/electronics9030483.
- [6] A. Lighthart, C. Catal, and B. Tekinerdogan, "Systematic reviews in sentiment analysis: a tertiary study," Artificial Intelligence Review, 2021/03/03 2021, doi: 10.1007/s10462-021-09973-3.
- [7] S. A. Lone and A. Ahmad, "COVID-19 pandemic – an African perspective," Emerging Microbes & Infections, vol. 9, no. 1, pp. 1300-1308, 2020/01/01 2020, doi: 10.1080/22221751.2020.1775132.
- [8] Z. Mukandavire, F. Nyabadza, N. J. Malunguza, D. F. Cuadros, T. Shiri, and G. Musuka, "Quantifying early COVID-19 outbreak transmission in South Africa and exploring vaccine efficacy scenarios," PloS one, vol. 15, no. 7, p. e0236003, 2020.
- [9] Y. Yu, X. Si, C. Hu, and J. Zhang, "A review of recurrent neural networks: LSTM cells and network architectures," Neural computation, vol. 31, no. 7, pp. 1235-1270, 2019.
- [10] H. Yin, X. Song, S. Yang, and J. Li, "Sentiment analysis and topic modeling for COVID-19 vaccine discussions," World Wide Web, vol. 25, no. 3, pp. 1067-1083, 2022.
- [11] Z. B. Nezhad and M. A. Deihimi, "Twitter sentiment analysis from Iran about COVID 19 vaccine," Diabetes & Metabolic Syndrome: Clinical Research & Reviews, vol. 16, no. 1, p. 102367, 2022.
- [12] E. T. Khalid, E. B. Talal, M. K. Faraj, and A. A. Yassin, "Sentiment analysis system for COVID-19 vaccinations using data of Twitter," Indonesian Journal of Electrical Engineering and Computer Science, vol. 26, no. 2, pp. 1156-1164, 2022.
- [13] R. Marcec and R. Likic, "Using Twitter for sentiment analysis towards AstraZeneca/Oxford, Pfizer/BioNTech and Moderna COVID-19 vaccines," Postgraduate Medical Journal, vol. 98, no. 1161, pp. 544-550, 2021, doi: 10.1136/postgradmedj-2021-140685.
- [14] Y. Fan, X. Li, L. Zhang, S. Wan, L. Zhang, and F. Zhou, "SARS-CoV-2 Omicron variant: recent progress and future perspectives," Signal transduction and targeted therapy, vol. 7, no. 1, p. 141, 2022.
- [15] C. Villavicencio, J. J. Macrohon, X. A. Inbaraj, J.-H. Jeng, and J.-G. Hsieh, "Twitter Sentiment Analysis towards COVID-19 Vaccines in the Philippines Using Naïve Bayes," Information, vol. 12, no. 5, p. 204, 2021.
- [16] C. Shofiya and S. Abidi, "Sentiment Analysis on COVID-19-Related Social Distancing in Canada Using Twitter Data," International Journal of Environmental Research and Public Health, vol. 18, no. 11, p. 5993, 2021.
- [17] J. Samuel, G. G. M. N. Ali, M. M. Rahman, E. Esawi, and Y. Samuel, "COVID-19 Public Sentiment Insights and Machine Learning for Tweets Classification," Information, vol. 11, no. 6, p. 314, 2020.
- [18] N. Chintalapudi, G. Battineni, and F. Amenta, "Sentimental Analysis of COVID-19 Tweets Using Deep Learning Models," (in eng). Infect Dis Rep, vol. 13, no. 2, pp. 329-339, Apr 1 2021, doi: 10.3390/idr13020032.
- [19] X. Zhou, "Sentiment Analysis of COVID-19 Tweets," 2021.
- [20] X. Yu, C. Zhong, D. Li, and W. Xu, "Sentiment analysis for news and social media in COVID-19," Proceedings of the 6th ACM SIGSPATIAL International Workshop on Emergency Management using GIS, 2020.
- [21] H. Kaur, S. U. Ahsaan, B. Alankar, and V. Chang, "A proposed sentiment analysis deep learning algorithm for analyzing COVID-19 tweets," Information Systems Frontiers, vol. 23, no. 6, pp. 1417-1429, 2021.
- [22] F. Rustam, M. Khalid, W. Aslam, V. Rupapara, A. Mehmood, and G. S. Choi, "A performance comparison of supervised machine learning models for Covid-19 tweets sentiment analysis," Plos one, vol. 16, no. 2, p. e0245909, 2021.
- [23] M. Rahman and M. N. Islam, "Exploring the performance of ensemble machine learning classifiers for sentiment analysis of covid-19 tweets," in Sentimental Analysis and Deep Learning: Springer, 2022, pp. 383-396.
- [24] K. N. Alam et al., "Deep learning-based sentiment analysis of COVID-19 vaccination responses from Twitter data," Computational and Mathematical Methods in Medicine, vol. 2021.
- [25] T. T. Mengistie and D. Kumar, "Deep Learning Based Sentiment Analysis On COVID-19 Public Reviews," in 2021 International Conference on Artificial Intelligence in Information and Communication (ICAIC), 13-16 April 2021 2021, pp. 444-449, doi: 10.1109/ICAIC51459.2021.9415191.
- [26] Y. Didi, A. Walha, and A. Wali, "COVID-19 Tweets Classification Based on a Hybrid Word Embedding Method," Big Data and Cognitive Computing, vol. 6, no. 2, p. 58, 2022.
- [27] M. Arbane, R. Benlamri, Y. Brik, and A. D. Alahmar, "Social media-based COVID-19 sentiment classification model using Bi-LSTM," Expert Systems with Applications, vol. 212, p. 118710, 2023/02/01/ 2023, doi: https://doi.org/10.1016/j.eswa.2022.118710.
- [28] M. E. Basiri, S. Nemat, M. Abdar, S. Asadi, and U. R. Acharrya, "A novel fusion-based deep learning model for sentiment analysis of COVID-19 tweets," Knowledge-Based Systems, vol. 228, p. 107242, 2021.
- [29] C. Sitaula, A. Basnet, A. Mainali, and T. B. Shahi, "Deep learning-based methods for sentiment analysis on Nepali covid-19-related tweets," Computational Intelligence and Neuroscience, vol. 2021, 2021.
- [30] S. S. Aljameel et al., "A sentiment analysis approach to predict an individual's awareness of the precautionary procedures to prevent

- COVID-19 outbreaks in Saudi Arabia," International journal of environmental research and public health, vol. 18, no. 1, p. 218, 2021.
- [31] M. Singh, A. K. Jakhar, and S. Pandey, "Sentiment analysis on the impact of coronavirus in social life using the BERT model," Social Network Analysis and Mining, vol. 11, no. 1, pp. 1-11, 2021.
- [32] E. Saad et al., "Determining the efficiency of drugs under special conditions from users' reviews on healthcare web forums," IEEE Access, vol. 9, pp. 85721-85737, 2021.
- [33] V. Rupapara, F. Rustam, A. Amaar, P. B. Washington, E. Lee, and I. Ashraf, "Deepfake tweets classification using stacked Bi-LSTM and words embedding," PeerJ Computer Science, vol. 7, p. e745, 2021.
- [34] R. R. Aryal and A. Bhattarai, "Sentiment Analysis on Covid-19 Vaccination Tweets using Naïve Bayes and LSTM," Advances in Engineering and Technology: An International Journal, vol. 1, no. 1, pp. 57-70, 2021.
- [35] A. A. Reshi et al., "COVID-19 Vaccination-Related Sentiments Analysis: A Case Study Using Worldwide Twitter Dataset," in Healthcare, 2022, vol. 10, no. 3: MDPI, p. 411.
- [36] N. Aslam, F. Rustam, E. Lee, P. B. Washington, and I. Ashraf, "Sentiment Analysis and Emotion Detection on Cryptocurrency Related Tweets Using Ensemble LSTM-GRU Model," IEEE Access, vol. 10, pp. 39313-39324, 2022.

Validate the Users' Comfortable Level in the Virtual Reality Walkthrough Environment for Minimizing Motion Sickness

Muhammad Danish Affan Anua¹, Ismahafezi Ismail², Nur Saadah Mohd Shapri³, Wan Mohd Amir Fazamin Wan Hamzah⁴, Maizan Mat Amin⁵, Fazida Karim⁶

School of Multimedia-Faculty of Informatics and Computing, Universiti Sultan Zainal Abidin, Besut, Malaysia¹⁻⁵

School of Management Sciences-Faculty of Business and Management, Universiti Sultan Zainal Abidin, Besut, Malaysia⁶

Abstract—Motion sickness is a common scenario for users when they are exposed to a virtual reality (VR) environment. It is due to the conflict that occurs in the brain that tells the user that they are moving in the environment, but the fact is that the user's body is sitting still causing them to get symptoms of motion sickness like nausea and dizziness. Therefore, motion sickness has become one of the main reasons why users still do not prefer to use VR to enhance their productivity. Motion sickness can be overcome by increasing the user's comfort level of walkthrough in the VR environment. Meanwhile, a popular VR simulation which is widely used in many industries is a walkthrough in a VR environment at a certain speed. This paper is focused on presenting the result of walkthroughs in a VR environment using movement speed and based on frame rates performance and adopting the unified theory of acceptance and use of technology (UTAUT) model construct variables namely performance expectancy (PE) and effort expectancy (EE) to measure the user's comfort level. A mobile VR, 'VR Terrain' application software was developed based on the proposed framework. The application software was tested by 30 users by moving around in a VR environment with 4 different movement speeds that were implemented into four colored gates using a head-mounted display (HMD). A descriptive and coefficient analysis was used to analyze all the data. The blue gate revealed the most comfortable, outperforming all other three gates. Overall, the most suitable speed to use for VR walkthrough is 4.0 km/h. The experiment result may be used to create a parameter for the VR developers to reduce the VR motion sickness effect in the future.

Keywords—Virtual reality; motion sickness; head-mounted display; head lean movement; mobile VR; walkthrough technique; UTAUT; frame rate

I. INTRODUCTION

A. Background Study

Virtual Reality (VR) is the technology that provides almost real and believable experiences in a synthetic or virtual way [1]. Recent research says VR objects typically act similarly to their real-world counterparts. The user can interact with these things in line with the actual physics principles [2]. Users that use hardware devices like goggles, headphones, and special gloves encounter virtual worlds created and served by software [3]. Experiencing VR technology requires a system with designed computer software that can implement the technology and a head-mounted display (HMD) that allows users to see the VR environment [4]. The first HMD was invented and

developed by Sutherland in 1965, which was introduced as the ultimate display [5] and became available to the public commercially as the revival of VR with HMD at a low price [6].

However, there are common issues that occur involving the use of VR technology. Users are reported to often experience motion sickness when experiencing VR [7][8]. Motion sickness is generally detected after the symptoms appear. This indicates that the sickness has begun, and this effect is already uncomfortable for the user [9]. When it comes to motion sickness or VR related, many report that women are more susceptible than men. One of the classic examples is seasickness with a ratio of approximately 5:3 by Lawther and Griffin, (1988) [10]. Several factors of motion sickness in VR are associated with the HMD such as motion, field of view (FOV), latency, duration of use, and the VR environment [11][12][13][14][15]. These uncomfortable effects will inhibit future experiences of VR.

In this paper, researchers attempted to minimize VR motion sickness by measuring the comfortable level of experience in mobile VR through movement speed of walkthrough. The users must perform a walkthrough by moving around in the VR environment using the four levels of speeds that were implemented in the developed Android application software, 'VR Terrain'. The test evaluation was carried out by adopting some parts of Unified Theory of Acceptance and Use of Technology (UTAUT) [16]. For evaluation, the theory's construct variables namely performance expectancy (PE) and effort Expectancy (EE) were involved, as well as observing the frame rate value of the used device.

B. Paper Contribution

The contributions of the paper are the following:

- The description and the implementation of the VR walkthrough movement speed parameter.
- An Android application software to measure the users' comfortable level experience of VR walkthrough.
- Results of UTAUT and frame rate values for users' comfortable level experience of VR walkthrough.

II. RELATED WORK

A. Motion Sickness of Virtual Reality

One major theoretical issue that has dominated the field for many years concerns the motion sickness effect in VR. It is also known as cybersickness in terms of symptoms as they are similar. The only difference between the terms is that motion sickness is caused in the real world while cybersickness is caused by the virtual one [17]. With developers striving for higher constancy in VR, the content's specifics are becoming more complicated. Thus, this paper also highlights several factors and effects of VR motion sickness.

1) *Factor of VR motion sickness:* VR motion sickness is caused when the brain receives sensory input that does not match the movements of the body. Even though VR technology is becoming more advanced, users still experience motion sickness. This paper focuses on the factors that cause motion sickness among VR users, specifically involving the use of head-mounted display (HMD). There are five main factors that contribute to VR motion sickness: motion and speed, field of view, latency, duration of use, and environment. Rapid and intense movements [18][10][19], a narrow field of view [12][20][21], high latency [22][23][24], prolonged use [25][26][27], and overly stimulating environments [28][29][30] can trigger motion sickness in VR users. By understanding these factors and taking steps to mitigate them, it is possible to reduce the risk of VR motion sickness and improve the VR experience for users.

2) *Effect of motion sickness:* On the other hand, VR motion sickness occurs when a person experiences discomfort or disorientation while using VR technology, due to a mismatch between what they see and feel in the VR environment and what their body is physically experiencing [31]. The effects of VR motion sickness can include nausea [32][33][31], dizziness [34][35][29], headache [8][36][37], fatigue [38][39][40], and eye strain [41][12][37]. These effects can vary from person to person and negatively impact a person's overall enjoyment and use of VR technology. Understanding the effects of VR motion sickness is important for developers and designers of VR technology, as it can help them create more comfortable and effective VR environments for users. The goal of this research is to minimize the effects of VR motion sickness.

B. Head Lean Movement Technique for VR Walkthrough

The head lean movement technique allows for hands-free navigation in VR by relying solely on head movement and six degrees of freedom (6-DoF) tracking [42]. The user simply leans their head to move forward, making it easy to use. The smartphone's gyroscope measures the head movement. Studies have shown that this technique can be effective, but there may be a pause gap between inputs due to the user's tendency to look forward and downward while navigating [43]. Studies have also measured the effects of this technique on factors such as sickness, presence, usability, and comfort when experiencing VR. Previous research has employed this method with a variety of VR headsets, including the Oculus Rift,

Samsung Gear VR, and even the HTC Vive, on both mobile and desktop platforms. Table I below shows the head lean technique for VR.

TABLE I. HEAD LEAN TECHNIQUE FOR VR

Head Lean Technique			
No	VR Device	VR Platform	Author
1	Oculus Rift DK2	Desktop	[44]
2	Samsung Gear VR	Mobile	[42]
3	Samsung Gear VR	Mobile	[45]
4	HTC Vive	Desktop	[43]

C. Unified Theory of Acceptance and Use of Technology

A concept called the Unified Theory of Acceptance and Use of Technology (UTAUT) integrates eight separate hypotheses to explain why people adopt and use technology [16]. It can account for 50% of the variations in technology usage and 70% of the variations in willingness to use technology [46]. UTAUT was initially developed for usage in business environments, but it can be used in other contexts as well. The Technology Acceptance Model (TAM), the Theory of Planned Behavior (TPB), and Social Cognitive Theory (SCT) are among the models on perception, acceptance, and preparedness to employ technology that are combined in this study. Performance Expectancy (PE), Effort Expectancy (EE), Social Influence (SI), and Facilitating Conditions (FC) are the four variables used by UTAUT to validate users' capacity and willingness to adopt new technology.

1) *Performance expectancy:* Performance expectancy (PE) is the degree to which a person expects that using technology will help them execute their jobs more effectively [16]. Age and gender also determine the impact of PE [47]. The first fundamental factor used to determine UTAUT for consumers is the degree to which consumers benefit from using technology [48].

2) *Effort Expectancy:* Technology's early acceptance behavior is influenced by how easily it may be used [49]. Effort Expectancy (EE) stands for "the degree of system usability" [16]. For older people, online technologies could be challenging. Age, gender, and experience are also determining the impact of EE [47]. Regardless of the consumers' technological competency, the EE can be used to determine whether they anticipate experiencing difficulties when utilizing the system.

3) *Social Influence:* Social Influence (SE) refers to how much a person values other people's opinions about whether they should use new technology [16]. Subjective norms in technology adoption indicate social influence [50]. Age, gender, experience, and voluntariness of use are determining the impact of SI [47]. The intention to use technology is impacted by having active social contacts [51].

4) *Facilitating Conditions:* Facilitating conditions (FC) refer to the extent to which consumers believe they have the resources and technologies to support their use of modern technology [16]. Besides, awareness and motivation are what

encourage people to overcome technological obstacles. According to UTAUT, FC is the internal attribute of a target customer [16][47]. Only age and experience determine the impact of FC [47]. The enabling condition influences effort expectancy or ease of use and can predict the use of technology.

D. Frame Rates for Mobile VR

For mobile VR, keeping a high frame rate is very crucial. Smartphones, which are less powerful than desktop PCs or gaming consoles, power mobile VR systems. Because mobile devices have limited capabilities, it might be challenging to attain a high frame rate with mobile VR systems [52][53]. Mobile VR systems must also work under the constraints of the and motion tracking sensors, which can further cut down on the resources available for producing and showing VR content [54]. Notwithstanding these difficulties, high frame rate mobile VR has made tremendous advancements.

The Oculus Quest 2 and Samsung Gear VR are two examples of the most recent mobile VR systems that have made considerable advancements in offering a high-quality VR experience with a high frame rate [55][51]. For instance, the Oculus Quest 2 can display VR material at 72 FPS, which is quite near to the optimal frame rate of 90 FPS.

III. DEVELOPMENT AND IMPLEMENTATION

A. Application Software Development

For this paper, suitable hardware and software have been chosen to avoid any difficulties during the development of this virtual rehabilitation application. Hence, it was important to have the level of expertise in selecting the hardware and the software to ensure any problems that arise during this development process can be solved.

The VR Terrain application software was designed in the Unity editor version 2018. This editor enables developers to create apps, games, or 2D and 3D experiments on a variety of platforms, including PC, Mac, Linux, iOS, Android, Windows, and web. Prior to importing the models, a configuration setting must be done to enable the VR environment settings and the Android build support. Thus, three software packages needed to be imported and installed to produce a working Android mobile application with Unity 3D software. Therefore, an Android Software Development Kit (SDK), Java Development Kit (JDK) and Google VR SDK were added to the Unity editor. The first two software packages convert the application built into an '.apk' file. This file can only be compatible with Android Operational System (OS) smartphone devices. On the other hand, the Google VR SDK software package enables the development of applications for VR HMD or glasses.

Next, the 3D models had been selected and downloaded from the Unity store which can be accessed through the Unity 3D website, were then imported into the editor. Then, the models were modified, scaled, and moved to form a terrain that later was turned into a VR environment. A function called 'Character Controller' by Unity was created. This function applies as the camera that will give user an ability to see the VR environment in first-person view and rotate their head to change the viewing angle, allowing presence of immersion for

VR user. After .apk file was created when the development had been completed, it will be installed in the smartphone. Users can view the VR environment through the lenses in the HMD.

The head lean movement technique was applied to enable the user to humanly move from one place to another in the VE. The user needed to lean their head to the front and downward to simulate the walking movement. The head rotation plays an important part as it was set to a certain degree using a script. The script had also been added and applied a condition, which changed the movement speed of the character controller. The speed changed when there was a collision between the character controller and the gates.

There are four gates designed with four color variations which are green, blue, purple, and red. In order for user to change the movement speed when performing the walkthrough in the VR environment, the gates were set as collision objects, which means when the user makes contact with one of the gates, the movement speed will take effect and make the user virtually move at a specific level of speed. There are four levels of speeds set in VR Terrain application software which are 3, 4, 5 and 6 kilometer per hour (km/h) and individually connected the gates. When the user penetrates the green gate, they will move at a speed of 3 km/h. When the user penetrated the blue gate, they will move at speed of 4 km/h. When the user penetrates the purple gate, they will move at the speed of 5 km/h. When the user penetrates the red gate, they will move at the speed of 6 km/h.

Finally, the speed in kilometers per hour (km/h) and the frame rate information display were added into the application. This information was displayed on the top corner of the screen in the application to keep tracking their current uses of the walkthrough speed movement in the VR environment. This information was used in evaluation phase to strengthen the results and provide a valid outcome. Fig. 1 shows the screenshot of the application view in the smartphone.

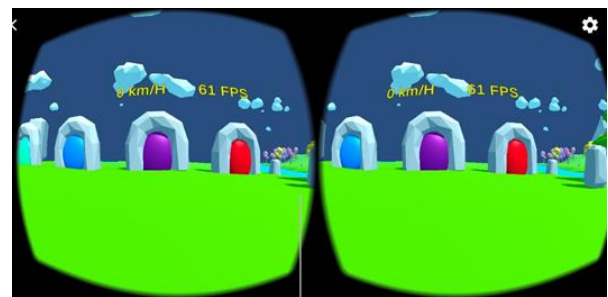


Fig. 1. Application screenshot from mobile.

B. Implementation

Prior to the test, the users will be instructed to fill in the required consent form to prove that the users agreed to be subjects in the test session and the data obtained from the test was used and studied. Each subject was later given a consent form signing that they voluntarily agreed to allow their individual data to be collected and analyzed after the session ended. Note that this testing procedure was using the same smartphone to ensure quality of the research is guaranteed and to avoid data confusion.

The users were asked to wear HMD with the smartphone attached to the front. The VR Terrain application software was previously installed and launched before the device was attached to the headset. After the HMD had been properly attached, the users were instructed to move through one of the four colored gates and moved around the terrain before moving through another colored gate one at a time before ending the session. The users were also instructed to lean their head down to a minimal degree as it is the way they can move in the VR environment. The users were also able to change direction by turning their head left or right. The session ended after the users had moved through all the gates and moved around the terrain. After taking a rest for five minutes, the users were asked to complete an online self-survey regarding their session before returning home. Fig. 2 shows the flowchart of the testing procedure.

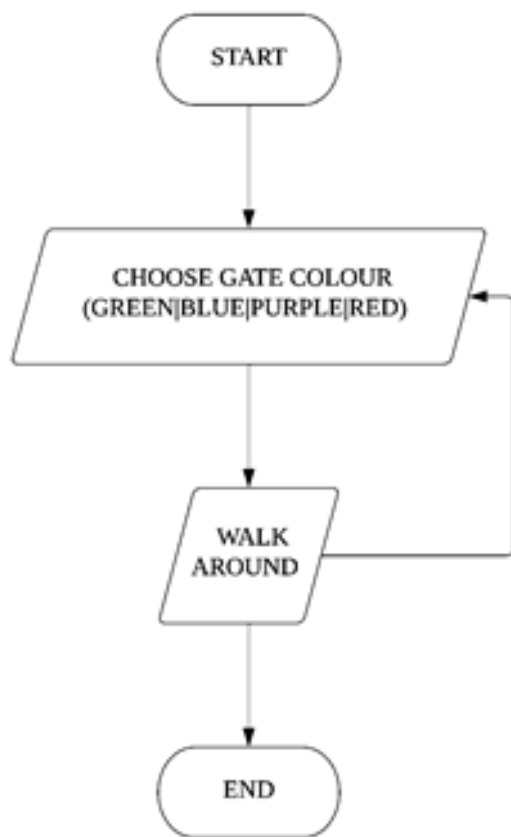


Fig. 2. VR terrain testing procedure flowchart.

C. Evaluation

In this paper UTUAT was used to validate the users' comfortable level walkthrough of VR Terrain application. Only two key constructs have been considered which are PE and EE. These two constructs were selected to validate the comfort level of the user when performing walkthrough in the VR environment using VR Terrain application. In addition, the frame rates of the VR Terrain were observed during the test was running.

1) *UTAUT evaluation:* Performance Expectancy (PE) and Effort Expectancy (EE) are two of the four components from the Unified Theory of Acceptance and Use of Technology included in this paper where each of the items brought out three items that were used to determine the users' comfort level of VR walkthrough for the reducing motion sickness in VR environment application software, VR Terrain application. For the evaluation of users' comfort level walkthrough of VR Terrain application, there are six variables used which are three variables each for PE and EE. The variables are different from one to another. These variables are as shown in Table II.

In the questionnaire, four colored gates (green, blue, purple, and red) are represented with six variables of PE and EE (PE1, PE2, PE3, EE1, EE2, and EE3), making the questionnaire to have a total of twenty-four items of UTAUT model variables assessment that the users are required to complete.

The data was collected and carried out to perform quantitative analysis using SPSS software. The analysis focused on finding PE and EE significant value (p-value) of the four colored gates. The least significant value was taken into account and declared as the most significant findings. Thus, the most comfortable movement speed to use in VR walkthrough was revealed. Table II below has also been adapted in the online questionnaire.

TABLE II. UTAUT RESEARCH VARIABLE

Variable	Item
Performance Expectancy (PE)	PE1: I find this movement speed of VR walkthrough useful for my study/future job.
	PE2: I find using this movement speed of VR walkthrough enables me to accomplish my task pleasantly.
	PE3: I find using this movement speed of VR walkthrough increases my chances of achieving things that are important to me.
Effort Expectancy (EE)	EE1: I find my interaction with this movement speed of VR walkthrough is clear and understandable.
	EE2: I find learning how to use VR walkthrough with this movement speed is easy for me.
	EE3: I find it is easy for me to become skillful at using with this movement speed of VR walkthrough.

2) *Frame rates evaluation:* The frame rates are measured in frames per second (FPS) as a measurement unit and dependent on the processing power of the device. Therefore, the same device has been used to test the VR Terrain application software as the display. This FPS was also validated to see whether it remained constant or underwent any change during the testing session.

IV. RESULT

A. Respondent Profiles

In this paper, the users' demographic profile (n = 30) was classified using frequency distribution. The profile of respondents consists of three descriptions of users that cover the respondents' age, gender, and experience with VR. Table III shows the profile of the respondents using the frequencies and percentages analysis.

TABLE III. RESPONDENT PROFILES

No	Respondent Profiles	Frequency (N)	Percentage (%)
1	Age		
	21	4.0	13.3
	22	19.0	63.3
	23	3.0	10.0
	24	4.0	13.3
2	Gender		
	Male	14.0	46.7
	Female	16.0	53.3
3	Experience with VR		
	Yes	24.0	80.0
	No	6.0	20.0

The demographic analysis provides a general view of the users, which is beneficial in understanding the respondents' background. Based on the result, the majority of the age of the respondents recorded was 22 years old (63.0%), followed by 21 years old (13.3%), and 24 years old (13.3%) participated for the prototype testing. Likewise, most of the respondents were females (53.3%) while males recorded with 46.7%.

Apart from this, the result shows that most of the respondents (80.0%), had experienced VR while only 20.0% of the respondents had not experienced VR before testing the VR Terrain application.

B. Performance and Effort Expectancy Descriptive Analysis

TABLE IV. PERFORMANCE EXPECTANCY DESCRIPTIVE ANALYSIS

No	Item	Frequency (N)	Min	Mean	Max	Std. Dev.
Green Gate						
1	PE1	30.0	1.00	3.07	5.00	1.60
2	PE2	30.0	1.00	3.00	5.00	1.58
3	PE3	30.0	1.00	3.23	5.00	1.59
Blue Gate						
4	PE1	30.0	1.00	3.27	5.00	1.51
5	PE2	30.0	1.00	3.37	5.00	1.56
6	PE3	30.0	1.00	3.33	5.00	1.53
Purple Gate						
7	PE1	30.0	1.00	3.73	5.00	1.28
8	PE2	30.0	1.00	3.50	5.00	1.41
9	P3	30.0	1.00	3.50	5.00	1.36
Red Gate						
10	PE1	30.0	1.00	2.93	5.00	1.51
11	PE2	30.0	1.00	2.90	5.00	1.47
12	PE3	30.0	1.00	3.03	5.00	1.47

TABLE V. EFFORT EXPECTANCY DESCRIPTIVE ANALYSIS

No	Item	Frequency (N)	Min	Mean	Max	Std. Dev.
Green Gate						
1	PE1	30.0	1.00	3.30	5.00	1.58
2	PE2	30.0	1.00	3.20	5.00	1.58
3	PE3	30.0	1.00	3.10	5.00	1.60
Blue Gate						
4	PE1	30.0	1.00	3.33	5.00	1.56
5	PE2	30.0	1.00	3.27	5.00	1.52
6	PE3	30.0	1.00	3.23	5.00	1.48
Purple Gate						
7	PE1	30.0	1.00	3.57	5.00	1.38
8	PE2	30.0	1.00	3.37	5.00	1.47
9	P3	30.0	1.00	3.40	5.00	1.38
Red Gate						
10	PE1	30.0	1.00	3.07	5.00	1.57
11	PE2	30.0	1.00	3.07	5.00	1.53
12	PE3	30.0	1.00	2.97	5.00	1.54

In Table IV, the mean score of performance expectancy ranges between 2.90 to 3.73, and the standard deviation ranges between 1.28 and 1.63. Meanwhile, Table V shows the mean score of effort expectancy which ranges between 2.97 to 3.57, and the standard deviation ranges between 1.38 and 1.60.

Both standard deviation scores are relatively small (less than 3). Thus the degree of performance expectancy and effort expectancy variations is within the normal distribution as proposed by Burn and Bush, (2010) [56].

C. Performance and Effort Expectancy Analysis Coefficient Result

TABLE VI. COEFFICIENT RESULTS OF PERFORMANCE EXPECTANCY

Model	Performance Expectancy (PE)	Unstandardized Coefficients		Standardized Coefficients	t	Sig.(p)
		B	Std. Error	Beta (β)		
1	(constant)	0.707	0.372		1.903	0.069
	PE_G	0.041	0.118	0.063	0.350	0.729
	PE_B	0.316	0.112	0.482	2.819	0.009
	PE_P	0.222	0.118	0.285	1.876	0.072
	PE_R	0.195	0.099	0.281	1.972	0.060

Dependent Variable: CL

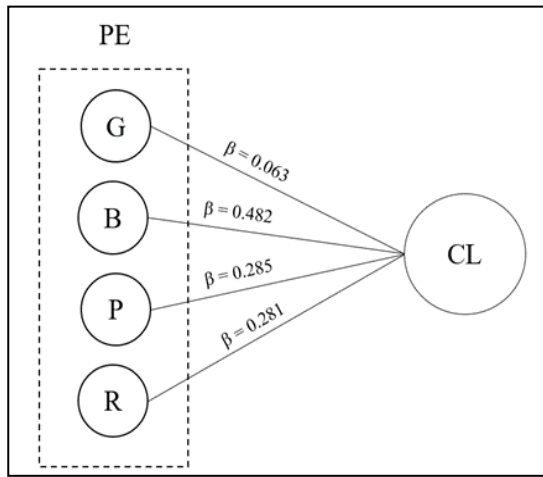


Fig. 3. Performance expectancy (PE) beta (β) result summary.

Table VI shows the result coefficients table for performance expectancy (PE) on comfort level (CL). From the result, it shows that the highest beta value of performance expectancy (PE) is blue gate (0.482), followed by purple gate (0.285), red gate (0.281) and green gate (0.063).

The result concludes that the strongest comfort level for walkthrough in VR environment using VR Terrain application in terms of performance expectancy is Blue Gate ($\beta = 0.482, p = 0.009$). From the finding, walkthrough from Blue Gate (B) has helped the users to accomplish their task pleasantly. Moreover, the movement speed of Blue Gate has increased the user learning in VR environment.

In contrast, Green Gate (G) is not significantly achieving the comfort level for walkthrough ($\beta = 0.063$), as it did not reach the significant level ($p = 0.729 > 0.05$). Similarly Purple Gate and Red Gate received the same beta value ($\beta = 0.285, 0.281$), which are also not significant ($p = 0.072, 0.060$) to achieve comfort level in walkthrough. Fig. 3 above shows the summary of Performance Expectancy Beta results.

TABLE VII. COEFFICIENT RESULTS OF EFFORT EXPECTANCY

Model	Effort Expectancy (EE)	Unstandardized Coefficients		Standardized Coefficients	t	Sig.(p)
		B	Std. Error	Beta (β)		
1	(constant)	0.742	0.341		2.173	0.039
	EE_G	0.009	0.131	0.013	0.066	0.948
	EE_B	0.341	0.121	0.507	2.827	0.009
	EE_P	0.237	0.107	0.326	2.204	0.037
	EE_R	0.170	0.091	0.253	1.863	0.074

Dependent Variable: CL

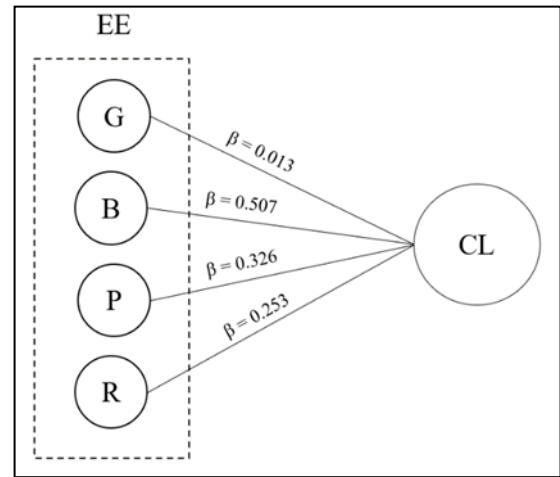


Fig. 4. Effort expectancy (EE) beta (β) result summary.

Table VII shows the result coefficients table for effort expectancy (EE) on comfort level (CL). From the result, it shows that the highest beta value of effort expectancy (EE) is blue gate (0.507), followed by purple gate (0.326), red gate (0.253) and green gate (0.013).

The result concludes that the strongest comfort level for walkthrough in VR environment using VR Terrain application in terms of performance expectancy is Blue Gate ($\beta = 0.507, p = 0.009$). From the finding, walkthrough from Blue Gate (B) has helped the users to accomplish their task pleasantly. Moreover, the movement speed of Blue Gate has increased the user learning in VR environment.

In contrast, Green Gate (G) is not significantly achieving the comfort level for walkthrough ($\beta = 0.013$), as it did not reach the significant level ($p = 0.948 > 0.05$). Similarly Purple Gate and Red Gate received the same beta value ($\beta = 0.326, 0.253$), which are also not significant ($p = 0.037, 0.074$) to achieve comfort level in walkthrough. Fig. 4 above shows the summary of Effort Expectancy Beta results.

D. Result of Frame Rate Observations

The frame rate of a VR experience is a critical factor in determining the overall quality of the experience. A low frame rate can cause motion sickness and detract from the immersion of the experience, while a high frame rate can provide a smoother and more realistic experience. For this reason, the frame rate is one of the most important metrics to be measured and validated during VR Terrain application software testing. From the observation during the test, the collected frame rate data was analyzed to assess the performance of the VR Terrain software application.

TABLE VIII. RESULT OF FRAME RATE OBSERVATIONS

VR Walkthrough Movement Speed (km/h)	Frames Per Second (FPS)		
	59	60	61
	Frequency (N)		
3	0	0	30
4	0	0	30
5	2	25	3
6	2	27	0

According to Table VIII, there are four different movement speeds: 3, 4, 5, and 6 km/h. All 30 users performed the VR walkthrough at all four movement speeds, and the frame rate was recorded through the observation for each user at each speed. The data shows that for 3 and 4 km/h, all users experienced 61 FPS, with no one experiencing 59 FPS or 60 FPS. This indicates that the frame rate was consistently high for these speeds and that the VR experience was smooth and immersive.

For 5 km/h movement speed, however, the data tells a different story. Two users experienced 59 FPS, which is below the generally accepted standard for smooth VR experiences. 25 users experienced 60 FPS, which is generally considered to be the minimum frame rate for a smooth VR experience. Three users experienced 61 FPS, which is higher than the generally accepted standard, indicating that the VR experience was exceptionally smooth and immersive for these users.

Finally, for 6km/h movement speed, the data again shows some variability in the frame rate. Two users experienced 59 FPS, which is below the generally accepted standard for smooth VR experiences. 28 users experienced 60 FPS, which is generally considered to be the minimum frame rate for a smooth VR experience. No users experienced 61 FPS, indicating that the VR experience was not exceptionally smooth or immersive for any users at this speed.

Likewise, for 3km/h and 4km/h, the average frame rate was 61fps, as all users experienced this frame rate. For 5km/h, the average frame rate was 60fps, as 25 users experienced this frame rate, and it is generally considered to be the minimum standard for a smooth VR experience. For 6km/h, the average frame rate was 60fps, as 28 users experienced this frame rate.

In conclusion, the frame rate of a VR experience is a critical factor in determining the overall quality of the experience. In the context of a VR walkthrough, the frame rate has become even more important, as users are moving through a VR environment and a low frame rate can cause motion sickness and detract from the immersion of the experience. Based on the data collected, the average frame rate was consistently high for 3km/h and 4km/h, indicating a smooth and immersive VR experience in mobile device. However, for 5km/h and 6km/h, the frame rate was more variable, with some users experiencing frame rates below the generally accepted standard for a smooth VR experience (Kopczynski, 2021; Menakhin, 2016; Fuchs, 2017). Overall, this data highlights the importance of maintaining a high frame rate in VR walkthroughs to ensure a smooth and immersive experience for all users.

E. Result Summary

Overall, the blue gate (B) received the most positive result compared to the purple gate (P), red gate (R), and green gate (G). Likewise, the most comfortable walkthrough in VR recorded in terms of performance expectancy and effort expectancy was with the blue gate (B). Therefore, the result of PE and EE to support CL proposed that, 4 km/h is the most suitable movement speed to use for walkthrough in VR environment. Additionally, the VR walkthrough using VR Terrain application is smooth with the average frame rate of 61 FPS which is acceptable for VR experience for mobile device.

V. FUTURE WORK AND CONCLUSION

A. Future Work

The VR walkthrough application software can and should be further enhanced, so that the system is better, more efficient, and more effective. The suggestions for providing a better result are as stated below:

1) *Enhanced interfaces:* Future researchers can provide more information for users to see in the display such as the time engagement and the distance travelled for users to understand more about their current situation. Besides, the researchers can also use higher quality 3D object and create more realistic VR environment such as city environment, houses, and walking tracks to enhance the VR experience for the user.

2) *Cross-sectional study:* Currently, the research and development of VR environment using the walkthrough technique for minimizing motion sickness is only limited to technological science study. This research and development can be further studied across many regions especially in medical and human science since it involves motion sickness, which is the biological condition for humans. This indirectly contributes to more insight and increases the diversity of knowledge of VR.

3) *Expanding platform:* There are still some limitations to this technology that can be addressed in future versions of mobile-based VR Terrain. Therefore, future work is suggested to expand the testing of VR Terrain software application in desktop-based VR device such as using HTC Vive Pro or in a standalone VR device such as Oculus Quest 2. Desktop-based VR devices offer more advanced hardware and capabilities compared to mobile-based VR experiences. With more powerful hardware, these VR experiences can offer higher quality graphics, more complex interactions, and greater immersion. In addition, stand-alone VR devices offer the convenience of mobile-based VR without the need for an external device like a smartphone.

B. Conclusion

The developed VR Terrain application software was used for VR walkthrough to measure the users' comfort level VR walkthrough experience and was validated by adopting UTAUT model construct variables which are performance expectancy (PE) and the effort expectancy (EE) as well as the frame rate observations. The result of comfort level shows all

users found that VR Terrain application was useful and convenient to use. Overall, VR Terrain application was considered satisfying for the user to use for measuring comfort level of walkthrough in VE. The UTAUT analysis and FPS rate shows 4 km/h is the most comfortable to use for walkthrough in VE, in which the result indicates that 4.0 km/h movement speed can be used in walkthrough in VR environment to minimize the effect of VR motion sickness. Therefore, it can be said that the users' comfort level walkthrough measurement of VR Terrain application is useful for VR developer to use as a reference in minimizing the VR motion sickness effect.

ACKNOWLEDGMENT

This research paper is supported by Universiti Sultan Zainal Abidin (UniSZA) using FUNDAMENTAL RESEARCH GRANT SCHEME (FRGS), project number: FRGS/1/2021/ICT10/UNISZA/02/2. Special Thanks to the Ministry of Higher Education Malaysia (MOHE) and Centre for Research Excellence & Incubation management (CREIM) UniSZA for providing financial support for the research.

REFERENCES

- [1] Furht, B. (2008). Immersive virtual reality. Encyclopedia of Multimedia. Springer, Boston, MA.
- [2] Popov, V. V., Kudryavtseva, E. V., Katiyar, N. K., Shishkin, A., Stepanov, S. I., & Goel, S. (2022). Industry 4.0 and Digitalisation in Healthcare. *Materials* 2022, 15, 2140.
- [3] Mokhtar, M. N. A. B. D., Ismail, I., Hamzah, W. M. A. F. W., Shamsuddin, S. N. W., & Arsad, M. A. M. (2022, July). Real-Time Dream House Decorator in the Virtual Reality Environment. In *Sustainable Finance, Digitalization and the Role of Technology: Proceedings of The International Conference on Business and Technology (ICBT 2021)* (pp. 525-537). Cham: Springer International Publishing.
- [4] Zolkefly, N. N., Ismail, I., Safei, S., Shamsuddin, S. N. W., & Arsad, M. A. M. (2018). Head gesture recognition and interaction techniques in virtual reality: a review. *International Journal of Engineering & Technology*, 7(4.31), 437-440.
- [5] Fluke, C. J., & Barnes, D. G. (2016). The ultimate display. arXiv preprint arXiv:1601.03459.
- [6] Wang, J., & Lindeman, R. (2015). Coordinated hybrid virtual environments: Seamless interaction contexts for effective virtual reality. *Computers & Graphics*, 48, 71-83
- [7] Jerald, J. (2015). The VR book: Human-centered design for virtual reality. Morgan & Claypool.
- [8] Keshavarz, B., Hecht, H., & Lawson, B. D. (2014). Visually Induced Motion Sickness: Causes, Characteristics, and Countermeasures.
- [9] Liao, C. Y., Tai, S. K., Chen, R. C., & Hendry, H. (2020). Using EEG and deep learning to predict motion sickness under wearing a virtual reality device. *Ieee Access*, 8, 126784-126796.
- [10] Lawther, A., & Griffin, M. J. (1988). A survey of the occurrence of motion sickness amongst passengers at sea. *Aviation, space, and environmental medicine*, 59(5), 399-406.
- [11] Riener, R., Harders, M., Riener, R., & Harders, M. (2012). Virtual reality for rehabilitation. *Virtual Reality in Medicine*, 161-180.
- [12] Chang, E., Kim, H. T., & Yoo, B. (2020). Virtual reality sickness: a review of causes and measurements. *International Journal of Human-Computer Interaction*, 36(17), 1658-1682.
- [13] Brunnström, K., Dima, E., Qureshi, T., Johanson, M., Andersson, M., & Sjöström, M. (2020). Latency impact on quality of experience in a virtual reality simulator for remote control of machines. *Signal Processing: Image Communication*, 89, 116005.
- [14] Qu, J., Wang, W., Yuan, S., Li, F., & Cai, R. (2018, July). Analysis of Simulator Sickness and Performance in Virtual Training. In *Journal of Physics: Conference Series* (Vol. 1060, No. 1, p. 012030). IOP Publishing.
- [15] Park, S., Mun, S., Ha, J., & Kim, L. (2021). Non-contact measurement of motion sickness using pupillary rhythms from an infrared camera. *Sensors*, 21(14), 4642.
- [16] Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS quarterly*, 425-478.
- [17] Martirosov, S., & Kopecek, P. (2017). Virtual reality and its influence on training and education-literature review. *Annals of DAAAM & Proceedings*, 28.
- [18] Iijima, A., Kiryu, T., Ukai, K., & Bando, T. (2007). Vergence eye movements elicited by non-disparity factors in 2D realistic movies. In *Proceedings of The First International Symposium on Visually Induced Motion Sickness, Fatigue, and Photosensitive Epileptic Seizures (VIMS 2007)* (pp. 59-66).
- [19] Rahimi, K., Banigan, C., & Ragan, E. D. (2018). Scene transitions and teleportation in virtual reality and the implications for spatial awareness and sickness. *IEEE transactions on visualization and computer graphics*, 26(6), 2273-2287.
- [20] Wu, F., & Rosenberg, E. S. (2019, March). Combining dynamic field of view modification with physical obstacle avoidance. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (pp. 1882-1883). IEEE.
- [21] Keshavarz, B., Hecht, H., & Zschuschke, L. (2011). Intra-visual conflict in visually induced motion sickness. *Displays*, 32(4), 181-188.
- [22] Kundu, R. K., Rahman, A., & Paul, S. (2021). A study on sensor system latency in vr motion sickness. *Journal of Sensor and Actuator Networks*, 10(3), 53.
- [23] Halarnkar, P., Shah, S., Shah, H., Shah, H., & Shah, A. (2012). A review on virtual reality. *International Journal of Computer Science Issues (IJCSI)*, 9(6), 325.
- [24] Torres Vega, M., Liaskos, C., Abadal, S., Papapetrou, E., Jain, A., Mouhouche, B., ... & Famaey, J. (2020). Immersive interconnected virtual and augmented reality: a 5G and IoT perspective. *Journal of Network and Systems Management*, 28, 796-826.
- [25] Wang, Z., Foo, K., Yan, S., Gonzalez, V. A., & Giacaman, N. (2022). Refresh Rate and Graphical Benchmarks for Mobile VR Application Development. *Journal of Mobile Multimedia*, 1561-1598.
- [26] Cao, Z., Jerald, J., & Kopper, R. (2018, March). Visually-induced motion sickness reduction via static and dynamic rest frames. In *2018 IEEE conference on virtual reality and 3D user interfaces (VR)* (pp. 105-112). IEEE.
- [27] Cheung, B., & Nakashima, A. (2006). A review on the effects of frequency of oscillation on motion sickness.
- [28] Chatha, U. A., Janjua, U. I., Anwar, F., Madni, T. M., Cheema, M. F., & Janjua, S. I. (2020). Motion sickness in virtual reality: An empirical evaluation. *IEEE Access*, 8, 130486-130499.
- [29] Berti, S., & Keshavarz, B. (2020). Neuropsychological approaches to visually-induced vection: an overview and evaluation of neuroimaging and neurophysiological studies. *Multisensory Research*, 34(2), 153-186.
- [30] Saredakis, D., Szpak, A., Birckhead, B., Keage, H. A., Rizzo, A., & Loetscher, T. (2020). Factors associated with virtual reality sickness in head-mounted displays: a systematic review and meta-analysis. *Frontiers in human neuroscience*, 14, 96.
- [31] Keshavarz, B., & Golding, J. F. (2022). Motion sickness: current concepts and management. *Current opinion in neurology*, 35(1), 107-112.
- [32] Wickham, R. J. (2020). Revisiting the physiology of nausea and vomiting—challenging the paradigm. *Supportive Care in Cancer*, 28, 13-21.
- [33] Zhong, W., Shahbaz, O., Teskey, G., Beever, A., Kachour, N., Venketaraman, V., & Darmani, N. A. (2021). Mechanisms of nausea and vomiting: current knowledge and recent advances in intracellular emetic signaling systems. *International journal of molecular sciences*, 22(11), 5797.
- [34] Cha, Y. H., Golding, J. F., Keshavarz, B., Furman, J., Kim, J. S., Lopez-Escamez, J. A., ... & Lawson, B. D. (2021). Motion sickness diagnostic

- criteria: consensus document of the classification committee of the Bárány society. *Journal of Vestibular Research*, 31(5), 327-344.
- [35] Hromatka, Bethann S., Joyce Y. Tung, Amy K. Kiefer, Chuong B. Do, David A. Hinds, and Nicholas Eriksson. "Genetic variants associated with motion sickness point to roles for inner ear development, neurological processes and glucose homeostasis." *Human molecular genetics* 24, no. 9 (2015): 2700-2708.
- [36] Marcus, D. A., Furman, J. M., & Balaban, C. D. (2005). Motion sickness in migraine sufferers. *Expert opinion on Pharmacotherapy*, 6(15), 2691-2697.
- [37] Hettinger, L. J., & Riccio, G. E. (1992). Visually induced motion sickness in virtual environments. *Presence: Teleoperators & Virtual Environments*, 1(3), 306-310.
- [38] Tychsen, L., & Foeller, P. (2020). Effects of immersive virtual reality headset viewing on young children: visuomotor function, postural stability, and motion sickness. *American journal of ophthalmology*, 209, 151-159.
- [39] Zhang, C. (2020, September). Investigation on motion sickness in virtual reality environment from the perspective of user experience. In 2020 IEEE 3rd International Conference on Information Systems and Venkatesh, V., Thong, J. Y., & Xu, X. (2012). Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. *MIS quarterly*, 157-178. *Computer Aided Education (ICISCAE)* (pp. 393-396). IEEE.
- [40] Chattha, U. A., & Shah, M. A. (2018, September). Survey on causes of motion sickness in virtual reality. In 2018 24th International Conference on Automation and Computing (ICAC) (pp. 1-5). IEEE.
- [41] Chen, J., Yoon, I., & Bethel, E. (2005). Interactive, Internet Delivery of Scientific Visualization via Structured, Pre-rendered Multiresolution Imagery (No. LBNL-57528). Lawrence Berkeley National Lab.(LBNL), Berkeley, CA (United States).
- [42] Tregillus, S., Al Zayer, M., & Folmer, E. (2017, May). Handsfree omnidirectional VR navigation using head tilt. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (pp. 4063-4068).
- [43] Buttussi, F., & Chittaro, L. (2019). Locomotion in place in virtual reality: A comparative evaluation of joystick, teleport, and leaning. *IEEE transactions on visualization and computer graphics*, 27(1), 125-136.
- [44] Kitson, A., Hashemian, A. M., Stepanova, E. R., Kruijff, E., & Riecke, B. E. (2017, March). Lean into it: exploring leaning-based motion cueing interfaces for virtual reality movement. In 2017 IEEE Virtual Reality (VR) (pp. 215-216). IEEE.
- [45] Anggoro, P. D. W. (2018). Kajian Interaksi Pengguna untuk Navigasi Aplikasi Prambanan VR berbasis Virtual Reality. *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*, 5(2), 239-246.
- [46] Nazmi, N. A. M., Rizhan, W., & Rahim, N. Developing and Evaluating AR for Food Ordering System based on Technological Acceptance Evaluation Approach: A Case Study of Restaurant's Menu Item Selection.
- [47] Venkatesh, V., Morris, M. G., & Ackerman, P. L. (2000). A longitudinal field investigation of gender differences in individual technology adoption decision-making processes. *Organizational behavior and human decision processes*, 83(1), 33-60.
- [48] Venkatesh, V., Thong, J. Y., & Xu, X. (2012). Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. *MIS quarterly*, 157-178.
- [49] Cimperman, M., Brenčič, M. M., & Trkman, P. (2016). Analyzing older users' home telehealth services acceptance behavior—applying an Extended UTAUT model. *International journal of medical informatics*, 90, 22-31.
- [50] Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Q.* 13 (3), 319-340 (1989).
- [51] Rice, R. E., & Aydin, C. (1991). Attitudes toward new organizational technology: Network proximity as a mechanism for social information processing. *Administrative science quarterly*, 219-244.
- [52] Raaen, K., & Kjellmo, I. (2015). Measuring latency in virtual reality systems. In *Entertainment Computing-ICEC 2015: 14th International Conference, ICEC 2015, Trondheim, Norway, September 29-October 2, 2015, Proceedings 14* (pp. 457-462). Springer International Publishing.
- [53] Ho, K. T., King, C. T., Das, B., & Chang, Y. J. (2018, March). Characterizing display QoS based on frame dropping for power management of interactive applications on smartphones. In *2018 Design, Automation & Test in Europe Conference & Exhibition (DATE)* (pp. 873-876). IEEE.
- [54] Hwang, C., Pushp, S., Koh, C., Yoon, J., Liu, Y., Choi, S., & Song, J. (2017, October). Raven: Perception-aware optimization of power consumption for mobile games. In *Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking* (pp. 422-434).
- [55] Hillmann, C., & Hillmann, C. (2019). Comparing the gear vr, oculus go, and oculus quest. *Unreal for Mobile and Standalone VR: Create Professional VR Apps Without Coding*, 141-167.
- [56] Burns, A. C., & Bush, R. F. (2010). *Marketing Research* Pearson Education.

Fusion Privacy Protection of Graph Neural Network Points of Interest Recommendation

Yong Gan¹, ZhenYu Hu²

School of Computer and Communication Engineering, Zhengzhou University of Light Industry, Zhengzhou, China

Abstract—For the rapidly developing location-based web recommendation services, traditional point-of-interest(POI) recommendation methods not only fail to utilize user information efficiently, but also face the problem of privacy leakage. Therefore, this paper proposes a privacy-preserving interest point recommendation system that fuses location and user interaction information. The geolocation-based recommendation system uses convolutional neural networks (CNN) to extract the correlation between user and POI interactions and fuse text features, and then combine the location check-in probability to recommend POIs to users. To address the geolocation leakage problem, this paper proposes an algorithm that integrates k -anonymization techniques with homogenized coordinates (KMG) to generalize the real location of users. Finally, this paper integrates location-preserving algorithms and recommendation algorithms to build a privacy-preserving recommendation system. The system is analyzed by information entropy theory and has a high privacy-preserving effect. The experimental results show that the proposed recommendation system has better recommendation performance on the basis of privacy protection compared with other recommendation algorithms.

Keywords—Recommendation algorithms; location protection; graph convolutional neural networks; k -anonymity

I. INTRODUCTION

With the rapid development of mobile Internet, a series of location-based social network (LBSN) services such as e-commerce and social software have attracted millions of users, who establish online links with other users on LBSN, share experiences and comments of visited points of interest[1], and the service platform analyzes user information and mines user preferences to recommend POIs to users. However, as the volume of data and the complexity of information, the traditional text-based collaborative filtering method is no longer applicable to the development of recommendation systems. In addition, the naturally existing geographic distance factor also has an impact on user activities, as human activity areas have the phenomenon of geographic aggregation[2]. Therefore, the geographic impact between users and POIs and between POIs is as important as the impact between users. Therefore, it makes sense to incorporate the geographic distance factor into a collaborative user or content-based filtering approach.

Compared with traditional methods, neural network-based recommendation algorithms can tap deeper features with local relevance and location invariance to extract potentially valid information from Euclidean spatial data[3]. Therefore, it is the mainstream practice in recommendation systems to use CNNs to extract hidden features from data such as users' social

relationships, comment texts and visit frequencies and fuse them, and then combine the check-in probabilities of geographic distance factors[4]. At this stage, the emergence of graph neural networks (GNNs) further improves the performance of recommendation systems[5]. As a natural bipartite graph structure (user-interest point nodalization), traditional collaborative filtering methods based on neural networks, although having powerful feature representation, are weak in handling higher-order interaction information of graph-structured data in non-Euclidean space[6]. Therefore, introducing GCN into the recommendation system to extract the connection features of user-interest point interactions through information transfer between graph nodes brings better entity representation and stronger interpretation capability to the recommendation system, thus improving the prediction accuracy of the model[7][8].

Since the recommendation system in LBSN lacks location protection measures for users when recommending POIs to them, they may face the threat of location leakage. Existing location protection methods for LBSN are mainly divided into three types: generalization[9][10], k -anonymity and encryption[11][12]. k -anonymity is a method that constructs an anonymous region containing $k-1$ users after removing the identification information from their real locations and query contents, so that the probability of identifying a user does not exceed $1/k$. k -anonymity is widely studied and applied because of its high privacy protection effect compared with other methods. However, existing approaches to location privacy protection exhibit significant limitations. First, recommendation systems utilize trusted third-party anonymous servers to protect user location information. This operation inherently leads to information leakage problems. Second, the generalization-based location protection mechanism generalizes the user's real location so that the user's virtual location exists at any point in the constructed larger virtual area, and the recommendation system calculates the recommendation result based on the user's virtual location, which leads to inaccurate recommendation results. k -anonymity technique, although it is effective for location protection, is less capable of protecting the user's location during continuous requests, because an attacker will analyze the user's action trajectory based on the user's location requests at multiple moments, and then infer the location visited by the user, exposing the user to security threats. Moreover, k -anonymity is achieved by sacrificing the quality of service in exchange for privacy protection [13].

Since the existing location-based point-of-interest recommendation system mainly calculates recommendation

results by analyzing users' social relationships and geographical factors, the deep interaction information between users and POI and users' comment information have more user characteristic information compared with the case of sparse social relationships of users, which can effectively improve the accuracy of the recommendation system. Therefore, how to improve both location privacy protection capability and recommendation system performance is the focus of this paper.

The main contribution of this paper is the design and analysis of a recommendation system that takes location protection into account. Because previous recommender systems did not focus on the protection of user location information, after continuous exploration it is found that privacy protection does not need to sacrifice the greater accuracy of the recommender system. User location information as the most important factor of the recommendation system calculation results, there exists the attack method against location information. The KMG method proposed in this paper is a controllable generalized geolocation algorithm, which calculates virtual location based on all user locations in the anonymous region and divides subanonymous regions based on the distance between real location and virtual location to achieve effective protection of user location without significantly reducing recommendation accuracy. The contribution of this paper in the recommendation system is mainly to fuse the deep interaction information and comment information between users and POI to obtain a more efficient and deeper recommendation algorithm, and integrate it with the location protection algorithm into one system.

II. RELATED WORK

First LBSN-based recommendation systems usually considers three factors: user similarity, social influence, and geographic influence[14]. The traditional method of recommendation by extracting user ratings and review features suffers from cold start and data sparsity. Park et al.[15] proposed ConvMF recommendation model, which integrates CNN into probability matrix decomposition can effectively capture contextual information and improve recommendation accuracy. Zhao et al.[16] proposed hierarchical dichotomous graph neural network, by stacking multiple graph neural networks(GNN) and alternating with clustering algorithm to obtain the potential preference information of users, which substantially improves the recommendation performance compared to the recommendation system with fused CNNs. Lin et al.[17] fused multilayer graph convolutional models with recurrent neural networks to capture user interaction graph information, which can accurately capture the rich potential implicit information between users-items. Shafqat et al.[18] transformed the similarity into the interaction probability between the neighbors of user nodes with different probability distributions using KL scatter to find the distance between them and perform clustering operations on neighboring nodes, and this method improves the efficiency of neighborhood aggregation for GCN models. The graph neural collaborative filtering model (NGCF) proposed by Wang et al.[19] is one of the classical recommender systems incorporating the GCN model, which embeds and models the user interaction information to obtain the higher-order connectivity between nodes and recommends items of interest to users. He et al.[20]

proposed the LightGCN model, which eliminates the feature transformation and nonlinear activation in GCN and improves the efficiency of neighborhood aggregation. However, existing location-based point-of-interest recommendation algorithms generally use fused social relationships without deep mining of user and POI node interaction information, and although GCN can extract the association relationship between users and POI well, most location-based recommendation algorithms do not take into account the advantages of deep learning for processing textual information. Therefore, this paper will use both GCN and deep learning methods to extract interaction information and text information respectively.

There is a correlation between user check-in to points of interest and geographic distance. Ye et al.[21] analyzed the effect of distance on user check-in on the Foursquare and Whrrl datasets and constructed probability distributions based on the POI distance of the same user check-in. Most of the points of interest checked-in by the same user are in a range of small mutual distance, i.e., the check-in locations have a geographic cluster area effect. They fit the check-in probabilities to a power-law distribution, and found that they could cover most (90%) of the POI pairs after eliminating those that did not fit the power-law distribution. Therefore, the power-law distribution can be used to model the distance between POI visited by the same user and calculate the check-in probability.

k -anonymity is one of the main methods for privacy preservation. Song et al.[22] proposed k -anonymity method, whose basic idea is to remove user identification information from data so that each user in the same equivalence class is identified with probability A , so it has high privacy preserving ability. Gruteser et al.[23] introduced the idea of k -anonymity into LBSN privacy preserving by generalizing the real location of the user into an anonymous region and fuzzy the spatial location information of the user. Liu et al.[24] proposed the B-privacy method, which delineates the area of the region according to the user's location no less than s and generates k virtual locations. Since the B-privacy method statically sets the parameters k and s , it leads to weak protection in areas with small population density. Since the above methods use virtual locations to obtain query results, which leads to low precision of results. Ji et al.[25] proposed to split the anonymized region into several dispersed subanonymized regions, which not only reduces the scope of the anonymized region, but also improves the precision of user queries. Although all these location protection methods effectively protect the user's location, they construct as large a virtual region as possible in order to enhance the security strength, resulting in uncertainty in the distance between the user's virtual location and the real location, which will have an uncertain impact on the recommendation effect when combined with the recommendation algorithm. Therefore, in this paper, by limiting the distance between the real location and the virtual location and fusing the construction of subanonymous regions, the virtual location can meet the security and ensure that it does not have a great impact on the accuracy of the recommendation system.

We construct an anonymous region containing k users based on their locations, and replace the center location with

the user location, and the recommendation algorithm recommends users based on the center location. If the distance between the center location and the user location is large, the anonymous region is divided into sub-anonymous regions, and the interest points are recommended again based on the center location of the sub-anonymous region. The recommendation algorithm extracts and fuses high-order interaction features, text features and rating features by LightGCN, CNN and latent factor model (LGCL) respectively, and then recommends POIs jointly with geographic distance factor check-in probability.

III. RECOMMENDATION ALGORITHM

The recommendation algorithm in this paper consists of two parts: the LGCL module and the geographic distance probability. The LGCL module is a nodalization of user-POI through the LightGCN model, which uses GCN to propagate aggregated user-item interactions on the graph and obtain an embedded representation of all user-item association relationships on the graph; the text feature extraction module is to process the text data into a collection of embedding vectors by BERT pre-training model, and then extract the text features by CNN; the rating data processing module obtains the implicit features of the rating matrix by the implicit semantic model (LFM). The geographical distance probability p^g is calculated using the naive Bayesian method based on the user's historical check-in distance. Finally, the features processed by the similarity module are fused and normalized to obtain \hat{f}'_{u_i} , and the POI is recommended to the user jointly with p^g . The structure of the recommended algorithm is shown in Fig. 1.

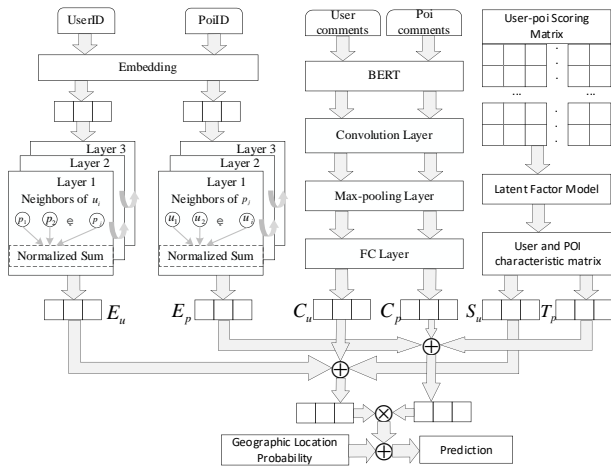


Fig. 1. Recommended algorithm structures.

A. LGCL Model

The ID features of the user and POI are first fed into the embedding layer, and the set of embedding vectors is represented as:

$$e_u^{(0)} = [e_{u_1}^{(0)}, e_{u_2}^{(0)}, \dots, e_{u_N}^{(0)}], e_p^{(0)} = [e_{p_1}^{(0)}, e_{p_2}^{(0)}, \dots, e_{p_M}^{(0)}] \quad (1)$$

where N users, M POI, $e_{u_i}^{(0)}$ and $e_{p_j}^{(0)}$ denote the c dimensional embedding vectors of user i , POI j , respectively.

Then the IDs of all POI nodes in all neighboring nodes of user u_i are embedded in the vector for aggregation operation. The propagation rule for the embedding representation of user u_i in the LightGCN model for layers L to $L+1$ by:

$$e_{u_i}^{(L+1)} = \sum_{p_j \in N_u} \frac{1}{\sqrt{|N_u|} \sqrt{|N_p|}} e_{p_j}^{(L)} \quad (2)$$

$$e_{p_j}^{(L+1)} = \sum_{u_i \in N_p} \frac{1}{\sqrt{|N_p|} \sqrt{|N_u|}} e_{u_i}^{(L)} \quad (3)$$

where N_u and N_p denote the set of neighboring nodes containing all u_i and p_j , respectively; $e_{u_i}^{(L+1)}$ and $e_{p_j}^{(L+1)}$ denote the expression of the embedding of u_i and p_j in LightGCN obtained after the propagation of L layers in $L+1$ layers.

In order to obtain better embedding aggregation, this paper introduces an attention mechanism to get the importance weight of each layer of propagation:

$$Attention_{u_i}^l = softmax(\mu_1 \otimes sigmoid(\mu_2 \otimes e_{u_i})) \quad (4)$$

where μ_1, μ_2 are adjustable hyperparameters, $Attention_{u_i}^l$ is the layer L weight; weighted summation of the embedding vectors for each layer yields the embedding expression for the association between u_i and p_j :

$$E_{u_i} = \sum_{l=0}^L Attention_{u_i}^{(l)} e_{u_i}^{(l)}, E_{p_j} = \sum_{l=0}^L Attention_{p_j}^{(l)} e_{p_j}^{(l)} \quad (5)$$

B. Text Feature Extraction Module

All comments of the same user or POI constitute the comment collection, respectively. Because the extraction of user and POI comment features is composed of two similar parallel network structures, only the extraction of user comment features is described in this paper. To simplify the calculation, let the number of comment texts of each user and POI be n and the length of each comment text be q . The comment set r_{u_i} of user u_i is input to the BERT pre-training model to obtain the comment embedding vector $e_{u_i} \in R^{n \times q \times d}$, where d is the output dimension of the BERT pre-training model.

The user comment embedding vector e_{u_i} is input to CNN for extracting semantic information features. Let the m convolutional kernels with step size s extract contextual features:

$$g_j = Relu(M_i * R + b) \quad (6)$$

where $Relu()$ is the activation function, $*$ denotes the convolution operation, and b_i denotes the bias.

After the convolution operation, the features g_j are fed into the max-pooling layer to generate features with higher values, which take the maximum vector w_i for the region $max(g_1, g_2, \dots, g_{d-s+1})$ corresponding to the convolution kernel, preserving the characteristics of the original feature vector and reducing the dimensionality.

Due to the fact that each comment in the comment collection expresses user preference features to a different degree, this paper introduces an attention mechanism that applies different weights to the comments and normalizes them using the softmax function:

$$h_i = softmax(Relu(\mu_3 \otimes g_i + b)) \quad (7)$$

Finally, the text information feature representation of the user u_i is obtained after processing in the fully connected layer:

$$C_{u_i} = \sum_{i=1}^n a_i h_i \omega_i + b_i \quad (8)$$

where a_i and b_i denote the amount of deviation from the weights of the fully connected layer, respectively.

C. Rating Data Processing Module

The rating can be regarded as a direct feedback, and the level of rating can express the user's liking of the POI. In this paper, we use the Latent Factor Model (LFM) to process the user rating matrix, the essence of which is to decompose the rating matrix to get the user feature matrix and the POI feature matrix, so that the implicit features can be expressed.

The LFM model is to decompose the scoring matrix R into a user feature matrix S_{N_t} and a POI feature matrix T_{tM} such that the matrix multiplication of S_{N_t} and T_{tM} is approximately equal to R , where t is the hidden feature vector dimension:

$$\hat{R}_{NM} = \sum_{t=1}^t S_{N_t} T_{tM} \quad (9)$$

To obtain accurate potential feature matrices of users and POIs, this paper obtains the feature matrices S_{N_t} and T_{tM} by iteratively minimizing the loss function, and the expressions are shown below:

$$Loss(S, T) = \sum_i^N \sum_j^M (R_{ij} - S_i T_j)^2 + \omega (\sum_i^N \|S_i\|^2 + \sum_j^M \|T_j\|^2) \quad (10)$$

To avoid overfitting add the regularization term, ω is the regularization parameter. To minimize the loss function, this paper iteratively optimizes the parameters by iteratively moving the variables along the direction of the negative

gradient of the loss function through the gradient descent method until convergence to the true scoring matrix.

D. Feature Integration

The final feature representations of users and POIs are obtained by fusing the associative relationship feature vectors E_u, E_p and text feature vectors C_u, C_p and scoring feature vectors S_u, T_p obtained from the LightGCN processing module, the text feature extraction module and the scoring data processing module:

$$U_{u_i} = E_{u_i} \oplus C_{u_i} \oplus S_{u_i}, P_{p_j} = E_{p_j} \oplus C_{p_j} \oplus T_{p_j} \quad (11)$$

The prediction of user u_i rating of interest point p_j is:

$$\hat{f}_{u_i} = U_{u_i} \otimes P_{p_j}^T \quad (12)$$

which is normalized to obtain \hat{f}'_{u_i} .

To minimize the difference between the predicted score \hat{f}_{u_i} and the true score f_{u_i} in the dataset, this paper uses a loss function to adjust the model parameters, which is calculated as follows:

$$loss = \sum_{i,j \in O} (\hat{f}_{i,j} - f_{i,j}) + \lambda \|\Theta\|^2 \quad (13)$$

where O denotes the number of samples in the training set; $\lambda \|\Theta\|^2$ is the canonical term and Θ is all trainable model parameters in the model, where λ is the adjustable coefficient of the canonical term, which is used to control the model parameters to prevent overfitting. This paper uses Adam to optimize the model parameters and minimize the loss function.

E. Geographical Distance Probability

The probability of a user signing up to a POI follows a power-law distribution of geographic distance probabilities, so this paper calculates the probability of a user signing up to a POI by modeling the distance between POIs with a power-law distribution, which is calculated as follows:

$$p[d(l_m, l_n)] = a \cdot d(l_m, l_n)^b \quad (14)$$

where a, b are the parameters of the power-law distribution and $d(l_m, l_n)$ is the distance between the point of interest l_m and l_n .

For the POI set $l_k (k=1, 2, \dots)$ around l_i , it has an effect on the probability of a user checking in to l_i . Therefore, under the influence of the set l_k , the probability of a user visiting l_i is calculated using the plain Bayesian method:

$$p^s = p[l_i | l_k] = \frac{p[l_k] \cdot \prod_{k=1, 2, \dots} p[d(l_i, l_k)]}{p[l_k]} \quad (15)$$

where the total number of check-ins is greater than l_i and the closest interest point is judged to have some influence on the set l_k of user check-ins to l_i . Finally, this paper normalizes the check-in probabilities as follows:

$$S_{i,j}^g = \frac{p_{i,j}^g}{z_i^g}, \text{ where } z_i^g = \max_{j \in L_k} p_{i,j}^g \quad (16)$$

F. Joint Recommendation Result

This paper integrates user similarity factors and geographic distance probabilities into a linear function, and then calculate the combined probability P of user u_i signing up to p_j :

$$P = (1-\alpha)\hat{f}_{i,j}' + \alpha S_{i,j}^g \quad (17)$$

where α and $1-\alpha(0 \leq \alpha \leq 1)$ represent the weights of the two factors, respectively, and when α is 0 means that the recommendation result is independent of the geographical distance factor.

IV. GEOLOCATION PROTECTION ALGORITHM

The KMG algorithm in this paper inherits the ideas of k-anonymity and location generalization, and changes the user's real location dynamically according to the locations of other users around. The individual location k-anonymity method has better protection for users who make location request services non-continuously, and poor protection for requesting continuous services, such as attackers who connect check-in records at different times to form trajectories, and can infer the user's true location by combining the direction and distance of the trajectories. Therefore, this paper combines k-anonymity techniques with location generalization techniques to provide better privacy-preserving capabilities for both individual service requests and consecutive requests.

A. Description of KMG Algorithm

The KMG protection algorithm is shown in Algorithm 1. The anonymous parameters k and L are set to 15 and 10, respectively. The specific scheme is as follows:

1) After receiving the location $l(x, y)$ of user u , the KMG protection algorithm finds the nearest remaining $k-1$ users according to his location and obtains their location information to generate the anonymous region, as in Fig. 2(a), and if the number of users in the anonymous region is less than k , a virtual user is generated. The center coordinates of k users in the anonymous region are calculated, and when the distance l between user u and the center coordinates is less than L , the center coordinates are replaced with the real locations of all users in the anonymous region and sent to the recommendation algorithm. Where x and y denote longitude and dimension respectively.

2) If there exists a user whose distance from the center coordinate is greater than L , the anonymous region is divided into n subanonymous regions, as shown in Fig. 2(b). Suppose the anonymous region is divided into 3 subanonymous regions.

The number of users k' in each subanonymity region is 5, and the first subanonymity region is formed by finding the nearest 4 users with user u as the center, and repeating the process of forming a subanonymity region by selecting a user as the center among the remaining users at random. The center coordinates are calculated based on the position of the user in each subanonymous region and sent to the recommendation algorithm instead of the position of the user in the subanonymous region.

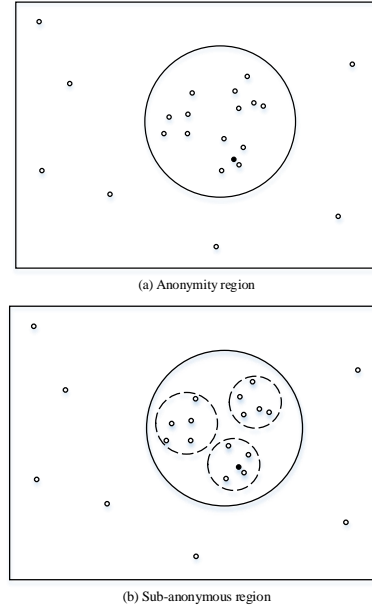


Fig. 2. Constructing anonymous regions and sub-anonymous regions.

3) *Virtual user generation*: Generating virtual users satisfies the principles of user distribution similarity and distance similarity. Firstly, the distance between existing users in the anonymous region is calculated and sorted, and the minimum distance between users is s . Virtual users are added in the space of any two users so that the distance between virtual users and users is not less than $s/2$ and does not overlap with the real user position.

Algorithm 1. KMG Algorithm

Input: k user locations $(L_1(x_1, y_1), L_2(x_2, y_2), \dots, L_k(x_k, y_k))$
Output: Returns the generalized location of all users in the anonymous region or sub-anonymous region (x', y')

```

1: if  $sum(L_i) < k$ :
2:      $sort \leftarrow dis(L_i, L_j), add(L_{k-i+1}, \dots, L_k)$ 
3:  $AR_0 \leftarrow Gen(L_i(x_i, y_i))$ 
4: if  $L \geq dis(L_i(x_i, y_i), AR_0(x_o^{(0)}, y_o^{(0)}))$ 
5:     return  $(x_o^{(0)}, y_o^{(0)})$ 
6: else:
7:      $k \% n == 0$  and  $n \geq 2$ 
8:      $(AR_1, AR_2, \dots, AR_n) \leftarrow Gen(L_a(x_a, y_a), L_b(x_b, y_b), \dots, L_n(x_n, y_n))$ 
9:     return  $(x_1^{(1)}, y_1^{(1)}), (x_2^{(2)}, y_2^{(2)}), \dots, (x_n^{(n)}, y_n^{(n)})$ 
10: end

```

B. Privacy Analysis based on Information Entropy

The uncertainty of information is proportional to the information entropy. Therefore, this paper uses Shannon entropy as a measure of the privacy-preserving ability of the model, and a larger Shannon entropy value indicates that it is more difficult for an attacker to distinguish the user's true location from the k locations. Shannon entropy is defined as follows:

Definition 1. Assuming that the random variable x takes values on a finite set X , the entropy of the random variable x is defined as

$$H(x) = -\sum_{x \in X} p(x) \log_2 p(x) \quad (18)$$

where $p(x)$ denotes the probability when the variable takes the value x .

The attacker successfully attacks and obtains the user location information in the anonymous region as an event set X . The successful attack on a single user's location information is an event x within the event set, and $p(x)$ denotes the probability of obtaining the user location information, which has a high uncertainty. The location protection method against this paper, this process needs to be divided into two cases. In the first case, if there is no sub-anonymous region, the attacker successfully obtains the anonymity region corresponding to the user and successfully identifies the user; in the second case, if there is a sub-anonymous region, the attacker successfully identifies the user on top of successfully obtaining the anonymity region and the sub-anonymous region. Since anonymous regions cannot be distinguished directly, the probability that the attacker successfully obtains the anonymous region where the user is located is $1/N$ (assuming there are N anonymous regions). If there is a sub-anonymous region, the probability that the attacker succeeds in obtaining the sub-anonymous region is $1/n$. The probability that the attacker succeeds in identifying the user from the anonymized region or sub-anonymous region in the second step is $1/k'$.

Therefore the probability of an attacker successfully obtaining user information is

$$p_u = \begin{cases} \frac{1}{kN} \\ \frac{a}{nk'N} \end{cases} \quad (19)$$

where a denotes the weight value for dividing sub-anonymous regions, and in this model if there exists $l > L$, then sub-anonymous regions are divided.

The entropy of the user can be obtained from (18) as:

$$H(u) = -p_{u_r} \log_2 p_{u_r} \quad (20)$$

The entropy of all users in the anonymous region is:

$$H(u) = \begin{cases} \sum_i H(u_r) \\ \sum_i \sum_j H(u_r) \end{cases} \quad (21)$$

Therefore, when a higher value of anonymization k is chosen, the higher the entropy value of the whole anonymization region, the higher the degree of privacy.

V. ANALYSIS OF EXPERIMENTAL RESULTS

A. Dataset Description

In this paper, we use the Yelp dataset, which is widely used for location-based social network research, as the experimental data. In this paper, the data set is divided by the filter condition that the total number of check-ins is greater than 10 and the number of comments is greater than 3. The data sparsity is 2.63×10^{-3} , and the check-in dataset contains user ID, POI ID, longitude, latitude, comments, ratings and time information. This paper randomly selects 80% of the check-in data from the dataset as training data, and the other 20% as test data.

B. Evaluation Metrics

To assess the recommendation quality of the recommendation $Top-K$, this paper uses $Precision@K$ and $Recall@K$ as evaluation criteria, where K denotes the number of recommended POIs ($K=5,10,15,20$).

$$Precision@K = \frac{A \cap B}{B} \quad (22)$$

$$Recall@K = \frac{A \cap B}{A} \quad (23)$$

Where A denotes the POI of user check-in in the test set and B denotes the POI in $Top-K$.

C. Comparison Approach

To verify the performance of the algorithm proposed in this paper, this paper compares the algorithm with the following three algorithms, all of which involve location privacy protection and point-of-interest recommendation. In this paper, the experimental dataset operation of the compared methods is the same as the method proposed in this paper, while generalizing the user location according to the location protection algorithm and using virtual location to make recommendations to users.

1) *USG[21]*: This method blends user preferences for POIs, social relationship influence and geographic influence, and calculates the probability of recommended POIs by the Naive Bayes method. Note that the USG model does not have location privacy protection.

2) *USD[26]*: This method is based on k-coordinate generalized user location and then POI recommendation, where the recommendation system is similar to the USG method. However, in terms of geolocation influencing factors, USD calculates the recommendation POI probability based on the check-in frequency.

3) *GLP[14]*: The GLP approach generalizes user locations based on population density (i.e., check-in density) and uses virtual locations to recommend POIs to users.

D. Experimental Parameters Adjustment

In order to achieve the best recommended performance, the experiment-related parameters are set in this paper as follows:

embedding vector dimension 64; BERT model output dimension d is 128; number of graph convolution iterations is 3; attention mechanism dimensions are 64 and 128, respectively; regularization coefficient λ is 1×10^{-4} . In addition, the experimental comparison model uses the parameter settings with the best results in the corresponding literature.

E. Regularization Coefficient

The regularization coefficient λ takes a range of $[1 \times 10^{-1}, 1 \times 10^{-2}, 1 \times 10^{-3}, 1 \times 10^{-4}, 1 \times 10^{-5}]$, which is essentially to adjust the hyperparameters in the LGCL model to make feature extraction more accurate and prevent overfitting. The parameters K , L and α are set to 5, 10 and 0.2, respectively. As shown in Table I, the values of $Precision@K$ and $Recall@K$ of the recommendation algorithm show a trend of increasing and then decreasing with the decrease of the regularization coefficient. When λ is 1×10^{-4} , both $Precision@K$ and $Recall@K$ reach the maximum; when λ is 1×10^{-5} , $Precision@K$ and $Recall@K$ decrease, and considering the overfitting problem, the regularization coefficient is 1×10^{-4} .

TABLE I. REGULARIZATION COEFFICIENT COMPARISON EXPERIMENT

Recommended performance	Regularization coefficient				
	1×10^{-1}	1×10^{-2}	1×10^{-3}	1×10^{-4}	1×10^{-5}
Precision@K	0.1473	0.1754	0.2041	0.2452	0.2270
Recall@K	0.0158	0.0182	0.0214	0.0232	0.0226

F. Geographical Distance Factor Weighting Analysis

The weight of the geographic distance factor has a significant effect on the performance of the recommendation algorithm. As shown in Fig. 3, the recommendation parameters K and L are set to 5 and 10, respectively. As the geographic weight α increases, $Precision@K$ and $Recall@K$ gradually decrease, and the recommendation algorithm achieves the best performance when α is 0.2.

G. L-value Analysis

The L value indicates the critical value of the distance between the real location and the center location of the anonymous region, which has a certain influence on the performance of the recommendation algorithm. The parameter α is set to 0.2. The range of the value of L is set to $[10, 60, 100, 150, 300]$. As shown in Fig. 4, with the fixed $Top-K$, the $Precision@K$ and $Recall@K$ of the recommendation algorithm gradually decrease as the parameter L increases. When the value range of L is $[10, 60, 100, 150]$, the recommendation performance is weakened, but there is no significant impact on the accuracy of the recommendation results. When K is 5, the Recall values when L is 60, 100 and 150 are 2.5%, 6.89%, 15.9% and 38.8% lower than the Recall values when L is 10, and the Precision is 0.9%, 1.2%, 2.3% and 5.8% lower, respectively. Therefore, $[10, 150]$ is a reasonable interval for the parameter L .

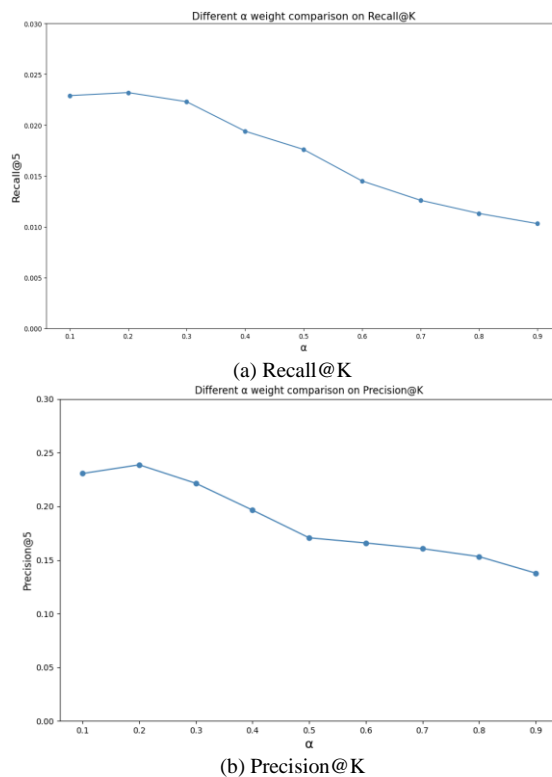


Fig. 3. Recall@K and precision@K of recommendation algorithm.

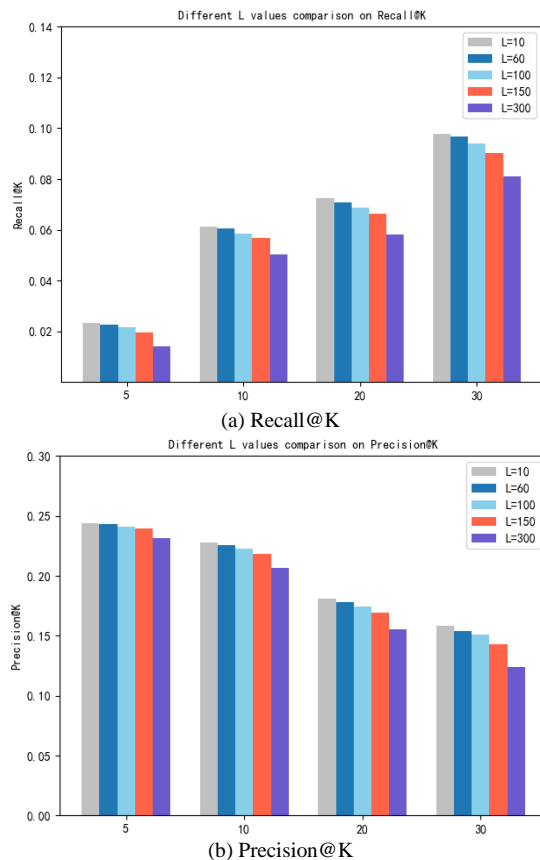


Fig. 4. Recall@K and precision@K of different l values.

H. Performance Comparison

Fig. 5 shows the recommended performance of Top-K (5,10,20,30) for all methods with α set to 0.2, respectively. Fig. 5(a) shows the *Recall@K* performance and Fig. 5(b) shows the effect of *Precision@K*. The performance of our method degrades as K increases. From the figure, it can be seen that the method proposed in this paper outperforms USG, USD and GLP in terms of accuracy and recall. By analyzing Fig. 5, it can be seen that the performance of the recommendation algorithms proposed in this paper are both higher than other comparison algorithms. Although the performance improvement of the method proposed in this paper is small compared with the USG method when K is taken as 5 (Recall and Precision are improved by 5.4% and 6.3%, respectively), the method proposed in this paper generalizes the processing for user location, and from the perspective of information entropy, the method in this paper has a better privacy protection ability. Thus a good balance between recommendation performance and privacy protection can be achieved by sacrificing a small recommendation performance.

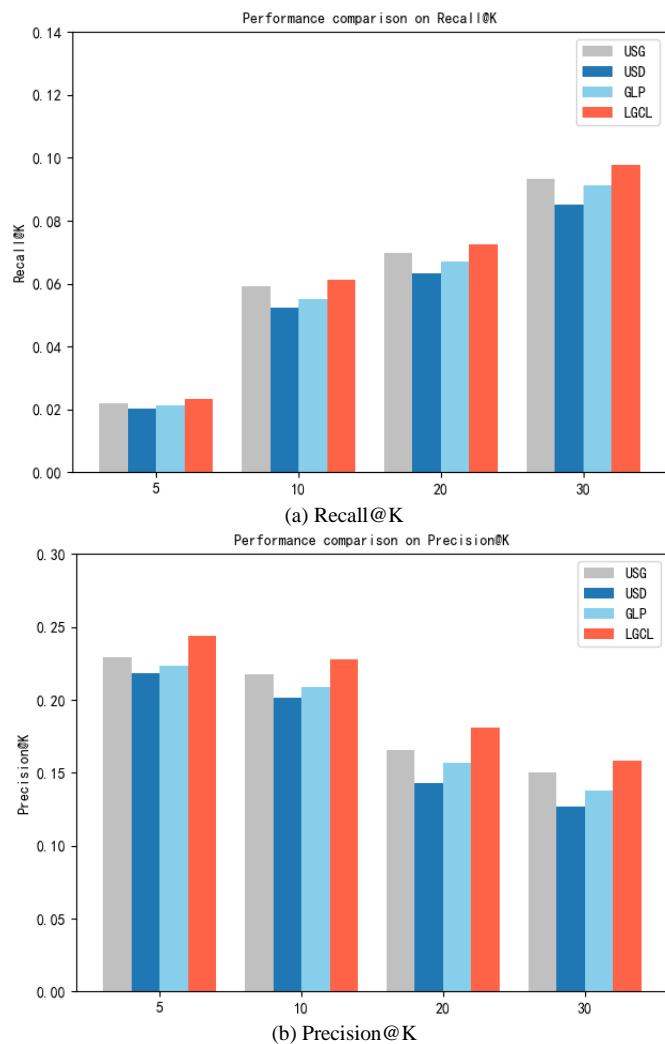


Fig. 5. Performance comparison of different methods.

VI. CONCLUSION

In this paper, we propose a recommendation algorithm that integrates GNN and geographic distance based on geographic location privacy protection. The algorithm learns implicit features from interaction information, comment information, and geographic distance for user recommendations, and integrates k-anonymity and generalization techniques for user location privacy protection. Finally, the information entropy theory analysis and experiments on real datasets show that our proposed recommendation system with integrated location protection can better adapt to interest in LBSN while effectively protecting user location privacy compared to other algorithms point recommendation. Therefore, it is meaningful and feasible to integrate privacy protection in point-of-interest recommendation systems. Since recommendation systems in location services involve geographic location and information records, the next step of research will focus on improving information record protection on the basis of ensuring recommendation performance.

REFERENCES

- [1] Hao P Y, Cheang W H, Chiang J H. Real-time event embedding for POI recommendation[J]. Neurocomputing, 2019, 349: 1-11.
- [2] Jannach D, Manzoor A, Cai W, et al. A survey on conversational recommender systems[J]. ACM Computing Surveys (CSUR), 2021, 54(5): 1-36.
- [3] Da'u A, Salim N. Recommendation system based on deep learning methods: a systematic review and new directions[J]. Artificial Intelligence Review, 2020, 53(4): 2709-2748.
- [4] Seo S, Huang J, Yang H, et al. Interpretable convolutional neural networks with dual local and global attention for review rating prediction[C]//Proceedings of the eleventh ACM conference on recommender systems. 2017: 297-305.
- [5] Sun J, Guo W, Zhang D, et al. A framework for recommending accurate and diverse items using bayesian graph convolutional neural networks[C]//Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2020: 2030-2039.
- [6] Hu K, Wu J, Li Y, et al. Fedgen: Federated learning-based graph convolutional networks for non-euclidean spatial data[J]. Mathematics, 2022, 10(6): 1000.
- [7] Ying R, He R, Chen K, et al. Graph convolutional neural networks for web-scale recommender systems[C]//Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining. 2018: 974-983.
- [8] Wang S, Hu L, Wang Y, et al. Graph learning approaches to recommender systems: A review[J]. arXiv preprint arXiv:2004.11718, 2020.
- [9] Wu S, Sun F, Zhang W, et al. Graph neural networks in recommender systems: a survey[J]. ACM Computing Surveys, 2022, 55(5): 1-37.
- [10] Yu Z, Wong R K, Chi C H. Efficient role mining for context-aware service recommendation using a high-performance cluster[J]. IEEE transactions on services computing, 2015, 10(6): 914-926.
- [11] Wu W, Liu J, Wang H, et al. Secure and efficient outsourced k-means clustering using fully homomorphic encryption with ciphertext packing technique[J]. IEEE Transactions on Knowledge and Data Engineering, 2020, 33(10): 3424-3437.
- [12] Zhang G, Qi L, Zhang X, et al. Point-of-interest recommendation with user's privacy preserving in an iot environment[J]. Mobile Networks and Applications, 2021, 26(6): 2445-2460.
- [13] Wang H, Huang H, Qin Y, et al. Efficient location privacy-preserving k-anonymity method based on the credible chain[J]. ISPRS International Journal of Geo-Information, 2017, 6(6): 163.
- [14] Huo Y, Chen B, Tang J, et al. Privacy-preserving point-of-interest recommendation based on geographical and social influence[J]. Information Sciences, 2021, 543: 202-218.

- [15] Kim D, Park C, Oh J, et al. Convolutional matrix factorization for document context-aware recommendation[C]//Proceedings of the 10th ACM conference on recommender systems. 2016: 233-240.
- [16] Li Z, Shen X, Jiao Y, et al. Hierarchical bipartite graph neural networks: Towards large-scale e-commerce applications[C]//2020 IEEE 36th International Conference on Data Engineering (ICDE). IEEE, 2020: 1677-1688.
- [17] Lin S, Runger G C. GCRNN: Group-constrained convolutional recurrent neural network[J]. IEEE transactions on neural networks and learning systems, 2017, 29(10): 4709-4718.
- [18] Shafqat W, Byun Y C. Incorporating similarity measures to optimize graph convolutional neural networks for product recommendation[J]. Applied Sciences, 2021, 11(4): 1366.
- [19] Wang X, He X, Wang M, et al. Neural graph collaborative filtering[C]//Proceedings of the 42nd international ACM SIGIR conference on Research and development in Information Retrieval. 2019: 165-174.
- [20] He X, Deng K, Wang X, et al. Lightgcn: Simplifying and powering graph convolution network for recommendation[C]//Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval. 2020: 639-648.
- [21] Ye M, Yin P, Lee W C, et al. Exploiting geographical influence for collaborative point-of-interest recommendation[C]//Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval. 2011: 325-334.
- [22] Song F, Ma T, Tian Y, et al. A new method of privacy protection: random k-anonymous[J]. IEEE Access, 2019, 7: 75434-75445.
- [23] Gruteser M, Grunwald D. Anonymous usage of location-based services through spatial and temporal cloaking[C]//Proceedings of the 1st international conference on Mobile systems, applications and services. 2003: 31-42.
- [24] Liu J, Jiang X, Zhang S, et al. FADBM: Frequency-aware dummy-based method in long-term location privacy protection[C]//2019 IEEE 25th International Conference on Parallel and Distributed Systems (ICPADS). IEEE, 2019: 384-391.
- [25] Ji Y, Gui R, Gui X, et al. Location privacy protection in online query based-on privacy region replacement[C]//2020 10th Annual Computing and Communication Workshop and Conference (CCWC). IEEE, 2020: 0742-0747.
- [26] NING Xueli, LUO Yonglong, XING Kai, ZHENG Xiaoyao. Frequent location privacy-preserving algorithm based on geosocial network[J]. Journal of Computer Applications, 2018, 38(3): 688-692.

Identity Authentication Protocol of Smart Home IoT based on Chebyshev Chaotic Mapping

Jingjing Sun^{1*}, Peng Zhang², Xiaohong Kong³

Hebi Institute of Engineering and Technology, Henan Polytechnic University, Hebi 458000, China^{1,2}

School of Mechanical and Electrical Engineering, Henan Institute of Science and Technology, Xinxiang 453003, China^{1,3}

School of Automobile and Transportation, Tianjin University of Technology and Education, Tianjin 300222, China²

Abstract—With the rapid development of the Internet of Things technology, the security of the Internet of Things is becoming increasingly important. Internet of Things (IoT) identity authentication is an important means to ensure network security. However, common identity authentication protocols have problems such as insufficient security factor and low efficiency. A smart home IoT identity authentication protocol based on Chebyshev chaotic map is proposed to improve the security of identity authentication. To solve the problem of low security of session key, the LAoCCM identity authentication protocol based on Chebyshev chaotic map is proposed to update session key. To solve the problem that the number of chaotic maps is too high, AEAD algorithm is introduced to reduce the number of chaotic maps. The results show that the average authentication error of LAoCCM authentication protocol is 0.00085, which is significantly smaller than that of EDHOC and ZKOP authentication protocols. Therefore, the proposed LAoCCM identity authentication protocol based on Chebyshev chaotic map has higher security performance and authentication efficiency, which can effectively meet people's needs for information security of smart furniture.

Keywords—Chebyshev; chaotic map; internet of things; identity authentication; LAoCCM protocol; key agreement; EDHOC protocol

I. INTRODUCTION

The fast advancement of connectivity, automation, and sensing technologies has made the IoT increasingly popular. In recent years, many different areas have made extensive use of the Internet of Things, such as smart home, industry, healthcare, and public safety, enhancing daily convenience for people [1]. Among them, the smart home is based on the family house as the carrier, integrating various communication technologies, network information technology and automatic control technology, etc., to realize the intelligence, comfort and high efficiency of home life. While realizing the smart home, a large amount of personal information is required to be registered and used, and the network system is subject to various malicious attacks and illegal intrusions. The resulting network security problems have become increasingly prominent. How to protect the privacy of individuals in the Internet of Things environment has become a need right challenge [2]. The most straightforward approach entails enhancing the IoT network system's security performance. Chaos, as a special form of movement, has significant randomness in the movement process. It is one of the commonly used methods to improve the cryptographic security

factor by using chaotic transformation in information encryption systems [3]. As a kind of chaotic sequence, the chaotic map based on Chebyshev has unique advantages in the construction of cryptosystem [4]. Based on the above background, the main issues that need to be addressed are as follows. Firstly, the key performance of IoT devices used in smart homes is weak, and in complex and ever-changing network environments, their passwords are easily cracked, posing a threat to system security. In addition, most IoT identity authentication protocols have limited computing and communication capabilities, resulting in significant resource consumption during the computing process, which limits the actual usage environment. Therefore, the research purpose of this article is to improve the weak security performance and low efficiency of the smart home IoT identity authentication process by designing an improved Chebyshev identity authentication protocol. At the same time, it reduces the computation and energy consumption in the identity authentication process. With the gradual opening up of the Internet, the problem of privacy leakage has become increasingly serious. Combining Chebyshev chaotic mapping with privacy protection meets user needs while enhancing protocol security. Through this identity authentication protocol, the security performance of IoT devices in smart homes is improved, ensuring the privacy and property security of internet users, and meeting people's higher network security needs. It has significant theoretical and practical significance for ensuring the stable development of network information security.

This study describes a smart home IoT identity authentication protocol for identity recognition. This identity authentication protocol is applicable to identity authentication and recognition of various types of smart homes, achieving the maintenance of network security and ensuring the safe development of network information. The main contributions of this study are as follows. Firstly, based on the Chebyshev chaotic map, a key agreement protocol is constructed using its characteristics to encrypt and decrypt information. Secondly, in response to the problem of high frequency of Chebyshev chaotic mapping, the study introduces the AEAD algorithm to reduce the frequency of chaotic mapping. By optimizing and updating the session key through Chebyshev chaotic mapping, the security and effectiveness of smart home identity authentication are ensured.

II. RELATED WORK

Chebyshev polynomials are extensions of cosine and sine functions derived from multiple angles, and are widely used in mathematics, physics, and science and technology. Abbasinezhad-Mood D et al. [5] proposed to construct a public key cryptosystem based on Chebyshev chaotic map for the security problems of shared keys in existing V2G networks, and anonymize the key agreement scheme. The results show that the efficiency of the proposed key agreement is significantly higher than other commonly used key agreements. Using Chebyshev polynomial theory, Yxh A et al. [6] studied the motion change of a dynamic model in an intelligent structure. Studies have found that under certain conditions, the model will lose its stability and produce periodic and chaotic motion. Qi RX et al. [7] presented a Chebyshev-based identity authentication scheme for real-time access to solve security and privacy issues. The results show that the method studied has less computational overhead and higher security performance. When Joachimiak M et al. solved the Laplace equation's inverse Cauchy-type puzzle, they used Chebyshev polynomials to deal with the problem, and thought about regularizing the issue, so that the problem was effectively solved [8]. Safdari H et al. used the Chebyshev configuration method to discretize the spatial fractions when solving the spatiotemporal fractional advection-diffusion equation (STFADE), which proved to be more accurate than other solutions [9]. Bozkaya C et al. [10] used the Chebyshev collocation method to solve the magnetohydrodynamics problem of an incompressible electric fluid in a square pipe, and discretized the governing equation by implementing the pre-assigned collocation points in the physical space. Body mechanics analysis can achieve relatively accurate results.

An information network system based on Internet technology and communication technology is known as the Internet of Things, which can break through the limitations of time and space to achieve barrier-free information interaction. As the result of information development, the Internet of Things technology brings with it problems such as network security and information leakage that need to be solved urgently. To stop harmful assaults that the Internet of Things could experience, Wan Z et al. [11] designed an IoT node roaming authentication model to improve the Internet of Things' security authentication performance. The findings indicate that this method can successfully fight off many network threats and has lower energy consumption. Liang W et al. [12] designed a radio frequency identification system based on multiple selection arbitrators, aiming to realize two-way identity authentication between the server and the user, and improve the security performance of the Internet of Things. The test results show that the identity authentication protocol has a better ability to resist external attacks, while improving system stability. Prasad SK and others used the cryptographic primitives based on unclonable functions to develop the ID card identification protocol of the IoT platform, and used the characteristics of unclonable function keys that are difficult to copy and unpredictable to enhance the security performance of ciphers. Experiments have proved that the identity authentication method is feasible, and it provides effective support in the device authentication and data security of the medical Internet of Things network [13]. Al-Naji F and others

classified different attacks on the IoT and proposed corresponding solutions, and used continuous identity authentication instead of static identity identification to improve the Internet of Things system's security performance [14]. Zhang Q et al. [15] create a secure channel using the key agreement for the intelligent terminal of the Internet of Things, to guarantee the intelligent terminal's security throughout information transfer. The findings demonstrate that the major agreement has increased security and lower time cost and energy consumption. Aiming at the problem of limited wireless sensor network nodes, Sharif AO et al. redesigned a key agreement scheme to build a secure channel between users and sensor nodes. The lightweight identity Internet of Things-based authentication protocol wireless sensor proposed by the research has been widely used [16].

To sum up, there are many related researches on using Chebyshev chaotic map to solve the problem. At the same time, corresponding improvement methods have also been proposed for the network security problems existing in the Internet of Things. However, for the Chebyshev polynomial in the password design system, the study of the identity authentication protocol for the Internet of Things is relatively insufficient, and the research on the use of this advantage to create the identity authentication protocol for the Internet of Things is relatively insufficient. Therefore, the research aims to develop a system for Internet of Things identity authentication based on Chebyshev polynomial chaotic mapping for this problem to improve security and effectiveness in IoT authentication process.

III. IDENTITY AUTHENTICATION PROTOCOL FOR SMART HOME IOT BASED ON CHEBYSHEV CHAOTIC MAPPING

A. IoT Authentication Protocol based on Chebyshev Chaotic Map

The perception layer, network layer, and application layer make up an IoT system. Commonly used IoT authentication protocols include identity authentication based on hash operation, IoT lightweight authentication protocol based on elliptic curve (Ephemeral Diffie-Hellman Over COSE, EDHOC) and IoT authentication protocol based on Chebyshev chaotic mapping. However, the identity authentication protocol based on hash operation does not use public key system, and only uses hash function for information saving calculation, so its security performance is low. Although the EDHOC identity authentication protocol has achieved some improvement in security, its computational cost is much higher than the hash authentication protocol, and the operation process is more complex. Among them, chaotic cryptography is widely used due to its low computational difficulty and high security. Chebyshev chaotic maps, as a class of chaotic sequences, have unique advantages in constructing cryptographic systems. Chebyshev polynomials have good pseudo-random characteristics, sensitivity to initial conditions and system parameters, and can effectively meet the principles of confusion and divergence in cryptographic design systems. Compared to traditional public key cryptography, this method has more advantages in computational storage [17]. Chebyshev polynomial chaotic maps are divided into three types, and the one used in the study belongs to the first type of Chebyshev

polynomial. A Chebyshev polynomial is a sequence of orthogonal polynomials arranged recursively. The Chebyshev polynomial's expression $T_n(x) : [-1, 1] \rightarrow [-1, 1]$ is shown in formula (1).

$$T_n(x) = \cos(n \cdot \arccos(x)) \quad (1)$$

In formula (1), n defined as an integer, $x \in [-1, 1]$, is a variable. The Chebyshev chaotic map's iterative connection is shown in formula (2).

$$T_n(x) = 2xT_{n-1}(x) - T_{n-2}(x), n \geq 2 \quad (2)$$

In formula (2), $T_0(x) = 1$, $T_1(x) = x$. The initial Chebyshev polynomial is expressed as $T_2(x) = 2^2 - 1$, $T_3(x) = 4x^3 - 3x$, $T_4(x) = 8x^4 - 8x^2 + 1$. From the initial expression, it can be known that the first kind of Chebyshev chaotic map contains two characteristics, namely, the chaotic property and the semigroup property. When integer $n \geq 1$, the polynomial map $T_n(x) : [-1, 1] \rightarrow [-1, 1]$ is n a density-invariant chaotic map based on Lyapunov exponents. The semigroup property means that $r, s \in N$ the relation in formula (3) exists for any time.

$$T_r(T_s(x)) = T_s(T_r(x)) = T_{rs}(x) \quad (3)$$

In formula (3), $x \in [-1, 1]$. When x in the interval $(-\infty, +\infty)$, the extended Chebyshev polynomial can be obtained, as shown in formula (4).

$$T_n(x) = (2xT_{n-1}(x) - T_{n-2}(x)) \pmod{p} \quad (4)$$

In formula (4), $n \geq 2$, P represents a large prime number, mod representing the P modulo operation. When using the Chebyshev polynomial chaotic map for identity verification, it is necessary to design the corresponding public-key encryption, and the Diffie-Hellman key agreement method of the specific public key cryptosystem needs to convert the encryption method used into the corresponding cryptographic algorithm [18]. During identity authentication, the information needs to be identified, but to guarantee the security and effectiveness of the information, the information should be encrypted. Party A and Party B randomly select two digital sums, a and b after calculation, Party A obtains the relationship between the polynomial and a as shown in the formula (5) shown.

$$R_a = T_a(x) \pmod{p} \quad (5)$$

After calculation, Party B obtains the relationship between Chebyshev polynomial and b formula (6)

$$R_b = T_b(x) \pmod{p} \quad (6)$$

The calculation results are passed to both parties alternately. A encrypts the calculation result of B and the calculation result of his own random number that he has received a , and obtains a new encryption method as shown in formula (7).

$$SR_a = T_a(b) = T_a(T_b(x)) \quad (7)$$

In the same way, B encrypts the calculation result of A and the calculation result of its own random number, and obtains a new encryption method as shown in formula (8).

$$SR_b = T_b(a) = T_b(T_a(x)) \quad (8)$$

Formula (7) and formula (8) represent $T_n(x)$ Chebyshev polynomials with integer n order and parameters. x After the information passed is encrypted, the computation between the nonce and the session key SR is indistinguishable. On this basis, a Chebyshev chaotic mapping-based smart home IoT identity authentication technique is created. The authentication process is combined with the zero-knowledge proof proposed by Schnorr, which is defined as the ZKOP protocol in the study, so that the user's identity may be verified between them and the server. The specific process is shown in Fig. 1.

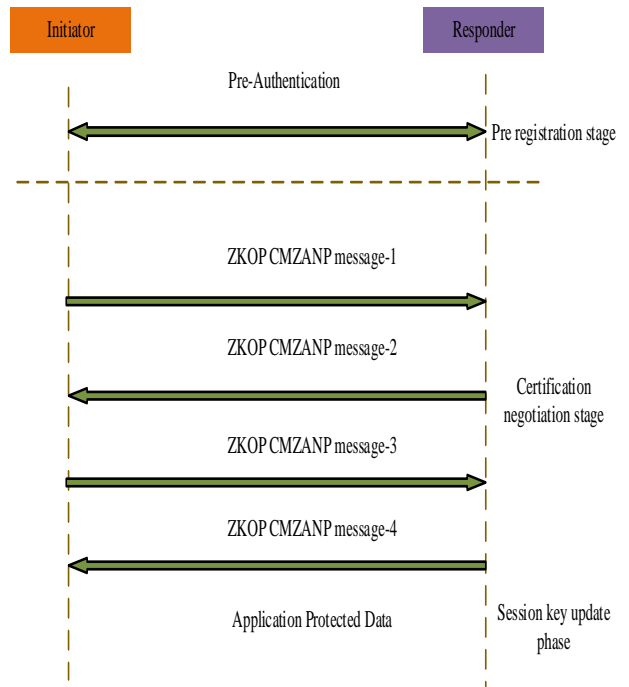


Fig. 1. Identity authentication protocol based on zero knowledge proof.

The identity authentication protocol process in Fig. 1 needs to experience multiple information interactions, which will affect the efficiency of identity authentication; at the same time, combined with zero-knowledge proof, it is concluded that the number of chaotic maps of Chebyshev chaotic maps is too high, which will aggravate the identity of IoT devices. Authentication pressure; in addition, in the Chebyshev chaotic map authentication process combined with zero-knowledge

proof, the session key only relies on automatic update, which cannot guarantee the security of information authentication.

B. LAoCCM IoT Identity Authentication Protocol based on Improved Chebyshev Chaos Mapping

Chebyshev chaotic map has good pseudo-random characteristics, which can meet the principle of confusion and diffusion in cryptographic design system. The application of chaotic cryptography based on Chebyshev polynomials has received more attention. However, the identity authentication protocol of Chebyshev chaotic map also has some shortcomings in the application process. First of all, too much information round-trip interaction directly affects the operation efficiency of the protocol. Secondly, the mapping times of chaotic mapping are too high, which will increase the pressure of IoT devices. Aiming at the above problems, the Lightweight Authentication over Chebyshev Chaotic Map (LAoCCM) is proposed to enhance the security capabilities of keys. The LAoCCM identity authentication protocol reduces the computation time of Chebyshev chaotic map by redesigning the authentication information generation and verification process, and completes the key update [19]. The LAoCCM protocol consists mostly of three steps, namely the pre-registration phase, the authentication negotiation phase and the session key update phase after the authentication negotiation. Table I displays several parameters related to the LAoCCM protocol procedure.

TABLE I. PARAMETER SYMBOLS AND MEANINGS IN LAOCCM PROTOCOL

Symbol	Meaning
DEV, GWN	IoT device and gateway GWN
P_A, P_B	Temporary public key of Chebyshev on both sides of the device and gateway
P_D, P_G	Authentication public key of both device and gateway
$T_n(x)$	Chebyshev polynomials of order n with parameter x
AEAD (K; Plaintext)	Encrypt additional data using the key K generated from the shared key
tD	Current time point of equipment
Extract	Production function of random key
Expand	Production function of symmetric key

During the LAoCCM protocol authentication process, both the IoT and the gateway hold the authentication public key $\langle D, P_D \rangle$ and authentication private key used by $\langle G, P_G \rangle$ them, where D and G represent the authentication private key of both, P_D and P_G represent the authentication public key of both. At the same time, the identifier ID_CRED_D sum of the corresponding authentication key is also required ID_CRED_G to retrieve the authentication key. In the authentication and key negotiation stage of LAoCCM, the first step is to judge the timeliness of the authentication information. At this time, it is necessary to use the Internet of Things to generate the current

timestamp t_{D1} for judgment, define the random number A , and calculate the Chebyshev polynomial such as formula (9) shown.

$$P_A = T_A(x) \pmod p \quad (9)$$

Before key negotiation, both parties need to determine the cipher suite SUITE-1 to use. Each SUITE-1 determines a set of cipher algorithms, including AEAD algorithm, hash algorithm, ECDH algorithm, etc. to encrypt information. However, in order to reduce the number of protocol round trips and the number of messages, the message round trip process can be simplified and the application auxiliary data can be transmitted with the message. Therefore, the application can use the AEAD algorithm in the selected cipher suite to protect the data to ensure the information security during the message round-trip process [20]. The IoT device will determine the used cipher suite suite-D, and pass the generated parameters to the gateway GWN, which is defined as Message 1; the second step is that the gateway checks the timeliness of the received information, and checks whether it conforms to the message in Message1. The cipher suite suite-D is used to judge the information. If the verification fails, the entire protocol is terminated; the gateway determines random number B and computes the Chebyshev polynomial if the verification is successful, as shown in formula (10).

$$P_B = T_B(x) \pmod p \quad (10)$$

In this stage, use Extract function and Expand function to decrypt the key, and use Extract to generate intermediate key (PRK). The gateway calculates the public information P_AB according to formulas (9) and (10), and thus two PRKs can be obtained, which are defined as PRK-1 and PRK-2 respectively. After the intermediate key is generated on the gateway side, the symmetric key K to be used needs to be generated, and the symmetric keys generated by the GWN are respectively defined as K-1 and K-2. After completing the above operations, use the gateway to construct a message authentication code, and use the AEAD algorithm in suite-D to encrypt K-1 and K-2, generate Mesange2 after encryption, and transmit the encrypted and authenticated information to the device for verification, and Verify Message2. The verification process is shown in Fig. 2.

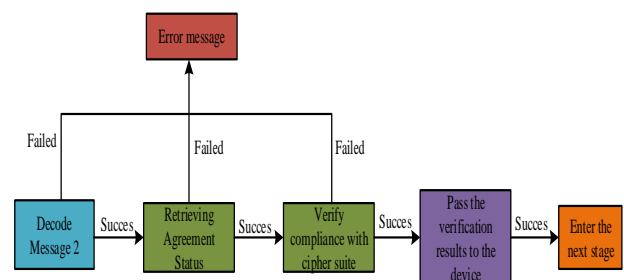


Fig. 2. Message 2 validation process.

From the process in Fig. 2 that if the verification fails, the authentication protocol is terminated; if the identities of both parties pass the verification, formula (11) is obtained.

$$P_{AB} = T_A(P_B) = T_B(P_A) \quad (11)$$

After the server completes the PRK generation, the client must decrypt and authenticate the PRK, that is, the device ought to use the hash function to generate the symmetric key K-1, as shown in formula (12).

$$K_{1'} = \text{Expand}(PRK_{1}, H(\text{Message}_{1} \| t_{G1} \| P_{-}B)) \quad (12)$$

In formula (12), the $H(\cdot)$ hash function is represented, and the same can be obtained through the hash function to regenerate the symmetric key K-2, which PRK_{2} represents the intermediate key, as shown in formula (13).

$$K_{2'} = \text{Expand}(PRK_{2}, H(\text{Message}_{2} \| t_{G1} \| P_{-}B)) \quad (13)$$

The encrypted information is then decrypted and verified. If the decryption is successful, the gateway's identity is legitimate, and following identity authentication, the next stage of information transmission can be entered; if the decryption fails, it is considered that the identity of the gateway cannot pass the authentication, and the identity authentication protocol is terminated. After the device completes the processing of the gateway data, the encrypted information is decrypted and verified by the key K. If the verification is passed, the gateway also considers that the identity information is legal and can proceed to the next session key construction; if the verification fails, If the device authentication fails, the gateway will immediately terminate the authentication process. After the two parties complete the authentication and key negotiation process, the LAoCCM session key update phase is entered, that is, the two communicating parties perform subsequent data encryption and authentication through the session key. When updating the session key, after any party requests a key change, select a random number n , calculate its temporary public key and private key results, and pass the calculation results to GWN; and the receiver first needs to verify the information Timeliness and random challenges are calculated as shown in Equation (14).

$$P_{A} = T_A(x) \pmod{p} \quad (14)$$

Finally, verify the validity of the calculation's outcome. If the result is not valid, the information receiver terminates the key update operation; if the result is valid, it is considered that the applicant of the application has passed the identity verification and can start to generate the updated session key, as shown in formula (15) shown.

$$P_{GU} = T_G(T_u(x)) \quad (15)$$

After the session key is updated, the two parties realize the identity verification, and the update of the session key is completed, and the updated channel can be entered to continue the identity information authentication. In Fig. 3, the particular procedure is displayed.

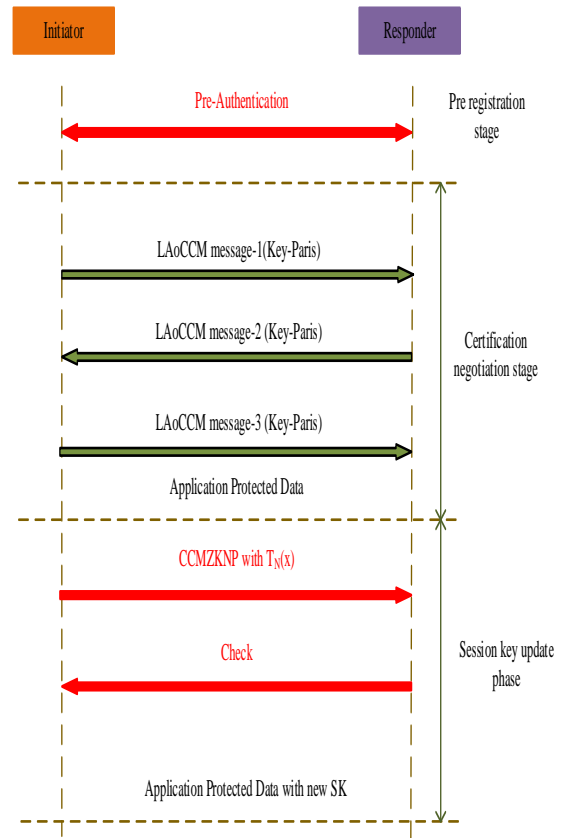


Fig. 3. LAoCCM authentication protocol process.

After completing bidirectional authentication and key negotiation based on the above process, both parties encrypt and communicate subsequent data through shared session keys. If the session key needs to be updated at this time, either party will initiate a request to change the key. Since the above process has already implemented the complete process of identity authentication, it is necessary to ensure the security of the current session key, the legality of the identity of the session key applicant, and a brief update process when updating the session key. The LAoCCM IoT identity authentication protocol based on Chebyshev chaotic mapping realizes identity authentication by constructing key pairs multiple times to transmit information and encrypting the information.

IV. PERFORMANCE ANALYSIS OF LAOCCM IoT AUTHENTICATION PROTOCOL BASED ON CHEBYSHEV CHAOTIC MAP

Before you begin to format your paper, first write and save the content as a separate text file. Keep your text and graphic files separate until after the text has been formatted and styled. Do not use hard tabs, and limit use of hard returns to only one return at the end of a paragraph. Do not add any kind of pagination anywhere in the paper. Do not number text heads-the template will do that for you. Finally, complete content and organizational editing before formatting. Please take note of the following items when proofreading spelling and grammar: Aiming at actual performance of the LAoCCM IoT Chebyshev chaotic mapping-based authentication protocol, the data in the

Libsodium cryptographic library is used to evaluate the effectiveness of the LAoCCM IoT authentication protocol based on Chebyshev chaotic mapping. In view of the scalability of the protocol proposed in the study, the client obtains data by accessing nodes. However, when the node data increases, the workload of the client will also increase. In the LAoCCM protocol, the server completes most of the work of the protocol. Therefore, when the number of nodes increases, it only increases the burden of the server and has a small impact on the client. Because the server performance is generally high, the protocol proposed in the study has good scalability. The errors in the authentication negotiation process of the commonly used identity authentication protocol EDHOC, the Chebyshev chaotic map identity authentication protocol combined with zero-knowledge proof, and the LAoCCM scheme are compared, shown in Fig. 4.

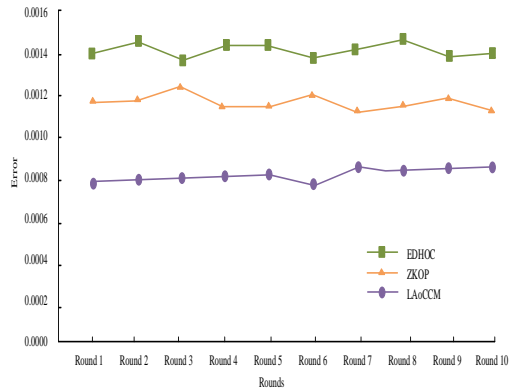


Fig. 4. Error comparison of three identity authentication protocols.

From Fig. 4 that the error range of the EDHOC IoT identity authentication protocol is between 0.0013 and 0.0015, and the average error is 0.0014; the ZKOP identity authentication protocol error range is between 0.0011 and 0.0013, and the average error is 0.0012. The LAoCCM IoT proposed by the study. The error range of the authentication protocol is between 0.0008 and 0.0009, and the average error is 0.00085. The average error is 0.00055 less than the average error of the EDHOC authentication protocol and 0.00035 less than the average error of the ZKOP authentication protocol. The proposed LAoCCM authentication protocol is more accurate. Comparing the execution times of the above three identity authentication protocols, the results obtained are shown in Fig. 5.

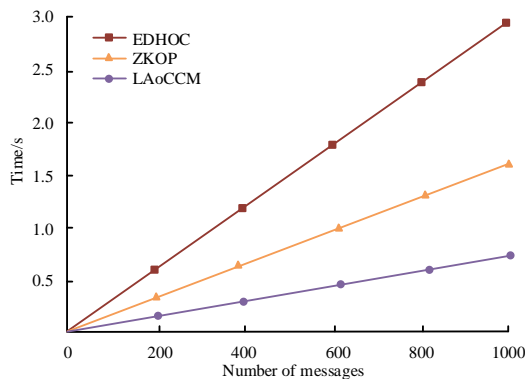


Fig. 5. Comparison results of execution time message authentication.

Fig. 5 shows that as the amount of information increases, the execution costs of the three identity authentication protocols will increase, but the time required by each is drastically different. The EDHOC identity authentication protocol takes the most time, and as the amount of information increases, the time pressure for its operation increases. Taking 1000 pieces of information as an example, the identity authentication protocol takes 3 seconds; the ZKOP identity authentication protocol is significantly lower than the EDHOC protocol, the time required for 1000 pieces of information authentication is 1.5 seconds; the LAoCCM identity authentication protocol proposed by the study takes the least time, and it only takes 0.75 seconds for 1000 pieces of information to run; it is 2.25 seconds lower than the EDHOC protocol and 0.75 seconds lower than the ZKOP protocol second. It can be concluded that the authentication efficiency of the LAoCCM identity authentication protocol is optimal, which can significantly increase the effectiveness of authenticating. Comparing the calculation overhead results of the three methods during the key agreement stage, the results are shown in Fig. 6.

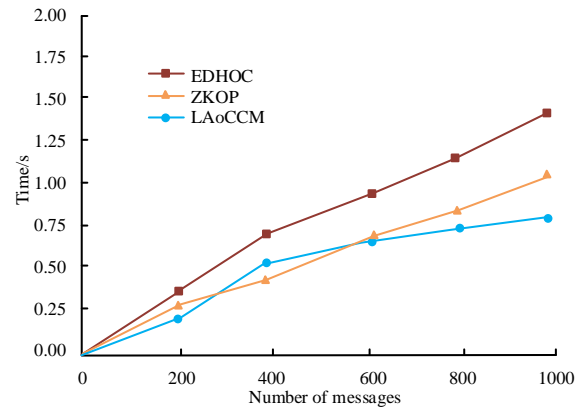


Fig. 6. Comparison results of computational overhead in key agreement phase.

From Fig. 6 that the time overhead of the three identity authentication protocols in the key negotiation stage increases with the increase of the amount of information. Among them, the EDHOC identity authentication protocol has the most time overhead, and the time overhead is 1.5 seconds when there are 1000 messages; the second is the ZKOP identity authentication protocol, which has more time overhead in the key negotiation stage, and the time overhead of 1000 messages is 1.15 seconds. The LAoCCM identity authentication protocol proposed by the research has the least time overhead in the key negotiation stage, and the time overhead is greater than that of the ZKOP identity authentication protocol only when there are about 400 pieces of information. The time overhead of the protocol at this stage is 0.75 seconds and 0.4 seconds lower, respectively. From the time cost of the key negotiation stage, the research proposes that the identity authentication protocol has faster key negotiation efficiency. In the realization process of the identity authentication protocol, the information's encryption, decryption, and interactive calculating programs are where the majority of the overhead is located. Table II displays the results of calculating the time overheads of the three authentication techniques.

TABLE II. COST COMPARISON OF DIFFERENT PROTOCOLS

Scheme	Time cost(ms)			Total	Communication cost(bits)
	User	GWN	SN		
EDHOC	/	10T CCM	10T CCM	10T CCM	/
ZKOP	2T CCM +1T E +9T H	1T CCM +5T H	/	3T CCM +2T E +14T H	994
LAoCCM	/	3T CCM +2T S +3T H	3T CCM +2T S +3T H	6T CCM +4T S +6T H	/

In Table II, TCCM represents the cost of Chebyshev chaotic mapping, TCCM=0.0025ms. TS represents symmetric encryption process, TS=0.0021ms. TH represents hash operation, TH=0.0947ms. TE represents the fuzzy extraction process, TE=0.1ms. In the above table, the overhead of the EDHOC protocol in the user phase Negligible, there are 10 times Chebyshev chaotic mapping operations in both the GWN process and the SN process, and the total overhead is 10 Chebyshev chaotic mapping operations; ZKOP protocol has 2 Chebyshev chaotic mapping operations in the User phase, 1 hash operation, 9 fuzzy extraction processes, 1 Chebyshev chaotic mapping operation and 5 hashing operations in the GWN stage; LAoCCM protocol has 3 Chebyshev chaotic mapping operations in the GWN stage and the SN stage respectively, 2 hash operations and 3 fuzzy extractions, in general, the proposed LAoCCM protocol has less time overhead and higher efficiency. To further evaluate the actual operational efficiency of the proposed solution, the study statistically analyzed the computational costs of the three protocols on the server and user sides, as shown in Fig. 7.

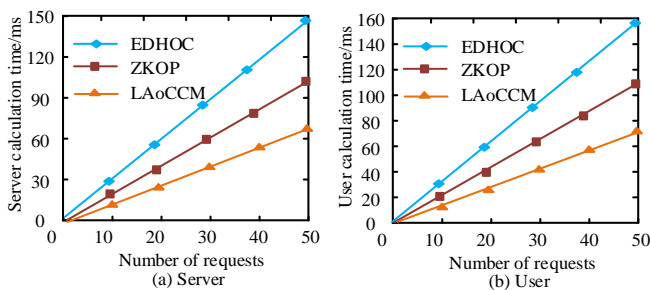


Fig. 7. Time cost of different protocols.

From Fig. 7, the time cost on both the server and user sides shows the same trend, with EDHOC having the highest time cost, followed by ZKOP. The proposed protocol scheme has the lowest time cost. When the number of visits reaches 50, the time cost of LAoCCM protocol on the server and user sides is 68ms and 75ms, respectively, significantly lower than the other two methods. Compared with the other two identity authentication protocols, in addition to saving the authentication time as much as possible, the identity authentication protocol's primary goal is to secure the privacy and security of information exchange, including forward security, post-item security, confidentiality, anti-camouflage attacks and Anti-replay attacks, etc. The security performance of the three authentication protocols is compared, and Table III presents the outcomes.

TABLE III. SECURITY ATTRIBUTES OF THREE AUTHENTICATION PROTOCOLS

Safety function	EDHOC	ZKOP	PAP	CHAP	EAP	LAoCCM
Forward secret	No	No	Yes	Yes	Yes	Yes
Backward security	Yes	Yes	Yes	Yes	Yes	Yes
Confidentiality	Yes	Yes	Yes	Yes	No	Yes
Replay attacks	Yes	Yes	No	Yes	Yes	Yes
Impersonation attack	Yes	Yes	Yes	Yes	Yes	Yes
Eavesdropping attacks	Yes	No	Yes	Yes	Yes	Yes
Denial of service attack	Yes	Yes	No	No	Yes	Yes

In Table III, Yes indicates that the protocol has the security performance, and No indicates that the protocol does not have the security performance. It can be seen from Table III that EDHOC protocol does not have forward security performance. ZKOP protocol does not have forward security and security against eavesdropping attacks. PAP protocol does not have replay security performance and denial of service attack performance. CHAP protocol does not have denial of service attack performance. The performance of EAP protocol in confidentiality is poor. Attackers can break the security performance of protocol transmission and obtain identity information through eavesdropping, modification, counterfeiting and other attacks. The LAoCCM protocol proposed in the study has passed the test in the aspects of forward security, backward security, confidentiality, replay attack, camouflage attack, eavesdropping attack and denial of service attack, and its security performance is significantly higher than that of EDHOC and ZKOP identity authentication protocols. The security of identity authentication protocols directly determines their availability in actual network environments. To verify the performance of the identity authentication protocol proposed in the research, 500 pairs of verification and response requests were constructed, and three identity authentication protocols were used to verify these data separately. The validation accuracy of the three methods in the application environment was statistically analyzed, and the results are shown in Fig. 8.

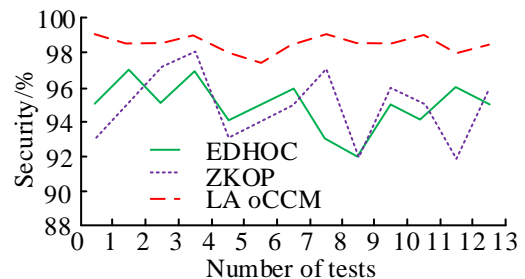


Fig. 8. Security analysis of identity authentication protocol.

As shown in Fig. 8, the overall security fluctuation of the EDHOC identity authentication protocol is the highest, followed by ZKOP. The LAoCCM identity authentication protocol proposed in the study has the highest security, with

the smallest difference in security in multiple experiments, with an average security performance of 99.12%. The average security performance of EDHOC and ZKOP in multiple experiments is 95.37% and 95.46%, which is significantly lower than the LAoCCM identity authentication protocol proposed in the study. Therefore, the security of this method can effectively meet the requirements of practical application environments. The cost of authentication will be impacted by how much energy is used by the identity authentication mechanism. The energy costs consumed by the three identity authentication protocols during communication and computing are tallied. Fig. 9 displays the outcomes.

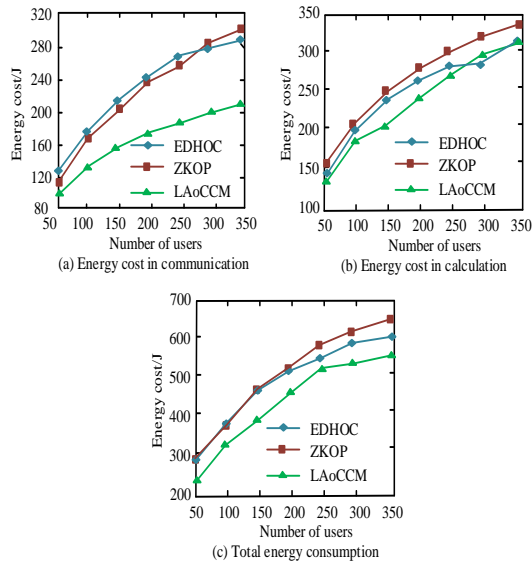


Fig. 9. Comparison of energy consumption of different authentication protocols.

Fig. 9 displayed that the quantity of energy used in the three different IoT identity authentication protocols is quite different during the operation process. From Fig. 9 (a), when the information is transmitted in the three methods, the number of users who are switched on causes a progressive rise in energy usage. When the number of users is less than 300, the energy consumption of the EDHOC protocol is greater than the ZKOP protocol. As the number of users increases, the energy consumption of EDHOC tends to grow slowly. After more than 300 users, the energy consumption is less than that of the ZKOP protocol; the study suggested that the LAo CCM identity authentication technique used in the communication process used energy is significantly smaller than the other two methods, and as the number of users has grown, the energy consumption cost gradually decreases. When the number of users reaches 350, the energy consumption of EDHOC is 290J, the energy consumption of ZKOP protocol is 300J, and the energy consumption of LAoCCM protocol is 210J, which are 80J and 90J lower than the other two methods respectively. From Fig. 9 (b), the energy consumption gap between the three authentication protocols is relatively small when calculating, and the ZKOP protocol has the highest energy consumption; the EDHOC protocol has a large fluctuation when the number of users approaches 300, and the energy consumption significantly decreased, and then gradually increased; the energy consumption of the LAOCCM identity authentication

protocol proposed in the study is still smaller than the other two methods. Fig. 9 (c) shows the total energy consumption of the three identity authentication protocols during the communication and computing stages. When the number of users is below 200, the energy consumption of the EDHOC and ZKOP protocols is basically the same. When the number of users is above 200, the energy consumption of the ZKOP protocol is higher than that of the EDHOC protocol. Overall, the proposed LAoCCM identity authentication protocol has the least energy consumption and the best performance. The study above shows that the energy cost of the communication's LAoCCM authentication methodology and calculation process is smaller than the other two methods, and has better application performance.

Identity authentication security is an important influencing factor for the development of IoT devices. As an important aspect of IoT security, the storage, computing, and communication capabilities of most IoT devices are limited. To compensate for the shortcomings of existing identity authentication protocols, the IoT identity authentication protocol based on improved Chebyshev chaotic mapping is proposed. By updating the session key and using the AEAD algorithm to solve the problem of high number of chaotic mappings, the security, computing, and storage capacity of the identity authentication protocol are optimized. The study compares the EDHOC and ZKOP protocols. Comparison shows that the average error of EDHOC is 0.0014, the average error of ZKOP is 0.0012, and the average error of LAoCCM is 0.00085, which is significantly better than EDHOC and ZKOP. When the number of visits reaches 50, the time cost of the LAoCCM identity authentication protocol proposed in the study on the server and user sides is 68ms and 75ms, respectively, which is significantly lower than the other two methods. The total energy consumption of the three identity authentication protocols in the communication and computing stages of the LAoCCM protocol. When the number of users is below 200, the energy consumption of the EDHOC and ZKOP protocols is basically the same. When the number of users is above 200, the energy consumption of the ZKOP protocol is higher than that of the EDHOC protocol. Through the above comparison, it was found that the proposed identity authentication protocol based on improved Chebyshev chaotic mapping has better performance and can effectively meet practical application requirements.

V. CONCLUSION

Network security is the main problem brought about by the development of information technology. At this stage, the key scheme used in IoT devices has limited security performance and is prone to leakage risks. Cryptography is an important means to ensure network security. By encrypting information and constructing a two-party identity information authentication protocol to ensure basic information security. Chebyshev chaotic map plays a good role in this aspect, but the traditional Chebyshev chaotic map has problems such as too many mapping times and low efficiency. Meet the needs of identity authentication in smart homes. According to the experimental findings, the average authentication mistake of LAoCCM authentication protocol is 0.00085, which is 0.00055 less than that of EDHOC and 0.00035 less than that of ZKOP,

which is significantly smaller than that of EDHOC and ZKOP. It only takes 0.75 seconds, which is 2.25 seconds lower than the EDHOC protocol and 0.75 seconds lower than the ZKOP protocol; the energy consumption of LAoCCM in communication is 210J, which is 80J and 90J lower than the other two methods respectively. Therefore, the Chebyshev chaotic map-based suggested LAoCCM identity authentication mechanism provides superior security performance, higher authentication efficiency and more ideal practical application performance. However, there are still deficiencies in the study. First of all, the research and design of the Internet of Things identity authentication protocol needs to go through many information round-trip steps, and the operation is too complex. Secondly, the public key system of Chebyshev polynomial chaotic map is still developing, so there is still a lack of comprehensive and effective security proof. In the follow-up research, this problem needs to be studied and optimized.

REFERENCES

- [1] W. L. Tai, Y. F. Chang, P. L. Hou, "Security analysis of a three-factor anonymous authentication scheme for wireless sensor networks in internet of things environments," *International Journal of Network Security*, vol. 21, no. 6, pp.1014-1020, 2019.
- [2] Z. Wan, Z. Xu, S. Liu, "An internet of things roaming authentication protocol based on heterogeneous fusion mechanism," *Journal of Network Security*, vol. 21, no. 6, pp.1014-1020, 2019.
- [3] M. Z. Talhaoui, X. Wang, M. A. Midoun, "Fast image encryption algorithm with high security level using the Biilban chaotic map," *Journal of Real-Time Image Processing*, vol. 18, no. 1, pp.85-98, 2021.
- [4] J. Attaullah, S. Aanan, Q. A. Tariq, M. Dept, P. Islamabad, "Cryptosystem techniques based on the improved Chebyshev map: an application in image encryption," *Multimedia Tools and Applications*, vol. 78, no. 22, pp.31467-31484, 2019.
- [5] D. Abbasinezhad-Mood, A. Ostad-Sharif, S. M. Mazinani, M. Nikooghadam, "Provably secure escrow-less Chebyshev chaotic map-based key agreement protocol for vehicle to grid connections with privacy protection," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 12, pp. 7287-7294, 2020.
- [6] A. Yxh, A. Kfz, B. Wz, B. Swy, "Nonlinear dynamics and dynamic instability of smart structural cross-ply laminated cantilever plates with MFC layer using zigzag theory," *Applied Mathematical Modelling*, vol. 79, pp.639-671, 2020.
- [7] R. X. Qi, S. Ji, J. Shen, P. Vijayakumar, N. Kumar, "Security preservation in industrial medical CPS using Chebyshev map: An AI approach," *Future Generation Computer Systems*, vol. 122, no. 1, pp.52-62, 2021.
- [8] M. Joachimiak, M. Ciakowski, A. Frckowiak, "Stable method for solving the Cauchy problem with the use of Chebyshev polynomials," *International Journal of Numerical Methods for Heat and Fluid Flow*, vol. 30, no. 3, pp.1441-1456, 2019.
- [9] H. Safdari, Y. E. Aghdam, J. F. Gomez-Aguilar, "Shifted Chebyshev collocation of the fourth kind with convergence analysis for the space-time fractional advection-diffusion equation," *Engineering with Computers*, vol. 38, no. 2, pp.1409-1420, 2022.
- [10] C. Bozkaya, N. Türk, "Chebyshev spectral collocation method for MHD duct flow under slip condition," *Progress in Computational Fluid Dynamics*, An International Journal, vol. 22, no. 2, pp.118-129, 2022.
- [11] Z. Wan, Z. Xu, S. Liu, W. C. Ni, S. T. Ye, "An internet of things roaming authentication protocol based on heterogeneous fusion mechanism," *IEEE Access*, vol. 8, no. 99, pp.17663-17672, 2020.
- [12] W. Liang, S. Xie, J. Long, "A double puf-based RFID identity authentication protocol in service-centric internet of things environments," *Information Sciences*, vol. 503, pp.129-147, 2019.
- [13] S. K. Prasad, B. Malarkodi, "A decentralized framework for device authentication and data security in the next generation internet of medical things," *Computer Communications*, vol. 180, no. 12, pp.146-160, 2021.
- [14] F. Al-Naji, R. Zagrouba, "A survey on continuous authentication methods in Internet of Things environment," *Computer Communications*, vol. 163, no. 11, pp.109-133, 2020.
- [15] Q. Zhang, L. Zhu, Y. Li, Z. R. Ma, J. L. Yuan, J. Zheng, S. Ai "A group key agreement protocol for intelligent internet of things system," *International Journal of Intelligent Systems*, vol. 37, no. 1, pp.699 -722, 2022.
- [16] A. O. Sharif, H. Arshad, M. Nikooghadam, D. A. Mood, "Three party secure data transmission in IoT networks through design of a lightweight authenticated key agreement scheme," *Future Generation Computer Systems*, vol. 100, no. 11, pp.882-892, 2019.
- [17] W. L. Tai, Y. F. Chang, P. L. Hou, "Security analysis of a three-factor anonymous authentication scheme for wireless sensor networks in Internet of things environments," *International Journal of Network Security*, vol. 21, no. 6, pp.1014-1020, 2019.
- [18] K. Nath, P. Sarkar, "Efficient elliptic curve diffie-hellman computation at the 256-bit security level," *IET Information Security*, vol. 14, no. 6, pp.633-640, 2020.
- [19] A. Kumar, P. Jain, "A lightweight encryption authentication scheme using rectangle and chaotic logistic map algorithm for smart grid," *International Journal of Advance Research, Ideas and Innovations in Technology*, vol. 7, no. 1, pp.126-130, 2021.
- [20] S. Y. Chen, Y. L. Liu, C. L. Lin, "Lightweight verifiable group authentication scheme for the Internet of things," *Acta Electronica Sinica*, vol. 50, no. 04, pp.990-1001, 2022.

Hybrid Optimization with Recurrent Neural Network-based Medical Image Processing for Predicting Interstitial Lung Disease

K.Sundaramoorthy¹, R.Anitha², Dr.S.Kayalvili³, Ayat Fawzy Ahmed Ghazala⁴, Prof. Ts. Dr. Yousef A.Baker El-Ebiary⁵, Sameh Al-Ashmawy⁶

Professor, Information Technology, Jerusalem College of Engineering, Velachery Road, Narayanapuram, Pallikaranai, Chennai-600100¹

Professor, Biomedical Engineering, Jerusalem College of Engineering, Velachery Road, Narayanapuram, Pallikaranai, Chennai-600100²

Professor, Department of Computer Science and Engineering, Velalar College of Engineering and Technology, Thindal, Erode - 638012. Tamil Nadu³

Assistant Professor, Department of Education and Psychology-College of Science and Arts in Al-Qurayyat, Jouf University - Saudi Arabia⁴

Assistant Professor Educational Technology - Faculty of Specific Education - Menoufia University – Egypt⁴

Faculty of Informatics and Computing, UniSZA University, Malaysia⁵

Imam AbdulRahman Bin Faisal University- Kingdom of Saudi Arabia⁶

Abstract—One of the dreadful diseases that shortens people's lives is lung disease. There are numerous potentially fatal consequences that can arise from interstitial lung disease, such as: Lung hypertension. This illness doesn't influence your overall blood pressure; instead, it only affects the arteries in your lungs. To prevent mortality, it is essential to accurately diagnose pulmonary illness in patients. Various classifiers, including SVM, RF, MLP, and others, are processed to identify lung disorders. Large datasets cannot be processed by these algorithms, which causes false lung disease identification. A combined new Spider Monkey and Lion algorithm is suggested as a solution to get around these limitations. Images of interstitial lung disease (ILD) were taken for the study from the publicly accessible MedGIFT database. The median filter is employed during the pre-processing step of ILD images to reduce noise and remove undesirable objects. The features are extracted using a hybrid spider Monkey and Lion algorithm. The lungs' damaged and unaffected regions are divided into categories using recurrent neural networks. Several metrics such as accuracy, precision, recall, and f1-score are used to evaluate the performance of the proposed system. The results demonstrate that this technique offers more precision, accuracy, and a higher rate of lung illness detection by processing a large number of computerized tomography representations quickly. When compared to other strategies already in use, the proposed model's accuracy is greater at 99.8%. This method could be beneficial for staging the severity of interstitial lung illness, prognosticating, and forecasting treatment outcomes and survival, determining risk control, and allocation of resources.

Keywords—*Interstitial lung disease; spider monkey lion optimization; recurrent neural network; medical image processing; diagnosis and identification; classification*

I. INTRODUCTION

In the U.S, lung disease is the most common cause of mortality. Lung disease affects the lives of over 400,000

Americans each year, and risk of death from this disease is rising even while they fall for other serious diseases like cancer. ILDs are a diverse group of more than 200 lung conditions that mostly affect the lung parenchyma but can also appear as respiratory or vascular symptoms. Numerous lung conditions, including interstitial lung disease, chronic obstructive pulmonary disease, and acute respiratory distress syndrome, are linked to the mechanical performance of the lung. ILD is a group of devastating disorders where fibrosis damages and stiffens lung tissue [1].

A class of extensive parenchymal lung ailments known as "interstitial lung diseases" are highly diseased and lethal. The establishment of a new categorization of idiopathic interstitial respiratory infections, which divides the condition into three categories—major, uncommon, and unclassified [2]. The updated version is unusual because it allows for the treatment of threatening matters in accordance with the disease classification. Foremost fatal interstitial lung disease, idiopathic pulmonary fibrosis exhibits a great degree of symptomatic variability. The therapist managing the impacted patient faces a substantial problem when dealing with ILD in an apparently healthy host that seems to be ordinary. A comprehensive group of specialists should examine a mix of diagnostic, radiology, and abnormal criteria that are used to diagnose an ILD [3]. The pathophysiology of these disorders has been linked to a number of variables, including genetic component, diseases, medications, thermal energy, and workplace and environmental durations.

Idiopathic pulmonary fibrosis, sarcoidosis, hypersensitivity pneumonitis, ILD as a symptom of connective tissue disease, drug induced ILD, and pneumoconiosis are the most common types of ILD. The most prevalent and serious type of ILD, IPF, has drawn the considerable attention in respiratory research. Interstitial fibrosis, also known as the common

technique of UIP, is visible on HRCT of the lung in IPF patients. Compared to other respiratory problems, ILD have a stronger link to health hazards. The pneumoconiosis is caused by asbestosis, silicosis, and coal worker's pneumoconiosis is typical instances of industrial disorders [4].

Infection and fibrosis of the lung tissue characterise the category of lung disorders known as ILD and pulmonary fibrosis. IPF is a specific type of ILD, and because of its unidentified origin, poor overall outcome, and moderate reaction to clinical treatment, IPF is frequently regarded as one of the most prevalent and significant ILDs. Although the treatment for ILD and pulmonary fibrosis can occasionally be complicated, it is based on the ILD's most likely origin. Additionally, the outcomes of the medical evaluation, auto-immune serologic analysis, chest CT imaging, and, if necessary, a lung sample must be added to the health information in order to make a precise ILD classification [5]. PFTs are crucial for care and typically reveal a sequence of a restrictive ventilator disorder with an aberrant continues to prove, but they mostly serve as a gauge for the extent of the condition and its prognosis by identifying a particular type of ILD [6].

A record-breaking volume of data has been generated as a result of the enormous progress in image capture technology, capacity growth, and the installation of bio-medical information collection equipment [7]. These data come from numerous (sometimes inconsistent) database systems, have a high dimension (Computed tomography, Magnetic resonance imaging, etc.), and are rich in variables. It provides healthcare information that is difficult, particularly on images [8]. DL assists in creating new ones. In addition, DL not only aids in disease detection but also evaluates the predicted goal and gives proactive forecast models to help doctors create efficient treatment protocols. Every scientific subject, especially medical image into regions, is involved in DL. The usage of a DNN model is implied by the phrase DL [9]. Various samples are inputted into the NN, and combines with parameters to produce digital output through non-linear processes. The classification of ILD is given in Fig. 2.

One of the crucial components of digital image processing used for medical planning in healthcare systems is medical imaging. Additionally, a significant amount of medical data is employed in medical field for scientific and educational purposes, such as clinical data visualisation. For quicker distribution, the clinical images are reduced and kept hidden. Due to their smaller system memory, compressed images have an extremely short transmission rate[10]. Typically, there are two types of compression: both lossy and lossless methods. With file format, input images are offered without any losses, and the output image is identical to the original image. However, compression ratio offers a lower compression ratio [11]. FFNN, which has loops in the hidden units, is often enhanced by recurrent neural network (RNN). It is capable of learning combinations. All stages and neurotransmitters share the same weights. The framework provides the model to take sample segments as input and determines the temporal link between the data. The representation of dynamical modification at a time series is handled using this Classifier [12].

The following criteria contribute identify interstitial lung disease:

- The research suggests the most appropriate deep-learning approach for identifying and classifying interstitial lung disease (ILD) from medical images.
- The dataset was gathered from patient medical images who had ILDs. The first scenario is pre-processing of medical images, the second is feature extraction and selection, and the third is categorization.
- The findings show that the proposed hybrid Spider Monkey and Lion algorithm has effectively selected the features and the deep learning-based recurrent neural network (RNN) is capable of recognizing the ILD classification.

The remaining part of this article has been organized as follows: Section II provides a detailed description of the most recent techniques for classifying interstitial lung disease. The dataset obtained from individuals with interstitial lung disease is described in Section III. Section IV covers the system structure and assessment method, which includes image pre-processing and the steps taken to develop the extraction of features for categorization. The findings and discussion sections are described in Section V accordingly. Lastly, Section VI presents the relevant outcomes of the study.

II. RELATED WORKS

In this paper [13], a convolutional neural network for classifying ILD sequences is evaluated. The network is composed of three dense layers, an averaged pooling with a size equivalent to the length of the final classification mappings, and five CNN layers with 22 kernels and LeakyReLU activations. Seven outcomes in the final deep network correspond to the classes taken into account: good health, grounded glassy opacity, micro nodules, standardisation, water supply systems, deformation, and a mixture of Aspects. This work used a sample of 14696 input images, obtained over 120 Computed tomography clinics, to train and test the CNN. This is the first deep Network created specifically for the issue. The efficiency of the utilised CNN in comparison to earlier techniques on a difficult collection was demonstrated by a similar evaluation. The classification results of Convolutional in evaluating respiratory signals (around 85.5 percentages) showed its potential. However, the work cannot involve CNN by expanding 3-dimensional data from Computed tomography volume scans and incorporating the utilised approach into a Computer aided system which helps radiologists by providing a differential diagnosis for Lung lesions. This study uses a limited dataset to evaluate the model, making it unsuitable for use in real-world conditions.

In order to extract relevant images, content-based image retrieval systems thoroughly analyse underlying image LL properties. This eliminates the need for speech recognition tags, textual descriptions, or phrases to be connected with the images. For resemblance identification and classification for a particular query image, a CBIR system preserves great image visualisations in the form of extracted features. The initial identification and categorization of lung disorders depending

on lung X-ray images are made possible by the CBIR system for the recovery of medical images suggested in the paper [14]. Balancing across several measurement approaches, extensive experiments on the benchmark dataset showed that the strategy improved precision by 49.7 %. Additionally, the area under the precision-recall curve values for all subclasses showed a 26.6 % enhancement. Due to its numerous levels and lengthy training period, this strategy is ineffective.

An intriguing development is the use of AI in the identification of chronic obstructive pulmonary disorders. Discovering commonalities in forensic test data allows AI systems to anticipate health outcomes or identify obstructionist traits. However, prospects for AI in the identification of acute respiratory disorders and to summarise recent trends has to be evaluated. The paper [11] provide an explanation for the use of AI in telehealth, breath research, lung sound assessment, and other management of cancer. By offering precise differentiated diagnosis, ML has yielded promising result in computerized pulmonary function interpretation. Modern DL models for obstructive classification tasks in Computed tomography include Convolutional neural network. Due of the extensive time and resource requirements, this approach is ineffective.

The paper [15] used a novel hybrid DL framework called VDSNet for the detection of lung illnesses in X-ray images. A database of National institutes of health chest X-ray images obtained from the Kaggle source is used for the assessment. The overall recognition accuracy for the full dataset is 73percentage points for VDSNet, whereas the maximum accuracy for natural grey, vanilla Color images, hybrid Neural VGG, basic CapsNet, and customized CapsNet are 68%, 69%, 69.4%, 61%, and 64%, respectively. VDSNet has a validation accuracy value of 73%, which is higher than the reference dataset's accuracy score of 71%. However, instead of 19 seconds for the measurement, VDSNet takes a time for training of 431 seconds for the whole data source. However, computerized chest X-ray diagnostic techniques is not enhanced in this work, so this paper focuses on the application of image data enhancement techniques such as colour space implants, kernel filtering, deep feature improvements, etc. The VDSNet approach can be used to analyse X-ray images of potential patients and healthy controls for the individuals who have pneumonia associated with COVID-19. When focusing on the large-scale dataset, this study research meets certain difficulties. Smaller datasets can therefore produce good accuracy, but they are ineffective for use in real-world scenarios.

An automatically generated method for the recognition and classification of ILD patterns is provided in the paper [16]. It is done by removing designs for the modest inter-class information variability and large intra-class value fluctuation by the use of translational and luminance effects, respectively. The Template-Matching Combined Sparse Coding approach is

a brand-new and effective dimensionality reduction technique that differs substantially from defined areas of interest inside lung parenchyma that are resistant to translational and illuminating changes. Using a framework comparison procedure, the converted image patches is matched to all potential models of the image. By minimising the objective function of the classification model between the translation enhanced image and the reference, the equivalent sparse matrices for the set of translation texture features and their corresponding framework are produced. For handling high-intra class characteristic variability challenges, a SVM is created, which improves classification performance. In this work, zone of interests of 5 lung tissue patterns - healthy, emphysematous, surface - are found and used. These designs were chosen from an internally multimedia record that includes high-resolution positron emission tomography image series. The method works more effectively than the majority of cutting-edge multi-class classification methods. Whenever the data set includes greater amounts of noise, it fails to function as well.

In the paper [17], authors classified lung HRCT extracted features into five classes: good material, emphysema, field glass, fibrosis, and micronodules using an enhanced DenseNet algorithm with small kernel DenseNet. The SK-design DenseNet's consists of two compact blocks and two different levels. Each thick block has six groups in it. The SK-DenseNet model is more efficient for extracting high level and minor pathologic information for ILD categorization utilising a convolution kernel in accordance with the properties of Computed tomography features of respiratory disease. According to experimental findings, the SK-DenseNet performed best than other Convolutional networks like DenseNet, AlexNet, VGGNet, and ResNet. However, iterations of the research will concentrate on images with ILD patter annotations. This method is not effective since the Indirect effects of hMIKO-1 on T and B cells and other cells of inflammation are also possible. Table I depicts the comparison of existing approaches' merit and demerits.

In an animal model of bleomycin-induced ILD, the paper sought to determine how the hybrid proteins known as human MIKO-1 affected murine macrophages activity and whether it had any protective effects. In order to do this, the phenotype of hMIKO-1-co-cultured thioglycolate-induced murine intraperitoneal lymphocyte was investigated. Since day 0 to day 14, mice were divided into regular and given different doses. Since day 28, the mice were put to death, and the organs are examined for collagen fibres, and for levels of gene transcription. In vitro, hMIKO-1 prevented murine macrophage from polarising toward an M2 preponderance [18]. In comparison to the BLM-alone category, the histologic grade of the lung pathology and the lung extracellular matrix components level were drastically decreased in the BLM + hMIKO-1 collective. This strategy is inappropriate for delivering a greater accuracy level.

TABLE I. COMPARISON OF EXISTING APPROACH MERIT AND DEMERITS

References	Method	Implementation	Advantage	Disadvantage
[13]	CNN	Python language	With no data loss, it lowers the considerable dimensionality of images.	This study uses a limited dataset to evaluate the model, making it unsuitable for use in real-world conditions
[14]	CNN	Python and MATLAB	It requires less work from humans to create its functions.	Due to its numerous levels and lengthy training period, this strategy is ineffective.
[11]	AI	PFT interpretation software	It can considerably boost accuracy while reducing errors.	Due to the extensive time and resource requirements, this approach is ineffective.
[15]	VDSNet	Tensorflow, Jupyter Notebook, and Keras	Much less time is needed for training.	When focusing on the large-scale dataset, this study research meets certain difficulties. Smaller datasets can therefore produce good accuracy, but they are ineffective for use in real-world scenarios.
[16]	B-MCSVM	-	It uses relatively little storage and performs well in high-dimensional areas.	Whenever the data set includes greater amounts of noise, it fails to function as well.
[17]	SK-DenseNet	GraphPad Prism version 7.0 software	It has more effectively utilized features	This method is not effective since the Indirect effects of hMIKO-1 on T and B cells and other cells of inflammation are also possible.
[18]	PSO and Fuzzy C-mean Clustering method	Python coding	In terms of processing efficiency, it is better in terms of speed and memory needs.	This strategy is inappropriate for delivering a greater accuracy level.

III. DATASET

Among the main obstacles to the development of DL in medical image analysis is the absence of enough training data, which are necessary to establish the accuracy of DL classifiers. The production of huge medical imaging datasets is difficult due to the annotation calls for a lot of input from medical professionals. In particular, several expert perspectives are needed to address the issue of human error. The collection includes medical information from patients with pathologically confirmed classifications of ILDs as well as high-resolution computerized tomography image series with 3-dimensional identified areas of diseased lung tissue. The goal of this effort is to make up for the dearth of publicly accessible datasets of ILD cases that may be used as a foundation for the creation and assessment of image-based computerised diagnostic tools. The present research makes use of the freely accessible MedGIFT database [19]. The database contains 108 HRCT images with annotations. It has 17 distinct ILD designs, each measuring 512×512 . A complete set of 1946 ROIs was offered from 108 HRCT imaging series. Five of the most common healthy and ILD patterns are taken into consideration in the present study. Since the database is multipattern, there is a chance that many patterns will coexist in a single slice.

IV. METHODOLOGY

The feature extraction phase of image recognition is crucial. A hybrid spider monkey and lion method is used for feature extraction, and RNN is used for classification. It is used to categorise the diseased portion of the lung for future prevention and find the fastest way for the infected region recognition utilising the distinctive hunting behaviour of lion and monkey. Fig. 1 represents the process of ILD prediction.

The basic process flow diagram for the steps of ILD prediction is shown in Fig. 1. The lung input image is first pre-

processed, and then features are retrieved using a combination of lion and spider monkey optimization. RNN was finally utilized to categorise the results.

A. Pre-processing

Pre-processing comes after the images has been chosen as the starting point for ILD diagnosis. The median filter has been used to improve the Lung images by reducing noise and removing unwanted items. A non-linear, well-organized simple image processing method called the median filter is frequently employed to lower noise in images. The result of the Median filter is given by Eq. (1).

$$\hat{g}(x, y) = \text{median}_{(a,b) \in T_{xy}} \{f(a, b)\} \quad (1)$$

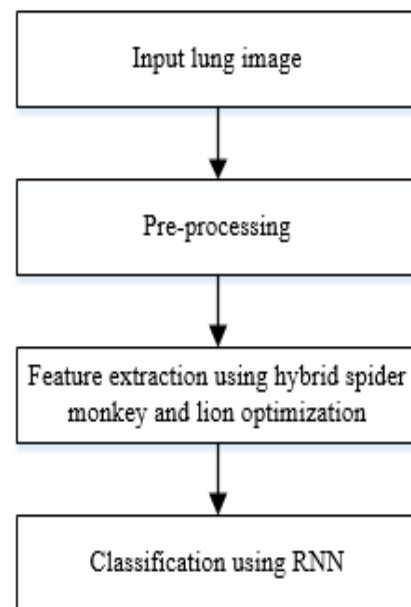


Fig. 1. Process of ILD prediction.

For the goal of detecting arrhythmias, segmentation is the most crucial topic in informatics and the health industry. The main aim of this study is to create and build a reliable classification method using a RNN classifier that is SM and lion-based. Data pre-processing, image segmentation, extraction of features, and classification are the 4 stages of the heartbeat detection algorithm. The standard will be eliminated from the ECG signal when it first enters the information processing phase. The pre-processed sensor will go through the classification step, which will execute the segmentation using the Frequency response. The segmented data move on to the feature extraction phase, where features including wave features and temporal features will be recovered. Pre-processing the information is important since it helps to remove the signal's undesirable components, ensuring that the characteristics that are recovered are correct and suitable for blood pressure and heart rate identification. The retrieved characteristics will be sent to the classification stage, where the Spider Monkey and lion -based Recurrent neural network will be used to classify the disease (RNN classifier).

The hybrid Spider Monkey lion algorithm, which combines the Spider Monkey Optimization (SMO) and the lion algorithms, will be used in conjunction with the RNN in the suggested method. The Lion algorithm is a generalisation of interaction and the activities of lion whereas the SMO leverages the foraging behaviour of spider monkeys in a fission-fusion like framework to address the optimization model. Additionally, combining SMO with lion effectively captures the hunting and intellectual behaviour of the lion. The goal of the optimization ensures that the RNN is effectively tuned for precise predicting and classifying. The suggested method will be put into practise using the MATLAB tool, and its performance will be assessed using metrics like accuracy and excellent put. As a result, when compared to the current strategies, such as DL approach, hierarchical categorization, and research driven technique, the suggested strategy successfully performs the ILD classification and will achieve superior accurateness and excellent output.

B. SMO

A recently introduced stochastic improved algorithm called Spider Monkey Optimization takes its cues from nature. A kind of monkey that is included in the group of creatures that have a fusion and fission social structure is referred to as "spider money." These spider monkeys are frequently observed in groups and have skilled hunting habits. They facilitate food gathering in a variety of ways by sharing pertinent information with the other group members [20]. The advanced food-seeking strategy used by spider monkeys served as motivation for the invention of this SMO method. Fitness function is given in Eq. (2).

$$fitness\ function\ (F_i) = \begin{cases} \frac{1}{1+F_i} & \text{if } F_i \geq 0 \\ 1 + xyt(F_i), & \text{if } F_i < 0 \end{cases} \quad (2)$$

Six distinct aspects of the numerical method of SMO foraging behaviour for optimization techniques are covered in the subsequent subsections. SMO first creates a population of N spider monkeys at random. Each S_{ij} in SMO has the following initialization which is shown in Eq. (3).

$$S_{ij} = S_{minj} + X(0,1) \times (S_{maxj} - S_{minj}) \quad (3)$$

Where S_{minj} and S_{maxj} are lower and upper bounds in jth direction for S_i and $X(0,1)$ denotes a random number in the range [0, 1]. The next section describes all six phases of SMO in detail.

1) *Local Leader Phase (LLP)*: In this step, a current hire for a service user is obtained utilising the information provided by the group members via Eq. (4) and the local leader. The fitness value of the solution determines its excellence. The following iteration will focus on the option with the highest fitness (the new location is preferable to the existing position).

$$S_{newij} = S_{ij} + X(0,1) \times (S_{lpj} - S_{ij}) + X(-1,1) \times (S_{qj} - S_{ij}) \quad (4)$$

Where, S_{pj} and S_{qj} denote the positions local group leader and randomly chosen qth spider monkey respectively.

2) *Global Leader Phase (GLP)*: During GLP, each member updates their standpoint depending on knowledge from the firm's global leader and all members, as stated in Eq. (5).

$$S_{newij} = S_{ij} + X(0,1) \times (S_{Gj} - S_{ij}) + X(-1,1) \times (S_{qj} - S_{ij}) \quad (5)$$

Where, S_{Gj} shows the jth direction of the global leader. Furthermore, the fitness values is expressed in Eq. (6).

$$P = \frac{fitness_i}{\sum_{i=1}^K fitness_i} \quad (6)$$

Similar to LLP, further processing uses the better solution from the newly created position and the old position of the SM.

3) *Global Leader Learning (GLL) phase*: In this stage, the global leader takes the role with the consistent top fitness, and a global limitation monitor is used to track changes in the role of the global leader.

4) *Local Leader Learning (LLL) phase*: Local leader is given the necessarily lead with the best fitness. Similar to the GLL phase, the local limitation counter is incremented by one if the new position of the local leader is like the information and establishing.

5) *Local Leader Decision (LLD) phase*: When a local leader's local limit counter hits a certain count, all group members are reset using Eq. (7).

$$S_{newij} = S_{ij} + X(0,1) \times (S_{Lj} - S_{ij}) + X(-1,1) \times (S_{qj} - S_{ij}) \quad (7)$$

6) *Global Leader Decision (GLD) phase*: If the global leader location is not changed after a certain amount of iterations, it forms tiny size of subcategories. Each group's local leaders in GLD are chosen through the LLL procedure. If the global leader's position is not changed by a

predetermined threshold, all subgroups are combined into one group. SMO imitates the FFS organization in this manner.

C. Lion Optimization Algorithm

Analysing the behaviour of the lion, the computational solution for the Lion optimization method was developed. Hunting, running away to safety, straying, mating, migrating, defending, population stability, and convergence are examples of animal behaviours [21]. Such characteristics are expressed mathematically, and an efficient method is developed. The initial communities are produced using lions, an entirely arbitrary solution. The additional lions are randomly separated into P pride subdivisions, with S percentage of the prides being declared a female lion. Norm's lions are picked as the original population N. The optimum place for each lion is described as the best explanation from previous iterations and it is constantly altered throughout the optimization process. Each lion's ideal place can be found in the proud territory.

Young boys who reach sexual maturity are separated from their parents' dignity and go on to remain nomads, where they have less power than regular males. A nomad lion roams the searching area aimlessly in search of a better location (solution). A powerful nomad male forces the resident male out of the prides when they engage in combat. The regional male lion will take the spot of the nomads' male. Female desert lions occasionally join prides, while other nomadic female lions occasionally leave them. The lesser lion will be mercilessly killed for a number of reasons, including malnutrition and intensity of competition.

The inhabitants in the global optimum were first created spontaneously by the LOA. It observed each distinct solution as a "lion." A lion is described as follows for an Nv multidimensional optimization method which is shown in Eq. (8).

$$Lion = x_1 + x_2 + x_3 + \dots x_{Nv} \quad (8)$$

And the assessment of the minimization problem, as shown below, determines the cost (each lion's fitness value) which is shown in Eq. (9).

$$fitness\ value\ of\ lion = f(lion) = f(x_1 + x_2 + x_3 + \dots x_{Nv}) \quad (9)$$

The N pop alternatives were developed primarily at random in the problem space. A random selection of % N options was chosen as NL. P prides were created from the remnant population. Throughout the optimization process, a certain sexual identity was constant for each response in the LOA. In each pride, roughly 75–90 percent of total of the estimated workforce created in the preceding stage were referred to as female lions, with the remaining lions being male, to mimic this nature. Each lion identified its most frequented place while searching. Each pride built its region based on the regions that had been indicated. As a consequence, each pride's region was created by its supporters marking certain sites.

1) *Hunting*: A specific number of females search for food source in a team in each and every P to provide food for the

other P individuals. To surround and capture their food source, these hunting lions employ certain tactics. Typically, as lions hunt, they stick to a same routine. Every lioness changes her hunting position based on where she is and where the other lionesses are. As a result, some hunter lions rotate their prey and assault it from the other direction, and LOA employs opposition-based learning. In Eq. (10) hunter equation is expanded. Three "wings" of hunters are established. The centre wing has the best overall fitness, whereas the left and right wings are determined at random. The prey flees to a new area during a search as the predator gets fitter.

$$hunter = \begin{cases} rand((2 \times PY - hunter, PY), \\ (2 \times PY - hunter)) \\ rand((2 \times PY - Hunter, PY), \\ (2 \times PY - Hunter) < PY \end{cases} \quad (10)$$

D. Hybrid Spider Monkey and Lion Optimization

For particular jobs, it has been demonstrated that the SMO system integrates more quickly than alternative techniques. As a result, it can decelerate down as it approached the global ideal point. The probability of becoming trapped in the region of local optima rises as the algorithm continuously incorporates with the world wide optimal position. The SMO algorithm's dependency on search parameters seems to be another flaw. Based on the choices made, the accumulation rates may change. The SMO approach has undergone several modifications to increase its efficacy, with the main goal of balancing the incremental and radical features. Among the suggestions are tweaks to the developmental technique, adjustments to the parameters, adjustments to the upgraded criteria, and the introduction of better developing methods. It has been demonstrated that the lion is more efficient and has a better rate of success in multipurpose scenarios than the SMO, despite the fact that the lion's algorithms does not take acceleration into account. The flow diagram for the proposed method is given in Fig. 2.

$$fitness\ function\ (Fi) = \begin{cases} \frac{1}{1+f(x_1+x_2+x_3+\dots x_{Nv})} & \text{if } Fi \geq 0 \\ 1 + xyt(f(x_1 + x_2 + x_3 + \dots x_{Nv})), & \text{if } Fi < 0 \end{cases} \quad (11)$$

The process flow diagram for the hybrid spider monkey and lion optimization is shown in Fig. 2. When features are successfully extracted from ILD images, the fitness value is calculated.

E. Recurrent Neural Networks

The foundation of NN topologies is the idea that all signals are consecutively redundant. However, this presumption is erroneous and might be harmful for many applications, like time report and natural language processing where the relationships among successive training instances are crucial. A type of artificial neural network known as an RNN adds loops to the feed - forward neural network to increase its functionality. A recurrent hidden state, whose activation at each step depends on that of the preceding step, allows an RNN, unlike a feed - forward neural network, to process the complex combination. The system can display variable

temporal behaviour in this way. Fig. 3 depicts the basic structure of RNN.

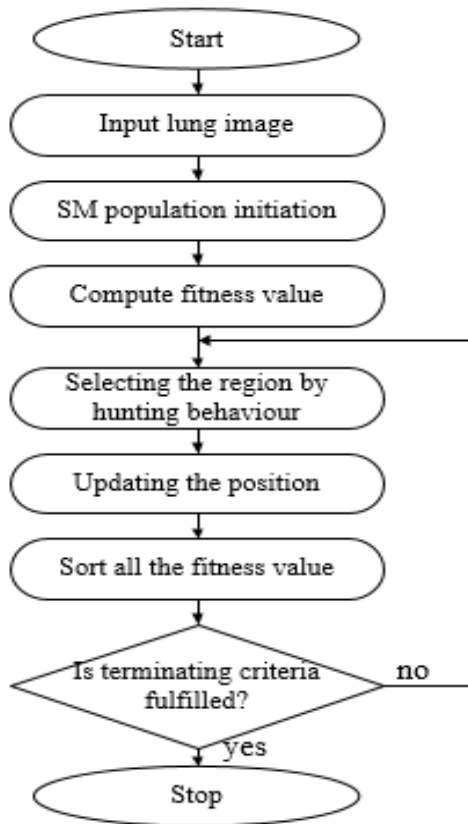


Fig. 2. Flow diagram for the proposed method.

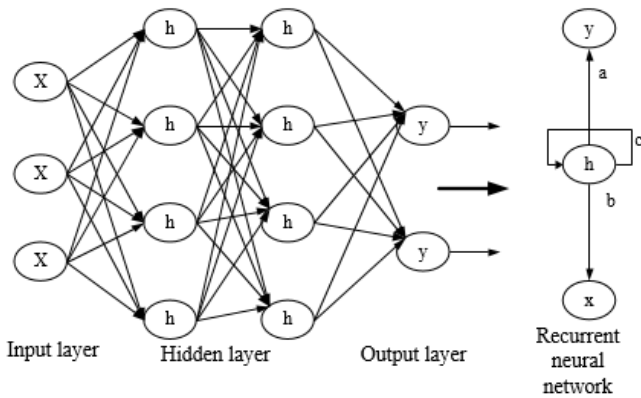


Fig. 3. Working process of RNN.

Fig. 3 depicts the fundamental layout of an RNN. Input, hidden, and output layers are all three layers. The ILD image results are successfully classified by using this approach.

V. RESULT AND DISCUSSION

Utilizing the images of the gathered plant leaves, the suggested strategy has been evaluated. To distinguish the diseased lung from the healthy lung, the suggested method employed hybrid spider monkey and lion optimization with recurrent neural network. The study assesses the 4 commonly

used evaluation metrics, including classification accuracy, precision, recall, and F1-score, which are described below.

With regard to the lung images that are obtainable, accuracy produces suitable results. In Eq. (12), accuracy is represented.

$$Accuracy = \frac{TN+TP}{FN+FP+TP+TN} \quad (12)$$

Precision thoroughly assesses a classifier's performance. Precision will be high if the lung image has low positives and low positives if the lung image has high positives. In Eq. (13) precision is represented.

$$precision = \frac{TP}{FP+TP} \quad (13)$$

Recall gauges how complete the classification is. More positive samples are found greater the recall. The Eq. (14) contains a representation of the recall formula.

$$Recall = \frac{TP}{FN+TP} \quad (14)$$

The F1-score measurement combines precision and recall. The F1-score measure is determined by calculating from the recognition rate. Eq. (15) represents the F1 measure.

$$F1 - score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (15)$$

From Table II, the accuracy of the proposed system is 99.8% which is high compared to other existing approaches such as ACO is 98.6%, CNN is 98.97% and SMB is 97%.

The existing and proposed analyses for accuracy measurements were compared in Table II and Fig. 4 represents the graph of accuracy measure, the suggested hybrid spider monkey and lion obtained more effective results.

The existing and proposed analyses for precision measurements were compared in Table II and Fig. 5 represents the graph of precision measure, the suggested hybrid spider monkey and lion obtained more effective results.

The existing and proposed analyses for recall measurements were compared in Table II and Fig. 6 represents the graph of recall measure, the suggested hybrid spider monkey and lion obtained more effective results.

The existing and proposed analyses for F1-score measurements were compared in Table II and Fig. 7 represents the graph of accuracy measure, the suggested hybrid spider monkey and lion obtained more effective results.

TABLE II. PERFORMANCE COMPARISON MATRIX

Method	Accuracy	Precision	Recall	F1-score
ACO	98.6 %	96%	96%	97.6%
CNN	98.97%	98.3%	98.7%	98.7%
SMB	97%	96%	97%	96%
Hybrid spider monkey and lion	99.8%	99.9%	99.7%	99.89%

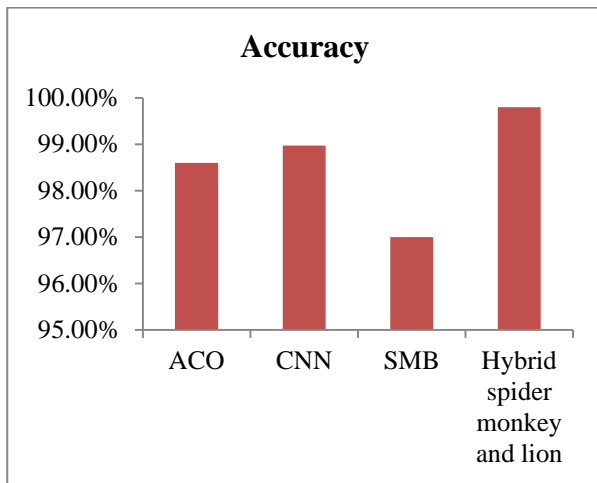


Fig. 4. Graph of accuracy measure.

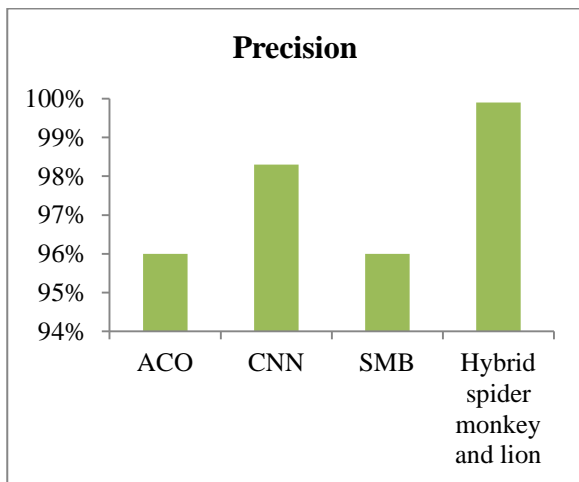


Fig. 5. Graph of precision measure.

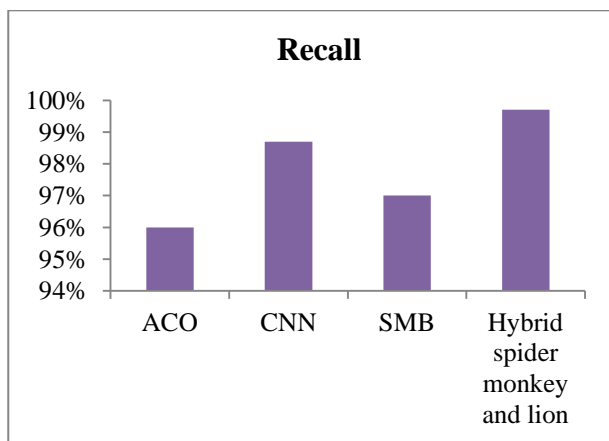


Fig. 6. Graph of recall measure.

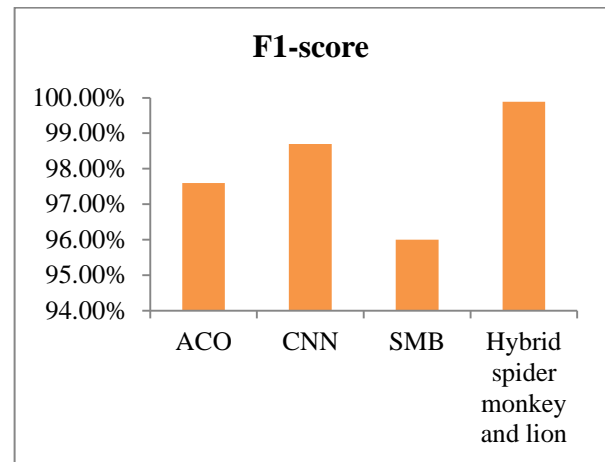


Fig. 7. Graph of F1-score measure.

In this instance, the hybrid spider monkey and lion with RNN model outperformed the ACO, CNN, and SMB models. The accuracy percentages for ACO, CNN, and SMB are 98.6%, 98.97%, and 97%, respectively. Precision, recall, and F1-score have better rates in the hybrid spider monkey and lion method compared to other models, and the F1-score has attained the greatest rate compared to other models due to the accuracy rate of 99.8% in the hybrid spider monkey and lion model.

VI. CONCLUSION

Deep learning (DL), which has significantly outperformed more conventional ML algorithms in recent decades, has taken centre stage in the mechanization of our daily lives. DL -based apps will replace humans in some roles, and autonomous robots will handle the majority of everyday chores. In contrast to other real-world issues, DL is just slowly making its way into the healthcare industry, particularly in the field of medical imaging. It has been suggested to use full HRCT images in a two-stage hybrid deep learning network screening method for interstitial lung disease (ILD). The accuracy, precision, recall, and F1 measure are determined by the suggested approach. The developed method yields a high prediction accuracy of 99.8%, and the proposed model is subsequently used for medical diagnostics. In this study, the IL disease prediction system is based on an improved spider monkey and lion optimization with RNN. The feature extraction was done with hybrid spider monkey and lion optimization, while the classification was done with RNN classifier. The proposed approach is utilized to distinguish between diseased and healthy Lung. With maximum accuracy of 98.8%, precision of 99.9%, recall of 99.7%, and F1-score of 99.89%, the suggested spider monkey and lion optimization with RNN is better than the existing approaches. Future research will build on this work by attempting to forecast the incidence of other serious illnesses including cancer and other heart-related diseases with high computational speed.

REFERENCES

- [1] J. Ker, L. Wang, J. Rao, and T. Lim, "Deep Learning Applications in Medical Image Analysis," *IEEE Access*, vol. 6, pp. 9375–9389, 2018, doi: 10.1109/ACCESS.2017.2788044.
- [2] B. Zhou, B. J. Bartholmai, S. Kalra, and X. Zhang, "Predicting lung mass density of patients with interstitial lung disease and healthy subjects using deep neural network and lung ultrasound surface wave elastography," *J. Mech. Behav. Biomed. Mater.*, vol. 104, p. 103682, Apr. 2020, doi: 10.1016/j.jmbbm.2020.103682.
- [3] G. Litjens et al., "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, Dec. 2017, doi: 10.1016/j.media.2017.07.005.
- [4] D. Bermejo-Pelamp, "Classification of Interstitial Lung Abnormality Patterns with an Ensemble of Deep Convolutional Neural Networks" *Scientific Reports*, 2020.
- [5] M. I. Razzak, S. Naz, and A. Zaib, "Deep Learning for Medical Image Processing: Overview, Challenges and the Future" 2018.
- [6] A. Nyma, M. Kang, Y.-K. Kwon, C.-H. Kim, and J.-M. Kim, "A Hybrid Technique for Medical Image Segmentation," *J. Biomed. Biotechnol.*, vol. 2012, pp. 1–7, 2012, doi: 10.1155/2012/830252.
- [7] M. L. Smith, "The histologic diagnosis of usual interstitial pneumonia of idiopathic pulmonary fibrosis. Where we are and where we need to go," *Mod. Pathol.*, 2022.
- [8] R. Borie et al., "The genetics of interstitial lung diseases," *Eur. Respir. Rev.*, vol. 28, no. 153, p. 190053, Sep. 2019, doi: 10.1183/16000617.0053-2019.
- [9] E. Bendstrup, J. Møller, S. Kronborg-White, T. S. Prior, and C. Hyldgaard, "Interstitial Lung Disease in Rheumatoid Arthritis Remains a Challenge for Clinicians," *J. Clin. Med.*, vol. 8, no. 12, p. 2038, Nov. 2019, doi: 10.3390/jcm8122038.
- [10] P. Hattikatti, "Texture based interstitial lung disease detection using convolutional neural network," in 2017 International Conference on Big Data, IoT and Data Science (BID), Pune, India: IEEE, Dec. 2017, pp. 18–22. doi: 10.1109/BID.2017.8336567.
- [11] N. Das, M. Topalovic, and W. Janssens, "Artificial intelligence in diagnosis of obstructive lung disease: current status and future potential," *Curr. Opin. Pulm. Med.*, vol. 24, no. 2, pp. 117–123, Mar. 2018, doi: 10.1097/MCP.0000000000000459.
- [12] S. P. Pawar and S. N. Talbar, "Two-Stage Hybrid Approach of Deep Learning Networks for Interstitial Lung Disease Classification," *BioMed Res. Int.*, vol. 2022, pp. 1–10, Feb. 2022, doi: 10.1155/2022/7340902.
- [13] S. Agrawal, A. Chowdhary, S. Agarwala, V. Mayya, and S. Kamath S., "Content-based medical image retrieval system for lung diseases using deep CNNs," *Int. J. Inf. Technol.*, vol. 14, no. 7, pp. 3619–3627, Dec. 2022, doi: 10.1007/s41870-022-01007-7.
- [14] M. Anthimopoulos, A. Christe, and S. Mougiakakou, "Lung Pattern Classification for Interstitial Lung Diseases Using a Deep Convolutional Neural Network," 2015.
- [15] S. Bharati, P. Podder, and M. R. H. Mondal, "Hybrid deep learning for detecting lung diseases from X-ray images," *Inform. Med. Unlocked*, vol. 20, p. 100391, 2020, doi: 10.1016/j.imu.2020.100391.
- [16] C. Helen Sulochana and S. Praylin Selva Blessy, "Interstitial lung disease detection using template matching combined sparse coding and blended multi class support vector machine," *Proc. Inst. Mech. Eng. [H]*, vol. 236, no. 10, pp. 1492–1501, 2022.
- [17] T. Kotani, M. Ikemoto, S. Matsuda, R. Masutani, and T. Takeuchi, "Human MIKO-1, a Hybrid Protein That Regulates Macrophage Function, Suppresses Lung Fibrosis in a Mouse Model of Bleomycin-Induced Interstitial Lung Disease," *Int J Mol Sci*, 2022.
- [18] P. Kavitha and S. Prabakaran, "A Novel Hybrid Segmentation Method with Particle Swarm Optimization and Fuzzy C-Mean Based On Partitioning the Image for Detecting Lung Cancer," 2019.
- [19] A. Depeursinge, A. Vargas, A. Platon, A. Geissbuhler, P.-A. Poletti, and H. Müller, "Building a reference multimedia database for interstitial lung diseases," *Comput. Med. Imaging Graph.*, vol. 36, no. 3, pp. 227–238, 2012, doi: <https://doi.org/10.1016/j.compmedimag.2011.07.003>.
- [20] H. Sharma, G. Hazrati, and J. C. Bansal, "Spider Monkey Optimization Algorithm" 2019.
- [21] M. Selvi, "Lion optimization algorithm (LOA)-based reliable emergency message broadcasting system in VANET" 2020.

Study on Tomato Disease Classification based on Leaf Image Recognition based on Deep Learning Technology

Ji Zheng*

Guilin University of Electronic Technology
School of Optoelectronic Engineering
541004, China

Minjie Du²

Guilin University of Electronic Technology
School of Electronic Engineering and Automation
541004, China

Abstract—The utilization of computer vision technology is of the utmost significance in the examination of plant diseases. Research utilizing image processing to investigate plant diseases necessitates the analysis of discernible patterns on plants. Recently, numerous image processing and pattern classification techniques have been employed in the construction of a digital vision system capable of recognizing and categorizing the visual manifestations of plant diseases. Given the abundance of algorithms formulated for the purpose of plant leaf image classification for the detection of plant diseases, it is imperative to assess the accuracy of each algorithm, as well as its potential to identify diverse disease types. The main objective of this study is to explore accurate deep learning architectures that are more effective in deploying and detecting tomato diseases, thus eliminating human error when identifying tomato diseases through visual observation. and get more widespread use. An initial model was constructed from the ground up using a convolutional neural network (CNN), which was trained with 22930 tomato leaf images, and then compared to VGG16, Mobile Net, and Inceptionv3 architectures through a fine-tuning process. The basic CNN model achieved a training accuracy of 90%, whereas the training accuracies of VGG16, Mobile Net, and Inceptionv3 were respectively observed to be 89%, 91%, and 87%. The VGG16 model has a greater computational complexity than other approaches due to its considerable quantity of predefined parameters. Despite to be simpler, MobileNet proved to be the most efficient in terms of accuracy and thus is the most suitable for this research, due to its lightweight structure, fast functioning and adaptability for mobile devices. In contrast to other architectures, the suggested CNN architecture exhibits shallower characteristics, facilitating faster training on the same dataset. This research will provide a solid foundation for future scholars to easily improve the categorization of plant diseases, which is to develop algorithms that are lighter, faster, easier to run, and have higher accuracy.

Keywords—Deep learning; convolutional neural network; image recognition; plant diseases

I. INTRODUCTION

Agriculture continues to play a significant role in the economy, contributing considerably to foreign exchange earnings and gross domestic product. Tomatoes are a highly sought-after vegetable commodity worldwide due to their rich nutrient content, including organic acids, vitamins, essential amino acids, and natural fiber. It is cultivated both outdoors

and indoors. Globally, the annual output of fresh tomatoes is about 160 million tons, which is three times that of potatoes and six times that of rice. In recent years, China's tomato production has steadily increased in terms of cultivated area. According to statistics, China's tomato cultivation area will increase by 1.6% to 1.104 million hectares in 2020. China's tomato cultivation area will increase by 0.7% to 1.113 million hectares in 2021. China is the world's largest producer of tomatoes, with perennial production accounting for as much as one-third of the global total. According to statistics, in 2020 China produced approximately 65.15 million tons of tomatoes, an increase of 3.63 percent. China produces approximately 66.09 million tons of tomatoes in 2021, an increase of 1.44 percent. Nonetheless, numerous maladies on tomato plants cause substantial harm to the quality and yield of tomatoes.[1] Pathogens including bacteria, fungi, and viruses are responsible for producing disease in plants and can be transmitted through soil, water, and air.[2] The use of pesticides and climate change, which can alter the stage and rate of pathogen development, complicate the management of diseases.[3] Common tomato plant maladies include leaf blight, early blight, late blight, target spot, Septoria leaf spot, and yellow leaf curl virus.[4-5] These diseases can affect the leaves, stems, fruits, and roots of the entire tomato plant. If diseases are not controlled in proper time, they will cause significant losses and injury to the tomato plant. Every year, producers incur significant losses. Protecting tomato plants from these maladies is therefore essential for increasing yields. Traditionally, plant diseases are identified through visual observation with the assistance of trained personnel [6-7].

Numerous advancements have been made in computer vision technology in recent years, and there are more examples of applications in identifying and classifying the diseases of fruits and vegetable. Traditional computer vision techniques still require professional knowledge of these agricultural hazards. The development of artificial intelligence technology eliminates the requirement for professionals to conduct constant field observation. [8-9]. It also serves to minimize the possibility of human observational errors. Recent case studies demonstrate that soft computing models like neural networks [10], decision trees [11], support vector machines [12], and Nave Bayes [13] have been applied to the automatic classification of plant diseases. In recent years, deep learning has dominated computer vision. The proliferation of

smartphones, the integration of high-performance processors into cameras and mobile devices, and the developments of computer vision technology, particularly with the application of deep learning strategies, have facilitated the automation of diseases diagnosis utilising image recognition. [14]. Deep neural networks are a combination of deep learning and neural networks that replicate the brain's general operating principles. A deep neural network typically comprises an input layer, an output layer, and multiple hidden layers in between. A neural network with a single hidden layer is considered shallow in comparison.

In this paper, the input layer is trained to discover the optimal filter weight values [15]. The initial layer of a CNN model that takes in pictures as input is referred to as the inputted layer. Image height, width, and depth are fed into the input layer of the neural network. The depth indicates how many color channels the image belongs to. If the picture is an RGB picture, for example, it comprises three color components: red, green, and blue, thus the size of the picture is 3. However, if it is grayscale, its size is 1.

In between the input and output layers are the hidden layers. They will be used to identify particular characteristics. The undisclosed layer is used in this research to detect color and texture features that can be used for the categorization of tomato plant maladies. The output layer comprises category labels and has full connection. If we wish to classify an image into ten categories, we must designate ten labels in the fully connected layer. Convolutional neural networks have been revolutionary in several pattern recognition-related disciplines over the past decade [16]. Various deep learning architectures, such as VGG16, Inceptionv3, Mobile Net, Alex Net, Dense Net, and Google Net, have been proposed and all feature slight differences in their respective hidden layers. [17]. Despite all of these learning architectures, the classification of plant maladies remains imperfect.

Currently, CNN has been proven to be a very good choice for achieving plant disease detection in terms of accuracy [18]. Each structure has its limitations. An illustration of this is that certain architectures can be more intricate because a greater number of parameters require more time for processing, yet aid in generating a high degree of precision. The primary objective of this research is to ascertain the most effective classification technique of plant diseases with a view to enhancing this domain further. This is not only in terms of accuracy, but also in terms of speed and difficulty of deploying and adjusting.

However, many plant disease detection methods proposed are not to replace the current detection methods, but to supplement these methods [19]. Despite the promising findings of multiple research studies, the practical implementation of

these technologies has yet to be adequately explored. There has been a proliferation of computer-based applications that have demonstrated great efficacy, yet the efficacy of these architectures has not been properly appraised. The primary objective of this research is to analyse the efficacy of modern deep learning systems in regard to plant disease classification with the intent of attaining the highest level of accuracy. This research seeks to provide a definitive answer, rather than a hypothesis, that can be used to assist novice researchers in selecting a current deep learning algorithm that is most advantageous for their initiatives in the field of automated classification of plant diseases.

II. MATERIALS AND METHODS

A. Methodology Overview

The methodology implemented in the research is shown in Fig. 1. The process of constructing a model can be divided into five distinct stages: data collection, in which images are acquired and divided into the three primary folders of training, verification and testing; pre-processing, to guarantee that relevant traits are present; the selection of an appropriate architecture; the training and verification stage, in which the model is trained and verified to generate the final model; and the prediction stage. This stage furnishes the classifier with characteristic test images to assess its practical efficacy. The ensuing text contains an elucidation of each phase.

B. Data Collection

Ten different categories of 256x256 resolution images of tomato leaves affected by various diseases were procured from the Plant Village Library located on the Kaggle platform. The dataset is segregated into three distinct subsections, namely training, validation, and testing, which contain 17184, 4585, and 1161 images respectively. The utility of testing and verifying using two distinct datasets is to confirm the model's capacity to process new data. The test data is utilised to assess the efficacy of the ultimate model, whereas the verification data is employed to assess the performance of the model during the training phase. Table I illustrates the total amount of tomato leaf images that have been subjected to training, testing, and validation, and classified accordingly. As hardware, the ACER Aspire3 computer uses Intel core TM i5-8265U 1.66Hz, 8GB DDR4 memory, and NVIDIA GeForce MX230 for this research.

In this paper, the method of classifying diseased tomato leaf images focuses on automatically identifying the marks appearing as spots on the leaves, as shown in Fig. 2. Each picture represents a tomato leaf image that belongs to a certain category.

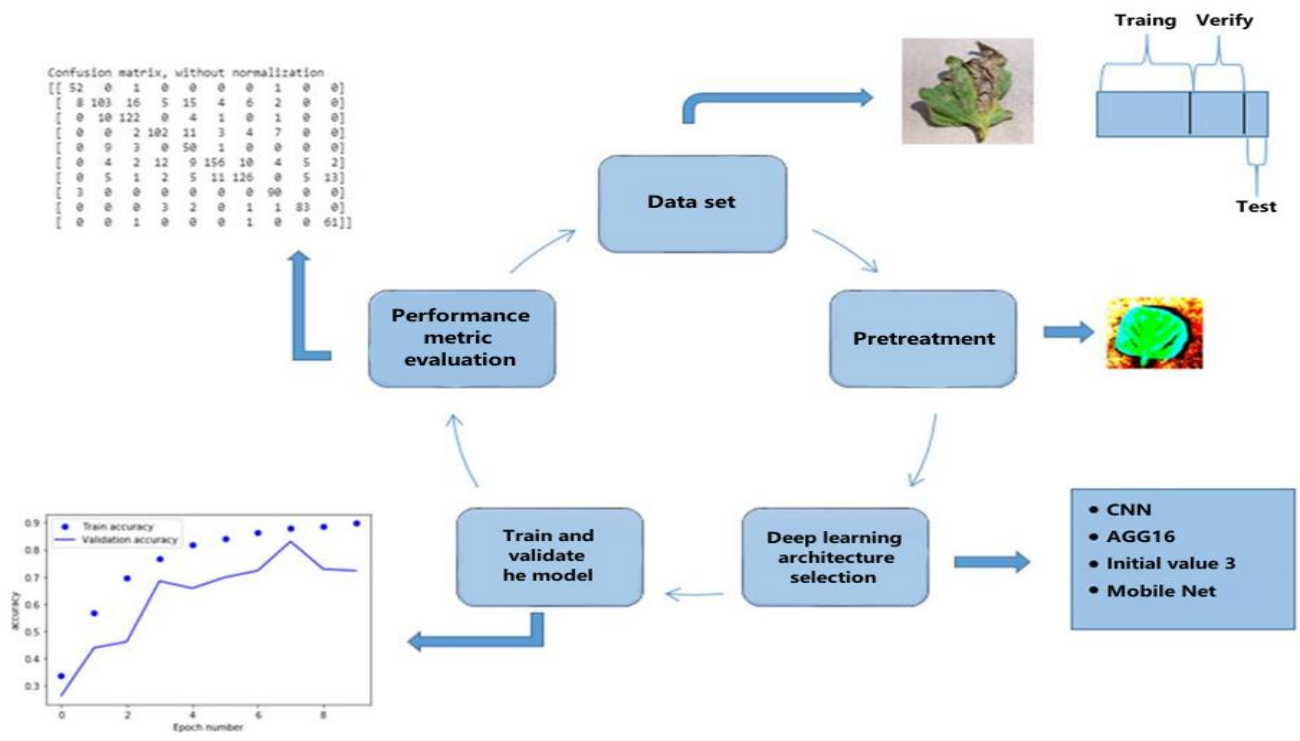


Fig. 1. Steps followed in the research method.

TABLE I. TOMATO LEAF IMAGES OBTAINED DURING THE STUDY

Disease type	Training data set	Validation data set	Test data set
Fine plaque	1684	425	54
Early blight	1761	480	159
Late blight	1713	463	138
Leaf mold	1753	470	129
Septicemic plaque	1682	436	63
Tetranychus	1537	435	204
Target spot	1659	457	168
Mosaic virus	1700	448	90
Yellow leaf curl virus	1868	490	93
Healthy leaves	1863	481	63
Total	17184	4585	1161

C. Image Preprocessing

Image preprocessing is a technique employed to improve the input image for subsequent manipulation, allowing for the pertinent characteristics to be accentuated for further

processing [20]. The utilisation of it aids in the alleviation of numerous issues, such as irradiance effects, illumination, and complications resultant from inadequate contrast. Fig. 3(a) and 3(b) demonstrate the difference in a diseased tomato image before and after pretreatment using the Keas input function, respectively. In the preprocessing stage, data standardization and data enhancement technologies are applied.

Data normalization is an essential step in guaranteeing that each pixel in a picture has a comparable dissemination of information. It facilitates rapid convergence of the network during the training process. To normalize the data for this study, the mean was subtracted from each pixel, and the output was divided by the standard deviation.

Following this step, each pixel was assigned a value between 0 and 1. Image enhancement is essential to the development of an efficient image classifier. To construct an accurate classifier, a large number of unique images is required. It is practically impossible to locate so much information. By employing data augmentation techniques, it is feasible to produce fresh data by transforming the current data. In this study, rotation, cropping, scaling, and horizontal and vertical rotating are used as enhancement options.

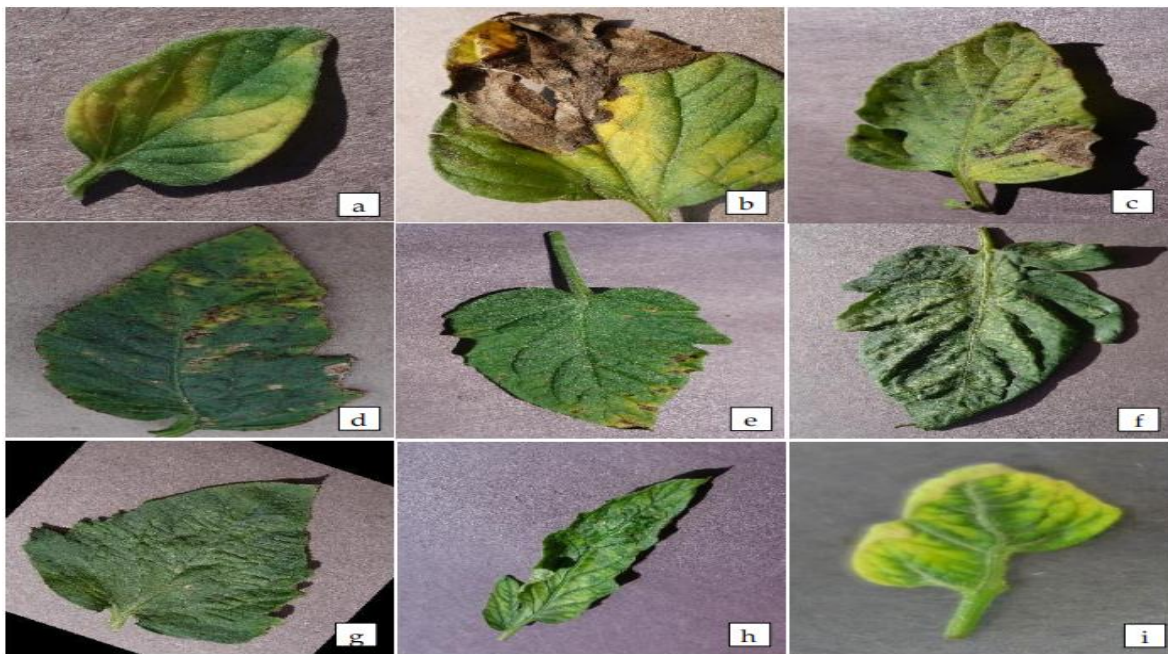


Fig. 2. Image of tomato leaves infected with the following virus, (Note: (a) fine plaque, (b) early blight, (c) late blight, (d) leaf mold, (e) septicemic leaf spot, (f) spider mite, (g) target spot, (h) Mosaic virus, (i) yellow leaf curl virus).

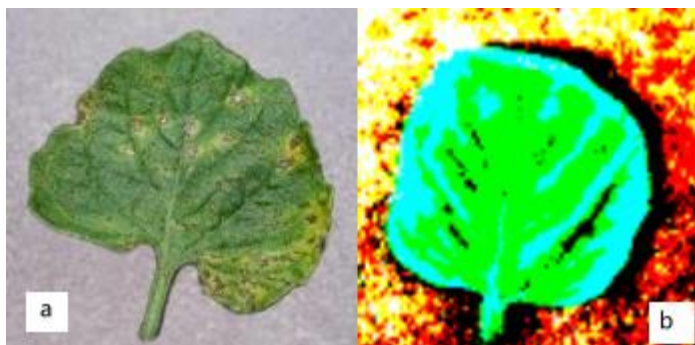


Fig. 3. Images of pretreated tomato leaves (a) before and (b) after.

D. Model Selection

Big data's availability and computational capacity are responsible for the success of deep neural networks. Therefore, it is essential to determine which designs are most effective and under what conditions they can be implemented.

1) *Convolutional neural networks (CNN)*: Convolutional neural network (CNN) is a deep learning model specifically designed to identify and analyse features from multidimensional data sets [21-22]. The selected CNN model consists of four convolutional layers, with subsequent implementation of batch normalization, maximal pooling, and discard layers. Moreover, there are two additional layers, which are dense and flat. Table II presents an overview of the network architecture, comprising of layer configurations and the corresponding number of parameters for training. Convolutional layers have the capability to detect distinct patterns of features, including edges and hues, among other characteristics. If one layer recognizes border patterns, the second layer will recognize line patterns. Each possesses an

abstraction level. Batch normalization resembles the preparatory normalization utilized in this study. The only distinction is that this paper employs a different type of layering. This paper makes the image more appropriate for the input layer by normalizing it during the preprocessing phase. In addition, batch normalization is used to make the image more suited for the hidden layer. Normalization of the activation layer's output is performed by subtracting the mean value and dividing by the standard deviation of the batch, resulting in improved stability and speed of the model. To prevent overfitting, the disintegrating and pooling layers are added [14]. In this model, the maximum pooling layer is utilized to reduce the overall size and thereby dynamically boost the computational performance. Each output is the corresponding data after applying maximum pooling to the inputs. Importantly, the maximum pooling contains only hyperparameters, so gradient descent does not necessitate learning. The dropout layer, as indicated by its name, drops out some neurons during training, but keeps them active during validation. As the model's complexity grows, it tends to overfit by memorizing all the weights, but the dropout layer solves this problem by randomly dropping out a portion of each layer during training.

The use of the Flatten Layer between the Merging Layer and the Dense Layer is generally recommended, as it facilitates the transformation of the Merging Feature Map into a single vector, subsequently allowing its passage to the Dense Layer. For each block, the Rectified Linear Unit (ReLU) activation function is employed, however, finally, after the final dense layer, the SoftMax activation function is implemented. The activation function checks which nodes meet the defined conditions in order to pass through the layer. A fixed learning

rate of 0.001 is used in the whole training process. The total quantity of epochs is designated to be ten, and the batch magnitude is established to be twenty-seven. The most advantageous amalgamation of parameter values is ascertained by modifying the super parameter. Once the model has been constructed, it must be compiled. The Adam optimization algorithm and the cross-entropy loss function are employed respectively in the compilation of the model as the optimizer and loss function. The model has been finally trained and stored in a .h5 file format.

TABLE II. OUTLINES THE PROPOSED CNN ARCHITECTURE MODEL

Layer	Output shape	Parameter
Conv2d	(None,254,254,32)	896
Max_pooling2d	(None,127,127,32)	0
Conv2d_1	(None,125,125,64)	18496
Max_pooling2d_1	(None,62,62,64)	0
Conv2d_2	(None,60,60,64)	36928
Max_pooling2d_2	(None,30,30,64)	0
dropout	(None,30,30,64)	0
Conv2d_3	(None,28,28,128)	73856
Max_pooling2d_3	(None,14,14,128)	0
Flatten	(None,25088)	0
Dense	(None,128)	3211392
Activation	(None,128)	0
Dropout_1	(None,128)	0
Dense_1	(None,10)	1290
Activation_1	(None,10)	0

2) *VGG16*: VGG16 is a famous CNN model, which was put forward by VGG of Oxford University. Compared with Alex Net [23], an improvement of VGG16 is to replace the larger convolution kernel (11x11, 5x5) in Alex Net with several successive 3x3 convolution kernels [24]. In this work, the VGG16 model which had been pre-trained was adjusted for the purpose of plant disease identification. In this study, Transfer Learning and Fine-Tuning are employed to classify diseased leaves, leveraging the fact that the VGG16 model has already learnt the characteristics of plant leaves. Utilizing the preprocessing input feature of Keras VGG16, preprocessing is applied before the image is supplied to the input layer. The VGG16 model contains a total of 138,357,544 parameters, which can be obtained from an online source. Given that the model has assimilated these characteristics, it is not indispensable to alter the weights throughout the entire learning procedure. Consequently, the weights are kept constant to impede modification. The number of parameters in the model has been reduced to 134,301,514 due to the freezing of weights. The ultimate model is in accordance with the Adam optimization algorithm and the cross-entropy loss function for classification. For this model, the learning rate is set to 0.001, the batch size is set to 27 and the number of epochs is set to 10. The model has been successfully trained and the weights have been stored in the .h5 format.

3) *MobileNet*: Mobile Net is an example of a lightweight and high-speed family of deep neural networks that can be deployed for categorization and pattern identification tasks, analogous to other convolutional neural networks. In spite of its diminutive size when compared to other models, it demonstrates remarkable performance. [25] The model's magnitude is primarily determined by the cumulative number of parameters. This research employed a model that was sourced from the web with the use of the Keras Library and pre-processed with Mobile Net's capabilities, which are commonly found to require fewer data augmentation procedures as compared to other models. After the process of refinement, the final model was composed of 3,239,114 parameters. The sample size was ascertained as 27, the number of epochs was established at 10, and the learning rate was maintained at 0.001. The model which had been trained was subsequently stored in .h5 format in order to be used at a later date.

4) *Inceptionv3*: Szegedy et al. discovered the initial concepts belonging to the family of deep neural networks [26] Several initial models were introduced, and each was designated initial VN, where N represents the version number.[27] In this investigation, Inceptionv3, the third version in the family, was utilized. A model with 21,802,784 parameters was obtained by utilising the Keras library. In addition to the initial architecture, a layer with global average pooling, a discarding layer, two layers with dense connections, and a densely connected output layer with ten nodes were included. The utilisation of a global average pooling layer facilitates the identification of features, while simultaneously reducing superfluous parameters. The number of epochs, learning rate, and batch size were set to ten (10), 0.001, and twenty-seven (27) respectively. The weights of the bottom 172 layers were kept constant while the higher layers were re-trained. Following the fine-tuning process, the total number of parameters in the final model was determined to be 23,387,434, which is slightly higher than the number of parameters present in the original model. The model is ultimately trained and stored in a .h5 format for potential later use.

E. Model Training and Validation

The fit function is used by inputting the train image, validating the image, setting the era size, and the steps of the era. Finally, the performance index is evaluated, and these steps are repeatedly performed by changing the parameter values using undiscovered data until the best combination of parameters is found to obtain a generalized accurate model.

III. RESULTS AND DISCUSSION

The objective of this research is to assess the effectiveness of different deep learning architectures commonly employed today, including CNN, VGG16, Mobile Net, and Inceptionv3. These models were constructed, optimized, and trained successfully. Table III presents the validation and training accuracy of each model, as well as the time taken for each model to complete training during each epoch.

TABLE III. MODEL ACCURACY

Model	Training situation	Verification situation	Mean value
	Accuracy	Accuracy	The time of each period
CNN	0.9018	0.8968	3652s
VGG16	0.8923	0.7894	6341s
Mobile Net	0.9112	0.9078	2949s
Inceptionv3	0.8734	0.8628	6068s

A confusion matrix was created for each of the deep learning models, as shown in Fig. 6, 7(a), and 7(b). The matrices represent the summary of predictions made by each model. The X-axis displays the predicted labels while the Y-axis displays the true labels. The dark labels in the matrix indicate the number of correct classifications made by the model. The test data was used to create the matrix, and it shows the prediction accuracy of each category. As is exemplified in Fig. 4, the bottom row of the confusion matrix indicates that all specimens with yellow leaf curl disease had been correctly categorized, consequently yielding a 100% predictive accuracy. The performance of the predictions for each disease category can be evaluated using these matrices. Despite being limited by space, the confusion matrices for Inceptionv3 and other CNN models can be viewed in Fig. 5(a) and 5(b).

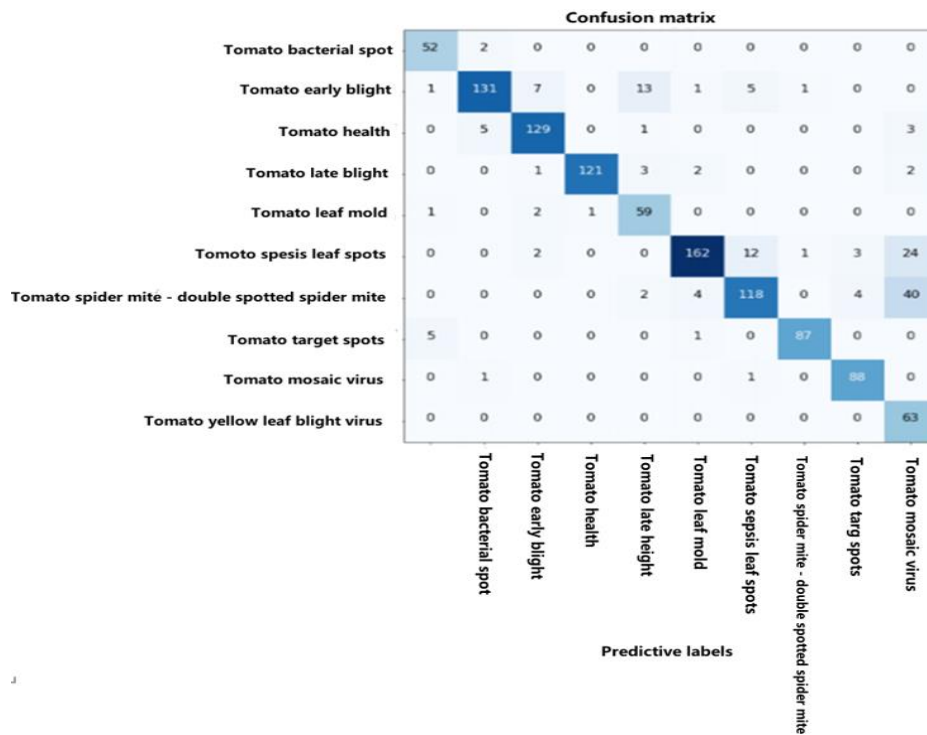


Fig. 4. Confusion matrix of MobileNet.

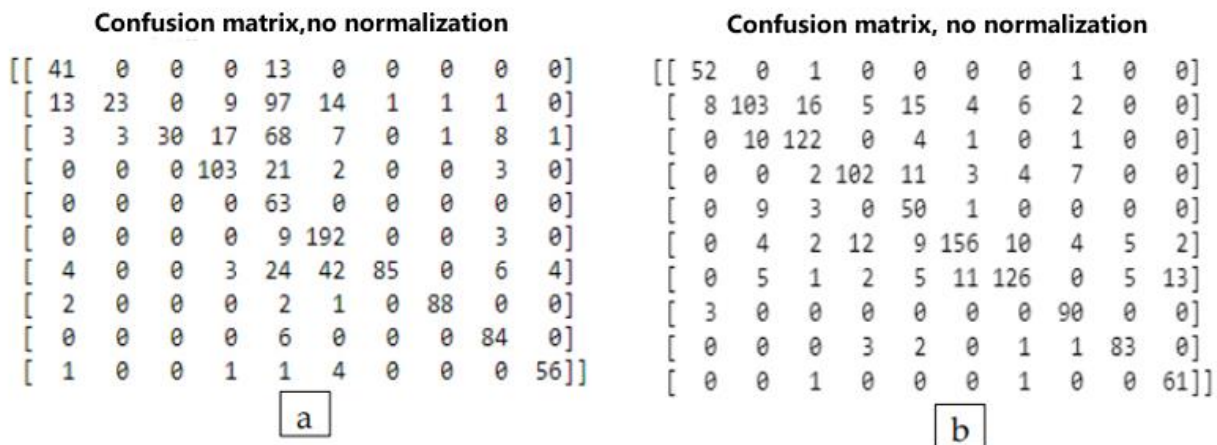


Fig. 5. Confusion matrix of (a) Inceptionv3 and (b) CNN.

The Matplotlib drawing library in Python was utilized for the creation of Fig. 6 and 7. Fig. 6 and 7 illustrate the

progression of training and validation accuracy/loss of each model over time.

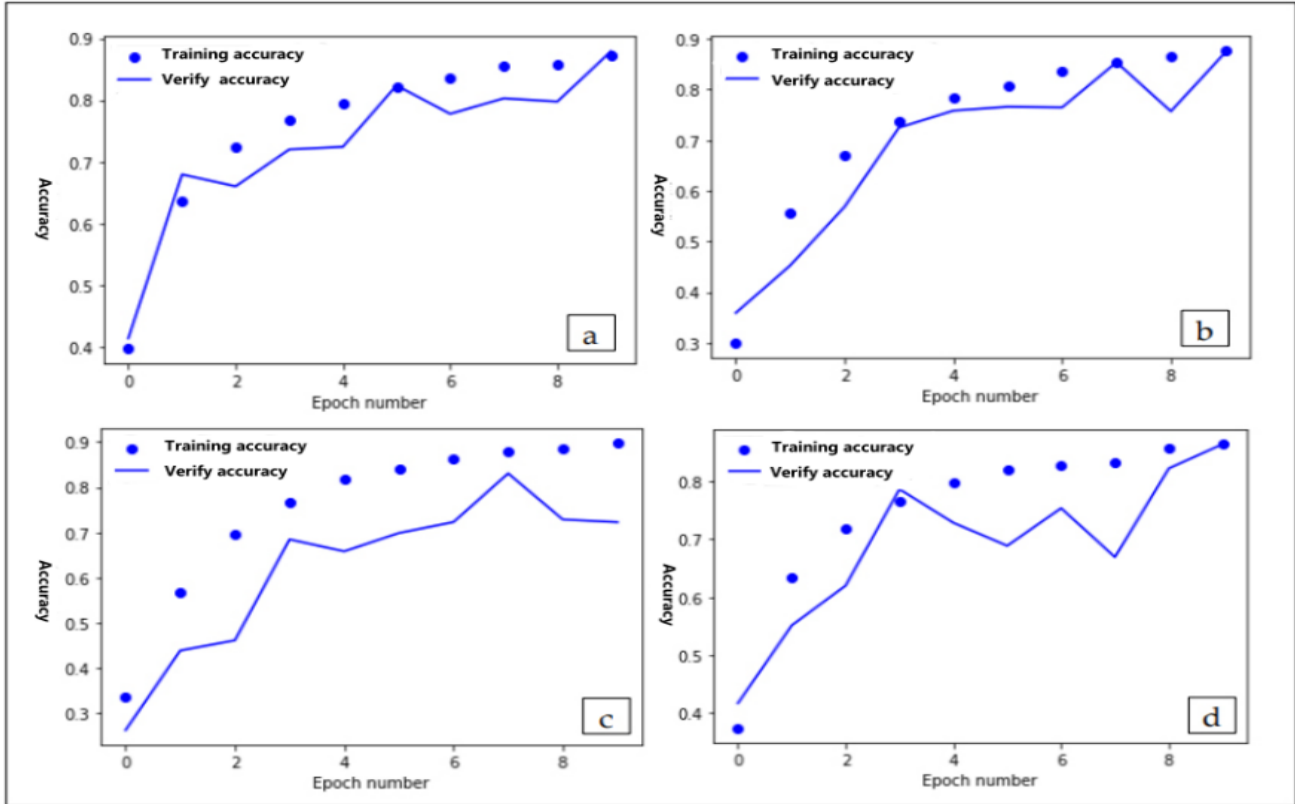


Fig. 6. Training and verifying the accuracy of the model: (a) CNN, (b) MobileNet, (c) Vgg16 and (d) Inceptionv3.

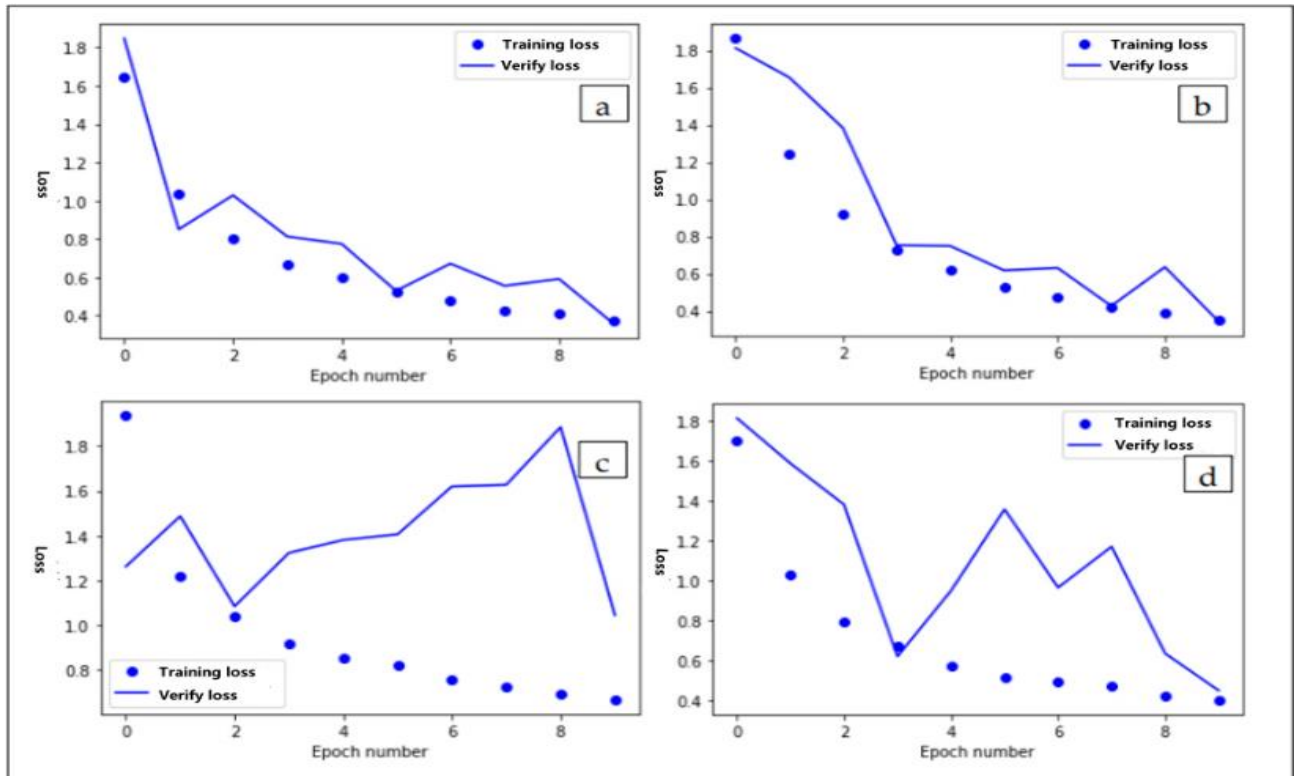
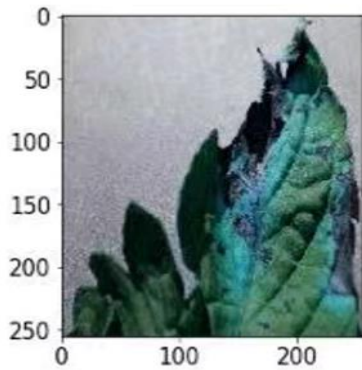


Fig. 7. Loss of training and validation models: (a) CNN, (b) MobileNet, (c) VGG16 and (d) Inceptionv3.

Referring to the diagrams, we can observe the fluctuations in verification accuracy and loss for each training epoch. To evaluate the performance of the final trained model, an unseen image is fed to the model, and the predicted category and confidence are obtained. This is achieved through the use of OpenCV, a Python library. The purpose of this is to verify if the final model can accurately predict the unseen image. In Fig. 8(a), an image of a leaf affected by early blight is shown as an example.

Source:Category:Tomato Early Blight, Document: Toamto Early Blight Forecast; Classification: Tomato early blight, confidence:0.999915



Source: Classification:Tomato-tomato mosaic virus, file:Tomato-tomato mosaic virus 8b 8 90deg.JPG; Classification:Tomato mosaic disease virus,confidence:1.000000

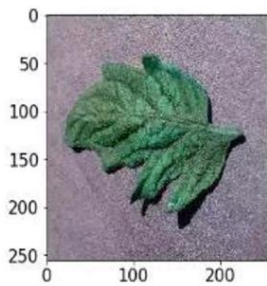


Fig. 8. Prediction of the MobileNet model for (a) Mosaic virus and (b) early blight.

It was correctly classified into the correct category with a confidence level of one hundred percent. Fig. 8(b) depicts a leaf infected by mosaic virus. In addition, it was confidently classified into the correct category.

A summary of correctly and incorrectly classified images is presented in Table IV. The initial column of the table denotes the various disease categories, while the second column illustrates the sum of images presented to the Mobile Net model from each category. The third column signifies the sum of images accurately classified, while the fourth column indicates the aggregate of images misclassified.

Table IV presents a synthesis of the correct and erroneous classification of images for each disease type, coupled with the overall number of images supplied to the ultimate Mobile Net model from each category. The last column of the table gives the average accuracy of each category. Research has demonstrated that the yellow leaf curl virus has been accurately categorized with a mean accuracy of 100%. Table IV can be utilized to likewise evaluate other diseases. The deep learning

model has demonstrated remarkable enhancement in the categorization of plant diseases. This study seeks to contrast the efficacy of a Convolutional Neural Network (CNN) model developed from the ground up with that of a fine-tuning pre-training model using the Plant Village dataset from Kaggle. Training the VGG16 model was time-intensive due to the large parameter count. It can be concluded that an increase in the number of parameters also increases the time required for training the model.

TABLE IV. SUMMARY OF CNN MODEL PREDICTION

Species	Total	Correct value	Error value	Accuracy (%)
	Image	Classification value	Classification value	
Fine plaque	54	52	2	96
Early blight	159	131	28	82
Healthy	138	129	9	93
Late blight	129	121	8	94
Leaf mold	64	60	4	94
Septicemic plaque	207	165	42	80
Spider mite	168	118	50	70
Target point	94	88	6	94
Mosaic virus	90	88	2	98
Yellow leaf curl virus	84	84	0	100

When examining the training and validation accuracy of VGG16, it is evident that the former significantly exceeds the latter, indicating the model's strong performance on training data but weak performance on test data. This phenomenon is known as overfitting, which frequently results from an overly complex model. With its relatively smaller number of parameters, Mobile Net is less complex than Inceptionv3 and VGG16, making it better suited for effective detection. The prediction accuracy of Mobile Net and CNN models is 0.90 and 0.89, respectively, indicating their superior suitability for test data compared to VGG16. Fig. 6, 7(a), and 7(b) show that the classification accuracy of Fig. 6 is 100% compared to other models for the last category of matrices (yellow leaf curl). Determining the layers to freeze and train for the fine-tuning of VGG16 and Inceptionv3 necessitates more time. These models are difficult to train due to their high complexity and sensitivity pattern identification. In contrast, the proposed CNN architecture is slightly shallow, making it quicker to train on the same dataset. Accuracy is dependent on parameter identification, which requires prior experience. Mobile Net outperforms Inceptionv3 in terms of size, delay, and accuracy, and could easily be implemented in mobile devices and embedded vision applications. Although the CNN model in this study is less deep than other models, it is still easily applicable due to its reduced computational complexity.

IV. CONCLUSIONS

This study suggests that a basic convolutional model with a minimum of four layers can be enough for extracting the relevant features for categorising tomato plant diseases in comparison to other tuned models. VGG16 and Inceptionv3 have relatively more parameters, but do not reflect better recognition accuracy. This requires more time to refine, train and run on the one hand, and limits its deployment and fast recognition on smaller systems on the other. The suitability of Mobile Net and CNN for this study is due to its compactness, speed, and ability to run smoothly on mobile devices. With guaranteed high accuracy recognition, MobileNet and simple CNN structures are not only easier to deploy, adjust and work with in the areas covered by this study, their lighter weight structure allows them to have further development potential.

The aim of this investigation is to deliver a conclusive answer instead of a supposition, which will support upcoming scholars in opting for the most innovative and efficient deep learning algorithms and networks for their further research in the realm of automatic plant disease classification. Currently only tomato leaf images obtained from the internet were examined in this paper, and they will be further tested on real-time images detection program in the future, employing the CNN model developed throughout the research process.

V. FUNDING STATEMENT

Project support: This project is funded by the national Guilin University of Electronic Science and Technology Student Innovation and Entrepreneurship Training Program (No.202110595052).

REFERENCES

- [1] P. Das & P. Misra. Artificial intelligence based automated detection of plant diseases: A review. Archives of Computational Methods in Engineering, 2019, 26(6), 1463-1486.
- [2] K. Kumar, N. Belwal, and A. Singh. "A Review on Detection and Diagnosis of Plant Leaf Diseases using Image Processing Techniques", International Journal of Advanced Research in Computer Science and Software Engineering, vol. 7, no. 10, 2017, pp. 63-68.
- [3] M. Pautasso. Emerging infectious diseases in crops: the plant health challenge. Annals of Applied Biology, 2013, 163(1), 1-8.
- [4] R. Michael Davis, Michael D. Littrell, Tomato Disease Management: A Practical Guide"
- [5] R. M. Davis, R. N. Raid, Integrated Management of Tomato Diseases"
- [6] S. S. Chouhan, (2018). Feature extraction and classification using Deep convolutional Neural Networks. Journal of Cyber Security and Mobility, 2018, 8(2), 261-276.
- [7] Retrieved 2nd December 2020, from http://www.tomatonews.com/en/background_47.html Chouhan, S. S.
- [8] A. K. Rangarajan, R. Putushothaman, &A. Ramesh, Tomato crop disease classification using pre-trained deep learning algorithm. Procedia Computer Science, 2018, 133(133), 1040-1047.
- [9] A. Kaul, U. P. Singh, & S. Jain. Bacterial foraging optimization based radial basis function neural network (BRBFNN) for identification and classification of plant leaf diseases: An automatic approach towards plant pathology. IEEE Access, 2018, 6, 8852-8863.
- [10] A. F. Fuentes, S. Yoon, J. Lee, and D. S. Park. High-performance deep neural network-based tomato plant diseases and pests' diagnosis system with refinement filter bank. Frontiers in Plant Science, 2018, 9: 1162.
- [11] J. Singh, A. Kaur & A. Singh, An efficient technique for classification of plant leaf diseases using decision tree. International Journal of Engineering and Technology, 2018, 7(4.34), 107-110.
- [12] Y. Dandawate, and R. Kokare, "An automated approach for classification of plant diseases towards development of futuristic decision support system in Indian perspective." International Conference on Advances in Computing, Communications and Informatics (ICACCI 2015), 2015, 794-799.
- [13] K. Mohanapriya, M. Balasubramani, "Recognition of Unhealthy Plant Leaves Using Naive Bayes Classifier." IOP Conference Series: Materials Science and Engineering, 2019, 561(1).
- [14] S. N. Khan Meera, A. A. Shaikh, H. Ansari, N. Ansari, Disorder Detection in Tomato Plant Using Deep Learning. SSRN Electronic Journal, 2019, 2154-2160.
- [15] H. Durmus, E. O. Gunes, M. Kirci, Disease detection on the leaves of the tomato plants by using deep learning, 6 th International Conference on Agro-Geoinformatics, 2017.
- [16] S. Albawi, T. A. M. Mohammed, S. Alzawi, "A Data-Driven Approach To Precipitation Parameterizations Using Convolutional Encoder-Decoder Neural Networks Pablo." IEEE, 2017.
- [17] M. A. Hossain, & G. Muhammad. Comparative performance analysis of convolutional neural networks for plant disease identification. Computers and Electronics in Agriculture, 2020, 171, 105275.
- [18] H. S. Nagamani, & H. Sarojadevi. Tomato Leaf Disease Detection using Deep Learning Techniques, 2022, 13(1), 305-311.
- [19] S. P. Mohanty, D. P. Hughes, and M. Salathé, "Using Deep Learning for Image-Based Plant Disease Detection." Frontiers in Plant Science, 2016. 7:1419.
- [20] A. R. Patil, & R. S. Sonawane, Brain tumor classification using CNN with improved training dataset. Journal of Ambient Intelligence and Humanized Computing, 2021, 12, 4573-4583
- [21] Available at: <https://wiki.tum.de/display/lfdv/Layers+of+a+Convolutional+Neural+Network> Bodapati, J.D. and Veeranjaneyulu, N. (2019).
- [22] Y. LeCun, Y. Bengio & G. Hinton, Deep learning. Nature, 2015, 521(7553), 436-444.
- [23] A. Krizhevsky, I. Sutskever, & G. E. Hinton. ImageNet classification with deep convolutional neural networks. In Advances in neural information processing systems, 2012. 1097-1105.
- [24] K. Simonyan, & A. Zisserman. Very deep convolutional networks for large-scale image recognition, 2015.
- [25] A. G. Howard, M. Zhu, B. Chen, et al. "Mobilenets: Efficient convolutional neural networks for mobile vision applications," 2017, 1704.04861.
- [26] C. Szegedy, V. Vanhoucke, S. Ioffe, Rethinking the Inception architecture for computer vision. In Proceedings of the IEEE conference on computer vision and pattern recognition, 2016. 2818-2826.
- [27] J. P. Too, M. A. U. Khan, M. A. Rana, & A. M. Khattak, Impact of deep learning algorithms on plant disease identification: A review. Computers and Electronics in Agriculture, 2019, 162, 707-723.

Research on Recommendation Model of College English MOOC based on Hybrid Recommendation Algorithm

Yifang Ding*, Jingbo Hao

North China Institute of Aerospace Engineering
School of Foreign Languages
Langfang, 065000, China

Abstract—Establishing a reasonable and efficient compulsory education balance index system is very important to boost the all-around of compulsory education development, and then realize the course recommendation for students with different attributes. Based on this, the research aimed at the problems in college English education and evaluation, aimed to establish a college English MOOC education and evaluation system based on the improved neural network recommendation algorithm. The research first constructed the college English MOOC education and evaluation data elements, and then established a genetic algorithm improved neural network algorithm (BP Neural Network Optimization Algorithm Based on Genetic Algorithm, GA-BP), and finally analyzed the effect of the assembled model. These results show that the fitness of the GA-BP model reaches the set expectation when the evolutionary algebra reaches 10 times, and its fitness is 0.6. The corresponding threshold and weight are obtained, and the threshold and weight are substituted into the model. After repeated iterative training, the model finally reached an error of 10⁻³ when it was trained 12 times, and the expected accuracy was achieved. The R value of each set hovered around 0.97, and the fitting degree was high, which showed that the GA-BP model proposed in the study had a better fitting degree. The difference between the expected value and the output value is mainly distributed in the [-0.08083, 0.06481] interval. To sum up, the GA-BP model proposed in the study has an excellent effect on college English education and evaluation. This evaluation model has a faster learning rate and a higher prediction accuracy and more stable performance.

Keywords—Genetic algorithm; education quality assessment; BP neural network; college English MOOC

I. INTRODUCTION

With the advent of the Internet age, due to the continuous increase of users, the amount of digital information has increased rapidly, and it is difficult for traditional technical means to process it in a short period of time, resulting in information overload [1]. At present, information overload is an urgent problem to be solved. Traditional information retrieval methods cannot quickly and accurately find the most valuable learning resources from massive data [2]. The existing teaching methods cannot fully stimulate the subjects' interest in learning. The above characteristics of the recommendation system can quickly improve the user's retrieval speed for information, and at the same time ensure the validity of the information obtained by the user in the recommendation

system. It eliminates the distraction of users during use and improves the credibility of users [3]. In terms of learning and teaching, introducing the recommendation system into the learner's learning situation can make it easier for the subjects to obtain appropriate learning resources, so that the subjects can learn more focused and in-depth [4-5]. MOOC stands for Massive Open Online Course, also known as Massive Open Online Course. MOOC is a curriculum system with a large number of participants, no admission conditions, and an online carrier. Due to the high technical requirements of platforms that support large-scale registration for the operation of MOOC, it is difficult for a university to independently complete it. Therefore, in the MOOC process of implementing blended English teaching in universities, internationally mature MOOC platforms such as China University MOOC and Xuetang Online can be utilized, and teaching materials can be reasonably selected and rigorously designed based on actual teaching needs. Based on this, the study proposes a college English MOOC education and evaluation system based on the improved neural network recommendation algorithm, which can then realize the evaluation of MOOC courses and recommend corresponding courses to students with different attributes. The paper constructs a data mining algorithm and establishes a corresponding system. Based on this, it proposes an improved neural network model. The first part of the paper is an introduction, which introduces the background and motivation of the research; The second part is a literature review, summarizing the research of relevant scholars in different fields; The third part is the research method, which constructs the data mining algorithm in college English education and evaluation system, including the establishment of college English education and evaluation data elements, and the improved neural network algorithm based on genetic algorithm; The fourth part is the analysis of the application effect of improved neural network algorithms in college English MOOC education and evaluation; The fifth part is the conclusion, and the last part is the references.

II. RELATED WORK

Chinese and foreign scholars have conducted extensive discussions and researches on this. Features based on artificial intelligence algorithms. On this basis, in-depth research has been carried out, and it has been widely used in artificial intelligence. Through the simulation experiment of this model, it is proved that this method is correct [6]. Lunde et al in aim of

improving students' subject-related skills, it combines methods such as theoretical education on research topics, proximal validity and validity, and group discussions with social education students on Mental Development of SA-SH-Ext [7]. Sotomayor-Moriano et al. proposed a virtual laboratory environment for online practice using the standard process FDI system. Thanks to VLE, trainees can set the errors of sensors, according to the model of FDI, and perform performance tests on FDI systems. This paper uses a four-pot process (4 TP) FDI system as an example to demonstrate its application under various failure scenarios. At the end of the article, the practical experience of using these two patterns in FDI design is given [8]. Putra et al. collected by questionnaire survey. In the group test there are 9 respondents and in the larger group test there are 12 respondents. It was found that the textbook experts scored 4.06, the media professional scored 4.78, and the language professional scored 4.19, all of which are relatively feasible. The scores for the group test and the large group test are 3.36, 4.21, which shows that this set of materials is very interesting. This set of textbooks can be used as a reference for learning trigonometric functions [9]. Nursalam et al. The purpose of this paper is to conduct a study on the quantum education model established by SDN Salupompong, Level 4, Mamuju Regency, West Sulawesi. This thesis is based on narrative method and sequence method, and carried out 4 investigations and 2 experiments. The study found that in English language learning, students' learning performance was 3.7, while motivation was 3.53, and response to students was 3.44. Among the three effective scales, the effective scale is 3.55, which is a very effective teaching method [10].

Chaves GL et al. through four simulation experiments, the course allows students to explore the effects of glycolysis, TCA cycle, genetic modification on carbon redistribution, and other metabolic pathways. metabolism. there. Students responded positively to using OptFlux as an adaptive questionnaire [11]. Under the concept of "teaching factory", Mourtzis et al. proposed a hybrid model, which has been proven in practice. In the case study of the hybrid power laboratory, engineering students remotely guided the laboratory staff to make and assemble a custom remote-control vehicle. The advantages and limitations of existing methods are discussed, and their development trends are prospected[12]. A manual simulation experiment was conducted to test whether using durable, inexpensive yarn can help pre-med students understand fetal lateral folding. Changes in knowledge are evaluated before and after use. The survey showed that students found the activity to be rewarding and fun, and the model work was easy to carry out. The above research results show that 3D dynamic 3D modeling is an effective teaching method [13]. Sultana et al examines the role of quantitative-based teaching materials in primary school students. Experiments have proved that quantitative teaching methods are more effective than conventional teaching methods in teaching students to think critically. From the critical thinking ability gain test of the experimental class, it can be found that the N-gain value is 0.59. Then, add a new criterion to a test category. The results of the experiment showed that the score of critical thinking in the experimental group increased by 0.20 percentage points compared with the

control group. Then, include a gain of n in a lower standard [14]. C Gonzalez-Velasco et al. adopts the flipped classroom teaching method, conducts a comparative analysis of students at different education stages [15].

In order to improve the efficiency of cross-cultural teaching of college English, scholars such as Yin H have constructed an improved MOOC teaching platform based on cloud computing and artificial intelligence technology, and established relevant functional modules. The research results indicate that the model has better performance and can improve the efficiency of English cross-cultural teaching [16]. Yin H proposes a new MOOC remote learning resource recommendation method for college English, which addresses the issue of low efficiency in traditional college English remote education recommendation resources. The interest feedback model is applied to the recommendation system. The research results indicate that this method can improve the accuracy of recommendation results [17].

Therefore, in teaching evaluation, the method of using neural network technology for teaching evaluation is rare, and the application research in college English MOOC teaching and evaluation is even rarer. To this end, in the paper, the neural network model improved by genetic algorithm has been applied in English teaching and evaluation, thus improving the teaching system.

III. RESEARCH ON DATA MINING ALGORITHMS IN COLLEGE ENGLISH EDUCATION AND EVALUATION SYSTEM

A. Establishment of College English Education and Evaluation Data Elements

How to evaluate college English teaching should start from the following aspects and analyze the factors affecting English teaching [18].

As can be seen from Fig. 1, the research focuses on the factors affecting the teachers themselves and the factors affecting teaching evaluations from students. It is the influence of the teaching quality evaluation of universities established and selected by the research institute. Factors, and then label them. The order of labeling is from top to bottom. The ten influencing factors are used as the input variable data required by the model below. The specific labeling is shown in formula (1).

$$x = (X_1, X_2, \dots, X_n)^T \quad (n=10) \quad (1)$$

As shown in formula (1), x is the vector of influencing factors of English education and evaluation in colleges and universities X_i , and the value of its influencing factors. According to the relevant factors of English education and evaluation in colleges and universities shown in Fig. 1, experts can obtain the teaching quality level of each college, through the analysis of various influencing factors, the structure of the model of the result is obtained, and the initial value. On this basis, using the number of ANN nodes to find the number of hidden layer nodes. On this basis, combined with the relevant factors of college English teaching and evaluation, the network is trained until the expected training deviation is achieved.

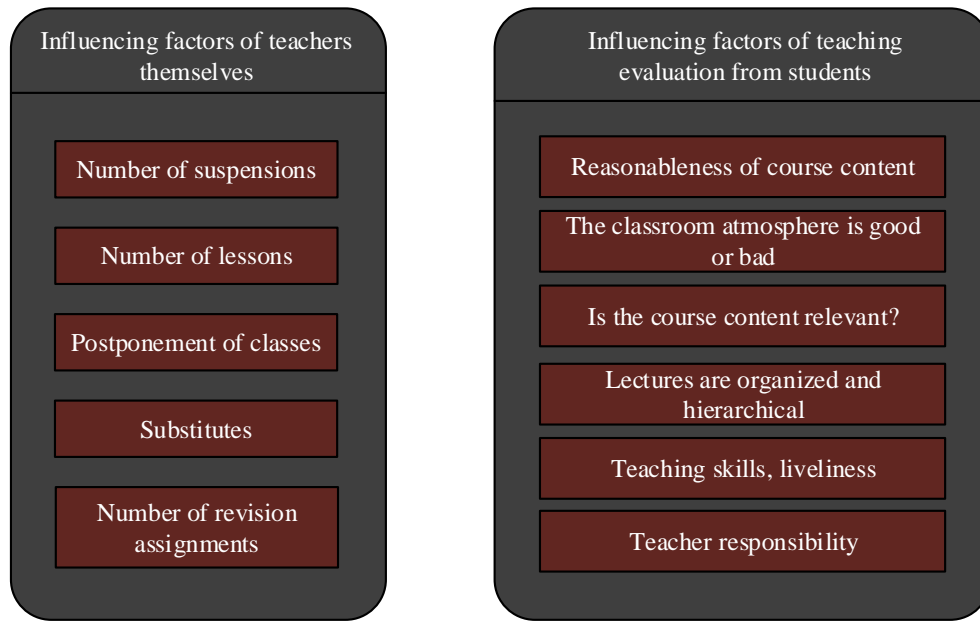


Fig. 1. The influencing factors of teaching quality evaluation in colleges and universities.

B. Improved Neural Network Algorithm based on Genetic Algorithm

The BP neural network has similar characteristics to human neurons, and can store a large amount of massive data Input and output training can be performed in the network without knowing the mapping relationship between variables. The basic idea of BP neural network is to reversely modify the forecast of the output layer, so that the deviation between the forecast result and the real data is closer to reality [19]. Its three-level network structure is shown in Fig. 2.

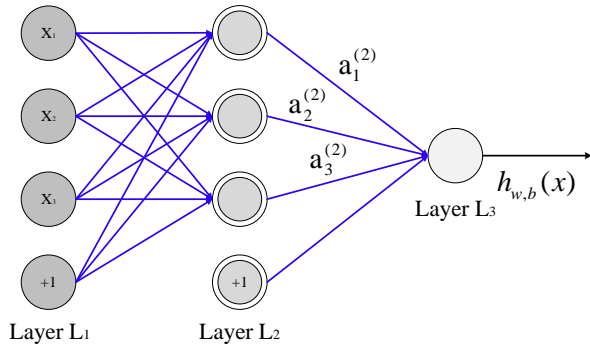


Fig. 2. Three-layer neural network topology.

As shown in Fig. 2, the BP neural network is divided into two stages: pre-activation and activation. Substituting it into the activation function [20], its mechanism and function are shown in formula (2) and formula (3).

$$g(x) = w_{ij}x_i + b_j \quad (2)$$

$$h(x) = f(g(x)) \quad (3)$$

In formula (2) and formula (3), x_i represents i the output value of the first neuron of the previous layer, w_{ij} represents the weight of the first neuron of the upper layer and the first j neuron of the previous layer, and represents the weight i of the first b_j neuron of this layer. j The critical threshold of the element, the relationship between its output x'_j and n input x_1, x_2, \dots, x_N can be expressed as formula (4).

$$X_j = \sum_{i=1}^N w_i \cdot x_i + S_j \quad (4)$$

In formula (4), S_j is the feedback signal. The BP neural network performs an iteration from the forward and backward directions, that is, starting from the input layer, performing continuous operations on the existing parameters, and then transmitting to the output layer along the forward direction of the network, and predicting it, and calculated its actual loss. The reverse transmission refers to the deviation between the previous forecast and the actual loss. The BP neural network performs an iteration from the forward and backward directions, that is, starting from the input layer, performing continuous operations on the existing parameters, and then transmitting to the output layer along the forward direction of the network, and predicting it, and calculated its actual loss. Backpropagation refers to the loss error between the predicted value and the real value obtained last time. Starting from the output layer, Using the gradient descending method, the parameters of each layer are reversely corrected until the deviation between the target [21]. The specific operation steps are shown in Fig. 3.

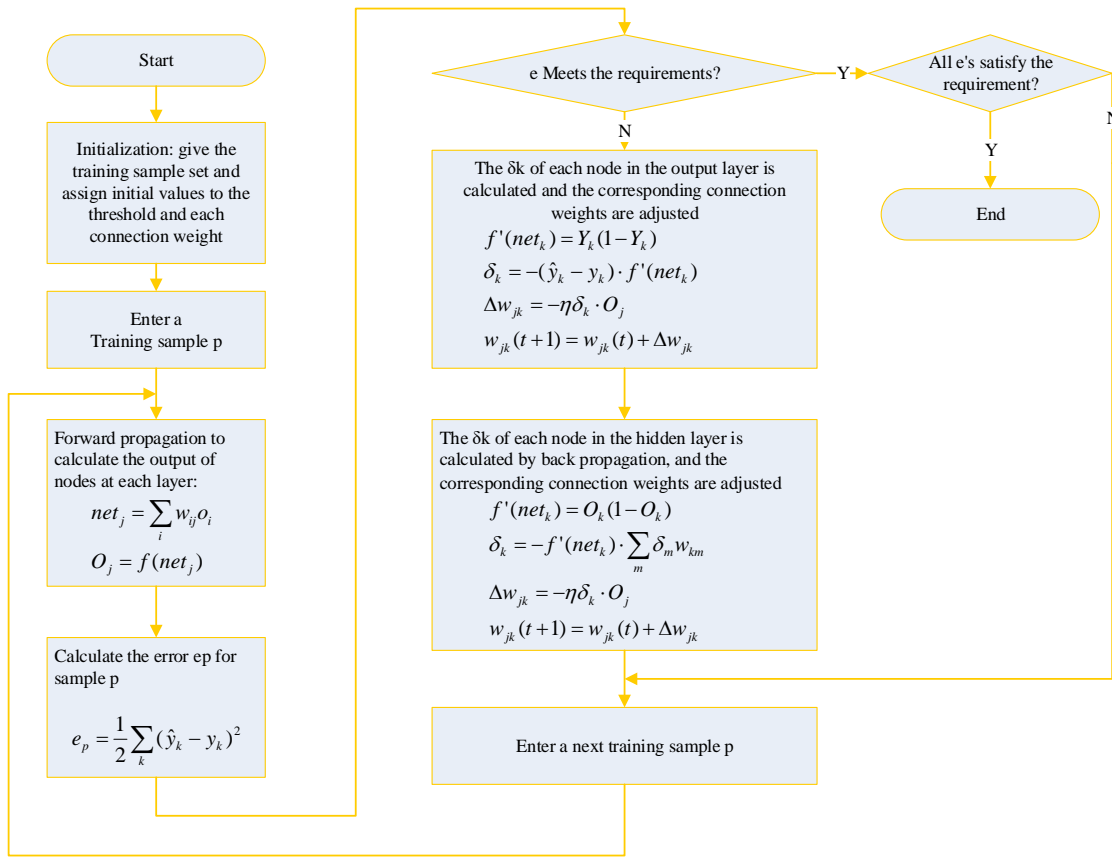


Fig. 3. BP neural network training flow chart.

Hidden layer output variable is shown in formula (5).

$$h_j = (h_1, h_2, \dots, h_p) = f\left(\sum_{i=1}^n w_{ij} x_i + a_j\right), j = 1, 2, \dots, p \quad (5)$$

In formula (5), a_j is the bias, and the actual output variable of the output layer is shown in formula (6).

$$y_k = (y_1, y_2, \dots, y_q) = \sum_{j=1}^p h_j w_{jk} + b_k, k = 1, 2, \dots, q \quad (6)$$

In Equation (7), the expected output variable of the output layer is shown in Equation (8).

$$d_k = (d_1, d_2, \dots, d_q), k = 1, 2, \dots, q \quad (7)$$

The error function is shown in formula (8).

$$E = \frac{1}{2} \sum_{k=1}^q (d_k - y_k)^2 \quad (8)$$

Assuming m training samples, the error function is shown in formula (9):

$$E = \frac{1}{2m} \sum_{l=1}^m \sum_{k=1}^q (d_k(l) - y_k(l))^2 \quad (9)$$

weight update formula is shown in formula (10).

$$\begin{cases} w_{ij} = w_{ij} + \eta h_j (1 - h_j) x_i \sum_{k=1}^q w_{jk} (d_k - y_k) \\ w_{jk} = w_{jk} + \eta h_j (d_k - y_k) \end{cases} \quad (10)$$

In formula (10), η is the learning rate, and the update formula of the bias is shown in formula (11).

$$\begin{cases} a_j = a_j + \eta h_j (1 - h_j) \sum_{k=1}^q w_{jk} (d_k - y_k) \\ b_k = b_k + \eta (d_k - y_k) \end{cases} \quad (11)$$

In the 1970s, genetic algorithms were introduced into computers and achieved great success in many practical problems in the 1980s, thus arousing great interest from people. Genetic algorithm refers to and imitates the genetic mechanisms and natural selection of organisms, adopting a "survival of the fittest" approach to continuously evolve the solution to a problem in competition, ultimately obtaining a satisfactory solution. It screens out random information exchange and fragments with strong adaptability in the sequence of the previous generation and reconstructs them into a new population, which is called "survival of the fittest". Sometimes new bits and fragments are added to the sequence, which is called "variation". By performing genetic operations such as selection, crossover, and mutation on the population, the optimization degree of the population during the evolution

process is continuously improved, ultimately approaching global optimization. Genetic algorithm has the advantages of global optimization, while BP neural network has better local search performance. However, since the BP network is randomly given the initial weight value, it will cause a situation. After the BP training, the total weight of each training and the total weight are different. In order to improve the quality of English teaching, this paper uses BP neural network and genetic algorithm to establish a new teaching evaluation system. In the weight optimization of the neural network, first initialize the group P and determine the size of the initial group; the second step is to sort the fitness function of each person, and then determine the individual in the network according to the probability value [22]. The selection method is shown in formula (12):

$$P_i = f(i) / \sum_{i=1}^n f(i) \quad (12)$$

In formula (12), is $f(i)$ the fitness of individual i ; The third step is to cultivate a new generation of individuals through genetic modification, and eliminate unqualified parents; the fourth step is to eliminate the new generation of individuals, insert them into the original population P, and perform fitness function calculations on them, and then use the error of the neural network to determine whether it meets the expectation, if it is satisfied, go to the sixth step; otherwise, go to step 3 and perform genetic manipulation; the sixth step, according to the error accuracy required by the neural network, on this basis, the best individual is selected as the weight of the neural network and operated on[23-24]. The probability of selecting individual i is shown in formula (13).

$$P_i = f_i / \sum_{k=1}^M f_k \quad (13)$$

Formula (13), the initial population is M and the fitness of individual i is f_i . After obtaining the probability of each individual, it will determine which individuals to mate with by random numbers between [0,1] Activities [21]. Since the interleaving mode has a great relationship with the actual coding situation, this article only discusses the coding method of floating-point numbers, and the form of arithmetic interleaving is as formula (14)[25].

$$\begin{cases} X_A^{t+1} = \partial X_B^t + (1-\partial)X_A^t \\ X_B^{t+1} = \partial X_A^t + (1-\partial)X_B^t \end{cases} \quad (14)$$

Formula (14), X is an individual and ∂ a parameter. If it is constant, it is an equation operation. If it is determined by evolutionary algebra, it is inconsistent. The last one is mutation, which uses the method of uniform mutation. Homogeneous mutation is to replace the gene values of all genes on a single chromosome with random numbers that conform to the average distribution with a certain probability[26]. The gene on each chromosome is regarded as a change point, and then according to different mutation points,

the initialization and termination of the genetic algorithm will be iterated repeatedly, and fitness calculation, replication and crossover will be performed in each iteration, genetic operation, such as a mutation, will not stop repeating until the end condition is reached[27]. Depending on the actual situation, different applications will have different end conditions. The end condition of this article is that it will stop when the number of repetitions is reached. The flowchart of the GA-BP neural network algorithm is shown in Fig. 4.

The initial value is compared with the adaptive function of BP neural network; obtain new populations through operations such as crossover and mutation until the goal is optimized, and then repeat the genetic operation until the genetic iteration is completed. By optimizing the BP neural network[28]. On this basis, the optimal initial value and minimum threshold can be obtained. This method mainly includes the initial and threshold of BP neural network, the learning and experiment of BP neural network[29-30]. Finally, the average cross-entropy of the verification set is calculated, and the model with the smallest average cross-entropy error of the verification set is determined as the final evaluation model, see formula (15).

$$Acc = \frac{1}{k} \sum_{i=1}^k acc_i \quad (i=1,2,\dots,k) \quad (15)$$

In formula (15), acc_i is the cross-entropy error of the first i verification set. Finally, the accuracy rate of college English teaching evaluation quality evaluation is recorded as formula (16).

$$P = \frac{N_{right}}{N_{all}} * 100\% \quad (16)$$

In Equation (16), N_{right} is several factors affecting the quality of English teaching, and N_{all} is the total number of items in the test set. Common accuracy evaluation indicators include recall rate (Recall) and F1 value[31-32]. The calculation method is shown in the formula (17) and formula (18).

$$Recall = \frac{TP}{TP + FN} \quad (17)$$

$$F1 = \frac{2 * Precision * Recall}{(Precision + Recall)} \quad (18)$$

Formulas (17) and (18), it TP represents the number of the samples that were actually positive and those that were predicted to be positive, referred to as true positive numbers, and so on TN , FP , FN are true negative numbers, false positive numbers, and false negative numbers[33-34]. Considering that in English teaching, it is of great significance to identify correct pronunciation and wrong pronunciation, so the accuracy rate is selected as the performance evaluation index of each model.

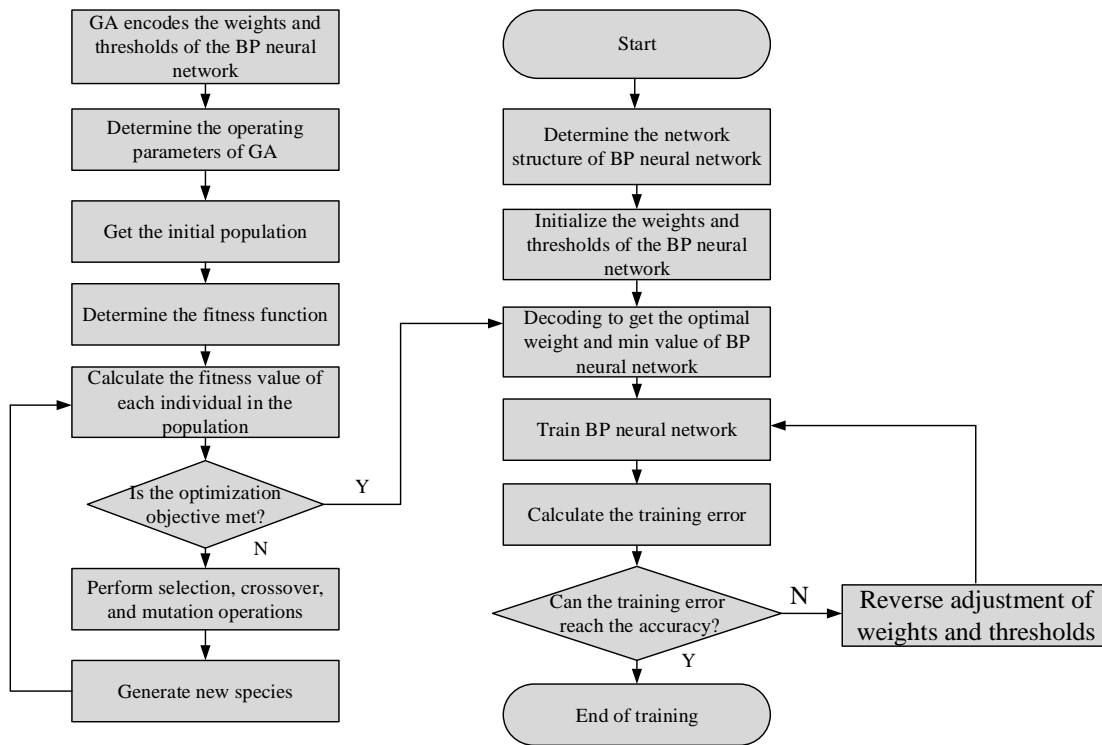


Fig. 4. GA-BP neural network algorithm flow chart.

IV. APPLICATION EFFECT ANALYSIS OF IMPROVED NEURAL NETWORK ALGORITHM IN COLLEGE ENGLISH MOOC EDUCATION AND EVALUATION

The study uses ten evaluation indicators of five universities as the input data of this study, and vectorizes the time to obtain the relevant teaching measurement score input matrix. Due to the difference in the dimension of each indicator, the deceleration rate of each indicator is different; adopting the same learning rate makes it difficult for the network training to achieve the optimal solution. At the same time, it avoids the gradient explosion caused by the index data with too large value, and the index with too small value is swallowed. Among them, in the experiment, the memory of each algorithm was 512Mb, the encoding tool was C++, and the operating system was Windows XP. The learning of BP network is an important content of BP neural network. In order to ensure the consistency of the distribution of numbers and models, this paper proposes a unified standard to combine the training set with the test set and standardize the data. Table I shows the details after normalization.

In the process of establishing the model, the Sigmoid function and the tanh function carry out the input data in the interval [0,1]. Most of them are in a suppressed state, and the differential interval is below 1. The gradient value shows an exponential decline with the increase of network depth, and there may be a gradual phenomenon. The Relu function is activated when the input value exceeds 0. It has a wide activation range and can well avoid the gradient disappearance phenomenon in network training. Neither the Reluc function nor the Reluc differential has performed any complicated mathematical operations, and will it be directly set to 0, thereby avoiding the participation of restricted neurons, thus making the learning of the neural network faster. However, Relu sets the initial state of the neuron to 0, so that it cannot perform subsequent operations, thus causing damage to the neuron during the learning process. Leaky_Relu is a modification of the Relu function, setting the inhibition state of Relu to a very small parameter α , which can well alleviate the weakness of the neural network. It needs to be adjusted continuously through network training, and α the setting is not easy to master. On this basis, different training errors and training times are obtained, as shown in Table II.

TABLE I. NORMALIZED DATA

-	x1 _	x2 _	x3 _	x4 _	x5 _	X 6	X 7	X 8	X 9	X 10
School 1	0.11	0.21	0.54	0.54	0.43	0.19	0.55	0.61	0.54	0.63
School 2	0.13	0.35	0.32	0.38	0.13	0.36	0.19	0.35	0.64	0.38
School 3	0.68	0.18	0.48	0.21	0.23	0.81	0.81	0.31	0.94	0.72
School 4	0.61	0.16	0.39	0.64	0.78	0.44	0.52	0.25	0.15	0.14
School 5	0.54	0.09	0.21	0.64	0.46	0.64	0.64	0.12	0.11	0.52

TABLE II. COMPARISON OF EXPERIMENTAL RESULTS BETWEEN RELU FUNCTION AND LEAKY_RELU FUNCTION

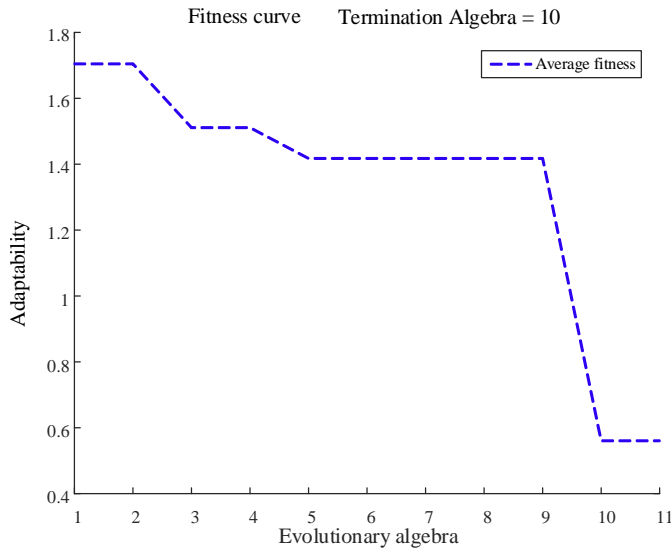
	Relu function		Leaky_ReLU function	
	Training error	Training time	Training error	Training time
The first time	0.0218	105	0.0238	118
The second time	0.0235	102	0.0240	123
The third time	0.0216	101	0.0209	109
The fourth time	0.0211	102	0.0210	115
fifth time	0.0219	103	0.0256	117

The BP neural network uses the Relu function as the initial function. BP neural network has a high recognition rate in the (0,1) interval. The output of the BP neural network is the graded gradient of the quality of school education, and its grade gradient is 0~8, which belongs to multi-category problems. Generally, it is a softmax function. This function is the probability of mapping the output of multiple neurons to the (0,1) interval, and the total probability sum of each neuron is 1. After determining the GA-BP model, import the normalized data into the model for further analysis, and first obtain its fitness curve and the training error curve of the neural network, as shown in Fig. 5.

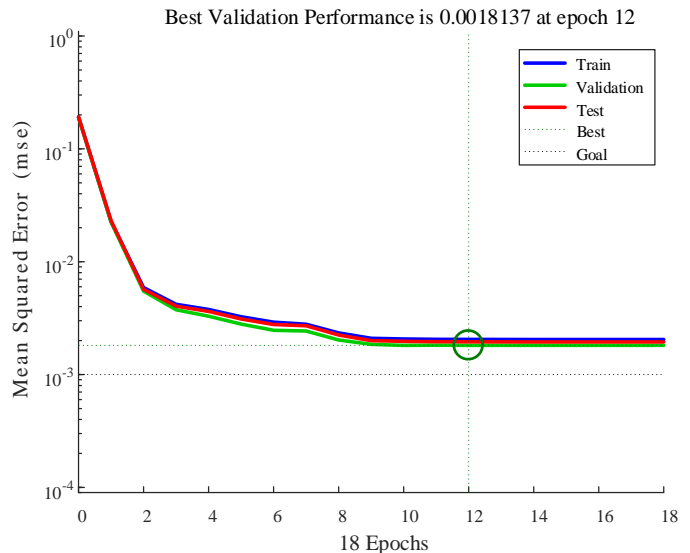
As shown in Fig. 5, the improved BP neural network model of the genetic algorithm has the desired accuracy; as shown in

Fig. 5(a), when the fitness value is 10 times, the evolution algebra of the GA-BP model is 10 times. The set expectation has a fitness of 0.6, and the corresponding threshold and weight are obtained. As shown in Figure 5(b), the threshold and weight are substituted into the model, and repeated iterative training is carried out. Finally, when the training reaches 12 times, the model error of 10^{-3} was achieved, achieving the expected accuracy. The error histograms of each set are shown in Fig. 6.

As can be seen from the graph, the error of the training set is mainly concentrated at the zero point, and the difference between the expected and output is mainly at [-0.08083,0.06481]. The predictor results show that the model based on the GA-BP neural network has a higher predictive ability and better stability.



(a) Fitness curve of GA-BP neural network



(b) Graph of the training result of GA-BP neural network

Fig. 5. GA-BP neural network training performance renderings.

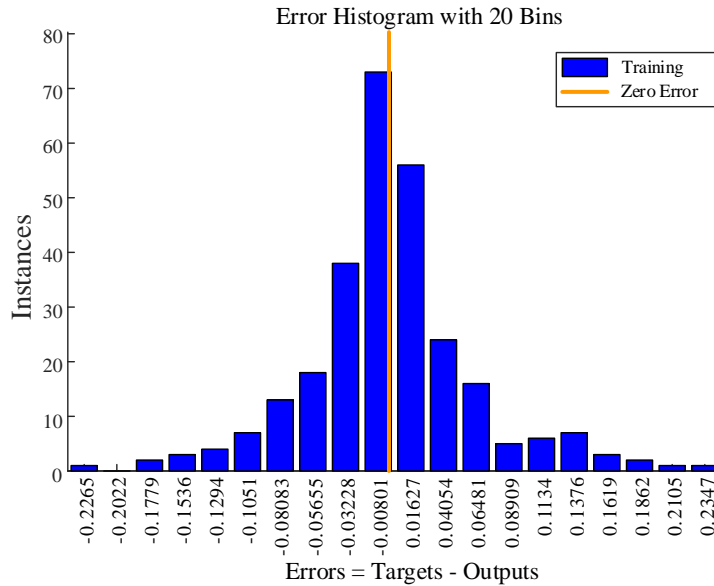


Fig. 6. The training error histogram of GA-BP neural network.

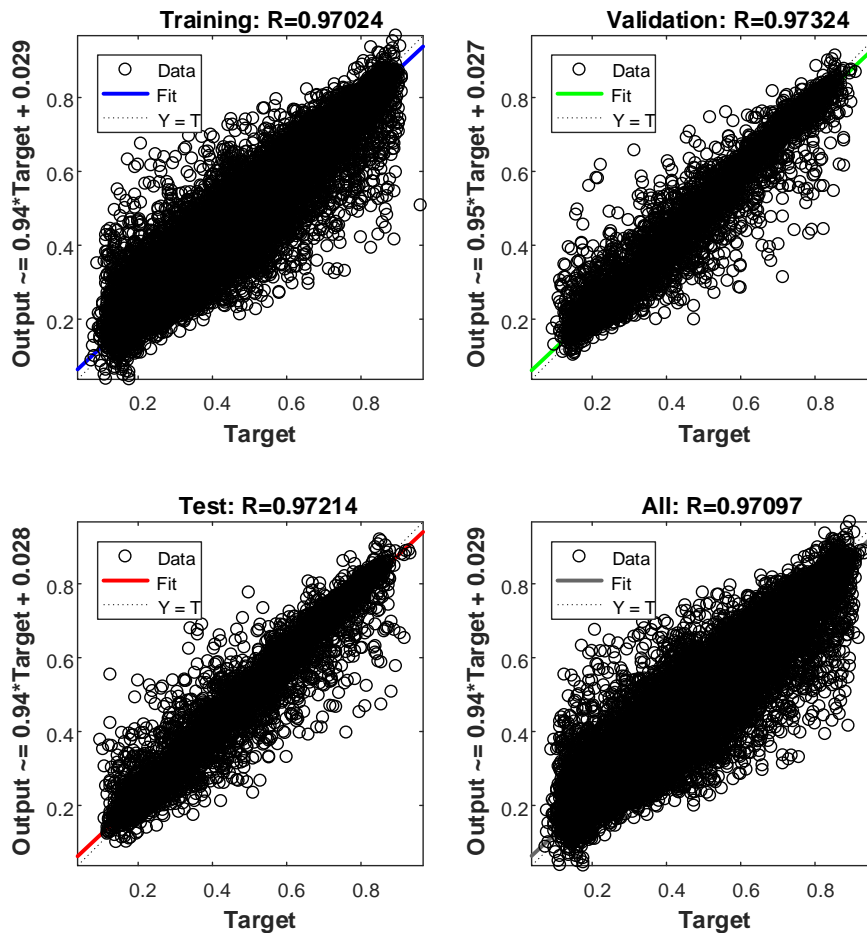


Fig. 7. Fitted curve after training.

Can be seen from Fig. 7 that the predicted value of the training set is relatively high, and deviations will occur in a few cases. The R values of each set hover around 0.97, and the fitting degree is high. The R value of the training set is

0.97024. The results show that the fitting effect of the GA-BP model established in this paper is better. By comparing the BP network and the GA-BP network, a conclusion similar to that in Fig. 8 is drawn.

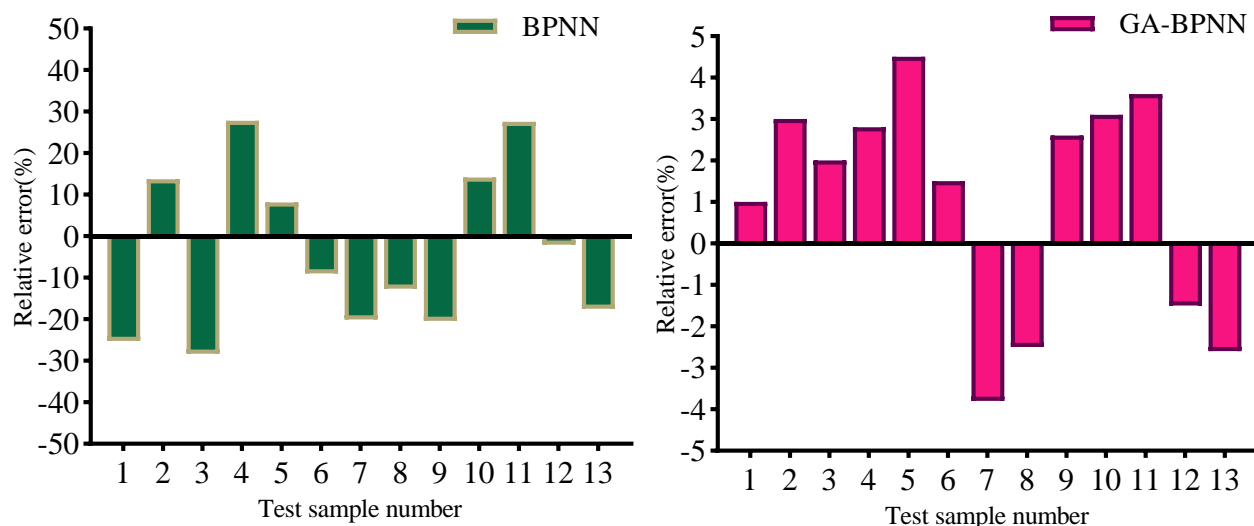


Fig. 8. The average relative error of the two models.

From the relative error histogram in Fig. 8, it can be clearly seen that the absolute value of the relative error of the GA-BPNN model proposed in this study does not exceed the 5% limit of each evaluation index. Only the relative errors of individual evaluation indicators have big problems, their values are 4.52% and -3.84%, and the relative errors of other indicators are kept at a low level, and their errors are much smaller than those of the BPNN model. The GA-BP model proposed in the study has an excellent effect on college English education and evaluation. This evaluation model has a faster learning rate, higher prediction accuracy and more stable performance.

V. CONCLUSION

Under the current education model, traditional English learning methods are mostly teaching-based. Fostering student autonomy is a hot topic in education at the moment. How to let students focus more and learn in a broader range has also become a topic for scholars—the focus of attention. In the traditional classroom environment, with classrooms and libraries as the main body, the emergence of multimedia classrooms has increased the diversity of classroom teaching to some extent due to the wide application of computers. To ensure the fairness of compulsory education, it is highly relevant in terms of assessment of the teaching quality of the College English MOOC. This thesis aims to introduce genetic algorithm into college English teaching and evaluation and improve it. Deviations between the expected and the real values are essentially $[-0.08083, 0.06481]$. It can be known that the prediction performance of the model is robust. The R values of each set hover around 0.97, and the fitting degree is high. Finally, a comparative analysis of the relative prediction errors of the two models shows that the absolute value of the relative error of the proposed GA-BP model does not exceed that of each evaluation. The limit value of 5% of the index is better than a single algorithm. In short, the research can reflect the actual situation of delivering and assessing English in colleges and universities, and has substantial practical application value.

REFERENCES

- [1] X. Jia, "A Review of the Teaching Model of Mental Health Education Courses in Primary and Secondary Schools in the New Media Environment." *Education Study*, vol. 3(2), pp. 225-227, 2021.
- [2] M. Abdullah, S. B. Thalib, G. D. Dirawan, "The Quality of Integrating, Rationale & Approach Participation (IRA) Teaching Model Based on Environmental Education: A Research & Development." *Universal Journal of Educational Research*, 2021, vol. 9(8), pp. 1560-1571, 2021.
- [3] C. A. Salazar, C. I. Cantillo, H. Muoz, "Contribution of the implementation of an alternative teaching model in the teaching of mathematics: solving arithmetic problems in primary basics." *Espacios*, vol. 42(1), pp. 119-130, 2021.
- [4] J. Wexler, C. Lyon, E. K. Hogan, M. Devin Kearns. "Individualizing Literacy Instruction in Co-Taught Classrooms Through a Station Teaching Model." *Intervention in School and Clinic*, 2021, vol. 56(4), pp. 224-232, 2021.
- [5] C. Jiny, M. You-Mi, "A Study on the Development of Pedagogic Corpus and Suggestions for Teaching Model Using it: Focused on the Low-intermediate Students." *Chinese Studies*, 2020, vol. 71, pp. 221-242, 2020.
- [6] Y. Wang, "Ideological and political teaching model using fuzzy analytic hierarchy process based on machine learning and artificial intelligence". *Journal of Intelligent and Fuzzy Systems*, 2020, vol. 40(6), pp. 1-13, 2020.
- [7] G. H. Lunde, A. Bakke, K. Areskoug-Josefsson. "Piloting a Research-Oriented Teaching Model in a Bachelor Program for Social Educators – A Way to Increase Competence in Research Methodology and Sexual Health." *Uniped*, 2020, vol. 43(3), pp. 260-274, 2020.
- [8] J. Sotomayor-Moriano, G. Pérez-Zúiga, M. Soto, "Teaching Model-based Fault Detection and Isolation using a Virtual Laboratory Environment." *IFAC-PapersOnLine*, 2020, vol. 53(2), pp. 17350-17355, 2020.
- [9] R. Putra, S. Suherman, B. S. Anggoro. "Alqurun Teaching Model-Based Trigonometry Teaching Material." *Indonesian Journal of Science and Mathematics Education*, 2020, vol. 3(2), pp. 219-227, 2020.
- [10] M. Nursalam, H. E. Fitriana, J. Jusmawati, "Efektifitas Model Quantum Teaching Terhadap Pembelajaran Matematika Siswa di Sekolah Dasar." *Journal Basicedu*, 2021, vol. 5(2), pp. 506-516, 2021.
- [11] G. L. Chaves, R. S. Batista, J. Cunha, "Teaching cellular metabolism using metabolic model simulations." *Education for Chemical Engineers*, vol. 38, pp. 97-109, 2022.
- [12] D. Mourtzis, N. Panopoulos, J. Angelopoulos. "A Hybrid Teaching Factory Model for Supporting the Educational Process in COVID-19 era." *Procedia CIRP*, vol. 104, pp. 1626-1631, 2021.

- [13] S. McConnell, C. Mooney. "A Crocheted Model Activity for Teaching Embryonic Lateral Folding to Medical Students." *Anatomical sciences education*, vol. 14(5), pp. 666-674, 2021.
- [14] N. Sutarna, N. Nurfirdaus. "Bahan Ajar Berbasis Model Quantum Teaching Untuk Meningkatkan Kemampuan Berpikir Kritis." *Naturalistic Journal Kajian Penelitian Pendidikan dan Pembelajaran*, vol. 4(1), pp. 5.17-42, 2020.
- [15] C. González-Velasco, Feito-Ruiz I., M. G. Fernández. "Does the teaching-learning model based on the flipped classroom improve academic results of students at different educational levels". *Revista Complutense de Educacion*, vol. 32(1), pp. 27-39, 2021.
- [16] Yin H. The recommendation method for distance learning resources of college English under the MOOC education mode. *International journal of continuing engineering education and life-long learning*, 2022,32(2):265-278.
- [17] Xie H, Mai Q. College English cross-cultural teaching based on cloud computing MOOC platform and artificial intelligence. *Journal of Intelligent & Fuzzy Systems: Applications in Engineering and Technology*, 2021,40(4):7335-7345.
- [18] D. G. Bratt, C. Berridge, M. Young, "A simple novel training model for teaching suprapubic catheter (SPC) exchange." *Actas Urológicas Españolas (English Edition)*, vol. 44(8), pp. 549-553, 2020.
- [19] X. Chang, "An Integrated Model of Teaching Theory and Action Research in POA-based Textbook Writing." *Chinese Journal of Applied Linguistics*, vol. 43(3), pp. 359-372, 2020.
- [20] M. Keller, D. Ritter, L. Schmitt. "Teaching Nonlinear Model Predictive Control with MATLAB/Simulink and an Internal Combustion Engine Test Bench". *IFAC-Papers on Line*, vol. 53(2), pp. 17190-17197, 2020.
- [21] D. Park, J. Ahn, J. Jang. "The Development of Software Teaching-Learning Model based on Machine Learning Platform." *Journal of the Korean Association of Information Education*, vol. 24(1), pp. 49- 57, 2020.
- [22] E. Lee, Developing a Low-Cost Microcontroller-Based Model for Teaching and Learning. *European Journal of Educational Research*, vol. 9(3), pp. 921-934, 2020.
- [23] C. Slab, D. Assa, E. Fq, "Use of a virtual 3D anterolateral thigh model in medical education: Augmentation and not replacement of traditional teaching" - ScienceDirect. *Journal of Plastic, Reconstructive & Aesthetic Surgery*, vol. 73(2), pp. 269-275, 2020.
- [24] B. Abdelkader, "Optimization of the Geometric Model Neuronal (BPNN) of a Polyarticulated Arm." *Journal of Advanced Research in Dynamical and Control Systems*, vol. 12(4), pp. 1137-1146, 2020.
- [25] J. Chen, "DDoS Attack Target Detection Based on AM+BPNN." *Scientific Journal of Technology*, vol. 4(8), pp. 45-49, 2022.
- [26] H. Fu, Y. Liu. "A deep learning-based approach for electrical equipment remaining useful life prediction". *Autonomous Intelligent Systems*, vol. 2(1), pp. 1-12, 2022.
- [27] J. Liu, H. Shao, Y. Jiang, X. Deng, "CNN-Based Hidden-Layer Topological Structure Design and Optimization Methods for Image Classification. *Neural Processing Letters*, vol. 54(4), pp. 2831-2842.
- [28] M. Mahboubkhah, A. Barari, M. Aliakbari, "Computer integrated work-space quality improvement of the C4 parallel robot CMM based on kinematic error model for using in intelligent measuring." *International Journal of Computer Integrated Manufacturing*, vol. 35(4-5), pp. 444-461, 2022.
- [29] J. Pu, Z. Liu. "Analysis and research on intelligent manufacturing medical product design and intelligent hospital system dynamics based on machine learning under big data". *Enterprise Information Systems*, vol. 16(2), pp. 193-207, 2022.
- [30] J. Zhang, J. Fan, Yang J., J. Yu. "Semisupervised image classification by mutual learning of multiple self-supervised models." *International Journal of Intelligent Systems*, vol. 37(5), pp. 3117-3141, 2022.
- [31] A. Chen, S. Hong, Y. Wang, C. Li., C. Yang, H. Chen. "Rapid Assessment of Gasoline Quality by near-Infrared (NIR) Deep Learning Model Combined with Fractional Derivative Pretreatment." *Analytical Letters*, vol. 55(11), pp. 1745-1756, 2022.
- [32] X. Lu, W. Liao, Y. Zhang, Y. Huang. "Intelligent structural design of shear wall residence using physics-enhanced generative adversarial networks." *Earthquake Engineering and Structural Dynamics*, vol. 51(7), pp. 1657-1676, 2022.
- [33] M. B. Umair, Z. Iqbal, F. Z. Khan, M. Khan, S. Kadry, "A Deep Learning Based Method for Network Application Classification in Software-Defined IoT." *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 30(03), pp. 463-477, 2022.
- [34] T. A. M. Devi, P. Darwin, "Hyper Spectral Fruit Image Classification for Deep Learning Approaches and Neural Network Techniques." *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 30(03), pp. 357-383, 2022.

Employee Information Security Awareness in the Power Generation Sector of PT ABC

Ridwan Fadlika¹, Yova Ruldeviyani², Zenfrison Tuah Butarbutar³,
Relaci Aprilia Istiqomah⁴, Achmad Arzal Fariz⁵

Faculty of Computer Science, Universitas Indonesia, Depok, Indonesia^{1,2,3,4,5}

Abstract—Presidential Regulation No. 82 of 2022 demonstrates the Indonesian government's dedication to protecting Vital Information Infrastructure, which has become increasingly susceptible to cyber attacks. Intrusion detections at PT ABC reached 79,575 in 2021, and malware, botnets, targeted attacks, malicious websites/domains, and ransomware attacks may cause considerable financial losses. The implication of these incidents is that employees' awareness of information security is critical, in addition to security technologies like firewalls and monitoring tools. To enhance employees' knowledge of information security, this study aims to evaluate the information security awareness among PT ABC personnel using the HAIS-Q survey instrument alongside ISO/IEC 27001:2013 criteria. The study will provide valuable recommendations to improve the organization's security protocols. This research intends to investigate the correlation between employees' knowledge, attitude, and behavior towards information security. Data was collected through a questionnaire and analyzed using the Pearson Correlation, Cronbach's Alpha, descriptive statistics, linear regression, and Kruskal-Wallis test method. The study findings suggest that the overall information security awareness level among employees is "Good". However, certain areas like internet usage, information handling, asset management, incident reporting, and the use of mobile devices need improvement. To address these areas, the study recommends promoting information security awareness according to employee categories.

Keywords—Security awareness; data; information; ISO/IEC 27001:2013

I. INTRODUCTION

The Indonesian government through Presidential Regulation No. 82 of 2022 [1] pays attention to and is committed to protecting Vital Information Infrastructure due to the abuse of information and electronic transactions. Threats to the security of vital objects such as power plants have been experienced by the Gundremmingen nuclear power plant in Germany in 2016 where the "W32.Ramnit" and "Conficker" viruses were attacked through an employee's USB device¹.

The 2021 BSSN Report on Cybersecurity Monitoring reports that one of the background causes of data leaks is phishing [2]. The phishing method is where the hacker infiltrates malicious codes through an e-mail or website page

during internet browsing [3][4]. Monitoring data from PT ABC states that the number of intrusion detections during 2021 was 79,575. Cyber attacks such as malware, botnets, targeted attacks, malicious websites/domains, and ransomware attacking the company can result in significant financial losses [3][4]. The lesson learned from these incidents is the need for information security awareness among employees at PT ABC, as security technologies such as firewalls or monitoring tools play an important role in security, but the human factor must also be considered [4].

The measurement of awareness of information security has been the subject of numerous prior studies. Vina Effendy et al. (2022) conducted a study utilizing the HAIS-Q modeling to evaluate the level of information security awareness at XYZ polytechnic. The findings of the study revealed that the level of awareness was at a medium level at the research site, indicating the need for further monitoring to enhance the level of awareness. However, the authors did not provide recommendations based on employee criteria [5]. Another study by Aulia Zulfia et al. (2019) employed the HAIS-Q method to measure information security awareness at PT PQS. Nevertheless, the authors did not provide recommendations based on employee criteria [6]. In a similar vein, Rahardi Prakoso et al. (2020) measured awareness of information security among online transportation users using the HAIS-Q method. The authors identified the areas that require improvement, but did not provide recommendations based on sub-area categories among respondent demographics [7].

The Human Aspects of Information Security-Questionnaire (HAIS-Q) is a widely recognized tool for evaluating global information security awareness. Numerous studies, including [5][6][7][8][21], have utilized the HAIS-Q in various contexts, spanning commercial enterprises, academic institutions, and government agencies. Despite its extensive adoption, previous research has yet to integrate the HAIS-Q with the ISO/IEC 27001:2013 standard, and no research has specifically investigated the extent of awareness of information security among employees of PT ABC.

The motivation described above has instigated a research initiative aimed at assessing the awareness of information security of the PT ABC personnel. The HAIS-Q survey instrument, in conjunction with the ISO/IEC 27001:2013 criteria will be utilized to achieve this goal. The outcomes of this investigation will furnish recommendations for enhancing the organization's security protocols. It is expected that these

¹S. Christoph and A. Eric, 'German nuclear plant infected with computer viruses, operator says', *Reuters*, 2016, <https://www.reuters.com/article/us-nuclearpower-cyber-germany-idUSKCN0XN2OS>, (accessed 10 October 2022)

insights will have a favorable influence on the information security awareness level at PT ABC.

The ensuing section constitutes the second component of this paper and aims to expound upon the literature review. Subsequently, the third section delineates the theoretical framework, followed by the fourth section which explicates the research methodology. The fifth section comprises a thorough analysis and discourse of the outcomes. Finally, a conclusion will be presented to summarize the findings.

II. LITERATURE REVIEW

A. Information Security Awareness

The field of information security is concerned with safeguarding both information and the systems utilized to transmit, store, or manipulate it [9]. Management of information security entails not only considerations of technology, but also concerns pertaining to human users of the system, as evidenced by the importance of information security awareness [6]. Information security awareness encompasses two distinct dimensions, namely the degree to which users comprehend information security practices and the extent to which they are willing to adhere to organizational policies, rules, and guidelines [5]. The 2021 Annual Cybersecurity Monitoring Report [2] identifies weak human awareness as the primary factor contributing to anomalous network traffic. Cybersecurity training and awareness programs can be broken down into three components [10].

1) *Education*, which aims to impart wisdom on the importance of information security for the organization.

2) *Training*, which teaches users how to use security functions in the information system and in their work processes; and

3) *Awareness*, which builds on the foundation provided by education and training to promote individuals' knowledge of and adherence to best security, safety, and privacy practices [11].

B. HAIS-Q and KAB (Knowledge-Attitude-Behavior) Model

The Human Aspects of Information Security Questionnaire (HAIS-Q) is a validated assessment tool that enables the evaluation of individuals' level of awareness related to information security [12]. The HAIS-Q encompasses seven distinct domains, namely, password management, email usage, internet usage, social media utilization, mobile devices usage, information controlling, and incident reporting [4]. Furthermore, these seven areas are classified into three dimensions that are commonly known as KAB (Knowledge, Attitude, and Behavior) [13]. Each dimension can be elucidated as follows: a) Knowledge pertains to an individual's comprehension of information, b) Attitude denotes an individual's opinion, and c) Behavior pertains to an individual's disposition to undertake actions.

Users with high scores according to HAIS-Q perform better in phishing experiments, showing that HAIS-Q is a good framework for measuring users' information security awareness level [14].

C. ISO/IEC 27001:2013

The ISO/IEC 27001:2013 standards are universally employed for managing information security. These requirements dictate the establishment, implementation, maintenance, and continuous improvement of an organization's strategic decisions [15]. Moreover, they govern the application of management systems based on the PDCA approach, along with supplementary information security controls.

ISO/IEC 27001:2013 consists of 7 clauses and 14 information security control areas comprising 114 control points. Like other ISO standards based on high level PDCA this information security management system standard has a difference in clause 8, which is operation. The main point of attention is how to control information security risks outlined in Annex A.

This research focuses on some information security controls found in Annex A, within the scope of individuals' awareness of information security. Out of the 114 existing controls, some relevant to individual awareness will be selected, such as mobile devices (A.6.2.1), password management (A.9.4.3), email usage (A.13.2.3), internet usage (A.13.2.1), social media usage (A.18.1.4), information handling (A.8.3), incident reporting (A.16.1.3), and asset management (A.12.3 and A.12.5).

D. Validation and Reliability Test Method

1) *Validation test*: The method used to identify the validity of the questionnaire data is the Pearson Method [16]. Bivariate Pearson Correlation is used to determine the correlation between two variables x and y based on Eq. (1).

$$r = \frac{(\sum xy - \frac{\sum x \sum y}{N})}{\sqrt{(\sum x^2 - \frac{(\sum x)^2}{N})(\sum y^2 - \frac{(\sum y)^2}{N})}},$$
$$-1 \leq r \leq +1 \quad [16] \quad (1)$$

The variables in the formula are defined as follows: r is the Bivariate Pearson Correlation coefficient, N represents the number of data points, while x and y represent the first and second variables, respectively.

2) *Reliability test*: In identifying the reliability value of the questionnaire data, this research uses the Cronbach's Alpha method [16]. The Cronbach's Alpha value is used to measure internal consistency based on Eq. (2).

$$\alpha = \frac{k}{k-1} \left(1 - \frac{\sum s_i^2}{s_t^2} \right) \quad [16] \quad (2)$$

where k is the number of questions, s_i is the variance of each question item, s_t is the variance of the group, and α is the reliability value.

3) *Linear regression*: In the realm of hypothesis testing for the K-A-B relationship, the Linear Regression approach (as detailed in research [16]) is utilized to ascertain the degree to which the independent variable x can affect variations in the dependent variable y . Eq. (3) is employed to perform regression analysis and determine the R-squared value, which indicates the extent to which the independent variable can

account for variability in the dependent variable. The formula for R-squared, denoted as r^2 , is presented in study [16] as follows.

$$r^2 = \frac{(\sum xy - \frac{\sum x \sum y}{N})^2}{(\sum x^2 - \frac{(\sum x)^2}{N})(\sum y^2 - \frac{(\sum y)^2}{N})} \quad [16] (3)$$

In Eq. (3), N refers to the number of data points, x denotes the first variable, and y represents the second variable.

4) *Descriptive statistics*: The Mean value is used to determine the average value of a variable based on Eq. (4).

$$\bar{x} = \frac{\sum x}{N} \quad [16] (4)$$

where \bar{x} is the mean, N is the number of data, and x is the value of the variable.

The computation of the degree of variability in a variable is ascertained by the employment of the Standard Deviation (SD) value in accordance with Eq. (5), which is represented as:

$$\sigma = \sqrt{\frac{\sum x^2 - \frac{(\sum x)^2}{N}}{N}} \quad [16] (5)$$

In this formula, σ refers to the standard deviation value, x represents the value of the variable, N signifies the total number of data [16].

E. Significant Difference Test

The Kruskal-Wallis Test [16] is employed in statistical analysis to assess the degree of variation between two or more

groups, pertaining to a particular area of interest, by examining values that signify significant differences. As a non-parametric, rank-based test, this method utilizes the mean rank to determine the extent of variation between groups.

The mean rank value is calculated based on Eq. (6):

$$\bar{R}_A = \frac{\sum_{i=1}^{n_A} R_{Ai}}{n_A} \quad [16] (6)$$

where n_A is the number of samples in a particular group for a focus area, R_{Ai} is the rank of a focus area for a sample in a specific respondent group, \bar{R}_A is the mean rank of a focus area for a single respondent group, N is the number of data.

The Kruskal-Wallis Test for a single respondent group is calculated using Eq. (7):

$$H = \frac{n_A[\sum_{i=1}^{n_A} (R_{Ai} - (N+1)/2)^2]}{N(N+1)/12} \quad [16] (7)$$

III. THEORETICAL FRAMEWORK

The theoretical frameworks used in this research are the KAB Dimensions, HAIS-Q, and ISO/IEC 27001:2013. To measure the level of awareness of information security, it is required to measure the levels of Knowledge, Attitude, and Behavior from the employees' perspective, this is based on the theory proposed by Schrader & Lawless (2004) [13]. The researcher then focuses the measurement area on some measurement items based on HAIS-Q and ISO/IEC 270001:2013. The theoretical framework is shown in Fig. 1.

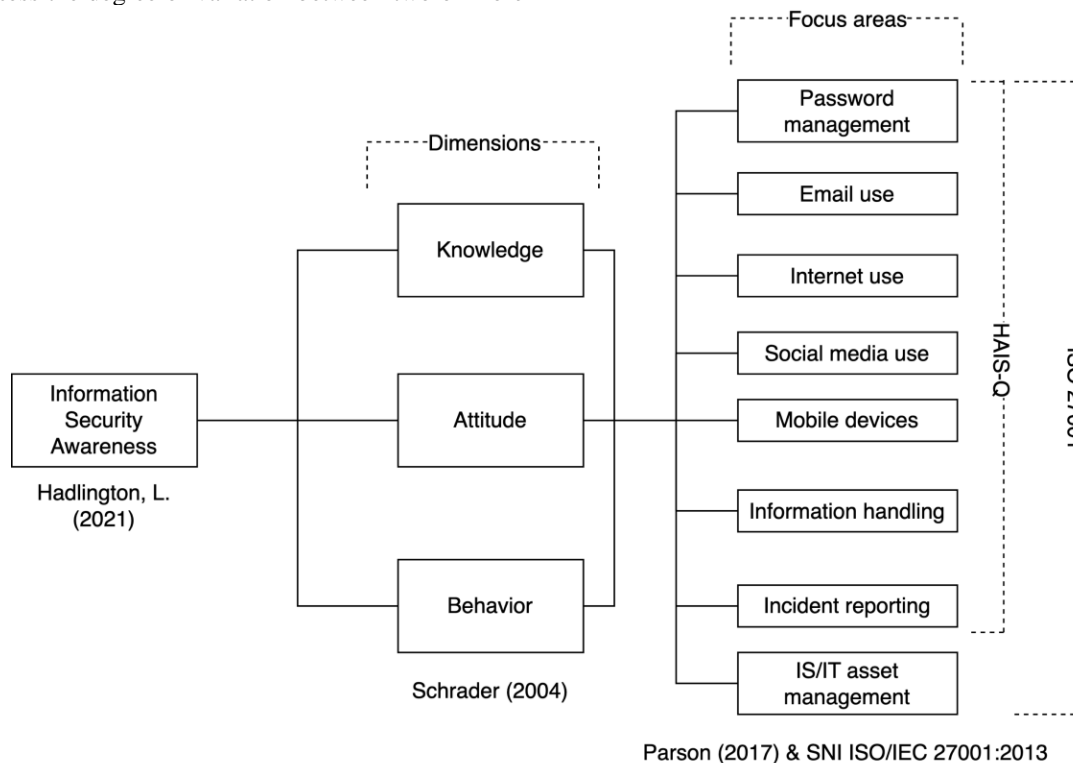


Fig. 1. Theoretical framework.

The research hypothesis consists of:

- H1: The knowledge dimension has a significant effect on the attitude dimension,
- H2: The knowledge dimension has a significant effect on the behavior dimension,
- H3: The attitude dimension has a noteworthy impact on the behavior dimension.

The visual representation of the proposed research hypothesis is depicted in Fig. 2.

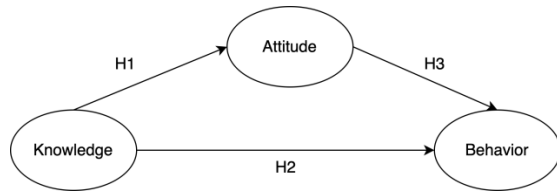


Fig. 2. The research hypothesis.

IV. RESEARCH METHODOLOGY

This section explains four important things for collecting evidence that justifies the conclusion made, which are A) research instrument, B) stages of research, C) data collection method, and D) data processing and analysis method.

A. Research Instrument

This study utilizes a questionnaire as a research tool, incorporating insights from prior research works, including [5][6][7][8]. The questionnaire is comprised of question components from seven key areas of the HAIS-Q [12] and eight areas of the ISO/IEC 27001:2013 standard [15]. Notably, seven of these areas exhibit significant overlap, encompassing password management, email usage, internet usage, social media usage, mobile device usage, information controlling, and incident reporting. However, one area - pertaining to IT/IS asset management - does not share this overlap. Consequently, the focus areas for this study encompass password management, email usage, internet usage, social media usage, mobile device usage, information controlling, incident reporting, and IT/IS asset management. Additionally, each of these focus areas is further segmented into three distinct dimensions, namely Knowledge, Attitude, and Behavior (K, A, B), as outlined in Table I.

The research tool employed in this study comprises a comprehensive questionnaire consisting of 48 items that are designed to assess the levels of information security awareness among the respondents. The questionnaire is conducted to evaluate the knowledge, attitude, and behavior of the participants, pertaining to eight key areas of focus. Answers are given using the Likert scale 1-5 (1: Strongly Disagree, 2: Disagree, 3: Neutral, 4: Agree, 5: Strongly Agree).

B. Stages of Research

The stages of research of this study can be outlined as follows:

1) *Research problem identification*: Identify the research problem, which is the need to measure employee information security awareness in PT ABC.

2) *Literature review*: Conduct a comprehensive review of relevant literature on information security awareness, employee behavior, and security culture in the power generation sector.

3) *Research design*: Determine the research design, including the research approach, data collection methods, sample size, and data analysis techniques.

4) *Data collection*: Collect data from the employees using survey questionnaires.

5) *Data processing and analysis*: Analyze the data using the Pearson Correlation, Cronbach's Alpha, descriptive statistics, linear regression, and Kruskal-Wallis test method.

6) *Results interpretation*: Interpret the results and draw conclusions based on the findings, highlighting the level of employee information security awareness in PT ABC.

7) *Discussion*: Discuss the implications of the findings for PT ABC's information security management system and suggest recommendations for improving employee information security awareness.

8) *Conclusion*: Summarize the main findings and conclusions of the study, and highlight its contributions to the field of information security awareness in the power generation sector.

TABLE I. FOCUS AREA OF INFORMATION SECURITY AWARENESS

Focus Area	Sub-Area	Focus Area Code	Indicator Code (K, A, B)
Password management	Sharing password	MP	KMP, AMP, BPM
	Safe password usage		
Email use	Clicking on a link in an email from an unknown sender	EM	KEM, AEM, BEM
	Opening an email attachment from an unknown sender		
Internet use	Downloading a file	IN	KIN, AIN, BIN
	Entering information into the internet		
Social media use	Social media privacy settings	MS	KMS, AMS, BMS
	Posting about work		
Mobile devices	Sending sensitive information over Wi-Fi	PM	KPM, APM, BPM
	Hacking technique: shoulder surfing (observation)		
Information handling	Disposal of sensitive document printouts	PF	KPF, APF, BPF
	Use of USB/other removable media		
Incident reporting	Reporting suspicious behavior	PD	KPD, APD, BPD
	Reporting all incidents		
IS/IT asset management	Regulations regarding the installation of software on agency-owned IT assets	MA	KMA, AMA, BMA
	Data backup		

C. Data Collection

The current research utilized a questionnaire comprising 48 items, distributed via Google Form and administered to 150 employees of PT ABC using a sampling strategy. The questionnaire consisted of eight focus areas, each of which contained two questions pertaining to the dimensions of Knowledge, Attitude, and Behavior. The final sample size consisted of 130 participants, from whom the research team successfully obtained data.

D. Data Processing and Analysis Method

The method for testing validity is using Bivariate Pearson Analysis (Pearson Product Moment Correlation). The Pearson value for all variables is greater than the critical value of 0.172 based on the Pearson Critical Value Table with a sample size of 130 and a significance value of 0.05 2-tailed.

The reliability test was done using Cronbach Alpha, to test the reliability of the measurement indicators used in the research. According to J. Hair (2017), a Cronbach's Alpha value above 0.7 is considered reliable [17].

Linear regression is used to test the hypothesis with the help of SPSS software. The output from SPSS is then processed using Microsoft Excel for descriptive statistical analysis that produces mean and standard deviation data for each sub-area in the knowledge, attitude, and behavior dimensions. The index value is obtained by dividing the total score by the maximum Likert scale value (Y) multiplied by 100% as in Equation (8).

$$Indeks = \frac{Total\ Skor}{Y} \times 100\% \quad [18] \quad (8)$$

The subsequent course of action involves the computation of mean values for each dimension, based on the gathered index values. This leads to the assessment of level of information security awareness. Kruger and Kearney (2006) [18] have classified information security awareness levels into three tiers: Good (80 – 100%), Moderate (60 – 79.99%), and Poor (≤ 59.99%), as illustrated in Table II.

TABLE II. INFORMATION SECURITY AWARENESS CLASSIFICATION [18]

Awareness	Value (%)
Good	80 – 100
Average	60 – 79.99
Poor	≤ 59.99

The Kruskal-Wallis test is a statistical tool commonly utilized to ascertain the significance of categorical variables. Specifically, it is employed to evaluate and compare two or more groups within a particular domain, through the calculation of a significant difference value. This test is a non-parametric procedure, relying on the ranking of the observations (i.e., mean rank) [13].

V. RESULT AND DISCUSSION

A. Demographic of Respondents

The questionnaire was collected from 130 respondents. The demographics of the respondents in this research included job field, job title, education, and length of work at PT ABC. The composition of the respondents can be seen in Table III.

TABLE III. RESPONDENT DEMOGRAPHY

Categories	Total	Percentage
Field	Operation	49 38%
	Maintenance	25 19%
	Engineering	26 20%
	Administration	21 16%
	Other	9 7%
Job title	Structural	33 25%
	Functional	97 75%
Education	S2 (Master)	3 2%
	D4/S1 (Bachelor)	80 62%
	D3 (Three-year diploma)	18 14%
	SMA (High school)	29 22%
Job tenure	≤ 5 years	57 44%
	6 – 10 years	20 15%
	11 – 15 years	24 18%
	≤ 16 years	29 22%

In terms of job field, 38% are in the operations field followed by engineering, maintenance, and lastly, 16% are in administration, while 7% are in other fields. For job title, 75% are functional while the rest are structural. The most common education level among the respondents is D4/S1 with 62%. For length of work, the majority of the respondents have experienced for less than 5 years, accounting for 44%.

B. Validity and Reliability Test

The present study has conducted validity and reliability tests on the variables of interest. The outcome of these tests has been included in the appendix. Specifically, the Pearson correlation coefficient for the APD2 indicator code was found to be 0.108, which falls below the critical value of Pearson. Therefore, the APD2 variable has been excluded from further analysis. On the other hand, Table XV in the appendix presents evidence of the validity of all indicators related to the research variables. Subsequently, a reliability test was conducted, and the results have been presented in Table IV.

This study obtained a Cronbach's Alpha value of 0.919, indicating a high level of internal consistency reliability among the variables assessed. Based on the results of the validity and reliability tests, it can be inferred that the measurement variables utilized in this study exhibit strong validity and reliability.

C. Hypothesis Testing Results

The hypothesis testing was conducted using linear regression method which tests the significance between dimensions in the KAB modeling, which are the Knowledge dimension towards Attitude, the Knowledge dimension towards Behavior and the Attitude dimension towards Behavior.

1) Knowledge – Attitude: The regression coefficient value for the dimension of Knowledge towards Attitude shown in Table VI confirms the significance between the two dimensions (Sig: 0.000).

The findings presented in Table V and Table VI indicate a notable influence of the Knowledge dimension on the Attitude

dimension. The analysis reveals a strong positive correlation ($\beta = 0.803$) and a significant level of statistical significance ($sig < 0.001$) between the two dimensions, with a coefficient of determination (*R-squared*) of 0.644. The empirical relationship between the Knowledge and Attitude dimensions can be expressed by the equation $y = 8.492 + 0.849x$, where y represents Attitude and x denotes Knowledge.

TABLE IV. RELIABILITY TEST RESULT

Cronbach's Alpha	N of Items
0.919	48

TABLE V. COEFFICIENT OF DETERMINATION K-A

Model Summary				
Model	R	R-squared	Adjusted R-squared	Std. Error of the Estimate
1	0.803	0.644	0.642	4.596
Predictors: (Constant), K				

TABLE VI. REGRESSION COEFFICIENTS K-A

	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	8.492	3.747		2.266	0.025
K	0.849	0.056	0.803	15.230	0.000

Based on these results, it can be concluded that Hypothesis 1 is supported, as the Knowledge dimension has a significant impact on the Attitude dimension.

2) *Knowledge – Behavior*: Table VIII shows the regression coefficient values for the dimension of Knowledge towards the dimension of Behavior.

Table VII and Table VIII present the findings of the analysis that demonstrate the Knowledge dimension's impact on the Behavior dimension (correlation $\beta = 0.685$, significance $sig < 0.001$, coefficient of determination *R-squared* = 0.470). The Knowledge-Behavior (K-B) dimensions' relationship can be described by the equation $y = 15.008 + 0.767x$, where y represents the Behavior dimension, and x represents the Knowledge dimension. Thus, the results indicate that an improvement in the Knowledge dimension can lead to a corresponding increase in the Behavior dimension.

Therefore, it can be concluded that Hypothesis 2 can be accepted. The dimension of Knowledge has a significant impact on the dimension of Behavior.

3) *Attitude – Behavior*: Table X shows the regression coefficient values for the relationship between the Attitude dimension and the Behavior dimension.

The analysis conducted on the data presented in Tables IX and X indicate that Attitude significantly influences Behavior, as evidenced by a strong positive correlation ($\beta = 0.807$) and a high level of statistical significance ($sig < 0.001$), with a coefficient of determination (*R-squared*) value of 0.652. These findings suggest that enhancing employees' attitude towards information security may effectively lead to a positive change

in their behavior, which can ultimately lead to better security practices in the organization. The Attitude-Behavior (A-B) relationship can be stated with the equation $y = 10.540 + 0.855x$ where y represents Behavior and x represents Attitude.

Therefore, it can be concluded that Hypothesis 3 can be accepted. The Attitude dimension has a significant impact on the Behavior dimension. The results of the hypothesis test as shown in Fig. 3.

TABLE VII. COEFFICIENT OF DETERMINATION K-B

Model Summary				
Model	R	R-squared	Adjusted R-squared	Std. Error of the Estimate
1	0.685	0.470	0.466	5.943
Predictors: (Constant), K				

TABLE VIII. REGRESSION COEFFICIENTS K-B

	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	15.008	4.845		3.098	0.002
K	0.767	0.072	0.685	10.650	0.000

TABLE IX. COEFFICIENT OF DETERMINATION A-B

Model Summary				
Model	R	R-squared	Adjusted R-squared	Std. Error of the Estimate
1	0.807	0.652	0.649	4.816
Predictors: (Constant), A				

TABLE X. REGRESSION COEFFICIENTS A-B

	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	10.540	3.627		2.906	0.004
A	0.855	0.055	0.807	15.480	0.000

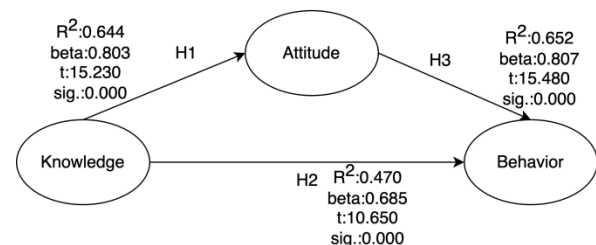


Fig. 3. The results of the hypothesis test.

D. Results of the Information Security Awareness Level Measurement

Table XI presents the findings of the information security awareness level measurement conducted at PT ABC.

The password management, email use, and social media use areas have all met the Good category with regards to the knowledge, attitude, and behavior dimensions. This suggests that employees have a high level of understanding, positive attitude, and appropriate conduct when it comes to password management, email use, and social media use. The results for

the internet use area have been classified under the moderate category, which requires significant attention. As highlighted in the introduction, the internet represents a vulnerable entry point for cyber attacks, including phishing, viruses, and data leakage [8]. Consequently, ensuring that employees exhibit wise and safe internet practices is essential to safeguard the company's information security. Given the moderate category classification, it is imperative for PT ABC to implement measures to improve information security awareness and enhance employee knowledge and behavior in this area.

The mobile device usage area has attained the good category concerning the knowledge and attitude aspects. However, with regards to the behavioral aspect, it is still classified in the moderate category.

In relation to information handling area and incident reporting, both knowledge and behavioral dimensions have achieved the good category; however, in terms of the attitude dimension, it is categorized as moderate.

The assessment of asset management area reveals that it has achieved the Good category with regards to the attitude and behavioral dimensions. However, in terms of knowledge, it has only reached the Moderate category, indicating a need for further improvement.

TABLE XI. RESULTS OF THE INFORMATION SECURITY AWARENESS LEVEL MEASUREMENT

Area/Dimension	Knowledge	Attitude	Behavior
Password management	89.54	83.15	88.92
Email use	83.54	85.69	81.46
Internet use	70.54	75.62	76.00
Social media use	87.23	85.00	81.08
Mobile devices	83.23	82.23	79.92
Information handling	89.62	79.62	87.69
Incident Reporting	85.15	75.31	83.54
IS/IT asset management	79.77	85.69	84.46

Indicators with moderate measurement results can be a priority improvement area that can be enhanced by the company.

1) Behavior dimension based on job tenure category: Table XII shows a significant difference in behavior levels based on the job tenure category in the focus area of email usage, where employees with working experiences of less than five years (mean rank 73.07) have a high value of behavior compared to employees with job tenure 6 to 10 years (mean rank 49.43). Within the information handling focus area, the highest value is in employees with more than 15 years of job tenure (mean rank 77.04) and the lowest is in employees with length of employment 5 to 10 years (mean rank 53.48). Within the incident reporting focus area, the highest value is in employees with more than 15 years of job tenure (mean rank 86.88; mean rank 82.43) and the lowest is in employees with job tenure of less than 10 years (mean rank 58.06; mean rank 58.70).

2) Behavior Dimension based on education level: Table XIII shows a significant difference in the habit of using social media; the lowest is in employees with a Master's degree background and the highest is in employees with a High School education background. The highest incident reporting habit is in employees with a SMA (High School) education background and the lowest is in employees with a Master's degree background.

TABLE XII. BEHAVIOR BASED ON JOB TENURE

Indicator		≤ 5 yr (mean rank)	6 – 10 yr (mean rank)	11 – 15 yr (mean rank)	> 15 yr (mean rank)	Sig*
Password management	MP1	64.31	59.53	69.90	68.33	0.733
	MP2	65.12	59.38	65.75	70.26	0.691
Email use	EM1	73.07	49.43	60.96	65.47	0.052
	EM2	63.58	59.98	64.58	73.84	0.528
Internet use	IN1	62.00	60.15	66.25	75.45	0.372
	IN2	66.64	54.48	57.58	77.41	0.075
Social media use	MS1	62.39	69.08	54.21	78.48	0.083
	MS2	68.34	51.33	61.33	73.14	0.119
Mobile devices	PM1	59.61	64.58	69.19	74.66	0.297
	PM2	60.62	64.15	65.35	76.14	0.286
Information handling	PF1	61.28	53.48	70.88	77.64	0.041
	PF2	66.93	49.80	63.83	74.90	0.087
Incident reporting	PD1	58.06	59.60	62.25	86.88	0.003
	PD2	61.06	58.70	61.25	82.43	0.028
IS/IT assets management	MA1	60.66	58.05	68.00	78.09	0.084
	MA2	68.38	50.50	58.94	75.62	0.067

*Kruskal-Wallis test

TABLE XIII. BEHAVIOR BASED ON EDUCATION LEVEL

Indicator		SMA (mean rank)	D3 (mean rank)	D4/S1 (mean rank)	S2 (mean rank)	Sig*
Password management	MP1	68.48	60.78	65.03	77.50	0.812
	MP2	65.95	80.56	62.69	45.83	0.109
Email use	EM1	56.12	71.28	67.84	59.00	0.360
	EM2	62.40	76.00	64.45	60.50	0.597
Internet use	IN1	63.02	74.67	65.18	43.00	0.489
	IN2	73.17	69.25	62.68	44.00	0.337
Social media use	MS1	76.17	76.08	60.47	33.00	0.046
	MS2	66.97	76.42	61.44	94.00	0.151
Mobile devices	PM1	72.16	73.78	61.67	53.67	0.362
	PM2	62.34	71.86	64.94	72.67	0.809
Information handling	PF1	73.64	70.75	61.12	72.17	0.284
	PF2	61.03	77.14	64.70	60.17	0.441
Incident reporting	PD1	83.69	71.08	58.03	55.50	0.007
	PD2	77.24	74.86	59.06	67.67	0.057
IS/IT assets management	MA1	64.41	77.81	62.84	73.17	0.356
	MA2	76.09	69.42	60.96	60.67	0.243

*Kruskal-Wallis test

TABLE XIV. BEHAVIOR BASED ON JOB POSITION

Indicator		Structural (mean rank)	Functional (mean rank)	Sig*
Password management	MP1	65.67	65.00	0.922
	MP2	64.04	69.79	0.360
Email use	EM1	67.41	59.88	0.272
	EM2	64.97	67.06	0.773
Internet use	IN1	62.62	73.95	0.123
	IN2	63.60	71.08	0.280
Social media use	MS1	63.78	70.55	0.354
	MS2	65.50	65.50	1.000
Mobile devices	PM1	63.27	72.06	0.224
	PM2	62.16	75.30	0.064
Information handling	PF1	63.46	71.48	0.225
	PF2	64.03	69.82	0.401
Incident reporting	PD1	61.84	76.26	0.041
	PD2	62.40	74.61	0.078
IS/IT assets management	MA1	63.62	71.03	0.264
	MA2	68.12	57.80	0.149

*Kruskal-Wallis test

3) Behavior dimension based on job position category:

Table XIV shows the behavior level based on job categories. The focus area with a significant difference value less than 0.05 is the incident reporting focus area. In the incident reporting focus area, the highest habit value is in functional employees (mean rank 76.26) and the lowest is in structural employees (mean rank 61.84).

E. Implications

Based on the research results, the following are the practical and theoretical implications:

1) *Practical implication:* The measurement of information security awareness at PT ABC is presented in Table XI. The results indicate that certain focus areas do not meet the Good category, which suggests that PT ABC should take necessary measures, such as training or actions, to enhance information security awareness. This can serve as a guide for the company to identify the specific areas that require improvement and facilitate the implementation of targeted interventions. Utilizing training programs can be an effective method for improving knowledge and awareness of potential security threats and risks [19]. The focus areas and dimensions that need to be improved are:

- Internet use

It is recommended that PT ABC offer training, socialization, and seminars to employees as a means of enhancing their knowledge and awareness of information security. Additionally, it is advised that the company establish regulations, actions, and punishments against employees who engage in behavior or activities that threaten the security of the company's information system within the internet use focus area. What is included in the internet use focus area such as downloading files carelessly to office devices, accessing

suspicious online sites, and entering information into online sites to assist work.

- Mobile devices

PT ABC is authorized to issue warnings, prescribe regulations, and take appropriate actions against employees whose practices may put the company's information system security at risk within the software device use focus area. What is included in the mobile device use focus area such as physical security of devices such as leaving laptops/mobile phones carelessly, sending sensitive information via online networks, and opening sensitive documents near strangers.

- Information handling

PT ABC has the authority to provide notifications, implement regulations, and execute appropriate measures towards employees whose conduct may jeopardize the security of the organization's information system within the information handling focus area. What is included in the information handling focus area such as putting/throwing sensitive documents carelessly, and inserting USB/removable media into office PC/laptops carelessly.

- Incident reporting

PT ABC has the authority to issue warnings, set rules, and take appropriate actions against employees who exhibit behaviors that pose a threat to the security of the company's information system within the incident reporting focus area. What is included in the incident reporting focus area such as reporting suspicious behavior in the office, reporting if there is a dangerous action from a colleague related to the information system security, reporting all incidents/events related to the information system security.

- IS/IT Assets Management

PT ABC can provide training/socialization/seminars to employees to improve their knowledge of asset management such as carelessly installing software, especially pirated software and related data backups.

Table XII - XIV shows the results of the measurement of the dimension of the habits based on the categories of work experience, education, and job type. This categorization can provide insight to the company to identify the strengths and weaknesses of awareness of information security and also facilitate the development of a customized information security training program for employees [20].

Based on the table, PT ABC can provide rules/actions in the more focused areas of information security system based on categories.

If based on work experience category, it is as follows:

- Within the focus area of email use, employees who have worked for over 5 years demonstrate inadequate knowledge of information security systems. Therefore, there is a pressing need to prioritize information security awareness-raising efforts in the email use sector for this category of employees.

- Within the focus area of information handling, employees who have served for 6-10 years exhibit reduced knowledge of information security systems. Consequently, there is a necessity to prioritize awareness-raising efforts on information security within the information handling sector for this group of employees.
- In the incident reporting focus area, employees who have worked for less than 11 years have a lower awareness of information security systems. Consequently, there is a need to focus more on enhancing information security awareness within the incident reporting sector.

Based on education, the focus of increasing awareness of information security can be more focused on the use of social media and incident reporting areas. The details are as follows:

- In the focus area of social media use, employees with S2 education have a lower awareness of information security systems; hence, it is imperative to give more consideration to the awareness of information security within the social media usage domain.
- In the incident reporting focus area, employees with D4/S1 and S2 education have a lower awareness of information security systems, therefore the incident reporting sector requires greater emphasis on information security awareness.

Then, based on job type, PT ABC can focus more on incident reporting on structural employees.

2) Theoretical implications

- In this research, it can be concluded that knowledge has a significant impact on attitudes and habits. Furthermore, the attitude dimension also has a significant effect on behavior.
- Measurement based on respondent category can be done to categorize participants in training/seminar/awareness-raising activities on information security.

VI. CONCLUSION

Based on research results, the overall average level of information security awareness among employees at PT ABC is considered good. The findings also highlight specific areas that require improvement to increase awareness about information security, such as internet usage, information handling, asset management, incident reporting, and the use of mobile devices.

The research also showed that the dimension of knowledge influences attitude and behavior, so the information security awareness improvement program can focus on increasing the dimension of knowledge, such as socialization, seminars, and training, as well as punishment systems or monitoring if it focuses on the attitude dimension.

REFERENCES

[1] "Peraturan Presiden Nomor 82 Tahun 2022 tentang Perlindungan Infrastruktur Informasi Vital (IIV)," 2022.

[2] BSSN, "Laporan Tahunan Monitoring Keamanan Siber 2021," 2021.

[3] F. Salahdine and N. Kaabouch, "Social engineering attacks: A survey," *Future Internet*, vol. 11, no. 4. MDPI AG, 2019. doi: 10.3390/FI11040089.

[4] I. Ghafir et al., "Security threats to critical infrastructure: the human factor," *Journal of Supercomputing*, vol. 74, no. 10, pp. 4986–5002, Oct. 2018, doi: 10.1007/s11227-018-2337-2.

[5] V. A. Effendy, Y. Ruldeviyani, M. M. Rifa'i, V. A. Rahmatika, W. Nur'aini, and Y. P. Sagala, "Measurement of Employee Information Security Awareness on Data Security: A Case Study at XYZ Polytechnic," in *2022 1st International Conference on Information System and Information Technology, ICISIT 2022*, 2022, pp. 272–276. doi: 10.1109/ICISIT54091.2022.9873077.

[6] A. Zulfia, R. Adawiyah, A. N. Hidayanto, and N. F. A. Budi, "Measurement of Employee Information Security Awareness Using the Human Aspects of Information Security Questionnaire (HAIS-Q): Case Study at PT. PQS; Measurement of Employee Information Security Awareness Using the Human Aspects of Information Security Questionnaire (HAIS-Q): Case Study at PT. PQS," 2019.

[7] R. Prakoso, Y. Ruldeviyani, K. F. Arisya, and A. L. Fadhilah, "Measurement of Information Security Awareness Level: A Case Study of Online Transportation Users," in *2020 3rd International Seminar on Research of Information Technology and Intelligent Systems, ISRITI 2020*, Dec. 2020, pp. 170–175. doi: 10.1109/ISRITI51436.2020.9315375.

[8] E. Kritzinger, A. da Veiga, and W. van Staden, "Measuring organizational information security awareness in South Africa," *Information Security Journal*, 2022, doi: 10.1080/19393555.2022.2077265.

[9] L. Hadlington, J. Binder, and N. Stanulewicz, "Exploring role of moral disengagement and counterproductive work behaviours in information security awareness.," *Comput Human Behav*, vol. 114, Jan. 2021, doi: 10.1016/j.chb.2020.106557.

[10] Nurbojatmiko, A. Fajar Firmansyah, Q. Aini, A. Saehudin, and S. Amsariah, "Information Security Awareness of Students on Academic Information System Using Kruger Approach," in *2020 8th International Conference on Cyber and IT Service Management, CITSM 2020*, Oct. 2020. doi: 10.1109/CITSM50537.2020.9268795.

[11] Y. Salem, M. Moreb, and K. S. Rabayah, "Evaluation of Information Security Awareness among Palestinian Learners," in *2021 International Conference on Information Technology, ICIT 2021 - Proceedings*, Jul. 2021, pp. 21–26. doi: 10.1109/ICIT52682.2021.9491639.

[12] K. Parsons, D. Calic, M. Pattinson, M. Butavicius, A. McCormac, and T. Zwaans, "The Human Aspects of Information Security Questionnaire (HAIS-Q): Two further validation studies," *Comput Secur*, vol. 66, pp. 40–51, May 2017, doi: 10.1016/j.cose.2017.01.004.

[13] P. G. Schrader and K. A. Lawless, "The knowledge, attitudes, & behaviors approach how to evaluate performance and learning in complex environments," *Performance Improvement*, vol. 43, no. 9, pp. 8–15, Sep. 2004, doi: 10.1002/pfi.4140430905.

[14] D. Fujs, S. Vrhovc, and D. Vavpotic, "Know Your Enemy: User Segmentation Based on Human Aspects of Information Security," *IEEE Access*, vol. 9, pp. 157306–157315, 2021, doi: 10.1109/ACCESS.2021.3130013.

[15] "SNI ISO/IEC 27001:2013 Standar Nasional Indonesia Badan Standardisasi Nasional," 2016. [Online]. Available: www.bsn.go.id.

[16] S. Dowdy, S. Weardon, and D. Chilko, *Statistics for Research*, Third. Hoboken, New Jersey, USA: John Wiley & Sons, Inc., 2004.

[17] J. F. Hair, G. T. M. Hult, C. M. Ringle, and Marko. Sarstedt, *A primer on partial least squares structural equation modeling (PLS-SEM)*. 2007.

[18] H. A. Kruger and W. D. Kearney, "A prototype for assessing information security awareness," *Comput Secur*, vol. 25, no. 4, pp. 289–296, Jun. 2006, doi: 10.1016/j.cose.2006.02.008.

[19] K. Khando, S. Gao, S. M. Islam, and A. Salman, "Enhancing employees information security awareness in private and public organisations: A systematic literature review," *Comput Secur*, vol. 106, Jul. 2021, doi: 10.1016/j.cose.2021.102267.

[20] A. McCormac, T. Zwaans, K. Parsons, D. Calic, M. Butavicius, and M. Pattinson, "Individual differences and Information Security Awareness,"

Comput Human Behav. vol. 69, pp. 151–156, Apr. 2017, doi:
10.1016/j.chb.2016.11.065.

[21] Al-Shanfari I, Warusia Yassin, Nasser Tabook, Roesnita Ismail, Anuar
Ismail, "Determinants of Information Security Awareness and

Behaviour Strategies in Public Sector Organizations among Employees,"
(IJACSA) International Journal of Advanced Computer Science and
Applications, Vol. 13, No. 8, 2022.

APPENDIX

TABLE XV. INSTRUMENT VALIDITY TEST RESULTS

Indikator	Pearson Correlation	Sig. (2-tailed)	N	Indikator	Pearson Correlation	Sig. (2-tailed)	N
KMP1	0.374	0.000	130	APM1	0.555	0.000	130
KMP2	0.402	0.000	130	APM2	0.631	0.000	130
KEM1	0.285	0.001	130	APF1	0.234	0.007	130
KEM2	0.290	0.001	130	APF2	0.550	0.000	130
KIN1	0.552	0.000	130	APD1	0.512	0.000	130
KIN2	0.562	0.000	130	APD2	0.108	0.220	130
KMS1	0.456	0.000	130	AMA1	0.590	0.000	130
KMS2	0.390	0.000	130	AMA2	0.398	0.000	130
KPM1	0.416	0.000	130	BMP01	0.506	0.000	130
KPM2	0.486	0.000	130	BMP02	0.615	0.000	130
KPF1	0.370	0.000	130	BEM1	0.447	0.000	130
KPF2	0.521	0.000	130	BEM2	0.437	0.000	130
KPD1	0.318	0.000	130	BIN1	0.627	0.000	130
KPD2	0.492	0.000	130	BIN2	0.513	0.000	130
KMA1	0.563	0.000	130	BMS1	0.382	0.000	130
KMA2	0.213	0.015	130	BMS2	0.359	0.000	130
AMP1	0.491	0.000	130	BPM1	0.539	0.000	130
AMP2	0.461	0.000	130	BPM2	0.399	0.000	130
AEM1	0.539	0.000	130	BPF1	0.599	0.000	130
AEM2	0.565	0.000	130	BPF2	0.549	0.000	130
AIN1	0.568	0.000	130	BPD1	0.583	0.000	130
AIN2	0.596	0.000	130	BPD2	0.591	0.000	130
AMS1	0.467	0.000	130	BMA1	0.619	0.000	130
AMS2	0.514	0.000	130	BMA2	0.397	0.000	130

ECAH: A New Energy-Aware Coverage Method for Wireless Sensor Networks using Artificial Bee Colony and Harmony Search

ZHOU Bing^{1*}, ZHANG Zhigang²

College of Artificial Intelligence
Jiaozuo University, Jiaozuo, Henan, 454000, China

Abstract—Wireless Sensor Networks (WSNs) offer diverse applications in the research and commercial fields, such as military applications, medical science, waste management, home automation, habitat monitoring, and environmental observation. WSNs are generally composed of a large number of low-cost, low-power, and multifunctional sensor nodes that sense, process, and communicate data. These nodes are connected by a wireless medium, allowing them to collect and share data with each other. To achieve network coverage in a WSN, a few to thousands of tiny and low-power sensor nodes should be placed in an interconnected manner. Over the last decade, deploying sensor nodes in a WSN to cover a large area has received much attention. Coverage, regarded as an NP-hard problem, is an essential parameter for WSNs that determines how the deployed sensor nodes handle each point of interest. Various algorithms have been proposed to tackle this problem. However, they often come with a trade-off between energy efficiency and coverage rate. Moreover, the scalability of the algorithms needs to be considered for large-scale networks. This paper proposes a novel energy-aware method combining Artificial Bee Colony (ABC) and Harmony Search (HS) algorithms to address the coverage problem in WSN, called ECAH. The proposed ECAH algorithm has been tested with various network scenarios and compared with other existing algorithms. The results show that ECAH outperforms the existing methods in terms of network lifetime, coverage rate, and energy consumption. Additionally, the proposed algorithm is also more robust and efficient as it can adjust to dynamic network environment changes, making it suitable for various network scenarios.

Keywords—Artificial bee colony; coverage; deployment; harmony search; wireless sensor network

I. INTRODUCTION

Recent years have seen a significant advancement in wireless and emerging technologies, particularly the Internet of Things (IoT) [1], Wireless Sensor Networks (WSNs) [2], artificial intelligence [3, 4], machine learning [5-7], smart grids [8], Blockchain [9, 10], 5G connectivity [11], and cloud computing have successfully brought many benefits to society.

WSNs are composed of economical, energy-efficient, and tiny sensor nodes which are widely employed in many applications [12-14]. For instance, in a military application, the WSN can be utilized for battlefield surveillance and asset monitoring [15]. Environmental monitoring, building monitoring, forest fire detection, and natural disaster prevention are examples of civil applications of WSNs [16].

The sensors have three prominent roles; sensing, computing, and wireless communication [14]. The sensors can shift separately after a random placement indicating the sensors' self-deploys. Some sensors have the self-repair ability where nodes can reposition themselves when another node fails [17]. Some environmental information, such as temperature, humidity, or other environmental data, is measured by each node periodically. This information should be sent to a central node named sink [18].

Coverage problems pose a major challenge to the development of effective sensor networks. In a sensor network, the coverage problem serves as an indicator of QoS [19]. The coverage is classified into three groups such as area, point, and barrier coverage. The WSNs aim to cover and manage an environment completely. There must be at least one sensor placed at each point in the area. In the area coverage, the best coverage occurs when the target area is fully covered by the lowest number of sensors [20]. The point coverage aims to cover the specific points of the environment. Sensors do not cover the whole area in this model; just some areas are covered. So, some particular points specify goals that need to be controlled. In barrier coverage, the sensors do not cover the entire area. At least one sensor controls the goals in barrier coverage, and all existing plans are controlled by that area. Maintaining connectivity and coverage requires finding the optimal number of active nodes [21].

Coverage in WSNs is an NP-hard problem, and many heuristic methods are proposed to increase the coverage rate in these networks. One of the developed metaheuristic algorithms is Harmony Search (HS). The HS algorithm has fewer mathematical requirements compared to other metaheuristic algorithms. With the HS algorithm, efficient regions are found within a reasonable timeframe. However, it suffers from non-convergence to an optimal solution, low convergence accuracy, and more time to reach optimal response. It is possible to improve accuracy and convergence rates by combining HS with the Artificial Bee Colony (ABC) algorithm. The ABC algorithm performs better than other evolutionary algorithms and optimizes harmony memory with its variants. This paper presents an algorithm combining HS and ABC algorithms to resolve the coverage optimization problem in WSNs. Briefly, the paper makes the following contributions:

- Maximizing the coverage rate in WSN;

- Minimizing the energy consumption in the WSN;
- Increasing the lifetime of the WSN.

The paper is organized in the following manner. The related work is discussed in Section II. The proposed hybrid algorithm is presented in Section III. Section IV reports the experimental results. The conclusion is discussed in the last part.

II. RELATED WORK

Many meta-heuristic algorithms have been adopted by researchers in recent years in order to improve the performance of WSNs, with various effects on the coverage area, network lifetime, the routing strategy, and node distribution. Each algorithm has its own advantages and drawbacks, and selecting the most suitable algorithm for a given application is an important step for successful deployment of WSNs. Therefore, it is important to understand the characteristics of each algorithm in order to make an informed decision. Additionally, many algorithms can be combined to further improve network performance. Finally, it is important to continuously evaluate and monitor the performance of a WSN to ensure optimal performance.

Ab Aziz, et al. [22] have utilized the PSO algorithm and Voronoi diagrams to optimize the coverage problem in the WSNs. In this work, the target area is a two-dimensional square environment, considered a convenient and practical model that can be used for different applications. The sensors are placed in the region of interest using the Voronoi diagram instead of the grid and PSO algorithm. In this work, the Voronoi diagram is produced initially on an unbounded area considering the position of sensors in a specific region of interest. This way, each boundary is covered with a set of randomly selected points. The proposed method performs well regardless of the number of sensors used or the size of the region of interest. On the other hand, there are some disadvantages to using the Voronoi diagram. One disadvantage is that it can be more difficult to generate the diagram for more complex shapes. Another disadvantage is that the diagram can be sensitive to changes in the position of the sensors, which can lead to inaccurate results.

Also, Panov and Koceski [23] have solved the area coverage problem in WSNs using the HS algorithm. The HS algorithm determines the location and number of the optimal sensor nodes in this study. Therefore, the network coverage and cost can be improved. To start the HS algorithm, harmony memory stores an initial population of solution vectors. Each vector represents a sensor node location denoted by its real number. Based on the obtained results, the HS algorithm requires more iterations than the PSO algorithm. Thus, HS achieves faster convergence and rapid evaluation than other heuristic methods. However, the HS algorithm also has some drawbacks. One such drawback is that the HS algorithm can be easily trapped in a local optimum. This is due to the fact that the HS algorithm relies on the harmony memory to generate new solutions. Thus, the HS algorithm is not as effective as other heuristic methods in terms of escaping local optima.

In addition, Fidanova and Marinov [24] developed an algorithm based on the ACO algorithms for addressing the coverage issue in WSNs, focusing on reducing nodes and increasing coverage. Pheromone trails are bound by a fixed upper and lower bound in the proposed method. Thus, part of the possible movements and repetition of the similar and identical solution will prevent the accumulation of high pheromones. Their method implementation utilizes two main features of the ACO algorithm, an in-depth exploration of space in search of the best solution and extensive research into the best solution. Compared to simulated annealing and cataclysmic mutation algorithms, the ACO optimization algorithm needs fewer evaluations to achieve good results. In addition, the ACO algorithm reaches similar or higher fitness values with 2000 times less effort than simulated annealing or cataclysmic mutation.

Maleki, et al. [25] have suggested a new hybrid method based on the differential evolution (DE) and PSO algorithms for the distribution of the sensors for area coverage in WSNs. The hybrid method considers two factors, suitable distribution of the sensors and their energy consumption. In the hybrid method, the most optimized points for n sensors are searched according to the PSO algorithm for covering p -aimed points. Searching takes place based on the particles near the goal and can cover the ground suitably, and then the mutation and crossover operations on n sensors take place using the DE algorithm. The hybrid method increases the network's lifetime, reduces energy use and covers all points optimally. The authors have compared the proposed hybrid method with the PSO algorithm to demonstrate its efficiency. Results prove that the hybrid algorithm has a better coverage ratio and network lifetime than the PSO algorithm.

To obtain a specific coverage ratio, Qin and Chen [26] developed an area coverage algorithm based on differential evolution. WSNs are monitored to maximize their lifetime using the proposed algorithm. To realize the optimization process by WSNs, coverage of continuous areas is translated into coverage of discrete points. The main objective is to maintain coverage performance while consuming minimal energy. Binary differential evolution is redeveloped in the area coverage algorithm to find an improved node subset and satisfy coverage requirements. It has been demonstrated that the differential evolution area coverage algorithm provides 90% network coverage and is energy efficient and computation efficient.

He, et al. [27] have developed a model for optimizing WSN coverage using an improved marine predator algorithm. The algorithm improves its solution accuracy by introducing a dynamic inertia weight adjustment approach to the exploration and exploitation stages of the global exploration and local exploitation phases. A multiple random leading method is also employed in the improved algorithm to improve information exchange between population members and to jump out of local optimums. Wu, et al. [28] studied a sensor-based angle coverage judgment method. To monitor the planting region from k angles, using topological analysis, a multi-objective optimization function is developed based on the relationship between targets and sensors. In experiments, judgement was found to be more efficient than the other methods.

According to the above discussion, the majority of existing research focuses on reducing energy consumption or improving network coverage. In the proposed method, both enhancements to the lifetime of the networks and coverage issues are addressed in order to maintain a balance between the amount of energy consumed and the total area covered by the sensors in the sensing region. The proposed method optimizes the deployment of sensors in order to maximize the coverage while minimizing energy consumption. It also addresses the problem of network heterogeneity and scalability, making it an attractive choice for large-scale WSNs. In the first step, the optimal position of the sensors is calculated. The radius of the sensing region required for maximum coverage is then determined. In the next step, the hybrid algorithm is used to place the nodes in their optimal positions. The network lifetime is then improved by optimizing the transmission range of the nodes. Finally, the coverage and lifetime of the network are evaluated and the performance is compared with other algorithms.

III. PROPOSED METHOD

The sensor position has the main factor in the coverage problems. Sensors should be placed to ensure the total usage of the sensing ability, increasing the covered area. On the other hand, optimal coverage increases network lifetime and also reduces network power consumption.

A. Network Model and Assumptions

The problem is premised on the following assumptions:

- All sensors have a similar sensing radius;
- The nodes are mobile and homogeneous;
- The deployment strategy is a random type;
- The sensing model is a probabilistic model;
- All sensors cover an area of two dimensions.

According to the problem assumptions, the sensing area contains N randomly distributed nodes. The density of all nodes is homogeneous and uniform. All nodes are capable of performing sensing, receiving, and transmitting. For each sensor, three radius types of the same size are considered, sensing radius represented as R_s , communication radius defined as R_c , and uncertainty sensor detection radius illustrated as R_e . The network model of the suggested method is shown in Fig. 1.

B. Coverage Problem Definition

The optimum position of the network sensors in the area is considered the main challenge in WSN coverage [29]. After the sensors are positioned for the first time in the static version, they cannot move in the network, which is a limitation for this version of sensor deployment [30]. However, the sensors in dynamic networks can move along coordinates within the mission space. Sensors gather data around an area that falls under their detection ranges (like their location) and then share that information with their neighbors. The communication of sensors with each other and sharing of their information causes effective detection in a network [31]. Therefore, in the coverage problem, the positional flexibility

of the mobile sensors is used to enhance the ratio of the covered area. Generally, the coverage ratio of the WSN is computed via Eq. 1 [31], in which t denotes the overall size of the area of interest, s signifies the set of the nodes, and c_i refers to the coverage of a sensor i .

$$CR = \frac{\sum c_i}{t} \quad i \in s \quad (1)$$

a) Sensing Models

Searching the adequate coverage and monitoring ability depends on the sensing model. Two kinds of sensing models, such as binary and probability sensing models, are mainly applied in the study of WSN [32].

Binary sensing model: An area with radius R_s is the coverage area of sensor nodes in a two-dimensional plane. This circular area is called the sensing disk. R_s is called the sensing radius, which is determined by the physical properties of the sensing unit. Consider the coordinates of the node S_i are $S(X_{si}, Y_{si})$ in the binary sensing model, for any point P at (x, y) on the plane, the probability of that node S_i can detect the events that occur on P calculated by Eq. 2. Where $d(S_i, P)$ stands for the average Euclidean distance from the sensor node to the demand point in the plane, calculated by Eq. 3.

$$C_{xy}(S_i) = \begin{cases} 1 & x < 0 \\ 0 & otherwise \end{cases} \quad (2)$$

$$d(S_i, P) = \sqrt{(x_{si} - x)^2 + (y_{si} - y)^2} \quad (3)$$

Probability sensing model: The binary sensing model enables the detection of events. Nevertheless, the sensor node detection capability is unstable in virtual application environments due to interference from ambient noise and a decrease in signal strength. Implementing the probability sensing model in real-world environments is more practical since, instead of considering the sensor's radius and distance from the target, multiple dependent factors must be considered. Fig. 2 illustrates the probability sensing model. Probability is determined by Eq. 4 [32].

$$C_{xy}(S_i) = \begin{cases} 1 & d(S_i, p) \leq R_s - R_e \\ e^{-(\alpha_1 \lambda_1 \beta_1) / (\lambda_2 \beta_2 + \alpha_2)} & R_s - R_e < d(S_i, p) < R_s + R_e \\ 0 & d(S_i, p) \geq R_s + R_e \end{cases} \quad (4)$$

In Eq. 4, R_e ($0 < R_e < R_s$) refers to the uncertainty in the sensor detection radius. Sensor node characteristics are represented by $\alpha_1, \alpha_2, \beta_1, \beta_2, \lambda_1$, and λ_2 , where λ_1 and λ_2 are expressed as Eq. 5 and Eq. 6.

$$\lambda_1 = R_e - R_s + d(s_i, i) \quad (5)$$

$$\lambda_2 = R_e + R_s + d(s_i, i) \quad (6)$$

Due to overlapping sensors of sensing areas, this type of sensing model compensates for low detection probability potential. Eq. 7 determines this joint sensing probability when a demand point i lies in the overlap area of a set of sensors (s_{ov}).

$$C_{xy}(k_{sov}) = 1 - \prod_{s_i \in s_{ov}} (1 - C_{xy}(s_i)) \quad (7)$$

In Eq. 7, if $C_{xy}(k_{sov}) > C_{th}$, demand point p is considered to be effectively covered, where C_{th} refers to the coverage threshold, and its value is application dependent. The proposed method uses a probabilistic sensor detection model that includes points of different probabilities in the area. In this model, the area coverage is determined using probabilistic terms, so the results are realistic.

b) Coverage Model

Eq. 8 calculates the area coverage rate of the node set, where, $R_{area}(k_{sov})$ stands for the total ratio of the coverage area and $m \times n$ is the size of the sensing area.

$$R_{area}(k_{sov}) = \frac{\sum_{x=1}^m \sum_{y=1}^n C_{xy}(k_{sov})}{m \times n} \quad (8)$$

c) Energy Consumption Model

Energy efficiency is a very important issue in WSNs due to the limitation of battery resources. We need mechanisms that keep the energy resources because these mechanisms affect the network lifetime. In general, there are two mechanisms for conserving energies in a WSN, reducing the sensing range and planning of the sensor node activity. In cases where setting the sensing range is permissible, active sensors should dynamically adjust their sensing ranges so as to meet the entire sensing objective [33]. Energy savings can be achieved by reducing the communication range when an option to adjust the range of sensor communication radius would exist. In the proposed method, the introduced energy model in [33] is used to calculate the energy. According to the model, the consumption energy of the sensors is calculated via Eq. 9-11, where l denotes the length of the data transmitted, d_0 signifies threshold transmission distance, d represents the distance between sensor nodes, $E_{T,x}$ refers to the consumed energy for transmission, $E_{R,x}$ stands for the consumed energy for reception, E_{elec} specifies the amount of dissipated energy per bit to run the transmitter or the receiver circuit, and E_{Fs} and E_{Tr} depend on the transmitter amplifier model. The value of E_{elec} , E_{Fs} , and E_{Tr} are $E_{elec} = 50nJ/bit$, $E_{Fs} = 10pJ \times bit^{-1} \times m^{-2}$, $E_{Tr} = 0.0013pJ \times bit^{-1} \times m^{-4}$, and d = transmission radius or communication range.

$$E_{T,x}(l, d) = l \times E_{elec} + l \times E_{Fs} \times d^2 \text{ .if } d < d_0 \quad (9)$$

$$E_{T,x}(l, d) = l \times E_{elec} + l \times E_{Tr} \times d^4 \text{ .if } d > d_0 \quad (10)$$

$$E_{R,x}(l, d) = l \times E_{elec} \quad (11)$$

d) Lifetime Model

The length of time between the initial deployment of the network and running out of the energy of the first relay node defines the network lifetime. In this work, lifetime is expressed according to the seconds, and for a single node, it can be estimated with Eq. 12 [33], where $e_{initial}$ is the initial energy of sensor nodes and e_{total} refers to the amount of consumed energy for data transmission and reception calculated by Eq. 13.

$$\text{Lifetime} = \frac{e_{initial}}{e_{total}} \quad (12)$$

$$e_{total} = E_{T,x}(l, d) + E_{R,x}(l, d) \quad (13)$$

e) Objective Function

Objective functions are one of the most vital factors in optimization algorithms. According to the energy and lifetime models, both of these models depend on the distance between the sensors. Also, according to the network coverage definition, the network becomes an optimal coverage when the distance between the sensors is less than their sensing radius. On the other hand, as the coverage rate increases, more sensors will be active, and reducing the number of active sensors will result in a lower coverage rate. The problem is mathematically described as follows: given a set of N potential sensors, $k = \{s_1, s_2, s_3, s_4 \dots s_N\}$. In this regard, the optimization goal is calculated as follows.

$$z = \max(f_1, f_2) + \min(f_3) \quad (14)$$

In Eq. 14, f_1 , f_2 , and f_3 stand for the coverage rate, network lifetime, energy consumption, respectively. The fitness function is calculated using Eq. 15, where w_1, w_2 and w_3 are weight coefficients used to normalize coverage, energy, and network lifetime values.

$$\text{Fitness} = w_1 \left(\frac{\sum_{x=1}^m \sum_{y=1}^n C_{xy}(k_{sov})}{m \times n} \right) + w_2 \left(\frac{1}{(l \times E_{elec} + l \times E_{Fs} \times d^2) + (l \times E_{elec})} \right) + w_3 \left(\frac{e_{initial}}{(l \times E_{elec} + l \times E_{Fs} \times d^2) + (l \times E_{elec})} \right) \quad (15)$$

C. Proposed Coverage Algorithm

Due to the NP-Hard nature of the area coverage problem in WSN, the optimal approximate solution could be suitable. The traditional greedy algorithm is inefficient in solving the coverage problem, and finding the best solution is impossible. In the previous methods, the optimal results are not included in the coverage rate, energy consumption, and network lifetime. In this work, a new hybrid method is suggested based on the ABC and HS algorithms for resolving the coverage problem of the WSNs. The HS algorithm is one of the latest and the easiest methods to optimize the issues inspired by music. This algorithm has premature and slow convergence problems, particularly over the multimodal fitness landscape. Also, convergence to the optimal solution occurs slower [34].

In the present work, the ABC algorithm is used to improve the accuracy and convergence of the HS algorithm and to optimize harmony memory. Honey bees' intelligent foraging behavior is crucial in generating an ABC algorithm as a new swarm intelligence method. Harmony memory can be improved using the ABC and its variants. A significant advantage of the ABC algorithm is its ability to seek global and local results throughout each iteration. This advantage is compared to GA, PSO, and other intelligent computing methods. Therefore, finding optimal solutions increases and is avoided in the stuck of local optimization. The hybrid method uses the HS algorithm as the primary global search technique. In contrast, the ABC algorithm is the subprocedure of the local search and optimizes the harmony memory.

The coverage problem of the WSNs is proportional to the behavior of honey bees to find the optimal position of food sources and how the musician plays music. The position of the food sources represents the coordinates of the sensors in the sensing area, the value of each food source indicates the

network coverage rate, and the maximum value for the fitness function represents the highest coverage rate for each sensor. In terms of the parameters of the HS algorithm, each harmony vector represents an optimal position for the sensors. The harmony vector components also represent the decision variables, i.e., coverage rate, energy consumption, and sensor lifetime.

a) Preparing the Hybrid Method

This step consists of two sub-steps, the production of the initial population and the initialization of the harmony memory. In general, the HS algorithm has five main steps. Steps 1 and 2 are used in the first step of the proposed method, and steps 3, 4, and 5 are used to improve the newly generated solutions and update the harmony memory in the third step of the proposed method.

Generation of an initial population: This step generates the initial population (initial solutions) randomly. The initial population for the coverage problem is the coordinates of the sensors based on (x,y) in the sensing area. At this step, each sensor's coordinates (x,y) are generated, and the sensors are distributed under these coordinates in the sensing area. The size of the population should be equal to the size of the harmony memory.

Initialization of harmony memory: After generating the initial population of the sensors in steps 1-1 and distributing these sensors in the sensing area based on their x and y coordinates, a harmony vector is generated for each sensor. After generating the harmony vector for each sensor, the fitness function for each vector is calculated. Then the harmony memory is initialized with these vectors and their fitness values. Each harmony vector consists of three decision variables, coverage rate, energy consumption, and network lifetime. The harmony vector structure and harmony memory are described as follows.

- Harmony vector: Each solution in the HS algorithm is referred to as a harmony vector, representing the number of parameters of the optimization problem in a D-dimensional vector. Since the optimization parameters in the coverage problem are the coverage rate, energy consumption, and the sensors' lifetime, the harmony vector is considered a three-dimensional vector.
- Harmony memory: The HS algorithm uses a harmony memory to store possible solutions (harmony vectors) and their objective function values [35]. The number of initial populations determines harmony memory size. Eq. 16 represents harmony memory as a matrix.

$$\text{Harmony memory} = \begin{matrix} & \text{Decision variables} & & \text{Fitness function} & \\ \left[\begin{array}{cccc|c} x_1^1 & x_2^1 & x_3^1 & \dots & x_D^1 & f(x^1) \\ x_1^2 & x_2^2 & x_3^2 & \dots & x_D^2 & f(x^2) \\ x_1^{HMS} & x_2^{HMS} & x_3^{HMS} & \dots & x_D^{HMS} & f(x^{HMS}) \end{array} \right] & & & & (16) \end{matrix}$$

- Improving the harmony memory using the ABC algorithm

In this step, harmony memory is improvised by the ABC algorithm. The solution to the optimization problem is represented by the position of food sources in the ABC algorithm. A food source's nectar amount corresponds to the associated solution's fitness (quality). The employed artificial bees are located in the artificial bee colony's first half, and the onlookers are placed in the second half. Also, each food source has only one employed bee. Conforming to the above description, in the coverage problem, the location of the sensor nodes in the sensing area identifies the food sources, and each food source's value is defined as the sensor nodes' coverage value. In this step, a bee is assigned to each position of sensors in the sensing area. These sensors perform a series of movements and change their current position in the worker and onlooker bee phases. Changing the sensor's position makes the sensors find their optimal position, increasing the network's coverage. Generally, the ABC algorithm has four main phases: initialization, employed bees, onlooker bees, and scout bees' phase. The initial population of ABC algorithms is the harmony vectors generated and evaluated in step 1.

Employed bee phase: The particular group of bees is called employed bees that use the available food sources. Each of the employed bees keeps the profitability of the associated food

source and then returns to the hive and performs the waggle dance. Worker bees dance in different parts of the hive area to communicate with other bees. Generally, there are three types of dances: round, waggle, and tremble. In this phase, the new solution's fitness value (nectar amount) is essential in modifying the current solution by employing bees. When the fitness value of the new food source is higher than the old food source, the bee updates her position with the new one and rejects the old one. A position update equation for the j^{th} dimension of each i^{th} candidate in this phase is shown in Eq. 17 [34].

$$V_{ij} = X_{ij} + \varphi_{ij}(X_{ij} - X_{kj}) \quad (17)$$

In Eq. 17, $\varphi_{ij}(X_{ij} - X_{kj})$ is called step size. Although k is determined randomly ($k \neq i$), φ_{ij} is a random number between $[-1, 1]$. In the coverage problem, these working bees display the sensors. In this phase, these sensors change their current coordinates by Eq. 17. In the new position, the fitness function (objective function) is calculated, and a greedy selection is applied to the existing coordinates and old coordinates of the sensors.

Onlooker bee phase: Following the employed bee stage, the onlooker bee stage begins. In this stage, the employed bees share their fitness and position information with the onlookers in the hive [43]. Analyzing the available information, the

onlooker bees choose (roulette wheel selection method can be used) a solution with a probability P_i , related to its fitness. The probability P_i is calculated using Eq. 18 [34], where fit_i refers to the fitness value of the i^{th} solution.

$$P_i = \frac{fit_i}{\sum_{i=1}^{HMS} fit_i} \quad (18)$$

Like the employed bees, the onlooker bees change the position of their memory and check the fitness of the candidate source. Like an employed bee, the bee also memorizes the new position and forgets the old one when the new one has a higher fitness value. In the coverage problem, after the sensors have changed their current positions in the worker's bee phase and calculated the fitness value for each of their new positions, in this phase, for each of these sensors, a probability value based on Eq. 18 and the fitness value of each of the sensors is calculated. Then, using a roulette wheel selection method, a random position is selected, the sensors are positioned in the chosen position, and then the same as in the employed bee phase, by Eq. 17, the current position of sensors is changed. Then the new position is compared with the current position, and choose a better position by selecting a greedy selection.

Scout bee phase: The scout bees are generally responsible for finding new food sources around the hive. Food sources are expected to be abandoned after a specified amount of time (called "limit") if their position has not been updated [36]. These bees choose new food sources based on Eq. 19, where lb and ub are lower and higher bounds of the decision variables, respectively.

$$X_{ij} = lb_j + rand(0.1)(ub_j - lb_j) \quad (19)$$

According to the above principle, an indicator is defined as a trial index for each network sensor in the coverage problem.

$$x_i^{new} = \begin{cases} x_i(k) \in \{x_i(1), x_i(2), \dots, x_i(k)\} & r_1 > HMCR \\ x_i(k) \in \{x_i^1, x_i^2, \dots, x_i^{HMS}\} & r_1 \leq HMCR \\ x_i(k) + r_2 \times BW & r_1 \leq PAR \end{cases} \quad (20)$$

Evaluation and updating of harmony memory: Comparing the new candidate harmony and the worst harmony vector in the harmony memory leads to updating the harmony memory. New candidate vectors replace the worst harmony vectors if they are improved.

HS algorithm termination check: The termination condition in the HS algorithm depends on the number of

In the worker and onlooker bees phase, if the new position of sensors has a higher value for coverage rate, these sensors memorize the new position and forget its current position; otherwise, these sensors keep their current position, and one unit adds to the trial index of these sensors. In this phase, these trial indexes will be examined for all sensors; if the trial index value for each sensor is higher than the specified limit, it means this position does not have a good value for the coverage, and then this position should be abandoned. The sensors should create a new position based on Eq. 19.

b) Improving New Harmony Vectors and Updating Harmony Memory

At this step, a new solution (new harmony vector) generated by the three phases of the ABC algorithm was developed and stored in harmony memory. Three sub-steps are included in this step.

Improvisation of new harmony vectors: Three rules, including a Harmony Memory Consideration Rate (HMCR), a Pitch Adjustment Rate (PAR), and random numbers (r_1 and r_2), are used to improve a new harmony vector X^{new} according to the HS algorithm, in which random numbers r_1 and r_2 ranging from 0-1 are generated. Memory consideration generates the decision variable $X^{new}(k)$ if r_1 is less than HMCR; otherwise, random selection achieves $X^{new}(k)$. If r_1 is greater than HMCR, the decision variable $X^{new}(k)$ is generated randomly using Eq. 19. A slight adjustment can be made by perturbing once $x_i(k)$ and selecting one of the stored good values, with a probability of PAR. Lastly, fret width (FW or BW) determines the maximum pitch variation allowed. The new harmony vector is described in Eq. 20 [34].

decision variables, namely, sensor coverage, sensor energy consumption and sensor lifetime. In other words, the repetition loop for optimizing the new harmony vector with PAR probability is repeated three times. The first repetition is for the first component, the second repetition is for the second component, and finally, the third repetition is for the third component of the new harmony vector.

TABLE I. VARIABLES IN SIMULATION

Parameters	Definition	Value
A (m×n)	Terrain of experiment	20×20(m) - 500×500(m)
N	Number of sensor nodes	20 -210
D	Transmission radius (nodes distance)	Changeable for variant scenarios
E_{total}	Total energy for data transmission	Changeable for variant scenarios
MaxIT	Maximum iteration	Changeable for variant scenarios
HMS	Harmony memory size	Changeable for variant scenarios
R_s	Sensing radius	Changeable for variant scenarios
L	Data generated by each node	100 bit
HMCR	Harmony memory consideration rate	0.96
PAR	Pitch adjustment rate	0.68
Limit	Abandonment criteria	100
R_e	Uncertainty radius for sensor detection	$R_s/2$
$E_{initial}$	Initial energy	2 J
E_{elec}	Energy consumed by radio electronics	50nJ/bit
E_{Fs}	Energy consumed by the power amplifier	$10pj \times bit^{-1} \times m^{-2}$
E_{Tr}	Energy consumed by the power amplifier	$0.0013pj \times bit^{-1} \times m^{-4}$
V_{ij}	New position for sensors	-
X_{ij}	Current position for sensors	-
P_l	Roulette wheel selection probability	-
fit_i	New position fitness	-
r1 and r2	Random numbers between 0 and 1	-

TABLE II. FIRST SCENARIO PARAMETERS

Simulation parameters									
Sensing radius					Number of sensors		Network size		
1.5	2	2.5	3	4	5	20	20×20		
Harmony search parameters									
HMCR=0.9		PAR=0.4		FW=0.2		$\alpha_1=1$	$\alpha_2=0$	$\beta_1=1$	$\beta_2=0.5$

B. Dataset

The performance of the proposed method is compared with the performance of the HS [23], PSO [22], ACO [24], multi-objective GA [37], ABC [31], and hybrid PSO with DE [25] algorithms. Therefore, we considered the datasets and parameters used in these studies.

C. Obtained Results

This section compares our method with previous methods in six scenarios and shows its performance. In each scenario, the results are compared with those of previous methods. Lastly, the final scenario presents the effect of the number of sensors and the sensing radius on the network's coverage rate, energy, and lifetime.

In the first scenario, we used the parameters that are listed in Table II. Table III shows the results of applying our method and HS algorithm in area coverage for the different radii. The near-optimal solution for our approach is provided in fewer iteration than for the HS algorithm, as shown in Table III. The network environment for ECAH and HS algorithms with radius=2 m are shown in Fig. 3 and 4, respectively. The results demonstrate that our proposed approach can provide an optimal or near-optimal solution with faster convergence in

terms of time and number of iterations. Furthermore, our proposed method can achieve a better coverage rate in a shorter time compared to the HS algorithm.

TABLE III. RESULTS FROM APPLYING THE PROPOSED METHOD AND HS FOR AREA COVERAGE WITH DIFFERENT RADIUSSES

Radius (m)	Iteration (HS)	Iteration (ECAH)	Coverage rate (HS)	Coverage rate (ECAH)
1.5	7000	100	24.66%	53.71%
2	7000	100	45.97%	55.39%
2.5	7000	850	85.25%	77.66%
3	7000	1000	93.16%	93.34%
4	7000	100	99.50%	100%
5	7000	150	100%	100%

TABLE IV. SECOND SCENARIO PARAMETERS

Simulation parameters		
Sensing radius	Number of sensors	Network size
5	45	50×50

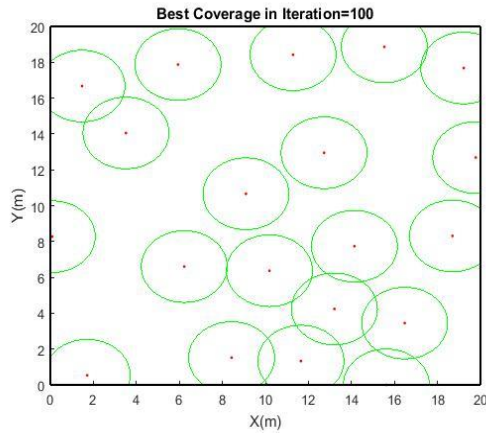


Fig. 3. The network environment in ECAH.

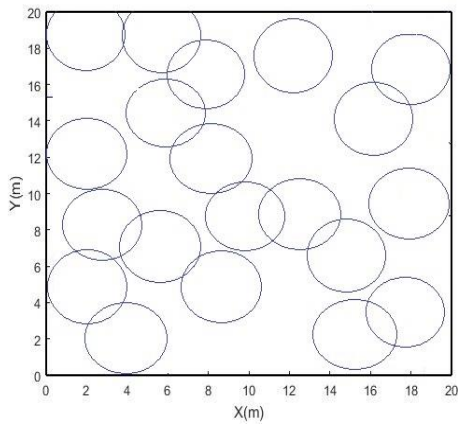


Fig. 4. The network environment in HS.

Table IV shows the parameters used in the second scenario. Our method and PSO algorithm were applied to the coverage of an area with a radius of 5 meters, as shown in Table V. The network environment for the proposed algorithm and the PSO algorithm with radius=5 m is shown in Fig. 5 and 6, respectively. PSO algorithm found the best coverage after the 100th iteration, whereas our method found it after 46 iterations. This shows that our method is significantly more efficient and accurate in finding the best coverage in a given area. It is able to find the best coverage in fewer iterations, indicating that it is more precise in its calculations and requires less time and resources to achieve the desired results.

TABLE V. RESULTS FROM APPLYING THE PROPOSED METHOD AND PSO FOR AREA COVERAGE WITH RADIUS=5

Radius (m)	Iteration (PSO)	Iteration (ECAH)	Coverage rate (PSO)	Coverage rate (ECAH)
5	300	100	80.08%	94.82%

TABLE VI. THIRD SCENARIO PARAMETERS

Simulation parameters		
Sensing radius	Number of sensors	Network size
22	100	278×278

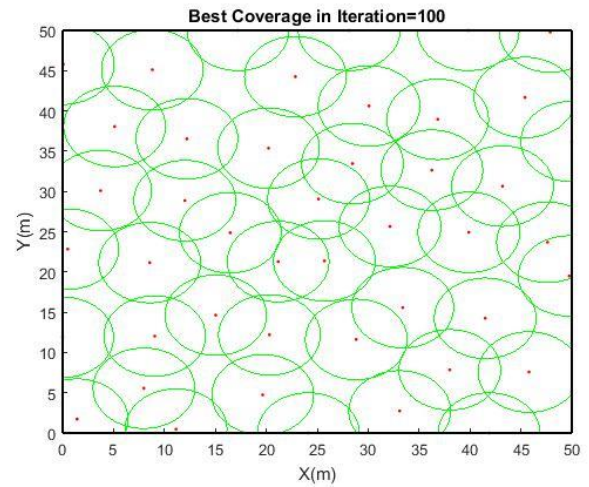


Fig. 5. The network environment in ECAH.

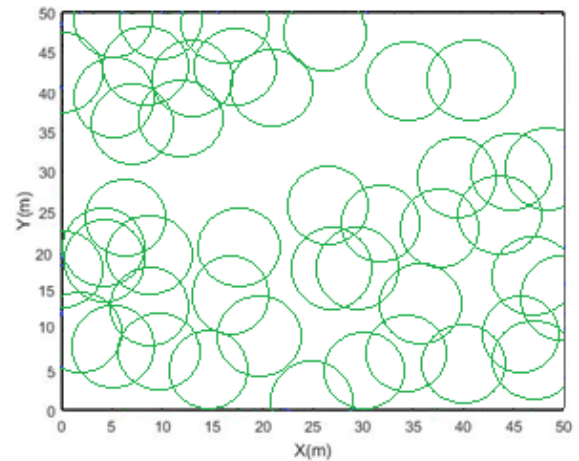


Fig. 6. The network environment in PSO.

In the third scenario, the parameters of [24] are used, shown in Table VI. Table VII gives the results of applying ECAH and ACO algorithms to area coverage for radius=22 m. The network environment for ECAH and ACO algorithms with radius=22 m is shown in Fig. 7 and 8, respectively. The ECAH has found the best coverage in the 85th iteration, while the best coverage of the ACO algorithm is achieved in the 100th iteration. The results show that the ECAH algorithm was able to achieve the best coverage in fewer iterations than the ACO algorithm, indicating that it is more effective and efficient in finding the best coverage in this network environment.

Table VIII shows the parameters adopted from reference [37] for the fourth scenario. Several sensors and iterations are used to conduct the experiment, which shows the detailed results in Table IX. The ECAH has found the best coverage in the 85th iteration, while the best coverage of the GA algorithm is achieved in the 100th iteration. According to the obtained results, the GA with 39 sensors and 100 iterations has reached the optimal coverage, but the proposed method has been optimized with 30 sensors and 100 repeats. The network environment for ECAH and GA algorithms with radius=13 m and N=40 are shown in Fig. 9 and 10, respectively.

TABLE VII. RESULTS FROM APPLYING ECAH AND ACO FOR AREA COVERAGE WITH RADIUS=22

Radius (m)	Iteration (ACO)	Iteration (ECAH)	Coverage rate (ACO)	Coverage rate (ECAH)
22	100	100	87.08%	99.82%

TABLE VIII. FOURTH SCENARIO PARAMETERS

Simulation parameters		
Sensing radius	Number of sensors	Network size
13	100	100×100

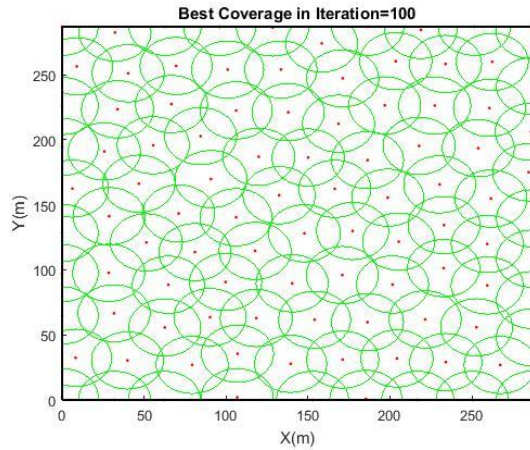


Fig. 7. The network environment in ECAH.

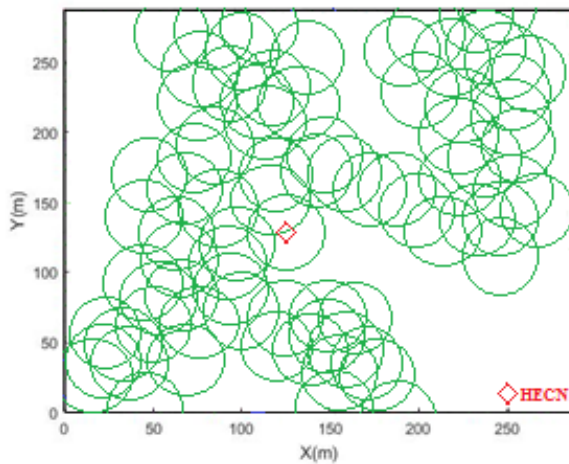


Fig. 8. The network environment in ACO.

TABLE IX. RESULTS FROM APPLYING ECAH AND GA FOR AREA COVERAGE WITH DIFFERENT RADIUS

Number of nodes	Best iteration for optimal coverage in GA		Best iteration for optimal coverage in ECAH	
	Iteration	Coverage rate	Iteration	Coverage rate
N=39	60	93.92%	60	99.92%
	120	95.6%	120	99.95%
N=40	80	94.40%	49	100%
N=37	120	95.6%	102	99.92%

TABLE X. FIFTH SCENARIO PARAMETERS

Simulation parameters				
Sensing radius		Number of sensors		Network size
7		100		100×100
Artificial bee colony parameters				
$\lambda 1 = 1$	$\lambda 2 = 0$	$\beta 1 = 1$	$\beta 2 = 0.5$	“limit”=100

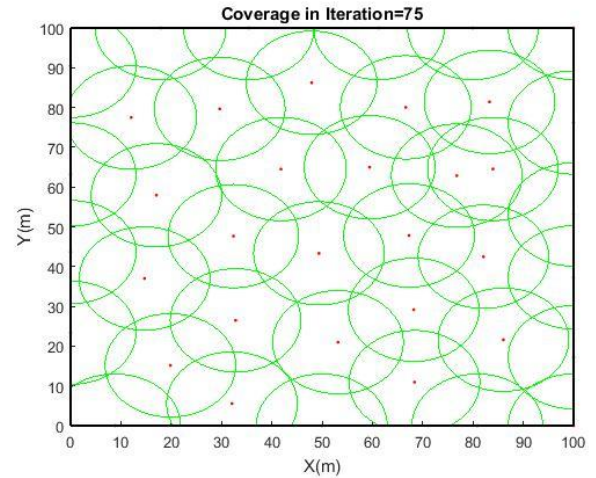


Fig. 9. The network environment in ECAH.

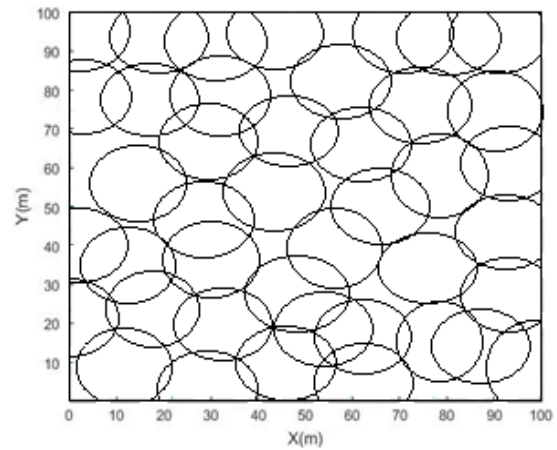


Fig. 10. The network environment in GA.

In the fifth scenario, we used the parameters of [31], which are shown in Table X. Table XI gives the results from applying ECAH and ABC algorithms for area coverage with radius=7 (m). According to the obtained results, ABC has found the best coverage in the 703rd iteration, while the optimal coverage of the ECAH algorithm is achieved in the 652nd iteration. The network environment for ECAH and ABC algorithms with radius=13 m and N=40 are shown in Fig. 11 and 12, respectively.

In the sixth scenario, we used the parameters of [25]. These parameters are shown in Table XII. Table XIII gives the results of applying ECAH and hybrid DEPSO algorithms for area coverage with radius=40 m. The network environments

for ECAH and hybrid DE and PSO algorithms with radius=40 m and N=40 is shown in Fig. 13 and 14, respectively. Fig. 15 shows an overview of the coverage rate of the existing methods. As shown in Table XIV, the proposed method requires fewer iterations for convergence to optimal coverage than existing methods, which the proposed method has inferior runtime and higher run speed compared to existing methods.

TABLE XI. RESULTS FROM APPLYING ECAH AND ABC FOR AREA COVERAGE WITH RADIUS=7

Radius (m)	Iteration (ABC)	Iteration (ECAH)	Coverage rate (ABC)	Coverage rate (ECAH)
7	1000	1000	97.52%	98.59%

TABLE XII. SIXTH SCENARIO PARAMETERS

Simulation parameters						
Sensing radius	Number of sensors				Network size	
40	20	25	30	35	40	450×450

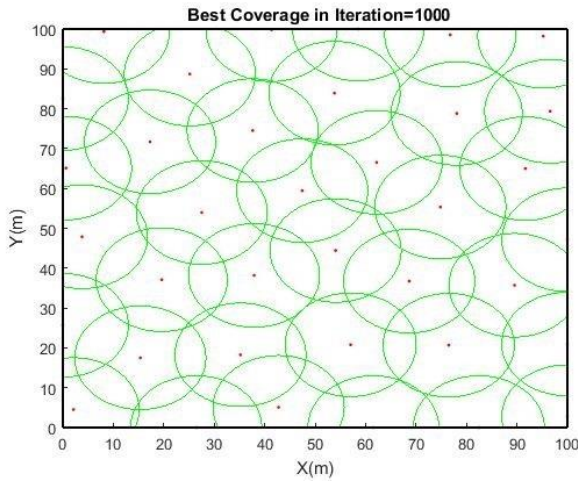


Fig. 11. The network environment in ECAH.

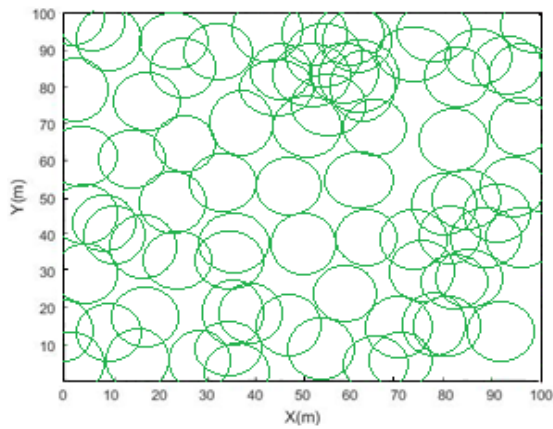


Fig. 12. The network environment in ABC.

TABLE XIII. RESULTS FROM APPLYING ECAH AND HYBRID DE AND PSO FOR AREA COVERAGE WITH RADIUS=40

Number of nodes	Best iteration for optimal coverage in hybrid DE and PSO		Best iteration for optimal coverage in ECAH	
	Iteration	Coverage rate	Iteration	Coverage rate
N=20	100	27%	54	49.25%
	100	32%	99	61.15%
N=25	100	37%	98	70.92%
N=30	100	42%	48	78.56%
N=35	100	47%	31	84.44%

TABLE XIV. THE NUMBER OF ITERATIONS OF THE PROPOSED METHOD AND EXISTING METHODS IN DIFFERENT SCENARIOS

Scenario	Methods reviewed	Iteration	Scenario	Methods reviewed	Iteration
First scenario	ECAH	900	Forth scenario	ECAH	49
	HS	7000		GA	120
Second scenario	ECAH	100	Fifth scenario	ECAH	1000
	PSO	300		ABC	1000
Third scenario	ECAH	100	Sixth scenario	ECAH	31
	ACO	100		DE and PSO	100

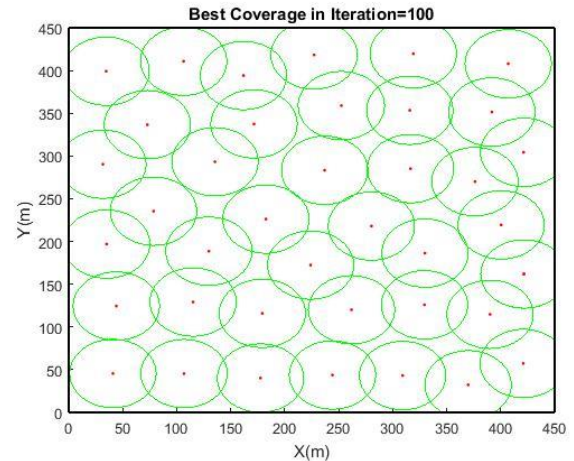


Fig. 13. The network environment in ECAH.

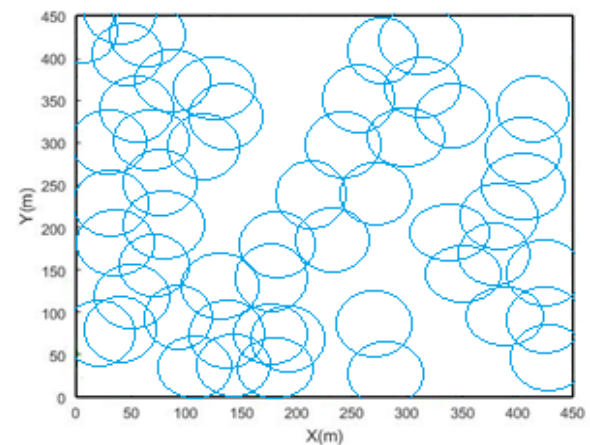


Fig. 14. The network environment in DEPSO.

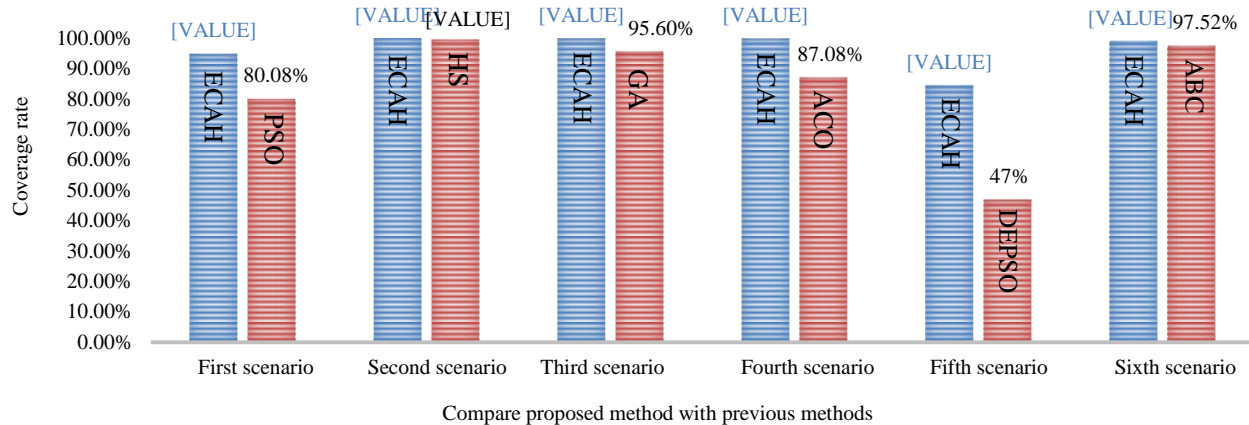


Fig. 15. Overview of the coverage rate of the existing methods.

V. CONCLUSION

Coverage problem is one of the most critical research fields for WSNs. The network's lifetime is increased by creating optimized coverage in WSNs. In addition, the optimal coverage causes the optimal consumption of the network energy. This paper combined the HS and ABC algorithms to increase the area coverage and the lifetime of the WSNs. The efficiency of the hybrid method is demonstrated by comparing it with HS, ABC, ACO, GA, hybrid DE, and PSO algorithms. Our method outperforms previous methods according to the simulation results. We hope that we will be able to find better and more optimized methods using other meta-heuristic algorithms for WSNs coverage problems. This algorithm will be improved in the future and applied to WSN coverage optimization in more complex environments. Additionally, the algorithm will be extended to a wide range of IoT optimization problems.

ACKNOWLEDGMENTS

This work was supported by the training plan for young backbone teachers in Henan Province (No.2018GGJS267)

REFERENCES

- [1] A. Mehdodniya et al., "Energy-Aware Routing Protocol with Fuzzy Logic in Industrial Internet of Things with Blockchain Technology," *Wireless Communications and Mobile Computing*, vol. 2022, 2022.
- [2] F. Zhu and W. Wang, "A Distributed Unequal Clustering Routing Protocol Based on the Improved Sine Cosine Algorithm for WSN," *Journal of Sensors*, vol. 2022, 2022.
- [3] S. A. Saeidi, F. Fallah, S. Barmaki, and H. Farbeh, "A novel neuromorphic processors realization of spiking deep reinforcement learning for portfolio management," in *2022 Design, Automation & Test in Europe Conference & Exhibition (DATE)*, 2022: IEEE, pp. 68-71.
- [4] F. Vahedifard, S. Hassani, A. Afrasiabi, and A. M. Esfe, "Artificial intelligence for radiomics; diagnostic biomarkers for neuro-oncology," *World Journal of Advanced Research and Reviews*, vol. 14, no. 3, pp. 304-310, 2022.
- [5] R. N. Jacob, "Non-performing Asset Analysis Using Machine Learning," in *ICT Systems and Sustainability: Proceedings of ICT4SD 2020*, Volume 1, 2021: Springer, pp. 11-18.

- [6] J. Akhavan, J. Lyu, and S. Manoochehri, "A deep learning solution for real-time quality assessment and control in additive manufacturing using point cloud data," *Journal of Intelligent Manufacturing*, pp. 1-18, 2023.
- [7] S. R. Abdul Samad et al., "Analysis of the Performance Impact of Fine-Tuned Machine Learning Model for Phishing URL Detection," *Electronics*, vol. 12, no. 7, p. 1642, 2023.
- [8] S. H. Haghshenas, M. A. Hasnat, and M. Naeini, "A Temporal Graph Neural Network for Cyber Attack Detection and Localization in Smart Grids," *arXiv preprint arXiv:2212.03390*, 2022.
- [9] S. Meisami, M. Beheshti-Atashgah, and M. R. Aref, "Using Blockchain to Achieve Decentralized Privacy In IoT Healthcare," *arXiv preprint arXiv:2109.14812*, 2021.
- [10] A. Morteza, M. Ilbeigi, and J. Schwed, "A blockchain information management framework for construction safety," in *Computing in Civil Engineering 2021*, 2021, pp. 342-349.
- [11] S. Vairachilai, A. Bostani, A. Mehdodniya, J. L. Webber, O. Hemakesavulu, and P. Vijayakumar, "Body Sensor 5 G Networks Utilising Deep Learning Architectures for Emotion Detection Based On EEG Signal Processing," *Optik*, p. 170469, 2022.
- [12] C. Lersteau, A. Rossi, and M. Sevaux, "Minimum energy target tracking with coverage guarantee in wireless sensor networks," *European Journal of Operational Research*, 2017.
- [13] A. J. Al-Mousawi and H. K. Al-Hassani, "A survey in wireless sensor network for explosives detection," *Computers & Electrical Engineering*, 2017/12/01/ 2017, doi: <https://doi.org/10.1016/j.compeleceng.2017.11.013>.
- [14] V. Markevicius et al., "Two thermocouples low power wireless sensors network," *AEU - International Journal of Electronics and Communications*, vol. 84, pp. 242-250, 2018/02/01/ 2018, doi: <https://doi.org/10.1016/j.aeue.2017.11.032>.
- [15] A. Oracevic, S. Akbas, and S. Ozdemir, "Secure and reliable object tracking in wireless sensor networks," *Computers & Security*, vol. 70, pp. 307-318, 2017.
- [16] P. B. Rasband, "Wireless sensor network," ed: Google Patents, 2017.
- [17] S.-C. Wang, H. C. Hsiao, C.-C. Lin, and H.-H. Chin, "Multi-objective wireless sensor network deployment problem with cooperative distance-based sensing coverage," *Mobile Networks and Applications*, pp. 1-12, 2022.
- [18] W. Osamy, A. M. Khedr, A. Salim, A. I. Al Ali, and A. A. El-Sawy, "Coverage, deployment and localization challenges in wireless sensor networks based on artificial intelligence techniques: a review," *IEEE Access*, 2022.
- [19] S. J. Bhat and S. KV, "A localization and deployment model for wireless sensor networks using arithmetic optimization algorithm," *Peer-to-Peer Networking and Applications*, vol. 15, no. 3, pp. 1473-1485, 2022.

- [20] P. Tirandazi, A. Rahiminasab, and M. Ebadi, "An efficient coverage and connectivity algorithm based on mobile robots for wireless sensor networks," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1-23, 2022.
- [21] Q. Li and N. Liu, "Coverage optimization algorithm based on control nodes position in wireless sensor networks," *International Journal of Communication Systems*, vol. 35, no. 5, p. e4599, 2022.
- [22] N. A. B. Ab Aziz, A. W. Mohemmed, and B. D. Sagar, "Particle swarm optimization and Voronoi diagram for wireless sensor networks coverage optimization," in *Intelligent and Advanced Systems, 2007. ICIAS 2007. International Conference on*, 2007: IEEE, pp. 961-965.
- [23] S. Panov and S. Koceski, "Area coverage in wireless sensor network by using harmony search algorithm," in *Embedded Computing (MECO), 2014 3rd Mediterranean Conference on*, 2014: IEEE, pp. 210-213.
- [24] S. Fidanova and P. Marinov, "Optimal wireless sensor network coverage with Ant Colony Optimization," in *International conference on swarm intelligence*, 2011.
- [25] I. Maleki, S. R. Khaze, M. M. Tabrizi, and A. Bagherinia, "A new approach for area coverage problem in Wireless Sensor Networks with hybrid particle swarm optimization and differential evolution algorithms," *International Journal of Mobile Network Communications and Telematics (IJMNCT)*, vol. 3, no. 6, pp. 61-76, 2013.
- [26] N.-n. Qin and J.-l. Chen, "An area coverage algorithm for wireless sensor networks based on differential evolution," *International Journal of Distributed Sensor Networks*, vol. 14, no. 8, p. 1550147718796734, 2018.
- [27] Q. He, Z. Lan, D. Zhang, L. Yang, and S. Luo, "Improved Marine Predator Algorithm for Wireless Sensor Network Coverage Optimization Problem," *Sustainability*, vol. 14, no. 16, p. 9944, 2022.
- [28] H. Wu, H. Zhu, and X. Han, "An improved k-angle coverage algorithm for multimedia wireless sensor networks based on two-layer tabu search," *Peer-to-Peer Networking and Applications*, vol. 15, no. 1, pp. 28-44, 2022.
- [29] A. Zrelli and T. Ezzedine, "A new approach of WSN deployment, K-coverage and connectivity in border area," *Wireless Personal Communications*, vol. 121, no. 4, pp. 3365-3381, 2021.
- [30] S. Sumitha Pandit and B. Kalpana, "Coverage Using Swarm Intelligence Family in Wireless Sensor Networks," in *Soft Computing for Security Applications: Springer*, 2022, pp. 351-363.
- [31] C. Ozturk, D. Karaboga, and B. Gorkemli, "Probabilistic dynamic deployment of wireless sensor networks by artificial bee colony algorithm," *Sensors*, vol. 11, no. 6, pp. 6056-6065, 2011.
- [32] Z. Fan and W. Zhao, "Network Coverage Optimization Strategy in Wireless Sensor Networks Based on Particle Swarm Optimization," 2011.
- [33] M. N. Rahman and M. Matin, "Efficient algorithm for prolonging network lifetime of wireless sensor networks," *Tsinghua Science & Technology*, vol. 16, no. 6, pp. 561-568, 2011.
- [34] B. Wu, C. Qian, W. Ni, and S. Fan, "Hybrid harmony search and artificial bee colony algorithm for global optimization problems," *Computers & Mathematics with Applications*, vol. 64, no. 8, pp. 2621-2634, 2012.
- [35] K. S. Lee and Z. W. Geem, "A new structural optimization method based on the harmony search algorithm," *Computers & structures*, vol. 82, no. 9, pp. 781-798, 2004.
- [36] J. C. Bansal, H. Sharma, and S. S. Jadon, "Artificial bee colony algorithm: a survey," *International Journal of Advanced Intelligence Paradigms*, vol. 5, no. 1-2, pp. 123-159, 2013.
- [37] J. Jia, J. Chen, G. Chang, and Z. Tan, "Energy efficient coverage control in wireless sensor networks based on multi-objective genetic algorithm," *Computers & Mathematics with Applications*, vol. 57, no. 11, pp. 1756-1766, 2009.

Patient Health Monitoring System Development using ESP8266 and Arduino with IoT Platform

Jamil Abedalrahim Jamil Alsayaydeh^{1*}, Mohd Faizal bin Yusof², Muhammad Zulkhakim Bin Abdul Halim³,
Muhammad Noorazlan Shah Zainudin⁴, Safarudin Gazali Herawan⁵

Department of Electronics & Computer Engineering Technology-Fakulti Teknologi Kejuruteraan Elektrik & Elektronik (FTKEE),
Universiti Teknikal Malaysia Melaka (UTeM), Melaka, Malaysia^{1,3}

Cetri-Faculty of Electronics and Computer Engineering, Universiti Teknikal Malaysia Melaka, Hang Tuah Jaya, Melaka,
Malaysia⁴

Department Homeland Security-Faculty of Resilience, Rabdan Academy, Abu Dhabi, United Arab Emirates²
Industrial Engineering Department-Faculty of Engineering, Bina Nusantara University, Jakarta, Indonesia 11480⁵

Abstract—The Internet of Things (IoT) has emerged as a transformative technology that has revolutionized the field of healthcare. One of the most promising applications of Internet of Things (IoT) in healthcare is patient health monitoring, which allows healthcare providers to remotely monitor patients' health and provide prompt medical attention when needed. This research work focuses on developing an Internet of Things (IoT)-based patient health monitoring system aimed at providing a solution for patients, particularly the elderly, who face the risk of unexpected death due to the lack of medical attention. The proposed system utilizes a heartbeat sensor and an Infrared IR temperature sensor connected to Arduino UNO and Nodemcu, respectively, to monitor the patient's vital signs. The sensors collect the data, which is then sent to an Internet of Things (IoT) web platform via a Wi-Fi connection. The Internet of Things (IoT) platform displays the real-time data of the patient's health status, including the temperature and heartbeat rate, which can be monitored by doctors and nurses. The system is designed to send alerts to healthcare providers in the event of any medical emergency, ensuring that prompt medical attention can be provided to the patient. The significance of this research work lies in its potential to revolutionize the healthcare industry by providing a more efficient and effective means of patient health monitoring. The system can be used to monitor a large number of patients simultaneously, which is particularly beneficial in hospitals with a large patient load. Moreover, it can reduce the workload of healthcare providers, allowing them to focus on other critical tasks. This innovative system has the potential to improve the overall quality of healthcare services and lead to better health outcomes for the society.

Keywords—Patient health monitoring; Internet of Things (IoT); Arduino UNO; Nodemcu ESP8266; thingspeak; wearable device; temperature value; heartbeat value; remotely

I. INTRODUCTION

In today's digital age, the integration of the internet has transformed various industries, including healthcare. The emergence of Internet of Things (IoT) technology in health monitoring systems has revolutionized the way we approach healthcare [1]. With the widespread usage of the internet, coupled with the increasing efficiency of devices and gadgets, we can now monitor patients around the clock using Internet of Things (IoT)-based health monitoring systems. These devices constantly examine factual data and produce vital signs that

can be accessed remotely via web connectivity. This not only enables real-time patient monitoring but also facilitates prompt crisis response services. As such, Internet of Things (IoT)-based gadgets provide both recognition and emergency response services, which are critical components in ensuring optimal patient care.

The traditional approach to patient monitoring in hospitals and healthcare facilities can be tedious and often results in unequal treatment of patients. Medical staff is required to manually check multiple vital signs of patients, which can be time-consuming and prone to errors. Furthermore, patients' families or relatives need to visit them for updates on their condition, which can be inconvenient and burdensome [2]. Additionally, many hospitals still rely on outdated paper-based recording systems, which are prone to errors and can result in missing or illegible data [3]. These issues can lead to a waste of time and resources, as well as compromise patient care. However, the implementation of an Internet of Things (IoT)-based patient health monitoring system can revolutionize healthcare by providing real-time remote monitoring of patients, digital storage of patient information, and immediate check-ups for individuals to maintain optimal health [4]. The proposed work presented in this study marks a new pathway towards solving numerous problems that have been plaguing the healthcare industry and society at large. The current system of healthcare services has been marred with inefficiencies and management issues, which necessitates the need for a more effective and sophisticated approach to healthcare services. The IoT system is the perfect match for this modern era, with its real-time monitoring, sophisticated data collection and storage, and alert systems that have proven to be highly effective. This proposed work represents a step ahead of the current healthcare system, with its ability to solve numerous issues and monitor patients accurately and in real-time. Despite the impressive results of previous studies and proposed projects, the medical industry has been slow to adopt Internet of Things (IoT)-based systems. This study seeks to create an updated, reliable, and innovative healthcare system that will meet the growing expectations of society in terms of saving lives and advancing medical technology development [5][6].

To achieve the lofty goal of this endeavor, it is imperative to concoct a novel and innovative approach to the management

of hospital and clinic system operations, with a thorough and effective methodology in place. The objectives of this work are to assist health staff in monitoring and recording patient data frequently at specific times, which is especially important for hospitals with a large number of patients where unexpected fluctuations can occur. Additionally, with the numerous data collections that occur for each monitored patient, all recorded data will be saved in one place - a cloud-based system - which is more efficient than the current method used. Lastly, an alert system functions as an emergency detection tool that can send notifications via email to the administrator or health staff, allowing immediate action to be taken based on the condition of the patient during panic situations.

This study has made several contributions to the field of healthcare by developing an Internet of Things (IoT)-based patient health monitoring system that can remotely monitor patients' vital signs and provide prompt medical attention in case of any medical emergency. The system uses heartbeat and temperature sensors connected to Arduino UNO and Nodemcu, respectively, and sends data to an Internet of Things (IoT) web platform that displays real-time data of the patient's health status. This innovative system has the potential to revolutionize the healthcare industry by providing a more efficient and effective means of patient health monitoring, which will ultimately lead to a better quality of life and improved health outcomes for the society.

The remainder of this paper has been organized as follows: Section II discusses the related works. The background of the study is described in Section III. Section IV describes the system implementation methods. Section V describes the results, Section VI describes the discussion and finally, the conclusion is described in Section VII.

II. RELATED WORK

To achieve the intended target, it is crucial to delve into the fundamentals and knowledge garnered from prior studies. Previous works have delved into health monitoring systems that utilize wireless technology. As the health monitoring system is an integral part of the Internet of Things (IoT) system, it is imperative to scrutinize past works that employ comparable strategies and aspire to similar objectives as this research endeavor.

A. Classic Health Monitoring

The classic method of patient health monitoring has been a stalwart in healthcare institutions such as hospitals and clinics since its inception in the early 2000s. Despite its extensive use over the years, this approach falls short in terms of its ability to accurately monitor a large number of patients within a set time frame. Health staff, comprising doctors and nurses, is tasked with physically examining and recording patient data. Patients are required to queue or wait their turn for checkups, while those confined to their beds require round-the-clock monitoring. Patient data is painstakingly recorded by hand in a patient booklet and stored in physical form on shelves [7].

B. Wearable Smart Health Monitoring

Pradhan, S., Zainuddin, A. A., and Sahak's research project [8] offers an intriguing insight into the myriad methods

employed to monitor health status. The authors devised a plethora of sensors to keep a watchful eye on the real-time health of animals. Their study utilized an array of sensors, including those that measured temperature, respiration rate, heart rate, and humidity rate [9][10]. To control this project, they employed the Raspberry Pi, complete with its built-in Wi-Fi module. Additionally, they created a cloud-based platform to store and transmit sensor data, while an Android mobile application system was used to gather real-time data on the animal's health state. This project effectively underscores the feasibility of wearable devices, which may be a fitting solution for our research endeavor.

C. Real Time Health Monitoring Devices

A trustworthy patient monitoring system that allows medical professionals to keep track of patients in the hospital or during their daily activities has been developed by A. Abdullah, A. Ismael, A. Rashid, A. Abou-Elnour, and M. Tarique [11][12]. The system employs mobile devices for wireless patient monitoring and enables real-time transmission of data on the patient's physical condition. The project aims to collect and analyze vital physiological data from patients to accurately determine their health and fitness levels. Additionally, the system can send text messages or email reports with important health information to patients. Healthcare professionals can provide valuable medical advice by utilizing the information in these messages. The authors utilized multiple sensors, an Arduino microcontroller, and LabVIEW software to develop the system [13][14].

D. Focused Health Monitoring

Masud, Muhammad and Alhumyani proposed a health monitoring system that focuses on three vital signs of patients: temperature, blood pressure, and heartbeat [15]. The system uses a microcontroller and several sensors to collect and visualize the data on a Liquid Crystal Display (LCD) and store it in cloud-based storage. This system is designed to assist healthcare professionals in monitoring multiple patients simultaneously. The authors employed microcontrollers such as Arduino and Nodemcu, as well as Ubidots software, to implement the project [16][17]. The system has been tested and has demonstrated good accuracy in monitoring patients' vital signs introducing a novel tree-based deep model for automatic face recognition in a cloud-based environment that strikes a balance between computational efficiency and high accuracy. The model divides the input volume into multiple volumes and creates a tree structure for each volume, where each branch of the tree is represented by a residual function comprising of a convolutional layer, batch normalization, and a non-linear function. Rigorously evaluated on publicly available databases, the proposed model achieves remarkable accuracies, surpassing the current state-of-the-art deep models for face recognition.

E. Smart Patient Monitoring using Databases and Physically Sensors

Chee Yuan, L. and Kong have proposed a health monitoring system that uses physical sensors [18]. The system involves attaching sensors to the patient's body to monitor their health data, which is then stored in a database built using Personal Home Page (PHP). While this method is effective for storing patient data remotely, it requires the use of contact

sensors that must physically touch the patient's body. Despite this limitation, the system is functional and can be implemented in health departments [19].

F. Internet of Things IoT Health Monitoring

V. Akhila, Y. Vasavi, K. Nissie, and P. V. Rao have utilized IoT and sensor technologies to develop a system that can monitor a patient's health condition in real-time [20]. The authors tested the system by monitoring the health condition of a real individual. The project involved the use of Internet of Things (IoT) and microcontroller like Arduino to gather data from analogue sensors. In the healthcare industry, the combination of IoT with Arduino microcontroller has proven to be effective in monitoring and collecting current patient health condition data [21]. This work describes the use of Arduino UNO as a sensor node and gateway in a remote health monitoring system. The system utilizes GSM for communication between doctors and patients, and Wi-Fi for transmitting sensor data to a web-based IoT platform. The ATMEGA 328p microprocessor, installed within the sensor nodes, measures temperature, pulse rate, and blood pressure readings using the temperature sensor and HP/BP sensor. The data is displayed on an LCD attached to the Arduino and transmitted to the cloud server via the Wi-Fi module. This technology allows for continuous monitoring of the health of severely ill patients and seniors suffering from heart and blood pressure disorders.

G. Contactless Internet of Things IoT Patient Health Monitoring

The framework suggested by Dipti S. Gandhamwar and Sunil Kuntawaris [22] is used to remotely monitor patients' wellbeing. Sensors such as the heartbeat sensor, temperature sensor, and SpO₂ sensor are used to estimate important parameters. The suggested model enables medical professionals to check on patient wellbeing from any location and helps people consult experts anywhere in the world. The system combines the Internet of Things (IoT) and remote sensor technology for effective health monitoring, and sensor data must be available consistently. The data is archived and viewable on the web server [23]. The system is set up to send a message to the specialist if the sensor data exceeds the threshold values. The main benefit is a reduction in the amount of time specialists and patients need to intervene in times of crisis. By recommending a minimally expensive framework for saving lives, with the hope that individuals will be amenable, the goal is achieved. The limitation of the specialist's accessibility is considered, and the proposed model does not include the pulse-checking framework [24].

III. BACKGROUND OF THE STUDY

Smart technology is a combination of sensor-based, data-driven, and programmable technologies that also incorporate artificial intelligence. The integration of traditional medical scaling with advanced technology has resulted in patient health monitoring, which offers numerous benefits, such as monitoring vital signs like heartbeat rate, temperature rate, and respiration rate, which can help facilitate faster medical intervention. Currently, the Internet of Things (IoT) is considered the leading solution for developing innovative health monitoring systems [27]. Patient health monitoring

systems provide a more accessible and cost-effective alternative to the traditional clinical conduct, allowing for improved access to healthcare facilities.

The goal of this proposed research is to develop a patient health monitoring system based on the Internet of Things (IoT). By minimizing doctor visits, hospital stays, and daily diagnostic tests, this system aims to reduce healthcare costs [28]. The sensors are connected to an Arduino Uno microprocessor that tracks and transmits the data collected to ThingSpeak, which can generate alerts. Health staff can monitor patients' health using IoT technology and identify any unusual readings in terms of temperature and heartbeat. ThingSpeak can display patient data in real-time. This system can save a lot of time in measuring and monitoring a patient's health at a specific time, making IoT technology reliable and efficient. The use of IoT-based patient monitoring systems has enormous potential for healthcare professionals to observe tiny aspects of a patient's well-being [29]. Therefore, the system must check and analyze the data in detail, with ThingSpeak serving as an IoT platform.

Medical facilities in remote areas are often inaccessible to local communities [20]. As a result, many people ignore minor health issues that could indicate larger problems, such as changes in body temperature or heart rate. When these issues progress to the point of being life-threatening, medical attention is sought, resulting in unnecessary expenses. This is especially important to consider in the event of an epidemic spreading to an area without easy access to doctors. Providing patients with smart sensors that can be remotely monitored could save many lives and help prevent the spread of illness [30].

The main goal of this proposed work is to develop a practical and effective system for Internet of Things (IoT)-based patient health monitoring, and to study its wellness, operation methods, theoretical framework, and impacts [31]. The recommended system is depicted in Fig. 1, and it uses two sensors to detect the patient's body temperature and heartbeat. The sensors are connected to a microcontroller unit that processes the input data from both sensors. The collected data is then sent to or stored in an IoT database. This database allows doctors or health staff to retrieve the data and monitor the patient's current health condition from anywhere via the base station.

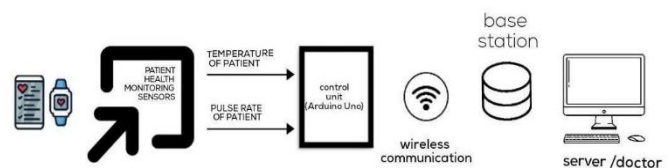


Fig. 1. Smart health monitoring system process flow.

The flowchart of the object recognition process is shown in Fig. 2.

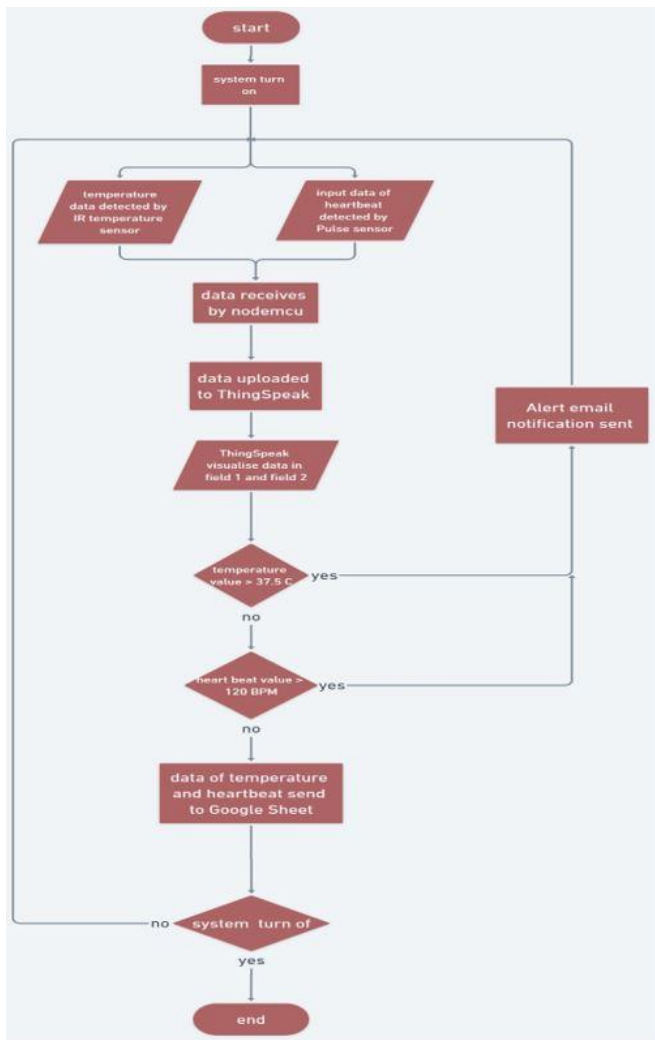


Fig. 2. Flowchart of the system.

IV. SYSTEM IMPLEMENTATION METHODS

Fig. 2 illustrates the steps involved in developing this system. Firstly, the hardware components necessary for the project were carefully selected. The Arduino Integrated Development Environment (IDE) was chosen as the software platform for coding due to its user-friendly interface and ability to process input signals and generate corresponding output signals. The project involves placing sensors on the patient's body to record vital signs. The MLX90614 sensor will be used to measure the patient's body temperature, while the Heartbeat module or pulse sensor will monitor the patient's pulse rate. The microcontroller will process the data obtained from both sensors, checking the condition of the patient against pre-set parameters. All data collected will be converted to a list and notification of any abnormal readings will be sent to the administrator. This feature is particularly useful in cases where the patient's health is at risk, and prompt action is required.

A. Hardware Development

This section provides an explanation of the hardware components used in the system. The Arduino UNO is used as the microcontroller and receives and sends data from the pulse sensor to Nodemcu through UART serial communication. The pulse sensor is not compatible with Nodemcu and needs to be connected to Arduino UNO instead. The Arduino UNO is powered by a 9V battery which is also used to power up the pulse sensor. The connections between the pulse sensor and Arduino UNO are established by connecting the VCC pin to 3.3V of Arduino, GND pin to GND of Arduino, and analog pin to A0 of Arduino. The paragraph also mentions that the hardware components are functioning properly as they indicate light when turned on.

Nodemcu is a microcontroller with a built-in Wi-Fi module that is well-suited for this Internet of Things (IoT) system as it can transmit data via the internet to the desired IoT platform. ThingSpeak platform is used in this system. Nodemcu is powered by a 9V battery. Serial communication with Arduino is established via the Receive RX and Transmit TX pins on each microcontroller, allowing pulse sensor data from Arduino to be read. Infrared (IR) temperature sensor is connected to Nodemcu to read temperature data. The IR temperature sensor requires a voltage of 3.3V, which is provided by connecting VCC to 3.3V of Nodemcu and GND to Ground GND of Nodemcu. The Serial Clock (SCL) and Serial Data (SDA) pins are connected to D1 and D2 of Nodemcu, respectively.

B. Software Development

This section describes the coding and data transmission process for the IoT-based patient health monitoring system. The coding is necessary for the pulse sensor to read the heart beat data, which is then transmitted to the Nodemcu microcontroller via serial communication. The pulse sensor library is included in the coding, and the data is read using the library's provided read data value. The data is then sent from Arduino UNO to Nodemcu, where it is received and combined with the temperature data from the Infrared IR temperature sensor. The combined data is then transmitted to the ThingSpeak platform for monitoring. The system settings provide the ThingSpeak host and Application Programming Interface API key necessary for obtaining the data. Fig. 3 shows how to access this information. ThingSpeak is a useful technology for IoT applications as it allows for remote monitoring and management of data processing. It collects, analyzes, and visualizes sensor data, and can also initiate responses.

ThingSpeak in this work is platform to visualize data receives from Nodemcu which are temperature value and heartbeat value for monitoring purpose in real time using internet. IFTTT is also used in this work by connecting it to ThingSpeak as it acts for alert message via email when there is misreading over the value or threshold that have been set when reading data from Nodemcu such as for temperature value when exceeding 37.5 and for heart beat when exceeding 120 BPM.


```
19
20 const char *ssid = "BEECHMANTUL-2.4GHz"; //ENTER YOUR WIFI SETTINGS <<<<<<<<<
21 const char *password = "P@ssw@rdbeec";
22 WiFiClient Client;
23 unsigned long myChannelNumber = 1765756 ;
24 const char * myWriteAPIKey = "P5RBUBP1QH5W056I";
25 //Web address to read from
26 const char *host = "api.thingspeak.com";
27 String apiKey = "P5RBUBP1QH5W056I"; //ENTER YOUR API KEY <<<<<<<<<
28
```

Fig. 3. Coding for setup IoT data transmission.

V. RESULTS

Upon completing the hardware and software development for this work, it is imperative to conduct testing to determine if the desired goals have been achieved. To this end, an Infrared IR temperature sensor is connected through ESP8266, while the pulse sensor is connected through Arduino UNO, as depicted in Fig. 4. Both microcontrollers are powered by a 9V battery to initiate the circuit and system. Serial communication, specifically universal asynchronous receiver/transmitter (UART) communication, facilitates interaction between the two microcontrollers. In this transmission of serial data, various serial protocols are employed and adhered to.

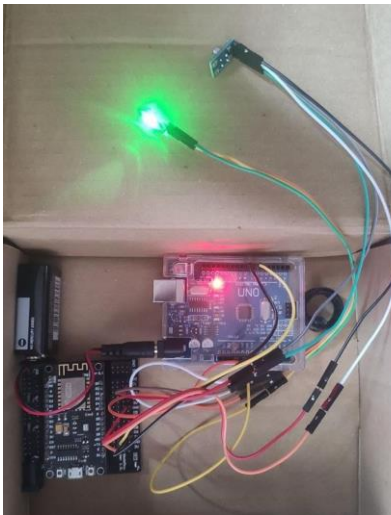


Fig. 4. Complete hardware circuit communication.

To collect data from the patient, the sensors are positioned in a way that aligns with the patient's position on the bed. The pulse sensor is placed on the fingertip while the Infrared IR temperature sensor is placed on the patient's forehead. The sensors are placed in these locations due to their compatibility with the patient's environment and the specifications required for data collection. In Fig. 5, the collected data is uploaded from the Nodemcu to ThingSpeak, with the temperature value on field 1 and the patient's heartbeat value on field 2. The uploaded data is then sent to a pre-set email on IFTTT and uploaded to a Google spreadsheet.

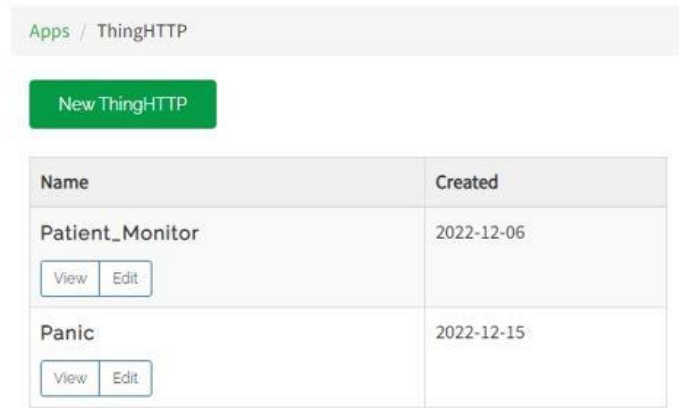


Fig. 5. ThingSpeak channel created.

To set up the IFTTT account, login is required followed by creating a new applet with two actions for triggering. The first action, "Patient_info," involves sending patient data, such as heartbeat and temperature, to a spreadsheet row via email. The second action, "Panic," alerts via email if needed (see Fig. 7). The actions should then be reviewed to ensure proper functioning. To connect the data from ThingSpeak to email, ThingHTTP and React must be configured, and the actions created in IFTTT must be triggered. Finally, save all the data and test the system.

	A	B	C	D
5	January 14, 202	Patient_info	35.6 (C)	88 (BPM)
6	January 14, 202	Patient_info	34.6 (C)	92 (BPM)
7	January 14, 202	Patient_info	35.7 (C)	86 (BPM)
8	January 14, 202	Patient_info	35.1 (C)	76 (BPM)
9	January 14, 202	Patient_info	36.4 (C)	79 (BPM)
10	January 14, 202	Patient_info	35.4 (C)	72 (BPM)
11	January 14, 202	Patient_info	34.9 (C)	72 (BPM)
12	January 14, 202	Patient_info	35.8 (C)	75 (BPM)
13	January 14, 202	Patient_info	36.4 (C)	83 (BPM)
14	January 14, 202	Patient_info	36.9 (C)	98 (BPM)
15	January 14, 202	Patient_info	37.4 (C)	100 (BPM)
16	January 14, 202	Patient_info	35.6 (C)	88 (BPM)
17				

Fig. 6. Google spreadsheet showing data collected.

When the data uploaded to ThingSpeak meeting the condition set in ThingHTTP the alert message will be send through email saying it is panic condition for the patient as well as sending the data values of temperatures and also heartbeat as can see in Fig. 6. The condition set for temperature is when value of data is exceeding 37.5 Celsius while for heartbeat is when value of data is exceeding 120 Beat Per Minute BPM.



Fig. 7. Alert notification send via email.

Based on the detailed results and implementation of this work, it can be concluded that it holds promising potential for future development in the medical sector. Comparing it to other related works listed in Table I, this system shows good accuracy in collecting data through sensors and efficiency when implemented in the medical sector. However, there are a few areas for improvement to ensure the sustainability of this work. For instance, incorporating additional sensors can enable tracking of more health conditions, while high-quality sensors can stabilize the readings of data for longer periods. Additionally, using MySQL and Personal Home Page PHP as the database can cover a lot of patient data.

TABLE I. COMPARISON BETWEEN CURRENT AND RESEARCHED SYSTEM

YEAR	METHOD	NUMBER OF PATIENT TEST FOR MONITORING PER HOUR	NUMBER OF PATIENT CAN COVERED/ MONITORED PER HOUR	PERCENTAGE OF COVERED PATIENT	DATA RECORDED ACCURACY COMPARED TO ACTUAL DATA	REFERENCE
2009	CURRENT SYSTEM THAT USED WRITING METHOD FOR COLLECTING PATIENT DATA WHEN CHECK UP	50	14	28%	85%	[7]
2015	HEALTH MONITORING USING MOBILE DEVICE	50	36	72%	90%	[11]
2020	MONITORED THREE BASIC HEALTH CONDITION WITH DIFFERENT MODES	50	33	66%	89%	[15]
2019	USING PHYSICALLY SENSORS USING ARDUINO UNO AND NODEMCU WHICH THEN BE UPLOADED TO MYSQL AND PERSONAL HOME PAGE PHP	50	44	88%	93%	[18]
2018	IOT SMART HEALTH MONITORING USING ARDUINO UNO ATTACHED TO PERSONAL COMPUTER	50	40	82%	94%	[20]
2019	DATA FROM SENSORS IS ATTACHED TO ARDUINO UNO AND SEND TO CLOUDS BY RASPBERRY PI 3.	50	42	84%	91%	[25]
2021	CONTACTLESS SENSORS DATA TRANSMITTED BY ARDUINO UNO TO NODEMCU WHICH THEN BE UPLOADED ONLINE TO CLOUD	50	46	92%	94%	[26]
PROPOSED RESEARCH						
PROPOSED SYSTEM OF IOT PATIENT HEALTH MONITORING USING NODEMCU AND ARDUINO UNO	50	47	93%	96%	N/A	THIS WORK

VI. DISCUSSION

The Patient Health Monitoring System, developed with the objective of utilizing ESP8266 and Arduino with an Internet of Things (IoT) platform, has been successfully implemented and tested through various iterations. As indicated in the results section, the hardware circuit connection in Fig. 4 provided the best and most functional means of collecting patient data, including temperature and heartbeat. The collected data is shown in real-time on ThingSpeak, making it easier for medical staff to monitor patients, while also being stored in Google Sheets for future reference. Despite its success, the system still has some limitations, particularly in terms of the reliability of the sensor data due to varying price and quality. This system has significant potential for further development and implementation in the medical sector, including clinics, hospitals, and pharmacies. It is imperative to continue improving the system by adding more features, reliable hardware, and ensuring that it is user-friendly and affordable for commercial use.

VII. CONCLUSIONS

In conclusion, the implementation of the Internet of Things (IoT)- Patient Health System brings a positive impact to the healthcare industry by providing an efficient and effective means of patient health monitoring. The Internet of Things (IoT)- Patient Health System and the proposed system presented in the research both aim to remotely monitor the vital signs of patients and provide prompt medical attention when needed. The system has successfully met all the objectives and has been able to monitor patients and alert healthcare professionals in real-time. The system integrates microcontrollers, sensors, and the ThingSpeak Internet of Things (IoT)- platform with email notifications to create a systematic and efficient healthcare system. This system provides a solution to the problem of handling a large number of patients during sudden outbreaks of diseases such as Covid-19, potentially improving patient outcomes and saving lives. Similarly, the proposed Internet of Things (IoT)-based patient health monitoring system presented in the research has the potential to improve the overall quality of healthcare services and lead to better health outcomes for society. The system's ability to remotely monitor the vital signs of patients, display real-time data on an Internet of Things (IoT)- platform, and send alerts in the event of a medical emergency can reduce the workload of healthcare providers, allowing them to focus on other critical tasks. The current system can be further improved by incorporating additional sensors like oximeters and blood pressure devices to provide a comprehensive analysis of the patient's health condition. Upgrading the sensors with newer models that offer advanced features can enhance the accuracy of the collected data. In the future, implementing this system in hospitals and clinics can help address patient data management and monitoring issues. The integration of Internet of Things (IoT)- technology in healthcare paves the way for healthcare professionals to concentrate on urgent issues and emergencies, expanding their roles in the healthcare industry. The utilization of Internet of Things (IoT)- capabilities can make healthcare more efficient and effective in providing high-quality patient care.

ACKNOWLEDGMENT

The authors would like to thank Centre for Research and Innovation Management (CRIM) for the support given to this research by Universiti Teknikal Malaysia Melaka (UTeM).

REFERENCES

- [1] Kulkarni, N.I Ransing, P., Patil, S., Pawase, S., Shinde, P., Pawar, S., & Patil, S. (2022). Healthcare Monitoring System Using IoT. In *International Journal for Research in Applied Science and Engineering Technology* (Vol. 10, Issue 12, pp. 341–345). *International Journal for Research in Applied Science and Engineering Technology (IJRASET)*. <https://doi.org/10.22214/ijraset.2022.47819>.
- [2] M. Javaid and I. H. Khan, "Internet of Things (IoT) enabled healthcare helps to take the challenges of COVID-19 Pandemic," *J. Oral Biol. Craniofac. Res.*, vol. 11, no. 2, pp. 209-214, Apr.-Jun. 2021, doi: 10.1016/j.jobcr.2021.01.015.
- [3] S. S. Raouf and M. A. S. Durai, "A Comprehensive Review on Smart Health Care: Applications, Paradigms, and Challenges with Case Studies," *Contrast Media Mol Imaging*, vol. 2022, p. 4822235, Sep. 2022, doi: 10.1155/2022/4822235.
- [4] S. Abdulmalek et al., "IoT-Based Healthcare-Monitoring System towards Improving Quality of Life: A Review," *Healthcare*, vol. 10, no. 10, p. 1993, Oct. 2022, doi: 10.3390/healthcare10101993.
- [5] P. Yadav, P. Kumar, P. Kishan, P. Raj, and U. Raj, "Development of Pervasive IoT Based Healthcare Monitoring System for Alzheimer Patients," *J. Phys. Conf. Ser.*, vol. 2007, p. 012035, 2021.
- [6] A. Kishor and C. Chakraborty, "Artificial Intelligence and Internet of Things Based Healthcare 4.0 Monitoring System," *Wirel. Pers. Commun.*, pp. 1-17, 2021. [Online]. Available: <https://link.springer.com/content/pdf/10.1007/s11277-021-08708-5.pdf>.
- [7] Sucher, J. F., Moore, F. A., Sailors, R. M., Gonzalez, E. A., & McKinley, B. A. (2009). Performance of a Computerized Protocol for Trauma Shock Resuscitation. *World Journal of Surgery*, 34(2), 216–222. <https://doi.org/10.1007/s00268-009-0309-7>.
- [8] Pradhan, S., Zainuddin, A. A., Sahak, R., & Yunus, M. F. A. M. (2022). Investigation into Smart Healthcare Monitoring System in an IOT environment. *Malaysian Journal of Science and Advanced Technology*. <https://doi.org/10.56532/mjsat.v2i2.53>.
- [9] Almotiri, S. H., Khan, M.I A., & Alghamdi, M. A. (2016). Mobile Health (m-Health) Sys-tem in the Context of IoT. In 2016 IEEE 4th International Conference on Future Internet of Things and Cloud Workshops (FiCloudW). 2016 IEEE 4th International Conference on Future Internet of Things and Cloud Workshops (FiCloudW). IEEE. <https://doi.org/10.1109/w-ficloud.2016.24>.
- [10] Fedorchenko, I., Oliinyk, A., Stepanenko, A., Svyrydenko, A., Goncharenko, D. "Genetic method of image processing for motor vehicle recognition". 2019 2nd International Workshop on Computer Modeling and Intelligent Systems, CMIS, 2019, Zaporizhzhia, April 15-19, CEUR Workshop Proceedings, Vol. 2353, pp. 211-226.
- [11] Abdullah, A., Ismael, A., Rashid, A., Abou-Elnour, A., & Tarique, M. (2015). Real Time Wireless Health Monitoring application using mobile devices. *International Journal of Computer Networks & Communications*, 7(3), 13–30. <https://doi.org/10.5121/ijcnc.2015.7302>.
- [12] Fedorchenko, I., Oliinyk, A., Stepanenko, Zaiko, T., Korniienko S., Kharchenko, A. Construction of a genetic method to forecast the population health indicators based on neural network models // *Eastern-European Journal of Enterprise Technologies*, 2020, 1 (4-103), P. 52–63. DOI: 10.15587/1729-4061.2020.197319.
- [13] Almaiah, M. A. (2020). Multilayer Neural Network based on MIMO and Channel Estimation for Impulsive Noise Environment in Mobile Wireless Networks. In *International Journal of Advanced Trends in Computer Science and Engineering* (Vol. 9, Issue 1, pp. 315–321). The World Academy I of Research in Science and Engineering. <https://doi.org/10.30534/ijatcse/2020/48912020>.
- [14] Fedorchenko, I., Oliinyk, A., Goncharenko, D., Stepanenko, A., Fedoronchak, T., Kharchenko, A., Langendorfer, P. "Development of a Genetic Method for the Recognition of Medical Packaging". 2021 IEEE 8th International Conference on Problems of Infocommunications,

- Science and Technology, PIC S and T 2021 - Proceedings, 2021, pp. 245 - 250. doi: 10.1109/PICST54195.2021.9772213.
- [15] Masud, M., I Muhammad, G., Alhumyani, H., Alshamrani, S. S., Cheikhrouhou, O., Ibrahim, S., & Hossain, M. S. (2020). Deep learningbased intelligent face recognition in IoT-cloud environment. In *Computer Communications* (Vol. 152, pp. 215–222). Elsevier BV <https://doi.org/10.1016/j.comcom.2020.01.050>.
- [16] Iranpak, S., Shahbahrami, A. & Shakeri, H. Remote patient monitoring and classifying using the internet of things platform combined with cloud I computing. *J Big Data* 8, 120 (2021). <https://doi.org/10.1186/s40537-021-00507-w>.
- [17] Naghshvarianjahromi, M., Kumar, S., & Deen, M. J. (2019). BrainInspired Intelligence for Real-Time Health Situation Understanding in Smart e-Health Home Applications. In *IEEE Access* (Vol. 7, pp. 180106–180126). I Institute of Electrical and Electronics Engi-neers (IEEE). <https://doi.org/10.1109/access.2019.2958827>.
- [18] Chee Yuan, L., Kong, L., Tunku, U., & Rahman, A. (2019). A project report submitted in partial fulfillment of the requirements for the award of Bachelor of Engineering (Honours) Electrical and Electronic Engineering.
- [19] Yew, H. T., Ng, M. F., Ping, S. Z., Chung, S. K., Chekima, A., & Dargham, J. A. (2020). IoT Based Real-Time Remote Patient Monitoring System. In *2020 16th IEEE International Colloquium on Signal Processing & Its Applications (CSPA). 2020 16th IEEE Inter-national I Colloquium on Signal Processing & Its Applications (CSPA)*. IEEE. <https://doi.org/10.1109/cspa48992.2020.9068699>.
- [20] T. Akhila, V., Vasavi, Y., Nissie, K., & Rao, P. v. (2018). An IoT based Patient Health Monitoring System using Arduino Uno.
- [21] Emayavaramban, A thingspeak IOT on Real Time Room Condition Monitoring System. (2020). Retrieved January 26, 2023, from <https://ieeexplore.ieee.org/document/9140127/>.
- [22] Internet of things and Nodemcu a review of use of NODEMCU Esp8266 (2019). Retrieved January 26, 2023, from [https://www.researchgate.net/publication/337656615_Internet_of_Thin gs_and_Nodemcu_A_review_of_use_of_Nodemcu_ESP8266_in_IoT products](https://www.researchgate.net/publication/337656615_Internet_of_Thin gs_and_Nodemcu_A_review_of_use_of_Nodemcu_ESP8266_in_IoT_products).
- [23] Mohamad Hadis, N. S., Amirnarazullah, M. N., Jafri, M. M., & Abdullah, S. (2020). IoT Based Patient Monitoring System using Sensors to Detect, I Analyse and Monitor Two Primary Vital Signs. In *Journal of Physics: Conference Series* (Vol. 1535, Issue 1, p. 012004). IOP Publishing. <https://doi.org/10.1088/1742-6596/1535/1/012004>.
- [24] Riazul Islam, S. M., Daehan Kwak, Humaun Kabir, M., Hossain, M., & Kyung-Sup Kwak. (2015). The Internet of Things for Health Care: A Comprehensive Sur-vey. In *IEEE Access* (Vol. 3, pp. 678–708). Institute of I Electrical and Electronics Engineers (IEEE). <https://doi.org/10.1109/access.2015.2437951>.
- [25] Swaroop, K. N., Chandu, I K., Gorrepotu, R., & Deb, S. (2019). A health monitoring system for vital signs using IoT. *Internet of Things*, 5, 116– 129. <https://doi.org/10.1016/j.iot.2019.01.004>.
- [26] Ijrst, I. J. of S. R. in S. and T. (2021). Review on Arduino Based Wireless Health Monitoring System for covid-19 Patients. *International Journal of Scientific Research in Science and Technology*. https://www.academia.edu/49658934/Review_on_Arduino_Based_Wir eless_Health_Monitoring_System_for_covid_19_Patients.
- [27] Senthamilarasi, C., Rani, J. J., Vidhya, B., & Aritha, H. (2018). A SMART PATIENT HEALTH MONITORING SYSTEM USING IOT. <http://www.acadpubl.eu/hub/>.
- [28] Gulraiz Joyia, G. J., I Liaqat, R. M., Farooq, A., & Rehman, S. (2017). Internet of Medical Things (IOMT): Applications, Benefits and Future Challenges in Healthcare Domain. In *Journal of Communications. Engineering and Technology Publishing*. <https://doi.org/10.12720/jcm.12.4.240-247>.
- [29] Perumal, K., & Manohar, I M. (2017). A Survey on Internet of Things: Case Studies, Applications, and Future Directions. In *Studies in Big Data* (pp. 281–297). Springer International Publishing. https://doi.org/10.1007/978-3-319-53472-5_14.
- [30] Rajendran T & Sridhar. (2021). Epileptic Seizure: Classification Using Autoregression Features. I In *International Journal of Current Research and Review* (Vol. 13, Issue 04, pp. 123–131). Radiance Research Academy. <https://doi.org/10.31782/ijcr.2021.13429>.
- [31] Louis, L. (2016). Working principle of Arduino and using it as a tool for study and research: Sciencegate. *International Journal of Control, Automation, Communication and Systems*. Retrieved January 26, 2023, from <https://www.sciencegate.app/document/10.5121/ijcacs.2016.1203>.

Integrated Methodology for Information Security Risk Management using ISO 27005:2018 and NIST SP 800-30 for Insurance Sector

Arief Prabawa Putra¹, Benfano Soewito²

Computer Science Department-BINUS Graduate Program-Master of Computer Science, Bina Nusantara University, Jakarta, Indonesia 11480^{1,2}

Abstract—The development of Information and Communication Technology (ICT) in the Industrial Revolution 4.0 era shows very fast and disruptive developments that encourage increased use of Information Technology (IT) services within organizations. However, there is a risk of creating vulnerabilities and threats to owned information systems. Plans and strategies are required to implement information security risk management to address vulnerabilities in threat events. This research is a case study of the Enterprise Resource Planning System in the Insurance Sector. The proposed methodologies for integrating information security risk management using ISO/IEC 27005:2018 as a risk management framework and NIST SP 800-30 Rev. 1 as guidance for risk assessments. The risk evaluation stage is the process of comparing the results of the risk analysis with the risk criteria to then determine whether the risk rating is acceptable or tolerable. For risk treatment and control using the ISO/IEC 27002:2022 framework.

Keywords—Risk management; information security; ISO/IEC 27005; NIST SP 800-30; ISO/IEC 27002

I. INTRODUCTION

The adoption of information and communication technology (ICT) during the fourth industrial revolution 4.0 shows very rapid and disruptive advances that support the growth of information technology services [1]. The use of ICT affects the development of a country through the development of the dissemination of knowledge, especially from developed and developing countries, and through innovations [2]. Almost all enterprises use ICT to obtain information or process data more quickly, precisely, and accurately [3].

However, along with the use of the internet, there are many sources of threats that come from inside and outside the organization. In different parts of the world, the number of cyberattacks is growing at an alarming rate and causing financial losses [4]. This can threaten the continuity of business activities in the organization, including the insurance sector.

ZZZ Insurance is part of the government of the Indonesian insurance sector, which uses an Enterprise Resource Planning (ERP) system to run its business. ERP is an integrated system used by organizations to manage day-to-day business activities, such as financial and accounting, cash management, procurement, and asset management. Their system integrates not only internal parties but also external parties, such as health social security agencies, hospitals, population and civil

registration agencies, and aggregators. One aspect of ICT that needs to concern every organization is information security. Information security can be formed by implementing controls, which include policies, processes, procedures, organizational structures, and functions of software and hardware [5]. These controls need to be implemented, monitored, reviewed, and required to ensure security and achieve the goals of the business organization.

Therefore, steps are needed in detecting vulnerabilities to threats inside and outside the organization. Therefore, procedures must be made to discover vulnerabilities to both internal and external threats. In addition, the importance of implementing organizational risk management in the software lifecycle is to produce proper supervision and responsibility and increase effectiveness and efficiency [6]. Planning and strategy are needed to overcome these vulnerabilities if a risk or threat occurs.

The need for planning and strategies to overcome these vulnerabilities in the event of a risk or threat if it disrupts Business Continuity (BC) [7]. Implementing Business Continuity Planning (BCP) is critical in an organization to anticipate if a disaster occurs and ensure that the business can continue to operate, or at the very least, that the organization can continue to provide its services after the disaster [8].

Therefore, it is necessary to implement information security risk management in enterprise systems organizations. Many standards are used in the implementation of risk management, including ISO 27005, which is widely applied in profit and non-profit organizations in other countries. Based on recommendations from another study, ISO 27005 is one of the international standards that is easy to implement in providing guidelines for information security risk management [9], [10]. Other studies also use the NIST 800-30 standard, where integrating qualitative and quantitative methodologies to give accurate and reliable risk data for decision making is the optimal method for risk assessment [11].

The purpose of this research was to conduct an information security risk assessment using ISO/IEC 27005:2018 as a risk management framework and NIST SP 800-30 revision 1 as a reference matrix of qualitative and quantitative risk levels. In accordance with the ISO/IEC 27002:2022 framework, the recommended controls include risk treatment and risk acceptance. This guideline is used to determine and implement information security risk

management controls inside an information security management system (ISMS) based on ISO/IEC 27001 [12].

II. LITERATURE REVIEW

A. Risk Management

Risk management is the process of identifying risks, assessing their relative magnitudes, and taking steps to reduce them to an acceptable level [13]. Information security risk management is a practical step in managing the risk of an organization's information security and aims to provide protection for organizational information and assets [14]. Risk management has three main processes: risk identification, risk assessment, and risk control.

The implementation of risk management in non-profit organizations provides several benefits, including planning basic information technology resources, providing decision-making support systems for leaders, and improving operational performance in terms of the maturity level of the risk management process. The combination of risk management processes based on ISO 31000:2018 and ISO 9001:2015 aims to provide guidelines for risk management principles and their application processes at the organizational, strategic, and operational levels [15]. Implementation of Enterprise Resources Planning (ERP) in the organization will increase the added value of the organization and support decision-making management [16].

Several other studies regarding the application of risk management standards in government agencies include the design of information security management for data communication applications at the XYZ Institute using ISO 27005 and NIST SP-800-30 [17]. Information security risk assessment with a combination of ISO 27005 information security standards and National Institute of Standards and Technology (NIST) SP 800-30 revision 1 adapted to organizational conditions [18]. The importance of conducting a cybersecurity risk assessment of the heart of electronic devices to determine the severity of the threat, prioritize the most significant risks and ensure effective risk management using a combination of ISO/IEC 27005 and NIST SP 800-30 [19]. The implementation of information security risk management (ISRM) in government agencies at the Bali Regional Police regarding System-Based Electronic Governance (EBGS) is an attempt to protect the risk of valuable assets [20]. In the industrial era 4.0, cloud computing has become widely used in the government sector, so security is now part of risk management [21]. The proposed cloud computing security model includes data security, risk assessment, regulation, compliance, and requirements.

An organization knows the importance of implementing risk management and making it the main step in minimizing the risks that will occur. The successful implementation of risk management in government agencies is influenced by several factors, namely risk management, policy development, and policy compliance [22].

B. ISO/IEC 27005

ISO/IEC 27005 is part of the ISO 27000 series. ISO 27005:2018 is a standard used to provide guidance for information security risk management [14]. ISO 27005 supports the general concepts described in ISO 27001 and is designed to assist in the proper implementation of information security based on a risk management approach. ISO 27005 has stages, namely context establishment, risk assessment (risk identification, risk analysis, risk evaluation), risk treatment, risk acceptance, risk communication and consultation, and monitoring and review.

C. NIST SP 800-30

NIST (National Institute of Standards and Technology) Special Publication (SP) 800-30 is a guide that aims to provide risk assessment of organizational and government information systems and is a complement to NIST SP 800-39 guidelines. The latest version of NIST SP 800-30 is revision 1. The risk assessment approach for NIST SP 800-39 revision 1 is supported by security standards and other guidelines to manage information security risks. The risk assessment approach to NIST SP 800-39 revision 1 is supported by security standards and other guidelines to manage information security risks [23]. NIST SP 800-30 can be used to complement the ISO 27005 standard in conducting risk assessments.

NIST 800-30 provides a basis for the development of an effective risk management program, as well as a definition and practical guidance for assessing and mitigating a risk that exists in an IT system. This framework has nine steps of risk management activities, starting with system characterization, threat identification, vulnerability identification, control analysis, likelihood determination, impact analysis, risk determination, control recommendations, and the results document of the risk assessment report.

III. RESEARCH DESIGN

This research uses a case study of ZZZ Insurance ERP system. As seen in Fig. 1, the proposed method uses the ISO/IEC 27005:2018 framework as the main risk management framework is integrated with the NIST SP 800-30 revision 1 risk assessment guideline. Recommendations for control using are the ISO/IEC 27002 framework.

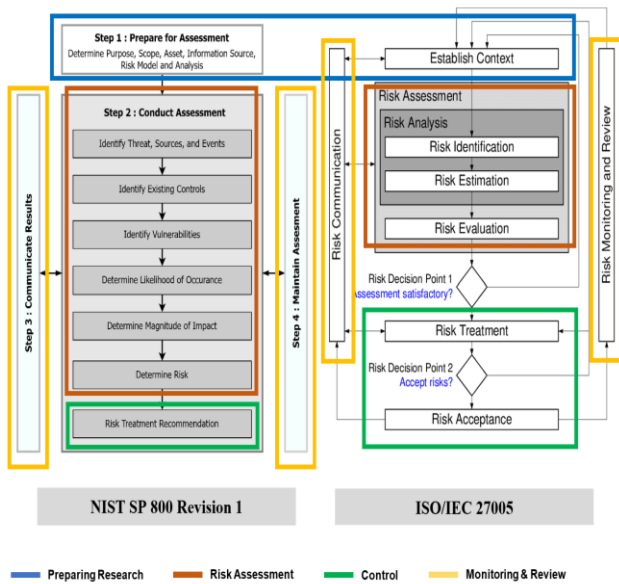


Fig. 1. Integrated methodology.

Beginning with research preparation includes identifying problems, collecting data, and context organization. In identifying problems, it is necessary to explain the problems faced by the organization, which are presented in the form of background and problem formulation, as well as solutions to the problems faced, which are presented in the form of goals and benefits.

The next stage is the collection of data obtained from document reviews, interviews, Forum Group Discussions (FGD), and observations. Document review is conducted to understand the organization in detail. During the interviews and FGD stages, it was conducted to find out the conditions and needs for information security in the organization.

Next, define the context of the organization to consider the impact that both internal and external factors have on the company's operational activities and its ability to achieve targets. The risk management process must be aligned with the corporate culture, processes, structure, and strategy. To prepare a risk assessment, it is necessary to first set the organizational context.

Risk assessment is a structured approach to identifying and analyzing uncertainties that exist in the achievement of organizational goals. Based on interviews with risk owners and IT risk officers in the organization, risk assessment aims to:

- Recognize the risks that may occur in the organization.
- Understand the risk so that the significance of the risk can be assessed, and the level of risk can be evaluated based on the organization's risk criteria.

- Identify possible risks that can be accepted or modified.
- Considering the relative impact of various risk-reduction treatment options.

At the risk treatment stage, risk scenarios that trigger risk appetite are mitigated and prioritized to receive risk treatment in the form of information security control recommendations and information security target setting based on ISO/IEC 27002:2022. The results of a series of risk management design stages are communicated to top management to confirm their suitability.

IV. RESEARCH AND DISCUSSION

A. Context Establishment

Risk criteria are used to rank risk levels as unacceptable or acceptable. Risk criteria can include several limits with a target risk scale that are adjusted to the needs of the organization. Based on the results of the interview with management, this research refers to the NIST SP 800-30 revision 1 framework standard based on the level of risk shown in Table I.

TABLE I. RISK SCALE

Scala	Description	Semi Quantitative Value	
Very High	Have a negative impact	$4,0 < x \leq 5,0$	5
High	Almost certainly have a negative impact	$3,0 < x \leq 4,0$	4
Moderate	A medium probability results in a negative impact.	$2,0 < x \leq 3,0$	3
Low	A Small probability gives a negative impact	$1,0 < x \leq 2,0$	2
Very Low	A low probability has a negative impact.	$x \leq 1,0$	1

In determining the context, impact criteria using the level option are based on the level description in NIST SP 800-30, shown in Table II. Likelihood criteria using impact considerations that allow the threat to occur, as well as the possibility of starting or occurring, are listed in Table III.

TABLE II. IMPACT OF THREAT EVENT

Scala	Description	Value
Very High	Several severe or catastrophic negative effects on organizational operations, assets, individuals, or the nation	5
High	Severe or catastrophic adverse effect on organization operation, assets, individuals, or the nation.	4
Moderate	Serious adverse effect on an organization's operations, assets, individuals, or the nation.	3
Low	Limited effect on organization operation, assets, individuals, or the nation.	2
Very Low	Negligible adverse effect on organization operations, assets, individuals, or the nation.	1

TABLE III. LIKELIHOOD OF THREAT EVENT RESULTING IN ADVERSE IMPACTS

Scala	Description	Value
Very High	If the threat event is initiated or occurs, it is almost certain to have adverse impacts.	5
High	If the threat event is initiated or occurs, it is highly likely to have adverse impacts.	4
Moderate	If the threat event is initiated or occurs, it is somewhat likely to have adverse impacts.	3
Low	If the threat event is initiated or occurs, it is unlikely to have adverse impacts.	2
Very Low	If the threat event is initiated or occurs, it is highly unlikely to have adverse impacts.	1

B. Risk Assessment

At this stage, an assessment of the identified risks has been carried out, and an evaluation has been carried out for each risk scenario [24].

1) *Risk identification*: Risk identification is the process of finding, recognizing, or describing risk attributes. Risk identification includes identification of risk sources, events, and causes in the organization.

a) *Asset identification*: The process of asset identification begins with a weighted factor analysis of all ERP assets. Each information asset is scored for each critical factor and assigned a weight for each criterion. The weighting value is obtained from risk owner and IT risk officer in the organization.

To calculate weighted factor analysis, where each asset is scored for a critical factor and given a weight for each criterion. Criteria of weighted factor analysis consist of criterion 1 (impact to revenue – 30%), criterion 2 (impact to profitability – 40%), and criterion 3 (impact to public image – 30%). For scoring a critical factor, scores range from 0.1 to 1.0, and criteria are weighted from 1 to 100; each is weighted to indicate the importance of the criteria set for the organization [13]. The range of values obtained with reference to NIST SP 800-30 revision 1. Table IV shows that example of a weighted factor analysis worksheet.

TABLE IV. WEIGHTED FACTOR ANALYSIS WORKSHEET

Information Asset	Criterion 1 (30)	Criterion 2 (40)	Criterion 3 (30)	Weighted Score
IT Procedures and Policies	0,8	0,9	0,6	69
Network Topology	0,4	0,4	0,4	40
Server	0,8	0,9	0,7	81
ERP System	0,8	0,9	0,7	81
...				
IT Operation	0,8	0,8	0,7	77

In this discussion, we first identify two types of assets based on organization conditions, namely main assets consisting of business processes and activities, and information, and supporting assets consisting of all assets:

hardware, software, network, site, personnel, and organizational structure [14].

Based on asset identification, the total number of identified primary assets is 9 assets, consisting of 3 business process and activities assets and 6 information assets. For identified supporting assets, 53 assets consist of 14 hardware assets, 18 software assets, 2 network assets, 11 site assets, and 8 personnel assets. Table IV is an example of asset identification with the following asset codes: A1-IT Procedures and Policies, A2-Network Topology, Source Code A3, Server-A4, A5-ERP System, until A62-IT Operation, as shown in Table V.

TABLE V. ASSET IDENTIFICATION

Asset Code	Asset Type	Asset Category	Risk Owner	Location
A1	Primary Asset	Business Processes and Activities	Head of ICT Division	Head Office
A2	Primary Asset	Information	Infrastructure Department	Head Office
A3	Primary Asset	Information	System Department	Head Office
A4	Supporting Asset	Hardware	Infrastructure Department	Head Office
A5	Supporting Asset	Software	System Department	Data Center
...				
A62	Supporting Asset	Personnel	Head of ICT Division	Head Office

b) *Threat Identification*: This research divides threat sources into two categories: adversarial and non-adversarial threat sources. Threat sources identified in this research were obtained from 12 adversarial sources and 18 non-adversarial sources.

In this discussion, the adversarial threat sources are as follows: S1: distributed denial of service, S2: injection, S3: intrusion, S4: malware, S5: social engineering, S6: sniffing, spoofing, or phishing, S7: website attack, S8: employee, S9: external stakeholder, S10: lack of employees, S11: unauthorized access, and S12: failure to maintain physical facilities.

Non-adversarial threat sources are as follows: S13: error requirement and design system, S14: human error (least privilege), S15: human error (personnel IT), S16-limited budget allocation for training, S17: obsolete technology, S18: lack of monitoring and control, S19: lack of expertise, skills, and employee behavior, S20: lack of employee information security awareness, S21: insecure password, S22: the application crashes, S23: error connection database, S24: operation system crashes, S25: web server crashes, S26: failure to backup data, S27: broken communication data, S28: failure data, S29: limited storage media, S30: devices end of support, S31: short-circuit, S32: power supply failure, S33: unstable power supply, S34: interruption of service from the provider, S35: rodent, S36: overhead, S37: maintenance fiber optics; S38: fire, S39: earthquake, and S40: thunderbolt.

Then, identify all threats that interfere with information security aspects such as Confidentiality (C), Integrity (I) and

Availability (A) on assets that have been identified. The following are questions to ask when identifying threats:

- What are the threats to the asset that you know or suspect?
- What are the most dangerous threats to the organization?
- What are the most expensive threats to recover from in the event of an attack?
- What are the threats that require the greatest expenditure to prevent them?

After getting the source of the threat in the ERP System, then identify this threat. For each asset, 35 threats have been identified, with different sources of threat. Table V explained that in the ERP system, T1: errors in making policies, procedures, or other relevant documents; T2: dissemination of information by unauthorized parties; T3: cybercrime; T4: broken access control; T5: failure of hardware, a network device, or physical facility assets; and until T35: unauthorized access.

Threat events are obtained from threat sources that have been defined through event logs and interviews with IT risk officers. Relevance was obtained according to NIST SP 800-30 revision 1, shown in Table VI.

TABLE VI. THREAT IDENTIFICATION

Asset Code	Threat Event	Threat Source	CIA
A1	T1	S8, S9, S18, S19	C+I
A2	T2	S8, S9, S15	C
A3	T2	S8, S9, S15	C
A4	T3	S1, S2, S3, S4, S5, S6	C+I+A
	T5	S11, S15, S27, S29, S31, S32	A
A5	T3	S2, S3, S4, S6, S7	C+I+A
	T4	S8, S9, S14, S21, S30	C+I+A
	T5	S13, S22, S23, S24, S25	I + A
...			
A62	T35	S8, S15, S18, S20	C+I+A

c) *Identification of existing controls:* The next step is to identify the security controls that the company has implemented to protect the organization's assets from threats. In this discussion through the observation method obtained 52 security controls on assets. For example, of the existing control in this case, namely: C1: periodically review internal IT policies, procedures, and circulars, C2: classified information (controlled restricted, unclassified information, controlled, and public), C3: non-disclosure agreement, C4-restriction of access control, C5: information security awareness, education, and training program, C6: rollback procedure, C7: periodic review of access rights, C8: using a strong password according to best practice recommendations, C9: implemented least privileges, C10: log access control

failures, C11: periodic maintenance of physical assets, C12: apply periodic updates or firmware to the most recent hardware, network devices, and software versions, and until C52: monitoring and controlling.

d) *Identification of vulnerabilities:* Identification of vulnerabilities means the extent to which the company has implemented controls to protect assets from threats. Vulnerabilities that have no corresponding threat may not require the implementation of control, but they do need to be identified and monitored. However, implementing ineffective controls or controls that don't work properly can be a vulnerability. In obtaining our vulnerability results, we used vulnerability sources references from the OWASP top ten [25].

The results of the study found 46 vulnerabilities, and Table IV shows the vulnerabilities for each asset based on the controls that have been implemented. The following is an example of controls in this discussion: V1: ineffective implementation of information security policies, V2: unencrypted documents and files, V3: vulnerable and end of support components, V4: insecure design system, V5: cryptographic failures, V6: no backup components, V7: software or hardware misconfigurations; and until V46: lack of information security practices.

TABLE VII. IDENTIFICATION OF VULBERABILITIES

Asset code	Existing controls	Vulnerability	Severity
A1	C1	V1	Low
A2	C2, C3	V2	Low
A3	C2, C3	V2	Low
A4	C5, C7, C8, C9, C12	V4	Moderate
	C6	V3, V5	Moderate
A5	C5, C7, C8, C9, C12	V4	High
	C4, C10	V5	Moderate
	C11, C52	V3, V5, V7	Moderate
...			
A62	C4, C7, C52	V46	Moderate

2) *Risk Analysis:* Risk analysis is the activity of mapping assets, asset values, threats, security controls, vulnerabilities, and impacts on CIA aspects. Risk analysis is intended to obtain the results of an impact assessment and identify possible information security risks.

In this research, we calculate the risk with a formula [13]: Risk is the probability of a successful attack on the organization (loss frequency = likelihood * attack success probability) multiplied by the expected loss from a successful attack (loss magnitude = asset value * probable loss) plus the uncertainty of estimates of all stated values.

Loss frequency is a measurement of the likelihood of an attack combined with the probability that it will succeed if it targets an organization. Loss magnitude is a combination of the asset value and the likelihood of its loss in an attack.

As shown in Table VIII, the risk analysis obtained 142 moderate level, 97 at low level, and 13 at very low level. levels of risk in the ERP system, with 2 at high level, 30 at

TABLE VIII. RISK ANALYSIS

Asset Code	Threat event	Threat Source	CIA	Existing Control	Vulnerability	Risk	Level of Risk
A1	T1	S8, S9, S18, S19	C+I	C1	V1	0,20	Very Low
A2	T2	S8, S9, S15	C	C2, C3	V2	1,44	Low
A3	T2	S8, S9, S15	C	C2, C3	V2	1,64	Low
A4	T3	S1, S2, S3, S4, S5, S6	C+I+A	C4, C5, C7, C8, C9, C12	V4	2,88	Moderate
	T5	S11, S15, S27, S29, S31, S32	A	C6	V3, V5	2,23	Moderate
A5	T3	S2, S3, S4, S6, S7	C+I+A	C4, C5, C7, C8, C9, C12	V4	3,12	High
	T4	S8, S9, S14, S21, S30	C+I+A	C4, C9, C10	V5	2,11	Moderate
	T5	S13, S22, S23, S24, S25	I+A	C11, S2	V3, V5, V7	2,47	Moderate
...							
A62	T35	S8, S15, S18, S20	C+I+A	C4, C7, C52	V46	2,73	Moderate

3) Risk Evaluation: Risk evaluation in this discussion aims to compare the results of risk analysis with risk criteria and then determine whether the risk rating is acceptable or tolerable. The stages of risk evaluation include compiling risk priorities based on the amount of risk, provided that:

- The highest level of risk gets the highest priority.
- If there is more than one risk with the same risk magnitude, the risk priority is determined based on the sequence of impact areas from the highest to the lowest according to the loss magnitude.
- If there is still more than one risk that has the same magnitude and area of impact, then the risk priority

is determined based on the order of the highest to the lowest risk category according to the loss frequency.

- If there is still more than one risk that has the same magnitude, loss magnitude, and loss frequency, then the risk priority is determined based on the judgment of the risk owner.

Table IX shows as risk appetite based on semi-quantitative based risk rating guidelines with NIST SP 800-30 revision 1 and two risk treatment criteria (likelihood and overall level of impact).

TABLE IX. RISK APPETITE

Overall likelihood (Threat event occurs and result in adverse impact)	Level of impact				
	Very Low (1)	Low (2)	Moderate (3)	High (4)	Very High (5)
Very High (5)	Accept	Mitigation	Mitigation	Mitigation	Mitigation
High (4)	Accept	Mitigation	Mitigation	Mitigation	Mitigation
Moderate (3)	Accept	Accept	Mitigation	Mitigation	Mitigation
Low (2)	Accept	Accept	Accept	Mitigation	Mitigation
Very Low (1)	Accept	Accept	Accept	Accept	Accept

Risk determination is the first step before risk prioritization. Priority risk matrix is classified based on NIST

SP 800-30 revision 1 and is a matrix of the relationship between assets and threats, show as in Fig. 2.

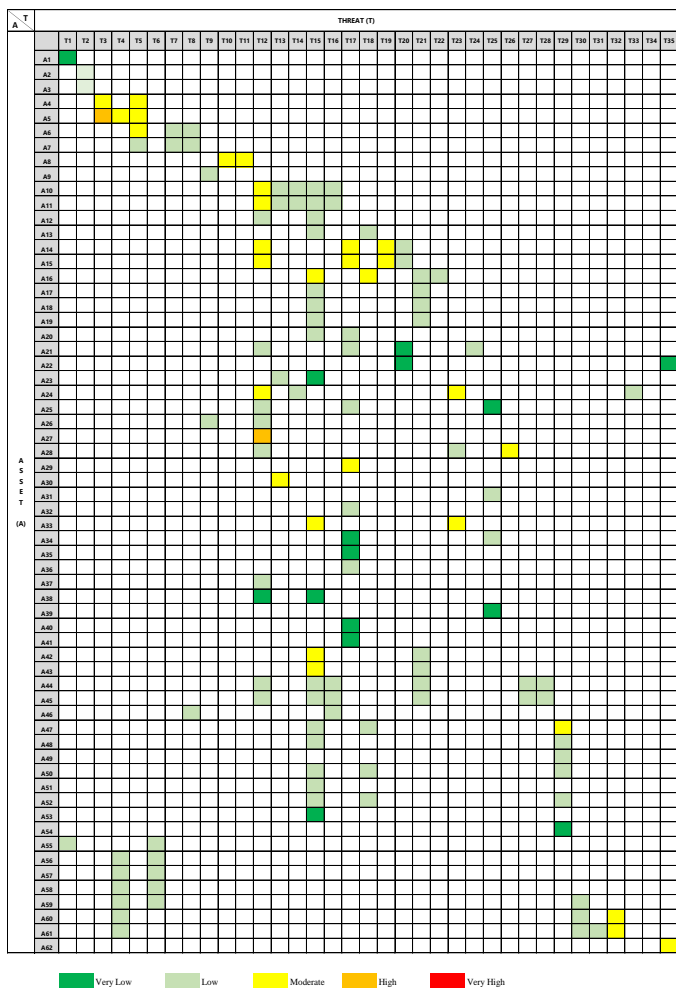


Fig. 2. Risk matrix.

C. Risk Treatment and Risk Acceptance Strategy

1) Risk treatment strategy: Risk treatment aims to control dangerous risks by developing relevant treatments to control the causes of risk, measuring the effectiveness of the treatment, and if the estimated risk value remains at an intolerable level, preparing alternative treatments.

According to ISO/IEC 27005:2018, there are four options available for risk treatment, namely risk modification, risk avoidance, risk sharing, and risk retention. In this discussion, we found 142 risks with unacceptable decisions for 32 modification risks. Total risk acceptance is 110, of which there are 90 risk retention, 2 risk avoidance, and 18 risk sharing.

In selecting the risk treatment that has been sorted based on risk priority from the highest to the lowest risk level. The following is an example of a risk priority in Table X.

TABLE X. RISK TREATMENT

Priority	Risk Scenario	Level of Risk	Decision	Risk Appetite
1	A5, T3	High	Mitigation	Risk Modification
2	A4, T3	Moderate	Mitigation	Risk Modification
3	A62, T35	Moderate	Mitigation	Risk Modification
...				
142	A1, T1	Very Low	Accept	Risk Retention

2) Risk Acceptance Strategy: This activity is carried out to describe more clearly some of the security controls that have been selected for risk treatment. In this discussion, we propose that an information security team be created to define the roles and responsibilities, or Person in Charge (PIC), of information security activities in every organization.

In establishing information security controls, PIC is required to be responsible for risk acceptance, as shown in Table XI.

TABLE XI. RISK ACCEPTANCE

Priority	Risk Scenario	Control with ISO/IEC 27002:2022	PIC
1	A5, T3	<p>Organizational controls: 5.1 Policies for information security 5.37 Documented operating procedures.</p> <p>People controls: 6.3 Information security awareness, education, and training</p> <p>Technology controls: 8.7 Protection against malware 8.20 Networks security 8.23 Web filtering</p>	Head of Department Information Technology System
2	A4, T3	<p>Organizational controls: 5.1 Policies for information security 5.37 Documented operating procedures.</p> <p>People controls: 6.3 Information security awareness, education, and training</p> <p>Technology controls: 8.7 Protection against malware 8.20 Networks security 8.32 Change management</p>	Head of Department Information Technology Infrastructure and Service
3	A62, T35	<p>Organizational controls: 5.1 Policies for information security 5.17 Authentication information</p> <p>People controls: 6.2 Terms and conditions of employment 6.3 Information security awareness, education, and training</p> <p>Physical controls: 7.7 Clear desk and clear screen</p> <p>Technology controls: 8.2 Privileged access rights</p>	Head of Department Information Technology Infrastructure and Service
...			

32	A5, T4	<p>Organizational controls: 5.1 Policies for information security 5.15 Access control</p> <p>People controls: 6.3 Information security awareness, education, and training</p> <p>Physical controls: 7.7 Clear desk and clear screen</p> <p>Technology controls: 8.2 Privileged access rights 8.3.2 Secure coding</p>	Head of Department Information Technology Infrastructure and Service
----	--------	--	--

According to research findings, there are 10 controls on 14 types of threats for organizational control (clause 5), 5 controls on 7 types of threats for people control (clause 6), 3 controls on 4 types of threats for physical control (clause 7) and 14 controls on 11 types of threats for technology control (clause 8), as shown in Fig. 3.

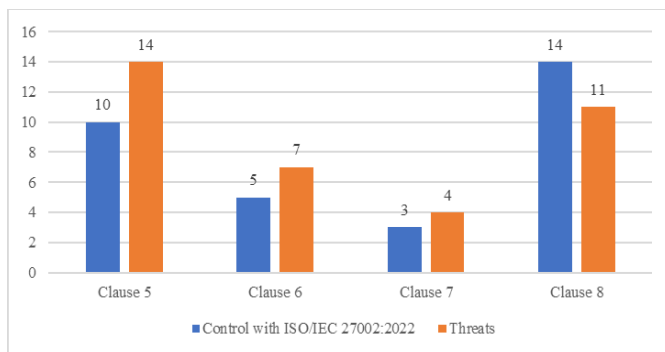


Fig. 3. Determination controls on risk categories.

For documentation and monitoring risks, we are using a risk register. The risk register provides holistic information about risks and enables stakeholders to make decisions regarding those risks and their management. The risk owner or PIC uses the risk register to document and manage risks to asset organizations.

V. CONCLUSION

Risk management principles are something to consider when preparing the framework and forming the foundation for risk management practices. The risk management process must also be aligned with the corporate culture, processes, structure, and strategy. We use an integrated ISO/IEC 27005:2018 and NIST SP 800-30 revision 1 framework, to make it easier to implement in organizations.

Context establishment is the process of determining the basic parameters in risk management by providing an understanding of the internal and external environments in management implementation. The risk assessment has three stages, namely risk identification, risk analysis, and risk evaluation. Risk identification includes risk sources, events, and causes, and impacts on each asset. Risk identification is sourced from historical data, theoretical analysis, expert opinion, and stakeholder needs.

Risk analysis is a systematic process to determine how often an event occurs, the risk of the impact that might occur, and the size of the consequences that arise from the event. The

results of risk analysis obtained: 142 levels of risk with a high level (high) of 2 risks, a moderate level (moderate) of 30 risks, a low level (low) of 97 risks, and 13 levels of very low (very low).

Risk evaluation is the process of comparing the results of risk analysis with risk criteria to then determine whether the risk rating is acceptable or tolerable. There are 142 identified risk priorities, of which 110 are acceptable and 32 are unacceptable. We have developed this research up to the risk treatment stage by providing control recommendations using ISO 27002:2022 guidelines. The risk management strategy in this case aims to eliminate the threat of risk so that it does not become an obstacle in efforts to achieve organizational goals.


Risk treatment option depends on the risk appetite and risk tolerance. In this case, we use several options to deal with these risks, such as risk modification to reduce risk through selecting controls so that the risk is acceptable, avoiding risks by avoiding activities or conditions that may pose risks, sharing risks by working with third parties who are able to deal with risks, and accepting risks without taking further action.

REFERENCES

- [1] J. Tupa, J. Simota, and F. Steiner, "Aspects of Risk Management Implementation for Industry 4.0," *Procedia Manuf.*, vol. 11, pp. 1223–1230, 2017, doi: 10.1016/j.promfg.2017.07.248.
- [2] I. Appiah-Otoo and N. Song, "The impact of ICT on economic growth-Comparing rich and poor countries," *Telecomm Policy*, vol. 45, no. 2, Mar. 2021, doi: 10.1016/j.telpol.2020.102082.
- [3] D. W. Jorgenson and K. M. Vu, "The ICT revolution, world economic growth, and policy issues," *Telecomm Policy*, vol. 40, no. 5, pp. 383–397, May 2016, doi: 10.1016/j.telpol.2016.01.002.
- [4] Sharif, M. H. U., & Mohammed, M. A., "A literature review of financial losses statistics for cyber security and future trend," *World Journal of Advanced Research and Reviews*, vol. 15, no. 1, pp. 138–156, Jul. 2022, doi: 10.30574/wjarr.2022.15.1.0573.
- [5] ISO, ISO/IEC 27002:2022, Information security, cybersecurity and privacy protection — Information security controls, 2022. [Online]. Available: <https://www.iso.org/standard/75652.html>.
- [6] J. Masso, F. J. Pino, C. Pardo, F. Garcia, and M. Piattini, "Risk management in the software life cycle: A systematic literature review," *Computer Standards and Interfaces*, vol. 71. Elsevier B.V., Aug. 01, 2020, doi: 10.1016/j.csi.2020.103431.
- [7] M. Niemimaa, J. Järveläinen, M. Heikkilä, and J. Heikkilä, "Business continuity of business models: Evaluating the resilience of business models for contingencies," *Int J Inf Manage*, vol. 49, pp. 208–216, Dec. 2019, doi: 10.1016/j.ijinfomgt.2019.04.010.
- [8] K. Charoenthammacheke, N. Leelawat, J. Tang, and A. Kodaka, "Business continuity management: A preliminary systematic literature review based on sciencedirect database," *Journal of Disaster Research*, vol. 15, no. 5. Fuji Technology Press, pp. 546–555, 2020, doi: 10.20965/jdr.2020.p0546.
- [9] M. Fahrurrozi, S. A. Tarigan, M. A. Tanjung, and K. Mutijarsa, "The Use of ISO/IEC 27005: 2018 for Strengthening Information Security Management (A Case Study at Data and Information Center of Ministry of Defence)," in *ICITEE 2020 - Proceedings of the 12th International Conference on Information Technology and Electrical Engineering*, Oct. 2020, pp. 86–91, doi: 10.1109/ICITEE49829.2020.9271748.
- [10] M. Brunner, C. Sauerwein, M. Felderer, and R. Brey, "Risk management practices in information security: Exploring the status quo in the DACH region," *Computer Security*, vol. 92, May 2020, doi: 10.1016/j.cose.2020.101776.
- [11] P. Wang and M. Ratchford, "Integrated methodology for information security risk assessment," *Advances in Intelligent Systems and*

- Computing, vol. 558, pp. 147–150, 2018, doi: 10.1007/978-3-319-54978-1_20.
- [12] ISO, ISO/IEC 27001:2022, Information security, cybersecurity and privacy protection — Information security management systems — Requirements, 2022. [Online]. Available: <https://www.iso.org/standard/82875.html>.
- [13] M. E. Whitman and H. J. Mattord, “Principles of Information Security Sixth Edition,” 2018. [Online]. Available: www.cengage.com.
- [14] ISO, ISO/IEC 27005:2018, Information technology — Security techniques — Information security risk management, 2018.
- [15] I. Akkiyat and N. Souissi, “Modelling risk management process according to ISO standard,” *International Journal of Recent Technology and Engineering*, vol. 8, no. 2, pp. 5830–5835, Jul. 2019, doi: 10.35940/ijrte.B3751.078219.
- [16] G. H. S. Rampini, H. Takia, and F. T. Berssaneti, “Critical success factors of risk management with the advent of ISO 31000 2018 - Descriptive and content analyzes,” in *Procedia Manufacturing*, 2019, vol. 39, pp. 894–903. doi: 10.1016/j.promfg.2020.01.400.
- [17] H. Setiawan, F. A. Putra, and A. R. Pradana, “Design of Information Security Risk Management Using ISO/IEC 27005 and NIST SP 800-30 Revision 1: A Case Study at Communication Data Applications of XYZ Institute,” in *In 2017 International Conference on Information Technology Systems and Innovation (ICITSI)* (pp. 251-256). IEEE., 2018, pp. 251–256.
- [18] M. al Fikri, F. A. Putra, Y. Suryanto, and K. Ramli, “Risk assessment using NIST SP 800-30 revision 1 and ISO 27005 combination technique in profit-based organization: Case study of ZZZ information system application in ABC agency,” in *Procedia Computer Science*, 2019, vol. 161, pp. 1206–1215. doi: 10.1016/j.procs.2019.11.234.
- [19] M. Ngamboé, P. Berthier, N. Ammari, K. Dyrda, and J. M. Fernandez, “Risk assessment of cyber-attacks on telemetry-enabled cardiac implantable electronic devices (CIED),” *Int J Inf Secur*, vol. 20, no. 4, pp. 621–645, Aug. 2021, doi: 10.1007/s10207-020-00522-7.
- [20] I. M. M. Putra and K. Mutijarsa, “Designing Information Security Risk Management on Bali Regional Police Command Center Based on ISO 27005,” in *3rd 2021 East Indonesia Conference on Computer and Information Technology, EIConCIT 2021*, Apr. 2021, pp. 14–19. doi: 10.1109/EIConCIT50028.2021.9431865.
- [21] O. Ali, A. Shrestha, A. Chatfield, and P. Murray, “Assessing information security risks in the cloud: A case study of Australian local government authorities,” *Gov Inf Q*, vol. 37, no. 1, Jan. 2020, doi: 10.1016/j.giq.2019.101419.
- [22] H. Okonofua and S. Rahman, “Evaluating the Risk Management Plan and Addressing Factors for Successes in Government Agencies,” in *Proceedings - 17th IEEE International Conference on Trust, Security and Privacy in Computing and Communications and 12th IEEE International Conference on Big Data Science and Engineering, Trustcom/BigDataSE 2018*, Sep. 2018, pp. 1589–1592. doi: 10.1109/TrustCom/BigDataSE.2018.00230.
- [23] NIST, “Guide for conducting risk assessments,” Gaithersburg, MD, 2012. doi: 10.6028/NIST.SP.800-30r1.
- [24] G. Stoneburner, A. Goguen, and A. Feringa, “Risk Management Guide for Information Technology Systems Recommendations of the National Institute of Standards and Technology,” 2012.
- [25] OWAPS, “Upcoming OWASP Global Events,” 2022. [Online]. Available: <https://owasp.org/www-project-top-ten/#>

An Effectivity Deep Learning Optimization Model to Traditional Batak Culture Ulos Classification

Rizki Muliono¹, Mayang Septania Iranita², *Rahmad BY Syah³ 

Faculty of Engineering, Informatics Department, Universitas Medan Area, Medan, Indonesia^{1, 2, 3}
Excellent Centre of Innovations and New Science, Universitas Medan Area, Medan, Indonesia^{1, 2, 3}

Abstract—Ulos is one of the Batak culture's traditional heritage fabrics. Ulos cloth is divided into several types, each with a distinct function. Ulos Ragi Hotang, Ulos Pinunsaan, Ulos Tumtuman, Ulos Ragi Hidup, and Ulos Sadum are the five Batak ulos motifs. The Batak ulos motif has evolved over time and is now well-known in other countries. However, many ordinary people have difficulty distinguishing between ulos cloth and other fabrics. This study categorizes the different types of ulos cloth so that it can be used by ordinary people who are unfamiliar with the different types and functions. The Convolutional Neural Network is the method used (CNN). CNN is used to recognize and classify images. CNN's main feature is that it detects feature patches from locations in the input matrix and assembles them into high-level references. The Modular Neural Network (MNN) is then used to break down large and complex computational processes into smaller components, reducing complexity while still producing the desired output. 80% of the data for the training process, 20% for testing. The accuracy value achieved is 97.83%, the loss value is 0.0793, the val loss is 2.1885, and the val accuracy is 0.7429.

Keywords—Ulos; classification; convolutional neural network; modular neural network; deep learning

I. INTRODUCTION

Indonesia has a rich cultural diversity. Language, ethnicity, religion, and community beliefs are all part of the community. The Toba Batak are a tribe found throughout North Sumatra [1]. In everyday life, the Batak tribe maintains culture, as evidenced by the performance of traditional events with a distinct feature, namely, ulos [2].

Ulos are used in Toba Batak traditional ceremonies, with different types and functions depending on their roles. At first glance, Batak Ulos appear to have the same motif or size, but this is not the case [3]. If you want to learn more about the different types of Batak Ulos, you can talk to weavers or those who work with ulos. The motif of the Batak ulos has, of course, evolved over time, and it is now well known not only in Indonesia, but also in other countries [4].

Toba Batak has five different types of ulos. Ulos Ragi Hotang is used in wedding traditions. Ulos Pinunsaan is worn by kings and used in large customary events. The first child of sedition wore Ulos Tumtuman. Ulos Yeast Life as a metaphor for finding happiness in life. Ulos Sadum is commonly given as a gift to officials.

One of the long-standing problems in computer vision that has yet to be solved is the classification of objects in images in general [5]. How to replicate the human ability to understand image information so that computers, like humans, can recognize objects in images. The feature engineering process is generally very limited in that it can only apply to specific datasets and cannot generalize to any type of image. This is due to various differences between images, such as differences in viewing angles, scale, light conditions, object deformation, and so on [6,7].

There are several algorithms that can be used in image processing. They include Nave Bayes, Support Vector Machines, and Neural Networks. A Neural Network is a commonly used algorithm [8]. The workings of neural networks in the human brain inspired the development of neural networks. Digital image processing algorithms have also been developed in tandem with technological advancements. The Convolutional Neural Network is a deep learning development [9,10].

Deep learning is an artificial intelligence (AI) method that teaches computers to process data in ways similar to how the human brain does. Deep learning models are capable of recognizing complex patterns in images, text, sound, and other data in order to generate insights and accurate predictions [11]. Deep learning methods can be used to automate tasks that would otherwise require human intelligence, such as describing an image or transcribing a sound file into text. In digital image recognition, the Convolutional Neural Network method produces the most significant results. This is due to the fact that CNN is based on an image recognition system in the human visual cortex [12].

A modular neural network is a collection of different neural networks that work independently to produce output without interacting with one another. Each neural network performs different sub-tasks by obtaining distinct inputs from other networks [13]. The benefit of this artificial neural network is that it divides large and complex computational processes into smaller components, reducing complexity while still producing the desired output. Interfaces between modules in such a modular architecture can be viewed as "information relays" that encode, delimit, and disseminate critical information [14].

II. RESEARCH METHODOLOGY

A. State of the Art

The following is a classification model formation scheme for CNN and MNN.

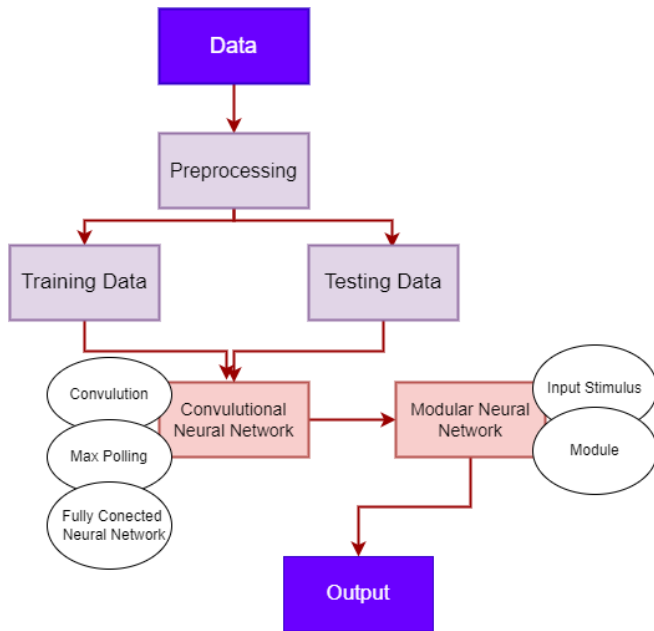


Fig. 1. State of the art.

The stages of the research can be seen in Fig. 1 [15,16].

- Data
- Data pre-processing to sort.
- Training to identify the features of each image, followed by labelled neurons to activate the image, and finally classification. Batch size and epoch are initialization parameters in the training process.
- The image is multiplied by a 32x32 kernel on three filters in the first convolution. The image is multiplied by the 64x64 kernel in the second convolution. The third convolution is an image multiplied by a 64x64 kernel with three filters. The resulting feature will make use of the same padding as well as the ReLu activation function. Using three dense layers, the first and second dense 128 layers, and the third 64 layers, the ReLu activation function was used.
- Testing to determine how well the system detects and calculates the number of images in the image. The fitting models obtained from the previous training stages were compared during validation.
- Using a circular control area around each output, the Modular Neural Network selects the number and location of inputs that affect each output. The control area uses each input's response range to identify the critical input and discard the rest.

B. Data Collection

Images obtained from observations at UD. Ulos and

Songket Tamariska Tarutung, North Tapanuli, North Sumatra Province. Images in the JPEG format were captured with a smartphone camera. There are 200 images in total, including Ulos Ragi Hotang, Ulos Pinunnaan, Ulos Tuntuman, Ulos Ragi Hidup, and Ulos Sadum [17]. See Fig. 2 for types of ulos.



Fig. 2. Types of Ulos. (a) Ulos ragi hotang, (b) Ulos Pinunnaan, (c) Ulos tuntuman, (d) Ulos ragi hidup, (e) Ulos sadum [17].

C. Convolutional Neural Network (CNN)

CNN consists of several layers. The CNN architecture has the following components [18,19,20]:

1) *Input layer*: This layer is the first to load and enter data, which is then carried over to the next layer.

2) *Convolution layer*: The convolution layer performs the convolution operation on the output from the previous layer. This layer serves as the foundation for CNN operations.

3) *Activation layer*: To convert a node's input signal into an output signal. The output results will be carried over to the next layer [21].

$$A(x) = \max(0, x) \quad (1)$$

$$C_i = \frac{a^{xi+\text{LOG}(b)}}{\sum_{k=1}^S a^{xi+\text{LOG}(b)}} \quad (2)$$

xi is the strength value of the neuron. $\text{Log}(b)$ constant value that can be determined [22].

$$\frac{\partial ai}{\partial aj} = C_i(\delta ij - Cj) \text{ with } \delta ij$$

$$= \begin{cases} 1, & \text{if } i = j \\ 0, & \text{if } i \neq j \end{cases} \quad (3)$$

- 4) *Max polling*: minimizing the number of parameters and calculations needed.
- 5) *Dropout*: To prevent overfitting problem.
- 6) *Fully connected layer*: All neuron activation layers in the previous layer are fully connected.

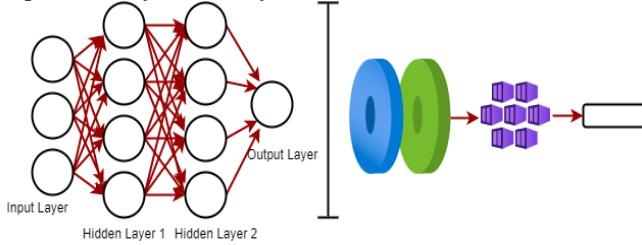


Fig. 3. CNN architecture.

Fig. 3 depicts the relationship between weighted units used to determine the effect of one object on another. CNN has one or more convolution layers that perform the convolution operation on the input before passing the results in the form of output to the next layer [23].

Several learning parameters, including batch size and epoch, must be set. Table I shows an explanation of the parameters [24].

TABLE I. INITIATION PARAMETERS

Parameter	Information
Batch Size	The number of data samples sent to the neural network in a single epoch.
Epoch	The number of rounds completed from the start of the first dataset to the end.

D. Modular Neural Network

Modular is a self-contained system that interacts with the overall architectural function to perform complex tasks. To achieve local computation, an explicit Action must be performed, so that the system being modeled becomes a meaningful function. In eq. (4), (5) and (6), static functions are represented in two dimensions [25, 26].

$$A(c(x), h(x)) = [3x] + [-x^2 + x] \quad (4)$$

$$c(x) = 3x \quad (5)$$

$$h(x) = -x^2 + x \quad (6)$$

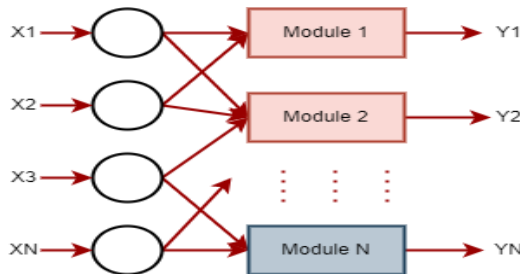


Fig. 4. Modular neural network structure.

A modular neural network structure in Fig. 4 [27].

- Tasks should be broken down into subtasks.
- Modular architecture organization
- Communication between modules

A vector can be used as the target response. On the network, each input vector is represented. Eq. (7) represents the MNN output [28].

$$h = \sum_{i=1}^k c_i y_i \quad (7)$$

c_i shows the weight on the number of network outputs resulting from the response, as explained by the deterministic transformation. Eq. (8) can be used to express the response [29].

$$a = R_i(x) + \epsilon_i \quad i = 1,2,3, \dots, n \quad (8)$$

III. RESULT AND DISCUSSION

Training data can account for up to 80% of the total. In the training process, the iteration parameter is set to 100 epochs, and the batch size is set to 32. The training procedure is carried out and is repeated 100 times in order to obtain the desired feature extraction. Table II shows the amount of training and testing data.

TABLE II. TRAINING DATA AND TESTING DATA

Distribution of Testing & Training Data	Amount of data
Training 80%	100
Testing 20%	35

The test employs 35 test data, with 7 images in each Ulos category. The model is used directly in the process of calculating accuracy. fit() method by calling the validation folder's directory and then executing it with the CNN method.

TABLE III. RESULTS BASED ON PARAMETERS

Epoch	Loss	Accuracy	Val Loss	Val Accuracy	Time
50	0.3315	0.8768	1.4779	0.6571	4 minute 13 second
75	0.2361	0.9203	0.9707	0.6571	20 minute 45 second
100	0.0793	0.9783	2.1885	0.7429	18 minute 40 second

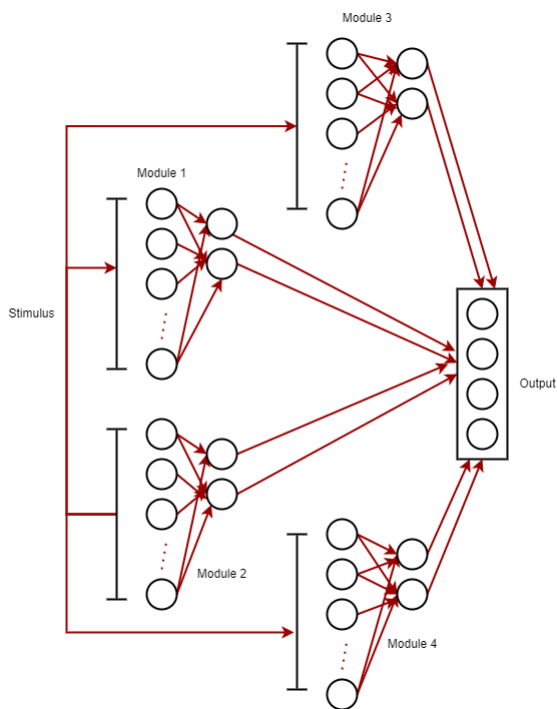


Fig. 5. MNN modeling paradigm [26].

MNN modeling paradigm is shown in Fig. 5. Table III shows the goal of determining model parameters: to compare the best models based on parameter values. It is re-optimized based on these parameters by dividing the smallest modules into the greatest number. The best accuracy results are found in the 100th epoch, which is 97.83%.

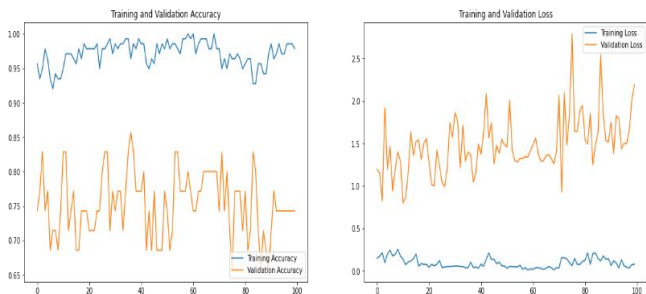


Fig. 6. Accuracy and loss at epoch 100.

Fig. 6 shows that the training and validation graphs on accuracy have the same pattern as the training and validation graphs on loss.

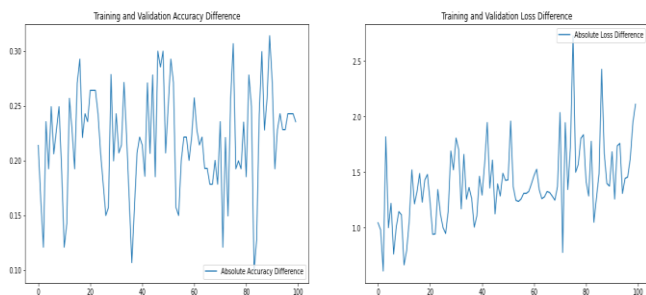


Fig. 7. Visualization of trends in trains and tests in epoch 100.

The accuracy and loss graphs have diverse patterns, as seen in Fig. 7. Due to insufficient data, this occurs and the pattern keeps repeating.



Fig. 8. Accuracy and loss at epoch 75.

An accuracy validation chart that follows the pattern of training results is shown in Fig. 8. Similar results can be obtained with the loss validation graph that mimics the training outcomes.

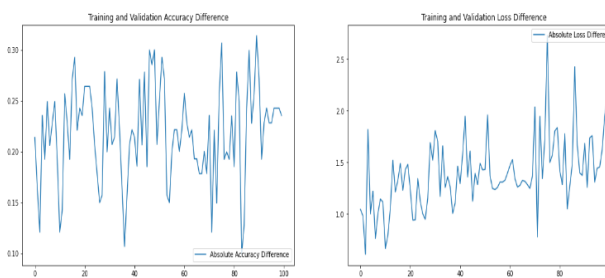


Fig. 9. Visualization of trends in trains and tests in epoch 75.

Fig. 9 shows the accuracy and loss graphs have different patterns. This is due to too little data so the pattern repeats.

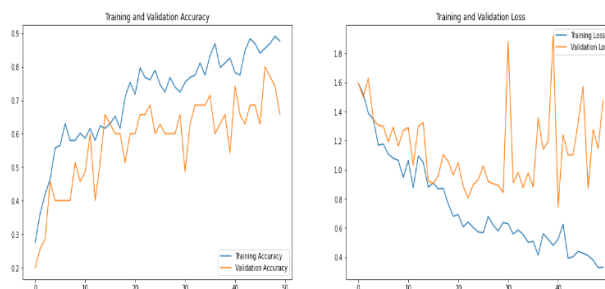


Fig. 10. Accuracy and loss at epoch 50.

Fig. 10 displays a validation graph for accuracy and a validation graph for losses that follow the pattern of training results.

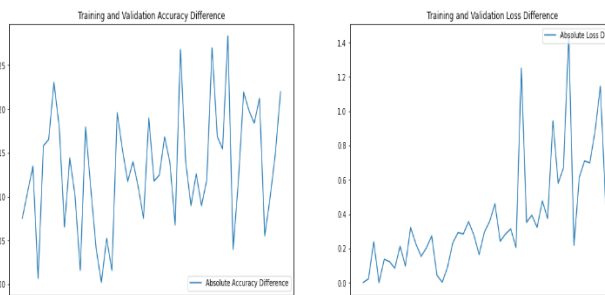


Fig. 11. Visualization of trends in trains and tests in epoch 50.

Fig. 11 shows the accuracy and loss graphs have different patterns due to too little data.

IV. CONCLUSION

The testing procedure is by using 35 test data with 7 images in each Ulos category. Because CNN already has a feature learning phase or process, it can classify objects without the need for additional feature extraction. When the process achieves accuracy, the use of MNN facilitates classification, where the process makes it easier to understand behavior or features in images to make them easier to recognize, and it increases in terms of time. Accuracy was 97.83%, loss was 0.0793, val loss was 2.1885, and val accuracy was 0.7429.

REFERENCES

- [1] E. Fatmawati, "Strategies to grow a proud attitude towards Indonesian cultural diversity," *Linguistics and Culture Review*, vol. 5, no. S1, pp. 810–820, Sep. 2021, doi: 10.21744/lingcure.v5nS1.1465.
- [2] T. M. Panggabean and A. Christy Barus, "Combining Local and Global Descriptors Through Rotation Invariant Texture Analysis for Ulos Classification," in 2019 7th International Conference on Robot Intelligence Technology and Applications (RiTA), Nov. 2019, pp. 153–159. doi: 10.1109/RITAPP.2019.8932823.
- [3] C. Nugroho, I. K. Nurhayati, K. Nasionalita, and R. M. U. Malau, "Weaving and Cultural Identity of Batak Toba Women," *J Asian Afr Stud*, vol. 56, no. 6, pp. 1165–1177, Sep. 2021, doi: 10.1177/0021909620958032.
- [4] Zulkifli and M. Ridwan, "Revitalization of the traditional values lost due to the commodification of art/crafts: a case study of Bataknese traditional Ulos," *Asian Ethnicity*, vol. 20, no. 4, pp. 541–554, Oct. 2019, doi: 10.1080/14631369.2019.1608812.
- [5] X. Feng, Y. Jiang, X. Yang, M. Du, and X. Li, "Computer vision algorithms and hardware implementations: A survey," *Integration*, vol. 69, pp. 309–320, Nov. 2019, doi: 10.1016/j.vlsi.2019.07.005.
- [6] A. O. Salau and S. Jain, "Feature Extraction: A Survey of the Types, Techniques, Applications," in 2019 International Conference on Signal Processing and Communication (ICSC), Mar. 2019, pp. 158–164. doi: 10.1109/ICSC45622.2019.8938371.
- [7] A. Al-Kaff, D. Martín, F. García, A. de la Escalera, and J. María Armingol, "Survey of computer vision algorithms and applications for unmanned aerial vehicles," *Expert Syst Appl*, vol. 92, pp. 447–463, Feb. 2018, doi: 10.1016/j.eswa.2017.09.033.
- [8] X. Li et al., "A comprehensive review of computer-aided whole-slide image analysis: from datasets to feature extraction, segmentation, classification and detection approaches," *Artif Intell Rev*, vol. 55, no. 6, pp. 4809–4878, Aug. 2022, doi: 10.1007/s10462-021-10121-0.
- [9] S. Sutrisno, N. Khairina, R. B. Y. Syah, E. Eftekhari-Zadeh, and S. Amiri, "Improved Artificial Neural Network with High Precision for Predicting Burnout among Managers and Employees of Start-Ups during COVID-19 Pandemic," *Electronics (Basel)*, vol. 12, no. 5, p. 1109, Feb. 2023, doi: 10.3390/electronics12051109.
- [10] M. Ramanan et al., "Secure blockchain enabled Cyber- Physical health systems using ensemble convolution neural network classification," *Computers and Electrical Engineering*, vol. 101, p. 108058, Jul. 2022, doi: 10.1016/j.compeleceng.2022.108058.
- [11] K. Choudhary et al., "Recent advances and applications of deep learning methods in materials science," *NPJ Comput Mater*, vol. 8, no. 1, p. 59, Apr. 2022, doi: 10.1038/s41524-022-00734-6.
- [12] Y. Dong, Q. Liu, B. Du, and L. Zhang, "Weighted Feature Fusion of Convolutional Neural Network and Graph Attention Network for Hyperspectral Image Classification," *IEEE Transactions on Image Processing*, vol. 31, pp. 1559–1572, 2022, doi: 10.1109/TIP.2022.3144017.
- [13] Y. Jiao, B. Xue, C. Lu, M. S. Avidan, and T. Kannampallil, "Continuous real-time prediction of surgical case duration using a modular artificial neural network," *Br J Anaesth*, vol. 128, no. 5, pp. 829–837, May 2022, doi: 10.1016/j.bja.2021.12.039.
- [14] J. Kauer-Bonin et al., "Modular deep neural networks for automatic quality control of retinal optical coherence tomography scans," *Comput Biol Med*, vol. 141, p. 104822, Feb. 2022, doi: 10.1016/j.combiomed.2021.104822.
- [15] S. Tripathy and R. Singh, "Convolutional Neural Network: An Overview and Application in Image Classification," 2022, pp. 145–153. doi: 10.1007/978-981-16-4538-9_15.
- [16] Y. Bhatia, A. H. Bari, G.-S. J. Hsu, and M. Gavrilova, "Motion Capture Sensor-Based Emotion Recognition Using a Bi-Modular Sequential Neural Network," *Sensors*, vol. 22, no. 1, p. 403, Jan. 2022, doi: 10.3390/s22010403.
- [17] F. R. Suwanto, D. Kartika, and D. Y. Niska, "Ethnomathematics: An analysis of frieze and crystallographic patterns on Ulos," 2022, p. 110021. doi: 10.1063/5.0113269.
- [18] Y.-L. Chang et al., "Consolidated Convolutional Neural Network for Hyperspectral Image Classification," *Remote Sens (Basel)*, vol. 14, no. 7, p. 1571, Mar. 2022, doi: 10.3390/rs14071571.
- [19] A. Sellami and S. Tabbone, "Deep neural networks-based relevant latent representation learning for hyperspectral image classification," *Pattern Recognit*, vol. 121, p. 108224, Jan. 2022, doi: 10.1016/j.patcog.2021.108224.
- [20] A. S. Paymode and V. B. Malode, "Transfer Learning for Multi-Crop Leaf Disease Image Classification using Convolutional Neural Network VGG," *Artificial Intelligence in Agriculture*, vol. 6, pp. 23–33, 2022, doi: 10.1016/j.aiia.2021.12.002.
- [21] Y. Kong, X. Ma, and C. Wen, "A New Method of Deep Convolutional Neural Network Image Classification Based on Knowledge Transfer in Small Label Sample Environment," *Sensors*, vol. 22, no. 3, p. 898, Jan. 2022, doi: 10.3390/s22030898.
- [22] T. Hassanzadeh, D. Essam, and R. Sarker, "EvoDCNN: An evolutionary deep convolutional neural network for image classification," *Neurocomputing*, vol. 488, pp. 271–283, Jun. 2022, doi: 10.1016/j.neucom.2022.02.003.
- [23] L. Chen, S. Li, Q. Bai, J. Yang, S. Jiang, and Y. Miao, "Review of Image Classification Algorithms Based on Convolutional Neural Networks," *Remote Sens (Basel)*, vol. 13, no. 22, p. 4712, Nov. 2021, doi: 10.3390/rs13224712.
- [24] X. Cao, J. Yao, Z. Xu, and D. Meng, "Hyperspectral Image Classification With Convolutional Neural Network and Active Learning," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 7, pp. 4604–4616, Jul. 2020, doi: 10.1109/TGRS.2020.2964627.
- [25] J. Soto, O. Castillo, P. Melin, and W. Pedrycz, "A New Approach to Multiple Time Series Prediction Using MIMO Fuzzy Aggregation Models with Modular Neural Networks," *International Journal of Fuzzy Systems*, vol. 21, no. 5, pp. 1629–1648, Jul. 2019, doi: 10.1007/s40815-019-00642-w.
- [26] M. Dindin, Y. Umeda, and F. Chazal, "Topological Data Analysis for Arrhythmia Detection Through Modular Neural Networks," 2020, pp. 177–188. doi: 10.1007/978-3-030-47358-7_17.
- [27] M. Amer and T. Maul, "A review of modularization techniques in artificial neural networks," *Artif Intell Rev*, vol. 52, no. 1, pp. 527–561, Jun. 2019, doi: 10.1007/s10462-019-09706-7.
- [28] R. Atefinia and M. Ahmadi, "Network intrusion detection using multi-architectural modular deep neural network," *J Supercomput*, vol. 77, no. 4, pp. 3571–3593, Apr. 2021, doi: 10.1007/s11227-020-03410-y.
- [29] M. K. Dahouda and I. Joe, "Neural Architecture Search Net-Based Feature Extraction With Modular Neural Network for Image Classification of Copper/ Cobalt Raw Minerals," *IEEE Access*, vol. 10, pp. 72253–72262, 2022, doi: 10.1109/ACCESS.2022.3187420.

Enhancing Customer Relationship Management Using Fuzzy Association Rules and the Evolutionary Genetic Algorithm

Ahmed Abu-Al Dahab¹, Riham M Haggag², Samir Abu-Al Fotouh³
Business Information Systems, Helwan University, Helwan, Egypt^{1,2}
Accounting Information Systems, Mansoura University, Mansoura, Egypt³

Abstract—The importance of Customer Relationship Management (CRM) has never been higher. Thus, companies are forced to adopt new strategies to focus on customers, given the competitive climate in which they operate. Also, companies have been able to maintain customer data within large databases that contain all information related to customers, thanks to the tremendous technological development seen recently. Multilevel quantitative association mining is a significant field for achieving motivational associations between data components with multiple abstraction levels. This paper develops a methodology to support CRM to improve the relationship between retail companies and their customers in the retail sector to retain existing customers and attract more new customers, by applying data mining techniques using the genetic algorithm through which an integrated search is performed. The proposed model can be implemented because the proposed model does not need the minimum levels of support and trust required by the user, and it has been confirmed that the algorithm proposed in this research can powerfully create non-redundant fuzzy multi-level association rules, according to the results of these experiments.

Keywords—Customer relationship management (CRM); fuzzy association rule mining; multilevel association rule; quantitative data mining

I. INTRODUCTION

Customer Relationship Management (CRM) is a critical component of an organization's information system architecture on which organizations fully rely to improve business relationships with customers. Depending on information technology, including the Internet, data warehouses, data mining, etc. After-sales service achieves a lot of the results that are obtained, which enable companies to talk, link, and exchange goods and services with their customers through various of media, including the Internet, call centers, phone calls, faxes, sales personnel, and surveys to attract new customers and retain customers and increase the happiness of customers, and gain the trust and loyalty of customers, which mainly contribute to achieving the greatest return on profitability.

One of the most important factors in the survival of companies is the relationship that binds the company to their customers. It is important to understand that when companies collect a large volume of unmanaged data, such as consumer transactions or sales, decision-makers will not benefit much from this type of data. Companies must use modern

technologies that help analyze a huge amount of data, process it, and convert it into useful information that can be understood and analyzed, which will give them good management skills and ideal basic competitiveness in the market [1].

In the information age, data mining is one of the newest and fastest-growing sub-solutions in machine learning. Knowledge discovery from data (information) is a variety of techniques for extracting common patterns from large or high-dimensional data.

Sets are technologies that provide us with accurate information that we may apply in a variety of fields, including business, engineering, and medical sciences. To find new knowledge, different strategies can be used to produce interesting rules known as association rule mining [2].

The Genetic Algorithm (GA) is a representative model that draws inspiration from the theory of natural evolution. Genetic algorithms are used because they are the most advanced and least complex algorithms when compared to other algorithms, for the variety of applications that they can be used for, and because they employ the scientific research method to identify the set of repetitive elements. This approach is better and easier to use than other genetic algorithms this heuristic method is used routinely to maintain high-quality responses in order to simplify the research of problems, dealing with them, the ability to address them, and the selection of the best solutions and results. These algorithms are widely used in mining important data to determine association rules. They are also used to find association rules in practical issues such as business databases and fraud detection [3].

Data mining efforts aim to identify association rules at the individual concept level, finding more comprehensive and meaningful knowledge by processing data and extracting the required information at the same time by exploring the different levels of ideas. Classifications of related elements can be thought of as hierarchical trees, pre-mapped for real-world uses. Inner nodes define classes or concepts that are created from lower-level nodes; [4] the leaf nodes of the hierarchical tree represent the real elements that the transactions are searching for. Fig. 1 provides a straightforward illustration of this. There are many reasons to mine multilevel association rules, including (a) the fact that these rules are more logical and user interpretable. (b) We can find answers to unwanted and

undesirable rules using the rules of the multilevel association. Applications that use spatial data analysis are encouraged [5].

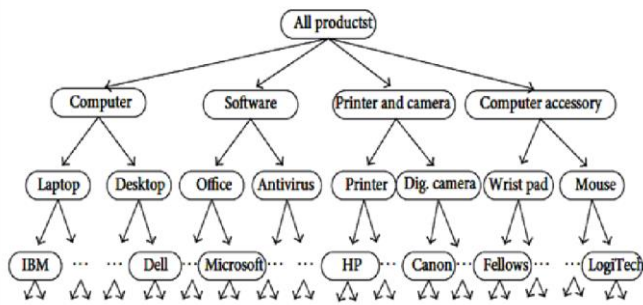


Fig. 1. The established taxonomy.

A. Motivation and Rationale

Customer contact and after-sales services have become major components of corporate strategies due to the increasing competition for retail activities. In the past, companies used to focus more on selling products and services than on looking into the details of the people who made those purchases and neglecting their wants and needs. And because of the abundance of competition, companies needed to intensify efforts to retain their existing customers, because it was difficult to obtain new ones. In addition, when economic and social conditions were adjusted for, customers were less likely to respond positively to any marketing communications from companies due to their lifestyles [6]. Companies have evolved from product/service-focused strategies to customer-focused strategies as a result of this environment. Thus, developing lasting relationships with clients has become a major strategic objective. In fact, to maintain a positive business relationship with customers, companies that want to stay ahead of the curve must constantly improve service standards. Some companies have invested in building large databases, which enable them to keep a large amount of data related to customers. Numerous data are collected for each customer, allowing analysis of the consumer's complete purchase history. But the knowledge gathered is rarely used to build enterprise processes such as Customer Relationship Management (CRM). In fact, most companies do not incorporate knowledge into their decision-support processes. Information overload and knowledge starvation are recurring problems caused by massive amounts of data. The rate at which analysts can process data has lagged behind in terms of whether the data is usable and relevant to the application [6].

Several trials have been directed at effectiveness (number of bases) and performance (speed). Quality has been neglected in the context of mining the rules of quantitative correlation. The mining algorithms in modern multilevel quantitative correlation rules completely rely on intensive looks at the database to obtain regular models that transcend all the different levels of abstraction [7]. Only users can plan the minimum support related to personal databases and this is based on the assumption of mining algorithms [8] Mining quantitative association rules is not a manageable expansion of mining categorical association rules. Since the search space is unlimited, we aim to detect a measurable set of exciting

solutions (quantitative rules), near the optimal answers. This illustrates why we have decided to solve this search problem with meta-heuristic routines, mainly genetic algorithms [9].

Each of the rules can be regularly defined or inferred and this can confuse the user when managing quantitative correlation rules mining. But more importantly, and some of these rules do not produce new knowledge and these rules may be redundant; some attempts to sell redundant controls in flat datasets. However, datasets can have multiple concept levels or a hierarchy/taxonomy, so the redundancy in these datasets requires modification. This topic is one of the stages of this research. Currently, the approach here is to identify all redundant rules and eliminate them immediately, as a result of which the number of rules that the user has to deal with will be reduced and the information content will not be reduced [10]. To locate familiar elements at different levels of abstraction, this paper presents a modified version of the Apriori algorithm with the aim of mining fuzzy multilevel association rules in large databases [11].

B. Research Problems

Understanding and meeting consumer needs are essential responsibilities for a business to compete in this day of intense competition. As a result, customer relationship management (CRM) now ranks highly among company concerns. How businesses attract new clients and retain existing ones is crucial in today's cutthroat culture. CRM is thought to use appropriate analytical devices with limited materials to attract the most valuable consumers and heighten their desire to increase purchases. The manager can search and examine a vast amount of data using data mining methods to find significant patterns and rules. This approach may also be seen as a means of acquiring knowledge to locate ambiguous association rules to carry out CRM since interactions between the expert system and users can be facilitated by fuzzy knowledge representation. The purpose of this study is to introduce the field of data mining, explicitly mining multi-level quantitative fuzzy correlation rules, to support CRM managers. To achieve precisely defined excellence, it is imperative that by answering these questions, it is probable to be provided in a basic manner. For our research, there are several dominant difficulties that we determined to gather around and seek to crack these challenges and they are:

- Boolean features can be studied as an exceptional example of categorical features and is an important student for generalizing Boolean data mining algorithms for quantitative features, however, we have to either obtain different algorithms or somehow transform the problem of quantitative correlation rules into a Boolean problem. Accordingly, we will use a novel approach to discover quantitative correlation rules emerging from a dataset with multiple concept levels.
- Due to the number of bases, this expands greatly with the number of elements. But this complication will be done by using some advanced algorithms that can cut the search space very efficiently [18]. Choosing this issue mainly helps the user when examining the ruleset. However, the process of developing more valuable quality measures for the rules through the use of the

genetic algorithm with the fitness function (relative confidence of the association rule) and the aim to confirm the most intriguing association rules is indicative of the advanced goals to solve this problem [22].

- Without prior knowledge, the purpose of these valid (estimated) intervals is to extract quantitative data which can be a very complex task. Moreover, these time intervals may not be sufficiently agreeable for experts to quickly gain non-intuitive knowledge from these generated rules. This fuzzy organic function can help with these problems.
- We can completely and effectively specify that there are no redundant association rules in hierarchical datasets.

C. Problem Statement

Market information in the actual world typically contains measurable quantities; therefore developing a sophisticated data-mining algorithm capable of working with quantitative data presents a difficulty to researchers in that field [7] The multilevel association rules mining problem can be defined as follows:

The following elements represent $I = \{i_1, i_2, \dots, i_n\}$ and Γ is a classification tree through which the multilevel taxonomic relationships between these elements can be clarified as an example of that field awareness. Element i_1 is the parent of i_2 and element i_2 is the descendant of i_1 if and there are some advantages of element i_1 over i_2 . Through which only leaf nodes are displayed within the database. The symbols DD represent a database of transactions where each transaction in the database DD is a set of elements as $T \subseteq I$. Each of the transactions is associated with an identifier $TTIIDD$. The symbol P represents the set of positive integers, and the symbol IV denotes the set $I \times p$. Couple $\langle x, v \rangle \in IV$, which means the quantitative attribute x , with the respective values and that $v \in IV$, $\{\forall \langle x, l, u \rangle \in I \times p \times p / l \leq u\}$, l represents the lower bound and u stands for the upper bound of p . The next trio $\langle x, l, u \rangle \in IV$ and denotes the quantitative x with a value within the interval $[l, u]$. Note that the coefficient TT contains the item $x \in I$ if x in TT or x is the parent of some of these elements in TT . In addition, the transaction X includes $\subseteq I$ with the condition that TT holds each element of X .

The multilevel association rule is a product of the $X \Rightarrow Y$ model, where $X \subseteq I$, $Y \subseteq I$, and $X \cap Y = \emptyset$. No element in YY is the source of any element in X ; This is $Y \cap C$ estors $(x) = \phi$ because the pattern rule " $x \Rightarrow$ ancestors $(x) = \phi$ " is fairly true with 100% confidence, and is redundant. X and Y may each contain Γ elements of any level of Γ [12][13][14].

Quantitative association rule mining still has some restrictions, such as [11]: (1) the design's adoption of a separation of the quantitative attribute prevents it from being usable by all users and attributes. (2) Users, and even specialists, frequently find it challenging to supply certain thresholds, such as the minimal amount of support, curiosity, and confidence. (3) If we use quantitative features, the search space may be very big. Fourth, the algorithm's specified rules may be too numerous to handle.

II. LITERATURE REVIEW

In this section, we compare quantum correlation rule mining algorithms taking into account the shape of these rules, and discuss the drawbacks and advantages of each technique, and the type of database that can be used. [15] [16]:

1) *Discretization*: The primary objective of this is to convert quantities of data into Booleans by examining the separation of numerical features into groups of intervals. Then this algorithm can be processed to detect the logical association rules to prepare the rules of quantities. Two main representations of the sections are included. Fixed partition, in which the groups of spacers are separated, and another type, where the ends of the spacers are overlapped with each other.

Beyond being the first effort on this topic, the main advantage of this approach is that it can manipulate numerical and categorical data in the same way. Disjoint sets incur harm from the Min_Sup and Min_Conf thresholds, whereas overlapping sets experience the cutting boundary problem. Situations (disjointed or overlapped), however, produce complications. Information will always be lost if intervals are used instead of the actual continuous data. The guidelines we provide will simply be an assessment of the ideal outcomes. Another issue is the expansion of the characteristics dimension; here, the issue is the requirement for additional memory and processing time for these data.

2) *Adjusted difference analysis*: This approach is based on using discretization and adjusted difference analysis to identify relationships between two properties. Any combination of numbers and categories might be used for the two qualities. This method may distinguish between positive and negative association rules without the requirement for user support or confidence criteria. The fact that it does not require any user criteria and can acquire a new substantial objective measure of the association rules are two of its benefits. Similar to the first strategy, this one has discretization issues that are a drawback. Additionally, this method is unmistakably thought of as producing a special case rule because the rules are always between just two characteristics.

3) *A fuzzy approach based on integrating the concepts of fuzzy logic and fuzzy sets with the Apriori algorithm*: It reforms numerical data into fuzzy member between $[0,1]$ with a membership function; then operates with the fuzzy member with an adjusted Apriori technique that can comfortably extract the rules, which are stated in linguistic terms. These approaches are based on the fuzzy additions to the classical association rules mining by establishing support and confidence in the fuzzy rule. While the mining results are straightforward to interpret by human operators, two shortcomings still insist on implementing such fuzzy approaches to the original problems. One is the computational time for mining from the database, and the other is the precision of deduced rules. A more formal description, as well as a survey of the existing methods of quantitative association rule mining, can be found in [17].

Numerous scholars have absorbed fuzzy multilevel association rules mining in the literature [13–17]. Without specifying the actual minimum support, certain of these solutions suggested multilayer membership functions via systems of ant colonies and genetic algorithms. Setting the functions for each item and then determining the minimal supports are used to increase computing performance. Other projects benefited from the efficiency and versatility that were done using OLAP and data mining techniques [12].

Up-to-date, there exist only a few algorithms for quantitative multilevel fuzzy association rule mining (QMLFRL). For example, in [25] the authors advised a QMLFRL based on the idea that the minimum support for an item at a higher taxonomic concept is valued as the minimum of the minimum supports of the items pertaining to it, and an item minimum support for an itemset is established as the maximum of the minimum supports of the items enclosed in the itemset. Under this limitation, the characteristic of downward closure is conserved, such that the original Apriori algorithm can be simply prolonged to find fuzzy large item sets..

The authors of [26] provide a brand-new, very innovative genetic-based strategy for choosing criterion values for common item sets. That is the method, an advanced coding technique is chosen, and both the fitness functions have some assurance. The user-specified minimum support is not necessary for our model. Using the genetic algorithm, a thorough search can be carried out. The experiment's findings show that the suggested approach is able to generate fuzzy multilevel association rules that are non-redundant.

The authors of [5] proposed a new technique for extracting quantitative association rules that may concurrently learn rules and quantize an attribute by using a clustering algorithm. They carried out clustering utilizing all qualities simultaneously ahead of time, and they extracted the clusters in the rules from the "association" feature. The authors' technique is superior to the traditional Cartesian product-type quantization technique in terms of total Rule extraction and quantization accuracy, which was supported by the numerical experiments.

The concept threshold - common item groups that are generated by the evolutionary algorithm - is used in the additional pertinent work provided in [3] to develop the quantitative dataset-based rules. In this example, crossover and mutation are used to unify the rule in various ways and can detect the co-occurrence of item sets. Here, comprehensibility, interestingness, and confidence are the three objectives being investigated. As a result, the created laws have developed the principle of multi-objective association. By achieving these goals, the search space for fitness functions is reduced. Finally, distribution-based optimal criteria are developed regarding the numerically valued attribute (A rule's right side displays the distribution of the values for numerical qualities like the mean or variance.)

The advantage of the previous systems is that they contain language terms that make established rules seem much more natural to human specialists; nonetheless, they may produce a

significant amount of fascinating association rules. However, since the sparseness of data in three dimensions, makes it not always easy to establish efficient association rules (meeting the lowest level of confidence and assistance) between data points at low (basic) variations in abstraction. Other related issues include (1) inadequate support for hierarchies that are needed to change over time (2) Real-world application requirements cannot be satisfied by algorithm efficiency; (3) the potential to eliminate the connection between several notion levels; (4) Their approach allowed users to specify various things with differing minimal supports[20][27].

To the book [18] that it is necessary to have a customer relationship management system from the requirement of a comprehensive analysis of the market and the study of consumer needs, product developments, and product life cycles. Information mining is the collection of data on the conduct of product sales over time to analyze different market trends. With this information, a product life cycle can be established and new products with some new improvements can be developed according to market trends and customer desires. Although it is usually a good idea to keep arbitrators global, fuzzy arbitrators operate on a group parameter that can be either general or class-specific. They base their choices on a comprehensive review of the product's sales figures in addition to analyzing other data.

Customer relationship management (CRM) seeks to create a "Learning Relation" with customers to help businesses concentrate on their needs, which are the cornerstone of all corporate operations. Businesses put their consumers at the center of management and operations by monitoring client reactions to specific goods and services. Businesses gain the knowledge necessary to raise the caliber of their goods and services by doing this. In other words, businesses find innovative strategies to keep their current clients, attract new customers, and encourage customers' contributions and loyalty to businesses through regular contact and complete knowledge [19].

A. Research Contribution

By examining several analytical characteristics of customer relationship management in the retail industry, the thesis idea presented in this study contributes to the marketing literature. In addition, it provides recommendations for companies on how they can benefit from using CRM analytics to help customers achieve understanding, which enhances customer connection. This paper incorporates the idea that was developed and inspired in part by work on QMLFRL, but that a genetic algorithm is used to calculate a minimum level of confidence and a minimum level of support for each level in this classification regardless of the nature of this data; this makes the system automated. Previous studies, it has fully investigated single-level association rule mining with GA, such as multi-target mining rules and single-level association mining rules. However, when setting up big data analysis, multilevel forms of correlation rules are regularly powerful. Multilevel association rules require big data mining to be a more efficient and effective method. GA-based multilevel association rule mining in this paper is only one attempt to discover multilevel association rules in big data with high efficiency.

III. THE PROPOSED MODEL

This project's main goal was to create practical Heuristic techniques used to mine multilevel association rules in huge datasets, with the primary goal of determining the minimal support and minimum confidence levels for each taxonomic level automatically. By utilizing the genetic algorithm's capacity to swiftly identify without completing exhaustive searches, numerous solutions can be found simultaneously in a huge multidimensional problem. our proposed technique can improve mining efficiency while maintaining the desired precision but avoiding the exhaustive list of potential association rules The definitions listed below are those connected to multilevel association rules [12][13][14]:

The first definition: the element set defines that X is the set of data objects known as the set of data elements $\{X_i, X_j\}$, where $X_i, X_j \in I$. which supports this set X in set S , $\sigma(X/S)$ is the number of transactions (in S) containing X versus the total number of transactions in S . The confidence of $X \Rightarrow Y$ in S , $\phi(X \Rightarrow Y/S)$, is the fraction of $\sigma(X \cup Y/S)$ is in competition with $\sigma(X/S)$, i.e., the probability that element set Y occurs in S when element mining occurs that X occurs in S .

The second definition: is that the group of elements X is very large in the groups at level L if the size of the support X is not less than the corresponding minimum of the support σ_L' Confidence in the $X \Rightarrow Y/S$ rule is very high at L if its confidence ratio is not lower than the equivalent minimum confidence threshold for ϕ_L'

Third identification: The rule $X \Rightarrow Y/S$ is very strong if $X \Rightarrow Y/S$ is large at the current level and the confidence of $X \Rightarrow Y/S$ is high at the current level.

The fourth definition: This mysterious transaction, denoted by the symbol T , is presented by that

$$\bar{T} = \{(x, \mu(x)) \mid \forall x \in I, 0 \leq \mu(x) \leq 1, \mu: I \rightarrow [0,1], \bar{T} \subseteq T$$

Where T is a general set of those coefficients, and $\mu(x)$ is a degree of membership in x .

Fifth definition: A The set of soft quantitative coefficients is denoted by symbols T_q' . Let (F, E) is the soft set over the universe U and $X \subseteq E$, F stands for the fuzzy energy set of U , and E is the set of parameters. It is defined as the set of X attributes that support a transaction if:

$$T_q' = \{(\langle x, l, u \rangle, e) \mid \forall \langle x, l, u \rangle \in I \times p \times p \mid l \leq u, e \in E\}$$

In general, the focus is always on digging into association rules on a single conceptual level. And some applications cannot locate the link in the multiple abstract levels in the large databases of transactions, where each of those transactions consists of a set of elements and its classification (hierarchy) on these elements, it is expected that the links between these elements will be discovered in any level of this classification. To investigate the process of searching for multi-level association rules, everyone wants to bear the cost of this data in a multi-level association and that on the multiple levels of this abstraction and these effective methods of mining in multi-level rules. We can achieve the first specifications by

producing classifications of concepts, from the primitive level of concepts to the higher level of those concepts, or we can find more efficient and effective methods for the process of exploration and research in the rules of the multi-level association.[13].

Only one modification of the Apriori algorithm, which deals with datasets within databases, particularly transaction records, or records including a certain number of fields, or which uses a multi-level "bottom-up approach" has been made to ML_T2L1 [21] [30]. Which runs the ML_T2L1 algorithm from the transaction table and these tables contain the hierarchical information in which it is encoded. Each of these levels deals with the data set completely separately. The next first: level 1 search (which is the highest level in the levels of the hierarchy) is done for very large groups of 1 item using the Apriori algorithm. Secondly, we then use the list of large group1 items from level 1 to correct, revise and cut the data set of this transaction for any element that does not have any predecessor or recurrence in the large group1 list of level 1 , which eliminates any transaction that may It has no items in common (so it only has rare items when grading using the Level 1 Big List of 1 Item). This is through the large 1 -item level 1 list, after which the large 2 -item level 1 list is completed (this using the dataset cleaned from before). Then large item sets are inferred from level 1 and this process is repeated until there are no duplicate item groups to find in level 1 . This is because ML_T2L1 only selects items that are offshoots of these duplicate items in level 1 (essentially they must be descended). From this level this set can be for large 1 element) and is always recursive by itself, and the collections of elements are completed at level 2 of the iterative transaction table. For level 2 the large 1 -item groups are built, from which the 2 large groups are determined and then the 3 large item groups, and so on. From the same well-filtered datasets) and so on. The fixes for ML_T2L1 so that all these levels are explored using the Apriori algorithm or the large 1itemsets are not displayed at the level of the proposed system core steps are as follows: [13][14][23][20] [27] [30-32] :

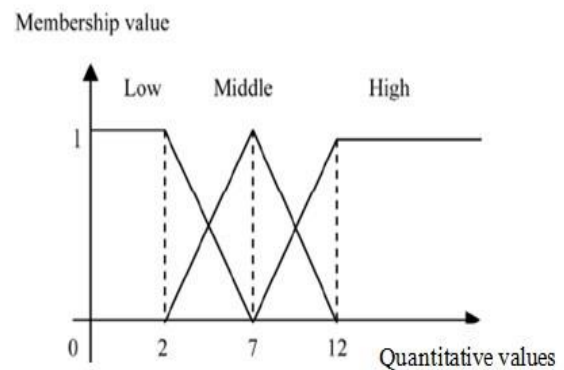


Fig. 2. Shows how the elements in perform as members.

Input: A collection of N transactional quantitative data D , a predetermined Γ with the sum of the parent element groups $\{i_1, i_2, \dots, i_n\}$, the set of membership functions for each of the elements in the different levels. In our example, all the organic functions have the same form as we can see in Fig. 2; but the primary x-axis for each element is determined based on the

highest quantitative value of each element with which it is associated. Finally, the parameter f is determined by the minimum support α_k and the minimum confidence λ_k that is gained by the genetic algorithm.

The Output: a set of fuzzy multi-level association rules without the constraints of minimum support and optimal trust.

Step 1: All previously defined labels are compiled using their order symbol "*" and numbers based on the formula, $C = \rho * 10 + i$, where i is the position number of the node at the current level l , C denotes the symbol of the node i^{th} at this current level and ρ is the parent symbol of node i^{th} at that current level.

Step 2: All clause terms are interpreted in each transaction statement and the encryption scheme is approved. Then we define $k = l$ and $r = l$ where $k, l \leq x$ is the number of the last level, x is the number of the level in a specific classification and r indicates the number of elements that have been saved in the set of recurring elements.

Step 3: We group all elements with the same first k number in each D_i transaction, and enter all quantities of elements in these similar groups into D_i . This stands for the j -th group I_j^k of D_i as v_{ij}^k .

Step 4: We have explored a lot of membership functions for each of the different data items where each data item has its features and each has its membership function, and then we convert the v_{ij}^k value of each transaction D_i of all encrypted group I_j^k to fuzzy set f_{ij}^k (Eq. 1) by plotting a v_{ij}^k over all the given membership functions, where I_j^k is the j -th element in level k , v_{ij}^k is the quantum value of I_j^k in D_i and h_j^k is the number of fuzzy regions of I_j^k , R_{jl}^k ($1 \leq l \leq h_j^k$) is the l -th fuzzy region of I_j^k , f_{ij}^k is the fuzzy organic value of R_{jl}^k

$$\left(\frac{f_{ij1}^k}{R_{j1}^k} + \frac{f_{ij2}^k}{R_{j2}^k} + \dots + \frac{f_{ijh}^k}{R_{jh}^k} \right) \quad (1)$$

Step 5: We construct the candidate set C_1^k by summing all the fuzzy regions (through linguistic terms) with organic values greater than zero. Where the numerical principal values k_{jl} for each fuzzy region are R_{jl}^k in the

$$\text{Transactional data as } s_{jl}^k = \sum_{i=1}^n f_{ijl}^k$$

Step 6: Check to see if the value $S_{k_{jl}}$ for each R_{jl}^k region in C_1^k is greater than or equal to the lower bound of α_k which represents the lower bound of the best support for level k that can be obtained by applying this genetic algorithm to the set of coefficients Included at all this level according to Γ (see. Genetic algorithm. 1). Whether R_{jl}^k is boundary matched, then we put it into the large l -items L_1^k collection of level k . that:

$$L_1^k = \{R_{jl}^k \mid S_{k_{jl}} \geq \alpha_k, R_{jl}^k \in C_1^k\} \quad (2)$$

Step 7: If L_1^k is empty, we can go to step 3, where $k = k + 1$; If it doesn't, then we create a C_2^k requester array of $L_1^1, L_1^2, \dots, L_1^k$ to catch the "level crossing" group of elements. Therefore, the pool of C_2^k applicants created must fulfill the following conditions: (1) each pool must consist of two items

in the C_2^k and contain at least one item only. L_1^k . (2) the two regions in a group consisting of two elements may not have the same name as this element. (3) that in a hierarchical relationship in this classification, the names of the elements in the set may not be made up of only two elements. (4) the two bulk element sets each containing a candidate element set of 2 must have support values greater than or equal to the minimum From the support, $\alpha_{k=2}$, in each case.

Step 8: If L_1^k is null, then increase k by one, $r = l$, and go to step 3 else set $r = r + l$.

(A) create the candidate set C_2^k in the case of $r = 2$, where C_2^k is the set of elements that have been nominated with the two elements at the k level of the set.

$L_1^1, L_1^2, L_1^3, \dots, L_1^k$ to learn the "level intersection" of the set of repeating elements. Therefore, each of these groups must contain or contain at least two elements in C_2^k ; but at the same time L_1^k contains only at least one element; Note that the first element must not be the same as the next in the classification. The two possible combinations are made in C_2^k .

(B) create the candidate group C_r^k if $r > 2$, where C_r^k is the group of elements that were nominated with r -items at the k level of L_{r-1}^k and this is done in the same way and the same steps mentioned in the previous steps.

Step 9: For each of the obtained r -itemset s with element set (S_1, S_2, \dots, S_r) in C_r^k :

The value of fuzzy S is computed in all data for each D_i transaction by using the minimum operator

$$\text{as } f_{is} = \min (f_{is1}, f_{is2}, \dots, f_{isr})$$

B) Estimate the scalar cardinality of S in all of the Transactional data as $\text{count}_s = \sum_{i=1}^n \int i_s$

If those numbers are greater than or equal to the set of elements you predetermined including $S = (S_1, S_2, \dots, S_r)$, $r > 2$ as follows: minimum support α_k place S into L_r^k .

Step 10: In the case of L_r^k is equal to nothing, then we increase the K by only one and move on to the next step for that; if you did not increase r by one, we move to step 8 immediately.

step11: If $k > x$ the next STEP, else set $r = 1$ go to step 3

step12: produced fuzzy association rules for every common r - Itemset, including $S = (S_1, S_2, \dots, S_r)$, $r > 2$ as follows:

- Catch all the rules $A \rightarrow B$ where $A \subset S$, $B \subset S$ and $A \cap B = \emptyset$, $A \cup B = S$.
- Calculate the confidence level for each association rule using by $\frac{\sum_{i=1}^n \min (f_{is})}{\sum_{i=1}^n \min (f_{iA})}$.

Step 13: Bases that have high confidence values λ_k and are not less than a pre-determined confidence threshold are selected, where λ_k represents the pre-determined least

confidence value of the k -level that was found from the application of the genetic algorithm.

Step 14: We delete all redundant rules in multilevel data sets, and this rule R_1 is considered redundant in rule R_2 if (1) the set of elements X_1 consists of at least two elements and at least one of them consists of a descendant of On the contrary, (2) the group of elements in X_2 consists of at least two elements, provided that at least one of them is a descendant of these elements in X_1 , and (3) all other non-original elements in X_2 are all It is in the X_1 element group. In the additional case (4) the confidence interval for R_1 (C_1) is less than or equal to the confidence interval for R_2 (C_2).

A. The Extraction of Parameters using the Genetic Algorithm

We define a genetic algorithm as a type of deep search algorithm that is used to mine the minimum support and minimum support for each taxonomic level. It searches a variety of options to deal with a given problem [26]. The algorithm generates a "big set". One of the scientific solutions to each problem that allows this algorithm to "evolve" over many generations, to find the right optimal solution for each problem. This algorithm begins by selecting a set of solutions (known as chromosomes) on which to start the algorithm. To create a new population, solutions from a single population are selected and maintained. The structure of the basic genetic algorithm is as follows: (see Fig. 3).

Procedure genetic algorithm

```
begin (1)
  t = 0;
  initialize P(t);
  evaluate P(t);
  While (Not termination-condition) do
    begin (2)
      t = t + 1;
      select P(t) from P(t - 1);
      recombine P(t);
      evaluate P(t);
    end (2)
  end (1)
```

Fig. 3. Structure of the genetic algorithm [30].

1) [Start] Make arbitrary n-chromosome samples (relevant solutions to the issue).

2) [Fitness] Examine the population's fitness (qualifying) function $f(x)$ for each chromosome x .

3) [The new population] evolves frequently and generates a new population once the new population is complete, by repeating these following steps.

Selection: Through the population, two of the parents' chromosomes are selected based on their fitness (the higher the level of fitness to the better, the higher the possibility of selection).

Crossover: The crossing over of parents to produce a new generation (children) if there is a possibility of a crossover, and the offspring is the exact reflection of the parents in the case of no crossover possibility.

Mutation: The GA creates a fresh generation of mutations at each spot with a mutation probability (site on the chromosome).

Accepting: In the new society, the privileged new generation is stored.

1) [Replace] control freshly produced population to improve the algorithm's path.

2) [TEST] If the final condition is met, the program terminates and gives the best solution in this current set to them.

3) [Loop] Go the step 2.

The GA maintains a population of results $p(t)$ during repetition t . r'_i, r'_n , where r'_i refers to the rule set that is produced at random for each level. The function is used to evaluate each answer. $E(\bullet)$ and $E(r'_i)$ is an indicators of how suitable a solution is. A person's fitness value determines whether they have the necessary ability to live and procreate in the next generation. The next iteration is $(t+1)$. A new resident is formed based on the procedures (2) and (3) [30].

B. Data Encoding

A population often consists of several sets of rules. The coding method used by the system is called the Michigan approach, in which each chromosome is treated as a set of all applicable rules. For each level, with this approach, randomly generated association rules are introduced into each level. Michigan technology uses binary coding for its encryption, which is (0, 1), where the number "1" indicates that the knowledge base is included in the knowledge base, while the number "0" indicates that it is excluded, that is, it does not use it. The main advantage of this technique is that all the rules of the rule are codified, or in other words encrypted, and accordingly, there is no necessity to conduct a quantitative analysis of these rules that cannot be dispensed with, to know and determine that the method in which it is dealt with works as planned or not because it is quite unlike the Bates method, which compiles all possible rules at runtime Genetic algorithm. The size of the chromosome is inversely related to the size of the base of the base, meaning that if the size of the base of the bases increases, the size of the number of groups of elements also increases. Noting that the large size of the chromosome is a major defect [8][29] [31].

C. Generic Operators

It is common to use genetic factors for reproduction, crossover, and mutation. To achieve genetic operators, it is not necessary to select a single member of the population to work on. The aggregation plan depends primarily on the level of physical fitness of the entire population. The system controls the elections in the style of the roulette wheel model. Multiple parents and mutants are selected depending on their fitness, that is, the higher the value of the fitness function a candidate has, the higher the probability of being selected. To implement the sampling of the roulette wheel, first, all applicant values are normalized so that they have the same chance range, then a random number between 0 and 1 is estimated using the chance number function, and a candidate is tested based on a match between this value and an adjustment value The value is based on suitability and, accordingly, the candidate is elected [31].

After an individual has been selected, cloning operators do not move the individual as it is selected; instead, they just replicated it using the existing population through the new population without transfer. Starting with two selected and selected individuals, the intersection operator moves to the transition point (a number among l and $L-l$ being an integer, anywhere L is the length of the strings) chosen at random. The mutation is the third genetic factor. Random shifts in population composition and arrangement can intermittently or irregularly produce beneficial results by deviating from the local optimum. In GA, the sole purpose of this mutation is to oppose each leg of this chain, *i.e.* change 0 to 1 and vice versa change 1 to 0 with the chance of p_m [31].

The algorithm stops working when it reaches a state of deterioration or decay - if the chromosome produced changes between the best and the worst in the population, and this is by less than 0.1%. The implementation is completed when the final number of these generations that the user previously specified reaches. In addition, the algorithm avoids randomly generating the initial set because it can appear in rules that exclude training data in cases where the fitness is too low. Accordingly, a population of the same rules that are locked and contain at least one training instance may result in overfitting in the relevance of this data. The evidence showed that initializing non-random methods can improve the accuracy of the solution and can significantly reduce the running time. [24]. As a result, we developed a new approach to adapting practice case selection to serve as a “seed and foundation” for generating these rules based on a change in the groups of elements within each level. [32].

In general, the genetic factor helps to manage population heterogeneity and prevent early alignment with the local optimum [27]. The primary goal is the exploration of fascinating association rules. As a result, the fitness function is very necessary to check the significance of the chromosome, and it also greatly influences the convergence of the genetic algorithm. In this case, the proposed system checks two different fitness functions. The initial function of fitness is the congruent confidence of the corresponding correlation rule as in Eq. 3, while the connection between the confirmation (conf) and support (sup) properties of the second fitness function, which is necessary to determine the correlation rule (see Equation 4) [8] [12] [27]. The parameters α and β are the important factors for the confidence balance, support in the fitness function, $\beta+\alpha=1$. Using the GA approach to mine confirmed association rules from the huge database, the threshold for the fitness function had to be determined in advance; In this case. $\alpha=\beta=0.5$.

$$f_1: \text{rconf}(X \rightarrow Y) = \frac{\text{sup}(X \cup Y) - \text{sup}(X) \times \text{sup}(Y)}{\text{sup}(X)(1 - \text{sup}(Y))} \quad (3)$$

$$f_2(x \rightarrow y) = \alpha \times \text{sup}(x \rightarrow y) + \beta \times \text{conf}(x \rightarrow y) \quad (4)$$

By implementing the suggested approach, only the most interesting rules are announced by the fitness function-defined interestingness measure, as opposed to standard mining models that produce an infinite number of interesting rules. Because GA performs a find and deals globally with attribute cooperation better than greedy rule selection algorithms, they

are the preferred method for learning high-level prediction rules [27].

Very briefly, the evolutionary method that has been proposed here for mining quantum correlation rules is driven particularly by the following: (1) the passed rules may be too many to handle; (2) the search space may be too large when we encounter quantitative attributes; (3) users, even experts, usually get bored when specifying the minimum support; and (iv) quantitative attribute splitting is not available for every attribute and user [22]. However, mining association rules fall short in terms of benefits; they also have several flaws, most notably a sophisticated algorithm. With the number of things, the number of rules grows exponentially. However, certain sophisticated algorithms are used to tackle this complexity and effectively reduce the search space. The second challenge is choosing intriguing rules from a set of rules, or attaining rules from rules.

The proposed work tackles the second issue, which essentially helps the user scan the rule set, and useful quality controls on the rules are implemented based on genetic algorithms. Usually, managing association rule mining results in a large number of rules being discovered or inferred, confusing the user. More importantly, some of these criteria might not be necessary and produce no new information. Some efforts have been made to address duplicate rules in flat datasets, but redundancy in these datasets needs to be focused on because they can contain a hierarchy/taxonomy or compound idea levels. One of the characteristics of this study is this problem.

IV. EXPERIMENTS AND RESULTS

We perform various experiments in this section to evaluate the performance of the proposed strategy and validate the improvements. Experiments are performed on a computer Intel® Core™ i5-2450M CPU @ 2.50 GHz, running on a Windows 7, 64-bit operating system, x64-based processor, and 6 GB of memory. All codes are implemented under MATLAB version 7.8.0.

A. Dataset

We have relied on the information contained in [28] because it is considered an important standard to be used for comparison. And this incoming information consists of the data of the goods, their quantities, and their elements declared in each purchase container that was marketed, which is known as the market basket data. Each set of data contains more than 1,000 sales receipts for sales in a food warehouse, and each asset contains 7 contracts (10,000 transactions). The pre-determined classification in the first level contains 7 nodes that describe the items that were made in this test, and the second level consists of 14 nodes related to flavor, taste, taste or different types of other stocks and comes in the third level, which consists of 48 nodes that express information about factories and production companies. These transactions contain a database that contains some data such as the name of the product, its shape, and the quantity of the goods that were purchased. However, it is not permissible in any way to use the same element more than once in one transaction.

B. Methodology

The methodology proposed by investigating various analytical CRM characteristics in the retail industry, this thesis contributes to the marketing literature. Additionally, it offers recommendations for businesses on how analytical CRM is used to facilitate consumers' acquisition of information, which in turn improves the relationship with customers. This thesis also emphasizes the potential of large-scale database data mining techniques for CRM.

It was compared to the method typically used in [28] which proposes a method for mining multi-level fuzzy quantitative association rules that applies the GA optimization approach. The goal was to (1) enhance multi-level optimum support and confidence employed to obtain interesting rules to unearth fresh in-depth data. (2) Removing unnecessary rules that were present in the conventional method. Both approaches consolidate fuzzy boundaries rather than explicit border intervals by inferring large item collections top-down and with increasing depth.

In contrast to the conventional method, which relies on specialists to decide these values manually the suggested approach enables the mining of association rules based on the most advantageous recalculated mining parameters (min sup, min conf). By using GA to determine these characteristics, the suggested system becomes more universal and independent from context. The min-sup and min-conf criteria were chosen at 0.29 and 1.8 for each taxonomic level in the investigations.

In the first experiment, by using the data set we test whether we can provide association rules within a given time frame with a fixed number of first generations. From the results presented in Table I, we see that the initial population size ranges from 30 to 100. We can infer the most robust association rules in the dataset but if the population is too large, and quite the opposite if the population is too small The GA-based algorithm is similar to the stochastic algorithm. Because the computational complexity is rising rapidly, we cannot obtain complete association rules immediately. However, as we can see, there is a high chance that most association rules have already been discovered, even with a small population and a short time frame. As a result, we decided to choose 50 cohorts as the default cohort for the data set because it is compatible with the approach.

TABLE I. THE CONNECTION BETWEEN THE AMOUNT OF MULTILEVEL ASSOCIATION RULES AND THE NUMBER OF INITIAL GA POPULATION (GENERATION NO. 10 APPLYING MICHIGAN ENCODING USING F1)

No. of Initial population	30	50	70	100
N0. Of association rule (Redundant)	3709	608	608	608
N0. Of association rule (non-redundant)	1734	348	348	348

As the Michigan primary coding approach is the proposed system, the primary set used is the chromosome. This is the most important reason for the stability of the number of bases extracted with the initial set, which consists of 50 chromosomes, and each chromosome contains a comprehensive depiction of the rules and laws based on the

Michigan method, and therefore each set will include the least number of these initial sets in the selected bases.

Through the following set of experiments, and using the method based on the GA algorithm, we have verified the applicability of the extracted association rules, instead of the traditional algorithm without GA. And that is to measure its value through the 10,000 transactions, using the f_1 function formula through the initial group = 50, the generation number = 10, and the mutation rate = 0.1. The data set results are shown in Table II.

TABLE II. CONTRAST STUDY

Methods	No. Of Non-Redundant Rules	Calculated min-sup	Calculated min-conf	Time (Sec)
Proposed Method with GA	348	L1=0.95 L2=0.67 L3=0.27	L1=1.12 L2=1.45 L3=1.92	450 s
Traditional method without GA [8]	2282	0.28 (L=1 to 3)	1.7 (L=1 to 3)	400 s

TABLE III. COMPARISON STUDY OF THE 2 FITNESS

Fitness Function	levels	Computed Fitness	No. of association rules (non-redundant)
f_1 (Eq. 3)	level 1	1.90	348
	level 2	1.97	
	level 3	2.06	
f_2 (E.q.4) $\alpha = \beta = 0.5$	level 1	1.07	2248
	level 2	1.26	
	level 3	1.47	

TABLE IV. EVALUATION OF USING DIFFERENT GA PARAMETERS, THE PROPOSED SYSTEM F1

Parameters ratio	No. of association rules (.redundant)	No. of association rules (non-redundant.)
Mutation= 0. 9 Crossover = 0.1	750	418
Mutation=0.8 crossover=0.2	750	418
Mutation= 0.7 crossover= 0.3	608	348

As a result of the above, we find that the approach proposed by us, which is based on the GA algorithm, takes a little more time, by an estimated rate of about 13%, so that it can identify the rules of association with high relevance compared to the traditional method. But concerning quality, the system proposed by us can extract rules of higher quality and more interesting by about 17% of the total rules that were extracted from the comparative system based on the traditional method. In general, the decision maker will be hindered if there are a large number of rules extracted from the market basket analysis. The proposed system based on the GA algorithm

presents us with the most interesting rules according to the fitness function, which is responsible for the evaluation that simulates how perfect the answer is: the higher the number provided by the proposed system, the better the solution, and vice versa.

Through the implementation of the third group of experiments, to compare the suitability of the mined association rule using either the fitness function that takes both support and confidence characteristics into account (Eq. 4). or even a fitness function that takes the relative confidence of the associated association rule into account (Eq. 3). Note that the experiment was performed based on the previous configuration of the GA. Accordingly, and through the results presented in Table III, we find that the use of f_1 improves the mining of association rules and generates an additional improvement in the rate of mined association rules while reducing the amount of extracted rules by 83%. Through this experiment, we conclude that the fitness function is a very crucial issue in the success of GA. This was shown clearly as GA did not get any benefit or advantages from using the f_2 state, and thus we got the same amount of bases extracted as we did use the traditional method (about 2248).

The use of f_1 mainly helps in improving the performance of GA, and this is through the optimal extraction of interesting rules, because of the support of each of the elements separately, as well as the calculation of the union support of the elements in each of these rules, and this is quite the opposite of the function that performs based on use support and confidence for each rule, according to the standard case used in current mining algorithms (such as the Apriori algorithm).

By studying the effect of the GA factor settings on the proposed system, which includes both the mutation rate and the crossover rate, and after it has been compared with several multilevel quantitative mining algorithms. Then we'll just change the setting to just one parameter at a time leaving the setting for another parameter at its default value to keep the number of blend settings small. Table IV shows us the number of (non-redundant) association bases that the system was able to identify in the used dataset with the mutation rate changed from 0.7 to 0.9. The above table shows that decreasing the mutation rate will reduce the number of bases extracted by 17%. This regression is very clear. Managing mutations to maintain genetic heterogeneity from one generation to the next was the main reason for this decline. In GA, intersection operators are used very extensively to supervise the population to focus on one of the best solutions yet. Mutation factors are frequently used to provide exploration (exploitation). As a result, while an intersection tries to focus on a specific area of the scene, a mutant does its best to avoid convergence and explore additional areas.

V. CONCLUSION

The proposed system provides rules based on interesting metrics, and we can easily extend the capabilities of the system if needed by changing the fitness function, and the final results show us the possibility of using quantitative correlation rules to support CRM managers. As the proposed system based on association rule mining maintains a very high accuracy when

comparing the proposed algorithm with traditional methods. In addition, the extracted rules are very close to reality. This is because different organic functions are adopted for each unique element. The minimum level of support and confidence is improved. Finally, the non-redundant algorithm was used to improve the quality and application of rules to facilitate interactions between the expert system and the users. This method can be viewed as a knowledge acquisition tool for market basket discovery and analysis (MBA), which helps CRM managers to improve their decision-making process. Future work involving the use of CRM is supported by an association rule based on genetic taboos that use GA to modify the fuzzy membership function of each element.

REFERENCES

- [1] V. Shrivastava, "Customer Relationship Management System Developed using Data Mining and Fuzzy Arbitrator", International Journal of Research in Engineering, Science and Management, vol.2, issue 2, pp.437-440, 2019.
- [2] T. Rahman, J. Kabir, and M. Kabir, "Performance Evaluation of Fuzzy Association Rule Mining Algorithms", International Conference on Electrical Information and Communication Technology, pp. 1-4, 2019.
- [3] T. Alam, S. Qamar, A. Dixit, and M. Benaida, "Genetic Algorithm: Reviews, Implementations, and Applications", International Journal of Engineering Pedagogy, vol. 10. Issue 6, 2020.
- [4] E. Mahmoudi, E. Sabetnia, M. Torshiz, M. Jalal, and G. Tabrizi, "multi-level fuzzy association rules mining via determining minimum supports and membership functions," the Second International Conference on Intelligent Systems Modeling, and Simulation, Iran, pp.55-61, 2013.
- [5] G. Garai, "Application of Genetic Algorithm in Numerous Scientific Fields", Genetic Algorithms, IntechOpen, 2022.
- [6] D. Pravinchandra, S. Shah, and M. Vasavada. Analytical CRM for Google Edge Data Mining Framework with Reference to Pharmaceuticals Industry in India, "International Journal of Management, Public Policy and Research", Vol. 2 issue 1, pp.40-61, 2023.
- [7] X. Yang, M. Zeng, Q. Liu, and X. Wang, "A genetic algorithm based multilevel association rules mining for big datasets" Mathematical Problems in Engineering, Vol. 2014, 2014, pp.1-10.
- [8] X. Yan, C. Zhang, and S. Zhang, "genetic algorithm-based strategy for identifying association rules without specifying actual minimum support," Expert Systems with Applications vol. 36, no. 2, 2009, pp. 3066-3076.
- [9] S. Aouissi, A. Vrain, and C. Nortet, "Quantminer: a genetic algorithm for mining quantitative association rules", the 20th International Conference on Artificial Intelligence, 2007, pp. 1035-1040.
- [10] G. Shaw, Y. Xu, and S. Geva, "Utilizing non-redundant association rules from multi-level datasets" the IEEE International Conference on Web Intelligence and Intelligent Agent Technology, vol. 03, Australia, 2008, pp.681-684.
- [11] P. Gautam, and K. R. Pardasani, "Algorithm for efficient multilevel association rule mining", International Journal on Computer Science and Engineering, vol. 2, no.5, 2010, pp.1700-1704.
- [12] G. Garai, Application of Genetic Algorithm in Numerous Scientific Fields. In: Genetic Algorithms. IntechOpen, 2022.
- [13] V. Ramana, M. Rathnamma, and A. Reddy, "Methods for mining cross level association rule in taxonomy data structures" International Journal of Computer Applications, vol. 7, no.3, 2010, pp.28-35.
- [14] S. Saraf, N. Adlakha, and S. Sharma, "Soft set approach for mining quantitative fuzzy association patterns in databases," International Journal of Advanced Research in Computer Science and Software Engineering, vol. 3, issue 11, 2013, pp.359-369.
- [15] M. Delgado, N. Manín, M. J. Martín-Bautista, D. Sánchez, and M. -A. Vila" Mining fuzzy association rules: an overview, "Studies in Fuzziness and Soft Computing, Springer Berlin Heidelberg, vol. 164, 2005, pp.351-373.

- [16] G. Attila, "A fuzzy approach for mining quantitative association rules, " ACTA CYBERN, vol. 15, no.2, 2001, pp.305-320.
- [17] R. Sridevi, and E. Ramaraj, "A general survey on multidimensional and quantitative association rule mining algorithms, " International Journal of Engineering Research and Applications, vol. 3, issue 4, 2013, pp.1442-1448.
- [18] ABHILASHA, Vishal Shrivastava. Customer Relationship Management System Developed using Data Mining and Fuzzy Arbiter. (2019)". "Volume-2, Issue-6.
- [19] C. Wang, C. Pang, "Applying Fuzzy Data Mining for an Application CRM", Bulletin of Networking, Computing Systems and Software Vol. 1, Num. 1, pp. 46-51, 2012.
- [20] P. Gautam, N. Khare, and K. R. Pardasani," A model for mining multilevel fuzzy association rule in database, " Journal of Computing, vol. 2, issue 1, 2010, pp.58-64.
- [21] C. Wang, Y. Li, J. Du and G. Corzo, "Mamdani Fuzzy Inference System for Rating the Performance of Sponge City Programme", International Conference on Systems and Informatics, pp. 1-6, China, 2022.
- [22] S. Prakash, M. Vijayakumar, and R.Parvathi, " A Novel Method of Mining Association Rule with Multilevel Concept Hierarchy", International Journal of Computer Applications, vol. 5, no. 5, 2011, pp.26-29.
- [23] G. Shaw, Y. Xu, and S. Geva, "Eliminating redundant association rules in multi-level datasets, " the 4thP P International Conference on Data Mining, USA, 2008pp.1-8.
- [24] E. Ayetiran, and A. Adeyemo, "A data mining-based response model for target selection in direct marketing," International Journal Information Technology and Computer Science, vol. 4, no. 1, 2012, pp. 9-18.
- [25] Y. C. Lee, T. P. Hong, and T. C. Wang, "Mining fuzzy multiple-level association rules with multiple minimum supports, " Expert Systems with Applications, vol. 34, no. 1, 2008, pp.459-468.
- [26] D. Kanani, and S. Mishra, "An optimize association rule mining using genetic algorithm," International Journal of Computer Applications, vol. 119, issue 14, 2015, pp.11-15.
- [27] Y. Wan, Y. Liang, and L. Ding, "mining multilevel association rules from primitive frequent itemsets," Journal of Macau University of Science and Technology, vol.3, issue 1, 2009,pp.10-19.
- [28] Q. Gao, F. Zhang, and R. Wang, "Mining Frequent Sets Using Fuzzy Multiple-Level Association Rules", Journal of Electronic Science and Technology, Vol. 16, Issue 2, pp.145-152, 2018.
- [29] S. Manish, A. Agrawal, and A. Lad, "Optimization of association rule mining using improved genetic algorithms," the International Conference on Systems, Man and Cybernetics, vol. 4, USA, 2004, pp.3725-3729.
- [30] M. Kayaa, R. Alhaji, "Genetic algorithm based framework for mining fuzzy association rules, " Fuzzy Sets and Systems, vol. 152, 2005, pp. 587-601.
- [31] S. Tiwari, M. K. Rao" Optimization in association rule mining using distance weight vector and genetic algorithm , "International Journal of Advanced Technology & Engineering Research, vol. 4, Issue 1, 2014, pp.79-84.
- [32] P. Wakabi-Waiswa, and V. Baryamureeba, "Mining high quality association rules using genetic algorithms", In Proceedings of the twenty second Midwest Artificial Intelligence and Cognitive Science Conference, USA, 2009, pp. 73-78.

Leader-follower Optimal Control Method for Vehicle Platoons to Improve Fuel Efficiency

Zhigang Li¹, Yushi Guo², Hua Wang³, Jianyong Li⁴, Yuye Xie⁵, Jingyu Liu⁶

School of Computer and Communication Engineering, Zhengzhou University of Light Industry, Zhengzhou, Henan China^{1, 2, 4, 5, 6}
School of Computer and Artificial Intelligence, Zhengzhou University, Zhengzhou, Henan China³

Abstract—The automotive industry has experienced swift development with the rapid growth of the economy. The increasing number of vehicles has led to deteriorating road traffic conditions, increased consumption of nonrenewable energy, and excessive vehicle emissions. To tackle the problems of fuel efficiency and safety control for vehicle platoons, this study suggests a novel leader-follower optimal control method for vehicle platoons to improve fuel efficiency. Depending on where vehicles are in a platoon, they are classified into two categories: the leader vehicle and follower vehicles. The differing driving circumstances of these two types of vehicles are considered in this paper by using various control methods. On the one hand, the leader vehicle uses an approximate fuel consumption model to improve computational efficiency. At the same time, speed and acceleration are constrained to obtain the speed optimization curve. The follower vehicle uses a distributed receding horizon control method, which calculates the vehicle's speed optimization profile online. On the other hand, a linear following model is used to prevent collisions between vehicles and ensure the safety of the vehicle platoon. The simulation experiment has demonstrated that this speed optimization control method can reduce the fuel consumption of the vehicle platoon and ensure the safety of the vehicles.

Keywords—Vehicle platoon speed optimization control; leader vehicles; follower vehicles; distributed re-eding horizon control method; vehicle platoon fuel consumption control

I. INTRODUCTION

In recent years, all industries have experienced swift development with the rapid growth of the economy. The automotive industry is no exception. According to the Ministry of Public Security, by the end of March 2022, the number of motor vehicles in the country was expected to reach 402 million, including 307 million cars. Moreover, this growth projected 487 million motor vehicle drivers, including 450 million car drivers [1]. The increasing number of vehicles has led to deteriorating road traffic conditions, increased consumption of nonrenewable energy, and excessive vehicle emissions [2]. A typical approach to addressing these problems is to accelerate the development of intelligent transport systems, and vehicle platoon control technology has received widespread attention as an important component of these systems [3].

One useful means to address to improve fuel efficiency problem is to control vehicles in a common lane to run in platoons with small inter-vehicle spacing. Arranging vehicles into a queue with a smaller distance between vehicles can effectively improve traffic flow efficiency and reduce traffic

congestion. At the same time, this strategy can also decrease the air resistance of vehicles in the vehicle platoon to reduce vehicle platoon fuel consumption and exhaust emissions. The coupling between vehicles during vehicle platoon operation is readily apparent, and vehicle platoon fuel consumption is greatly influenced by air resistance. Thus, appropriately controlling vehicle platoon speed can significantly reduce vehicle fuel consumption [4].

Platoon speed optimization control methods involve the use of advanced control algorithms and techniques for real-time monitoring and control of platoon speed. Common implementations of optimal platoon motion control include traditional proportional-integral-derivative control (PID), model predictive control (MPC), fuzzy control (FC), adaptive cruise control (ACC), rolling horizon control (RHC), and other methods. In addition, artificial intelligence techniques such as machine learning can be used to analyze and process the movement data of vehicles in a platoon, thereby improving the accuracy and efficiency of control. The platoon motion optimization control method has a wide range of promising applications in the fields of vehicle manufacturing and autonomous driving technology.

The remaining part of this paper is organized as follows: Section II briefly covers the related work. Section III is an introduction to the relevant models, including the fuel consumption model and the linear-following model. In Section IV, a leader-follower vehicle speed optimization control method for the vehicle platoon is given to reduce the overall fuel consumption of the platoon. This method includes two parts: the pilot vehicle speed optimization control method and the follower vehicle speed optimization control method. Section V presents the simulation validation and analysis. The practicality and effectiveness of the method are verified under simulation experiments and comparisons with other methods. Section VI presents the contributions and conclusions of this study.

II. RELATED WORKS

Initial research on fuel consumption-oriented speed optimization control methods focused on speed optimization control methods for a single vehicle. Many scholars have focused their attention on the influence of the road environment on vehicle fuel consumption. Regarding the study of fuel consumption of vehicles at signalized intersections, Xiaolin Tang et al. [5] proposed a multiobjective hierarchical optimal (MOHO) strategy by building a road model based on the real phase and position information of traffic signals. This

strategy can improve passenger ride comfort and fuel economy. According to Ding Feng et al. [6], the established vehicle dynamic model and an instantaneous fuel consumption model can yield the ideal constant speeds corresponding to circular curves of various radii. Jiaqi Ma [7] et al. proposed a fuel consumption optimization method for connected and automated vehicles (CAVs) on rolling terrains. This method is based on established vehicle dynamics and fuel consumption models. The newly developed algorithm was validated on a CVA platform, and the eco-drive's fuel-saving benefits were quantified.

As vehicle fuel consumption is closely related to aerodynamic drag [8], controlling vehicles to run in platoons with small intervehicle spacing in a common lane is reasonable. In fact, this approach can effectively reduce the fuel consumption of a platoon. Multiple studies [9-12] have shown that platooning can reduce fuel consumption.

In recent years, many scholars have focused their attention on the issue of vehicle platoon control. There are currently many control methods concerning vehicle platoons. The choice of vehicle In recent years, many scholars have focused on the issue of vehicle platoon control, and man control methods are currently available for vehicle platoons. The choice of vehicle control method usually depends on the specific control objectives, constraints, and environmental factors. Common control models include proportional-integral-derivative control [13, 14], adaptive cruise control [15, 16], and model predictive control [17-20]. These control methods have been extensively researched to reduce platoon fuel consumption. Communication delays will reduce the robustness of the vehicle platoon. Xu Zhu et al. [21] designed a distributed proportional-integral-derivative con-troller to improve the robustness of the vehicle platoon by obtaining strong stability conditions and upper bounds on communication delays. Cooperative adaptive cruise control (CACC) vehicles can improve the stability of the vehicle platoon and reduce emissions and fuel consumption. Yanyan Qin et al. [22] investigated the correlation between these two factors. Stability conditions for the vehicle platoon are calculated using car-following models to explore the relationship between stability and emissions. RHC and MPC are two methods based on predictive control, with the main difference being the design of the controller and how it is calculated. MPC achieves high-precision control of the system state by building a dynamic model of the system and using optimization techniques to solve for the state and control signals over several future time steps. RHC achieves high accuracy in the control of the system state by decomposing the control problem into a series of subproblems and solving each subproblem using a predictive model. Valerio Turri et al. [23] proposed a method that can dynamically calculate the optimal speed profile based on the fuel consumption of the vehicle platoon and enable real-time control of the vehicle platoon based on the DMPC method. In Chunjie Zhai et al.'s study [24], an ecological cooperative look-ahead control problem (Eco-CLC) based on distributed model predictive control (DMPC) was presented. The study addressed an autonomous vehicle platoon travelling on highways with varying slopes. Moreover, a particle swarm optimization approach with numerous dynamic populations increases

computational efficiency. To improve the fuel economy of a parallel hybrid electric vehicle (HEV), Bo Zhang et al. [25] formulated a receding horizon control problem based on a cost function of energy consumption and optimized it with a sequential quadratic programming algorithm. The results of the simulation platform validate that the pro-posed strategy can improve fuel economy.

Existing efforts to optimize vehicle platoon speed for fuel consumption often focus on setting the speed of the leader vehicle as a reference speed for the follower vehicle and optimizing the speed of the vehicles in the vehicle platoon on this basis. However, adopting a fixed reference speed for the leader vehicle is not reasonable for realistic road conditions. In addition, most of the speed optimization control strategies for vehicle platoons have complex cost functions and additional constraints, resulting in reduced computational efficiency. Therefore, the construction of the optimization problem and improvement of computational efficiency warrant exploration.

To address the aforementioned issues and lower the vehicle platoon's fuel consumption, this study proposes a fuel consumption-oriented vehicle platoon speed optimization control method based on a fuel consumption model. To improve the efficiency of the calculation, a simpler approximate fuel consumption model is used for the pilot vehicle. Because the motion states of the following vehicle are constrained by the vehicle ahead, the follower vehicles use a distributed receding horizon control method. This technology, which is both forward looking and real time, enables the vehicle platoon's ideal speed profile to be forecasted in advance based on the current driving traffic scenario and road environment for fuel consumption.

III. INTRODUCTION TO RELATED MODELS

A. Fuel Consumption Model

The fuel consumption model describes the relationship between vehicle driving and vehicle fuel consumption. Fuel consumption is influenced by the vehicle's performance (such as tractive effort, engine torque, mass, etc.), road features (such as the slope of a road), and traffic circumstances (traffic volume). Fuel consumption models can be broadly classified into three categories according to their form: instantaneous fuel consumption models, mode fuel consumption models, and average speed fuel consumption models. The instantaneous fuel consumption model is based on the relationship between fuel consumption and the instantaneous state of the vehicle. The fuel consumption model used in this study is such a model. The leader vehicle and the follower vehicle operate with different constraints and use different fuel consumption models. The leader vehicle is not constrained by the vehicle in front of it and uses a simpler fuel consumption model, which improves the efficiency of the calculation. Follower vehicles are also subject to more stringent constraints and use a more stringent fuel consumption model to enable the vehicle to travel safely and smoothly.

The fuel consumption model used in this study for the leader vehicle is an approximate fuel consumption model obtained by a quadratic approximation [26] to the Australian Rod Research Board (ARRB) model proposed by Akcelik. A

special case of an approximate fuel consumption model is where just the square of the acceleration can be used as a proxy for minimizing fuel consumption. This particular approximate fuel consumption model reduces computational effort and increases computational efficiency. At the same time, the fuel consumption calculated with this fuel consumption model does not significantly differ from the actual fuel consumption. The expression for this parameter is shown below.

$$\tilde{f}(v(t), a(t)) = \frac{1}{2} a(t)^2 \quad (1)$$

$\tilde{f}(v(t), a(t))$ denotes the secondary approximate instantaneous fuel consumption. $v(t)$ and $a(t)$ denote the velocity and acceleration of the vehicle at moment t .

However, when the vehicle is on the road, the fuel consumption of the vehicle depends not only on the acceleration of the vehicle but also on various other factors, such as engine speed, gear ratio, torque, and efficiency. The following approximate fuel consumption model is chosen for the fuel consumption model of the follower vehicle, as the movement ecology of the follower vehicle is more complex than that of the leader vehicle [27]. The fuel consumption model is expressed as follows:

$$f_{i,v}(t) = \begin{cases} b_0 + b_1 v_i(t) + b_2 v_i^2(t) + b_3 v_i^3(t) + \hat{a}_i(t)(q_0 + q_1 v_i(t) + q_2 v_i^2(t)) & a_i(t) \geq 0 \\ 0 & a_i(t) \leq 0 \end{cases} \quad (2)$$

where $b_0, b_1, b_2, b_3, q_0, q_1,$ and q_2 are parameters of fuel consumption model. When the acceleration $a_i(t)$ is greater than zero, the first part of the expression, $b_0 + b_1 v_i(t) + b_2 v_i^2(t) + b_3 v_i^3(t)$, represents the fuel consumption generated per unit time of the vehicle at velocity $v_i(t)$, and the second part, $\hat{a}_i(t)(q_0 + q_1 v_i(t) + q_2 v_i^2(t))$, represents the additional fuel consumption generated by the corresponding equivalent acceleration $\hat{a}_i(t)$ at velocity $v_i(t)$. Moreover, $\hat{a}_i(t) = a_i(t) + a_G(t)$, $\hat{a}_i(t)$ denotes the equivalent acceleration related to the acceleration $a_i(t)$ and the acceleration $a_G(t)$ caused by the ramp. In this study, the ramp of the road is not considered; therefore $\hat{a}_i(t) = a_i(t)$.

B. Linear Car-Following Model

A car-following model uses kinetic theory to investigate the change in motion of followers if the vehicle in front of them changes its motion state when the platoon is driving in a single non-overtaking lane. A mathematical model [28] is also used to create a model that is more consistent with the actual vehicle motion. In a platoon, the leader drives in traffic situations with low traffic density and large workshop distances without being influenced by surrounding vehicles and in a free-motion state. The follower vehicles are affected by traffic rules, the performance of the vehicle, and the surrounding vehicles. To

ensure safe driving, the speed of the follower vehicle is adjusted to avoid collisions with the vehicle in front of it; it cannot be driven at will by the driver's intentions and is not in a free-motion state. A diagram of the vehicle platoon's linear car-following model is shown in Fig. 1.

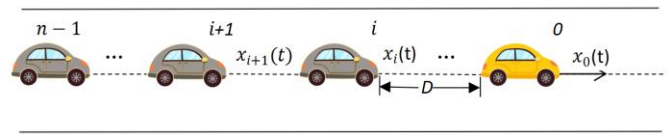


Fig. 1. Schematic diagram of the vehicle platoon's linear car-following model.

The vehicle platoon in Fig. 1 consists of $n-1$ vehicles, and $x_i(t)$ denotes the motion state of the follower vehicle i , $x_i(t) = [p_i(t), v_i(t), a_i(t)]^T$. The motion state $x_i(t)$ contains the position $p_i(t)$, the velocity $v_i(t)$, and the acceleration $a_i(t)$ of vehicle i . D denotes the distance between vehicles.

In the linear car-following model, when the motion of the previous vehicle changes, the motion of the follower vehicles also changes. If the reaction time is h , then the velocity of vehicle $i+1$ at moment $t+h$ can be expressed as:

$$\ddot{p}_{i+1}(t+h) = \lambda \left[\dot{p}_i(t) - \dot{p}_{i+1}(t) \right] \quad (3)$$

Where $\lambda = \frac{1}{h}$, λ and h denote the response sensitivity and delay time.

$\ddot{p}_{i+1}(t+h)$, $\dot{p}_i(t)$ and $\dot{p}_{i+1}(t)$ denote acceleration of the vehicle $i+1$ at time $t+h$, the velocity of the vehicle i and the velocity of the vehicle $i+1$ at time t respectively.

IV. OPTIMAL CONTROL METHOD OF VEHICLE PLATOON LEADING-FOLLOWING SPEED

The speed optimization problem for a platoon can generally be described as follows: given a starting point and an endpoint for each vehicle in the vehicle platoon, the overall vehicle platoon fuel consumption is minimized subject to certain constraints. In this study, a platoon of n vehicles is driven in a single lane for a distance on a road section without overtaking or changing behaviour. The communication topology chosen for this study is the predecessor-leader following type (PLF). This communication topology allows each follower vehicle to obtain information about the motion state of the leader vehicle and the previous vehicle. All vehicles in the platoon are divided into a leader vehicle and a follower vehicle. Different optimal control methods are employed for the leader vehicle and the follower vehicle to reduce the overall fuel consumption of the vehicle platoon along a specific segment of a one-way road. p_i , v_i , and a_i denote the position, velocity, and acceleration of vehicle i ($i=0,1,\dots,N-1$), respectively. l_i , d_i , and del_i denote the length of vehicle i , the desired workshop distance, and the error between the actual workshop distance and the desired workshop distance, with vehicle $i=0$ and others being followers. The leader vehicle must reach its final

state within a time range of T , where x_e and v_e are the final positions and the vehicle's driving speed, respectively. The time range T is within a reasonable range and can be obtained using constraints.

A. Optimal Speed Control Method for Leader Vehicle

To minimize the fuel consumption of the leader vehicle, the following fuel consumption optimization problem is used to control the leader in this study.

$$\min J_0 = \int_0^T f(v_0(t), a_0(t)) dt \quad (4)$$

where J_0 denotes the amount of fuel consumed by the leader vehicle while driving for T hours. $a_0(t)$ and $v_0(t)$ denote the leader vehicle's acceleration and speed at time t . T denotes the total driving time. $f(v_0(t), a_0(t))$ denotes the instantaneous fuel consumption of the leader vehicle. The instantaneous fuel consumption model used in this study is a special case, $f(v_0(t), a_0(t)) = \tilde{f}(v_0(t), a_0(t)) = \frac{1}{2} a_0(t)^2$.

Here, the Hamilton function is used to transform the optimization problem into an extreme value problem and to solve the leader vehicle equation of motion state, which contains the acceleration $a_0(t)$, the velocity $v_0(t)$, and the displacement equation $p_0(t)$, all of which are multiple functions of time t . The expressions are as follows:

$$a(t) = c_1 t + c_2 \quad (5)$$

$$v(t) = \frac{1}{2} c_1 t^2 + c_2 t + c_3 \quad (6)$$

$$p(t) = \frac{1}{6} c_1 t^3 + \frac{1}{2} c_2 t^2 + c_3 t + c_4 \quad (7)$$

Where c_1 , c_2 , c_3 , and c_4 are constants in the equation. These four constants can be solved by known conditions (velocity v_{00} , position P_{00} at the initial moment of the leader vehicle and velocity v_{0e} , position P_{0e} at the end moment) and constraints.

The leader vehicle has certain constraints on speed and acceleration during the actual driving process.

For the safety of the vehicle, the velocity $v_0(t)$ needs to be limited and the speed constraint of the leader vehicle can be expressed as:

$$0 \leq v(t) \leq v_{\max} \quad (8)$$

To ensure the comfort of the passengers riding in the process of the vehicle, it is necessary to constrain the acceleration of the leader vehicle, and the constraint can be expressed as:

$$a_{\min} \leq a(t) \leq a_{\max}, (a_{\min} \leq 0, a_{\max} \geq 0) \quad (9)$$

Where v_{\max} , a_{\min} , and a_{\max} are the maximum speed of the vehicle, the minimum acceleration, and the maximum acceleration possible, respectively.

To meet the above condition restrictions, the total duration T of the journey must be within a reasonable range rather than an arbitrary time. If the velocity of the leader vehicle at the initial moment $v_{00} = 0$, the initial position $p_{00} = 0$, the velocity at the end moment $v_{0e} = 0$ and the final position $p_{0e} = X$, where X is a constant and T denotes the total distance traveled by the leader vehicle. If the starting velocity v_{00} and the starting acceleration a_{00} of the leader vehicle are other values, it is still possible to solve for a range of values for the total time T . The equations of states (6) and (7) for the leader vehicle allow the solution of the unknown parameter

$$c_1 = -\frac{12X}{T^3}, c_2 = \frac{6X}{T^2}, c_3 = 0, c_4 = 0.$$

The acceleration for the leader vehicle is a monotonically decreasing function. When $t = 0$, equation (5) takes a maximum value. When $t = T$, the equation takes the minimum value. The range of T values can also be obtained using the acceleration constraint in (9).

$$\begin{cases} T \geq \sqrt{\frac{6X}{a_{\max}}} \\ T \geq \sqrt{\frac{6X}{a_{\min}}} \end{cases} \quad (10)$$

Similarly, the maximum and minimum values of the vehicle speed can be found in (6). When $t = 0$ or $t = T$, the speed equation takes the minimum value. The maximum point of the function is the maximum value when $t = -\frac{c_2}{c_1}$.

Combining this with the velocity constraint, (8) gives a range of values for T .

$$T \geq \frac{3X}{2v_{\max}} \quad (11)$$

Combining (10) and (11) can be obtained from the leader vehicle driving the total time T range of values.

B. Optimization of the Speed of the Follower Vehicles

The fuel consumption minimization problem of the follower vehicles is solved using the distributed receding horizon control method. The time slice of size T_0 is in the time domain of the total platoon driving time T , which is considered the look-ahead domain of the vehicle. At the same time, the look-ahead domain T_0 is divided into M time slices h (h is the reaction time of the follower vehicles). When the optimal control sequences for the follower vehicles in the look-ahead domain are solved, the first step of the optimal control sequences is applied. The dynamic planning diagram is shown in Fig. 2 below.

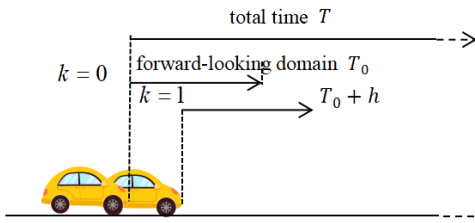


Fig. 2. Follower vehicle dynamic planning diagram based on the distributed receding horizon control method.

The total fuel consumption of the vehicles moving in the forward-looking domain at time T_0 is minimized if the total fuel consumption of each subsequent vehicle is minimized at time T_0 . The overall fuel consumption of the vehicles throughout all of the time T is minimized if the total fuel consumption of all the vehicles throughout the forward-looking domain T_0 is minimized. Thus, the speed optimization problem based on the combined fuel consumption of the vehicles in front of it can be changed into a speed optimization problem based on the fuel consumption of each vehicle. Eq. (12) shows the total fuel consumption J_1 for all follower vehicles driving T_0 . The optimal control problem for single-follower driving T_0 is shown in (13).

$$J_1 = \sum_{i=1}^{n-1} \int_t^{t+T_0} f_{i,v}(x_i(s), u_i(s)) ds \quad (12)$$

$$\min F = \min \sum_{i=1}^{n-1} \int_t^{t+T_0} F_i(x_i(s), u_i(s)) ds \quad (13)$$

where F_i denotes the cost function of the follower vehicle i , expressed as follows:

$$F_i = \psi_i f_{i,v}(x_i(t)) + \frac{1}{2} \beta_i (u_i(t))^2 + (r_i(t) - y_i(t))^T (r_i(t) - y_i(t)) \quad (14)$$

where the first term relates to the fuel consumption of the vehicle. β_i and ψ_i denote the corresponding weights and are constants. $f_{i,v}(x_i(t))$ denotes the immediate fuel consumption of the follower vehicle i . It is obtained through (2).

$u_i(t)$ is the vehicle's control input in the second period. The nonlinear model [29] was transformed into a linear model by using a linear feedback technique [30]. The computation is simplified. The structural formula is defined as follows:

$$u_i(t) = k_1 v_0(t) - k_2 v_i(t) + k_3 a_0(t) - k_4 a_i(t) \quad (15)$$

Where k_1 , k_2 , k_3 , and k_4 denote various parameters. $v_0(t)$ and $a_0(t)$ denote the leader's velocity and acceleration in a platoon, respectively.

The third term is related to minimizing the difference between the reference output and projected value. The prediction error is assumed to be the difference between the

predicted output $y_i(t)$ and the reference output $r_i(t)$. This quantity of feedback is added to achieve feedback correction. $x_i(t) = [p_i(t), v_i(t), a_i(t)]^T$ denotes the motion state of vehicle i in the platoon at time t . $p_i(t)$, $v_i(t)$ and $a_i(t)$ denotes the position, speed, and acceleration of vehicle i at time t . $y_i(t) = [v_i(t), a_i(t)]^T$ denotes the predicted output, including the speed and acceleration of the follower vehicle i at time t . $r_i(t) = [v_{oi}(t), a_{oi}(t)]^T$ denotes the output of the follower vehicles, i.e., the desired speed and acceleration of the follower. It is obtained using the linear following model.

With the PLF communication topology, the leader vehicle can broadcast messages to all follower vehicles. When the motion state of the leader vehicle changes the motion state of all followers also changes. Their reaction times are the same, both being h . The reference speed and acceleration output of the follower vehicle in a platoon at time t can be obtained from the leader vehicle and the current vehicle speed and position at the previous moment, as follows:

$$a_{oi}(t) = \lambda(v_0(t-h) - v_i(t-h)) \quad (16)$$

$$v_{oi}(t) = \lambda(p_0(t-h) - p_i(t-h)) \quad (17)$$

where $\lambda = \frac{1}{h}$, $a_{oi}(t)$, $v_0(t-h)$, and $v_i(t-h)$ signify the sensitivity, the follower vehicle's acceleration at time t , the leader vehicle's velocity, and the follower vehicle's velocity at time $t-h$, respectively. The symbols $v_{oi}(t)$, $p_0(t-h)$, and $p_i(t-h)$ denote the velocity of vehicle i at time t , the position of the leader vehicle, and the position of the follower vehicle i at time $t-h$.

The forward-looking domain T_0 is split into M time slices to numerically solve the optimal control issue. Euler's approach is then used to discretize the vehicle's motion state. A Hamiltonian function is constructed from the equation of the state and cost function of the vehicle. The prediction output equation can be solved according to the control equation, the costate equation, and the state equation. The constructed Hamiltonian function is solved iteratively to provide the predictive motion state equation, as shown in (18).

$$y_{i,k+1}(t) = (A_i h + I_3)^{k+1} x_{i,0}(t) + \sum_{j=0}^k (A_i h + I_3)^{k-j} B_i h u_{i,j}(t), \quad (18)$$

$$k = 0, 1, \dots, M-1$$

where $x_{i,0}(t) = [v_{0,i}(t), a_{0,i}(t)]^T$ denotes the current state, and I_3 is a third-order unit matrix.

According to (18), the predicted motion state equation can be represented by the current state information $x_{i,0}(t)$ and the future control sequence $u_{i,j}(t)$ together. The necessary condition for solving the optimal solution of the equation is

that the control equation $\frac{\partial H}{\partial u_i} = 0$. The optimal control series is obtained by an iterative algorithm. The first step in obtaining the optimal series is then applied as the control input for the next moment of the platoon.

V. SIMULINK SIMULATION VERIFICATION AND ANALYSIS

A. Experiments and Simulink Simulation

Three main methods are available for studying the fuel economy of vehicles: wind tunnel experiments, road experiments, and simulation methods. Among these three methods, wind tunnel experiments provide more accurate control of the influences that affect vehicle platoon fuel consumption. Road experiments are more realistic and reliable. Simulation experiments are more cost-effective, safer, and simpler than the other two methods. Therefore, this study chose to verify the feasibility of the method through simulation experiments.

The vehicle platoon in the simulation experiment consists of ten vehicles. The vehicle platoon traveled 2000m. The powertrain model parameters of the vehicles are shown in Table I.

TABLE I. VEHICLE POWERTRAIN MODEL PARAMETERS

Vehicle Parameters	Identifier	Data	Unit
Mass	m_i	1464	kg
Air Resistance	ρ	1.29	kg / m ³
Cross-sectional Area	A_i	2.2	m ²
Drag Coefficient	c_{di}	0.35	—
Mechanical Drag	d_{mi}	5	N
Time Constant of an Engine	ζ_i	0.25	—

Set the controller parameters to meet the requirements: $k_1 = 0.2516$, $k_2 = 0.51175$, $k_3 = 0.2601$, $k_4 = 0.1021$, $M = 150$. To ensure the safe movement of the vehicle platoon, according to the speed regulations, vehicles in a platoon should not exceed 50km/h on a one-way urban road. Therefore, the speed constraint for the leader vehicle is $0 \leq v_i \leq 13.9m/s$. At the same time, a typical vehicle can currently accelerate from zero to 100m/s in 10s. The average acceleration is $2.2778m/s^2$. The maximum acceleration can be up to $6m/s^2$. As a result, the acceleration of a typical vehicle when braking is $-6m/s^2$. Therefore, the acceleration constraint is $-6m/s^2 \leq v_i \leq 6m/s^2$. According to the literature [23], the acceleration restriction is typically set to $-2m/s^2 \leq v_i \leq 2m/s^2$ to make passengers feel comfortable. The leader vehicle's maximum velocity and acceleration are as stated above. The range of values for the total duration can be determined by combining (10) and (11). This study's total running time is $T = 250s$.

Under the direction of the integrated controller, Simulink simulations can be used to determine the acceleration-time

curve and velocity acceleration-time curve for each vehicle in the vehicle platoon, as shown in Fig. 3 and Fig. 4. From Fig. 3, it is noticed that that each vehicle in the vehicle platoon will swiftly finish the speed-up. All vehicles' acceleration values fall within the range $-0.8m/s^2 \leq v_i \leq 0.8m/s^2$. The vehicles' consistent acceleration improves the comfort of the passengers. From Fig. 4, the speed of all vehicles satisfies the speed constraint for vehicles traveling on a one-way urban road, which should not exceed 50km/h. The vehicles in the vehicle platoon converge at the same speed for all vehicles and stabilize within the vehicle platoon at around ten seconds.

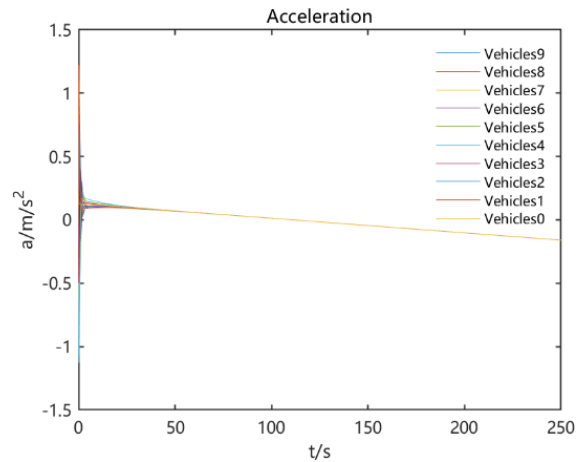


Fig. 3. Acceleration-time curve.

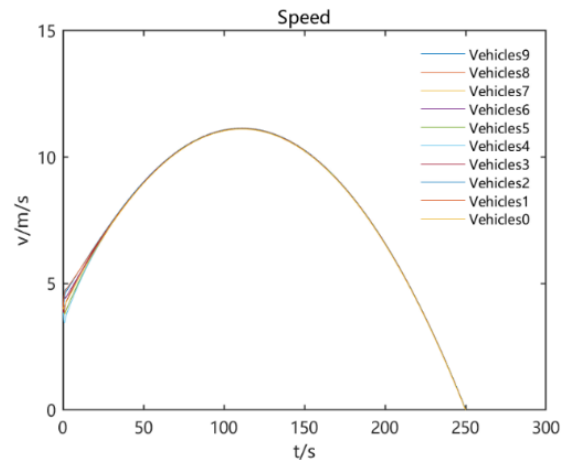


Fig. 4. Velocity-time curve.

For the follower vehicles in the platoon, use the parameters in the fuel consumption model: $b_0 = 0.1569$, $b_1 = 2.450 \times 10^{-2}$, $b_2 = -7.4152 \times 10^{-4}$, $b_3 = 5.975 \times 10^{-5}$, $q_0 = 7.224 \times 10^{-2}$, $q_1 = 9.681 \times 10^{-2}$, $q_2 = 1.078 \times 10^{-3}$. The vehicle length is $D = 3m$. According to the Road Traffic Safety Law of the People's Republic of China, when the speed is below 50km/h, the safe distance between vehicles is not less than 50m. Therefore, this study sets the desired vehicle distance $od_i = 60m$. The fuel consumption-time curve for the vehicle platoon is shown in Fig. 5. Compared to other controllers, the

vehicle platoon using the controller provided in this study consumes less fuel and saves significantly on fuel consumption. The curve of the workshop distance error over time is shown in Fig. 6, with the maximum workshop distance error not exceeding 8 meters. The workshop distance is still within the safe workshop distance range, ensuring the vehicle platoon's safety.

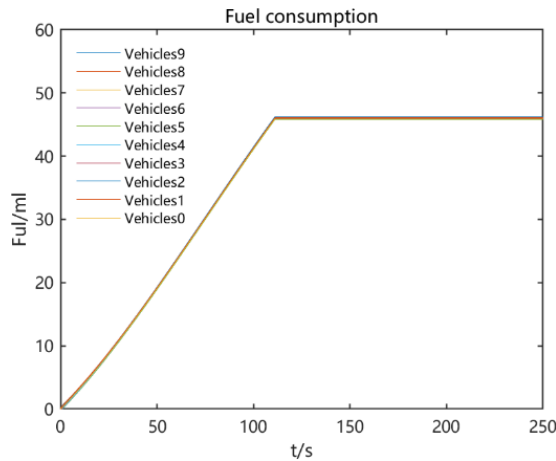


Fig. 5. Fuel consumption-time curve per vehicle.

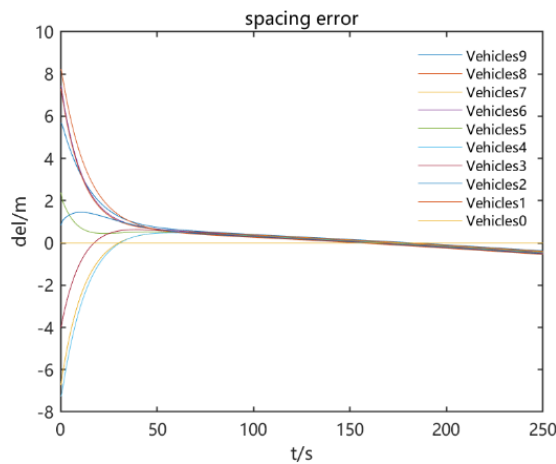


Fig. 6. Workshop distance error-time curve.

B. Analysis and Comparison

The following comparison will focus on five characteristics of the four controllers, including (1) the utilization of vehicle platoon technology. Using vehicle platoon technology, a collection of vehicles is arranged into a platoon beside a workshop. It is essential and of considerable practical value to use the platoon as a focal point for vehicle speed optimization problems because it has major benefits in the four key areas of enhancing traffic capacity, decreasing emissions and traffic accidents, and reducing fuel consumption. (2) The optimization of the speed of the leader vehicle is also considered. The current literature on optimal vehicle platoon speed control often focuses on setting the speed of the leader vehicle as a reference speed for the follower vehicles. Adopting a fixed reference speed for the leader vehicle is not reasonable in light of actual traffic conditions. In this study, speed optimization is applied to the lead vehicle of a vehicle platoon on this basis.

(3) The possibility of boosting the computational effectiveness is also considered. Increasing the computational efficiency is crucial since solving fuel consumption-focused speed optimization problems can be computationally taxing due to complex traffic conditions and a large number of moving vehicles. (4) Predictive control is then applied. The vehicle's movement is analysed and processed by predictive control. The speed of the vehicle can be quickly changed when unforeseen circumstances arise. (5) Lastly, the decision to increase fuel effectiveness is made. Since oil is a nonrenewable energy source, energy usage needs to be decreased. Fuel consumption and vehicle emissions should be lowered. It is essential to increase vehicle fuel efficiency and environmental protection.

In the same simulation environment, the integrated controller in this work is contrasted with the three controllers in the following literature based on five factors. Table II presents the outcomes.

TABLE II. COMPARISON WITH OTHER CONTROLLERS IN TERMS OF FUNCTIONALITY

Characteristics	this paper	Literature [27]	Literature [24]	Literature [23]
Using platoon technology	✓		✓	✓
Speed optimization of the leader vehicle	✓			
Improving computational efficiency	✓	✓	✓	
Enabling predictive control	✓		✓	✓
Improving fuel efficiency	✓	✓	✓	✓

The literature [27] focuses on enhancing the calculating efficiency. While the computations are simplified, significant fuel savings are achieved, but this process applies to a more demanding environment. The literature offers a state-constrained multi-objective switching control approach [24]. In addition to increasing fuel efficiency, it can also increase vehicle stability and safety. The leader vehicle's state of motion is known and used as a reference state for the follower vehicles in this procedure. State limitations are prevalent in the control strategies suggested in the literature [23]. State limitations decrease efficiency and raise computing complexity. By classifying the vehicle platoon's vehicles and streamlining calculations while enhancing computing efficiency, the controller in this study increases the vehicle platoon's fuel efficiency. Due to the lack of vehicles in front of the leader vehicle, it has a relatively simple traffic environment. The leader vehicle fuel consumption model is simplified to improve the efficiency of the calculation. Speed and acceleration constraints are also added to improve the safety of the vehicle. For the follower vehicles, the distributed receding horizon control method is used, which is forward looking and real time.

Vehicle fuel consumption is one of the most important indicators of vehicle fuel economy, reflecting the fuel consumption of the vehicle during the driving process. Vehicle fuel consumption as an evaluation indicator of the method plays an important role in evaluating vehicle performance and energy efficiency. Under the same simulation environment, the

fuel consumption of each vehicle in the platoon under the classification controller in this paper is calculated and compared with that under the controllers set up in other literature according to the fuel consumption model, as shown in Fig. 7. The data relating to the total fuel consumption of the platoon under the controllers set up in other literature is analyzed, and the results in Table III show that the reduction in fuel consumption of the platoon under the action of the classification controller is significant.

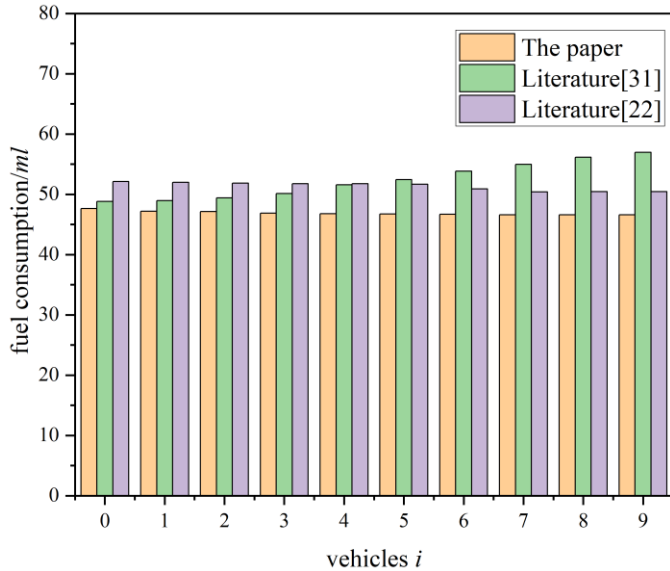


Fig. 7. The fuel consumption of the vehicles in the platoon based on the classification controller in this paper was compared with fuel consumption under controllers set up in other literature.

TABLE III. COMPARISON OF VALUES WITH OTHER CONTROLLERS IN TERMS OF FUEL CONSUMPTION

Fuel consumption	this paper	Literature [31]	Literature [22]
Total value	468.81	523.13	513.29
Average value	46.88	52.31	51.32
Minimum value	46.60	48.81	50.40
Maximum value	47.62	56.91	52.12
Median value	46.76	51.99	51.71
Standard deviation	0.33	3.03	0.70

VI. DISCUSSION

This section includes the main conclusions obtained during the experimentation. Furthermore, future work is included on how to continue the line of research, based on the premises of this article.

A. Conclusion

This paper is mainly oriented toward a vehicle platoon fuel consumption problem and the platoon fuel consumption problem. A novel leader-follower optimal control method for vehicle platoons to improve fuel efficiency is proposed. The vehicles are divided into leader vehicle and follower vehicles according to their position in the platoon and are controlled differently and optimally.

Firstly, the leader vehicle uses a relatively simple fuel consumption model to simplify the fuel consumption model of the vehicle and the optimization problem oriented to fuel consumption, improve the efficiency of the method calculation, and shorten the time for the vehicle to obtain the queue information. The vehicle transmits information about its movement status to other vehicles around it promptly so that the vehicle itself and surrounding vehicles can react accordingly to unexpected situations and reach safe driving purposes.

Secondly, the follower vehicles, being constrained by the vehicle in front, can choose a more complex and accurate fuel consumption model and a more sophisticated speed optimization control algorithm for real-time control based on the motion of the vehicle in front. The follower vehicles use a distributed rolling horizon control method. This method decomposes the global control problem into several local control problems through collaboration between vehicles in the platoon and achieves the global control objective through local control of each vehicle. It solves the optimal speed profile of the platoon online and is more real-time.

Finally, to ensure the safety of the platoon, the speed, and acceleration of the leader vehicle are constrained. The dynamic model of the vehicle queue is combined with a car-following model to ensure the safety of the platoon and a distributed rolling horizon control method to minimize the overall fuel consumption of the platoon.

The method reduces vehicle crashes and enhances vehicle platoon safety. This method has been developed based on experimental and simulation results to increase computational efficiency and decrease vehicle platoon fuel consumption. Additionally, the vehicle platoon's general safety is guaranteed.

B. Future Work

Due to the non-renewable nature of energy and the increase in air pollution, fuel consumption-oriented research, whether for vehicles or platoons, has practical significance and a wide range of application scenarios. However, as the actual road conditions are more complex and variable, there is still a gap in the actual application, and the research is not comprehensive enough. Combining the research results of this paper, the following outlook is given for the future research of fuel consumption-oriented vehicle and platoon movement optimization control methods:

1) *Further optimisation of the control algorithm:* For different vehicle types and driving conditions, more precise and efficient control algorithms should be developed to make the vehicle more energy efficient during the driving process. At the same time, it should be considered that the control method can be applied to roads with gradients, signals, etc. that are closer to reality.

2) *Exploring new optimisation strategies:* The current approach to optimized motion control is mainly based on model-based predictive control. In the future, more flexible and efficient optimization strategies, such as deep learning-based control methods and neural network-based control

methods, can be explored to further improve the efficiency of vehicle and platoon motion and fuel consumption reduction.

3) *Enhanced stability control of the platoon*: The platoon is more stable and does not transmit this disturbance, which is amplified by the disturbance that occurs in the leader vehicle.

ACKNOWLEDGMENT

We would like to thank all the reviewers for their valuable comments. This work was financially supported by Henan Provincial Science and Technology Research Project (232102210142).

REFERENCES

- [1] The Ministry of Public Security of the People's Republic of China, "The number of motor vehicles in the country reached 417 million with over 500 million drivers," Traffic Management Bureau, 2023, <https://app.mps.gov.cn/>.
- [2] R. Zhong, R. Xu, A. Sumalee, S. Ou and Z. Chen, "Pricing environmental externality in traffic networks mixed with fuel vehicles and electric vehicles," in IEEE Transactions on Intelligent Transportation Systems, vol. 22, no. 9, pp. 5535-5554, Sept. 2021.
- [3] X. H. Yu and G. Guo, "A general variable time headway policy in platoon control," Acta Automatica Sinica, vol. 45, pp. 1335-1343, 2019.
- [4] Y. Yang, F. Ma, J. Wang, S. Zhu and L. Guvenc, "Cooperative ecological cruising using hierarchical control strategy with optimal sustainable performance for connected automated vehicles on varying road conditions," Journal of Cleaner Production, vol. 275, 123056, 2020.
- [5] X. Tang, Z. Duan, X. Hu, H. Pu, D. Cao and X. Lin, "Improving ride comfort and fuel economy of connected hybrid electric vehicles based on traffic signals and real road information," in IEEE Transactions on Vehicular Technology, vol. 70, no. 4, pp. 3101-3112, April 2021.
- [6] F. Ding and H. Jin, "On the optimal speed profile for eco-driving on curved roads," in IEEE Transactions on Intelligent Transportation Systems, vol. 19, no. 12, pp. 4000-4010, Dec. 2018.
- [7] J. Ma, J. Hu, E. Leslie, F. Zhou, P. Huang and J. Bared, "An eco-drive experiment on rolling terrains for fuel consumption optimization with connected automated vehicles." Transportation Research Part C: Emerging Technologies, vol. 100, pp: 125-141, 2019.
- [8] A. A. Hussein and H. A. Rakha, "Vehicle platooning impact on drag coefficients and energy/fuel saving implications," in IEEE Transactions on Vehicular Technology, vol. 71, no. 2, pp. 1199-1208, Feb. 2022.
- [9] Q. Wen and B. J. Hu, "Integrated communication and control design for fuel-efficient vehicle platooning," Electronics, vol. 10, pp. 3117, 2021.
- [10] A. M. Mahbub, V. A. Le and A. A. Malikopoulos, "A safety-prioritized receding horizon control framework for platoon formation in a mixed traffic environment," arXiv preprint arXiv:2205.10673, 2022.
- [11] G. An and A. Talebpour, "Vehicle platooning control for merge coordination: a hybrid ACC-DMPC approach," 2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC), Macau, China, pp. 1611-1616, 2022.
- [12] M. Hu, C. Li, Y. Bian, H. Zhang, Z. Qin and B. Xu, "Fuel economy-oriented vehicle platoon control using economic model predictive control," in IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 11, pp. 20836-20849, Nov. 2022.
- [13] R. Li, S. Deng and Hu Y. "Autonomous vehicle modeling and velocity control based on decomposed fuzzy PID," International Journal of Fuzzy Systems, vol. 24, pp. 2354-2362, 2022.
- [14] S. Deng. "Vehicle control and active collision avoidance based on decomposed fuzzy PID," Beijing: Beijing Jiaotong University, 2021.
- [15] J. Wei, Y. J. Liu, H. Chen and L. Liu, "Fuzzy adaptive control for vehicular platoons with constraints and unknown dead-zone input," in IEEE Transactions on Intelligent Transportation Systems, 2023, in press.
- [16] H. Zhang, J. Liu, Z. Wang, C. Huang and H. Yan, "Adaptive switched control for connected vehicle platoon with unknown input delays," in IEEE Transactions on Cybernetics, vol. 53, no. 3, pp. 1511-1521, March 2023.
- [17] C. Earnhardt, B. Groelke, J. Borek and C. Vermillion, "Hierarchical model predictive control approaches for strategic platoon engagement of heavy-duty trucks," in IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 7, pp. 8234-8246, July 2022.
- [18] W. J. Liu, H. F. Ding, M F Ge and X. Y Yao, "Cooperative control for platoon generation of vehicle-to-vehicle networks: a hierarchical nonlinear MPC algorithm," Nonlinear Dynamics, vol. 108, pp. 3561-3578, 2022.
- [19] J. Luo, D. He, W. Zhu and H. Du, "Multiobjective platooning of connected and automated vehicles using distributed economic model predictive control," in IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 10, pp. 19121-19135, Oct. 2022.
- [20] Y. Jia, Z. Nie, W. Wang, and Y.Lian, "Eco-driving policy for connected and automated fuel cell hybrid vehicles platoon in dynamic traffic scenarios," International Journal of Hydrogen Energy, 2023 in press.
- [21] X. Zhu and Z. Zhang, "Distributed PID controller design and stability analysis for the vehicle platoon with communication delay and input delay," 2021 40th Chinese Control Conference (CCC), Shanghai, China, pp. 950-955, 2021.
- [22] Y. Qin, X. Hu, Z. He and S. Li, "Longitudinal emissions evaluation of mixed (cooperative) adaptive cruise control traffic flow and its relationship with stability," Journal of the Air & Waste Management Association, vol. 70, pp. 670-686, 2020.
- [23] V. Turri, B. Besselink and K. H. Johansson, "Cooperative look-ahead control for fuel-efficient and safe heavy-duty vehicle platooning," in IEEE Transactions on Control Systems Technology, vol. 25, no. 1, pp. 12-28, Jan. 2017.
- [24] C. Zhai, F. Luo, Y. Liu and Z. Chen, "Ecological cooperative look-ahead control for automated vehicles travelling on freeways with varying slopes," in IEEE Transactions on Vehicular Technology, vol. 68, no. 2, pp. 1208-1221, Feb. 2019.
- [25] B. Zhang, F. Xu and T. Shen, "Receding horizon optimal control of HEVs with on-board prediction of driver's power demand," IET Intelligent Transport Systems, vol. 14, pp. 1534-1545, 2020.
- [26] P. Typaldos, I. Papamichail and M. Papageorgiou, "Minimization of fuel consumption for vehicle trajectories," in IEEE Transactions on Intelligent Transportation Systems, vol. 21, no. 4, pp. 1716-1727, April 2020.
- [27] M. Kamal, M. Mukai, J. Murata and T.Kawabe, "Ecological driver assistance system using model-based anticipation of vehicle-road-traffic information," IET intelligent transport systems, vol. 4, pp. 244-251, 2010.
- [28] H. Zhou, L. Güvenç and Z. Liu, "Design and evaluation of path following controller based on MPC for autonomous vehicle," 2017 36th Chinese Control Conference (CCC), Dalian, China, pp. 9934-9939, 2017.
- [29] L. Grüne, J. Pannek, L. Grüne and J. Pannek, "Nonlinear model predictive control," Springer International Publishing, pp. 45-69, 2017.
- [30] Y. Zheng, S. E. Li, J. Wang, D. Cao and K. Li, "Stability and scalability of homogeneous vehicular platoon: Study on the influence of information flow topologies," IEEE Transactions on intelligent transportation systems, vol. 17, pp.14-26, 2015.
- [31] Y. Qin, X. Hu, Z. He and S. Li, "Longitudinal emissions evaluation of mixed (cooperative) adaptive cruise control traffic flow and its relationship with stability," Journal of the Air & Waste Management Association, vol. 70, pp. 670-686, 2020.

A Review of the Recent Progress on Crowd Anomaly Detection

Sarah Altowairqi, Suhuai Luo, Peter Greer

School of Information and Physical Sciences, The University of Newcastle, Newcastle, Australia

Abstract—Surveillance videos are crucial in imparting public security, reducing or avoiding the accidents that occur from anomalies. Crowd anomaly detection is a rapidly growing research field that aims to identify abnormal or suspicious behavior in crowds. This paper provides a comprehensive review of the state-of-the-art in crowd anomaly detection and, different taxonomies, publicly available datasets, challenges, and future research directions. The paper first provides an overview of the field and the importance of crowd anomaly detection in various applications such as public safety, transportation, and surveillance. Secondly, it presents the components of crowd anomaly detection and its different taxonomies based on the availability of labels, and the type of anomalies. Thirdly, it presents the review of the recent progress of crowd anomaly detection. The review also covers publicly available datasets commonly used for evaluating crowd anomaly detection methods. The challenges faced by the field, such as handling variability in crowd behavior, dealing with large and complex data sets, and addressing the imbalance of data, are discussed. Finally, the paper concludes with a discussion of future research directions in crowd anomaly detection, including integrating multiple modalities, addressing privacy concerns, and addressing crowd monitoring systems' ethical and legal implications.

Keywords—Crowd anomaly detection; advanced computer science; intelligent systems; video surveillance application; machine learning

I. INTRODUCTION

The field of crowd anomaly detection is rapidly growing, with increasing interest in identifying abnormal or suspicious behavior in crowds. Crowds can be found in various settings, such as public gatherings, transportation hubs, and shopping centers. The ability to detect anomalies in crowds has many important applications, including public safety, transportation management, and surveillance. The variety and frequency of indoor and outdoor activities that draw big crowds have been fast expanding, which has raised the likelihood of illegal gatherings, disturbance situations, mass stampedes, and other anomalous events greater than before [1]. The size and density of the crowds at huge gatherings have resulted in several public safety crises in recent years, such as the famous stampede on New Year's Eve of 2014 at Shanghai Bund group [2], Mina stampede during Hajj 2015 [3], etc. Surveillance applications are becoming more crucial for efficient crowd-control analysis. Such video surveillance systems must watch for unexpected crowd behavior, like crowd instability or chaos.

Video surveillance should be able to identify violent altercations or traffic accidents quickly and accurately. The

effectiveness of traditional methods is significantly constrained by the amount of human effort needed to make judgments manually. However, there is a rising need to express desired irregularities in an automatic and understandable manner as activities become more complicated and there are more possibilities to reason about. It is critical to comprehend the distinction between a crowd and a group. A group is any gathering of people for social interaction that can range in size from a few to many. On the other hand, a crowd is a group of people who have congregated in an uncontrolled or organized way for similar or dissimilar reasons. A crowd might thus be modelled in terms of many aspects, such as size, cohesiveness, structure, period, motive, and closeness. According to [4], the crowd is either structured or unstructured, the structured crowd moves in the same direction, speed, and pattern as in a marathon, a line of people waiting, or people using an escalator, etc. But the unstructured crowd exhibits complete uncertainty in their behaviors. Such crowds can be seen in marketplaces, shopping malls, etc., where the behavior is entirely uncertain.

In summary, this paper provides a comprehensive overview of crowd anomaly detection (CAD), beginning with a brief explanation of CAD systems and their importance in the research field. In Section III, we delve into the typical components of a CAD system, and the different taxonomies of CAD systems. The typical components include object density estimation, object tracking and object behavior analysis. Major taxonomies of crowd anomaly detection systems are presented based on the label availability and anomaly types used to categorize anomalies. Section IV provides an overview of recent advancements in crowd anomaly detection, while Section V presents a detailed analysis of the performance of various existing CAD systems and techniques. In Section VI, we outline publicly available datasets that can be used for crowd anomaly detection research. The challenges faced by existing systems are discussed in Section VII, including accuracy, computational complexity, and handling real-world scenarios. Finally, in Section VIII, we conclude with a summary of the key findings and potential future directions for research in this field. By the end of this paper, readers will have a comprehensive understanding of the current state of the art in crowd anomaly detection and will be equipped with the necessary knowledge to overcome existing challenges and push the boundaries of this research field even further.

II. CROWD ANOMALY DETECTION (CAD)

Crowd anomaly detection (CAD) is the process of understanding the overall characteristics of a crowd in a video,

such as density, flow, and demographic information. The crowd's density is the number of people per unit area and is used to measure the congestion level in a crowd. The flow of a crowd is the direction and speed of movement of people and is used to measure the level of mobility in a crowd. Several techniques have been developed to detect crowd anomaly in recent years. These techniques include image processing, computer vision, and machine learning. Image processing techniques, such as background subtraction and blob detection, are used to extract information about the density and flow of a crowd from video footage. Computer vision techniques, such as object detection and tracking, are used to extract information about the demographic characteristics of a crowd. Machine learning techniques, such as neural networks and, recently, deep learning models, are used to analyze the extracted information and make predictions about the crowd.

Anomaly detection in crowds can further be defined as identifying specific individuals or groups of people behaving abnormally, such as loitering, running, or moving against the flow of the crowd. This method involves tracking the movement of individuals in a crowd using computer vision techniques such as object detection and object tracking. Once an individual is tracked, their behavior can be analyzed to determine if it is normal or abnormal. Recently, anomaly detection and its analysis in social crowds have become a significant area of research. Due to the variety of anomalous events, crowd anomaly detection is a practical and challenging topic for computer vision. Automatic security analysis of crowd behavior is now possible when there are odd crowds or anomalous congestion. Because of activities like terrorist activities, fights, strange and suspicious movements, etc. automated detection of abnormal behavior in the crowd is of utmost relevance. In traditional systems, becomes the operators' responsibility to supervise the security surveillance to ensure safety closely. This is a significant challenge, resulting in costly and inaccurate decision-making. Therefore, creating a system that is free from errors and without any fatigue, providing real-time functionality, will have sufficient effects on managing crowd behavior.

The emergence of several sophisticated algorithms and the availability of high computational powers heightened the quantity and quality of the research in crowd anomaly detection. Computer vision algorithms that utilise image processing, machine learning and pattern recognition depict the challenges in crowd behavior analysis [5-7]. Some of its most crucial applications are crowd control, video surveillance, and the design of intelligent public spaces [8-14]. The intelligent environment, which is essential for public safety, can help to redirect the crowd and help the planner to design the public area with the most available space [4].

The increased number of research publications in the top publications related to crowd anomaly detection indicates the growing interest and demand in this field [5]. Fig. 1 presents the number of recent papers published on crowd anomaly detection. Crowd anomaly detection (CAD) is identifying unusual or abnormal behavior in a crowd using data analysis techniques, typically with the help of video cameras, sensors, or other monitoring devices. The goal of crowd anomaly detection is to identify situations that may pose a risk to public

safety or security, such as the presence of suspicious individuals or activities, overcrowding, or potential hazards. In order to achieve this goal, crowd anomaly detection systems employ machine learning algorithms and computer vision techniques to analyze data sources in real-time. By recognizing typical patterns of behavior, such as standing, sitting, or walking in specific areas, these algorithms can be trained to detect any deviations from the norm. In such cases, the system sends an alert to relevant authorities or security personnel, who can take the necessary steps to investigate the matter.

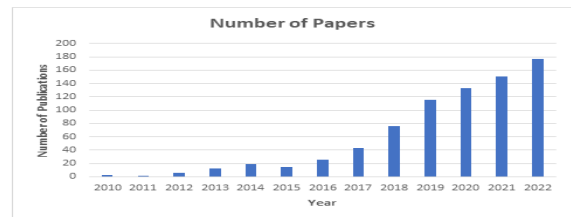


Fig. 1. Number of papers published on crowd anomaly detection.

III. COMPONENT AND TAXONOMY OF CROWD ANOMALY DETECTION

The general structure and flow of a crowd anomaly detection system is shown in Fig. 2. The anomaly detection system starts with the raw video data collected by the CCTV cameras. The sensor data are then pre-processed using various methods to lower signal noise and go through a feature extraction process. These features might include things like color, texture, motion, or shape. The goal is to identify patterns in the video that can help distinguish normal behavior from anomalous behavior.

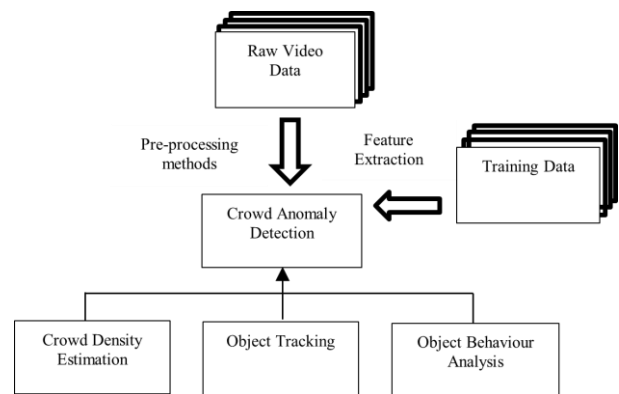


Fig. 2. Flow and structure of crowd anomaly detection.

Crowd anomaly detection system typically consists of three main components: crowd density estimation, object tracking, and object behavior analysis. Here is an overview of how these components work together. Crowd density estimation: The first step in crowd anomaly detection is to estimate the density of people in the scene. This can be done using various computer vision techniques such as background subtraction, foreground detection, or optical flow. The goal is to identify the regions of the scene where people are present and estimate the number of people in each region. Object tracking: Once the crowd density has been estimated, the next

step is to track individual objects (i.e., people) in the scene over time. This is typically done using a tracking algorithm that assigns a unique ID to each object and updates its position as it moves through the scene. Many different tracking algorithms are available, but some common techniques include Kalman filtering, particle filtering, and graph-based tracking. Object behavior analysis: Once individual objects have been tracked, the next step is to analyze their behavior to detect anomalies. This typically involves comparing the behavior of each object to some predefined normal behavior model. For example, the system might look for objects that are moving in an unusual direction, moving faster or slower than expected, or loitering in one area for an extended period. There are many different techniques for object behavior analysis, including rule-based methods, machine learning, and anomaly detection algorithms. The overall flow of the system is typically iterative, with the crowd density estimation and object tracking components running continuously in real time. The object behavior analysis component typically runs periodically (e.g. every few seconds) to detect any anomalies in the behavior of the tracked objects. When an anomaly is detected, the system can generate an alert or trigger some other action (e.g. turning on lights, sounding an alarm, or notifying security personnel).

Crowd density estimation is an important area of research with practical applications in managing and monitoring crowds in density populated locations such as subway stations, sports stadiums, and convention centers. With increasing population and urbanization, it is common for a large crowd to gather quickly. Precisely predicting the emergence of crowds and gauging their density is crucial for effective event planning and crowd management. The recent COVID-19 pandemic further highlights the importance of crowd density estimation, as social distancing policies were implemented to prevent the spread of the virus [15]. There are two primary methods for crowd density estimation: counting objects and estimating the density map [16]. CNN-based algorithms are preferred due to their better image and video sequence performance. A generic deep learning model for the automatic feature extraction from crowd scenes for crowd anomaly detection has been shown in Fig. 3. Techniques based on CNN, such as Scale Pyramid Module Network [17] and Attention Networks [18], are being used for crowd density estimation and counting. Attention Networks are capable of counting individuals in photos while considering scales by selecting appropriate global and local scales using the attention mechanism. Tracking the crowd is crucial in CAD systems, as it involves analyzing image sequences to determine the motion and trajectory of objects, specifically pedestrians. The process begins with detecting objects in a video and filtering them for tracking. The monitoring of pedestrian movement is an essential aspect of understanding crowd behavior. Object tracking can be challenging, as it requires following one or more objects over time. Automated tracking systems are needed to keep up with the movement of pedestrians in a crowd. Identifying and defining regions of interest (ROIs) is the first and most important step in this

process. This can be difficult due to various factors, such as the camera's view, orientation, and resolution. Once the ROI features have been extracted, the tracker can then follow the object of interest.

These models receive either supervised or unsupervised training. In order to take the necessary action, such as dispersing the crowd, when the crowd density exceeds a predetermined threshold, the degree of congestion can be calculated. The objects are tracked, and the anomaly is analyzed by utilizing the objects under tracking and their behavior. A final decision can be made in real-time whether the crowd state is normal or abnormal.

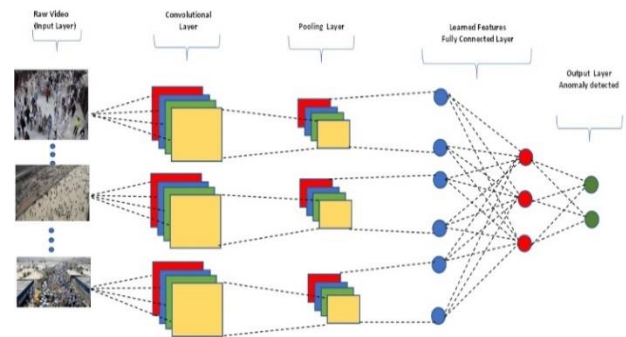


Fig. 3. A deep learning model for the automatic feature extraction from crowd scenes for crowd anomaly detection.

The research on crowd anomaly detection can be categorized into two types based on the availability of labels and the type of anomalies. The different types of crowd anomaly detection methods and related works on them are given in the following subsections.

A. CAD based on the Availability of Labels

Based on the availability of labels, the models used for crowd anomaly detection are divided into supervised/semi-supervised, unsupervised, hybrid, and one-class neural networks, as shown in Fig. 4. Supervised learning: This type of anomaly detection relies on labelled data, where the anomalies and normal behavior are defined beforehand. The model is trained on this labelled data, which can then be used to detect anomalies in new, unseen data [15], [16]. Semi-supervised learning: This type of anomaly detection also relies on labelled data, but it also uses unlabeled data to enhance the model's performance. The model is trained on labelled and unlabeled data, which can then be used to detect anomalies in new, unseen data. Unsupervised learning: This type of anomaly detection does not rely on labelled data. Instead, it uses techniques like clustering and dimensionality reduction to identify patterns in the data. Any deviation from these patterns can then be considered an anomaly. Given that they use labelled data, supervised anomaly detection approaches outperform unsupervised ones in terms of performance. From a series of annotated data instances (training), supervised anomaly detection learns the separation border. Using the learned model, it then divides a test instance into normal and anomalous classes (testing).

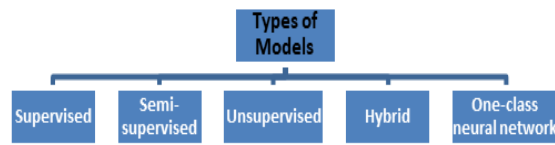


Fig. 4. Crowd anomaly detection based on the availability of labels.

Overviews of deep learning-based semi-supervised algorithms for anomaly identification were done in [15] and [6]. Unsupervised crowd anomaly detection is a significant study area, both for fundamental machine learning research and practical applications. One of the deep core architectures for unsupervised anomaly detection is autoencoders, as described in [17]. Discriminative boundary surrounding the majority class is learned by different anomaly detection techniques, as stated by Perera and Patel in their works [18], [19]. Any test instance that falls outside of this boundary is considered anomalous. To discover robust features, deep learning models are frequently utilized as feature extractors [20]. The hybrid models use two-step learning and are demonstrated to provide cutting-edge outcomes [21]. Robust characteristics are retrieved from the deep neural network's hidden layers to help separate them from the irrelevant features that can hide the existence of abnormalities. In deep hybrid models, one class Radial Basis Function (RBF) and Support Vector Machine (SVM) classifiers are used as inputs instead of more conventional techniques. A robust anomaly detection model needs both an anomaly detector and a feature extractor to be built on complicated, high-dimensional domains. One-class neural networks (OC-NN) combine deep network capabilities to extract more rich representations of data with a one-class objective, such as a hyperplane [22] or hypersphere [23], to distinguish all the typical data points from the outliers. Data representation in the hidden layer is learned by optimizing the objective function designed for anomaly detection. The experimental findings in [22] show that OC-NN may achieve equivalent or higher performance than current state-of-the-art approaches for complex datasets while having reasonable training and testing times in comparison to the existing methods.

B. CAD based on the Type of Anomaly

Crowd anomaly detection techniques can also be categorized based on the type of anomaly they handle. Mainly three types of anomalies are handled: point, contextual or conditional, and collective or group, as shown in Fig. 5. Most of the literature is devoted to pointing out anomalies. Point anomalies frequently signify an irregularity or deviation that occurs at random and may not have a specific meaning. A data instance that might be regarded as anomalous in a certain context is known as a contextual anomaly, also known as a conditional anomaly. By considering contextual and behavioral variables, a contextual abnormality is found. Time and space are two contextual elements that are frequently used. In contrast, the behavioral characteristics could be a pattern of financial spending, the occurrence of system log events, or any characteristic that characterizes typical behavior. Collective or group anomalies are abnormal groups of individual data points where each individual point, when

viewed separately, appears to be a typical instance of data but, when observed collectively, exhibit an unexpected characteristic.

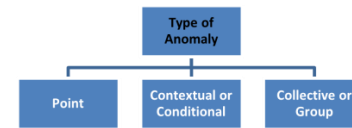


Fig. 5. Crowd anomaly detection based on the type of anomaly.

IV. RECENT PROGRESS ON CROWD ANOMALY DETECTION

In recent years, there has been a growing body of literature on crowd anomaly detection. Studies focus on various techniques and applications, including computer vision-based approaches, machine learning algorithms, and real-time anomaly detection. In this section, we review the most recent works on crowd anomaly detection and highlight their key findings and contributions to the field. The field of crowd anomaly detection and analysis has seen significant growth in recent years as the importance of understanding and managing crowd behavior has become increasingly apparent. With the proliferation of surveillance cameras and social media, there is a vast amount of data available for analyzing crowd behavior. As a result, numerous studies have been conducted on various aspects of crowd anomaly detection. However, with so much literature available, it can be difficult to gain a comprehensive understanding of the current state of research in this field. This literature review aims to address this issue by providing a comprehensive overview of the current state of research in crowd anomaly detection and analysis. By combining all available works on this topic, this literature review will provide a holistic view of the field, highlighting the most important contributions and identifying gaps in existing literature. A comprehensive examination of the various techniques used in the visual analysis of crowd behavior for surveillance purposes was conducted in [6]. In this study, recent research on crowd anomaly detection was classified based on the level of analysis and the types of anomalies observed. Several machine learning and deep learning approaches have been proposed to analyze crowd anomaly detection. A review of the different classical methods, such as the Spatial-Temporal Technique (STT), optical flow, Gaussian mixture model (GMM) and Hidden Markov model (HMM), has been done for the identification of the crowd abnormal behavior in [13]. Many recent studies in crowd anomaly analysis utilise deep learning methods. The different attributes of Convolution Neural Network (CNN) and the various optimization methods used in the context of crowd behavior analysis have been presented in [24]. Another literature review on intelligent video surveillance using various deep learning techniques for crowd detection was conducted in [17]. This work examined the use of Long-Short Term Memory (LSTM) networks, VGG16, and YOLO specifically. Additionally, [18] also conducted a detailed study of the various deep learning methods used for crowd counting and analysis, which are key components of crowd anomaly detection. Different methodologies utilized for counting the crowd have been reviewed and compared, presenting the various trends in CNN and traditional machine learning methods [25, 26].

Aggregation of Ensembles (AOE), a combination of four classifiers over sub-ensembles of three tuned Convolutional Neural Networks (CNNs) on crowd datasets, is proposed by [27] as a method to detect abnormalities in movies of crowded scenes via majority voting. A different classifier is used to process each sub-ensemble independently. A sub-ensemble of CNNs is composed of the CIFAR-10 AlexNet [28], GoogleNet [29], and VGGNet [30] networks. In this situation, CNNs acted as feature extractors, feeding Linear SVM, Quadratic SVM, Cubic SVM, and a SoftMax classifier and other classifiers. Several video frames are selected and analyzed to extract the features. A video is deemed abnormal if more than 10% of the batch's frames fall into this category. The publicly available datasets such as the Avenue dataset, UCSD Ped 1 and UCSD Ped 2, and AOE training and evaluation, were employed for this analysis.

A dual branch network was proposed in [35] as a framework for social multiple-instance learning. This approach uses a two-stream neural network consisting of an interactive dynamic stream and a spatiotemporal stream. The spatiotemporal stream processes RGB video clips. A 3D ConvNets (C3D) model that has already been trained on Kinetics and UCF-101 is used to extract features from the video after it has been divided into smaller pieces using a video segmentation method. The output is fed into a fully linked network after the features have been fed into a one-dimensional dependency attention capture module. The second channel receives force maps of social interactions as input. Maps of social interactions in a scene are created using the social force model [14]. They evaluate the effectiveness of their strategy on the UCF Crime dataset using the receiver operating characteristic (ROC) curve and area under the curve (AUC) metrics [31]. Comparisons with four more techniques were made throughout the evaluation. The dynamic network allows for a compact representation of moving video frames, which reduces false-positive anomaly alarms due to spatial limitations. This is the second advantage of their use. They employ the perturbation visual interpretation technique for identifying anomalies in order to give the results more credibility. The results presented were competitive with many of the similar works.

A general adversarial network-based abnormal behavior analysis in the massive crowd was proposed in [53], where a case study of the Hajj pilgrimage is considered. In this work, the dynamic features were extracted using optical flow. It uses a transfer learning strategy and U-Net and Flownet to distinguish large crowd scenes' normal and abnormal behaviors. This system has shown a very high accuracy in smaller video scenes such as UMN and UCSD, with 99.4% and 97.1%, respectively, whereas a considerably low accuracy on large datasets such as the Hajj dataset. The authors describe the need for improvement in anomaly behavior accuracy by collecting more annotated Hajj datasets and extracting complex features that utilise deeper models.

Behavior understanding becomes a difficult task with the fact that anomalies are not well defined and would occur very less frequently. Researchers have been trying to address these issues to make the learning algorithms robust in detecting the anomaly in the video. In [32], a new approach based on

Generative Cooperative Learning (GCL) has been employed for addressing the low frequency of anomalies and contributing towards the avoidance of manual annotations. The generator and discriminator networks in the model get trained in a cooperative style, thus enabling unsupervised learning. This approach has been found to be showing consistent improvement in UCF crime and ShanghaiTech, which are considered to be the well-cited large-scale video datasets.

A deep convolutional neural network (DCNN) architecture was proposed in [21] for detecting crowd anomalies. The architecture utilized the VGG16 model, which included ten convolutional layers and three max pooling layers. For crowd counting, they employed six convolutional layers with a dilation rate of 2 and a kernel size of 3x3. Another approach was proposed by Sagar [22], who presented a network architecture for crowd counting using a feature extractor based on ResNet. It extracts the details of an object at different scales by down-sampling the block with dilated convolutions and further up-samples the block using transposed convolution. In the analysis of crowd dynamics, Recurrent Neural Networks (RNNs) based models have been utilized [23]. In this work, the Bhattacharya distance was applied to detect a given frame's emotional state order to select the optimal keyframe for the video. To describe the scene, the Space-Time Interest Points (STIP) descriptor was used, with features being extracted from the keyframes. The RNN model was trained using an improved version of the Butterfly optimization algorithm, which enables it to distinguish between normal crowd behavior and behaviors associated with fighting, fleeing, walking, anger, happiness, and violence. The problem of vanishing and exploding gradients is addressed by Long-Short Term Memory (LSTM) networks [24], which are an extension of Recurrent Neural Networks (RNNs). LSTM has a longer memory capacity and can retain information for an extended period.

Most crowd anomaly detection robustness is analyzed based on the temporal consistency among the frames. The temporal features were considered for anomaly prediction based on the motion information using optical flow analysis [33, 34]. But there should be some system to distinguish between fake and real sequences for temporal consistency that lead to an anomaly or normal behavior. The anomaly behavior can be well detected by the optical flow methods, which are good for analyzing the short-term temporal relationship between adjacent frames. But eventually, these methods fail, especially in videos where events based on a long-term temporal relationship occur. To overcome such issues, a novel method based on a bi-directional architecture that introduces the inconsistencies on three different levels such as temporal-sequence, cross-modal, and pixel, has been proposed [35]. The bidirectional predictive network introduced in this work regularizes the predictive consistency. The long-term temporal relationship in the video sequences is identified by the discriminator developed in work. This method outperformed all other state-of-the-art learning methods on several datasets such as UCSD Pred2, ShanghaiTech and CUHK Avenue. Recent literature reviews have been conducted on the topic of crowd anomaly detection, providing in-depth coverage of the

various frameworks, taxonomies, methods, and techniques used. These reviews also included information on various datasets, such as the Hajj dataset used for video surveillance during the Hajj pilgrimage in Saudi Arabia. The Hajj pilgrimage is known to be the largest human gathering in the world, with an estimated of 2.5 to 3 million participants from various regions globally [9].

A summary of the different types of crowd anomaly detection is given in Table I.

TABLE I. SUMMARY OF THE CROWD ANOMALY DETECTION METHODS

Approach	Anomaly	Dataset	Performance
SSD-VGG16 [15]	Bullet train, pedestrian	PASCAL VOC, Railway	Overall accuracy=98.01% Detection accuracy=99.55%
SSD-VGG16 [16]	Small object	ILSVRC CLS-LOC, Railway	Accuracy =96.6%
3D-CNN LSTM [36]	Panics, fighting, protest	UMN, CAVIA, Web	Accuracy=0.995% Accuracy= 0.974% Accuracy= 0.926%
CNN RNN [37]	Use mobile in class, fighting, fainting	KTH, CAVIAR	Accuracy=87.15%
CNN Residual LSTM [38]	Fighting, explosion, accidents, shooting, robbery, shoplifting,	UCF-Crime, UMN, CUHK Avenue	Accuracy=78.43 % Accuracy=98.20 % Accuracy=98.80%
GAN [39]	Biking, fighting, vehicle, running	CUHK Avenue UCSD, ShanghaiTech Campus.	AUC=86.6% AUC=96.9% AUC=82.5% AUC=73.8%
Optical Flow GAN [40]	Standing, sitting, sleeping, running, moving in opposite, non-pedestrian	UMNScene1 UMNScene2 UMNScene3 UCSD, HAJJ datasets	Accuracy=99.4% Accuracy=97.1%, Accuracy=97.6%, Accuracy=89.26% Accuracy= 79.63%
Convolutional Neural Networks (CNNs) and Random Forests (RFs) [41]	standing, running, moving in opposite or different crowd directions, and non-pedestrian entities	UMN, UCSD, HAJJv2 dataset	Accuracy= 99.77% Accuracy= 93.71%. Accuracy=76.08%.
Convolutional Neural Network (CNN) [42]	Wheelchairs, skateboarders, motor vehicles, bicycles and crossing pedestrian tracks.	Violent Flows, UCSD, CUHK Avenue	Accuracy =90% Accuracy= 99.98% Accuracy=95%
Convolutional Long-Short-Term Memory (ConvLSTM) network and a Convolutional AutoEncoder. [43]	cyclists, skaters, cars	UCSD ped2, Shanghai Tech Campus.	AUC=95.6% AUC=73.1%

3DConv, Convolution Long Short-Term Memory (ConvLSTM) [44]	Vehicle and bicycle movement, throwing objects, running, Arrest, Abuse, Accident, Burglary, Explosion.	UCSD Ped1, UCSD Ped2, Avenue UCF-crime dataset	AUC=80.7% AUC=85.3% AUC=81.0% AUC=75.82%
CNN, RNN KNN, Optical Flow [45]	Bicycles, skateboards, wheelchairs	CUHK Avenue UCSD, UR fall Shanghai Tech Campus,	AUC=80.68% AUC=96.01% AUC=91.28% when k=10 AUC= 0.703 Optical flow module
Conv-LSTM [46]	Violence	Standard crowd anomaly	Accuracy =95.16%
Vgg-16 and LSTM [47]	Non-pedestrian	UCSD Ped2 CUHK Avenue	frame level: 95.0%, pixel level: 72.5% frame level:87.3%, pixel level: 93.8%
Cascaded attention model, Two Convolutional layers, Adam [48]	Fighting, Running, Robbery, lying down, crossing and car accident.	UCSD Ped2 CUHK Avenue, Shanghai Tech Campus	AUC =0.974, AUC=0.867, AUC= 0.736
3DCN, Transformer Adam [49]	Fighting, Running, Burglary, Fire and Assault.	Shanghai Tech Campus, UCF-Crime	AUC = 0.976 AUC = 0.832
GCN technique [50]	Fighting, Running, Burglary, Fire, Abnormal walking, lying down, group gathering and Assault.	UCSD, UCF-Crime, Shanghai Tech campus.	AUC =0.93, AUC=0.82 AUC=0.84.
motion attention, location attention. [51]	Fighting, Running, Burglary, Fire, Abnormal walking, lying down, group gathering, Assault, theft and Explosion.	UCSD Ped1, UCSD Ped2, CUHK Avenue, Shanghai Tech campus, UMN, Street Scene.	AUC=0.942, AUC=0.929, AUC=0.805, AUC=0.803, AUC=0.988 AUC=0.730.

V. PERFORMANCE COMPARISON

There are advantages and disadvantages to the crowd anomaly detection models that have been put forth so far. Most models just provide the output and estimate the findings in terms of accuracy, sensitivity, and specificity, failing to address the issue of output uncertainty. One of the issues that various anomaly detection techniques in a video frameset have in common is that they don't look at various situations including computational cost, pixel occlusion, noise and efficiency. To evaluate the models' performance and contrast it with other approaches, the following factors were considered:

1) The first is an analysis based on the amount of time required to run the algorithm for model estimation and a cost

analysis based on an estimate of the overall expenses associated with evaluation and error analysis.

2) Analysis of the uncertainty based on the dispersion of mean squares of error in different iterations and the average weight estimate of precision and recall on evaluating the performance.

3) Investigating the sensitivity to noise based on the classification of crowd behaviour, particularly in the presence of artifacts such as noise, a changing temperature, pixel occlusion, and low received frame quality.

4) Generalizations of the approaches for identifying and classifying person and group behavior in unseen frames.

Performance comparison of the most common datasets used in the CAD with different methods was shown in this section. The datasets used for the comparison are UMN [14], UCSD (University of California, San Diego) and UCSD Ped [53]. Performance comparison of UCSD on six different methods has been shown in the Fig. 6. CNN has the highest accuracy, with a score of 99.98%. GAN and CNN+KNN+Optical flow methods also have high accuracy, with scores of 96.90% and 96.01%, respectively. CNN+RF has a moderate accuracy score of 93.71%, while Optical Flow GAN has the lowest accuracy score of 89.25%.

In Table II, a performance comparison of UMN and UCSD Ped2 datasets are given. It can be observed that on the UMN dataset, the combination of CNN and random forest method achieved the highest performance, while on the UCSD Ped2 dataset; the Cascaded Attention CNN achieved the highest performance. It is also interesting to note that the UMN-Method achieved a higher performance than any of the methods applied to the UCSD Ped2 dataset.

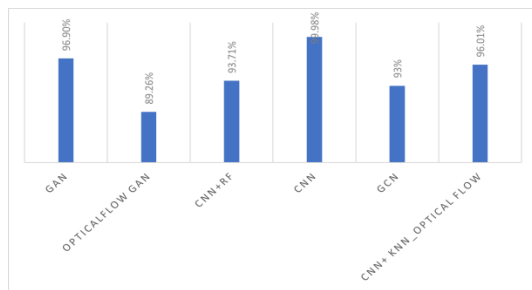


Fig. 6. Performance comparison of different methods on UCSD.

TABLE II. PERFORMANCE COMPARISON ON UCSD AND UMN DATASETS

UMN-Method	3D-CNN-LSTSM	CNN-Residual LSTM	Optical Flow GAN	CNN+RF
Performance	99.50%	98.20%	98.03%	99.77%
UCSD- Ped2 Method	Conv LSTM	Conv LSTM+ Conv Encoder	Vgg-16 LSTM	Cascaded Attention CNN
Performance	83.00%	95.20%	95%	97.40%

IV. PUBLICLY AVAILABLE DATASETS

Over the past few years, there has been a surge in the number of datasets devoted to crowd anomaly detection. These data sets can be used to analyze, compare and improve the performance of crowd anomaly detection systems.

Applications related to crowds, including counting, density estimates, categorization, activity recognition, and anomaly detection, use real crowd datasets. The majority of visual real crowd datasets have focused on counting tasks, including UCSD, PETS2009, UCF-CC-50, Mall, Shanghai Tech, etc. In a recent study, SIMulated Crowd Datasets (SIMCD) were presented for use in creating models for predicting and detecting crowd anomalies [52]. The details of these datasets along with many other datasets available for crowd anomaly detection have been given in Table III. The name of the dataset, the size of the dataset, a small description and a short scenario are mentioned in the table.

TABLE III. DETAILS OF THE PUBLICLY AVAILABLE DATASETS FOR CROWD ANOMALY DETECTION

Name	Scale	Description	Scenario
UMN [14]	Small	The collection consists of films from 11 different escape event scenarios shot in 3 various indoor and outdoor settings.	Crowd behavior anomaly detection
UCSD Peds 1 UCSD Peds 2 [53]	Small	Videos by a stationary camera gazing down on pedestrian walkways. peds1: perspective distortion and groups of people walking. Peds2: Camera-parallel pedestrian movement.	Abnormal crowd behavior detection
CVCS [54]	Medium	Cross-view, cross-scene, multi-view counting using a synthetic dataset. Each scenario in the dataset has roughly 100 camera views and consists of 31 scenes. For each scene, 100 crowd multi-view photos were taken.	Multi-view crowd counting
Grand Central [55]	Medium	It is collected from the New York Grand Central Station.	Crowd train station dataset
HAIJv1 [40]	Large	It is collected from pilgrims passing through hall passages in Haram masjid, Mecca, Saudi Arabia	Human abnormal behavior in Hajj
Shanghai Tech Part A Part B [56]	Large	It has 13 scenes with complex light conditions and different camera angles. It contains 130 abnormal events and over 270,000 training frames.	Crowd counting and density estimation
Violent flows [57]	Large	Video footage of crowd fighting together with industry-recognized benchmark standards for testing the accuracy of both the classification of violent and non-violent crowd behavior and the identification of violent outbreaks. 246 videos are included in the data collection. All of the videos	Classify and detect violent and non-violent behavior

		were downloaded from YouTube.	
WWW Crowd [58]	Large	A rich dataset for crowd understanding is provided by 10,000 videos with more than 8 million frames from 8,257 different scenes.	Crowd understanding
UCF-CC-50 [59]	Large	Extremely dense crowd dataset for crowd counting	for crowd counting
Multi-Task Crowd [60]	Large	A recent 100 image dataset that is completely annotated for crowd recognition, violent behavior detection, and categorization of density level.	Crowd counting, violence detection, and density level classification
UCF-Crime [31]	Large	It is made up of 1900 uncut, lengthy real-world surveillance movies that include 13 actual oddities like fighting, car accidents, robberies, and other crimes in addition to everyday occurrences.	Crowd behavior anomaly recognition, crowd behavior
Mall [61]	Medium	Webcams with public access are used to gather data. The video has 2000 frames, and each frame's annotations note the head position of each pedestrian.	Crowd counting
Street Scene [62]	Large	It is made up of video clips shot from a stationary USB camera looking down on a two-lane street with bike lanes and sidewalks, with another 35 clips used for testing.	Video anomaly detection.
CUHK-Avenue [63]	Small	It is a collection of short clips recorded by a single outdoor surveillance camera aimed at the side of a building facing a sidewalk. It has 15 sequences. Each one lasts approximately 2 minutes. There are 14 unusual events, including running, throwing objects, and loitering.	Abnormal event detection.

VI. CHALLENGES AND LIMITATIONS

In crowd scenarios with a variety of conditions, a reliable crowd anomaly detection algorithm tries to evaluate both local and global density and reliably anticipate crowd behaviour. The model's performance is significantly impacted by the changing circumstances. Therefore, it's crucial to first comprehend these difficulties and how they could affect the performance of the model. The development of more reliable models is aided by a thorough grasp of these difficulties.

a) Difficult to monitor crowd behaviour in various settings. For instance, it is significantly simpler to count and track the individuals in images captured by a single CCTV camera (such as those in the Mall dataset [61]) than to count people in images captured by numerous security cameras (such as those in the WorldExpo'10 dataset [64]). Drone surveillance often involves changing the scene, which makes it more difficult when combined with other variations such as scale variations.

b) When items belonging to the same class (in this case, humans) appear at various sizes in both a single photograph and across various images, it's referred to as scale variation. The distance (between the camera and the objects) and the perspective impact in the same image are the two factors that affect scale. Scale changes are additionally seen in photos with various resolutions. In crowd anomaly detection research, scale variation is one of the most prevalent issues that significantly affect model performance.

c) In different images, there are different numbers of people or other subjects of interest. Typically, low density visuals are simpler to understand than high density images. Similarly, another difficulty is when the same image shows various densities of people in multiple areas.

d) The distribution of objects in crowd photos may vary in different scenarios. For example, seats in a sports arena are evenly distributed among people, with constant spacing between objects, whereas in crowded streets, things might be distributed at random. In the absence of other features that would alter the accuracy, uniformly dispersed crowds can be estimated more precisely than non-uniform crowds.

e) Occlusion has been a major challenge in video analysis, and it is even difficult for crowd anomaly detection. The term occlusion describes how objects overlap. Intra-class occlusion refers to the overlapping of similar items (like humans), but inter-class occlusion refers to the overlapping of distinct objects (like automobiles, walls, and other people). Dealing with occlusion is frequently difficult. In the presence of occlusion, it is problematic for both the object detectors and the annotators to accurately annotate the objects as well as forecast them. Occlusion makes it challenging to distinguish between object borders in frames by interweaving semantic elements. It can also be challenging to learn when the object's pixel values are comparable to those of the background. Occlusion can limit the effectiveness of crowd anomaly detection systems as it makes it difficult to identify individuals in a crowd, leading to false positives or false negatives. Techniques like using multiple cameras or machine learning algorithms can help mitigate the effects of occlusion, but it can still be a significant challenge in dense and complex crowds.

f) Due to the different lighting conditions, the illumination in an image can change during the day and in various areas of the same image. This makes learning difficult because the same object (like people) in the same image will have varying pixel values. Other conditions which make CAD difficult include changes in the weather, noise and pixelation in images, rotation of objects, etc.

Some of the limitations of current crowd anomaly detection methods are given below:

- High computational cost: Some methods, such as deep learning-based methods, require a large number of computational resources to train and test, which can be a limitation in real-world applications where computational resources are limited.
- Need for annotated data: Many methods, such as deep learning-based methods, require a large amount of annotated data to train the models, which can be a limitation in real-world applications where data is limited or difficult to annotate.
- Ethical and privacy concerns: some methods, such as facial recognition-based methods, may raise privacy concerns, and it is important to ensure that the method is used in compliance with relevant laws and regulations.
- Limited ability to detect novel anomalies: Many methods are designed to detect known anomalies and may not be able to detect novel or unknown anomalies. Many methods focus on detecting anomalies but don't provide any insight into the root cause of the anomaly.
- Limited scalability: Many methods are not designed to handle large crowds and may not be able to scale to handle large amounts of data. Limited ability to handle multiple anomalies: Many methods focus on detecting a single type of anomaly and may not be able to handle multiple types of anomalies simultaneously.

VI. SUMMARY

This review paper provides an in-depth analysis of crowd anomaly detection and its importance in real-world surveillance and security. The latest research on crowd analysis is reviewed and summarized, with a focus on the major components of crowd anomaly detection, such as crowd density estimation, object tracking, and object behavior analysis. The paper also provides a summary of research based on different taxonomies, including the type of anomaly, and the type of dataset labels of the anomaly. Publicly available datasets used for crowd analysis were also reviewed, including the types of anomalies they address. Considerations for evaluating model performance and current challenges in the field were also discussed. In light of the review, the paper provides directions for future research, including the need for model generalization for different anomalies in different scenarios, designing application-specific crowd anomaly detection s, and effectively selecting the most appropriate models for analysis to reduce unnecessary resource usage and carbon emissions. The paper also provides directions for future research, including the incorporation of generative models, graph-based methods, reinforcement learning, transfer learning, online learning, ensemble methods, multi-task learning, domain adaptation, active anomaly detection, and meta-learning which have the potential to significantly improve the performance of crowd anomaly detection systems and address current limitations in the field. Overall, this review paper offers valuable insights and a comprehensive

understanding of the field of crowd anomaly detection, which will aid researchers in developing robust solutions to address the current limitations of the system.

VII. FUTURE DIRECTIONS

Some of the research directions based on the challenges identified by the various researches are mentioned next. The same benchmark datasets are used to train and evaluate the majority of crowd counting methods. So, model generalization hasn't been studied much, or studies are limited to the models that were fine-tuned on one dataset after being pretrained on another. However, the results that are normally published come from the refined model. This causes a significant disparity in the generalization of crowd counting models across different scenes, which warrants more research. It would be fascinating to observe model generalization in a variety of unusual situations, such as interior and outdoor videos, CCTV photos, drone photographs, etc.

An effective model must be able to run on a variety of hardware platforms with diverse computational capabilities, including servers, drones, cameras, mobile phones, etc., in order to support the potential applications of crowd anomaly detection. Applications (such as real-time or non-real-time), types of surveillance (such as CCTV-based surveillance or drone-based surveillance), and scenarios (shopping malls, metro stations, stadiums, etc.) all have different performance requirements. Therefore, it is ineffective to create a single optimal model with the highest accuracy for all applications, surveillance techniques, and circumstances. In actuality, such a model will be big, need a lot of computing power for fine-tuning, and have lengthier inference delays given the current trends in crowd anomaly detection model design. Applications requiring real-time inference, limited on-chip memory, and battery-powered devices will not be compatible with such a strategy. As a result, we anticipate and have also seen some recent attempts to have application-specific model designing, such as lightweight models for real-time applications on resource-constrained devices, and dense models for maximum accuracy over dense crowds in server-based systems.

The selection of deep or shallow neural networks is to be studied in detail. For sparse datasets with low crowd-density images, such as those from UCSD [53], Mall [61], and ShanghaiTech Part B [56], shallow models provide reasonably sufficient accuracy, and deeper models may not be necessary for the circumstances depicted in these datasets. Deeper models are typically used to achieve higher accuracy over large datasets, but these efforts often result in deeper and more complicated architectures. Unsurprisingly, single-scene crowd analysis and sparse multi-scene crowd detection are the two tasks with the fewest requirements. Even relatively tiny accuracy gains are the focus of most research efforts, and the ensuing model complexity is frequently disregarded. This leads to an increase in model complexity for a minor and frequently insignificant gain in accuracy. We think it's important to look into benchmarking for model training and inference times for crowd models. It's high time that researchers think about green computing and help to conduct low carbon emission systems for their research.

Incorporating domain knowledge, temporal information and multiple modalities into anomaly detection methods can improve their performance and make them more robust to different scenarios. Incorporating generative models, such as generative adversarial networks (GANs) and variational autoencoders (VAEs) can improve the ability to detect novel or unknown anomalies and can also be used to generate synthetic data for training models. Incorporating graph-based methods, such as graph convolutional networks (GCNs) and graph recurrent networks (GRNs), can improve the ability to model the relationships and interactions between individuals in the crowd. Ensemble methods, such as an ensemble of classifiers and an ensemble of experts, can improve the robustness and generalizability of the models by combining the predictions of multiple models. The concept of meta-learning can be used, which allows the model to learn how to learn, can improve the ability to adapt to new data and tasks, and can also improve the interpretability of the models. Developing interpretable models, such as decision trees and rule-based models, that can provide insight into why a certain behavior is considered abnormal can make the methods more understandable and trustworthy. Incorporating explainable AI(XAI) can reveal the reason for a certain decision and can improve the interpretability of the models and make the methods more trustworthy.

There are further many possible future directions in terms of techniques and methods used for crowd anomaly detection, such as incorporating generative models, graph-based methods, online learning, ensemble methods, multi-task learning, domain adaptation, active anomaly detection, and meta-learning. These techniques and methods can help to improve the performance and robustness of crowd anomaly detection models and make them more suitable for real-world applications.

REFERENCES

- [1] Y. Zhou, M. Qin, X. Wang, and C. Zhang, "Regional Crowd Status Analysis based on GeoVideo and Multimedia Data Collaboration," in 2021 IEEE 4th Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC), 2021, vol. 4: IEEE, pp. 1278-1282.
- [2] J. Duan, W. Zhai, and C. Cheng, "Crowd Detection in Mass Gatherings Based on Social Media Data: A Case Study of the 2014 Shanghai New Year's Eve Stampede," *Int. J. Environ. Res. Public Heal.* 2020, Vol. 17, Page 8640, vol. 17, no. 22, p. 8640, Nov. 2020, doi: 10.3390/IJERPH17228640.
- [3] M. Yamin, "Managing crowds with technology: cases of Hajj and Kumbh Mela," *Int. J. Inf. Technol.*, vol. 11, no. 2, pp. 229-237, Jun. 2019, <https://doi.org/10.1007/s41870-018-0266-1>.
- [4] J. Ma, Y. Dai, and K. Hirota, "A Survey of Video-Based Crowd Anomaly Detection in Dense Scenes," *Journal of Advanced Computational Intelligence and Intelligent Informatics*, vol. 21, no. 2, pp. 235-246, 2017, doi: 10.20965/jaciii.2017.p0235.
- [5] M. S. Zitouni, A. Sluzek, and H. Bhaskar, "Visual analysis of socio-cognitive crowd behaviors for surveillance: A survey and categorization of trends and methods," *Engineering Applications of Artificial Intelligence*, vol. 82, pp. 294-312, 2019.
- [6] B. Zhan, D. N. Monekosso, P. Remagnino, S. A. Velastin, and L.-Q. Xu, "Crowd analysis: a survey," *Machine Vision and Applications*, vol. 19, no. 5, pp. 345-357, 2008.
- [7] J. C. S. J. Junior, S. R. Musse, and C. R. Jung, "Crowd analysis using computer vision techniques," *IEEE Signal Processing Magazine*, vol. 27, no. 5, pp. 66-77, 2010.
- [8] S. D. Bansod and A. V. Nandedkar, "Crowd anomaly detection and localization using histogram of magnitude and momentum," *The Visual Computer*, vol. 36, no. 3, pp. 609-620, 2020.
- [9] V. J. Kok, M. K. Lim, and C. S. Chan, "Crowd behavior analysis: A review where physics meets biology," *Neurocomputing*, vol. 177, pp. 342-362, 2016.
- [10] B. Yogameena and C. Nagananthini, "Computer vision based crowd disaster avoidance system: A survey," *International journal of disaster risk reduction*, vol. 22, pp. 95-129, 2017.
- [11] X. Zhang, Q. Yu, and H. Yu, "Physics inspired methods for crowd video surveillance and analysis: a survey," *IEEE Access*, vol. 6, pp. 66816-66830, 2018.
- [12] N. Nayan, S. S. Sahu, and S. Kumar, "Detecting anomalous crowd behavior using correlation analysis of optical flow," *Signal, Image and Video Processing*, vol. 13, no. 6, pp. 1233-1241, 2019.
- [13] A. Afiq et al., "A review on classifying abnormal behavior in crowd scene," *Journal of Visual Communication and Image Representation*, vol. 58, pp. 285-303, 2019.
- [14] R. Mehran, A. Oyama, and M. Shah, "Abnormal crowd behavior detection using social force model," in 2009 IEEE conference on computer vision and pattern recognition, 2009: IEEE, pp. 935-942.
- [15] B. Guo, J. Shi, L. Zhu, and Z. Yu, "High-speed railway clearance intrusion detection with improved SSD network," *Applied Sciences*, vol. 9, no. 15, p. 2981, 2019.
- [16] L. Yundong, D. Han, L. Hongguang, X. Zhang, B. Zhang, and X. Zhifeng, "Multi-block SSD based on small object detection for UAV railway scene surveillance," *Chinese Journal of Aeronautics*, vol. 33, no. 6, pp. 1747-1755, 2020.
- [17] P. Baldi, "Autoencoders, unsupervised learning, and deep architectures," in Proceedings of ICML workshop on unsupervised and transfer learning, 2012: JMLR Workshop and Conference Proceedings, pp. 37-49.
- [18] P. Perera and V. M. Patel, "Learning deep features for one-class classification," *IEEE Transactions on Image Processing*, vol. 28, no. 11, pp. 5450-5463, 2019.
- [19] G. Blanchard, G. Lee, and C. Scott, "Semi-supervised novelty detection," *The Journal of Machine Learning Research*, vol. 11, pp. 2973-3009, 2010.
- [20] J. Andrews, T. Tanay, E. J. Morton, and L. D. Griffin, "Transfer representation-learning for anomaly detection," 2016: JMLR.
- [21] S. M. Erfani, S. Rajasegarar, S. Karunasekera, and C. Leckie, "High-dimensional and large-scale anomaly detection using a linear one-class SVM with deep learning," *Pattern Recognition*, vol. 58, pp. 121-134, 2016.
- [22] R. Chalapathy, A. K. Menon, and S. Chawla, "Anomaly detection using one-class neural networks," *arXiv preprint arXiv:1802.06360*, 2018.
- [23] L. Ruff et al., "Deep one-class classification," in International conference on machine learning, 2018: PMLR, pp. 4393-4402.
- [24] G. Tripathi, K. Singh, and D. K. Vishwakarma, "Convolutional neural networks for crowd behaviour analysis: a survey," *The Visual Computer*, vol. 35, no. 5, pp. 753-776, 2019.
- [25] Y. Luo, J. Lu, and B. Zhang, "Crowd counting for static images: a survey of methodology," in 2020 39th Chinese control conference (CCC), 2020: IEEE, pp. 6602-6607.
- [26] M. Bendali-Braham, J. Weber, G. Forestier, L. Idoumghar, and P.-A. Muller, "Recent trends in crowd analysis: A review," *Machine Learning with Applications*, vol. 4, p. 100023, 2021.
- [27] K. Singh, S. Rajora, D. K. Vishwakarma, G. Tripathi, S. Kumar, and G. S. Wallia, "Crowd anomaly detection using Aggregation of Ensembles of fine-tuned ConvNets," *Neurocomputing*, vol. 371, pp. 188-198, 2020, doi: 10.1016/j.neucom.2019.08.059.
- [28] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84-90, 2017.
- [29] X. Hu, Y. Huang, X. Gao, L. Luo, and Q. Duan, "Squirrel-cage local binary pattern and its application in video anomaly detection," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 4, pp. 1007-1022, 2018.

- [30] K. Simonyan and A. Zisserman, "VGGNet," in 3rd Int. Conf. Learn. Represent. ICLR, 2015.
- [31] W. Sultani, C. Chen, and M. Shah, "Real-world anomaly detection in surveillance videos," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 6479-6488.
- [32] M. Z. Zaheer, A. Mahmood, M. H. Khan, M. Segu, F. Yu, and S.-I. Lee, "Generative Cooperative Learning for Unsupervised Video Anomaly Detection," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 14744-14754.
- [33] R. Cai, H. Zhang, W. Liu, S. Gao, and Z. Hao, "Appearance-motion memory consistency network for video anomaly detection," in Proceedings of the AAAI Conference on Artificial Intelligence, 2021, vol. 35, no. 2, pp. 938-946.
- [34] G. Yu et al., "Cloze test helps: Effective video anomaly detection via learning to complete video events," in Proceedings of the 28th ACM International Conference on Multimedia, 2020, pp. 583-591.
- [35] C. Chen et al., "Comprehensive Regularization in a Bi-directional Predictive Network for Video Anomaly Detection," in Proceedings of the American association for artificial intelligence, 2022, pp. 1-9.
- [36] Y. Guan, W. Hu, and X. Hu, "Abnormal behavior recognition using 3D-CNN combined with LSTM," *Multimedia Tools and Applications*, vol. 80, no. 12, pp. 18787-18801, 2021.
- [37] C. Amrutha, C. Jyotsna, and J. Amudha, "Deep learning approach for suspicious activity detection from surveillance video," in 2020 2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA), 2020: IEEE, pp. 335-339.
- [38] W. Ullah, A. Ullah, T. Hussain, Z. A. Khan, and S. W. Baik, "An efficient anomaly recognition framework using an attention residual LSTM in surveillance videos," *Sensors*, vol. 21, no. 8, p. 2811, 2021.
- [39] Y. Hao, J. Li, N. Wang, X. Wang, and X. Gao, "Spatiotemporal consistency-enhanced network for video anomaly detection," *Pattern Recognition*, vol. 121, p. 108232, 2022.
- [40] T. Alafif, B. Alzahrani, Y. Cao, R. Alotaibi, A. Barnawi, and M. Chen, "Generative adversarial network based abnormal behavior detection in massive crowd videos: a Hajj case study," *Journal of Ambient Intelligence and Humanized Computing*, vol. 13, no. 8, pp. 4077-4088, 2021, doi: 10.1007/s12652-021-03323-5.
- [41] T. Alafif et al., "Hybrid classifiers for spatio-temporal real-time abnormal behaviors detection, tracking, and recognition in massive hajj crowds," arXiv preprint arXiv:2207.11931, 2022.
- [42] R. Lalit, R. K. Purwar, S. Verma, and A. Jain, "Crowd abnormality detection in video sequences using supervised convolutional neural network," *Multimedia Tools and Applications*, vol. 81, no. 4, pp. 5259-5277, 2021, doi: 10.1007/s11042-021-11781-4.
- [43] B. Wang and C. Yang, "Video Anomaly Detection Based on Convolutional Recurrent AutoEncoder," *Sensors*, vol. 22, no. 12, p. 4647, 2022.
- [44] X. Hu, J. Lian, D. Zhang, X. Gao, L. Jiang, and W. Chen, "Video anomaly detection based on 3D convolutional auto-encoder," *Signal, Image and Video Processing*, pp. 1-9, 2022.
- [45] K. Doshi and Y. Yilmaz, "A modular and unified framework for detecting and localizing video anomalies," in Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2022, pp. 3982-3991.
- [46] T. Saba, "Real time anomalies detection in crowd using convolutional long short-term memory network," *Journal of Information Science*, p. 01655515211022665, 2021.
- [47] L. Xia and Z. Li, "A new method of abnormal behavior detection using LSTM network with temporal attention mechanism," *The Journal of Supercomputing*, vol. 77, no. 4, pp. 3223-3241, 2021.
- [48] V.-T. Le and Y.-G. Kim, "Attention-based residual autoencoder for video anomaly detection," *Applied Intelligence*, vol. 53, no. 3, pp. 3240-3254, 2023.
- [49] D. Zhang, C. Huang, C. Liu, and Y. Xu, "Weakly supervised video anomaly detection via transformer-enabled temporal relation learning," *IEEE Signal Processing Letters*, vol. 29, pp. 1197-1201, 2022.
- [50] N. Li, J.-X. Zhong, X. Shu, and H. Guo, "Weakly-supervised anomaly detection in video surveillance via graph convolutional label noise cleaning," *Neurocomputing*, vol. 481, pp. 154-167, 2022.
- [51] S. Zhang et al., "Influence-aware attention networks for anomaly detection in surveillance videos," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 8, pp. 5427-5437, 2022.
- [52] A. Bamaqa, M. Sedky, T. Bosakowski, B. Bakhtiari Bastaki, and N. O. Alshammari, "SIMCD: SIMulated crowd data for anomaly detection and prediction," *Expert Systems with Applications*, vol. 203, 2022, doi: 10.1016/j.eswa.2022.117475.
- [53] A. B. Chan, Z.-S. J. Liang, and N. Vasconcelos, "Privacy preserving crowd monitoring: Counting people without people models or tracking," in 2008 IEEE conference on computer vision and pattern recognition, 2008: IEEE, pp. 1-7.
- [54] Q. Zhang, W. Lin, and A. B. Chan, "Cross-view cross-scene multi-view crowd counting," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 557-567.
- [55] B. Zhou, X. Wang, and X. Tang, "Understanding collective crowd behaviors: Learning a mixture model of dynamic pedestrian-agents," in 2012 IEEE Conference on Computer Vision and Pattern Recognition, 2012: IEEE, pp. 2871-2878.
- [56] Y. Zhang, D. Zhou, S. Chen, S. Gao, and Y. Ma, "Single-image crowd counting via multi-column convolutional neural network," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 589-597.
- [57] T. Hassner, Y. Itcher, and O. Kliper-Gross, "Violent flows: Real-time detection of violent crowd behavior," in 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 2012: IEEE, pp. 1-6.
- [58] J. Shao, K. Kang, C. Change Loy, and X. Wang, "Deeply learned attributes for crowded scene understanding," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 4657-4666.
- [59] H. Idrees, I. Saleemi, C. Seibert, and M. Shah, "Multi-source multi-scale counting in extremely dense crowd images," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2013, pp. 2547-2554.
- [60] M. Marsden, K. McGuinness, S. Little, and N. E. O'Connor, "Resnetcrowd: A residual deep learning architecture for crowd counting, violent behaviour detection and crowd density level classification," in 2017 14th IEEE international conference on advanced video and signal based surveillance (AVSS), 2017: IEEE, pp. 1-7.
- [61] K. Chen, C. C. Loy, S. Gong, and T. Xiang, "Feature mining for localised crowd counting," in *Bmvc*, 2012, vol. 1, no. 2, p. 3.
- [62] B. Ramachandra and M. Jones, "Street scene: A new dataset and evaluation protocol for video anomaly detection," in Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2020, pp. 2569-2578.
- [63] C. Lu, J. Shi, and J. Jia, "Abnormal event detection at 150 fps in matlab," in Proceedings of the IEEE international conference on computer vision, 2013, pp. 2720-2727.
- [64] C. Zhang, H. Li, X. Wang, and X. Yang, "Cross-scene crowd counting via deep convolutional neural networks," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 833-841.

Improving Brain Tumor Segmentation in MRI Images through Enhanced Convolutional Neural Networks

Kabirat Sulaiman Ayomide¹, Teh Noranis Mohd Aris², Maslina Zolkepli^{3*}

Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, Serdang, Malaysia

Abstract—Achieving precise tumor segmentation is essential for accurate diagnosis. Since brain tumors segmentation require a significant training process, reducing the training time is critical for timely treatment. The research focuses on enhancing brain tumor segmentation in MRI images by using Convolutional Neural Networks and reducing training time by using MATLAB's GoogLeNet, anisotropic diffusion filtering, morphological operation, and sector vector machine for MRI images. The proposed method will allow for efficient analysis and management of enormous amounts of MRI image data, the earliest practicable early diagnosis, and assistance in the classification of normal, benign, or malignant patient cases. The SVM Classifier is used to find a cluster of tumors development in an MR slice, identify tumor cells, and assess the size of the tumor that appears to be present in order to diagnose brain tumors. The proposed method is evaluated using a dataset from Figshare that includes coronal, sagittal, and axial views of images taken with a T1-CE MRI modality. The accuracy of 2D tumor detection and segmentation are increased, enabling more 3D detection, and achieving a mean classification accuracy of 98% across system records. Finally, a hybrid approach of GoogLeNet deep learning algorithm and Convolution Neural Network- Support Vector Machines (CNN-SVM) deep learning is performed to increase the accuracy of tumor classification. The evaluations show that the proposed technique is significantly more effective than those currently in use. In the future, enhancement of the segmentation using artificial neural networks will help in the earlier and more precise detection of brain tumors. Early detection of brain tumors can benefit patients, healthcare providers, and the healthcare system as a whole. It can reduce healthcare costs associated with treating advanced stage tumors, and enables researchers to better understand the disease and develop more effective treatments.

Keywords—MRI brain tumor; anisotropic; segmentation; SVM classifier; convolutional neural network

I. INTRODUCTION

All essential biological systems are controlled by the central nervous system (CNS), which is composed of the brain and spinal column which are the most important organs in our bodies' systems. The functions include speaking, moving, and thinking. They are made up of supporting cells called glial cells and nerve cells called neurons, which communicate with one another and with the rest of our body by both sending and receiving impulses via nerves. Magnetic resonance imaging (MRI), a category of imaging technology, provides accurate images of the anatomical anatomy of the human body, especially the brain, and provides useful data for biomedical research and clinical diagnosis. A brain tumor is the growth of malignant, abnormal cells in the brain [1]. There are numerous

varieties of brain tumors and these tumors may be malignant (cancerous) or benign (noncancerous). Cancer may begin in any other parts of the body before spreading to the brain or it may begin there (primary brain tumors) (secondary, or metastatic, brain tumors). The rate of growth of a brain tumor might vary greatly. A brain tumor will have an effect on how well your nervous system functions based on how quickly it develops and where it is. Your treatment options are influenced by the type of brain tumor you have as well as its size and location. A learning model for the classification of lung and pancreatic tumors was introduced [2] where a knowledge transfer served as the foundation for the learning model's 3D CNN architecture. The stated accuracy metrics of the transfer learning-based algorithms were superior to those attained by manually developed techniques. Applications specifically connected to neuro-oncology have increased interest in transfer learning. Deep data from brain MRI images have been extracted in research using trained networks. Transfer learning may be used to work with smaller datasets, as illustrated by studies that applied AlexNet and GoogLeNet to grade gliomas from MRI scans [1]. In terms of efficiency metrics, GoogLeNet performed better than AlexNet, where GoogLeNet's work on classifying brain anomalies used deep transfer learning and obtained impressive classification performance [2]. Wavelet transform (WT) [3] is an important strategy for feature extraction from MR brain images, but it necessitates a substantial amount of storage and is computationally expensive. Wavelet transform (WT) allows analysis of images at various levels of resolution due to its multiple-resolution analytic property [4]. Principal component analysis is appealing because it observe the impact the dimensionality of data and lowers the computational cost of processing new data [5]. The supervised classifier shows better performance than the unsupervised classifier in terms of classification success rates, regardless of the fact that all of these strategies led to successful outcomes. The classification accuracy of the preponderance of previous methods was less than 95%, therefore the purpose of the proposed research is to find a method that is significantly accurate. A CNN-based deep learning model was successfully used to solve the classification task for brain tumors. CNN-based classifier systems give a completely automated classifier that doesn't required manually segmented tumor regions, which is a benefit. For the purpose of identifying traits from brain MRI, a CNN architecture was introduced [6] where the effectiveness of the algorithm was enhanced by using CNN features and a classifier model from the category of extreme learning machines (ELM). Recall measures for the class meningioma were somewhat low for this trial; however they were quite high for the class pituitary

tumor. It illustrates that there are boundaries on the classifier's ability to discriminate. SVMs, which are state-of-the-art supervised classification methods based on machine learning theory, are currently used in many applications. Due to their extreme precision, exquisite mathematical tractability, and simple geometric interpretation, SVMs outperform existing methodologies like artificial neural networks, decision trees, and Bayesian networks by a massive margin [7]. Manual segmentation of brain tumors that is usually very challenging and time-consuming is frequently done by experts.

The remaining part of the paper is organized as follows: Section II the related works to the proposed approach. Section III describes the dataset use magnetic resonance brain image and the classifier settings which include the anisotropic filtering, support vector machine filtering, deep classification neural network feature and deep convolutional neural network. Section IV gives details of the experiment performed presents evaluation results and provides discussion on the results. Section V provides the conclusion of the study.

II. RELATED WORKS

The feature extraction approach uses a Stationary Wavelet Packet Transform (SWPT)[6]. The best feature is then selected by implementing a hybrid of adaptive black widow and moth flame optimisation (HABWMFO). The feature values undergo clustering in order to perform segmentation. For segmentation, the Adaptive Kernel Fuzzy C Means clustering technique (AKFCM) was developed. Finally, and certainly not least, the Dolphin-SCA deep learning method, which is based on deep CNN, implemented the Hybrid Convolution Neural Network-Long Short-Term Memory (CNN-LSTM) component to improve the accuracy of tumor classification.[7]. It helps with accuracy and enables efficient classification selections. Dolphin-SCA (Dolphin Echolocation based Sine Cosine Algorithm), along with the fuzzy deformable fusion model, were used to segment the data. The features were extracted using Power LDP and statistical traits (skewness, mean, and variance). The information was used to classify brain tumors using the Deep Convolution Neural Network, which trains using Dolphin-SCA. Brain tumors are segmented using a hybrid k-means algorithm and the fuzzy c-means technique [8]. For automatically identifying brain tumor cells in MRI MRI images, a deep CNN model has been created. On synthetic and real-time datasets, it has been tested [9]. Using the GoogLeNet and AlexNet architecture, the CNN network is trained on the training dataset, and the performance of the data model is assessed on the test dataset. The model's performance is evaluated using the metrics accuracy, sensitivity, specificity, and AUC integrating an implemented extreme learning machine with a hybrid feature extraction technique [10]. Using the RELM (regularised extreme learning machine), the tumor kind was classified. A Deep Wavelet Auto-encoder (DWA), which combines the auto-encoder, a fundamental feature reduction property, and the wavelet transform, an image decomposition method, is used for picture reduction [11]. A brain imaging dataset was produced using a DWA-DNN image classifier. A CNN-based U-net [12] for segmenting the tumor has been introduced where the transfer learning is based on a Vgg16 pre-trained convolutional foundation. To grade the tumor, a fully connected classifier was utilized.

III. MATERIALS AND METHOD

In the proposed approach, a novel upgraded classification model is proposed in an attempt to enhance tumor segmentation performance. Enhanced Convolutional Neural Networks are used to segment brain tumors. Automatic segmentation is carried out using the BAT algorithm, which strips the skull from the MRI image before applying the loss function to preprocess it. Deeper architectures are built utilising smaller kernels. Despite decreasing the probability of computational complexity given a network includes fewer weights [13].

The proposed methodology incorporated anisotropic diffusion, morphological operation, and SVM segmentation. On the figshare open dataset, the suggested approach had the best classification performance. Combining transfer learned deep CNN features and SVM classifier models improve performance. The combination excels even with less training samples utilised in the original input space on the dataset conducted in MATLAB, achieving the highest classification accuracy. The dataset comprises of 3064 brain MRI images taken from 233 patients who have been diagnosed with one of the three forms of brain tumors, which are meningioma, glioma, or pituitary tumors. The T1-CE MRI modality includes the axial, sagittal, and coronal images. It includes 1426 brain MRI images with glioma (corresponding to 89 patients), 708 meningioma photos (related to 82 patients), and the remaining 930 pituitary tumor images (belonging to 62 patients). Each image is 512x512 pixels in size and is accessible mat files. The input layer for Google Net's original RGB colour image design was 224x224x3. In the proposed approach, which aims to segment tumor using MRI images, automatic classification has been performed under Anisotropic Diffusion as tumor found and no tumor found. CNN, a successful deep learning method, was used as a classifier within the scope of the study. CNN architectures were used GoogLeNet. The classification results were obtained using the original MRI brain images, and the images obtained by using SVM. Then, new classification results were calculated using a framework from the obtained results.

A. Anisotropic Diffusion Filter

In order to perform morphological operations on the MRI image and determine if the patient's brain contains a tumor or not, the picture was first filtered using an anisotropic diffusion filter to lessen contrast between adjacent pixels. The image was then manually turned to black and white using a threshold value after being scaled. It is the first step in identifying potential tumor locations [14]. Morphological techniques have been used on the semi-processed image to acquire data on the areas and solidity of the likely places. The statistical average of many MRI pictures with tumor has been used to calculate the minimal value of both of these characters. The final detection result was then delivered using it [15]. Although the majority of the time the simulation technique can produce accurate results, it cannot function when the tumor is hollow or too small. getting an MRI scan, using an anisotropic diffusion filter to filter the image, processing the image with morphological operations in detail, the tumor is divided up by a specific border, if there is no tumor, image processing is finished; otherwise, the tumor's location is estimated. Tumor

identification is based on density and size [16]. Each parameterized image that results from shape-adapted smoothing is a combination of the original image and a filter that depends on the local content of the original image. As a result, anisotropic diffusion transforms the original image in a non-linear and space-variant way. The photos are blurred with the anisotropic filter without losing any edges.

$$\frac{\partial I}{\partial t} = \text{div} (c(x, y, t), \nabla I) = \nabla c \cdot \nabla I \quad (1)$$

where the diffusion coefficient is denoted by $c(x, y, t)$. It is chosen to maintain image edges and regulates the rate of diffusion. The ground-breaking Perona-Malik paper's projected diffusion co-efficient is as follows:

$$c(|\nabla I|) = \exp\left(-\frac{|\nabla I|}{K}\right)^2 \text{ or } c(|\nabla I|) = \frac{1}{1+\left(\frac{|\nabla I|}{K}\right)^2} \quad (2)$$

B. Data Pre-Processing

Morphology is a technique for removing visual cues like boundaries and skeletons that aid in the legation and recital of region shapes. It works by enlarging things, filling gaps, and then connecting all the broken things. The size of the thing is also diminished. The erosion process in a binary image eliminates the foreground background pixels. The image is subjected to morphological opening after being converted into a binary image [17]. To segment out the tumor spot from the image, a Binary tumor masked window must be created. Noise removal occurs during the pre-processing step and can be accomplished using a variety of spatial filters, including linear and nonlinear filters (Median filter). After segmentation, morphological processing is used to remove undesirable parts, and other artefacts, such as text, are removed.

C. Morphological Operation

It entails image opening, image closing, dilation, erosion operations, and a conclusion has been made regarding whether or not there is a tumor present in the MRI picture and whether it is normal or pathological. It is also where the RGB to grey conversion and reshaping happens. It has a median filter to reduce noise. The likelihood of noise appearing during a modern MRI scan is extremely low. It might come as a result of heat influence. The images frequently employs the following elements: threshold, edge, pixels, cluster, and neural network [18]. The study of morphology focuses on form-based image processing. An output image of the same size is produced by adding a structural element to the previously processed input image. Every surrounding pixel is compared to every input pixel, and the comparison's findings are used to determine the values of the corresponding output pixels. Erosion and dilation are the two primary morphological processes. Pixels are either added to the boundaries of the objects during dilation or subtracted from them during erosion, depending on the size and shape of the structuring elements. When comparing, the dilation action gives the output the greatest value of the adjacent surrounding pixels, while the erosion operation gives the output the lowest value of the nearby surrounding pixels. There are various clustering techniques, including the mountain, K-means, fuzzy C-means, and subtractive approach.

The most popular clustering method is called k-means clustering [19]. Compared to hierarchical clustering, it is more straightforward, speedier, and capable of handling a high number of variables.

The technique of segmenting an image into different parts or sections is known as picture segmentation. The main objective is to make the visual representation simpler or transform it into something else that is more straightforward and analytically insightful. It is employed to identify borders and objects in pictures. Every pixel in an image has a label assigned to it, and pixels with the same label have similar properties. Image classification is a well-known issue in image processing. The main objective of image classification is to foretell the categories of the input images using the features. Various classifiers exist, including ANNs (Artificial Neural Networks) and SVMs (Support Vector Machine). The segmentation stage is crucial for a thorough analysis of a picture because it determines how accurate the succeeding phases will be [20]. However, due to the wide variety of lesion shapes, sizes, and colours as well as various skin types and textures, effective segmentation is challenging. Additionally, some lesions have uneven borders, and occasionally, the border between the tumor and the skin is smooth. Many algorithms have been put up to solve the issue. Thresholding, edge-based or region-based, supervised and unsupervised classification approaches are some general categories that can be used to describe them. The normalisation intensity values for the MRI images in the dataset were likewise pre-processed [21].

```
%% Morphological Operation

label=bwlabel(sout);
stats=regionprops(logical(sout), 'Solidity', 'Area', 'BoundingBox');
density=[stats.Solidity];
area=[stats.Area];
high_dense_area=density>0.6;
max_area=max(area(high_dense_area));
tumor_label=find(area==max_area);
tumor=ismember(label,tumor_label);

if max_area>100
    figure;
    imshow(tumor)
    title('tumor alone','FontSize',20);
else
    h = msgbox('No Tumor!!!','status');
    %disp('no tumor');
    return;
end
```

Fig. 1. Morphological operation.

The intensity values between 0 and 1 were scaled using a min-max normalisation approach and resized to 224x224. The three channels were then produced by replicating the greyscale values three times because MRI pictures are greyscale images. A five-fold cross-validation at the patient level was followed by the evaluation of the created system using the Figshare dataset as shown in Fig. 1.

D. Convolutional Neural Network (CNN) for Tumor Classification in MRI Images

A deep neural network and deep learning architecture called a convolutional neural network (CNN) is applied by using transfer learning, where a deep neural network is utilized to both create new CNNs from scratch and to employ ones that already exist. The deep learning architecture is produced by combining sub-layers such as the convolution layer, activation function, pooling layer, flattening layer, and completely connected layer. The convolution layer is where the convolution procedure is carried out by segmenting the image[22]. On the other hand, activation functions are architectural elements that, depending on their function types, produce new outputs from their inputs. The layer on which pooling operations are carried out in order to reduce the larger image size caused by the convolution process is known as the pooling layer. Prior to entering the classification layer, the picture whose convolution procedures have been finished must be transformed from matrix to vector form. Feature matrices are converted into feature vectors by the process of flattening. [23][24]The classification procedure employing feature vectors and machine learning is the fully connected layer. One of the options, including support vector machines and artificial neural networks, might be chosen as machine learning.[3] Within the parameters of the study, brain tumors were segmented using transfer learning and a convolutional neural network. Feed-forward training of CNNs begins with the first input layer and continues to the final classification layer; following that, error back-propagation begins with the final classification layer and moves forward to the first convolutional layer. In a forward pass computed as follows, neuron j of layer l-1 provides input to neuron I in layer l.

$$\bar{I} n_i^l = \varepsilon_d^\Pi = 1^{w_{ij}^l \cdot x_j + b_i} \quad (3)$$

The output is computed by a nonlinearity ReLu function:

$$\text{Out}_i^l = \max(0, \text{In}_i^l) \quad (4)$$

Equations (3) and (4) are used by all neurons in the convolutional and fully connected layers to calculate the input and provide an output in the form of nonlinear activation. The K-by-K square window sliding on the N-by-N feature map is used by the pooling layer, which uses the maximum or average value of the features inside the window. It produces a single value from the K K region, reducing the spatial dimension of the feature map from N N to N/K N/K. The Softmax function is used in the last layer to calculate the classification probability for each tumor type.

$$\text{out}_i^l = \frac{e_i n_i^l}{\varepsilon_{i \in \text{Out}_k^l}} \quad (5)$$

By minimising the following cost function with regard to the unknown weights, back-propagation algorithms train

ACNNs. X I is the ith sample in the training set with the label yi, and The training set's overall training sample count is indicated by m.

$$C = -\frac{1}{m} \sum_i^m (p(y^i | X^i)) \quad (6)$$

E. Figshare Dataset of T1-CE MRI Modality MRI Images

The figshare dataset is freely accessible and frequently used to test classification and retrieval techniques. It consists of 3064 brain MRI pictures from 233 patients who have been identified as having one of the three types of brain tumors (meningioma, glioma, and pituitary tumors). The coronal, sagittal, and axial views are all part of the T1-CE MRI modality. It includes 1426 brain MRI images with glioma (corresponding to 89 patients), 708 meningioma photos (related to 82 patients), and the remaining 930 pituitary tumor images (belonging to 62 patients). Each image is 512x512 pixels in size and is accessible as.mat files. The input layer for GoogLeNet's original RGB colour image design had a size of 224x224x3.

F. Classifier Settings

1) *Anisotropic filtering*: The primary goal of image filtering is to eliminate noise from digital photographs. The noises severely impair the image's quality. The noise in the image can be removed in a variety of methods. In a noisy environment, the majority of image processing algorithms struggle to function. Fig. 2. Indicate the pre-processing tool is the image filter. Anisotropic Filter, one type of filter, is utilized in the propsoed approach for denoising. In order to explain the image diffusion process, the generic anisotropic diffusion equation is introduced as follows:

$$\frac{dx}{dy} = \text{div}(c(m, n, t) \nabla c \cdot \nabla x + c(m, n, t) \nabla x) \quad (7)$$

where, c (m, n, t) is the diffusion coefficient and ∇ is the image gradient.

The signal pixels exhibit weak diffusion action, noise pixels exhibit high diffusion action. As a result, noise can be reduced while maintaining the signal. The fixed step size for each iteration or the entire iterative process of the image can be adopted by a variety of diffusion models. Here, a more effective iteration step is suggested in the equation, where 1/4 is employed to guarantee the equation's convergence.

$$dt = \frac{1}{4} c \quad (8)$$

Iterative processes are used to produce the final output phase image. Iteration error (IE), whose formula is used to manage the iterative number during the iteration process.

$$1E = \frac{\|I^n - 1^{n-1}\|}{\|1^{n-1}\|} \leq T_i e \quad (9)$$

The iterative procedure is terminated by $IE \leq \text{Tolerance}$ (Tie).

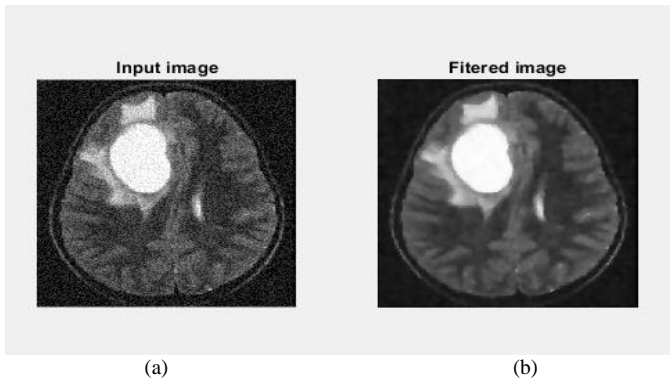


Fig. 2. Input and output of anisotropic filter.

2) *Support vector machine filtration:* In order to avoid over-fitting and local minima and improve generalisation capabilities, SVM relies on a different induction principle known as structural risk minimization. Algorithms for SVM use a set of mathematical functions referred to as the kernel. SVM has proven to have greater generalisation abilities in sparse samples and high dimensional space. The goal is to locate the best separating hyperplane that can accurately categorise all samples. Image segmentation is the process of dividing an image into different sections. Data is inputted into the kernel, which then transforms it into the desired form [25]. SVM employ a variety of kernel functions, including sigmoid, radial basis function (RBF), polynomial, linear, and nonlinear. For sequence data, graphs, text, pictures, and vectors, kernel functions may be introduced. However, RBF is the most popular type of kernel function because it has a localised and finite response along the entire x-axis, doesn't increase computational complexity, effectively overcomes the problem of dimensionality, and returns the inner product between two points in an appropriate feature space. The maximum-margin hyperplane can be fitted using KSVMs in a changed feature space. In Fig. 3, the classifier may be a hyperplane in the higher-dimensional feature space, but nonlinear in the original input space if the transformation is nonlinear and the transformed space is higher dimensional.

Formula for defining a hyperplane is

$$P(x) = \alpha_0 + \alpha^T z \quad (10)$$

where, α is the Weight vector, α_0 is the Bias and z is the training instances that are closest to the hyperplane. Best Possible Plane Legation is

$$|\alpha_0 + \alpha^T z| = 1 \quad (11)$$

The distinction between a point z and a hyperplane is revealed in the result.

$$(\alpha, \alpha_0): D = \frac{|\alpha_0 + \alpha^T z|}{\|\alpha\|} \quad (12)$$

The canonical hyperplane's numerator is one, and the difference between it and the support vectors is,

$$D = \frac{|\alpha_0 + \alpha^T z|}{\|\alpha\|} = \frac{1}{a} \quad (13)$$

For the canonical hyperplane, the numerator is equal to 1, and the difference from the support vectors is,

$$M = \frac{2}{a} \quad (14)$$

In the end, the maximising problem for M and the minimising problem for a function are the same. $R(\infty)$ is constrained in numerous ways.

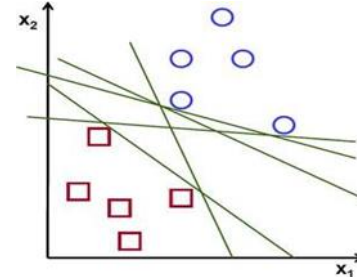


Fig. 3. Find a separating straight line for a set of 2D points that may be linearly divided into two classes.

3) *Deep convolutional neural network features:* On a variety of difficult visual analysis tasks, deep convolutional neural networks have been consistently outperforming other methods. Recently, deep convolutional neural networks have been used to classify images from enormous image datasets. In order to enable the description of latent concepts for pattern recognition, a deep CNN is able to automatically learn fundamental filters and combine them in a hierarchical manner. However, overfitting and lengthy processing times are issues with many deep CNNs. In order to learn the characteristics of brain MRI pictures with tumors, the suggested algorithm makes use of updated and improved GoogLeNet. In the remainder of the paper, the features retrieved using a modified version of GoogLeNet will be referred to as deep CNN features. GoogLeNet is referred to as a deep network, having a total of 22 learnable layers [26]. It has a fully connected layer, nine inception modules, two pooling layers, two convolutional layers, and two pooling layers. Once more, an inception module has a pooling layer in addition to six convolution layers. The architecture of the module includes filters in the sizes 1x1, 3x3, and 5x5. It is anticipated that filters with various kernel sizes will detect various data patterns. At the conclusion of each module, the feature maps corresponding to various filters are concatenated. Additionally, smaller kernel convolutions are carried out before bigger kernel convolutions. They are designed to reduce computations' dimension. The training set was used to train the modified GoogLeNet (after pre-processing). Heuristic adjustments were made to the network's hyperparameters to speed up the convergence of the loss function during training.

Deep convolutional neural network with features with support vector machine classifier: The modified GoogLeNet's pooling layer, which was implemented after the previous inception module, was used to extract features. After that, the traits were categorised using SVM. We used a multi-class SVM model with error-correcting output code (ECOC). A one-

vs-all strategy was used for multi-class categorization. There were three binary SVM learners with linear kernels.

Principles of Linear SVMs:

Given an N-size, p-dimensional training dataset of the following kind

$$\{(x_n, y_n) | X_n \in \mathbb{R}^p, y_n \in \{-1, +1\}\}, n = 1, \dots, N \quad (15)$$

where, either y_n is -1 or 1, which corresponds to class 1 or class 2. A p-dimensional vector represents each x_n . We want the support vector machine that separates class 1 from class 2 along the maximum-margin hyperplane.

$$wx^- = b0 \quad (16)$$

where, W is the hyperplane's normal vector and represents the dot product. The W and b should be chosen to maximise the margin between the two parallel hyperplanes while maintaining data separation. Thus, we use the following equations to define the two parallel hyperplanes:

$$wx^- = b1 \quad (17)$$

Binary SVMs pick up on the decision function $f(\cdot; w, b): \mathbb{R}^p \rightarrow \{-1, +1\}$ defined by $f(x; w, b) = \text{sign}(w \cdot x - b)$. Let Y now be a collection of $t > 2$ classes, which stands for the classes of tumor types. Now, we're going to take a look at t output functions, one for each class, that express the degree of certainty for each prediction. Equation defines the output function for a stylish $\in Y$.

$$F(x, \omega, b) = \langle w, \varphi(x) \rangle + b \quad (18)$$

To boost the forecast's level of confidence, the predicted class y_b for a given point x is determined, with the formula $y_b = \text{argmax}_y Y f_y(x; w, b)$. Finding w and b parameters that, at the very least, substantially result in accurate predictions and satisfy equation is what training is all about. $w, b; f_y(x) > f_u(x; w, b)$.

IV. RESULTS AND DISCUSSION

The performance of the proposed approach is validated using performance matrices like Accuracy, Precision, F-measure and Loss. The MATLAB application from MathWorks is used to carry out the experimental process. Brain tumors are recognised by a GoogLeNet deep learning algorithm. The Convolutional Neural Network and Sector Vector Machine are coupled to choose and classify the features (CNN-SVM). Anisotropic filter is utilized as a pre-processing tool, while morphological analysis is used as an optimisation tool. The values of the underlying variables should be generated using the continuity equation in order to evaluate the efficacy of the suggested approach.

A. Metrics for Performance

The measurements for the suggested model are provided in the section, some of which have been examined and verified. The following mathematical diagram illustrates the study of performance measures [6]:

1) Accuracy: Many correct patterns to the total number of patterns is how accuracy to precision ratio is calculated.

$$\text{Accuracy} = \frac{TP+TN}{TP+FN+FP+TN} \quad (19)$$

2) Precision: Precision is the ratio of positively anticipated values to all positively predicted values.

$$\text{Precision} = \frac{TP}{TP+FP} \quad (20)$$

3) Recall: The true positives and false negatives numbers are obtained using the recall measure. The equation below shows that the recall and sensitivity are both known quantities.

$$\text{Recall} = \frac{TP}{TP+FN} \quad (21)$$

4) F-score: The precision and recall values that are combined to get the F-score are used to calculate the score. The observation can be used to calculate the recall and precision weighted average.

$$\text{F-score} = 2 \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (22)$$

5) Loss: One maximum performance metric to evaluate the effectiveness of the SVM classifier. Moreover, it is evaluated using the error function.

B. Performance Analytics

The suggested model outperforms previous models, as shown by the statistical statistics provided above. The proposed approach is illustrated and flowchart in Fig. 4 in the approach recommended in the suggested technique.

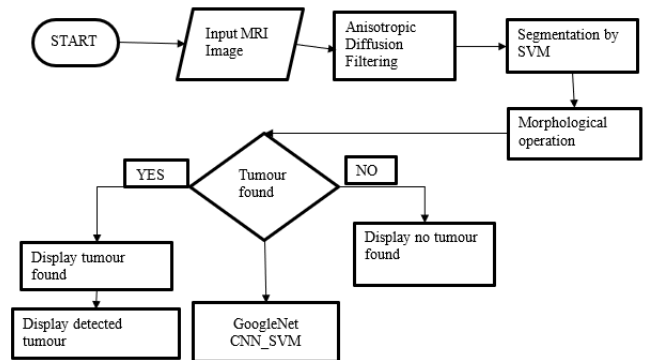


Fig. 4. The proposed methodology's flowchart process.

TABLE I. COMPARISON OF THE PROPOSED AND EXISTING METHODS' PERFORMANCE

Predictive Methods	Performance of Classifier				
	Accuracy	Precision	Recall	F-Measure	Loss
DNN	0.8641	0.8365	0.8254	0.86	1.6522
SVM	0.7225	0.7212	0.7356	0.73	1.2566
ANN	0.8316	0.7963	0.7792	0.83	0.9325
Hybrid CNN-LSTM	0.9785	0.9538	0.9781	0.96	0.7840
Proposed Method	0.9864	0.9317	0.9860	0.97	0.8630

In Table I, accuracy is represented by 0.9864, precision by 0.9317, recall by 0.9860, F-measure by 0.97, and loss using the proposed technique is represented by 0.8630. The accuracy rating for the Hybrid CNN-LSTM is 0.9785, the precision rating is 0.9538, the recall rate is 0.9781, the F-Measure is 0.96, and the loss rate is 0.7840. The accuracy rating for the

ANN method is 0.8316, the precision rating is 0.7963, the recall rate is 0.7792, the F-measure is 0.83, and the loss rate is 0.9325. Precision, recall, accuracy, and loss value for the SVM technique are 0.7225, 0.7212, 0.7356, and 1.2566, respectively, for the overall F-measure. Additionally, using the DNN technique, the accuracy is 0.8641, precision is 0.8365, recall is 0.8254, F-Measure is 0.86, and loss is 1.6522.

In Fig. 5, the accuracy of the CNN-SVM brain tumor detection is displayed. It takes about 32 seconds to complete the training process and displays the accuracy result with 58 epochs per iteration. Also, the accuracy is indicated by the blue graphical line. As a result, for a successful evaluation procedure, the performance of the proposal can be trained and validated. As a result, the performance is quite effective for the suggestion.

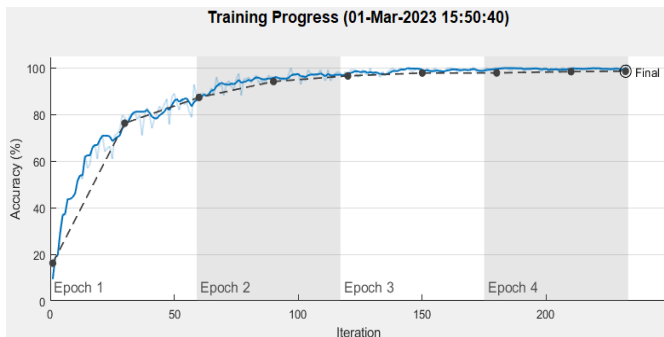


Fig. 5. The results that are accurate based on the graphical representation.

As seen in the graph below, the suggested attribute selection utilising the CNN-SVM model results in a loss with the least degree of detection error. When iteration is increased, the detection error of the suggested solution decreases. The suggested methodology is shown to improve accuracy and error function as a result. Fig. 6 Illustrates the loss function of the transfer learned model during training.

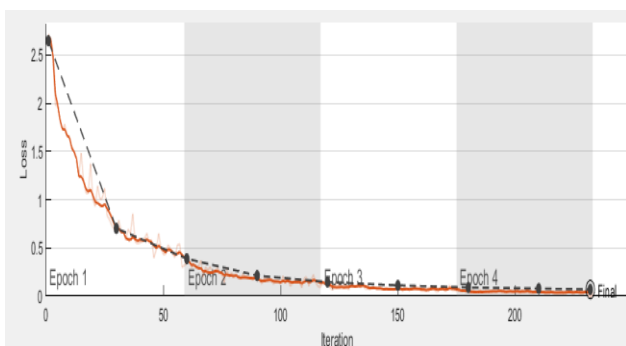


Fig. 6. The loss function of the transfer learned model during training.

The model's accuracy is demonstrated in Table II. By analysing the results and processing time. It may also be concluded that processing speed and results depend on the gradient of the pixels, the size of the image, and the quality of the image.

TABLE II. THE RESULTS OF THE TIME RESEARCH ON VARIOUS IMAGES OF BRAIN TUMORS SHOWING HOW TIME DEPENDS ON THE PREDICATED CRITERIA

No	MRI Brain Image.jpg	Preprocessing Time in seconds	Type of Image	Remarks	ACC %
i		16.268	-	Tumor not found	97
ii		19.989	Benign	Tumor found	99
iii		14.719	Benign	Tumor found	98
iv		29.401	Benign	Tumor found	99
v		11.451	Benign	Tumor found	96
vi		12.866	Malignant	Tumor found	95
vii		11.121	Malignant	Tumor found	97
viii		13.275	Malignant	Tumor found	99
ix		14.224	Malignant	Tumor found	96
x		16.263	-	Tumor not found	99

V. CONCLUSION

The proposed research combines Deep CNN characteristics and an SVM classifier powered by GoogLeNet's convolutional neural network (CNN) to segment brain tumors. The modified GoogLeNet's pooling layer, which was implemented after the previous inception module, was used to extract features. After that, the traits were categorised using SVM. Image processing has been successfully used to locate tumors in MRI brain scans, efficiently identifying the tumor's position by illuminating the area if it were there.

A database of 2D tumor image data is generated from the MRI scans acquired from different angles of a specific patient and the images are analyzed to pinpoint the precise 3D location of the tumor. The proposed method of classifying brain images using CNN and SVM has the potential to be a useful tool for computer assisted clinical diagnosis. The attributes increased performance when employed with tested classifier models. The most significant new finding of the study is a strategy for integrating them to distinguish between benign and malignant MRI brain as it produces a successful early tumor detection technique using computer-aided diagnosis. Early detection of brain tumors can benefit patients, healthcare providers, and the healthcare system as a whole. Patients can benefit from early detection by receiving timely treatment, which can improve their chances of survival and reduce the risk of complications. Healthcare providers can benefit by being able to provide more effective treatment and improve patient outcomes. Early detection can also reduce healthcare costs associated with treating advanced stage tumors. Additionally, early detection can enable researchers to better understand the disease and develop more effective treatments.

The future work can be advanced by regularly improving the model's accuracy by training it on larger and more comprehensive datasets. Several improvements are still feasible despite the successes presented in this research; there was a significant misclassification of samples from the dominant, and the phenomena of overfitting with smaller training data was noted. These problems should be addressed in the domain's future study, perhaps with data augmentation and additional fine-tuning of the transfer learnt model.

ACKNOWLEDGMENT

This research was not funded by any grant

REFERENCES

- [1] E. Hussein, S. Daoud, H. Alrabaiah, and R. Badawi, "Exploring undergraduate students' attitudes towards emergency online learning during COVID-19: A case from the UAE," *Child. Youth Serv. Rev.*, vol. 119, no. November, p. 105699, 2020, doi: 10.1016/j.childyouth.2020.105699.
- [2] F. Murat, O. Yildirim, M. Talo, U. B. Baloglu, Y. Demir, and U. R. Acharya, "Application of deep learning techniques for heartbeats detection using ECG signals-analysis and review," *Comput. Biol. Med.*, vol. 120, no. February, p. 103726, 2020, doi: 10.1016/j.combiomed.2020.103726.
- [3] E. Özbay and F. Altunbey, "Computer Methods and Programs in Biomedicine Interpretable features fusion with precision MRI images deep hashing for brain tumor detection," vol. 231, 2023, doi: 10.1016/j.cmpb.2023.107387.
- [4] R. Sille, T. Choudhury, P. Chauhan, and D. Sharma, "A systematic approach for deep learning based brain tumor segmentation," *Ing. des Syst. d'Information*, vol. 26, no. 3, pp. 245–254, 2021, doi: 10.18280/ISI.260301.
- [5] V. Bruni, M. L. Cardinali, and D. Vitulano, "A Short Review on Minimum Description Length: An Application to Dimension Reduction in PCA," no. Mdl, pp. 1–16, 2022.
- [6] R. S. Devi, B. Perumal, and M. P. Rajasekaran, "Advances in Engineering Software A hybrid deep learning based brain tumor classification and segmentation by stationary wavelet packet transform and adaptive kernel fuzzy c means clustering," *Adv. Eng. Softw.*, vol. 170, p. 103146, 2022, doi: 10.1016/j.advengsoft.2022.103146.
- [7] S. Kumar and D. P. Mankame, "Optimization driven Deep Convolution Neural Network for brain tumor classification," *Biocybern. Biomed. Eng.*, vol. 40, no. 3, pp. 1190–1204, 2020, doi: 10.1016/j.bbe.2020.05.009.
- [8] K. Maheswari, A. Balamurugan, P. Malathi, and S. Ramkumar, "Hybrid clustering algorithm for an efficient brain tumor segmentation," *Mater. Today Proc.*, vol. 37, no. Part 2, pp. 3002–3006, 2020, doi: 10.1016/j.matpr.2020.08.718.
- [9] C. Swarup, K. Udham Singh, A. Kumar, S. Kumar Pandey, N. varshney, and T. Singh, "Brain tumor detection using CNN, AlexNet & GoogLeNet ensembling learning approaches," *Electron. Res. Arch.*, vol. 31, no. 5, pp. 2900–2924, 2023, doi: 10.3934/era.2023146.
- [10] A. Gumaedi, M. M. Hassan, M. R. Hassan, A. Alelaiwi, and G. Fortino, "A Hybrid Feature Extraction Method with Regularized Extreme Learning Machine for Brain Tumor Classification," *IEEE Access*, vol. 7, pp. 36266–36273, 2019, doi: 10.1109/ACCESS.2019.2904145.
- [11] P. Kumar Mallick, S. H. Ryu, S. K. Satapathy, S. Mishra, G. N. Nguyen, and P. Tiwari, "Brain MRI Image Classification for Cancer Detection Using Deep Wavelet Autoencoder-Based Deep Neural Network," *IEEE Access*, vol. 7, pp. 46278–46287, 2019, doi: 10.1109/ACCESS.2019.2902252.
- [12] M. A. Naser and M. J. Deen, "Brain tumor segmentation and grading of lower-grade glioma using deep learning in MRI images," *Comput. Biol. Med.*, vol. 121, no. February, p. 103758, 2020, doi: 10.1016/j.combiomed.2020.103758.
- [13] M. M. Thaha, K. P. M. Kumar, B. S. Murugan, S. Dhanasekaran, P. Vijayakarathick, and A. S. Selvi, "Brain Tumor Segmentation Using Convolutional Neural Networks in MRI Images," *J. Med. Syst.*, vol. 43, no. 9, 2019, doi: 10.1007/s10916-019-1416-0.
- [14] R. Ranjbarzadeh, A. Bagherian Kasgari, S. Jafarzadeh Ghouschi, S. Anari, M. Naseri, and M. Bendeche, "Brain tumor segmentation based on deep learning and an attention mechanism using MRI multimodalities brain images," *Sci. Rep.*, vol. 11, no. 1, pp. 1–17, 2021, doi: 10.1038/s41598-021-90428-8.
- [15] A. Bhandari, J. Koppen, and M. Agzarian, "Convolutional neural networks for brain tumor segmentation," *Insights Imaging*, vol. 11, no. 1, 2020, doi: 10.1186/s13244-020-00869-4.
- [16] A. Alexander, A. Jiang, C. Ferreira, and D. Zurkiya, "An Intelligent Future for Medical Imaging: A Market Outlook on Artificial Intelligence for Medical Imaging," *J. Am. Coll. Radiol.*, vol. 17, no. 1, pp. 165–170, 2020, doi: 10.1016/j.jacr.2019.07.019.
- [17] A. S. Panayides et al., "AI in Medical Imaging Informatics: Current Challenges and Future Directions," *IEEE J. Biomed. Heal. Informatics*, vol. 24, no. 7, pp. 1837–1857, 2020, doi: 10.1109/JBHI.2020.2991043.
- [18] A. Chattopadhyay and M. Maitra, "Neuroscience Informatics MRI-based brain tumor image detection using CNN based deep learning method," *Neurosci. Informatics*, vol. 2, no. 4, p. 100060, 2022, doi: 10.1016/j.neuri.2022.100060.
- [19] Z. Lv, Y. Hu, H. Zhong, J. Wu, B. Li, and H. Zhao, "Parallel K-Means Clustering of Remote Sensing Images Based on MapReduce," in *WISM 2010*, 2010, vol. 6318, no. March, pp. 254–262, doi: 10.1007/978-3-642-16515-3.
- [20] K. Islam, S. Ali, S. Miah, and M. Rahman, "Machine Learning with Applications Brain tumor detection in MR image using superpixels, principal component analysis and template based K-means clustering algorithm," *Mach. Learn. with Appl.*, vol. 5, no. January, p. 100044, 2021, doi: 10.1016/j.mlwa.2021.100044.
- [21] R. Ranjbarzadeh, A. Caputo, E. Babae, S. Jafarzadeh, and M. Bendeche, "Brain tumor segmentation of MRI images: A

- comprehensive review on the application of artificial intelligence tools,” *Comput. Biol. Med.*, vol. 152, no. December 2022, p. 106405, 2023, doi: 10.1016/j.combiomed.2022.106405.
- [22] J. Kang, Z. Ullah, and J. Gwak, “Mri-based brain tumor classification using ensemble of deep features and machine learning classifiers,” *Sensors*, vol. 21, no. 6, pp. 1–21, 2021, doi: 10.3390/s21062222.
- [23] S. Deepak and P. M. Ameer, “Brain tumor classification using deep CNN features via transfer learning,” *Comput. Biol. Med.*, vol. 111, no. June, p. 103345, 2019, doi: 10.1016/j.combiomed.2019.103345.
- [24] K. Jaspin and S. Selvan, “Biomedical Signal Processing and Control Multiclass convolutional neural network based classification for the diagnosis of brain MRI images,” *Biomed. Signal Process. Control*, vol. 82, no. April 2022, p. 104542, 2023, doi: 10.1016/j.bspc.2022.104542.
- [25] W. Wu et al., “An Intelligent Diagnosis Method of Brain MRI Tumor Segmentation Using Deep Convolutional Neural Network and SVM Algorithm,” *Comput. Math. Methods Med.*, vol. 2020, 2020, doi: 10.1155/2020/6789306.
- [26] D. Daimary, M. B. Bora, K. Amitab, and D. Kandar, “Brain Tumor Segmentation from MRI Images using Hybrid Convolutional Neural Networks,” *Procedia Comput. Sci.*, vol. 167, no. 2019, pp. 2419–2428, 2020, doi: 10.1016/j.procs.2020.03.295.

Rain Streaks Removal in Images using Extended Generative Adversarial-based Deraining Framework

Subbarao Gogulamudi^{1*}

Research Scholar
Department of Computer Science
and Engineering
Annamalai University, Annamalai
Nagar, Chidambaram, Tamil Nadu
608002-India

V. Mahalakshmi²

Assistant Professor
Department of Computer Science
and Engineering
Annamalai University, Annamalai
Nagar, Chidambaram, Tamil Nadu
608002-India

Indraneel Sreeram³

Professor
Department of Computer Science
and Engineering
St. Ann's College of Engineering and
Technology, Chirala, Andhra
Pradesh 523187-India

Abstract—The visual quality of photographs and videos can be negatively impacted by various weather conditions, such as snow, haze, or rain, affecting the quality of the images and videos. Such impacts may greatly affect outdoor vision systems that rely on image/video data. It has recently drawn a lot of interest to remove rain streaks from a single image. Several deep learning-based methods have been introduced to address the issue of removing rain streaks from a single image. Still, the efficiency of rain streak removal with enhanced quality is challenging. Hence, a novel deep-learning method is introduced for rain streak removal. The proposed Extended Generative Adversarial based De-raining (Ex_GADerain) is the enhanced version of a traditional Generative adversarial network (GAN). The proposed Ex_GADerain introduced a Self-Attention based Convolutional Capsule Bidirectional Network (SA-CCapBiNet) based generator for enhancing the rain streaks removal process. Also, the loss function estimation using the adversarial loss and the mean absolute error loss minimizes the information loss during training. The minimal information loss enhances the generalization capability of Ex_GADerain, and hence the enhanced performance is acquired. The quality assessment of a derained image based on various assessment measures like SSIM, PSNR, RMSE, and DSSIM improved performance compared to the conventional rain streak removal methods. The maximal SSIM and PSNR acquired by the Ex_GADerain are 0.9923 and 26.7052, respectively. The minimal RMSE and DSSIM acquired by the Ex_GADerain are 0.9367 and 0.0051, respectively.

Keywords—Deep learning; rain streaks removal; image generation; quality measure; capsule network; adversarial learning

I. INTRODUCTION

In recent scenarios, outdoor vision systems are mainly impacted by awful weather conditions, including rain [1]. Due to the large light scattering and motion velocities, the raindrops generally produce bright streaks in the captured images or videos through cameras [2]. Such conditions can highly influence the image's visual quality and also degrades outdoor vision systems' efficiency [3]. Also, it affects the efficacy of several computational vision mechanisms like event detection, scene analysis, and object detection and action recognition. Thus, such effects due to rain can be resolved by performing an automated rain streak removal process [4]. During rainy conditions, the streaks of rain generate haziness and blurring impact in images because of

light scattering [5]. Thus, effective approaches are highly needed for many practical applications to remove rain streaks from the captured images or videos [6].

The primary goal of raindrop removal is to reduce the rain effects, which has been analyzed extensively [7]. Computer vision techniques mainly consider a pure image as input for understanding a scene. However, the available rain streaks can blur the scene and degrade the performance of such techniques [8]. The major problem in deraining approaches is exploiting the necessary attributes of rain streaks and the pure image [9-12]. The model-based methods define the removal of rain streaks as an optimization issue [13]. It contains hand-crafted regularizers that mention the preceding knowledge of a solution, like repeatability, high-frequency probability of the rain streaks and the image's piecewise smoothness [14]. Nevertheless, such model-based mechanisms are unsuitable for rainy conditions since the degradation process can become more complicated [15]. To solve this issue, learning-based schemes are utilized to learn the essential attributes from the data like convolutional filters, stochastic distributions and Gaussian mixture models (GMMs) [16-18]. In recent years, deep learning techniques are becoming more popular for detection analysis. Such techniques can acquire the data characteristics by a trained deep neural network and with an effective representation capability, attaining appropriate outcomes and promoting the data to a large extent.

The conventional deep learning techniques face two difficulties while removing rain streaks. At first, the trained network's efficiency is mainly based on the training data. For instance, the deep detail network (DDN) is mainly focused on understanding the nonlinear mapping in detail layer to the rain streaks with straightforward network structures. Secondly, because of the complicated network design, several deep learning techniques can suffer [19]. The rain streaks removal methods are generally categorized based on the input type. Single-image methods are employed when the input is in image form. On the other hand, the single image techniques demand the need of image priors to regain the fundamental background scene, such as nonlocal self-similarity prior, dictionary-based sparse prior, GMM-based layer prior and low rank prior [20]. However, the existing methods face several challenges while eliminating rain streaks in the images.

A. Research Questions

Some of the research questions concerning the proposed rain streak removal process are:

- 1) What is the need for including the noise removal technique through the filtering process?
- 2) How the efficient of the de-raining process enhanced with minimal computational complexity?
- 3) How the hybrid deep learning process generates the de-rained image by the generator module of the framework?
- 4) How the proposed method solves the problems faced by the existing methods?

The proposed study attempted to design a deep-learning model for removing rain streaks from input images. The major contributions of the research are:

- To develop a novel Extended Generative Adversarial based Deraining framework for eliminating awful rain streaks in input images.
- To attain needed information about rain streaks, a detailed layer is extracted in the proposed work with a hybrid filtering method.
- To obtain a de-rained image, a Self-Attention-based Convolutional Capsule Bidirectional Network is introduced in the generator block of the proposed framework.
- To evaluate different performance matrices for analyzing the efficacy of proposed framework with other existing methods.

II. RELATED WORKS

Some recent studies on rain streaks removal through different techniques are described as follows: Wang et al. [21] developed a kernel-guided convolutional neural network (KGCNN) for eliminating rain streaks from a single image. This existing study involved three important steps for performing rain streak removal. The motion blur kernel was initially learned through a plain neural network named parameter network from the raining patch's detail layer. Next, the learned motion blur kernel was stretched into a degradation map with a similar spatial size as the rainy patch. Finally, the developed deraining network, along with the ResNet design, was trained with the help of stretched degradation map with the detail patches. The simulation analysis shows that the developed model obtained optimal performance; however, the overall processing time is enhanced.

Lin et al. [22] introduced a sequential dual attention network for removing rain streaks in a single image. For this purpose, the proposed study designed a framework named Sequential dual attention based Single image DeRaining deep Network (SSDRNet) model. An inherent correlation between rain streaks in a given input image should be more powerful than that among the rain streaks and background of an image. Thus, a two-stage learning mechanism was applied to acquire the spread of rain streaks in an image effectively. The developed two-stage deep neural network contains three

varied blocks such as multi-scale feature aggregation modules (MAMs), sequential dual attention blocks (SDABs) and residual dense blocks (RDBs). The details about rain streaks of the given image were learned through the developed two-stage mechanism and perfectly eliminates the noise. The simulation results show that the developed model achieved more appropriate SSIM and PSNR values than other methods.

Hettiarachchi et al. [23] introduced conditional generative adversarial networks (CGANet) to ignore rain streaks in the provided single images. This existing work utilized the adversarial loss in generative adversarial networks (GANs), which affords a new component to the loss function. It helps to maintain the outcome and assists in obtaining higher performance. Here, a generator network was employed to map the rainy images to de-rained images and a discriminator network was utilized to categorize actual and created de-rained images. Varied performance measures were employed to compute the efficacy of a developed model through synthesized and realistic images. The result analysis shows that the developed CGANet model has more potential than other competing methods.

Darney et al. [24] presented the rain streak removal process through a dictionary-based sparsity process with MCA estimation. This existing study introduced a sparse coding process to eliminate rain streaks by applying morphological component analysis (MCA). Using MCA, estimation, of course, becomes effortless to manage the rainy streaks in the provided images. Through sparse decomposition, removing and estimating each sample redundancy is highly feasible. To obtain optimal MSE and PSNR outcomes from the recovered images, the developed MCA scheme is integrated with the process of sparsity coding. Furthermore, the developed model attained reduced MSE value; however, the computational complexity is enhanced.

Chang et al. [25] presented the curriculum learning model for eliminating the rain streaks from input images. This study uses a direction and residual awareness network to ignore unwanted rain streaks to obtain a clear image. To prove the efficacy of a developed model, a statistical analysis was performed on the extended-scale actual rainy images and plotted that rain streaks in normal patches exhibit principal directionality. By applying a direction-aware network, the directionality property was endowed and helps to differentiate rain streaks from an image edge. Finally, residual aware block (RAB) was introduced to analyze the relationship between the image and residual. This model assisted in learning the balance parameters to emphasize the necessary image features and provided positive outcomes.

Yang et al. [26] devised a rain streak removal process using the fractal band learning based strategy based on self-supervised approach. Here, the discriminative features were extracted from the input rained image using the fractal band learning strategy. Besides, the regularization of the network was devised for enhancing the generalization capability of the network through the cross scale learning approach. The robustness of the model was depicted through the quantitative analysis. Still, the failure in considering the texture features degrades the performance of the model.

Wang et al. [27] designed a deep learning based image de-raining model, wherein the shared source residual module was incorporated in the conventional deep convolutional neural network for making the skip connection to solve the vanishing gradient issues. The computational overhead evaluated by the designed model was minimal with enhanced outcome. Still, the method degrades its performance compared to state of art methods due to the negative perception.

Ran et al. [28] introduced an image de-raining through the patch analysis. In this, patch analysis was devised based on the task driven strategy, wherein K-shot learning was utilized for removing the rain streaks from the image. Besides, the computational overhead of the designed model was minimized through the skip connections and accomplished superior outcome in removing the rain streaks. Still, the computational complexity of the model was higher.

Wang et al. [29] suggested a rain streak removal method using the joint depth estimation approach. Here, the dilated residual network was incorporated in the generator module of the conventional GAN network for generating the depth map to remove the rain streaks. The robustness of the model was proved based on both the quantitative and qualitative analysis. Still, the processing time of the model was higher.

B. Problem Statement

The rainy weather environment highly influences the visibility of scene objects in acquired real-time images. Similarly, the visualization of a high-definition image captured by cameras degrades mainly under outdoor weather scenarios, including fog, snow and rain. These worst visual quality of real-time images can impact the nature of image surveillance, computer vision and multimedia applications. Hence, eliminating the rain streaks from the captured images is an essential demand in real-time multimedia applications. The rainy image generally defines the linear sum of the rain layer and a background image. In this case, restoring a blurred image due to rainy streaks is challenging because of several interconnection feasibilities among a specific image's rainy and background layers. Different techniques were developed to separate rainy streaks and clear images effectively. However, they failed to afford an appropriate performance because varied orientations and shapes of rain streaks degraded the images. Thus, to overcome limitations observed in previous studies, the proposed study introduced an effective rain streak removal mechanism by inspiring the higher efficiency of deep learning techniques. The proposed Ex_GADerain approach removes the rain streaks from the image more efficiently by incorporating the novel SA-CCapBiNet in the generator module of the conventional GAN technique. Besides, the computation overhead is reduced through the detail layer extraction through the CGBiF.

III. PROPOSED EX_GADERAIN-BASED RAIN STREAKS REMOVAL

Varying weather conditions like rain, snow, fog, and haze influence the visual quality of outdoor captured images. Rain patterns generate more effects on the image's visual quality than others. In addition, the rain patterns also degrade the forward and background information of an image. Thus,

removing rain streaks from the captured images improves the necessity. Most recently, deep learning models have gained more attention for removing rain streaks in images. Therefore, the proposed study planned to design a deep learning mechanism called Extended Generative Adversarial based De-raining (Ex_GADerain) framework to eliminate the rain streaks from the provided input images.

The proposed framework involves four stages: image acquisition, detail layer extraction, de-raining operation and classification. Initially, the input images are acquired from the publicly available dataset. The detail layer is extracted through Cross Guided Bilateral Filtering (CGBiF) method to attain essential information about rain streaks, removing unwanted noises. Then, the detail layer is provided as an input of the proposed Ex_GADerain framework. This framework contains two important blocks: generator and discriminator. The generator block produces the de-rained image through a new Self-Attention based Convolutional Capsule Bidirectional Network (SA-CCapBiNet) model. The discriminator block of the GAN model classifies whether the generated de-rained image is real or fake. Thus, the proposed deep learning framework effectively removes the rain streaks from the input dataset images. The block diagram for Ex_GADerain-based Rain Streaks Removal is depicted in Fig. 1.

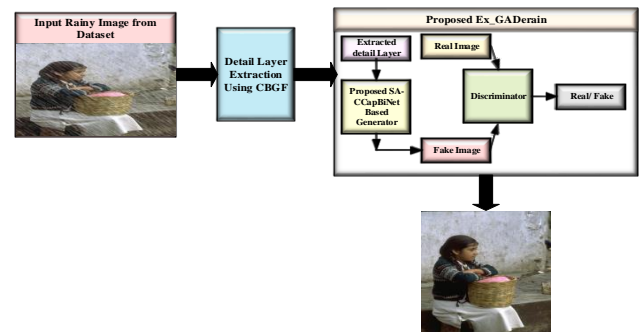


Fig. 1. Block diagram for Ex_GADerain-based rain streaks removal.

C. Image Acquisition

The rain streaks removal from the rained image using the proposed Ex_GADerain acquires the input rain-streaked image from the rainy image dataset [26]. Let A be the dataset with a total of t images, from which f^{th} image is taken for processing the proposed Ex_GADerain. It is expressed as,

$$A = \{A_1, A_2, \dots, A_f, \dots, A_t\} \quad (1)$$

Where, the f^{th} image in the dataset is indicated as A_f .

D. Detail Layer Extraction

The detail layer extraction is devised using cross guided bilateral filter (CGBF). In this, a guided image with weight and robust properties is utilized along with the bilateral filter. The CGBF comprises two different filters for performing filtering and kernel identification. In the traditional cross-bilateral filter (CBF), nearby pixels' geometric closeness and gray-level similarities are considered in the image to perform the filtering operation. In contrast, the guided bilateral filter

(GBF) uses a guided image indication based on similar pixels in the neighbourhood concerning an image. The hybridization of characteristic behaviour of both the GBF and CBF constitutes the CGBF [31]. The CGBF-based detail layer extraction is depicted in Fig. 2.

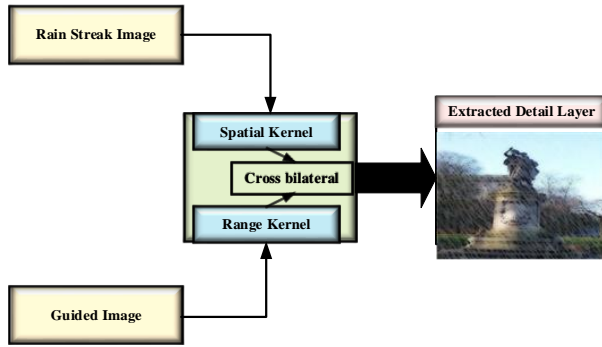


Fig. 2. CGBF-based detail layer extraction.

The outcome of a CGBF-based detail layer extraction is formulated as,

$$d_{DLE}(M,N) = P_{DLE} = \frac{1}{W_{M,N} \sum_{M,N \in W_{M,N}} S \cdot Y_{\sigma k}(|M-N|) Y_{\sigma g}(|P(M)-P(N)|) P(N)} I_{M,N} V_{M,N} + J_{M,N} \quad (2)$$

The outcome of a CGBF filtering is the detail layer extraction, which is represented as $d_{DLE}(M,N)$. M and N Refers to the adjacent pixels in the location $P(M)$ and $P(N)$, and the guided image is indicated as S . The range and spatial kernel functions concerning the CGBF are formulated as:

$$Y_{\sigma k}(|M-N|) = e^{-\frac{|M-N|^2}{2\sigma k^2}} \quad (3)$$

$$Y_{\sigma g}(|P(M)-P(N)|) = e^{-\frac{|P(M)-P(N)|^2}{2\sigma g^2}} \quad (4)$$

Where, the spatial kernel function is notated as $Y_{\sigma k}$, and the range function is notated as $Y_{\sigma g}$. The constraints are notated as σg and σk respectively. The smoothing of an image is employed for the reduction of error based on coefficients of the guided filter $I_{M,N}$ and $J_{M,N}$ and is formulated as,

$$I_{M,N} = \frac{\sum_{i \in W_{M,N}} V^i U^i - V^{M,N} U^i}{\sigma_{M,N}^2 + \gamma} \quad (5)$$

$$J_{M,N} = U^i - I_{M,N} V^{M,N} \quad (6)$$

Where, guided and original image's mean indicated as V^i and U^i respectively. The window function is indicated as W , and the regularization constant is indicated as γ . The variance

is notated as $\sigma_{M,N}^2$. The extracted detail layer is fed into the Ex_GADerain module for de-raining the image and classification.

E. Ex_GADerain-based Rain Streaks Removal

The rain streaks image filtered by the cross-guided bilateral filter is fed into the proposed Extended Generative Adversarial based Deraining (Ex_GADerain) for obtaining the de-rained image. The proposed Ex_GADerain is the improved version of traditional GAN. The GAN comprises networks, adversarial and generative components, each with specific functions.

a) *Networks*: The networks are utilized for training; hence, deep neural network (DNN) is utilized in GAN for learning.

b) *Adversarial*: The learning of a network is devised by the adversarial settings of a GAN.

c) *Generative*: The generation of data is performed by the generative model through the probabilistic approach.

GAN's work utilizes various network learning approaches termed discriminator and generator to make better decisions. Initially, GAN's generator produces a fake outcome based on the given input to confuse the discriminator. Hence, the role of a discriminator is to identify the original sample from the combined original and fake samples. Generating a fake outcome and identifying an original sample by the discriminator continues repetitively until optimal learning [32]. The proposed Ex_GADerain is depicted in Fig. 3.

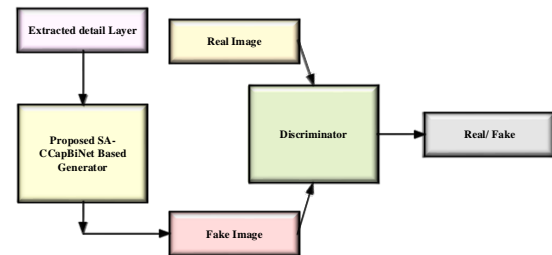


Fig. 3. Proposed Ex_GADerain.

The reason behind the consideration of GAN for acquiring the de-rained image is its high-quality image outcome for performing the image processing tasks. Here, the generator is utilized for generating the de-rained image. Hence, the rain streaks removal process is further enhanced by incorporating Self-Attention based Convolutional Capsule Bidirectional Network (SA-CCapBiNet) model on the generator side. The proposed SA-CCapBiNet is designed by hybridizing the Convolutional Capsule Network, Bidirectional Long Short Term Memory (BiLSTM), and the self-attention module. The illustration of a proposed SA-CCapBiNet for a generator is portrayed in Fig. 4.

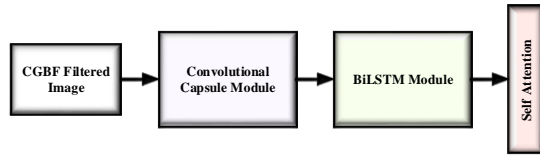


Fig. 4. Architecture of proposed SA-CCapBiNet for generator.

1) *Proposed SA-CCapBiNet-based generator module*: The traditional GAN utilizes DNN in the generator module to generate real and fake images. In the proposed Ex_GADerain technique of de-raining the rained image, SA-CCapBiNet is utilized in the generator for generation images. The newly devised SA-CCapBiNet is the hybridized structure of Convolutional Capsule Network (CCap), Bidirectional Long short term memory (BiLSTM), and self-attention mechanism. A detailed description is given below.

2) *Convolutional capsule network (CCap)*: The convolutional neural network (CNN) is a deep learning architecture commonly utilized for image processing tasks due to its promising solution. The two unique features utilized by CNN have shared weights and local connections. In addition, the parameter reduction is employed by CNN through the replicated weighting criteria. The outcome of the CNN is expressed as,

$$y_q^h = a \left(\sum_{p=1}^{F_{map}} y_p^{h-1} * r_{pq}^h + e_q^h \right) \quad (7)$$

$$a(y) = \max(0, y) \quad (8)$$

Where, the outcome of a past iteration $h-1$ concerning the p^{th} feature is indicated as y_p^{h-1} , the total features are notated as F_{map} , the q^{th} feature concerning the current iteration h is expressed as y_q^h , and the activation is notated as $a(\cdot)$. The convolution operation is indicated as $*$.

a) *Activation*: Rectified Linear Unit (ReLu) is utilized to transform the features nonlinearly to learn the complex features more accurately.

CNN's pooling operation is utilized for feature reduction, which causes information to be lost. Also, the CNN limits performance by changing scale, translational and rotational variance, but superior performance is acquired with the CapsuleNet. Thus, the combined Convolutional Capsule Network (CCap) is utilized in the proposed SA-CCapBiNet method.

b) *CapsuleNet*: The input to CapsuleNet is CNN's feature vector output. The CapsuleNet output is estimated using the Squashing function and is formulated as follows,

$$w_q = \frac{\|u_q\|^2}{\alpha + \|u_q\|^2} \frac{u_q}{\|u_q\|} \quad (9)$$

where, u_q refers to the input, the activation function is indicated as squashing function and is represented as w_q , in which the long vector is shrunken into the required length based on α and the short vector is shrunken to the length zero. The CapsuleNet acquires the input as,

$$u_q = \sum_p s_{pq} Wei_{pq} v_p \quad (10)$$

Where, the weight factor is notated as Wei_{pq} , and the outcome of a capsule is indicated as v_p .

Then, an expression for the coupling coefficient s_{pq} is formulated as,

$$s_{pq} = \frac{\exp(x_{pq})}{\sum_n x_{pn}} \quad (11)$$

Where, the probabilities of two coupled capsules are notated as x_{pq} and x_{pn} respectively.

The coupling coefficient is updated iteratively based on dynamic routing. The outcome of a subsequent layer is shared with the parent layer for each capsule's output, which is called dynamic routing. Here, the weight updation based on dynamic routing considers the adjacent capsule for enhancing the similarity between two coefficients. Thus, the combined CCap network replaces the single neuron with the neuron vector to enhance rain streaks removal efficiency. The features mapped using the CCap are fed into the BiLSTM to generate fake images to confuse the discriminator. The architecture of a CCap Network is depicted in Fig. 5.

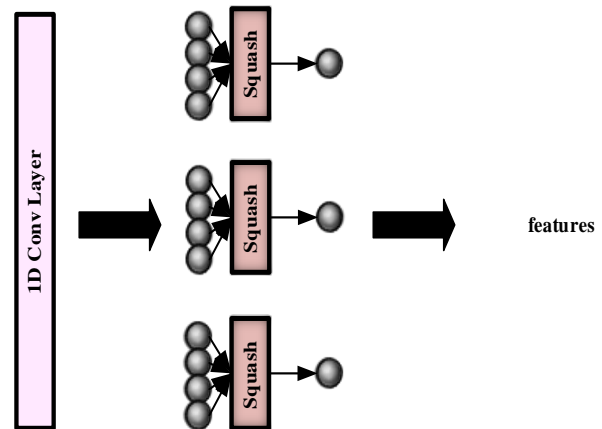


Fig. 5. Architecture of convolutional CapsuleNet (CCap).

3) *BiLSTM*: The outcome of CCap is taken as input to the BiLSTM module for capturing the long-term dependency information. The BiLSTM comprises several LSTM cells, wherein a gating mechanism is utilized to capture the essential

attribute mapping. The architecture of BiLSTM is depicted in Fig. 6(a), wherein the LSTM cell is portrayed in Fig. 6(b).

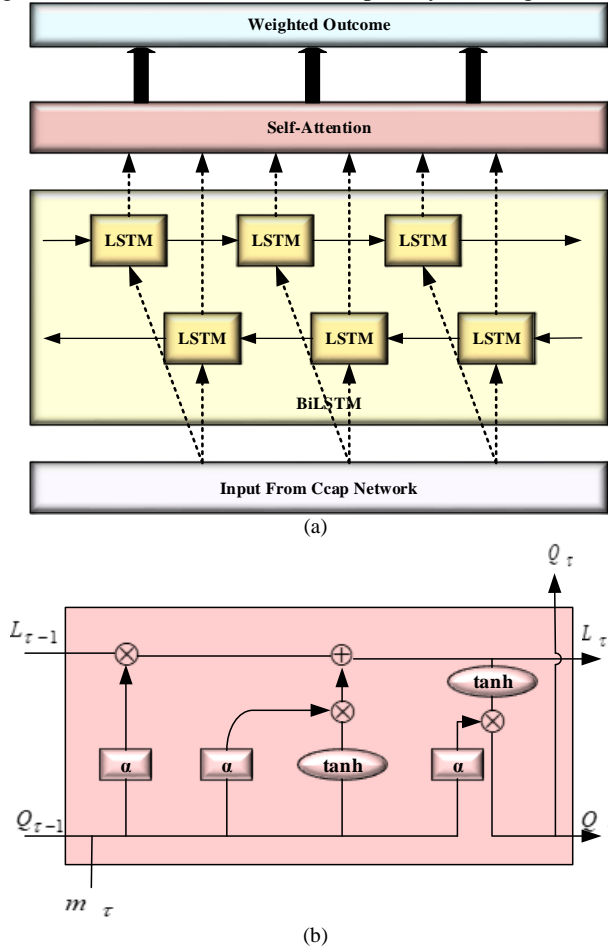


Fig. 6. Architecture of (a) BiLSTM and (b) LSTM cell.

a) *LSTM*: A recurrent neural network (RNN) with the memory unit constitutes the LSTM and is designed to overcome the issues concerning the vanishing gradient that limits the generalization capability. The gates and cell functioning are the major concept utilized in LSTM for making decisions. Cell state is utilized to maintain the information for further processing, which helps to minimize the vanishing gradient issue.

b) *Cell State*: The memory cell of LSTM constitutes the cell state, wherein the decision concerning the information removal or insertion is devised using the gating mechanism. The functioning is performed through sigmoid and point-wise multiplication for information maintenance. The cell state result is either '1' or '0', in which the information removal is devised for zero outcomes and the information maintenance is devised for the outcome '1'. Here, \tilde{L}_τ refers to the cell state and its new state is notated as L_τ . The formulation for the cell state based on the hidden state and input is expressed as,

$$\tilde{L}_\tau = \tanh(Z_L[Q_{\tau-1}, m_\tau] + X_L) \quad (12)$$

$$L_\tau = G_\tau * Q_{\tau-1} + R_\tau * \tilde{L}_\tau \quad (13)$$

Where, X_L refers to the bias and Z_L refers to the weight, hidden state is notated as $Q_{\tau-1}$, the input message is indicated as m_τ , forget gate's outcome is indicated as G_τ , and the outcome of the input gate is indicated as R_τ .

c) *Forget Gate*: The less informative features are removed by the forget gate, which is evaluated as,

$$G_\tau = \alpha(Z_G[Q_{\tau-1}, m_\tau] + X_G) \quad (14)$$

Where, the sigmoid activation is indicated as α the bias and weight concerning the forget gate are notated as X_G and Z_G respectively. Here, the outcome of a sigmoid function decides whether the information is remembered or forgotten.

d) *Input Gate*: The role of an input gate is to regulate the information that needs to be included in the cell state by using the sigmoid function. Feature filtering is employed for information maintenance by considering the input and hidden state. The sigmoid gate devises the filtering of information, and the expression for the input gate outcome is,

$$R_\tau = \alpha(Z_R[m_\tau, Q_{\tau-1}] + X_R) \quad (15)$$

Where, the outcome of an input gate is indicated as, and the bias and weight concerning the input gate are notated as X_R and Z_R , respectively.

e) *Output Gate*: The feature vector generation using the tanh function and filtering using the sigmoid function are devised in the output gate for better decision-making. The outcome of an output gate is,

$$G_\tau = \alpha(Z_G[Q_{\tau-1}, m_\tau] + X_G) \quad (16)$$

$$Q_\tau = D_\tau * \tanh(L_\tau) \quad (17)$$

Where, the outcome of a gate is indicated as G_τ , and the bias and weight concerning the input gate are notated as X_G and Z_G respectively.

f) *BiLSTM*: The information processing is devised in only one direction by the existing LSTM; hence, considering the past information is impossible. Thus, considering previous information, backward processing is essential, accomplished through backward information processing. Let \vec{P}_τ be the

forward direction-based processing and \overleftarrow{P}_τ be the backward direction processing of the BiLSTM. The architecture with both directional behaviours is depicted in Fig. 5. The outcome of BiLSTM is formulated as,

$$P_\tau = \vec{P}_\tau \oplus \overleftarrow{P}_\tau \quad (18)$$

Where, the outcome of BiLSTM is notated as P_τ , and the summation operation is notated as \oplus . The outcome of BiLSTM is more efficient due to considering both the subsequent and preceding information.

g) *Self-attention*: The assignment of weights among the extracted features is devised in the self-attention module. Here, the weights are employed for the features extracted by the BiLSTM, wherein the highest weight is assigned for the most significant attributes. The correlation between the currently hidden and high-dimensional vectors is utilized for weighting the attributes. The hidden vector formulation is expressed as,

$$\mu_{\tau} = \tanh(Z_z * Q_{\tau} + X_z) \quad (19)$$

Where, the hidden vector is notated as μ_{τ} , the hidden state is indicated as Q_{τ} , the bias is indicated as X_z and the weight is notated as Z_z . Then, the outcome of an attention module is expressed as,

$$Attn = \sum_{\tau} \eta_{\tau} * Q_{\tau} \quad (20)$$

Where, the variable η_{τ} is calculated as,

$$\eta_{\tau} = \frac{\exp(\mu_{\tau} * \mu_z)}{\sum_{\tau} \exp(\mu_{\tau} * \mu_z)} \quad (21)$$

Here, the high dimensional feature is indicated as μ_z . An outcome of the self-attention module is the de-rained image.

4) *Generator-discriminator operation*: Here, the proposed SA-CCapBiNet generator acquires the input data $p(m)$ and the noisy variable $p(c)$, in which the input is the rain streaks image. While performing the rain streaks removal process, the attributes like texture information and higher order colour are maintained during the image translation.

a) *Loss Function*: The loss of a proposed Ex_GADerain is minimized through the min-max optimization issue and is expressed as,

$$MIN^G MAX^D = K_{m \sim p(m)} |\log D(m)| + K_{c \sim p(c)} |\log(1 - G(c))| \quad (22)$$

Where, the discriminator is indicated as D , the generator is indicated as G , the expectation operator is indicated as K , the input rain streak image is indicated as $p(m)$ and the noisy image is indicated as $p(c)$. The role of a discriminator is to correctly identify the fake image generated by the generator and try to maximize $\log D(m)$. In contrast to the generator, SA-CCapBiNet tries to minimize the $\log(1 - G(c))$. Thus, the min-max optimization is devised in GAN for enhancing the image de-raining process.

In addition to the adversarial loss function, the mean absolute error is included for generating the blurred image. Thus, the loss function of the proposed Ex_GADerain is expressed as,

$$L_{MAE} = K_{m,b,c} \left[\|m - G(c,b)\| \right] \quad (23)$$

Where, the mean absolute error is notated as L_{MAE} , the generated image is indicated as $G(c,b)$ in the region (m,b,c) . Then, the total loss of a proposed Ex_GADerain is formulated as,

$$L_{Total} = L_{Ex_GADerain}(G.D) + \beta L_{MAE}(G) \quad (24)$$

Where, the parameter utilized for controlling the weights is indicated as β .

Thus, learning based on the loss function enhances the accuracy of a de-raining process and provides an efficient outcome.

F. Classification of Input Image

The outcome of the self-attention module is utilized for decision-making regarding whether the concerning input image is rained image or not. For this, this softmax activation is utilized and is formulated as,

$$C = \text{soft max}(Z_c s + X_c) \quad (25)$$

Where, the classification outcome is indicated as C , the input vector is indicated as s , the bias is indicated as X_c and the weight is notated as Z_c .

Hence, the outcome of a proposed Ex_GADerain is the classification of an input image as rained or not.

IV. RESULTS AND DISCUSSION

The proposed Ex_GADerain technique is analyzed by implementing the proposed methodology in PYTHON programming language using an 8GB RAM PC with Windows 10OS. The experimental outcome and various assessments are devised to portray the excellence of the Ex_GADerain method.

a) *Dataset Used*: The rainy image dataset [30] analyses the proposed method's performance. The dataset has 1000 normal images with 14 rainy images with different magnitudes and orientations of rain streaks. The rainy image was generated using photoshop.

A. Experimental Outcome

The experimental outcome of a proposed Ex_GADerain is illustrated in Fig. 7, wherein the rain-streaked image is portrayed in Fig. 7(a), the detail layer extracted image is portrayed in Fig. 7(b) and the de-rained image is portrayed in Fig. 7(c).





Fig. 7. Experimental outcome of Ex_GADerain: (a) Input rainy Image, (b) Detail layer extraction and (c) De-rained image

Fig. 7 shows that removing the rain streaks using the proposed Ex_GADerain technique obtained a visually better outcome. Still, the quality of a de-rained image is analyzed through image quality assessment measures.

B. Image Quality Assessment

The outcome of a proposed Ex_GADerain-based de-rained image is assessed based on various image quality assessment measures like structural similarity index measure (SSIM), Dissimilarity index measure (DSSIM), peak signal to noise ratio (PSNR), and root mean square error (RMSE).

1) *Structural similarity*: The similarity between the de-rained images acquired by the proposed Ex_GADerain and the original image is measured through the SSIM measure. By setting various windows, the quality of an image is measured. Let's consider a window size $B \times B$ for both the de-rained and original images. Then, the expression for the SSIM is formulated as,

$$SSIM_{t,o} = \frac{(2M_t M_o + s_1)(2\sigma_{to} + s_2)}{(M_t^2 + M_o^2 + s_1)(\sigma_t^2 + \sigma_o^2 + s_2)} \quad (26)$$

Where, the mean concerning the de-rained image is indicated as M_t , the mean concerning the original image is indicated as M_o , and the covariance is notated as σ_{to} . The variance for the de-rained and original image is notated as σ_t and σ_o respectively. The stabilization factors are notated as s_1 and s_2 , respectively. The analysis based on the similarity index of a proposed Ex_GADerain method is depicted in Fig. 8. The similarity between the derained outcomes of the proposed Ex_GADerain with an original image with 50% learning and 40 epoch is 0.7015, which is 0.9697 with 90% of learning data. The similarity value with 90% learning data is

closer to the maximal value 1, which depicts the excellence of a proposed method in removing the rain streaks in a rainy image. The detailed analysis of Ex_GADerain based on SSIM is depicted in Table I.

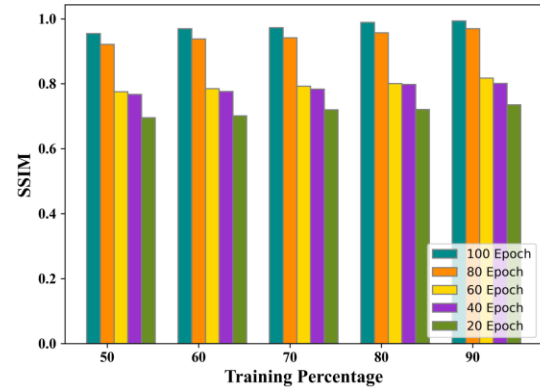


Fig. 8. Analysis of Ex_GADerain based on SSIM.

TABLE I. ANALYSIS OF EX_GADERAIN BASED ON SSIM

Epoch	Training Data				
	50	60	70	80	90
20	0.6955	0.7673	0.7756	0.9212	0.9545
40	0.7015	0.7765	0.7849	0.9378	0.9697
60	0.7197	0.7837	0.7926	0.9415	0.9726
80	0.7209	0.798	0.8005	0.9569	0.9888
100	0.7353	0.8009	0.8175	0.9695	0.9935

2) *Dissimilarity measure*: The dissimilarity between the de-rained image and an original image is measured through DSSIM. The formulation for finding the DSSIM is expressed as,

$$DSSIM_{t,o} = \frac{1 - SSIM_{t,o}}{2} \quad (27)$$

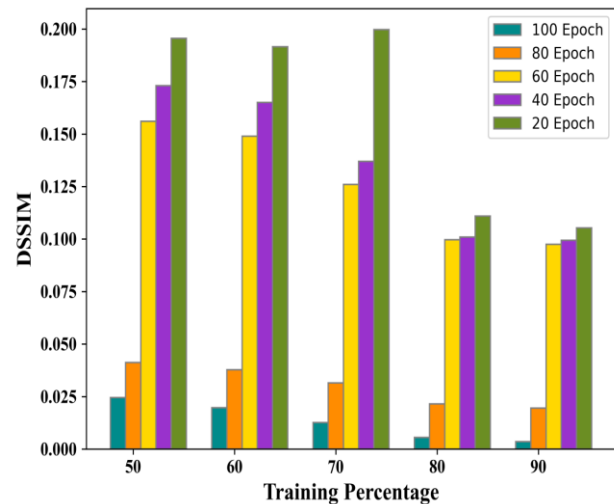


Fig. 9. Analysis of Ex_GADerain based on DSSIM.

TABLE II. ANALYSIS OF EX_GADERAIN BASED ON DSSIM

Epoch	Training Data				
	50	60	70	80	90
20	0.1955	0.173	0.156	0.0412	0.0245
40	0.1915	0.165	0.149	0.0378	0.0197
60	0.1997	0.137	0.126	0.0315	0.0126
80	0.1109	0.1009	0.0997	0.055	0.0215
100	0.1053	0.0994	0.0975	0.0195	0.0035

Where, the dissimilarity is notated as $DSSIM_{t,o}$. The analysis of Ex_GADerain based on DSSIM is portrayed in Fig. 9. With 60% of data learning and 20 epochs, the DSSIM evaluated by the Ex_GADerain is 0.173, which is 0.0994 for 100 epochs with the same training data. Thus, the epoch value enhancement improves the de-raining process by reducing the dissimilarity between the images. Likewise, for 70% of learning data and 80 epochs evaluated, the DSSIM of 0.0997 is 0.0215 with 90% of learning data. Thus, the increased amount of learning data minimizes the image's dissimilarity, indicating the more accurate removal of rain streaks. Thus, the minimal value of dissimilarity is acquired with higher learning data and epoch values. Also, the minimal DSSIM between the de-rained and original image depicts the better outcome of the proposed Ex_GADerain method. The detailed analysis of Ex_GADerain based on DSSIM is portrayed in Table II.

3) *Peak signal-to-noise ratio*: The reconstructed image quality from the rainy image based on the noise level is measured through the PSNR. The ratio between the original image and the noise evaluated based on the error defines the PSNR. Here, the error measure of the de-rained image is evaluated through the mean square error (MSE). Then, the formulation of PSNR is expressed as,

$$PSNR = 10 \log_{10} \frac{Value_{peak}}{MSE} \quad (28)$$

Where, the peak value of a de-rained image is notated as $Value_{peak}$. The MSE is formulated as,

$$MSE = \frac{\sum_{R,C} (I_t(R,C) - I_o(R,C))}{R * C} \quad (29)$$

Where, the number of rows is indicated as R , the number of columns is indicated as C , and the intensity of an original and derained image is notated as $I_o(R,C)$ and $I_t(R,C)$, respectively. The analysis based on PSNR is depicted in Fig. 10. The maximal PSNR acquired by the Ex_GADerain is 26.4052 with epoch 80 and 80% of learning data, which is 26.5457 with 90% of learning data with epoch 80. Here, the PSNR value also elevates with the epoch and learning data percentage enhancement. The larger value of PSNR measured in decibels depicts a better outcome, and the detailed analysis is portrayed in Table III.

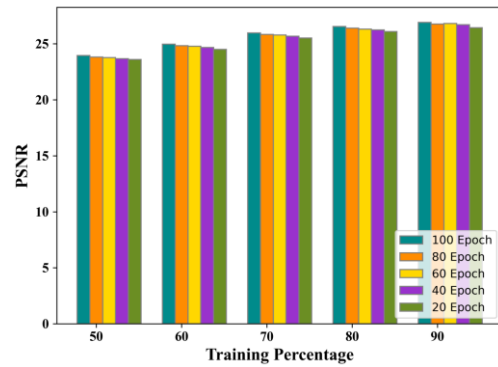


Fig. 10. Analysis of Ex_GADerain based on PSNR.

TABLE III. ANALYSIS OF EX_GADERAIN BASED ON PSNR

Epoch	Training Data				
	50	60	70	80	90
20	23.5955	23.6673	23.7756	23.8212	23.9545
40	24.5015	24.6765	24.7849	24.8378	24.9697
60	25.5197	25.6837	25.7926	25.8415	25.9726
80	26.0962	26.2372	26.3155	26.4052	26.5457
100	26.4353	26.7009	26.8175	26.7695	26.9135

4) *Root mean square error*: RMSE is an error measure that evaluates the error in the de-rained outcome based on the MSE. The formulation for the RMSE is defined as,

$$RMSE = \sqrt{MSE} \quad (30)$$

The analysis based on RMSE is depicted in Fig. 11. The RMSE evaluated by the Ex_GADerain method is 2.3955 with 20 epochs and 50% learning data, 2.1573 with learning data of 60%, 1.7312 with learning data of 80%, and 0.9445 with learning data of 90%. The analysis shows that the error gets minimized with more information learning. The higher amount of data the classifier learns enhances the generalization capability and minimizes the error in removing the rain streaks. The difference between the original and derained image based on the error magnitude accomplished minimal error that indicates the enhanced quality of the rain streaks removal process. The detailed analysis is depicted in Table IV.

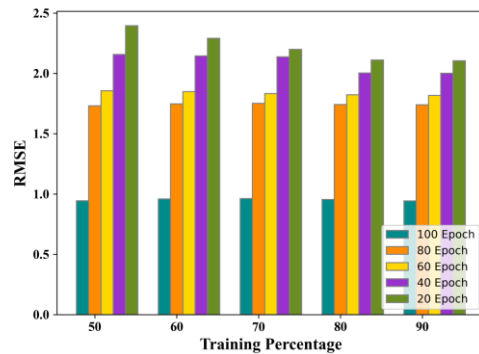


Fig. 11. Analysis of Ex_GADerain based on RMSE.

TABLE IV. ANALYSIS OF EX_GADERAIN BASED ON RMSE

Epoch	Training Data				
	50	60	70	80	90
20	2.3955	2.1573	1.856	1.7312	0.9445
40	2.2915	2.145	1.849	1.7478	0.9597
60	2.1997	2.137	1.8326	1.7515	0.9626
80	2.1109	2.0031	1.8223	1.7421	0.9545
100	2.1053	2.0009	1.8175	1.7395	0.9435

C. Comparative Analysis

The existing image de-raining techniques like CGAN [21], KGCNN [22] and SSDRNet [23] are compared with the proposed Ex_GADerain method to depict the performance enhancement. Fig. 12 depicts the comparative analysis. The SSIM estimated by the Ex_GADerain method is 0.9562 with 60% of data learning, which is higher than the conventional methods. Conventional KGCNN acquired the SSIM of 0.9336, CGAN of 0.7796 and SSDRNet of 0.7736, respectively. The dissimilarity measures of the EX_GADerain method is 0.0126 with 70% of training data; the traditional methods like KGCNN, CGAN, and SSDRNet accomplished the higher DSSIM value of 0.0315, 0.126, and 0.147, respectively. The error estimation based on RMSE estimated by the Ex_GADerain method is 0.9545 with 80% of training data, which is minimal compared to the traditional methods like KGCNN, CGAN and SSDRNet that acquired the RMSE value of 1.7421, 1.8223, and 2.0031 respectively. The maximal PSNR acquired by the Ex_GADerain method is 25.8552, with 50% of data learning. The traditional KGCNN, CGAN and SSDRNet methods acquired the minimal PSNR of 25.6549, 24.2887, and 23.9778, respectively. Thus, the Ex_GADerain method accomplished superior performance for all quality measures.

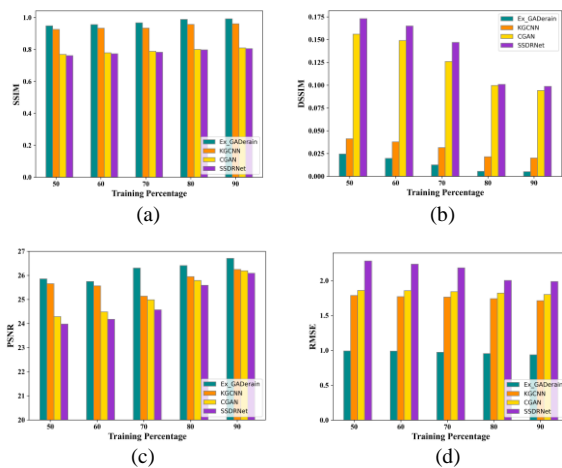


Fig. 12. Comparison in terms of (a) SSIM, (b) DSSIM, (c) PSNR and (d) RMSE.

D. Analysis based on the Classification

The detection of the rainy or non-rainy input image is devised by the proposed Ex_GADerain technique prior to the rain streaks removal.

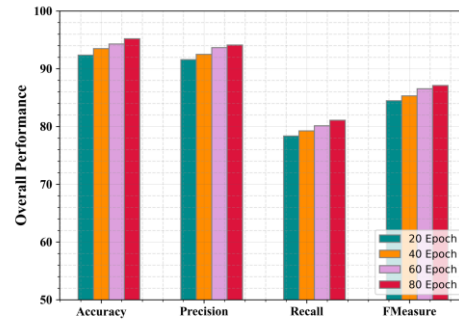


Fig. 13. Analysis based on the classification.

TABLE V. OVERALL PERFORMANCE ANALYSIS

Metrics/ Epoch	20	40	60	80
Accuracy	92.3576	93.479	94.286	95.1852
Recall	91.5763	92.482	93.655	94.1032
Precision	78.3517	79.252	80.127	81.0779
F-measure	84.4512	85.296	86.542	87.1063

The image detected as rainy is utilized for removing the rainy streaks to acquire the de-rained image. The outcome of the classification task based on the performance measures like accuracy, precision, recall and F-Measure is portrayed in Fig. 13. The outcome based on the classification task based on overall performance analyzed in Table V.

1) Accuracy and loss: The accuracy and loss analysis of a proposed Ex_GADerain method based on the testing and training data by varying the epoch is depicted in Fig. 14. The accuracy of training is higher compared to the testing process. Likewise, the loss concerning the testing process is higher than the training process.

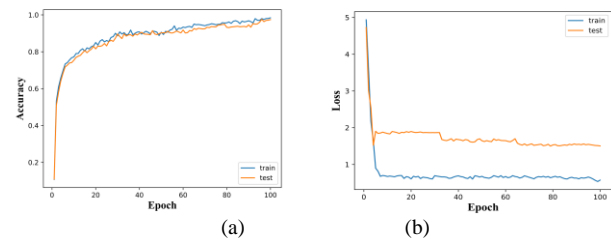


Fig. 14. Accuracy and loss analysis: (a) Accuracy and (b) Loss.

E. Complexity Analysis

The complexity analysis of a proposed Ex_GADerain with the conventional methods is depicted in Fig. 15. While training the data, the time complexity of the Ex_GADerain method is 65.34ms, which is 6.76%, 3.13%, and 1.21% improved performance compared to the traditional KGCNN, CGAN, and SSDRNet. Likewise, the time complexity of rain streaks removal methods while testing is 12.56ms, 13.91ms, 14.89ms, and 15.78ms for Ex_GADerain, KGCNN, CGAN, and SSDRNet methods. Here, the proposed Ex_GADerain acquired minimal time complexity compared to the traditional testing and training methods.

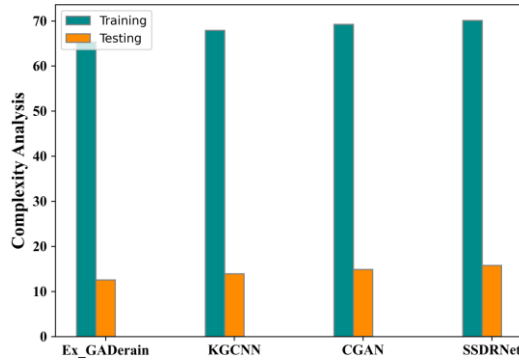


Fig. 15. Complexity analysis.

F. Comparative Discussion

The best outcome of Ex_GADerain based on various assessment measures along with the traditional KGCNN, CGAN and SSDRNet is depicted in Table VI. The maximal similarity of the de-rained image with the original image evaluated by the Ex_GADerain is 0.9923, which is 3.13%, 18.35%, and 18.84% enhanced performance compared to conventional KGCNN, CGAN and SSDRNet methods. The maximal PSNR evaluated by the Ex_GADerain is 26.752, 1.72%, 1.93%, and 2.31% enhanced performance compared to conventional KGCNN, CGAN and SSDRNet methods. The minimal RMSE acquired by Ex_GADerain is 0.9367, which is 45.30%, 48.03%, and 52.87% enhanced performance compared to conventional KGCNN, CGAN and SSDRNet methods. The minimal DSSIM acquired by Ex_GADerain is 0.0051, which is 74.63%, 94.59%, and 94.83% enhanced performance compared to conventional KGCNN, CGAN and SSDRNet methods. Here, Ex_GADerain accomplished excellent performance compared to the conventional rain streaks removal methods.

TABLE VI. COMPARATIVE DISCUSSION

Metrics/ Methods	KGCNN	CGAN	SSDRNet	Ex_GADerain
SSIM	0.9612	0.8102	0.8054	0.9923
PSNR	26.2464	26.1891	26.087	26.7052
RMSE	1.7125	1.8023	1.9875	0.9367
DSSIM	0.0201	0.0942	0.0987	0.0051

The analysis depicts the enhanced performance of the Ex_GADerain method based on various assessment measures. Rain streak removal using the Ex_GADerain utilizes the detail layer extraction for smoothing the rainy image. It also reduces the error through the guided filtering technique. The SA-CCapBiNet-based generator in generating fake images by removing the rain streaks. Here, the proposed SA-CCapBiNet reconstructs the de-rained image through the enhanced capability of feature representation. Also, the consideration of long term dependencies among the information enhances the rain streak removal capability of the model by solving the vanishing gradient issues. Besides, the texture and colour

information maintenance during the rain streaks removal process enhances the quality of a derained image. Considering mean absolute error with the adversarial error for evaluating the loss function minimizes the information learning and makes the generalization more efficient. Thus, the quality assessment of the derained image and the classification tasks acquired better performance.

V. CONCLUSION

Rain streaks removal using the proposed Ex_GADerain from the rainy image accurately reconstructed the derained image. The proposed SA-CCapBiNet-based generator enhances the deraining process by considering the long-term dependencies and the generated features' past information. Also, the weight updation based on the dynamic routing of a convolutional capsule network enhances the efficiency of rain streak removal through the neuron vector instead of a single neuron. The self-attention criteria weight the more appropriate features with higher weights, making the rain streak removal more efficient. The quality assessment of the derained image based on various assessment measures like SSIM, PSNR, RMSE, and DSSIM improved performance compared to the conventional rain streak removal methods. The maximal SSIM and PSNR acquired by the Ex_GADerain are 0.9923 and 26.7052, respectively. The minimal RMSE and DSSIM acquired by the Ex_GADerain are 0.9367 and 0.0051, respectively. However, the error evaluated by the proposed method is higher. Hence, in the future, a novel rain streak removal method with optimized deep learning will be designed for handling bad rainy conditions

REFERENCES

- [1] H. Wang, Y. Wu, Q. Xie, Q. Zhao, Y. Liang, S. Zhang, D. Meng, "Single image rain streaks removal: a review and an exploration," International Journal of Machine Learning and Cybernetics, vol. 11, pp. 853-872, 2020.
- [2] Y. Ding, M. Li, T. Yan, F. Zhang, Y. Liu, R. W. Lau, "Rain streak removal from light field images." IEEE Transactions on Circuits and Systems for Video Technology, vol. 32, no. 2, pp.467-482, 2021.
- [3] Y. Shen, Y. Wang, M. Wei, H. Chen, H. Xie, G. Cheng, and F.L. Wang, Semi-MoreGAN: A New Semi-supervised Generative Adversarial Network for Mixture of Rain Removal 2022. arXiv preprint arXiv:2204.13420.
- [4] L. Zhu, Z. Deng, X. Hu, H. Xie, X. Xu, J. Qin, P. A. Heng, "Learning gated nonlocal residual for single-image rain streak removal," IEEE Transactions on Circuits and Systems for Video Technology vol. 31, no. 6, pp.2147-2159, 2020.
- [5] K. Zhang, D. Li, W. Luo, W. Ren, B. Stenger, W. Liu, M. H. Yang, "Dual attention-in-attention model for joint rain streak and raindrop removal," IEEE Transactions on Image Processing vol. 30, pp.7608-7619, 2021.
- [6] T. Yan, M. Li, B. Li, Y. Yang, and R.W.H. Lau, "Rain Removal from Light Field Images with 4D Convolution and Multi-scale Gaussian Process." 2022 arXiv preprint arXiv:2208.07735.
- [7] H. Fazlali, S. Shirani, M. Bradford, T. Kirubarajan, "Single image rain/snow removal using distortion type information," Multimedia Tools and Applications vol. 81, no. 10, pp. 14105-14131, 2022.
- [8] N. Ahn, S. Y. Jo, S. J. Kang, "Eagnet: Elementwise attentive gating network-based single image de-raining with rain simplification," IEEE Transactions on Circuits and Systems for Video Technology vol. 32, no. 2, pp. 608-620, 2021.
- [9] Y. Wei, Z. Zhang, Y. Wang, M. Xu, Y. Yang, S. Yan, M. Wang, "Deraincyclegan: Rain attentive cyclegan for single image deraining and

- rainmaking," IEEE Transactions on Image Processing vol. 30, pp. 4788-4801, 2021.
- [10] D. T. Vu, J. L. Gonzalez, M. Kim, "Exploiting global and local attentions for heavy rain removal on single images," arXiv preprint arXiv: 2104.08126 2021.
- [11] X. Fu, J. Huang, X. Ding, Y. Liao, J. Paisley, "clearing the skies: A deep network architecture for single-image rain removal." IEEE Transactions on Image Processing vol. 26, no. 6, pp. 2944-2956, 2017.
- [12] S. Du, Y. Liu, M. Zhao, Z. Shi, Z. You, and J. Li. "A comprehensive survey: Image deraining and stereo-matching task-driven performance analysis." IET Image Processing vol. 16, no. 1, pp. 11-28, 2022.
- [13] W. Yang, R. T. Tan, S. Wang, Y. Fang, J. Liu, "Single image deraining: From model-based to data-driven and beyond." IEEE Transactions on pattern analysis and machine intelligence vol. 43, no. 11, pp. 4059-4077, 2020.
- [14] X. Hu, L. Zhu, T. Wang, C.W. Fu, P.A. Heng, "Single-image real-time rain removal based on depth-guided nonlocal features," IEEE Transactions on Image Processing vol. 30, pp.1759-1770, 2021.
- [15] P. Li, J. Tian, Y. Tang, G. Wang, C. Wu, Model-based deep network for single image deraining, IEEE Access vol. 8, pp. 14036-14047, 2020.
- [16] K. H. Lee, E. Rvu, J. O. Kim. "Progressive rain removal via a recurrent convolutional network for real rain videos," IEEE Access vol. 8, pp. 203134-203145, 2020.
- [17] X. Fu, Q. Qi, Z. Zha, X. Ding, F. Wu, J. Paisley, "Successive graph convolutional network for image de-raining," International Journal of Computer Vision vol. 129, pp. 1691-1711, 2021.
- [18] L. Cai, Y. Fu, W. Huo, Y. Xiang, T. Zhu, Y. Zhang, H. Zeng, and D. Zeng, "Multi-scale Attentive Image De-raining Networks via Neural Architecture Search." IEEE Transactions on Circuits and Systems for Video Technology 2022.
- [19] Y. Yu, W. Yang, Y. P. Tan, A.C. Kot, "towards robust rain removal against adversarial attacks: A comprehensive benchmark analysis and beyond," In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition pp. 6013-6022, 2022.
- [20] Frants, V., Agaian, S., & Panetta, K, "QSAM-Net: Rain streak removal by quaternion neural network with self-attention module," 2022 arXiv preprint arXiv: 2208.04346.
- [21] Y. T. Wang, X. L. Zhao, T. X. Jiang, L. J. Deng, Y. Chang, T. Z. Huang, "Rain streaks removal for single image via kernel-guided convolutional neural network." IEEE Transactions on Neural Networks and Learning Systems vol. 32, no. 8, pp. 3664-3676, 2020.
- [22] C.Y. Lin, Z. Tao, A. S. Xu, L. W. Kang, F. Akhyar, "Sequential dual attention network for rain streak removal in a single image," IEEE Transactions on Image Processing vol. 29, 9250-9265, 2020.
- [23] P. Hettiarachchi, R. Nawaratne, D. Alahakoon, D. De Silva, N. Chilamkurti, "Rain streak removal for single images using conditional generative adversarial networks," Applied Sciences vol. 11, no. 5, pp. 2214, 2021.
- [24] I.J. Jacob, P. E. Darney, "Rain Streaks Removal in digital images by Dictionary based sparsity process with MCA Estimation." Journal of Innovative Image Processing vol. 3, no. 3, pp. 174-189, 2021.
- [25] Y. Chang, M. Chen, C. Yu, Y. Li, L. Chen, L. Yan, "Direction and Residual Awareness Curriculum Learning Network for Rain Streaks Removal," IEEE Transactions on Neural Networks and Learning Systems 2023.
- [26] W. Yang, S. Wang, D. Xu, X. Wang and J. Liu, Towards scale-free rain streak learning via self-supervised fractal band learning. In Proceedings of the AAAI Conference on Artificial Intelligence Vol. 34, No. 07, pp. 12629-12636, 2020, April.
- [27] X. Wang, Z. Li, H. Shan, Z. Tian, Y. Ren, and W. Zhou, "Fastderainnet: A deep learning algorithm for single image deraining." IEEE Access vol. 8, pp. 127622-127630, 2020.
- [28] W. Ran, B. Yang, P. Ma, and H. Lu, "TRNR: Task-Driven Image Rain and Noise Removal With a Few Images Based on Patch Analysis." IEEE Transactions on Image Processing vol. 32, pp. 721-736, 2023.
- [29] Y. Wang, X. Yan, Y. Niu, L. Gong, Y. Guo, and M. Wei, "Joint Depth Estimation and Mixture of Rain Removal From a Single Image." 2023 arXiv preprint arXiv:2303.17766.
- [30] Rainy Image Dataset, <https://drive.google.com/file/d/10cu6MA4fQ2Dz16zfyZrQWDDhOWhRdhpq/view>
- [31] T. Joel, R. Sivakumar, "Nonsubsampled contourlet transform with cross-guided bilateral filter for despeckling of medical ultrasound images." International Journal of Imaging Systems and Technology vol. 31, no. 2, pp. 763-777, 2021.
- [32] H. H. N. Alrashedy, A. F. Almansour, D. M. Ibrahim, M. A. A. Hammoudeh, "BrainGAN: Brain MRI Image Generation and Classification Framework Using GAN Architectures and CNN Models," Sensors vol. 22, no. 11, pp. 4297, 2022.

A PSL-based Approach to Human Activity Recognition in Smart Home Environments

Yan Li

School of Intelligent Science and Information Engineering
Xi'an Peihua University
Xi'an 710125, China

Abstract—Human activity recognition is widely used in smart cities, public safety and other fields, especially in smart home systems where it has a pivotal role. The study addresses the shortcomings of Markov logic networks for human activity recognition and proposes a human activity recognition method in smart home scenarios - an activity recognition framework based on Probabilistic Soft Logic (PSL). The framework is able to deal with logical uncertainty problems and provides expression and inference mechanisms for data uncertainty problems on this basis. The framework utilizes Deng entropy evidence theory to provide an evaluation method for sensor event uncertainty, and combines event calculus for activity modeling. Comparing the PSL method with three other common recognition methods, Ontology, Hidden Markov Model (HMM), and Markov logic network, on a public dataset, it was found that the PSL method has a much better ability to handle data uncertainty than the other three algorithms. The average recognition rates on the ADL and ADL-E sub datasets were 82.87% and 80.33%, respectively. In experiments to verify the ability of PSL to handle temporal complexity, PSL showed the least significant decrease in the average recognition rate and maintained an average recognition rate of 81.02% in the presence of concurrent and alternating activities. The human activity recognition method based on PSL has a better performance in handling both data uncertainty and temporal complexity.

Keywords—Human activity recognition; probabilistic soft logic; MAP inference; temporal complexity; data uncertainty

I. INTRODUCTION

With the rise of smart cities and smart homes and the rapid development of related technologies, human activity recognition has become a hot topic of research for many researchers [1]. In the smart home environment, human activity recognition can help the smart home system to form a human “understanding” based on the activities being performed by the residents, and then provide better and smarter living services to the residents [2-3]. With the new iteration of wireless sensing devices, the research focus of smart home systems is gradually shifting from data collection to high-level information integration and activity recognition [4]. Human activity recognition includes sequential activity recognition and composite activity recognition, while the latter is more in line with the alternating and concurrent characteristics of daily activities [5-6]. Problems related to activity recognition have been classified into 12 main research types based on three different metrics: number of recognized users, activity

complexity and perceptual patterns [7]. In order to further optimize the structure and function of smart home systems and provide a more convenient, comfortable, and safe living environment for people with inconvenience, especially the elderly, the study focuses on the identification of complex activities based on dense sensing in a single user environment of smart home systems. Firstly, a human activity recognition framework based on Probabilistic Soft Logic (PSL) is proposed, and an event preprocessing mechanism is proposed based on the characteristics of dense sensing. The reduction of irrelevant and redundant data is achieved through fragment partitioning and event merging. In addition, an event uncertainty calculation method based on DS evidence theory is proposed for data uncertainty in activity recognition, and activity modeling is carried out by combining Event Calculus (EC) and PSL. It is expected that the ability to relax first-order logical constraints through PSL and the ability to describe event persistence through EC will allow for conflicts in the knowledge base, enhance the freedom of the activity model, and further enhance its ability to handle logical uncertainty and temporal complexity problems.

Section II of the article describes the relevant work, focusing on the current research status at home and abroad, and provides a detailed explanation of the improvements and technical roadmap of the research. The first section of Section III proposes a human activity recognition framework based on PSL, which provides a detailed introduction to segment partitioning, event merging, and calculation methods for event uncertainty. Subsection B of Section III provides an activity modeling method based on PSL-EC and proves the equivalence between PSL-EC and complex activity descriptions. Section IV verifies the effectiveness of the PSL method through experiments and compares it with current mainstream activity recognition methods. Section V summarizes the research methods and results, and concludes by summarizing and organizing them.

II. RELATIVE WORK

Human activity recognition is an important research field in the Internet of Things, especially in smart homes, which focuses on understanding human behavior and further predicting human action intentions and motivations. A deep neural network model using convolutional neural networks and gated recurrent units was proposed by Dua et al. for activity time series data collected by wearable sensors, and the model

was used to automatically extract and classify human activities. More than 95% accuracy was obtained on all three datasets, demonstrating the excellent recognition and classification performance of the model [8]. Zhang et al. proposed an approach combining convolutional neural networks and activity recognition attention mechanisms for sensors and mobile devices in smart healthcare applications and systems. The approach incorporates attention into a multi-headed neural network to improve the accuracy and feature extraction of activity recognition [9]. Bianchi et al. designed an activity recognition system combining a wearable device with deep learning, where the wearable system embeds an inertial measurement unit and WiFi to send the collected data to a cloud service. The system minimizes inference resources, saves cost, and achieves 97% accuracy in the recognition of 9 activities [10]. Agarwal and Alam propose a lightweight deep learning model for human activity recognition based on the feature that edge computing can reduce communication latency and network traffic, which overcomes the disadvantages of deep learning computationally intensive. Experimental results obtained on six daily activity data from testers show that the proposed model extends the ability to handle data uncertainty in activity recognition due to most existing machine learning and deep learning techniques [11]. Artikis et al. by defining the probability of maximum intervals and the confidence rate of such intervals. A linear time algorithm is then proposed to compute the full probabilistic time intervals for a given dataset and the performance of the method is evaluated on a benchmark activity recognition dataset [12].

Due to the uncertainty and temporal complexity of human daily activities, two main activity recognition methods, data-driven and knowledge-driven, have been derived. Data-driven methods mainly include Hidden Markov Model (HMM), support vector machine, dynamic Bayesian model, etc. Tran et al. investigated multiple recognition methods in multi-user scenarios and conducted evaluation experiments on the same dataset, while exploring the effectiveness and recognition efficiency of temporal learning algorithms using sequential data and non-temporal learning using temporal manipulation features the effectiveness and recognition efficiency of algorithms [13]. Li et al. proposed a method to analyze the significance of sensor data contribution based on sensor state frequency and inverse type frequency for daily behavior recognition of a single user in a multi-tenant smart home scenario Xi'an. The method is used to measure the contribution of specific types of sensors to a certain type of behavior recognition, and then construct a spatial distance matrix based on the layout of environmental sensors to achieve context awareness and reduce data noise. Based on this, an activity recognition algorithm based on wide time domain convolutional neural network and multi-environmental sensor data for daily activity recognition is also proposed [14]. Scholars such as Ashari P have conducted in-depth analysis of data fusion and multi classifier system technologies for human activity recognition, particularly systems based on mobile and wearable devices, focusing on sensor pattern based activity monitoring and classification methods used for behavior analysis, environmental monitoring, and other activities in smart home environments. They have identified the advantages, applications, and shortcomings of deep learning

fusion methods for human activity recognition [15]. Asghari et al. proposed an online application of hierarchical HMM to detect the current activity in a real-time stream of sensor events, and also to detect activities that occur during an activity, i.e., interrupted activities. The proposed approach is validated on two different smart home datasets and the experimental results demonstrate its effectiveness and superiority [16]. The knowledge-driven recognition approach reduces the dependence on data for activity recognition and usually uses Ontology or rules for activity modeling and reasoning. Zhang et al. proposed a knowledge-based multi-intelligence collaboration approach. This layered architecture for smart homes that combines Ontology and multi-intelligence technologies aims to automatically acquire semantic knowledge and support heterogeneous and interoperable services. A generic inference algorithm based on the properties of disordered actions and activity events is proposed in this architecture for real-time inference of continuous composite activities and personalized services. Then a new idea is introduced to allow intelligences to learn knowledge of human activities autonomously and to transform them. The feasibility, effectiveness and stability of the proposal are verified through an extensive experimental evaluation at [17].

Comprehensive domestic and international related research can find that most of the commonly used recognition models are hybrid-driven approaches that mix two modeling methods, such as Markov Logic Networks (MLN). However, although MLN is used as an effective framework to address uncertainty and complexity, it adopts hard constraints on closed atoms and cannot effectively describe continuous variables of sensor data classes, resulting in low efficiency and inability to meet real-time requirements. Therefore, in view of the shortcomings of MLN method for human activity recognition, a PSL based activity recognition framework is proposed. PSL adopts Lukasiewicz logic instead of Boolean logic to transform integer linear programming problem into convex optimization problem for solution. Then, DS evidence theory was used to compensate for the lack of measurement of event uncertainty in PSL, and the activity modeling method PSL-EC was proposed in conjunction with EC, aiming to achieve efficient and accurate human activity recognition.

III. PSL-BASED HUMAN ACTIVITY RECOGNITION IN SMART HOMES

A. PSL-Based Human Activity Recognition Framework

In smart home systems, especially in most voice-activated systems, a key part of implementing an intelligent control system is the recognition of human activities [18]. Many researchers have proposed many excellent recognition frameworks for different environments, but most of these frameworks focus on the monitoring of the user's own characteristics and are weak in monitoring situational information. To address the uncertainty and complexity issues in daily activities, a PSL based activity recognition framework is proposed. From an application perspective, the framework divides the smart home control system into data collection layer, event management layer, and application layer. This section mainly studies the event management layer, which is

divided into three sub tasks: event preprocessing, activity modeling, and activity inference. Event preprocessing is further divided into three parts: fragment partitioning, event merging, and event uncertainty calculation. Firstly, this section adopts a dynamic fragment partitioning method based on information quantity to address the characteristic of the unfixed sampling rate of raw sensor data. Then merge the redundant information within the fragments and consider the triggering frequency of the event; Finally, evaluate the uncertainty of these events based on DS evidence theory. The structure of the human activity recognition framework based on PSL is shown in Fig. 1.

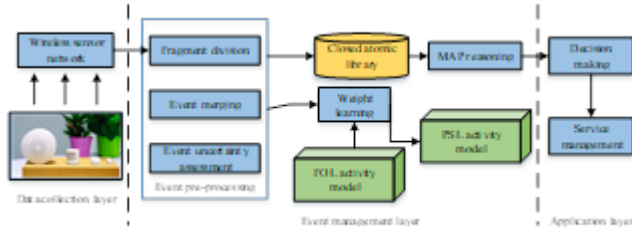


Fig. 1. Human activity recognition framework based on PSL.

In Fig. 1, the framework divides the smart home control system into a data collection layer, an event management layer, and an application layer. The data collection layer is responsible for collecting scenario information in the smart home environment and transmitting it to the event management layer through the wireless sensor network, which includes both sensors and network components deployed in the environment. The event management layer is responsible for transforming the received raw data into high-level scenario information, which can be divided into three parts by function: event pre-processing, activity modeling, and activity inference. The application layer is responsible for integrating the identified user activities and requests and is divided into two subtasks: decision making and service management. Event pre-processing is responsible for converting raw data into probabilistic events, which mainly includes three parts: fragmentation, event merging and event uncertainty calculation. Due to the sensor inexpensiveness and activity complexity in the smart home environment, usually a single behavior can be sensed by multiple sensors. Therefore, the study divides the relationship between sensors and activities into two categories: one sensor sensing multiple activities and multiple sensors sensing one activity. For the PSL model, the size of the problem increases exponentially with the number of formulas and the sensor information is susceptible to multiple factors. Therefore, the study first excludes irrelevant scenario information by segmenting the original sensor data, and then merges the sensor events within the segments to approximate the redundant information. Finally, the credibility of the observed evidence is measured by calculating the uncertainty of the events. Common segmentation methods include the interactive window method and the segmentation method based on unique attributes. The sensor trigger moments when a particular user performs a specific activity in the experimental dataset are shown in Fig. 2.

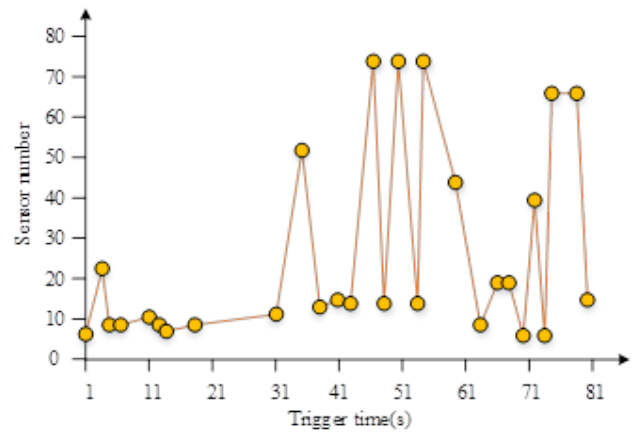


Fig. 2. Sensor trigger time chart when a user performs a specific activity.

By observing the sensor triggers when users perform different activities in the experimental dataset, it can be seen that the number of sensors triggered by different activities is stable at around 55. Therefore, the study adopts a fragmentation method based on the number of sensors based on the feature that the number of sensors triggered by different targets in the activity is more average. The method constitutes a range of values for the window length based on the average number of sensors triggered by the activities counted in the training data, and the window length is dynamically selected based on the current sensors, and the computational expression is shown in Equation (1).

$$L_s^* = \arg \max_{w_{s_l}} \{P(w_{s_l} / A_m)\} \quad (1)$$

In Equation (1), L_s^* is the optimal window length corresponding to the active A_m . The range of values is $[w_{s_l}, w_{s_l}]$, where $w_{s_l} = \min\{\bar{w}_s(A_1), \bar{w}_s(A_2), \dots, \bar{w}_s(A_m)\}$, $w_{s_l} = \text{median}\{\bar{w}_s(A_1), \bar{w}_s(A_2), \dots, \bar{w}_s(A_m)\}$, and $\bar{w}_s(A_m)$ represent the average of the number of sensors triggered by the activity A_m . The expression of the relationship between sensors and activities is shown in Equation (2).

$$A^* = \arg \max_{A_m} \{P(A_m / S_i)\} \quad (2)$$

A^* in Equation (2) represents the optimal activity of the sensor S_i . Combining Equation (1) and Equation (2) yields the probability estimation formula for the sensor and window length as shown in Equation (3).

$$w_s^* = \arg \max_{w_{s_l}} \{P(w_{s_l} / S_i)\} = \arg \max_{w_{s_l}} \{P(w_{s_l} / A_m) \times P(A_m / S_i)\} \quad (3)$$

There is still a large amount of redundancy in the segmented data, so event merging is required before modeling to avoid overly bloated recognition models. Most traditional merging methods address the case where multiple sensors are triggered at the same time and can therefore be merged into one event, generalizing the sensor data associated with one

event and ignoring the temporal impact range of sensor events. To address this problem, the study proposes an STF-EC algorithm that considers sensor trigger time and frequency in event merging. The algorithm first sets a time decay function to limit the impact range of the event, and then marks the recurring sensor events and records the number of times the sensor triggers. The impact range is calculated as shown in Equation (4).

$$R(e_i^a, e_j^a) = \begin{cases} 1, & e_i^a.SID \in Uact_i \ \& \ e_j^a.SID \in Uact_i \\ 0, & otherwise \end{cases} \quad (4)$$

In Eq. (4), $R(e_i^a, e_j^a)$ represents the sensor dependence between two atomic events e_i^a and e_j^a . $R(e_i^a, e_j^a) = 1$. Then it indicates that two sensors correspond to one event and can be combined if they are within the time constraint. The time dependence of the two atomic times is calculated as shown in Equation (5).

$$T(e_i^a, e_j^a) = \exp(-\delta |e_i^a.T - e_j^a.T|) \quad (5)$$

The final uncertainty event corresponding to that atomic time is generated and the number of occurrences of that event is recorded. The study calculates the event uncertainty based on the evidence fusion mechanism in DS evidence theory. First, we assign weights to sensor information based on Dun entropy theory, and then combine the weights and observations to calculate the uncertainty of the event. Dun entropy is a generalized version of information entropy, which is defined as shown in Equation (6).

$$DS = -\sum_{i=1}^N m(H_i) \log_2 \frac{m(H_i)}{2^{H_i} - 1} \quad (6)$$

In Equation (6), DS denotes the Dun entropy. H_i denotes the i th proposition in the identification framework, and H_i^* denotes the number of elements in the proposition. The relationship between sensor weights and Dun entropy is defined as shown in Equation (7).

$$r_i^{pst} = f(DS^*) = (1 - DS^*) \exp(DS^*) \quad (7)$$

In equation (7), DS^* is the normalized Dun entropy. In the problem of activity identification, the sensors used as evidence have different levels of reliability, so the uncertainty of the events is calculated using a weighted fusion as shown in Equation (8).

$$\hat{u} = \sum_{i=1}^n \alpha_i x_i \quad (8)$$

In Equation (8), \hat{u} is the uncertainty of the event, x_i indicates the sensor measured data, and α_i is the weight of the data in the fusion. n indicates the number of sensors corresponding to the sensor event x_i .

B. Combining PSL and EC for Human Activity Modeling and Inference Methods

Unlike general machine learning methods, PSL is essentially a first-order logical knowledge base with weights, so the process of activity modeling is similar to the construction of a knowledge base. In the application environment of smart home, the daily life has the characteristics of alternation and concurrency, and the activity model can directly determine the accuracy of activity recognition[19-20]. The way of rule definition also largely determines the scale of the problem in PSL reasoning, which in turn affects the recognition efficiency of the recognition framework. Therefore, the study proposes the activity modeling method PSL-EC, which combines PSL and EC, to solve the problem of logical uncertainty and temporal complexity in daily activities by using the logical expression uncertainty of PSL and the characteristics of EC for activity persistence modeling. The activity modeling part includes two sub-tasks: rule definition and weight learning, and the rules in PSL can be classified into soft rules and hard rules by rule type, which are used to describe the uncertainty relationship between “event” and “activity” and the domain knowledge of activity recognition, respectively. The activity modeling approach of PSL-EC uses the classical discrete event algorithm as a template to define rules based on specific problems in the activity recognition domain and to model the continuity of activities. Then existing parametric learning methods for first-order logic formulas are used to deal with logical uncertainty. In daily life, users perform activities characterized not only by temporal complexity, but also by diversity in daily life based on different execution habits. The sensor events and durations triggered by different users performing the same activity are shown in Fig. 3.

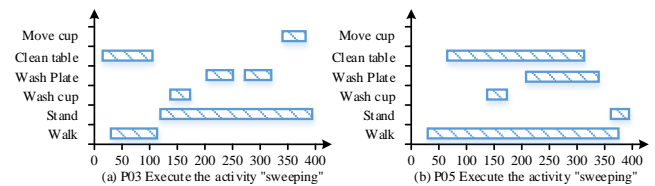


Fig. 3. Sensor events and duration triggered by different users performing the same activity.

Daily activities may seem to be irregular, but different users often have certain patterns embedded in their starting and ending actions when performing a particular activity. Assuming that daily activities are only related to start and end actions, the predicates in the event algorithm can be approximately reduced, thus streamlining the predicate structure. The study defines the start and end conditions of an activity as soft rules, and assigns weights to the formulas through parameter learning. The core of the event algorithm lies in the use of the law of inertia to describe the continuity of events, i.e., the state of an activity is determined only by its start and end conditions. The study defines hard rules based on the axioms of event algorithms and the law of the number of event triggers. The defined soft rules and hard rules are shown in Table I.

TABLE I. SOFT AND HARD RULES OF FRAMEWORK

Rule type	Number	Describe
Soft rule	1	$\forall Timestept : Happens(pa_1, t) \wedge Happens(pa_2, t+1) \rightarrow Initiates(a, t+1)$
	2	$\forall Timestept : Happens(pa_1, t) \wedge Happens(pa_2, t+1) \rightarrow Terminates(a, t+1)$
	3	$\forall Timestept : Happens(pa_1, t) \wedge [pa_1, frequency > K] \rightarrow Terminates(a, t+1)$
Hard rule	1	$\forall Timestept_1, t_2 : HoldsAt('close the front door', t_1) \wedge (t_1 + 1 \leq t_2) \rightarrow HoldsAt('open the front door', t_1)$
	2	$\forall Timestept_1, t_2 : HoldsAt('set the table', t_1) \wedge (t_1 + 1 \leq t_2) \rightarrow HoldsAt('eat breakfast', t_1)$
	3	$\forall Timestept_1, t_2 : HoldsAt('eat breakfast', t_1) \wedge HoldsAt('clear the table', t_1 + 1) \rightarrow HoldsAt('eat breakfast', t_1 + 1)$

The original version of the PSL model can be generated by rule definition, and these formulas are often given the same weights, so more appropriate weighting is needed through parameter learning. Commonly used weight learning methods are maximum likelihood estimation and maximum pseudo-likelihood estimation, and weight maximum pseudo-likelihood estimation uses pseudo-likelihood probabilities instead of likelihood probabilities. PSL-based inference needs to be performed in a closed PSL, so its inference problem is similar to that of probability maps. Activity inference is another important part of the study, and the activity inference mechanism in PSL includes maximum posterior probability inference and marginal probability inference. The PSL model constructed by activity modeling is defined as $P = (F_i, W_i)$, where F_i is the formula template in the model and W_i is the corresponding weights. The first step in performing Maximum A Posteriori Probabilistic Inference (MAP) is to convert the arithmetic and logical rules in PSL into a linear constraint form. This form of rule can be easily converted into the form of Horn clause as shown in Equation (9).

$$\left(\bigvee_{i \in I_j^+} l_i \right) \vee \left(\bigvee_{i \in I_j^-} \neg l_i \right) \quad (9)$$

In Eq. (9), L denotes the set of all words l_i included in a clause, and I_j^+ and I_j^- denote the set of non-negative and negative words, respectively. The PSL model includes soft and hard rules, both of which can be transformed into potential functions ϕ and inequality constraints φ as shown in Eq. (10).

$$\begin{cases} \phi(y, x) = \max \left\{ 1 - \sum_{i \in I_j^+} l_i - \sum_{i \in I_j^-} (1 - l_i), 0 \right\} \\ \varphi(y, x) = 1 - \sum_{i \in I_j^+} l_i - \sum_{i \in I_j^-} (1 - l_i) \leq 0 \end{cases} \quad (10)$$

In PSL, the MAP inference problem can be described as finding the maximum probability distribution of possible worlds with the set of variables Y given the sequence of observations $X = \{x_1, x_2, \dots, x_m\}$, where the normalization factor Z is fixed, as shown in Equation (11).

$$\arg \max_{s \in S} P(S = s) = \arg \max_{y \in Y} P(y|x) \quad (11)$$

According to Equation (11), it can be learned that the maximum probability distribution of possible worlds is equal to the distance between that world and the closure rule is minimized, so the MAP inference problem in PSL can be defined as shown in Equation (12).

$$\begin{cases} \arg \max_{s \in S} P(S = s) = \arg \min_y f_w(y, x) \\ s.t. \varphi_k(y, x) = 0, \forall k \in \varepsilon \\ \varphi_k(y, x) \leq 0, \forall k \in I \end{cases} \quad (12)$$

In Equation (12), S represents the set of possible worlds consisting of x and y , and ε and I represent the set of equational constraints and inequality constraints in PSL, respectively. The MAP inference problem in PSL is a convex optimization problem rather than an integer linear programming, thus giving birth to a consistent optimization algorithm for efficiently solving large-scale optimization problems. The core of this algorithm is the Alternating Direction Multiplier Method (ADMM). Assuming that the hidden predicates constituting the soft and hard rules in the model are different, let y_j be the state variable appearing in the potential function $\phi_j(y, x)$, $j = 1, 2, \dots, m$ and y_{k+m} be the state variable appearing in the hard constraint φ_k , $k = 1, 2, \dots, r$. For each hard constraint define an indicator function I_k as shown in Equation (13).

$$I_k [\varphi_k (y_{k+m}, x)] = \begin{cases} 0, & \text{the constraint is satisfied} \\ \infty, & \text{otherwise} \end{cases} \quad (13)$$

Finally, the variable Y_i is set to be a copy of the variable y_i $i = 1, 2, \dots, m+r$, so the MAP problem in PSL can be defined in the form shown in Equation (14).

$$\begin{cases} \arg \min_{y_i \in [0,1]} \sum_{j=1}^m w s_j \phi_j (y_i, x) + \sum_{k=1}^r I_k [\varphi_k (y(L, k+m), x)] \\ s.t. y_i = Y_i \end{cases} \quad (14)$$

Combining Eq. (12) and Eq. (14), it can be seen that it is feasible to solve the MAP problem in PSL according to the ADMM approach. The data flow of the human activity recognition framework based on PSL is shown in Fig. 4.

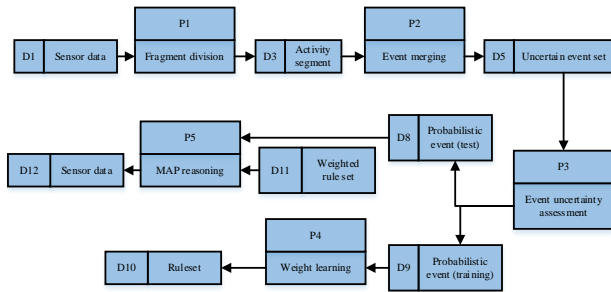


Fig. 4. Data flow of human activity recognition framework based on PSL.

As shown in Fig. 4, there are five core processing modules of the recognition framework, of which the weight learning and MAP inference modules use the platform’s original approach. The sensor data is segmented by calculating the window length corresponding to the sensor through the fragment partitioning module. Then, the sensor and time dependencies of the event are calculated in the event merging module, and uncertain events are generated by merging based on the event dependency. In the event uncertainty evaluation module, the corresponding relationship between the event and the sensor is first calculated and stored, and then the uncertainty is evaluated based on the calculated comprehensive weights. Finally, human activity recognition can be achieved through weight learning and MAP inference modules.

IV. RESULTS OF PSL-BASED HUMAN ACTIVITY RECOGNITION IN A SMART HOME ENVIRONMENT

In the superiority test of the PSL-based activity recognition method, the practicality test of the PSL method for solving the data uncertainty problem and the temporal complexity problem was focused on, and a comparative experiment was conducted with three existing recognition methods (Ontology, HMM, MLN) under two datasets. The datasets for the validity experiments are collected in the TWSTBED apartment of the WSU CASAS project, where 78 sensors including motion sensors, kettle sensors, faucet sensors, pillbox sensors, temperature sensors, etc. are deployed. The validation experiments of the PSL method’s ability to deal with data uncertainty use the ADL activity dataset containing the error data set, including both ADL and ADL-E. The ADL dataset

contains 6415 data obtained by 24 testers performing 5 different activities, while ADL-E is obtained by performing artificial activity omissions and errors based on ADL. The experiments to verify the ability of the PSL method to handle temporal complexity take the Interweaved ADL activity dataset with alternate execution activities, which consists of two parts of data obtained by 21 testers executing sequentially and by executing 8 activities in any order. Both sets of experiments were conducted using the ten-fold cross-check method, i.e. the data were equally divided into 10 parts, of which 9 parts were used as training data and 1 part was used as test data. The activities in the ADL dataset and the Interwoven ADL dataset are shown in Table II.

TABLE II. LIST OF ACTIVITIES IN THE ADL DATASET AND INTERWOVEN ADL DATASET

ADL dataset		Interweaved ADL Dataset	
Number	Activity	Number	Activity
SA1	Make a phone call	AC1	Fill medication dispenser
SA2	Wash hands	AC2	Watch DVD
SA3	Cook	AC3	Water plants
SA4	Eat	AC4	Answer the phone
SA5	Clean	AC5	Prepare birthday card
/	/	AC6	Prepare soup
/	/	AC7	Clean
/	/	AC8	Choose outfit

The performance evaluation indicators of other methods use the common F1 score, which is calculated as shown in formula (15).

$$F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (15)$$

In Equation (15), $\text{Recall} = TP / (TP + FN)$, $\text{Precision} = TP / (TP + FP)$, TP is the number of correctly identified activities, FN is the number of unidentified activities, and FP is the number of incorrectly identified activities. The experiments started with event pre-processing of the data, and the average duration of the activities and the number of triggered sensors in the CASAS dataset is shown in Fig. 5.

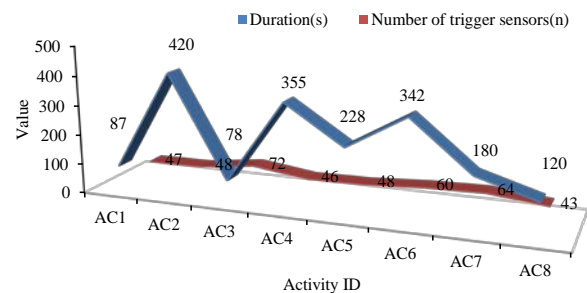


Fig. 5. Average duration of activity and number of triggered sensors in the CASAS dataset.

According to Fig. 5, the sensor corresponds to a window length in the range of [43, 54]. Then, event merging and uncertainty calculation are performed to transform the sensor data into probabilistic events. After defining the activity rules of the model, the PSL activity model is generated by learning the weights of the soft rules based on the training data. Finally, test data and PSL activity model are input, MAP inference is performed and the identification results are compared with the correct results. Experiment 1 verifies the ability of PSL method to handle data uncertainty, and the experimental results under ADL dataset and ADL-E dataset are shown in Fig. 6.

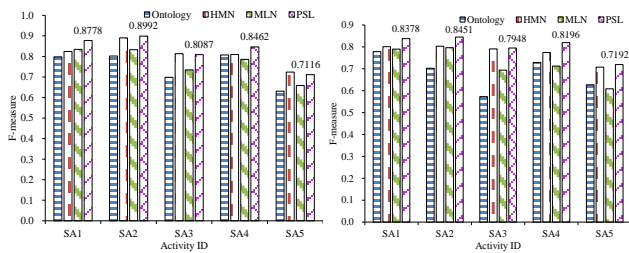


Fig. 6. Results to verify the ability of PSL method to deal with data uncertainty.

According to the experimental results of each method in Fig. 6(a) and Fig. 6(b) on the ADL dataset, it can be seen that the PSL method has good recognition performance for various activities, especially for activities SA1, SA2, and SA4. The average recognition rate for the three activities on the ADL dataset reached 87.44%, and even on ADL-E, it can maintain an average recognition rate of 83.42%. The execution methods of activities SA3 and SA5 have diversity, while the rules of PSL and MLN are manually defined, so the recognition performance of these two methods for SA3 and SA5 is slightly worse. The method of HMM is to establish recognition models based on training data, so the recognition rates for activities SA2 and SA5 are both above 70%. According to the experimental results on the ADL-E dataset in Fig. 6(b), it can be seen that the recognition rates of all methods have decreased, especially the Ontology method which decreased by 8.38%, followed by the MLN method which decreased by 5.76%, while the PSL method has the least significant decrease, with an average recognition rate of only 2.54%. The experimental results show that PSL has the best comprehensive recognition performance and can maintain good recognition performance even in the presence of erroneous activities, with a certain degree of recognition stability. The average recognition rate of each activity in two datasets using different recognition methods is shown in Fig. 7.

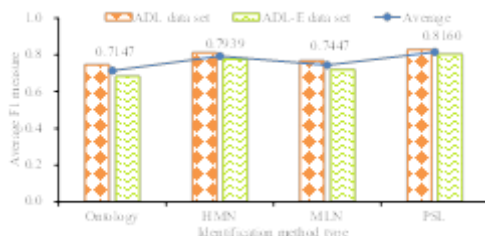


Fig. 7. Average recognition rate of activities in two data sets by different methods.

In Fig. 7, the PSL method achieved the highest average recognition rate, with an average recognition rate of 82.87% and 80.33% on the ADL and ADL-E datasets, respectively, while the Ontology method had the lowest recognition rate, with 74.73% (ADL) and 68.20% (ADL-E), respectively.

The average recognition rates of the four recognition methods on two datasets were 71.47% (Ontology), 79.39% (HMN), 74.47% (MLN), and 81.60% (PSL), respectively. The experimental results further demonstrate that the PSL method can maintain excellent recognition performance even in the presence of erroneous data, and verify the effectiveness of PSL in dealing with data uncertainty. The validation results of the ability of different algorithms to deal with temporal complexity are shown in Fig. 8.

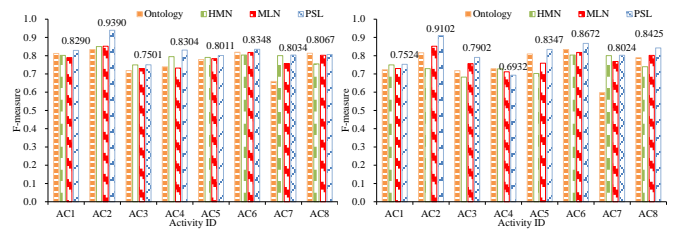


Fig. 8. Results of verifying the ability of different algorithms to deal with temporal complexity.

As shown in the experimental results of each method on the Sequential dataset in Fig. 8(a), PSL has the highest recognition rate among the four methods for 7 activities other than activity AC8, and the recognition rate for all other activities except for activity AC3 is above 80%. In the experimental results of each method on the interleaved dataset in Fig. 8(b), although the recognition rate of PSL for activities decreased under alternating and concurrent execution, the average recognition rate for activities is still the highest among the four methods, with an average recognition rate of 81.16%. The average recognition rate of each activity in two datasets using different recognition methods is shown in Fig. 9.

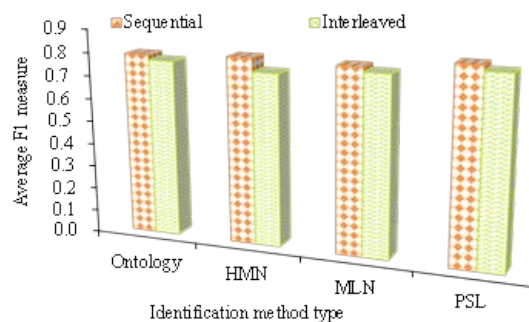


Fig. 9. Average recognition rate of each activity in two data sets by different recognition methods.

From Fig. 9, When there are alternating and concurrent recognition actions, the average recognition rate of the four activity recognition methods has decreased, with the HMM method showing the greatest decrease in recognition rate, with the average recognition rate decreasing from 80.32% to 74.65%. Although the average recognition rate of PSL on the Interleaved dataset has also decreased, it only decreased by

2.17%, and the decrease is not significant. It is still the highest average recognition rate among the four methods, maintaining an average recognition rate of 81.02%. The experimental results show that PSL can maintain the stability of activity recognition even when there are alternating and concurrent actions in recognition, verifying the effectiveness and superiority of PSL in processing temporal complexity data.

V. CONCLUSION

With the emergence and rapid development of smart cities and smart homes, ambient intelligence is widely used in various fields as an important part of artificial intelligence research. The study proposes a PSL-based activity recognition framework for human activity recognition in smart home scenarios, gives an evaluation method for sensor event uncertainty using DS evidence theory, and proposes an activity modeling method PSL-EC in conjunction with EC. the PSL method is compared with three other common recognition methods Ontology, HMN, and MLN on a publicly available dataset for experiments. to verify the ability of PSL to handle data uncertainty and temporal complexity. the ability of PSL to handle data uncertainty is far superior to the other three algorithms, with average recognition rates of 82.87% (ADL) and 80.33% (ADL-E) on the two sub-datasets. the average recognition rate of PSL decreases in the presence of concurrent and alternating activities is the least significant among the four methods, with declined by 2.17% and maintained the average recognition rate of 81.02%. The comprehensive experimental results show that the PSL-based human activity recognition method has excellent performance for both data uncertainty and temporal complexity. The different perception modes in the current smart home environment have their own advantages, so there is a trend towards a mixed use of multiple perception modes. However, these different patterns of scenario information exhibit heterogeneity, and current methods lack effective measures for processing these heterogeneous information. As an effective tool for eliminating information heterogeneity, ontology methods have been widely applied in other fields. However, the ability of ontology reasoning to handle uncertain information is poor, so future research can focus on the combination of ontology with methods such as MLN and PSL.

ACKNOWLEDGMENT

The research is supported by “The 14th Five-Year Plan” Project of Shaanxi Provincial Education Science Planning in 2022, “Research on the innovative application of project driven teaching mode based on OBE concept in application-oriented undergraduate colleges - taking “Android Program Design” as an example” (No. SGH22Y1838).

REFERENCES

[1] Y. Cui, L. Zhang, Y. Hou, G. Tian, Design of intelligent home pension service platform based on machine learning and wireless sensor network, *Journal of Intelligent & Fuzzy Systems*, 40(2): 2529-2540, 2021.

[2] S. Wan, L. Qi, X. Xu, C. Tong, Z. Gu, Deep learning models for real-time human activity recognition with smartphones, *Mobile Networks and Applications*, 25(2): 743-755, 2020.

[3] P. Asghari, E. Soleimani, E. Nazerfard, online human activity recognition employing hierarchical hidden Markov models, *Journal of Ambient Intelligence and Humanized Computing*, 11(3): 1141-1152, 2020.

[4] K. Ramirez-Amaro, Y. Yang, G. Cheng, A survey on semantic-based methods for the understanding of human movements, *Robotics and Autonomous Systems*, 119: 31-50, 2019.

[5] A. Prati, C. Shan, K. I. K. Wang, Sensors, vision and networks: from video surveillance to activity recognition and health monitoring, *Journal of Ambient Intelligence and Smart Environments*, 11(1): 5-22, 2019.

[6] L. K. Wang, R. Y. Liu, Human activity recognition based on wearable sensor using hierarchical deep LSTM networks, *Circuits, Systems, and Signal Processing*, 39(2): 837-856, 2020.

[7] D. Arifoglu, A. Bouchachia, Detection of abnormal behaviour for dementia sufferers using Convolutional Neural Networks, *Artificial Intelligence in Medicine*, 94: 88-95, 2019.

[8] N. Dua, S. N. Singh, V. B. Semwal, Multi-input CNN-GRU based human activity recognition using wearable sensors, *Computing*, 103(7): 1461-1478, 2021.

[9] H. Zhang, Z. Xiao, J. Wang, F. Li, E. Szczerbicki, A novel IoT-perceptive human activity recognition (HAR) approach using multihead convolutional attention, *IEEE Internet of Things Journal*, 7(2): 1072-1080, 2019.

[10] V. Bianchi, M. Bassoli, G. Lombardo, P. Fornacciari, M. Mordonini, I. D. Munari, IoT wearable sensor and deep learning: an integrated approach for personalized human activity recognition in a smart home environment, *IEEE Internet of Things Journal*, 6(5): 8553-8562, 2019.

[11] P. Agarwal, M. Alam, A lightweight deep learning model for human activity recognition on edge devices, *Procedia Computer Science*, 167: 2364- 2373, 2020.

[12] A. Artikis, E. Makris, G. Paliouras, A probabilistic interval-based event calculus for activity recognition, *Annals of Mathematics and Artificial Intelligence*, 89(1): 29-52, 2021.

[13] S. N. Tran, D. Nguyen, T. S. Ngo, X. S. Vu, L. Hoang, Q. Zhang, M. Karunanithi, On multi-resident activity recognition in ambient smart-homes, *Artificial Intelligence Review*, 53(6): 3929-3945, 2020.

[14] Y. Li, G. Yang, Z. Su, S. Li, Y. Wang, Human activity recognition based on multi-environment sensor data, *Information Fusion*, 91: 47-63, 2023.

[15] Nweke H F, Teh Y W, Mujtaba G, Al-garadi M A. Data fusion and multiple classifier systems for human activity detection and health monitoring: Review and open research directions. *Information Fusion*, 46: 147-170, 2019.

[16] P. Asghari, E. Soleimani, E. Nazerfard, online human activity recognition employing hierarchical hidden Markov models, *Journal of Ambient Intelligence and Humanized Computing*, 11(3): 1141-1152, 2020.

[17] Y. Zhang, G. Tian, S. Zhang, C. Li, A knowledge-based approach for multiagent collaboration in smart home: from activity recognition to guidance service, *IEEE Transactions on Instrumentation and Measurement*, 69(2): 317-329, 2019.

[18] X. Guo, Z. Shen, Y. Zhang, T. Wu, Review on the application of artificial intelligence in smart homes, *Smart Cities*, 2(3): 402-420, 2019.

[19] A. Keshavarzian, S. Sharifian, S. Seyedin, modified deep residual network architecture deployed on serverless framework of IoT platform based on human activity recognition application, *Future Generation Computer Systems*, 101: 14-28, 2019.

[20] Y. Xu, T. T. Qiu, Human activity recognition and embedded application based on convolutional neural network, *Journal of Artificial Intelligence and Technology*, 1(1): 51-60, 2021.

Design of Virtual Experiment Teaching of Inorganic Chemistry in Colleges and Universities Based on Unity3D

Xia Hu

Hunan University of Arts and Science, Chemistry and Materials Engineering College, Changde, 415000, China

Abstract—Focusing on the high-cost issues and many risk and uncontrollable factors in chemical experiment teaching in modern colleges and universities, a virtual reality inorganic chemistry (IC) simulation experiment strategy based on Unity3D technology is studied. Starting from the needs of IC experiment teaching, a collision detection algorithm (CDA) between the experimental environment and the virtual scene with Unity3D as the main technical framework is designed. The results show that the collision detection (CD) time of the CDA designed in the research is 99ms, the detection average value is 95.29%, and the accuracy variance is 0.021. The above values are better than the same type of algorithm. This shows that the CD accuracy and efficiency are higher, and the performance is stronger. In addition, the virtual chemistry experiment designed in the research can significantly improve the students' using attitude from the three main aspects of perceived ease of use (PEU), immersion and interactivity, and then enhance the students' learning effect. Therefore, the method of virtual experiment teaching of IC in colleges and universities based on Unity3D is effective and provides a new idea for modern teaching reform.

Keywords—Unity3D; inorganic chemistry; virtual reality; collision detection algorithm

I. INTRODUCTION

With the growth of contemporary computer technology, virtual reality (VR), a simulation technology based on computer technology, has gradually developed and matured. Driven by external experiential equipment and internal diverse functions, VR has gradually become a limited second reality environment. In this simulated environment realized with the external environment and equipment, users can get more intuitive experience and feedback than traditional computers. Therefore, VR technology is often used in interactive simulations such as online stores. Teaching experiment simulation is the main application direction [1-3]. It is very important to model the physical elements in the simulation environment. This is because the physical elements often need to interact with each other according to the operation of the experimenter during the experiment to completely restore the experimental process. Once there is a problem with the physical modeling, the experimental simulation will be unreal, and the feedback given to the operator will not be intuitive [4-6]. As the recently emerging 3D game production engine, Unity3D has its own particle system, physics engine, sky box and other engine elements that focus on interactivity. It includes functions such as scene view, game view, layer panel, engineering panel and monitoring panel, etc. which can

directly edit 3D virtual objects and mobilize 3D resources. Through these functions, it is easy to realize such functions as architectural visualization, 3D animation, 3D interactive environment editing, etc [7-9]. In teaching virtual experiment design, the application of Unity3D can not only give students a better virtual experiment experience, but also help schools save resources, realize self-development, and have stronger realizability.

Therefore, the research relies on Unity3D and other tools to use the interactive development environment to apply virtual reality technology to design the virtual experiment teaching of IC in colleges and universities. From the theoretical perspective of embodied cognition, the research integrates with virtual simulation technology, mobilizes learners' body movements and sensory perception systems in a highly realistic virtual learning environment, and helps learners to construct an instruction system. It not only enriches the practical research of embodied cognition in teaching applications in China, but also provides new ideas for the design of virtual simulation experiments. The virtual experiment teaching of IC in colleges and universities based on Unity3D has provided valuable theoretical and practical basis for the application of 3D technology in experimental teaching, and has played a certain role in promoting modern teaching reform.

II. RELATED WORK

The range of applications for Unity3D has grown significantly over the past few years, and the related research has also been increasing. Szłęg's team custom-developed a software that could simulate the operating environment of underwater autonomous robots using Unity3D. The software solved expensive hardware and high damage rate during the testing of underwater autonomous robots. Through this software, factors such as caustics and buoyancy of the underwater environment could be simulated to realize virtualized operation tests. The research results show that the software is robust in practical applications [10]. Zhu combined Unity3D with AI algorithms and applied them to game research and development. The researchers combined the behavior tree algorithm with the learning algorithm to establish a game NPC intelligent action model. By applying Unity3D technology to games, an intelligent game model was formed. The research results showed that the model could efficiently implement the intelligent behavior of NPC, and the system run more efficiently [11]. The Artal-Villa team applied Unity3D technology to traffic simulation, and proposed a framework

that could simulate the needs of pedestrians in traffic and restore the interactive relationship between vehicles and pedestrians. The research combined Unity3D with urban simulation open-source traffic simulator through transmission control protocol, and performed 3D simulation according to 2D scene. The findings showed that the traffic visualization ability of the system was strong [12]. Toriz applied Unity3D technology to the teaching system of the flipped classroom, and designed a laboratory classroom for landscape learning. The study analyzed the academic performance of the student group applying Unity3D technology through quantitative and qualitative dual analysis methods. The findings displayed that the average grades of these students significantly improved [13]. Wang's team used Unity3D technology to realize the research and development of the autonomous driving simulation system of unmanned ships, and designed the autonomous control simulation frameworks in the ocean and river driving scenarios in the system. Research used machine learning algorithms to enhance adaptive capabilities for autonomous navigation and control of unmanned ships. The research results showed that after algorithm optimization, the unmanned ship could successfully avoid obstacles in the simulation [14].

The research on virtual simulation experiments is also gradually enriched under the environment of rapid development of technologies such as Unity3D. For example, the Bogusevski team used 3D virtual technology to realize the simulation of campus experiments, thereby solving the problem that campus experiments were difficult to carry out under the condition of limited budget available on campus. The results showed that the study had a strong practicality [15]. Makransky et al. applied VR technology to the design of immersive simulation laboratory. The research used a pre-and post-test method to analyze whether the VR laboratory simulation program could improve students' self-learning effectiveness and scientific career vision. The results showed that students who used VR laboratory simulation programs had higher self-learning efficacy and scientific career vision [16]. Zhang combined VR technology with relay vibration protection algorithms, established a current protection simulation model in a VR simulation environment, and studied

the protection scheme under fault conditions. The results showed that the simulation scheme designed in the research could effectively simulate and demonstrate the micro electric protection model and current protection scheme [17]. The Schnack team applied VR technology to a simulation of a virtual store and compared it to a traditional computer desktop store. The research used hardware devices such as head-mounted displays, hand-held and body-sensing tracking devices to realize immersive store shopping simulations. The participants could interact on-site by moving their bodies. The research results showed that virtual store simulation technology was beneficial to promote customers' desire to buy [18]. O'Connor M's team proposed a simulation teaching strategy that combined 3D simulation technology and clinical radiology teaching experiments. The study used headsets and hand controllers in a VR kit to provide students with real-time experimental feedback, and simulated anatomy and radiology experiments through virtual ray simulation technology. The results showed that the teaching strategy of the research design gave students constructive feedback while reducing the risk of the experiment [19]. From the research related to Unity3D and virtual experimental teaching, the application area of VR and Unity3D in teaching experiments has greatly improved the teaching effect. However, most of the relevant researches are concentrated in the teaching subjects such as physics and medicine, which emphasize experimental operation. Therefore, the study applies Unity3D technology to chemistry teaching to establish a theoretical foundation for incorporating new technology into the teaching of chemistry experiments.

III. IC VIRTUAL EXPERIMENT TEACHING DESIGN BASED ON UNITY3D

A. Application Analysis of Virtual Experiment Teaching and Construction of Virtual Environment

When designing the virtual experiment teaching of IC, the research starts with students' interactive learning perception and experience cognitive habits in the VR learning environment. The learning of students is divided into three stages: VR, perceptual interaction and embodied cognition stage. The whole process is shown in Fig. 1.

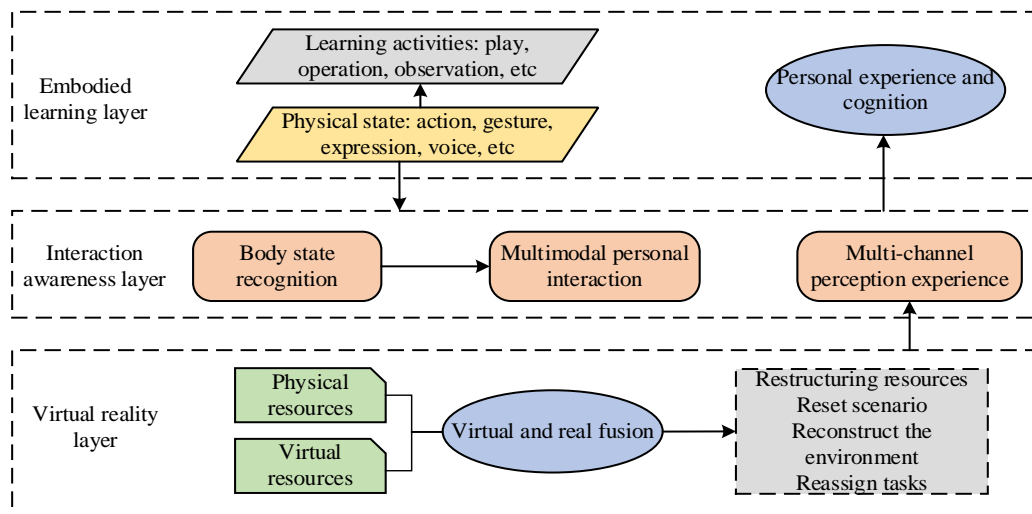


Fig. 1. Learning process.

The Unity3D platform is mainly used as the main technical carrier in the teaching design of IC virtual experiments. Unity3D is a game engine that includes functions such as virtual scenes, virtual 3D objects and physical calculations. This technology can realize the development of highly interactive applications such as architectural visualization, 3D games and animations, and support publishing on mobile platforms. This research focuses on using the physics engine to simulate Newtonian mechanics, and realizes the control, movement, and change of virtual objects in the virtual scene through scripts. This function can provide students with more intuitive and real experimental feedback. To realize the functional effect of the virtual experiment of IC, the study makes full use of VR technology to help students experience and establishes a computer simulation operating system in the virtual world. Through this system, the simulated natural environment required by users is constructed, so that students can feel the visual, auditory and tactile senses in the real scene in the virtual experiment. The research and design system is divided into 3D interactive characteristics, which are conception, immersion and interactivity. They are shown in Fig. 2.

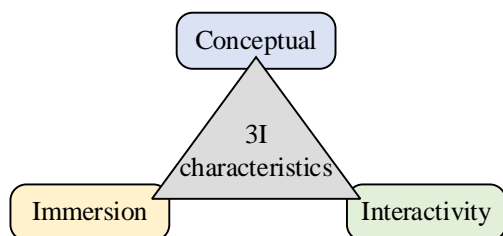


Fig. 2. 3D interactive features.

VR is widely employed in education, mainly including virtual laboratory, teaching, and skill training. The application of VR technology in these aspects can improve teaching quality, students' learning efficiency, expand students' thinking, and enhance students' practical ability. According to the different needs of users for the VR environment and the different implementation methods, there are four subsets of the VR system: immersive, augmented, distributed and desktop VR systems. The specific composition is displayed in Fig. 3.

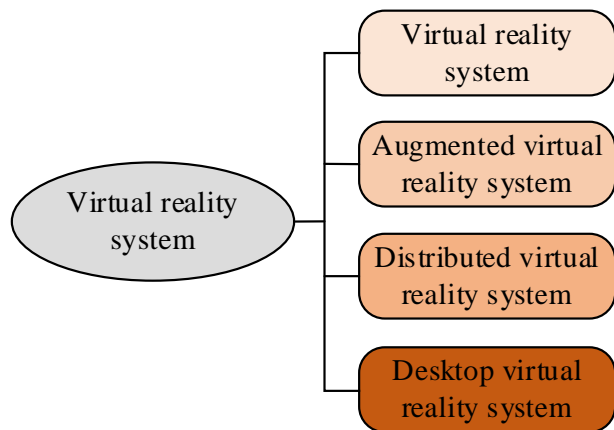


Fig. 3. Types of virtual reality systems.

The immersive VR system isolates the user's senses from the real world through devices like data gloves and helmets, and makes them fully immersed in the virtual environment. This type of system has a significant sense of immersion, but due to its high requirements for equipment and large investment, its development is also subject to certain restrictions. The augmented VR system refers to calculating the virtual information stored in the computer through the penetrating head-mounted display while the user observes the real environment. The virtual image obtained from the virtual information is superimposed on the real object to deepen the user's feeling of the real object. The distributed VR system relies on the immersive VR system, and enables people in different physical locations to share resources on the basis of the network to achieve a high collaborative work. Desktop VR systems use the computer screen as a channel for VR and user interaction. Users observe the VR environment through the screen, and interact with objects in the VR through simple devices. Through the device, the user can perform operations such as rotation and scaling on the controlled virtual object. Based on the cost of campus application, as well as the convenience of teaching when using the system for students, the research mainly uses desktop VR for virtual experiment design.

Virtual experiment is a VR learning resource that uses virtual equipment to conduct experimental exploration in a virtual environment constructed by a computer. The experimental environment of the virtual test is not limited by the site, which reduces the possible hazards in the real test and saves manpower and capital costs. The simulation experiment designed in the study is shown in Fig. 4.

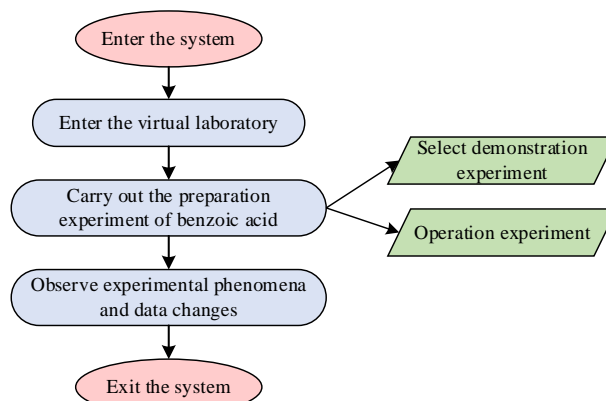


Fig. 4. Simulation experiment.

B. Virtual Anti-Collision Strategy Design

Aiming at the virtual experiment teaching of IC in colleges and universities, the research begins with the geometric modeling of chemical experiment instruments in VR. The objects in the virtual experiment are basically composed of geometric objects. Geometric modeling needs to process the topological and geometric information. Topological information is the relationship between each component that constitutes a geometric object. Geometry-related information is the position and size of objects in geometric space. The 3D graphics generate new graphics after operations such as

position movement and size adjustment. Such geometric transformation of 3D graphics is based on the coordinate origin and axes, and is realized through homogeneous coordinate transformation. When the 3D graphics are transformed, the point $P(x, y, z)$ is transformed $P'(x', y', z')$ in the geometric space. The coordinate transformation method is shown in formula (1).

$$\begin{cases} x' = x + C_x \\ y' = y + C_y \\ z' = z + C_z \end{cases} \quad (1)$$

In formula (1), C_x is translation amount of P on the X axis. C_y indicates the translation amount of P on the Y axis. C_z is translation amount of P on the Z axis. The matrix of translation of P is shown in formula (2).

$$C_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ C_x & C_y & C_z & 1 \end{bmatrix} \quad (2)$$

According to formula (2), the translation coordinate of P is displayed in formula (3).

$$[x' \ y' \ z' \ 1] = [x \ y \ z \ 1] \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ C_x & C_y & C_z & 1 \end{bmatrix} = [x \ y \ z \ 1] C_1 \quad (3)$$

When performing scale transformation on 3D graphics, if the center of the object is $P(x, y, z)$, then the point is transformed based on the coordinate axis. The scales of the transformation in the three mutually perpendicular directions of the coordinate system are u , v , and w . When the three transformation ratios are equal, the 3D graphics are transformed in equal proportions. When the $P(x, y, z)$ is changed $P'(x', y', z')$, the transformation matrix is shown in formula (4).

$$C_2 = \begin{bmatrix} u & 0 & 0 & 0 \\ 0 & v & 0 & 0 \\ 0 & 0 & w & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (4)$$

The scale transformation coordinate formula of the point P is as formula (5).

$$[x' \ y' \ z' \ 1] = [x \ y \ z \ 1] \begin{bmatrix} u & 0 & 0 & 0 \\ 0 & v & 0 & 0 \\ 0 & 0 & w & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = [x \ y \ z \ 1] C_2 \quad (5)$$

When a 3D object is rotated in the geometric space, the matrix of the point P' is different for coordinate axes. When the point P rotates around the X axis, the transformation of its coordinates on the X axis is shown in formula (6).

$$x' = x \quad (6)$$

The coordinate transformation of P on the Y axis is shown in formula (7).

$$y' = y \cos \alpha - z \sin \alpha \quad (7)$$

The coordinate change of P on the Z axis is shown in formula (8).

$$z' = y \sin \alpha + z \cos \alpha \quad (8)$$

In formula (7) and formula (8), α is rotation angle of P around the X axis. The transformation matrix can be obtained as formula (9).

$$C_3 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos \alpha & \sin \alpha & 0 \\ 0 & -\sin \alpha & \cos \alpha & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (9)$$

Then the coordinate transformation formula of P around the X axis is shown in formula (10).

$$[x' \ y' \ z' \ 1] = [x \ y \ z \ 1] \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos \alpha & \sin \alpha & 0 \\ 0 & -\sin \alpha & \cos \alpha & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = [x \ y \ z \ 1] C_3 \quad (10)$$

The transformation matrix of P rotating around the Y axis is as formula (11).

$$C_4 = \begin{bmatrix} \cos \beta & 0 & -\sin \beta & 0 \\ 0 & 1 & 0 & 0 \\ \sin \beta & 0 & \cos \beta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (11)$$

In formula (11), β is rotation angle of P around the Y axis. Then the coordinate transformation formula of P rotating around the Y axis is shown in formula (12).

$$[x' \ y' \ z' \ 1] = [x \ y \ z \ 1] \begin{bmatrix} \cos \beta & 0 & -\sin \beta & 0 \\ 0 & 1 & 0 & 0 \\ \sin \beta & 0 & \cos \beta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = [x \ y \ z \ 1] C_4 \quad (12)$$

The transformation matrix of P rotating around the Z axis is as formula (13).

$$C_5 = \begin{bmatrix} \cos \gamma & \sin \gamma & 0 & 0 \\ -\sin \gamma & \cos \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (13)$$

In formula (13), γ is rotation angle of P around the Z axis. Then the coordinate transformation formula of the P point rotating around the Z axis is shown in formula (14).

$$[x' \ y' \ z' \ 1] = [x \ y \ z \ 1] \begin{bmatrix} \cos \gamma & \sin \gamma & 0 & 0 \\ -\sin \gamma & \cos \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = [x \ y \ z \ 1] C_5 \quad (14)$$

Based on the basic transformation principle of 3D graphics in geometric space, the experimental scene of chemical experiments, the required equipment and their transformed positions in the geometric space of virtual experiments are geometrically modeled on Unity3D platform. After the geometric modeling of the experimental scene and objects is completed, CD technology is used to detect and process whether objects in the scene collide or not. The movement of objects in a virtual scene must also obey the objective law of the movement in reality, that is, two objects cannot occupy the same physical space. The CD technology will detect and obtain the information of the objects contact and cross-section. According to different virtual scenes, CDA are mainly divided into time-based and space-based. According to the different requirements of time changes, there are three types of time-based CDA, as shown in Fig. 5.

The static collision algorithm is suitable for use when objects in the virtual scene are stationary and do not collide. The discrete CDA detects whether objects in the virtual scene

collide at different discrete moments t_0, t_1, \dots, t_n . The continuous CDA is to continuously detect whether there is a collision between objects in the virtual scene within a time interval $[t_0, t_1]$. CDA based on space is broken up into two types according to image and object space. The algorithm solves the problem according to the 2D projection image information of the 3D object in the scene and the depth information of the image. The algorithm based on the object space solves the problem interactively according to the intersection of the 3D object with other 3D objects in the virtual space. The hierarchical bounding box algorithm (HBBA) is the most commonly used. It is suitable for object CD in more complex virtual 3D scenes. HBBA uses a bounding box slightly larger and simpler in shape than the object to wrap it around, making it a substitute for the geometric characteristics of the object. Common bounding boxes are spherical, AABB and OBB bounding boxes, as shown in Fig. 6.

In the virtual scene interaction, to enhance the speed of CD, the research proposes an improved CDA based on quantum technology ant colony (AC). The algorithm combines the AC algorithm with quantum computing technology, and improves the convergence speed of the AC algorithm through quantum computing technology. The representation of quantum AC algorithm is displayed in formula (15).

$$q_i = \left[\begin{array}{c} \cos(t_{i1}) \cos(t_{i2}) \dots \cos(t_{in}) \\ \sin(t_{i1}) \cos(t_{i2}) \dots \cos(t_{in}) \end{array} \right] \quad (15)$$

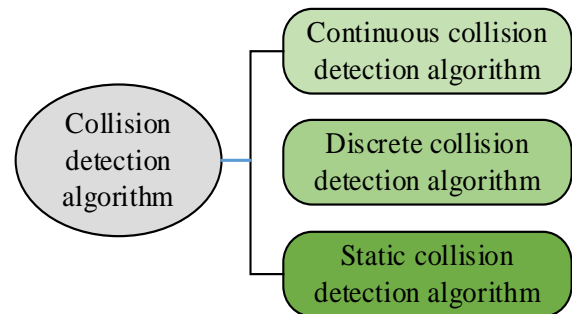


Fig. 5. Time-based collision detection algorithm.

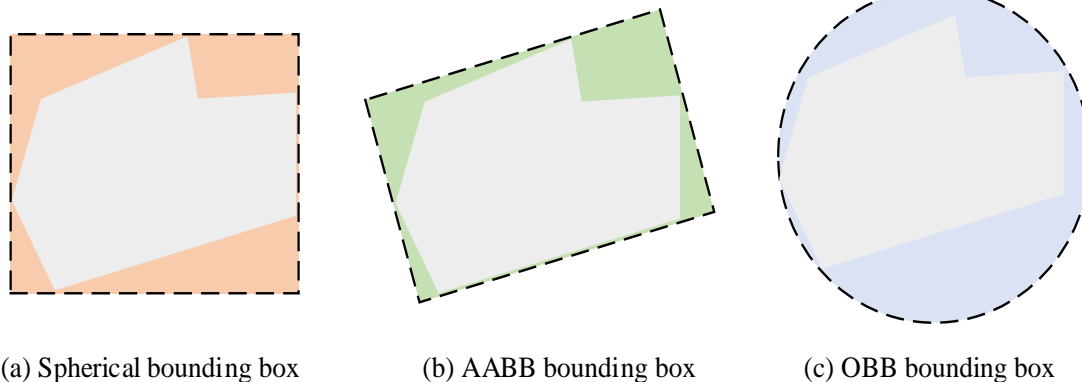


Fig. 6. Enclosure types.

In formula (15), $t_{ij} = 2\pi rand(i, j = 1, 2, \dots, m)$, the value range is a random value in $[0,1]$. m indicates the population size of the AC algorithm. In the AC algorithm, the ants calculate the path transition probability through the information appeal and heuristic information on the moving path, and then use the moving probability to calculate the movement of the ant to the next position. The path transition probability is shown in formula (16).

$$p(x_s) = \frac{[\tau(x_s)]^\alpha [\eta(x_s)]^\beta}{\sum_{x_s \in X} [\tau(x_s)]^\alpha [\eta(x_s)]^\beta} \quad (16)$$

In formula (16), X represents the position set of feature points of the ant colony in the search space. x_s represents the target position formed by the ant under the dual guidance of pheromone intensity and visibility. $\tau(x_s)$ means the pheromone intensity of the target position of the ant. $\eta(x_s)$ represents intensity of information generated by heuristics at the target location. α represents the weight of the information strength. β represents the weight of the heuristic information strength. The research mainly uses the quantum revolving door to update the position of the ants. The formula (17) demonstrates how the quantum revolving door works.

$$\begin{bmatrix} \cos t_{si} \\ \sin t_{si} \end{bmatrix} = \begin{bmatrix} \cos \theta_i & -\sin \theta_i \\ \sin \theta_i & -\cos \theta_i \end{bmatrix} \begin{bmatrix} \cos t_{ri} \\ \sin t_{ri} \end{bmatrix} \quad (17)$$

In formula (17), i is the serial number representing the probability amplitude of the qubit. θ_i represents the rotation angle of the probability amplitude. The study introduces mutation operators to enhance the diversity of races. The mutation operations are mainly realized by quantum NOT gates, as shown in formula (18).

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \cos \varphi_{ki} \\ \sin \varphi_{ki} \end{bmatrix} = \begin{bmatrix} \cos\left(\varphi_{ki} + \frac{\pi}{2} - 2\varphi_{ki}\right) \\ \sin\left(\varphi_{ki} + \frac{\pi}{2} - 2\varphi_{ki}\right) \end{bmatrix} \quad (18)$$

In formula (18), φ represents the state. The next position of the ant in the algorithm is determined by the strength and heuristic of the pheromone. The update rule of pheromone is shown in formula (19).

$$\begin{cases} \tau(x_s) = \tau(x_r) + \text{sgn}(\Delta f) * |\Delta f|^\alpha \\ \eta(x_s) = \eta(x_r) + \text{sgn}(\Delta \partial f) * |\Delta \partial f|^\beta \end{cases} \quad (19)$$

In formula (19), x_r represents the position that the ant is right now. x_s represents the new position formed after the ant changes its position. Once all the ants have completed the

position update, the pheromone intensity can be updated according to the formula (20).

$$\tau(x) = \begin{cases} (1-\rho)\tau(x) + \rho f(x), x = \tilde{x} \\ (1-\rho)\tau(x), x \neq \tilde{x} \end{cases} \quad (20)$$

The algorithm outputs the best individual with the collision threshold of the group by continuously updating until it reaches the iteration termination condition. When objects interact in a Unity3D scene, the situation is more complicated. For complex scenes, a simple CDA may result in penetrating objects or CD errors. If a grid collider is used, the computation of detection will be increased. Therefore, the improvement idea of the research is to combine the two algorithms. First, a spherical bounding box is established around the object to exclude objects that cannot be interacted with in the scene for a short time. Then a mesh collider is established for the object, so that the two CD calculations are performed on objects whose bounding boxes of two spheres have overlapped. The specific algorithm steps are as follows: establish a spherical bounding box for each object in the scene, and calculate the spherical bounding box according to the triangle information in the model. According to the i triangle the vertex vector of a , b , c and the patches n in a triangle, the calculation method of sphere enter o as shown in formula (21).

$$o = \frac{1}{3n} \sum_{i=1}^n (a+b+c) \quad (21)$$

The maximum distance d from the sphere center to all vertices is used as the radius of the spherical bounding box. Then a mesh collider is built for the 3D objects in the virtual scene, and the intersection calculation is performed for the spherical bounding box. If the intersection occurs, the next step of monitoring is not required. Finally, the object intersected by the spherical bounding box is monitored and calculated by the mesh collider. In interaction, the improved CDA is used to add colliders to the relevant instruments in the chemical experiment, and label the objects in the script to distinguish different objects.

IV. ANALYSIS OF TEACHING EFFECT OF VIRTUAL EXPERIMENT IN IC

A. Performance Analysis of Collision Algorithm

When analyzing the teaching effects of IC virtual experiments, the research first analyzes the collision algorithm designed for the research, and selects the SOLID CDA and the RAPID CDA as a comparative analysis. First, in the case of model collision and intersection, the detection intersection numbers of different algorithms are compared. Table I depicts the outcomes.

From Table I, the algorithm designed in the study always maintains the state with the largest number of detections in the range from 100 to 1000 intersecting facets. The detection ability of the algorithm designed in the research can be improved when detecting the number of intersecting triangles faces in the model. The algorithm designed in the study has the

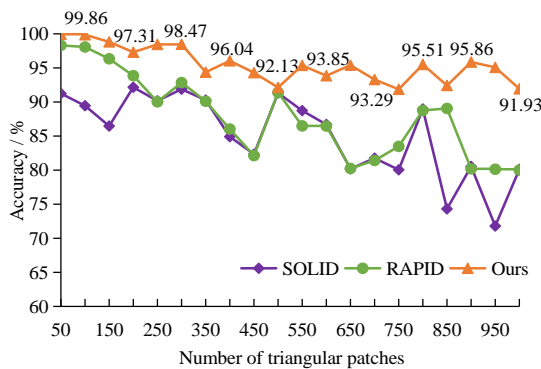
strongest detection capabilities. The model performance comparison is shown in Fig. 7.

TABLE I. COMPARISON OF MODEL INTERSECTION DETECTION

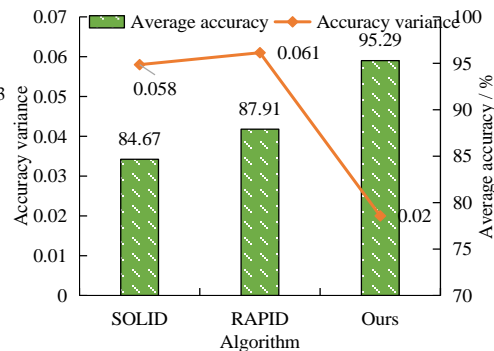
Number of Triangle Faces Intersected by the Model	Number of Intersections Detected by Different Algorithms		
	SOLID algorithm	RAPID algorithm	The algorithm designed in this research
100	88	98	100
200	186	190	194
300	278	282	294
400	342	370	386
500	462	444	474
600	512	512	562
700	576	582	656
800	694	658	762
900	718	730	858
1000	802	802	924

From Fig. 7, in CD time, the algorithm designed by the research only takes 99ms to complete the CD. The SOLID algorithm takes 105ms. The RAPID algorithm takes 132ms. The algorithm designed by the research has a significant advantage in CD efficiency. At the same time, the comparison of detection accuracy that the accuracy fold line of the algorithm designed in the study is at the highest position, in the range of 90% to 100%. The detection average is 95.29%, which is higher than the 84.67% of the SOLID algorithm and the 87.91% of the RAPID algorithm. The accuracy variance is 0.020, which is better than 0.058 of the SOLID and 0.061 of the RAPID. In the random collision performance test, the model designed by the research has advantages in both detection efficiency and accuracy.

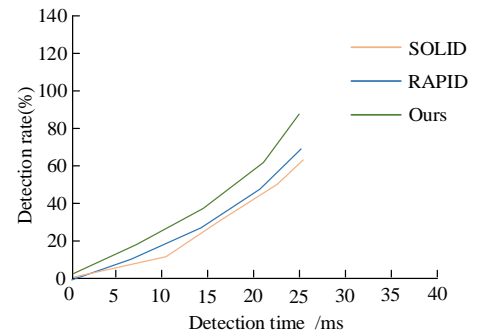
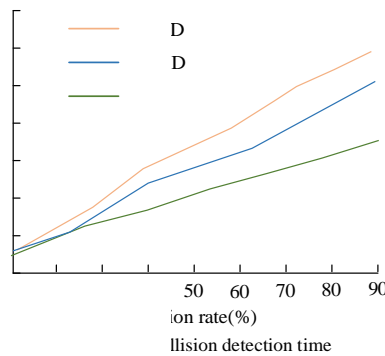
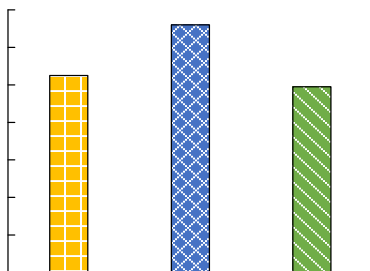
From Fig. 8, as the sampling features increases, the detection time also increases under the premise of achieving the same detection rate. Therefore, the feature samples should be adjusted in practical applications according to the requirements, and then simultaneously satisfy the effect of detection real-time and accuracy.



(a) Collision detection accuracy of three algorithms



(b) Mean and variance of collision detection accuracy of three algorithms



(e) Random collision detection rate comparison

Fig. 7. Model performance comparison.

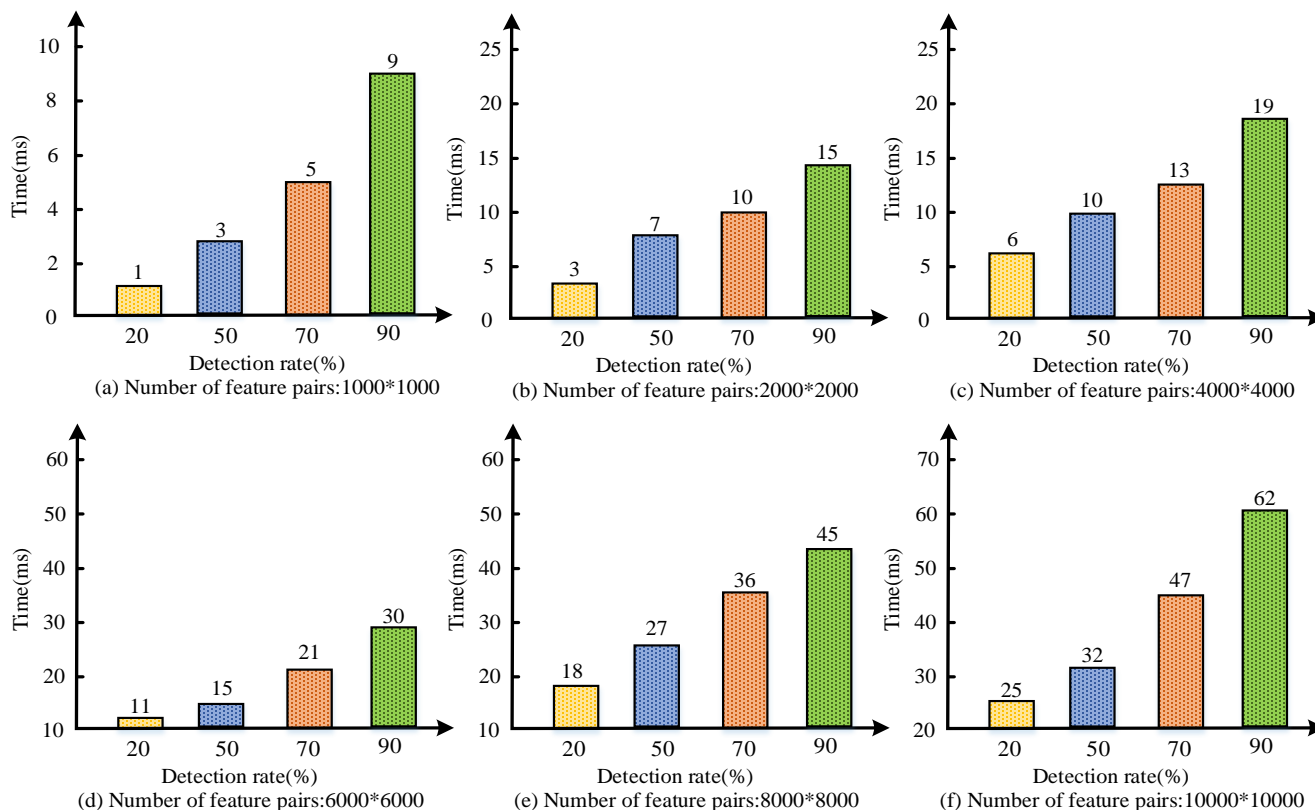


Fig. 8. Model performance comparison.

B. Analysis of Teaching Effect

In the effect analysis, the study adopts the method of teaching class experiment, and the experiment is conducted by grouping students. Questionnaires are distributed to investigate the teaching effect. Table II depicts the experimental settings.

TABLE II. TEST SETTINGS

Test subject	Type	Students at this chemistry course stage
	Quantity	52
	Group	2
	Number of people in each group	26
Test content		IC manipulative experiment
Virtual experiment hardware equipment		HTC VIVE
Virtual experiment software equipment		3ds Max
Data collection method		Questionnaire
data processing software		IBM SPSS Statistics 24

The research conducted a reliability and validity test on the distributed questionnaire. Table III depicts the results.

Cronbach’s Alpha coefficient of the research data is 0.782, which is above 0.75, indicating that the reliability is good. The KMO sampling suitability quantity is 0.652, and the significance is 0.000, which is above 0.6, and the validity is good. The study uses regression analysis to discuss the effects of virtual experiments on students’ learning motivation and use

attitude. The independent variables use four elements in the process of virtual experiments. They are PEU, perceived usefulness (PU), immersion and interactivity. Table IV displays the multiple regression analysis’s findings.

TABLE III. RELIABILITY AND VALIDITY TEST

Reliability test	Cronbach’s Alpha coefficient	0.782	
	Number of items	19	
Validity test	KMO Sampling Suitability Quantity	0.652	
	Bartlett’s test of sphericity	Approximate chi-square	443.843
		Significance	0.000
		Degrees of freedom	171

TABLE IV. MULTIPLE REGRESSION ANALYSIS RESULTS

Independent variable	Dependent variable	Standard coefficient	Adjusted fit	Significance
PEU	Learning motivation	0.691	0.467	0.000
PU		0.465	0.198	0.001
Immersion		0.578	0.318	0.000
Interactivity		0.412	0.156	0.003
PEU	Attitude	0.643	0.401	0.000
PU		0.107	0.172	0.002
Immersion		0.622	0.374	0
Interactivity		0.516	0.251	0.000

In learning motivation, the PEU and immersion of virtual experiments have the most significant impact on learning motivation, and the significance coefficients of the two variables are both 0.000. Followed by PEU, the significance coefficient is 0.001. The last is interaction, and the significance coefficient is 0.003. In use attitude, the PEU, immersion and interactivity of the virtual experiment have the most significant impact on it, and the significance coefficients of the three variables are all 0.000. Followed by PU, the significance coefficient is 0.002. Therefore, the Unity3D-based IC virtual experiment strategy in colleges and universities can significantly improve students' attitudes and use attitudes in chemical experiment teaching from the three main aspects of PEU, immersion and interactivity, thereby improving students' learning effects.

V. CONCLUSION

To solve the practical problems such as high consumption cost and many risk factors in chemical experiments in colleges and universities, starting with the characteristics of IC experiment teaching, a collision detection algorithm between virtual experiment environment and virtual scene based on Unity3D technology was designed. When designing the virtual experiment teaching of IC, it start with the students' interactive learning perception and experience cognition habits under the virtual reality learning environment, and divide the students' learning into three main levels: virtual reality, perception interaction and embodied cognition. At the same time, research was conducted on geometric modeling of chemical experimental instruments in virtual reality, and a virtual anti-collision strategy was designed, ultimately forming a virtual experimental teaching mode. The study analyzed experimental teaching through multiple regression analysis. The results show that in collision detection model performance, the algorithm designed in the study completed collision detection in just 99 ms, with the accuracy line at the highest position, within the range of 90% to 100%, with an average detection value of 95.29%; The accuracy variance is 0.020, and the numerical values are better than algorithms of the same type; And as the number of sampled features increases, the detection time of the algorithm increases while achieving the same detection rate. The collision detection algorithm designed in the research has superior performance. In teaching effectiveness, the survey data of the study has a Cronbach Alpha coefficient of 0.782, greater than 0.75, indicating good reliability. At the same time, the KMO sampling suitability scale is 0.652, greater than 0.6, with a significance of 0.000, indicating good reliability. The virtual experimental teaching designed for research can have a significant impact on learning motivation from PEU and immersion, with a significance coefficient of 0.000 for both variables. At the same time, it can have a significant impact on usage attitude from PEU, immersion, and interactivity, with a significance coefficient of 0.000 for all three variables. From this, the research designed model can effectively improve students' learning effectiveness in chemical experimental teaching. However, in practical applications, the number of feature samples should be adjusted in real-time according to the needs to simultaneously meet the real-time and accurate detection requirements. At present, 3D display mainly relies on the development of stereoscopic

display and sensing technology. The current system's three-dimensional sense cannot meet the needs of the system, and further research is needed in 3D display.

VI. DISCUSSION

The collision detection model designed in the study only completed collision detection in 99ms, with an average detection rate of 95.29%. Compared with previous studies, it significantly improved the collision detection speed and average detection rate in 3D displays. The study used regression analysis to analyze the impact of virtual chemistry experimental teaching on students' learning motivation and usage attitude. In learning motivation, the PEU and immersion of virtual experiments have the most significant impact on learning motivation, with significance coefficients of 0.000. In terms of usage attitude, the PEU, immersion, and interactivity of virtual experiments have the most significant impact on usage attitude, with significance coefficients of 0.000. From this, in virtual chemistry experimental teaching, the three aspects of PEU, immersion, and interactivity can significantly improve students' attitudes towards use, thereby enhancing their learning outcomes. From the theoretical perspective of embodied cognition, the research integrates with virtual simulation technology, mobilizes learners' body movements and sensory perception systems in a highly realistic virtual learning environment, and helps learners to construct an instruction system. It not only enriches the practical research of embodied cognition in teaching applications in China, but also provides new ideas for the design of virtual simulation experiments.

REFERENCES

- [1] P. M. G. Emmelkamp, K. Meyerbröker, "Virtual reality therapy in mental health." *Annual review of clinical psychology*, 2021(17): 495-519.
- [2] E. Chang, H. T. Kim, B. Yoo, "Virtual reality sickness: a review of causes and measurements." *International Journal of Human-Computer Interaction*, 2020, 36(17): 1658-1682.
- [3] A. D. Kaplan, J. Cruit, M. Endsley, P. A. Hancock. "The effects of virtual reality, augmented reality, and mixed reality as training enhancement methods: A meta-analysis". *Human factors*, 2021, 63(4): 706-726.
- [4] S. Poultasakis, S. Papadakis, M. Kalogiannakis, S. Psycharis, "The management of digital learning objects of natural sciences and digital experiment simulation tools by teachers." *Advances in Mobile Learning Educational Research*, 2021, 1(2): 58-71.
- [5] J. Wang, W. Zeng. "Research on the Realization Method of Augmented Reality based on Unity3D." *J. Robotics Netw. Artif. Life*, 2019, 6(3): 195-198.
- [6] Z. Jia, C. Gao, P. Dai, et al. "A study on a virtual simulation experiment with DNA in a biological evidence technology course during the COVID-19 pandemic." *Journal of Forensic Science and Medicine*, 2020, 6(4): 140-143.
- [7] G. Wen, Y. Xia, Y. Wang, et al. 2020, "Design of Virtual Training System for Horizontally Oriented Drill Based on Unity3D." *Journal of System Simulation*, 2020, 32(5): 801-807.
- [8] M. Akcaoglu, S. Dogan, C. B. Hodges. "Real Coding and Real Games: Design and Development of a Middle School Curriculum Using Unity 3D." *TechTrends*, 2022, 66(6): 931-937.
- [9] M. Lin, L. San, Y. Ding. "Construction of robotic virtual laboratory system based on Unity3D//IOP Conference Series: Materials Science and Engineering." *IOP Publishing*, 2020, 768(7): 072084.
- [10] X. Xu, "Design and application of intelligent manufacturing simulation system based on Unity3d", *International Symposium on Robotics*,

- Artificial Intelligence, and Information Engineering (RAIIE 2022). SPIE, 2022(12454): 134-141.
- [11] X. Zhu, "Behavior tree design of intelligent behavior of non-player character (NPC) based on Unity3D." *Journal of Intelligent & Fuzzy Systems*, 2019, 37(5): 6071-6079.
- [12] Q. Zhang, N. Chang, K. Shang. "Design and exploration of virtual marine ship engine room system based on Unity3D platform." *Journal of Intelligent & Fuzzy Systems*, 2020, 38(2): 1241-1247.
- [13] E. Toriz, "Learning based on flipped classroom with just-in-time teaching, Unity3D, gamification and educational spaces." *International Journal on Interactive Design and Manufacturing (IJIDeM)*, 2019(13): 1159-1173.
- [14] P. Szłęg, P. Barczyk, B. Maruszczak, et al. "Simulation Environment for Underwater Vehicles Testing and Training in Unity3D." *Intelligent Autonomous Systems 17: Proceedings of the 17th International Conference IAS-17*. Cham: Springer Nature Switzerland, 2023(577): 844-853.
- [15] D. Bogusevschi, C. Muntean, G. M. Muntean, "Teaching and learning physics using 3D virtual learning environment: A case study of combined virtual reality and virtual laboratory in secondary school." *Journal of Computers in Mathematics and Science Teaching*, 2020, 39(1): 5-18.
- [16] G. Makransky, G. B. Petersen, S. Klingenberg. "Can an immersive virtual reality simulation increase students' interest and career aspirations in science". *British Journal of Educational Technology*, 2020, 51(6): 2079-2097.
- [17] Q. Zhang, "Relay vibration protection simulation experimental platform based on signal reconstruction of MATLAB software." *Nonlinear Engineering*, 2021, 10(1): 461-468.
- [18] A. Schnack, M. J. Wright, J. L. Holdershaw, "Immersive virtual reality technology in a three-dimensional virtual simulated store: Investigating telepresence and usability." *Food Research International*, 2019(117): 40-49.
- [19] M. O'Connor, J. Stowe, J. Potocnik, Giannotti. N, "Rainford. 3D virtual reality simulation in radiography education: The students' experience." *Radiography*, 2021, 27(1): 208-214.

The Effective 3D MRI Reconstruction Method Driven by the Fusion Strategy in NSST Domain

Jin Huang¹, Lei Wang^{2*}, Muhammad Tahir³, Tianqi Cheng⁴, Xinping Guo⁵, Yuwei Wang⁶, ChunXiang Liu^{7*}

School of Computer Science and Technology, Shandong University of Technology, Zibo, 255000, China^{1,2,4,5,6}

Department of Computer Science, Mohammad Ali Jinnah University, Block 6 PECHS, Karachi, 75400, Pakistan³

School of Resources and Environmental Engineering, Shandong University of Technology, Zibo, 255000, China⁷

Abstract—The 3D reconstruction of medical images plays an important role in modern clinical diagnosis. Although the analytic-based, the iterative-based and the deep learning-based methods have been popularly used, there are still many problems to deal with. The analysis-based methods are not accurate enough, the iteration-based methods are computationally intensive, and the deep learning based methods are heavily dependent on the training of the data. To solve the default that only the single scan sequence is included in the traditional methods, a reconstruction method driven by the non-subsampled shearlet transform (NSST) and the algebraic reconstruction technique (ART) is proposed. Firstly, the multiple magnetic resonance imaging (MRI) sequences are decomposed into high-frequency and low-frequency components by NSST. Secondly, the low-frequency parts are fused with the weighted average fusion scheme and the high-frequency parts are fused with the weighted coefficient scheme that guided by the regional average gradient and energy. Finally, the 3D reconstruction is performed by using the ART algorithm. Compared with the traditional reconstruction methods, the proposed method is able to capture more information from the multiple MRI sequences, which makes the reconstruction results much clearer and more accurate. By comparing with the single-sequence reconstruction model without fusion, the experiments fully prove the accuracy and effectiveness.

Keywords—Multiple magnetic resonance imaging; 3D reconstruction; non-subsampled shearlet transform; the algebraic reconstruction

I. INTRODUCTION

Nowadays, the modern imaging technologies, such as the nuclear magnetic resonance imaging (MRI), the computer tomography (CT), sonography, have been popularly used by the clinicians to obtain the two-dimensional tomographic images of patients. These diagnostic techniques have great advantages over the traditional diagnostic methods in observing the structure of human tissue, extracting relevant information, and determining a reasonable course of treatment [1]. For example, CT can help to determine the size, shape, and extent of lesions in the examination of lung cancer and show the clear structure of the bone [2-3]. Ultrasound detection can be used to screen the fetus for malformations during the pregnancy check-up [4-5], and MRI can accurately show the tissue and tumors in the brain without interference from the skull during cranial examinations [6-7].

However, 2D images are isolated, static and abstract, making it difficult for physicians to infer the shape, size and

location of the lesions. Thus, it is necessary to convert 2D images into relational, dynamic, and concrete 3D reconstruction models [8-10]. On the other hand, the 3D reconstruction has been successfully applied in recent years, all of which can be classified into three schemes: the analytic scheme, the iterative scheme and deep learning scheme. For the first one, Feldkamp et al. proposed the filtered back-projection (FBP) [11-12], it solves the shape artifacts by convolving the projection before back-projection under each acquisition projection. FBP is very fast and supports the targeted reconstruction. But the filters are easily to produce obvious vibration and blur in the results. For the second scheme, Gordan et al. proposed the algebraic reconstruction technique (ART), which the reconstruction procedure is treated to be the problem of solving the mathematic equations [13-14]. It calculates the value of each voxel by solving the linear equations. In addition, the loop iterations will continuously optimize the objective function until the error is small enough to determine the simulated projection data is fully matched with the detection data. For the third scheme, Wang S et al. proposed the famous DeepVO model [15], where the recurrent convolutional neural network is employed to infer the pose directly from the source videos without any traditional visual odometry, improving the visual odometry loop in 3D reconstruction [16]. It can meet the real-time performance very well, but the accuracy is not satisfactory. Panigrahy C et al. proposed an improved PCNN model which uses the adaptive dual channel with a weighted parameter to fuse MRI sequences and SPECT images, and Seal A et al. proposed a wavelet transform method with random forest and à-trous strategy to fuse PET and CT images [17-18]. They fused different types of images, combining information from different images into one image, and have achieved very good results of reconstruction.

Considering the advantages and limitations of all these methods, with the aim of solving the problem that only one sequence can be included in the reconstruction, a pre-fused 3D reconstruction method is proposed. Firstly, multiple MRI sequences were decomposed into high-frequency and low-frequency parts, and then different strategies are used to fuse the low-frequency and high-frequency parts. Finally, ART reconstruction is performed on the images after fusion. This method concentrates the useful information of multiple sequences into one reconstruction model, which is helpful to observe the overall morphology of the lesions and their relationship with the surrounding structures in multiple directions and angles [19-20].

The contributions in this paper are summarized as follows: (1) The NSST transform based fusion algorithm is used to fuse the multiple MRI sequences and collect the useful information of each sequence; (2) The ART algorithm is used for 3D reconstruction, which makes the reconstruction be more accurate; (3) The Multiple sequences are pooled into a reconstruction model, which is more informative than the reconstruction model of a single sequence.

The rest of the paper is organized as follows. Section II briefly describes the related work and Section III, the proposed model. In Section IV, extensive experiments are used to evaluate the model. Finally, Section V gives a few concluding remarks.

II. RELATED WORK

A. The Nuclear Magnetic Resonance Imaging Sequences

In order to evaluate the various parameters of the detected tissue, weighted scans that highlight certain tissue characteristics can be obtained by adjusting the time parameters. Short echo time and repetition time can generate T1 sequence, long echo time and repetition time can generate T2 sequence and T1CE sequence can be generated by injecting contrast agent into the blood before MRI [21-22].

In Fig. 1, the differences between the three sequences are shown. In the T1-weighted image, the white is white matter, the gray is gray substance, and the black is cerebrospinal fluid, so the T1-weighted scans can be used for tomographic anatomy [23]. In the T2-weighted image, the highlighted area is usually the lesion, so the location and size of the lesion can be clearly shown from the T2-weighted scans [24]. As in T2-weighted image, the highlighted area in T1CE-weighted image is generally the lesion [25]. Furthermore, because of the contrast

agent, it can be used to reveal the intra-tumor conditions and discriminate the tumor and non-tumor lesions.

B. The Detail of the Non-Subsampled Shearlet Transform

The NSST is an advanced multi-scaled geometric analysis tool with translation invariance, simple computational cost and multi-dimension, which has been popularly used in many fields of computer vision [26-27].

As shown in Fig. 2, the process mainly includes two parts: non-subsampled pyramid (NSP) decomposition and direction localization [28]. The NSP uses a two-channel non-down-sampling filter bank to make the multi-scale [29]. The input image is decomposed by one level of NSP to produce a low-frequency part and a high-frequency part, and each subsequent level of NSP is iterated over the low-pass component to obtain the singularities in the image. So that, the two-dimensional image is decomposed by levels of NSP to obtain the subband images with the same size of the input image, which includes one low-frequency image and images of the same size but different scales. The direction localization is implemented by a shearlet filter (SF). From the pseudo-polarized coordinates to Cartesian coordinates, the Meyer wavelet is applied to construct a window function to obtain the shearlet filter, and the decomposed sub-band image is convolved with the Meyer window function in two dimensions to obtain the directional subband image.

Another good property is that there is no down-sampling operation in the process of the NSST, which makes it have translation invariance [30]. In addition, the NSST has very good localization properties and high directional sensitivity, and satisfies the parabolic scaling properties. All the above properties guarantee the great ability of capturing the important medical features from the MRI.

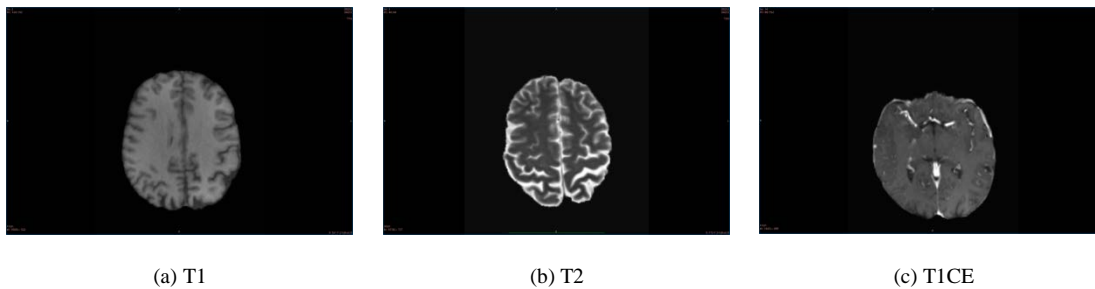


Fig. 1. Three examples of the MRI.

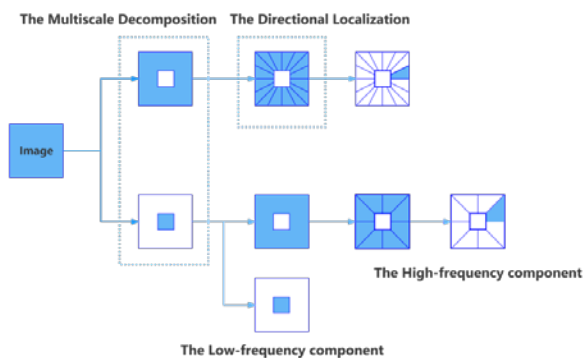


Fig. 2. The discrete process of the NSST.

C. The Algebraic Reconstruction

The basic idea of the Algebraic Reconstruction Technique (ART) is to transform the image reconstruction problem into a problem of solving a linear system of equations [31-32]. As the 2D section of a 3D object shown in the Fig. 3, it is divided into several discrete pixel regions, and set the image values inside each pixel region to be uniformly distributed, and denote the image values in the i -th pixel region by X_i .

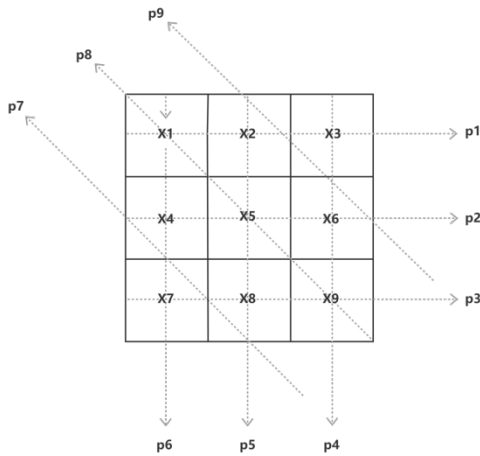


Fig. 3. The detection process to the 2D section.

Similarly, the detector plane is divided into the discrete projection cells and denote the projection value in the j projection cell by p_j . Then the detection process can be represented by a linear system as Equation (1).

$$\begin{pmatrix} x_{11} & x_{21} & \dots & x_{m1} \\ x_{12} & x_{22} & \dots & x_{m2} \\ \dots & \dots & \dots & \dots \\ x_{1n} & x_{2n} & \dots & x_{mn} \end{pmatrix} \begin{pmatrix} f_1 \\ f_2 \\ \dots \\ f_n \end{pmatrix} = \begin{pmatrix} P_1 \\ P_2 \\ \dots \\ P_n \end{pmatrix} \quad (1)$$

In theory, tomographic images can be obtained by solving the algebraic equations. But in practice this approach cannot be ideally achieved. Firstly, the system of equations has a unique solution only when the projection data is equal to the unknown image pixels, which is not actually satisfied. Secondly, there is seriously statistical noise in the projection data P obtained by nuclear medicine imaging, so when the ill-conditioned nature of the system transmission matrix C is serious, the direct solution method will be seriously disturbed by noise, and the obtained solution may be the same as the real situation. Therefore, an iterative method is usually used [33].

D. The Limitations of the Previous Research

In general, the 3D reconstruction methods can be classified into three types: the analytic-based methods, the iterative-based methods and the deep learning-based methods. The analytic-based methods are often only applicable to a certain scenario. They have achieved great performance, but there still have some problems. The iterative-based methods optimize the reconstruction process by using the idea of algebraic iteration, they require extensive calculations. The deep learning-based methods usually employee the complex neural networks, which require large amounts of data to train.

III. PROPOSED METHOD

A. The Whole Process

The whole process of the algorithm is shown in the Fig.4: (1) Perform the NSST decomposition on the three groups of MRI sequences to obtain their respective high and low frequency components; (2) For the low frequency part, the

weighted fusion method driven by guided filtering is used [34]; for the high frequency part, the fusion rule combining the average gradient, the regional energy index weighting coefficient and the larger absolute value is used [35]; (3) Perform ART on the fused information to obtain a 3D reconstruction model.

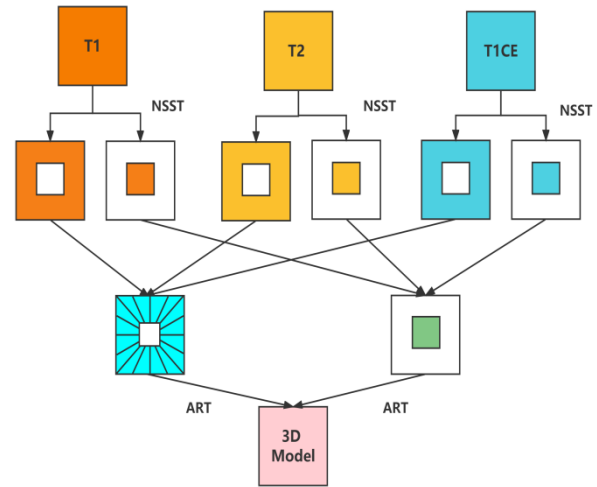


Fig. 4. The whole process of the proposed reconstruction method.

B. The Extraction and Fusion of the Information

The low-frequency part contains the energy of the image, and the coarse features in the image is correlated. Each pixel is not independent, and the pixels at any position of the pixel can show strong correlation. The low-frequency part is simply taken the maximum value or weighted average. Therefore, the weighted fusion method driven by the guided filtering is applied in this part. Taking the selected slices in the sequence as the guiding image and the other slices as the input image, use the low-frequency components of the original slices to subtract the respective output images of the guided filter to obtain the sharpened image. The improved regional sum of Laplacian energy (SML) of the sharpened image is used to determine the fusion weights, and then the low-frequency subband coefficients of the fused image are calculated by Equation (2). The A and B in Equation (2) can be expressed by Equation (3) and Equation (4).

$$A(m, n) = \frac{A}{B} \quad (2)$$

$$A = \sum_{i=1}^3 \omega_i(m, n) a_i(m, n) \quad (3)$$

$$B = \sum_{i=1}^3 \omega_i(m, n) \quad (4)$$

The high-frequency parts are mainly the feature information, such as the edge, textures of the input image. The human visual system is not sensitive to pixels, but sensitive to the edge, orientation and texture information of the image. The regional energy retains the input image details and also reflects the correlation of the input image. The larger the area energy,

the richer is the detailed information of the image. The average gradient of the image reflects the sharpness of the image, as well as the expressive ability of the image to the contrast of small details and the texture transformation characteristics of the image. Therefore, this paper proposes a fusion rule of regional average gradient and regional energy to jointly guide the weighting coefficient for high-frequency components. The following rules are adopted for high frequency components as Equation (5).

$$T^{l,k}(i,j) = \frac{1}{m \times n} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} \sqrt{A} \quad (5)$$

In Equation (5), A can be expressed as Equation (6).

$$A = \left(\frac{\partial f(i,j)}{\partial x} \right)^2 + \left(\frac{\partial f(i,j)}{\partial y} \right)^2 \quad (6)$$

In Equation (3), m and n are the rows and columns of the sequence slice. Respectively, $f(i,j)$ is the coefficient of the high-frequency component at pixel point (i,j) , and $T^{l,k}(i,j)$ is the average gradient at pixel (i,j) .

C. The Reconstruction Algorithm

It is well known that the reconstruction can be transformed into a system of linear equations as Equation (7).

$$\begin{cases} c_1 f = p_1 \\ c_2 f = p_2 \\ \dots \\ c_n f = p_n \end{cases} \quad (7)$$

$c_j = (c_{1j}, c_{2j}, c_{3j}, \dots, c_{nj})$ is the vector dot product operation. If we take the N components of the image to form an N dimensional space, then each set of estimates of the image is a point in this N dimensional space, and each of the linear equations in the set of equation is an $N-1$ dimensional hyperplane in this N dimensional space, which is orthogonal to the vector c_j . If the system of linear equations has a unique

solution, then these N hyperplanes have a unique intersection which is the solution we require.

From the initial image estimate f^0 , project this point onto the hyperplane represented by the first equation, use the projection point as the first image estimate f^1 , project the projection point onto the second hyperplane again to obtain a new projection point f^2 , continue to implement this procedure until the projection of all equations is carried out. Then the process of f^n can be expressed by the Equation (8).

$$f^n = f^{n-1} - \frac{(f^{n-1} \cdot c_n - p_n)}{c_n} \quad (8)$$

Finally, it can be further deduced into the Equation (9).

$$f^n = f_i^{n-1} - c_{in} \frac{\sum_i (f_i^{n-1} \cdot c_{in})}{\sum_i c_{in}^2} \quad (9)$$

IV. EXPERIMENTS AND DISCUSSION

The platform used for the experiments is the CentOS Linux x64, with CPU of Intel(R) Xeon(R) Silver 4114 and GPU of NVIDIA Tesla GPU in the memory of 16GB.

To verify the effectiveness of the fusion method and the effect of the 3D reconstruction result, two experiments with two datasets are implemented. Firstly, to verify the fusion method, the T1, T2, TICE sequences are compared with the fused sequences. The image definition (FD), standard deviation (STD), edge strength (EIN), information entropy (EN), and spatial frequency (SF) are the objective evaluation criterion for the effect of the fusion method. EN, STD and SF are the typical metrics of image fusion. EIN and FD can measure the segmentation and sharpness of the image edges, the clearer the edges; the more complete the shape of the individual organs will be after 3D reconstruction [36].

As shown in Table I, all the five metrics of the fused sequence are higher than the original sequence, proving the effectiveness of the proposed fusion method.

TABLE I. The comparison of T1, T2, TICE and the fusion results

Methods	EN	STD	SF	EIN	FD
T1	2.64	38.35	6.68	162.21	105.47
T2	2.55	35.42	6.74	164.14	107.31
TICE	2.56	40.94	7.17	165.45	108.95
Fusion	2.72	40.97	7.22	165.56	110.63

The ART is used to model the original sequences and the fused sequence separately, and observe the details of the model from the top, bottom, left and right directions to compare the texture, structure and clarity of the model.

As shown in Fig. 5 to Fig. 6, the images in (a)-(d) are 3D reconstructions of T1 sequence, (e)-(h) are 3D reconstructions

of T2 sequence, (i)-(l) are 3D reconstructions of TICE sequence and (m)-(p) are 3D reconstructions of fused sequence. The pictures of the top left, right and bottom directions are cropped. The bottom right corner is an enlarged image of the same region, so the details of the reconstructed model can be clearly compared. In the first three sets of reconstruction models, some areas are blurred and not smooth enough, but in

the proposed reconstruction model, the same areas are more smooth and clear without blurring. By comparing the enlarged area, the proposed reconstruction model is also clearly visible after zooming it, unlike the unfused reconstruction model that has blurring. The fused reconstruction model has clear texture, rich detailed information and good visual effect. In the comparison of the reconstruction model and the reconstruction

model before fusion, especially in the enlarged areas, it can be found that the results from the reconstruction method after fusion is much smooth, more feature details are remained, and the visual quantity is much better than the other results. All the facts fully prove the effectiveness and progressiveness of the proposed model.

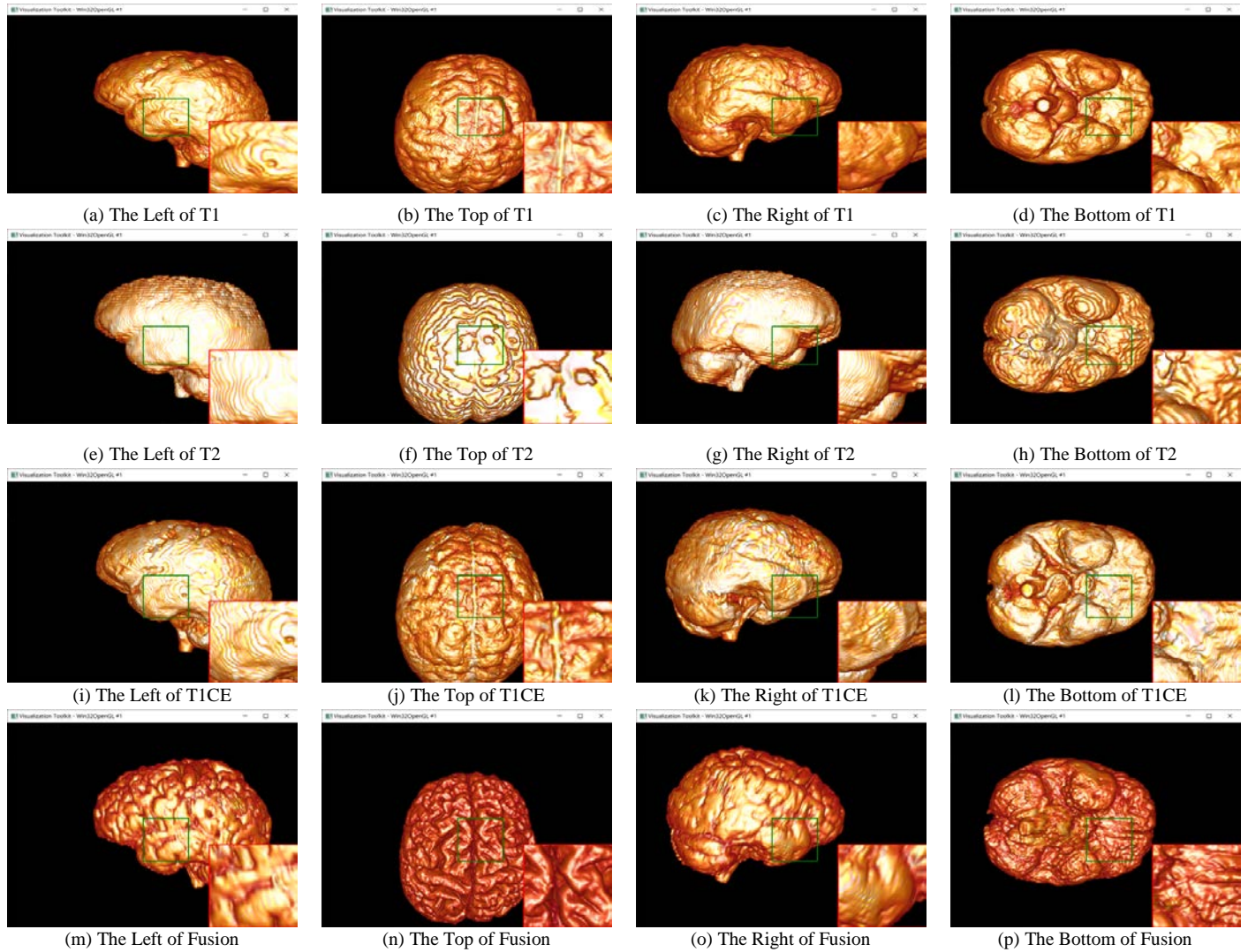
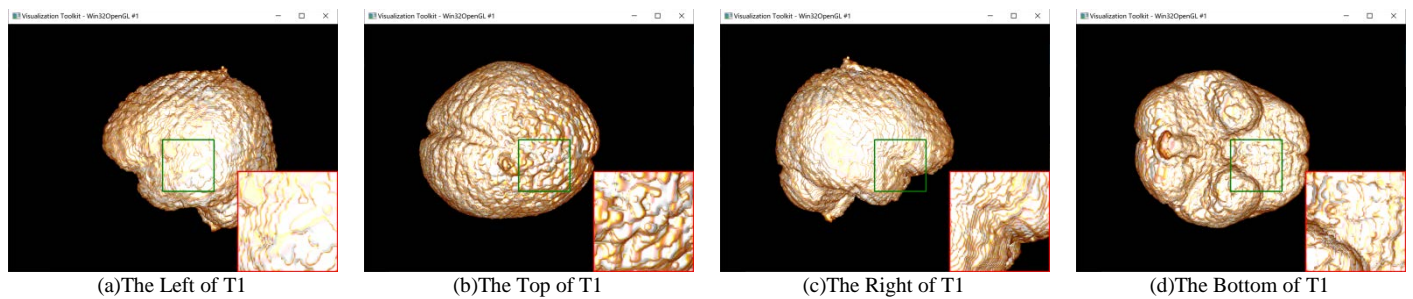


Fig. 5. The comparison of the 3D reconstruction results for the first group.



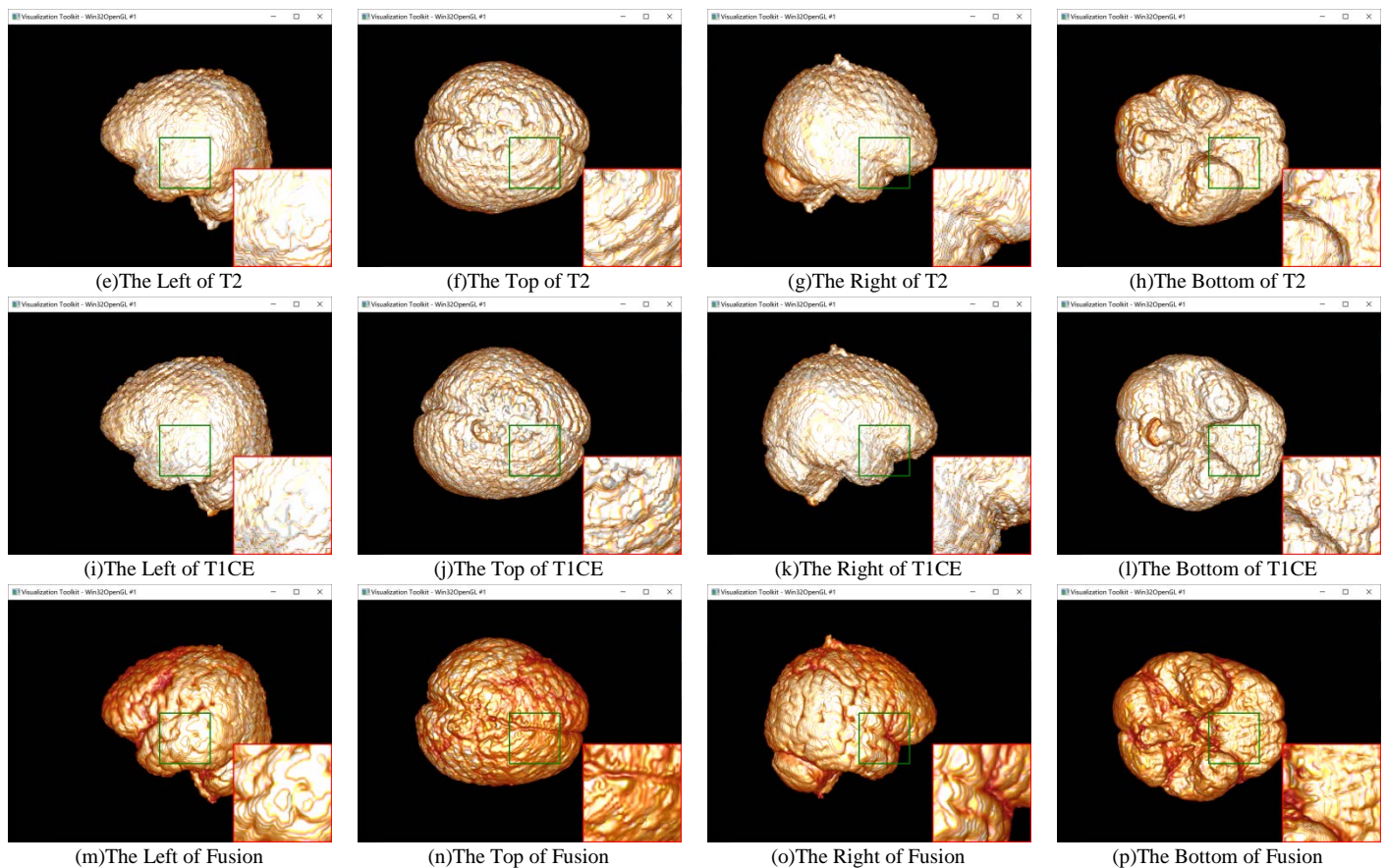


Fig. 6. The comparison of 3D reconstruction results for the second group.

V. CONCLUSION

An effective 3D reconstruction method, driven by NSST fusion scheme and ART, is proposed in this paper. Compared with other reconstruction methods, the reconstructed model can maintain more feature information by fusing multiple MRI sequences before reconstruction, which successfully meets the need that the reconstructed model contains more MRI sequence information. The post-fusion reconstructed model is compared with the pre-fusion reconstructed model by careful experiments. The proposed reconstructed model after fusion produces the good results with much clearer texture, smoother surface, and more detailed feature information.

In the future work, the proposed model will be optimized and extended to the other tasks of reconstruction, such as the Positron Emission Tomography (PET) and the Single-Photon Emission Computed Tomography (SPECT).

ACKNOWLEDGMENT

This study was supported by: A project ZR2021MF017 supported by Shandong Provincial Natural Science Foundation; The National Natural Science Foundation of China (61502282); A project ZR2020MF147 supported by Shandong Provincial Natural Science Foundation; The SDUT & Zibo City Integration Development Project(2020SNPT0055).

REFERENCES

- [1] Abadi E, Segars W P, Tsui B M W, et al. Virtual clinical trials in medical imaging: a review[J]. *Journal of Medical Imaging*, 2020, 7(4): 042805.
- [2] Kennedy K, Hulbert A, Pasquinelli M, et al. Impact of CT screening in lung cancer: Scientific evidence and literature review[C]//*Seminars in Oncology*. WB Saunders, 2022.
- [3] Zhang G, Jiang S, Yang Z, et al. Automatic nodule detection for lung cancer in CT images: A review[J]. *Computers in biology and medicine*, 2018, 103: 287-300.
- [4] Shainker S A, Coleman B, Timor-Tritsch I E, et al. Special Report of the Society for Maternal-Fetal Medicine Placenta Accreta Spectrum Ultrasound Marker Task Force: Consensus on definition of markers and approach to the ultrasound examination in pregnancies at risk for placenta accreta spectrum[J]. *American Journal of Obstetrics and Gynecology*, 2021, 224(1): B2-B14.
- [5] Leung Y, Shim H H, Wilkens R, et al. The role of bowel ultrasound in detecting subclinical inflammation in pregnant women with Crohn's disease[J]. *Journal of the Canadian Association of Gastroenterology*, 2019, 2(4): 153-160.
- [6] Wadhwa A, Bhardwaj A, Verma V S. A review on brain tumor segmentation of MRI images[J]. *Magnetic resonance imaging*, 2019, 61: 247-259.
- [7] Abd-Ellah M K, Awad A I, Khalaf A A M, et al. A review on brain tumor diagnosis from MRI images: Practical implications, key achievements, and lessons learned[J]. *Magnetic resonance imaging*, 2019, 61: 300-318.
- [8] Pichat J, Iglesias J E, Yousry T, et al. A survey of methods for 3D histology reconstruction[J]. *Medical image analysis*, 2018, 46: 73-105.

- [9] Singh S P, Wang L, Gupta S, et al. 3D deep learning on medical images: a review[J]. *Sensors*, 2020, 20(18): 5097.
- [10] Ahishakiye E, Bastiaan Van Gijzen M, Tumwiine J, et al. A survey on deep learning in medical image reconstruction[J]. *Intelligent Medicine*, 2021, 1(03): 118-127.
- [11] Yan Q, Dong H, Su J, et al. A review of 3D printing technology for medical applications[J]. *Engineering*, 2018, 4(5): 729-742.
- [12] Shi L, Liu B, Yu H, et al. Review of CT image reconstruction open source toolkits[J]. *Journal of X-ray Science and Technology*, 2020, 28(4): 619-639.
- [13] Greffier J, Frandon J, Larbi A, et al. CT iterative reconstruction algorithms: a task-based image quality assessment[J]. *European radiology*, 2020, 30(1): 487-500.
- [14] Shen R H, He Y T, Ming W Q, et al. Electron tomography for sintered ceramic materials by a neural network algebraic reconstruction technique[J]. *Journal of Materials Science & Technology*, 2022, 100: 75-81.
- [15] Wang S, Clark R, Wen H, et al. Deepvo: Towards end-to-end visual odometry with deep recurrent convolutional neural networks[C]//2017 IEEE international conference on robotics and automation (ICRA). IEEE, 2017: 2043-2050.
- [16] Girshick R. Fast r-cnn[C]//Proceedings of the IEEE international conference on computer vision. 2015: 1440-1448.
- [17] Han J, Cao Y, Xu L, et al. 3D Reconstruction Method based on Medical Image Feature Point Matching[J]. *Computational and Mathematical Methods in Medicine*, 2022, 2022.
- [18] Maken P, Gupta A. 2D-to-3D: A Review for Computational 3D Image Reconstruction from X-ray Images[J]. *Archives of Computational Methods in Engineering*, 2022: 1-30.
- [19] Panigrahy C, Seal A, Mahato N K. MRI and SPECT image fusion using a weighted parameter adaptive dual channel PCNN[J]. *IEEE Signal Processing Letters*, 2020, 27: 690-694.
- [20] Seal A, Bhattacharjee D, Nasipuri M, et al. PET-CT image fusion using random forest and à-trous wavelet transform[J]. *International journal for numerical methods in biomedical engineering*, 2018, 34(3): e2933.
- [21] Wolf M, de Boer A, Sharma K, et al. Magnetic resonance imaging T1- and T2-mapping to assess renal structure and function: a systematic review and statement paper[J]. *Nephrology Dialysis Transplantation*, 2018, 33(suppl_2): ii41-ii50.
- [22] Baessler B, Luecke C, Lurz J, et al. Cardiac MRI texture analysis of T1 and T2 maps in patients with infarctlike acute myocarditis[J]. *Radiology*, 2018, 289(2): 357-365.
- [23] Warnica W, Al-Arnawoot A, Stanimirovic A, et al. Clinical impact of cardiac MRI T1 and T2 parametric mapping in patients with suspected cardiomyopathy[J]. *Radiology*, 2022: 220067.
- [24] Yue Z, Wang X, Wang Y, et al. Clinical-Radiomics Nomogram from T1W, T1CE, and T2FS MRI for Improving Diagnosis of Soft-Tissue Sarcoma[J]. *Molecular Imaging and Biology*, 2022: 1-12.
- [25] Yue Z, Wang X, Wang Y, et al. Clinical-Radiomics Nomogram from T1W, T1CE, and T2FS MRI for Improving Diagnosis of Soft-Tissue Sarcoma[J]. *Molecular Imaging and Biology*, 2022: 1-12.
- [26] Huang Y, Bi D, Wu D. Infrared and visible image fusion based on different constraints in the non-subsampled shearlet transform domain[J]. *Sensors*, 2018, 18(4): 1169.
- [27] Tan W, Tiwari P, Pandey H M, et al. Multimodal medical image fusion algorithm in the era of big data[J]. *Neural Computing and Applications*, 2020: 1-21.
- [28] Yin M, Liu X, Liu Y, et al. Medical image fusion with parameter-adaptive pulse coupled neural network in nonsubsampled shearlet transform domain[J]. *IEEE Transactions on Instrumentation and Measurement*, 2018, 68(1): 49-64.
- [29] Radhakrishnan A, Uhler C, Belkin M. Downsampling leads to image memorization in convolutional autoencoders[J]. 2018.
- [30] Jose J, Gautam N, Tiwari M, et al. An image quality enhancement scheme employing adolescent identity search algorithm in the NSST domain for multimodal medical image fusion[J]. *Biomedical Signal Processing and Control*, 2021, 66: 102480.
- [31] Komarov D A, Samouilov A, Ahmad R, et al. Algebraic reconstruction of 3D spatial EPR images from high numbers of noisy projections: an improved image reconstruction technique for high resolution fast scan EPR imaging[J]. *Journal of Magnetic Resonance*, 2020, 319: 106812.
- [32] Vaniqui A, Schyns L E J R, Almeida I P, et al. The effect of different image reconstruction techniques on pre-clinical quantitative imaging and dual-energy CT[J]. *The British journal of radiology*, 2019, 92(1095): 20180447.
- [33] Komarov D A, Samouilov A, Ahmad R, et al. Algebraic reconstruction of 3D spatial EPR images from high numbers of noisy projections: an improved image reconstruction technique for high resolution fast scan EPR imaging[J]. *Journal of Magnetic Resonance*, 2020, 319: 106812.
- [34] Shehanaz S, Daniel E, Guntur S R, et al. Optimum weighted multimodal medical image fusion using particle swarm optimization[J]. *Optik*, 2021, 231: 166413.
- [35] Liu Y, Wang L, Cheng J, et al. Multi-focus image fusion: A survey of the state of the art[J]. *Information Fusion*, 2020, 64: 71-91.
- [36] Kaur H, Koundal D, Kadyan V. Image fusion techniques: a survey[J]. *Archives of computational methods in Engineering*, 2021, 28(7): 4425-4447.

Integrating Dropout Regularization Technique at Different Layers to Improve the Performance of Neural Networks

B. H. Pansambal¹, A.B. Nandgaokar²

Department of Electronics & Telecommunications Engineering,
Dr. Babasaheb Ambedkar Technological University, Lonere, 402103, India

Abstract—In many facial expression recognition models it is necessary to prevent overfitting to check no units (neurons) depend on each other. Therefore, dropout regularization can be applied to ignore few nodes randomly while processing the remaining neurons. Hence, dropout helps dealing with overfitting and predicts the desired results with more accuracy at different layers of the neural network like ‘visible’, ‘hidden’ and ‘convolutional’ layers. In neural networks there are layers like dense, fully connected, convolutional and recurrent (LSTM- long short term memory). It is possible to embed the dropout layer with any of these layers. Model drops the units randomly from the neural network, meaning model removes its connection from other units. Many researchers found dropout regularization a most powerful technique in machine learning and deep learning. Dropping few units (neurons) randomly and processing the remaining units can be considered in two phases like forward and backward pass (stages). Once the model drops few units randomly and select ‘n’ from the remaining units it is obvious that weight of the units could change during processing. It must be noted that updated weight doesn’t reflect on the dropped units. Dropping and stepping-in few units seem to be very good process as those units which step-in will represent the network. It is assumed to have maximum chance for the stepped-in units to have less dependency and model gives better results with higher accuracy.

Keywords—Convolutional layer; visible layer; hidden layer; dropout regularization; long short term memory (LSTM); deep learning; facial expression recognition

I. INTRODUCTION

Idea behind introducing the dropout layer (technique) helps classifying accuracies with minimized error rate than without having the same. Many applications process the large data and it is expected to have better accuracy with minimized loss and error rate. Applications will process data from different source e.g., speech, images, videos etc.

This adaptive nature helps developing a robust model. It is also possible to introduce the dropout layer with fully connected layers, recurrent layer such as long short-term memory layer (LSTM) and convolutional layer. There are visible layers, hidden layers and dropout layer can be added later to get better classification of the objects. Actually, dropout is a regularization technique which trains ‘n’ neural networks in parallel. Neural networks are complex due to various hidden layers. It is necessary to find the relationship between those neurons. If model fails to regularize the network

properly then relationship results into noise. There are many techniques that exist to avoid noise, and overfitting and are ‘early stopping’, ‘dropout regularization’, ‘weight penalties’, and ‘short weight sharing’. It’s a type of generalization and in some cases if the network is larger and complex may consume more memory and time. Dropping and selecting the neurons may consist of ‘n’ iterations and after each iteration network may get changed. Few researchers termed ‘dropout mechanism’ as an ‘ensemble mechanism’ because after each iteration network changes and can be termed as sub-network (new). Actually these are the structures and bunch of these sub-networks can be used for training purpose. Objective is to prevent the overfitting in the model and train the given dataset. To achieve an error rate with minimized losses dropout layer can be added in the proposed model. In some cases, adding dropout have effect of making training process noisy and one can say the model becomes ‘adaptive’ which learns from every phase. In dropout regularization method few neurons are randomly dropped and select other neurons.

Section II gives introduction to existing methods with advantages and gaps to implement the proposed system. Section III focused on proposed model with algorithmic details and methodologies. Section IV elaborates the results obtained in proposed approach with different dropout rates, epochs, and batch sizes. In the end conclusion remarks the implemented approach with future research directions.

II. LITERATURE REVIEW

Authors [1, 24] have developed a model which extracts features from given images and trained the model to process the upcoming features automatically. Authors have focused on important aspect of the emotion recognition i.e. working on static and dynamic images i.e. real-time images or images extracted (frames) from the video. Model processes the video as sequences of different frames and each frame consists of images. Model processes the features extracted from given images automatically by introducing 03 steps as follows –

- Image acquisition
- Feature extraction
- Expression recognition

Authors [2] also suggested, one can use support vector machine (SVM), naïve bayes, lexicon methods to recognize emotions from different conversations. Hence model generates

different vectors according to the different statements/conversations. Model helps identifying sentiments of people and knowing the depth of conversation. Model could help analyzing different styles of conversations, time required for certain conversation etc.

Authors [3, 22] have introduced robots and embed a proposed model to process real-life images. In this model robot extracts images from camera, video etc. and analyzes these frames. FERW model – facial expression recognition in the wild processes the features and gives corresponding feedback.

Authors [4, 11, 18] have used CNN and auto-encoders to extract different features from images for classification purpose. Authors have proposed 06 architectures out of which 02 are trained on Japanese female facial expressions and 04 with Berlin database. 02 architectures dealt with image-based emotions and 04 with speech recognition. CNN architecture consists of 03 convolution layers, 03 pooling layers, 03 dense layers with output layer. These convolution and pooling layers helps in extracting different features from emotions. Auto-encoder helps in reducing the dimensions. Authors [5, 6, 25] suggested that, ML machine learning and methods especially from the deep learning techniques can handle complex feature extraction from images and also helps in classifying them accordingly. “Gabor Filter” is applied to analyze the textures, edge detection, and extracting different features from the images. Authors pointed good response in edge detection can be obtained after application of the Gabor filter. Authors [7] have focused on 06 elementary emotions like happy, sad, fear, anger, surprise and disgust. Using MATLAB deep learning methods are implemented to extract and process the features from images. Model processes the JAFFE dataset which consists of 06 elementary emotions. System processes large volume of data. However, data could be labeled or unlabeled. Authors [8] have proposed a model to process unlabeled data. A hybrid label-less learning is proposed to automatically process the unlabeled data. Main advantage of this approach is human intervention is almost reduced and model automatically labels the data. Authors have proposed “LLEC” label less learning for emotion cognition. Authors suggested when labeling the unlabeled data prediction probability plays an important role. Entropy is used as a measure to assess the prediction uncertainty. Authors [9] have proposed a model using IoT devices which captures the images of the human and passes to the system. Model uses motion sensor to detect the motion of human. Camera attached with the module gets activated once the motion sensor detects the motion of human and captures the image using HAAR - feature based cascade classifier. Authors [10, 13, 26] have proposed a comparative approach for different deep learning architectures available in “Keras” for emotion recognition. Pre-trained models like VGC-16, ResNet152V2, Inception V3 and Xception etc. Authors have used 02 datasets Cohn-Kanade dataset and JAFFE Japanese female facial emotion. All the pre-trained models are evaluated on these datasets and their accuracies are measured.

A three-class annotation approach [12, 19, 23] is proposed to get the stimuli behind emotions. Authors have developed emotion classification framework using LSTM. Framework also consists of threshold scheme and categorizes emotions in

different segments like anxiety, anger, disgust, sadness and joy etc. and ratings are assigned in the range from 1 to 9. Authors have also used EEG signals and facial videos to extract the features from images. Authors have also concluded EEG feature-based classifier can give better accuracy when used with thresholding scheme. Authors [14, 27] have considered 05 emotions and developed a model based on deep convolutional neural network. Model is based on few layers like convolutional layer, dropout, fully connected layer and features are extracted from images. Using convolution filter represents the features of interests. Back-propagation algorithm can be embedded in the framework to reduce the errors. Activation function is introduced – ReLu (rectified linear unit) to make negative values zero. If any overfitting of the data is observed then dropout layer can be used to reduce the same.

An interesting topic “micro expression” in the emotion recognition is presented by the authors [15] using deep learning architecture. Authors have used FEREC-2013 database to extract the different features to identify the micro expressions. Model is based on cross-entropy loss function and uses “Adam optimizer”. This optimizer helps training and providing best results and also reduces the losses. To handle the overfitting of data, dropout layer can be introduced which reduces the overfitting of the model. With the confusion matrix authors have shown different emotions and their scores obtained. Authors [16] have proposed an LDL- label-distribution learning model with conditional probability function to reduce the entropy. A framework EDL-LBCNN is implemented using CNN on the s-JAFFE dataset. Basic emotions like happiness, sadness, fear, anger, surprise and disgust are accurately evaluated. Framework consists of 04 convolutional layers and single LBC (local binary convolutional layer. Initially framework extracts the features from grayscale images. Once the features are extracted next phase is to concatenate them and pass to the next phase. LBC (local binary convolutional layer) consists of few trainable parameters and fixed size of filters. Output of this layer is generation of feature map and in the next layer i.e. FC layer (fully connected layer) concatenates the feature maps generated by CNN and LBC layer. Authors [17, 18, 21] have used LSTM and CNN to learn speech emotion features. Interesting approach proposed by the authors is *combining facial and speech expression*. Advantage of this **fusion approach** is, it gives information of audiovisual features. High levels of features are extracted using deep neural networks. Audio features are extracted using *pyAudioAnalysis* an open-source library. Local features of data and global training features can be extracted using *CNN – convolutional neural network*. A multi-modal fusion is carried out on different features, also known as early-fusion or feature-level fusion. Model also consists of decision level fusion known as late-fusion. There is score level fusion implemented in the model to calculate the classification scores. Authors have combined happy and excite category into excite and works on remaining basic emotion like excite, sad and neutral. Authors [20] have implemented a pose-guide estimation model where features are extracted using few methods like pyramid histogram orientation gradient method (PHOG), edge histogram descriptor (EHD) and local binary pattern (LBP). Canny detector is introduced to extract the edge of the image. LBP helps in converting image into integer array. Authors have

evaluated their proposed model on the 04 datasets – CK+, JAFFE, CASIA and AR dataset. Model works on 03 steps - target pose estimation, template generation, and target matched.

III. PPROPOSED METHODOLOGY

‘Regularization’ is a technique that can be used in neural network to reduce the complexity of the proposed model. Once regularization is successfully implemented it helps generalizing the *new* data in efficient manner. The reason behind this approach is model randomly ignores the ‘n’ neurons and focus on the remaining neurons only. It is obvious that, there would be a significant difference in considering all the neurons and considering few one. With the selected neurons, model can run effectively and produce the results with better accuracy. Therefore, one can say with the help of regularization technique such as ‘dropout’ it is possible to obtain the useful and effective networks.

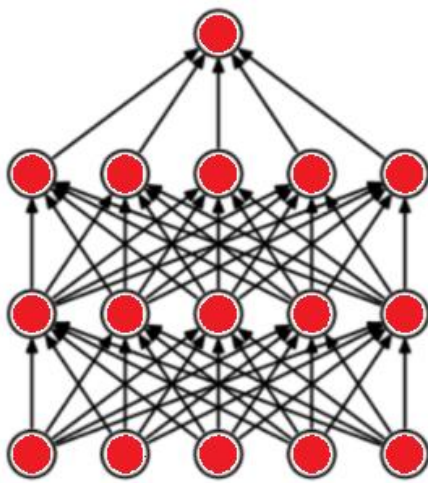


Fig. 1. Dropout regularization.

Drops hidden and visible units from the network, selects few from the remaining neurons as shown in the Fig. 1 and Fig. 2. Model temporarily removes the units as shown in the Fig. 2. One important point in this scenario is even if the weights of the neurons get changed it will not affect the skipped neurons. As its name indicates **‘dropout’** randomly drops few units from the given datasets. E.g. if the defined dropout rate is 0.2 then it randomly drops 20% units from the given collection. As discussed earlier, dropout layer can be added with the visible layer, hidden layer and other network layers. During processing the units/objects it creates combinations from all the defined layers. Model trains the dataset in iterations and randomly drops new units based on probability hyper parameter ‘P’. Many authors have termed this concept as formation of thinned network. Model further integrates these thinned networks to identify the *key properties*. However, this phase requires intensive processing the large datasets as there are many thinned networks. Objective behind adding the dropout layer in training the given data is to minimize the overfitting and achieve greater accuracy. Application of dropout method helps reducing the error rate significantly.

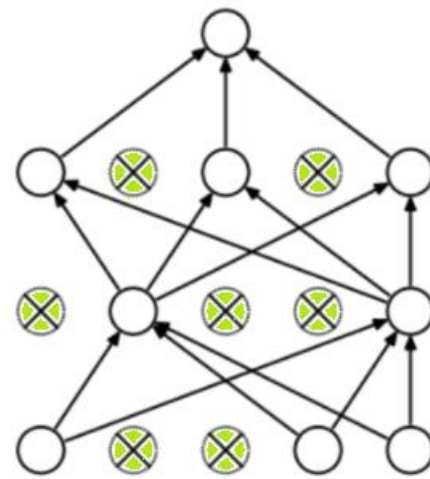


Fig. 2. Successful execution of dropout regularization.

A. Algorithmic Details

Idea behind application of dropout regularization is as follows-

- Consider there are ‘i’ inputs in the given network.
- If each of these ‘i’ units produces the output ‘o’ e.g. $O_1, O_2, O_3, \dots, O_n$
- Sum of these outputs $S_1 \dots S_n$
- Model should consider probabilities for these sums using individual scores like $P_1, P_2 \dots P_n$
- When model drops few neurons randomly and selects others from the remaining neurons, forms different sub-networks.
- Using the outputs and different sums calculated model can find possible probabilities of future sub-networks.

To implement the dropout function with a single layer, one must draw as many samples from a Bernoulli (binary) random variable as our layer has dimensions, where the random variable takes value 1 (keep) with probability $1-p$ and 0 (drop) with probability p . One easy way to implement this is to first draw samples from the uniform distribution $U [0, 1]$. Then, model can keep those nodes whose corresponding sample is greater than ‘p’ and drop the rest.

Table I explains the results obtained by applying the dropout layer with different dropout rates to the visible layer. Table I elaborates the different dropout rates i.e. 20%, 30%, 40% etc. applied to the visible layer and results obtained using the same. Model also rechecks the results obtained with a specific dropout rate i.e. 20% in the example as shown in the Table I. Similarly dropout rate can be applied to the hidden layer in the network. Table II elaborates the same with different dropout rates to measure the accuracy of the model. If the proposed model consists of ‘n’ iterations then each time new neurons (elements) will be dropped randomly with the dropout regularization technique. This technique is applied during training phase only.

IV. RESULTS AND DISCUSSION

In a network one can add ‘dropout layer’ to the hidden and visible layer etc. As shown in the Table I and II dropout regularization applied with different dropout rates, epochs at visible and hidden layer. Accuracy of the proposed model can be checked by using small as well as high dropout rates. It is obvious that, larger network will consist of ‘n’ neurons and with the dropout regularization technique one can get better performance. Table I and II shows the effective results obtained at different layers. It is also true that if the network is larger with ‘n’ neurons then training the neurons could be bit expensive.

Table I elaborates application of different dropout rates and epochs at the visible layer of the neural network. Overall the proposed model gives baseline efficiency with approx.85.55% (6.54%). One can observe as shown in Table I that batch size is very important along with epochs and dropout rates. Dropout rate 0.20 indicates 20% data can be randomly removed and remaining will be considered.

The first parameter dropout rate is the probability ‘p’ that a given unit (s) will drop out. In Table I, the different probabilities are used 0.2, 0.3, 0.4 and 0.5 which means roughly 20%, 30%, 40% and 50% units will drop out. The value 0.5 has been experimentally determined to be close to the optimal probability for a wide range of models, but feel free to experiment with other probabilities. To check the effectiveness, it is suggested to use variety of dropout rate in different cycles.

TABLE I. REGULARIZATION WITH DIFFERENT EPOCHS AND DROPOUT RATIOS

Dropout	Epochs	Batch size	Results mean and standard
0.20	300	16	Visible: 85.57% (6.18%)
0.30	400	18	Visible: 87.93% (7.57%)
0.40	500	20	Visible: 83.21% (8.26%)
0.20	300	16	Visible: 86% (7.98%)
0.20	300	16	Visible: 85.45% (7.81%)

It is a good approach to average the observations obtained by evaluating the sample data with different dropout rates, epochs and batch size. To improve the performance of neural networks especially in overfitting and regularizing the sample data one can integrate the dropout technique at different layers like hidden and visible layer. Fig. 5 and Fig. 7 give Bayes analysis for the sample data by weighting the settings for prior and posterior probabilities.

It is presumed that every neuron in the neural network has p% of dropout. Many researchers treat dropout technique as an ensemble method. Application of dropout technique at hidden and visible layers helps approximate the sample data and proved that it is computationally very cheap than other approaches.

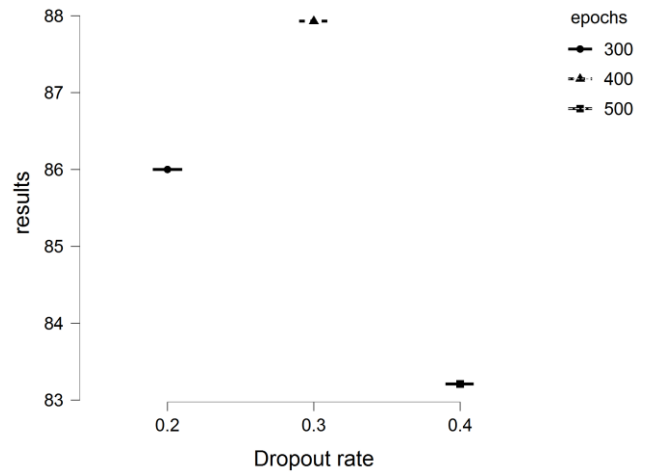


Fig. 3. Flexplot for the visible layer.

Fig. 3 Flexplot gives visual representation for the results obtained with different dropout rates.

A. Bayesian Correlation Pairwise Plots

Dropout rate - epochs

Scatterplot

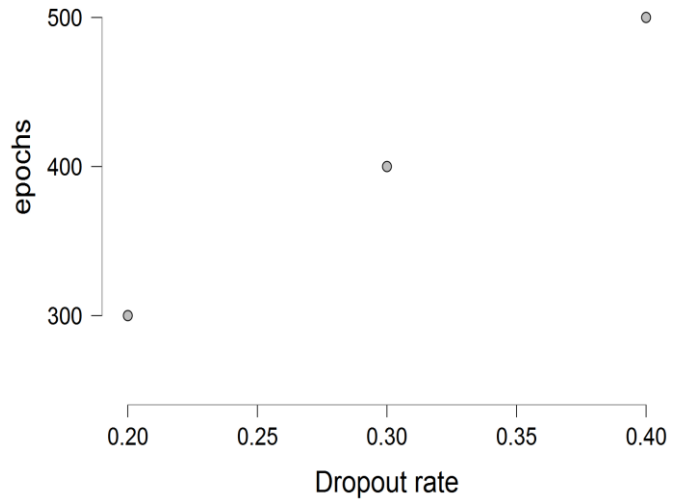


Fig. 4. Dropout rate v/s epochs.

Above Fig. 4 gives visual analysis for different dropout rates with ‘n’ epochs. Both the factors are very important if there are large distributions.

TABLE II. BAYESIAN PEARSON CORRELATIONS

	n	Pearson's r	BF ₁₀
Dropout rate-epochs	3	1.000	1.505
Dropout rate- batch size	3	1.000	1.505
Epochs- batch size	3	1.000	1.505

BF_{10} indicates bayes factor in favor of H_1 over H_0 . Bayesian Pearson correlation helps measuring strength of linear relationship between different variables. Table II gives sample observation for different approaches i.e. dropout rate vs epochs, dropout rate vs batch size, and epochs vs batch size as shown in Table III. The important factor in this observation is BF_{10} which may give different values for 'n' samples. Based on this factor one can categorize observed samples into *anecdotal evidence*, *moderate* and *strong evidences*.

TABLE III. REGULARIZATION WITH DIFFERENT EPOCHS AND DROPOUT RATIO (HIDDEN LAYER)

Dropout	Epochs	Batch size	Results mean and standard
0.20	300	16	83.62% (10.73%) performance degraded on hidden layers
0.30	400	18	Hidden: 83.71% (7.98%)
0.40	500	20	Hidden: 85.17% (5.81%)
0.20	300	16	Hidden: 84.12% (6.79%)
0.20	300	16	Hidden: 84.12% (6.79%)

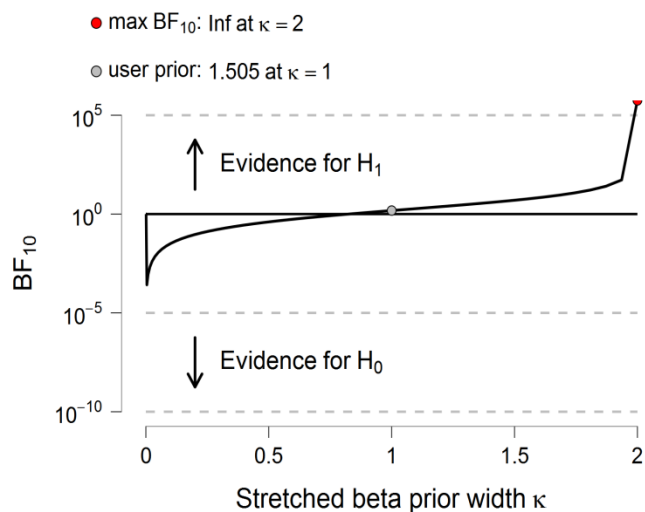


Fig. 5. Bayes factor robustness check.

Bayes factor robustness check is used for the wide range of prior distribution. It helps researchers to distinguish the obtained results in different segments like weak, moderate and strong evidence. Each iteration with different batch size, epochs and dropout rate one can get a new thinned neural network. For the large sample data proposed model with the dropout mechanism can have low classification errors. Even if the network is larger with the help of dropout technique one can reduce overfitting as shown in Fig. 6 to 8.

Dropout rate - batch size

Scatterplot

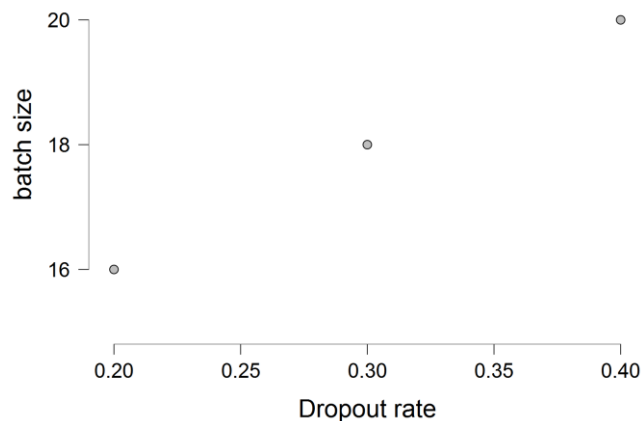


Fig. 6. Dropout rate v/s batch size.

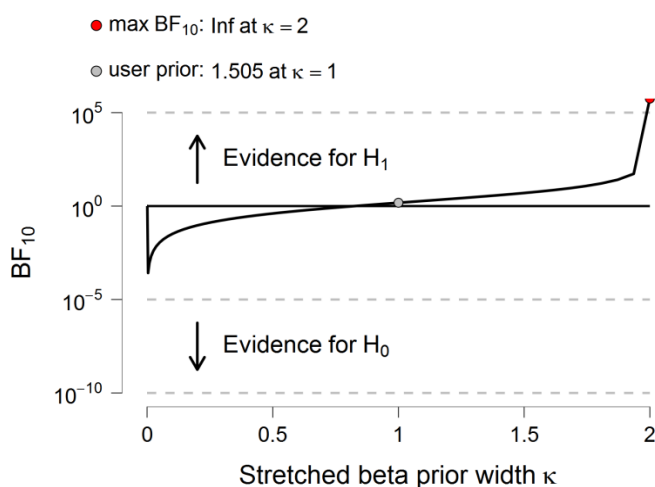


Fig. 7. Bayes factor robustness check for batch sizes.

Linear Modeling

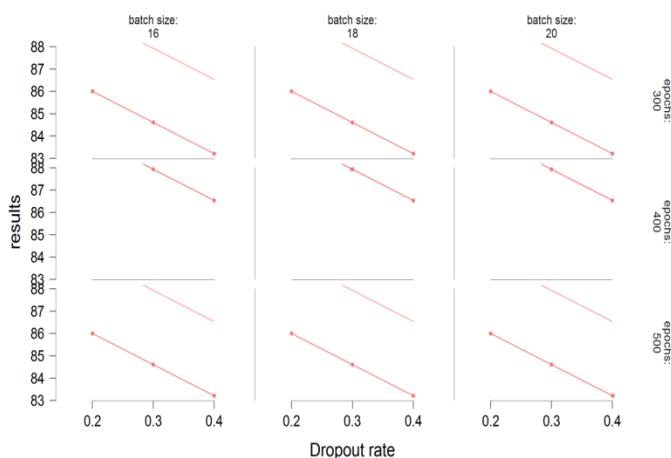


Fig. 8. Plot of the statistical model.

V. CONCLUSION

Although dropout is a potent tool, it has certain downsides. A dropout network may take 2-3 times longer to train than a normal network. Finding a regularizer virtually comparable to a dropout layer is one method to reap the benefits of dropout without slowing down training. This regularizer is a modified variant of L2 regularization for linear regression. An analogous regularizer for more complex models has yet to be discovered until that time when dropout drops out. One can propose a model with enhanced dropout mechanism i.e., having an effective method for randomly dropping neurons. Future work may include integrating the complex data from image, video etc. and improving the generalization of new data. Results can be improved with the hybrid approach like combining L2 regularization and dropout regularization techniques. If the network is larger and consists of maximum neuron then future research work can be directed to reduce the time complexity required for dropping and selecting the neurons. With the application of dropout technique every unit can act independently and this approach helps breaking co-adaptation in the neural networks.

ACKNOWLEDGMENT

This work is supervised by Dr. A. B. Nandgaokar, Department of Electronics and Telecommunications, Dr. Babasaheb Ambedkar Technological University, Lonere (India). Author thanks him for the continuous help and guidance provided to complete this research work. Author also thanks the Director of the Institution, Head of the Department and other technical staff of the University for their help and providing necessary infrastructure.

REFERENCES

- [1] Mohammadpour, Mostafa., Hossein, Khaliliardali., Seyyed, Mohammad, R, Hashemi., and Mohammad, AlyanNezhadi, "Facial emotion recognition using deep convolutional networks" In IEEE 4th international conference on knowledge-based engineering and innovation (KB EI), IEEE, Tehran, Iran, 2017, pp. 0017-0021.
- [2] Wang, Lei., Yiwei, Song., Jingqiang, Chen., Guozi, Sun., and Huakang, Li., "Emotion Recognition using Sequence Mining", In IEEE International Conference on Progress in Informatics and Computing (PIC), IEEE, Suzhou, China, 2018, pp. 129-133.
- [3] Chen, Hu., Ye, Gu., Fei, Wang., and Weihua, Sheng., "Facial expression recognition and positive emotion incentive system for human-robot interaction", In 13th World Congress on Intelligent Control and Automation (WCICA), IEEE, Changsha, China, 2018, pp. 407-412.
- [4] Kattubadi, Intihyaz, Bahsa., and Rama, Murthy, Garimella, "Emotion Classification: Novel Deep Learning Architectures", In 5th International Conference on Advanced Computing and Communication Systems (ICACCS), IEEE, Coimbatore, India, 2019, pp.285-290.
- [5] Zadeh, Milad, Mohammad., Taghi, Maryam, Imani., and Babak, Majidi, "Fast facial emotion recognition using convolutional neural networks and Gabor filters", In 5th Conference on Knowledge Based Engineering and Innovation (KB EI), IEEE, Tehran, Iran, 2019, pp.577-581.
- [6] Wentao, Hua., Fei, Dai., Liya, Huang., Jian, Xiong., and Guan, Gui, "HERO: Human emotions recognition for realizing intelligent Internet of Things", In IEEE Access vol. 7, 2019, pp. 24321-24332.
- [7] Harshitha, S., Sangeetha, N., Shirly, Asenath., and Abraham, Chandy, D., "Human facial expression recognition using deep learning technique", In 2nd International Conference on Signal Processing and Communication (ICSPC), IEEE, Coimbatore, India, 2019, pp. 339-342.
- [8] Chen, Min., and Yixue, Hao., "Label-less learning for emotion cognition", In IEEE transactions on neural networks and learning systems vol. 31.no.7, China, 2019, pp.2430-2440.
- [9] Yokoo, Kentaro., Masahiko, Atsumi., Kei, Tanaka., Haoqing, Wang and Lin, Meng., "Deep learning based emotion recognition IOT system", In International Conference on Advanced Mechatronic Systems (ICAMechS), IEEE, Hanoi, Vietnam, 2020, pp. 203-207.
- [10] Kondaveeti, Hari, Kishan., and Mogili, Vishal, Goud, "Emotion Detection using Deep Facial Features", In IEEE International Conference on Advent Trends in Multidisciplinary Research and Innovation (ICATMRI),IEEE, Buldhana, India,2020, pp.1-8.
- [11] Begaj, Sabrina., Ali, Osman, Topal., and Maaruf, Ali, "Emotion Recognition Based on Facial Expressions Using Convolutional Neural Network (CNN), In International Conference on Computing, Networking, Telecommunications & Engineering Sciences Applications (CoNTESA), IEEE, Tirana, Albania, 2020, pp.58-63.
- [12] Zhang, Su., and Cuntai, Guan, "Emotion recognition with refined labels for deep learning", In 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), IEEE, Montreal, QC, Canada, 2020, pp. 108-111.
- [13] Jaiswal, Akriti., A, Krishnama, Raju., and Suman, Deb, "Facial emotion detection using deep learning", In International Conference for Emerging Technology (INCET), IEEE, Belgaum, India, 2020, pp.1-5.
- [14] Pranav, E., Suraj, Kamal., Sathesh, Chandran, C., and Supriya, M, H, "Facial emotion recognition using deep convolutional neural network", In 6th International conference on advanced computing and communication Systems (ICACCS), IEEE, Coimbatore, India, 2020, pp.317-320.
- [15] Yadahalli, Srushti, S., Shambhavi, Rege., and Sukanya, Kulkarni, "Facial Micro Expression Detection Using Deep Learning Architecture", In International Conference on Smart Electronics and Communication (ICOSEC), IEEE, Trichy, India, 2020, pp.167-171.
- [16] Almomallad, Abeer., and Victor, Sanchez, "Human emotion distribution learning from face images using CNN and LBC features", In 8th International Workshop on Biometrics and Forensics (IWBF), IEEE, Porto, Portugal, 2020, pp.1-6.
- [17] Cai, Linqin., Jiangong, Dong., and Min, Wei, "Multi-modal emotion recognition from speech and facial expression based on deep learning", In Chinese Automation Congress (CAC), IEEE, Shanghai, China ,2020, pp.5726-5729.
- [18] Choi, Dong, Yoon., and Byung, Cheol, Song, "Semi-supervised learning for continuous emotion recognition based on metric learning", IEEE Access vol. 8, 2020, pp. 113443-113455.
- [19] L. Chen, Y. Ouyang, Y. Zeng and Y. Li, "Dynamic facial expression recognition model based on BiLSTM-Attention" In 15th International Conference on Computer Science & Education (ICCSE), IEEE, 2020, pp. 828-832
- [20] Liu, Jun., Yanjun Feng., and Hongxia Wang., "Facial expression recognition using pose-guided face alignment and discriminative features based on deep learning", In IEEE Access vol. 9, 2021, pp. 69267-69277.
- [21] Xiaoye Qu , Zhikang Zou; Xinxing Su Pan Zhou Wei Wei Shiping Wen and Dapeng Wu, "Attend to where and when: cascaded attention network for facial expression recognition" In IEEE Transactions on Emerging Topics in Computational Intelligence, vol. 6, no.3, 2021, pp. 580-592.
- [22] Yaoguang Ye , Yongqi Pan , Yang Liang and Jiahui Pan, "A cascaded spatiotemporal attention network for dynamic facial expression recognition", In Applied Intelligence, 2022, pp.1-14.
- [23] Subbarao, M. Venkata, Sudheer Kumar Terlapu, and Paladuga Satish Rama Chowdary, "Emotion Recognition using BiLSTM Classifier", In 2022 International Conference on Computing, Communication and Power Technology (IC3P) IEEE, Visakhapatnam, India, 2022, pp. 195-198.
- [24] Hartini, Sri, Zuherman Rustam, and Rahmat Hidayat. "Designing Hybrid CNN-SVM Model for COVID-19 Classification Based on X-ray Images Using LGBM Feature Selection." In International Journal on Advanced Science, Engineering and Information Technology (IJASEIT), vol.12. no.5, 2022, pp.1895-1906.

- [25] S. J. Prashantha and H.N. Prakash,"Two-Stage Approach of Hierarchical Deep Feature Representation and Fusion Frameworks for Brain Image Analysis" In International Journal on Advanced Science, Engineering and Information Technology (IJASEIT),vol.12 no.4, 2022, pp.1372-1378.
- [26] S. Kiruthika Devi and Subalalitha CN,"Intelligent Deep Learning Empowered Text Detection Model from Natural Scene Images" In International Journal on Advanced Science, Engineering and Information Technology (IJASEIT), vol.12 no.3, 2022, pp. 1263-1268.
- [27] Putu Arya Dharmaadi, Deden Witarasyah,Putu Agung Bayupati and Gusti Made Arya Sasmita,"Face Recognition Application Based on Convolutional Neural Network for Searching Someone's Photo on External Storage" In International Journal on Advanced Science, Engineering and Information Technology (IJASEIT) vol.12 no.3,2022, pp.1222-1228

Comparison Review on Brain Tumor Classification and Segmentation using Convolutional Neural Network (CNN) and Capsule Network

Nurul Fatihah Binti Ali¹, Siti Salasiah Mokri², Syahirah Abd Halim³, Noraishikin Zulkarnain⁴, Ashrani Aizuddin Abd Rahni⁵, Seri Mastura Mustaza⁶

Dept. of Electrical, Electronic and Systems Engineering-Faculty of Engineering and Built Environment,
Universiti Kebangsaan Malaysia, UKM Bangi,
Selangor, Malaysia

Abstract—Malignant brain glioma is considered as one of the deadliest cancer diseases that has a higher fatality rate than the survival rate. In terms of brain glioma imaging and diagnosis, the processes of detection and segmentation are manually done by the experts. However, with the advancement of artificial intelligence, the implementation of these tasks using deep learning provides an efficient solution to the management of brain glioma diagnosis and patient treatment. Deep learning networks are responsible for detecting, segmenting, and interpreting the tumors with high accuracy and repeatability so that the appropriate treatment planning can be offered to the patient. This paper presents a comparison review between two state of the art deep learning networks namely convolutional neural network and capsule network in performing brain glioma classification and segmentation tasks. The performance of each published method is discussed along with their advantages and disadvantages. Next, the related constraints in both networks are outlined and highlighted for future research.

Keywords—Deep learning; convolution neural network (CNN); capsule network; segmentation; classification; brain glioma

I. INTRODUCTION

Brain gliomas are the tumors that grow from the glial cells that support the brain and the spinal cord. There are three types of glial cells involved in the evolution of brain cancers: astrocytes in astrocytoma or glioblastoma tumors, oligodendrocytes in oligodendrogliomas, and ependymal cells in ependymomas tumors. Astrocytomas are mostly found in the cerebrum and cerebellum and can be in the form of slowly growing or fast-growing tumor. Slowly growing tumor is termed as low-grade glioma while fast growing tumor is called high grade glioma. The second type is oligodendroglioma. It is a rare type of brain tumor that attaches to the cerebrum. Lastly, is the ependymomas type that can occur in any section of the brain or spinal cord, most commonly near the cerebellum and lower spinal cord.

Brain gliomas are classified into two types: benign tumor and malignant tumor with cancerous cells [1,2,3]. These brain gliomas are also graded based on the staging endorsed by The World Health Organization. Low-grade gliomas (LGGs) are classified as WHO Grade II infiltrative brain tumor that typically appear solid and non-enhancing on magnetic resonance imaging (MRI) scans [3,4,5,6], in which the patient

may be treated with a watch-and-wait approach. High-grade gliomas (HGG), which can resemble LGG on an MRI scan, are typically classified as WHO Grade III or Grade V tumor. The mistakenly interpreted HGG as LGG causes the patient to suffer due to the watch-and-wait strategy although HGGs are extremely aggressive and advance very quickly.

Doctors may use a variety of methods to determine the type and the staging of the primary brain tumor and metastatic brain tumor. It is common that Magnetic Resonance Imaging (MRI) is used in the early stage of brain diagnosis to capture the presence of brain tumor and the necessary biopsy or surgery to obtain tissue sample in identifying the type of the tumor. Magnetic Resonance Imaging (MRI) employs magnetic fields to create a detailed image of the body with the assistance of a special dye known as a contrast medium (gadolinium) administered via intravenous, pill, or liquid to swallow. The dye enhances the MRI image by giving a more accurate depiction of the tumor. With the aid of the diffusion weighted imaging and perfusion imaging methods used in MRIs, soft tissue, joint, and organ abnormalities can be visualized in many different areas of the body, including the brain and joints.

Treatment of a brain tumor requires a multidisciplinary approach that includes surgery, radiotherapy, and chemotherapy. A neurosurgeon will remove the tumor tissue and have it examined to determine the type. To reduce intracranial pressure and relieve symptoms, any excess cerebrospinal fluid will be drained. If the tumor cannot be removed because of its location or if there is a high risk of complications after surgery, radiotherapy and chemotherapy will be used. Radiotherapy and chemotherapy are both commonly used to kill cancer cells by using high-energy x-rays or anti-cancer drugs [7]. Every person can have a different prognosis. The likelihood of survival is influenced by the type of tumor, the severity of the disease, the size and location of the tumor, the presence or absence of metastases, the tumor's response to treatment, the patient's age and general health, and the patient's tolerance for drugs, procedures, or therapies.

Deep Learning (DL), a subfield of Artificial Intelligence (AI), is one of the techniques used for analyzing raw data and extracting relationships in data sets, specifically in the

diagnosis and treatment planning of brain tumors. Deep Learning (DL) automatically learns representations and features from raw MRI image to perform the classification and segmentation of brain glioma. This approach overcomes the issue of manually computing and selecting relevant attributes from the images [8]. Since years ago, the deep learning method has showed its effectiveness to solve many issues and its application in biomedical imaging is outstandingly significant including in MRI brain glioma detection and analysis. As per current, the diagnostic, prognostic, and treatment evaluation of the brain glioma are manually handled. However, manual interpretation has its own limitations, facing the risk of ineffective patient management. Thus, deep learning approach will aid the clinical experts in making wise decisions regarding urgent clinical needs and the necessary treatments [9].

This paper discusses the previously proposed deep learning architectures utilized in classifying and segmenting the brain glioma in MRI images. We focus on CNN architecture and capsule network architecture [10], discussing their performance in terms of the advantages, and disadvantages as well as the prospective works.

II. LITERATURE REVIEW

A. Brain Glioma Classification

Various deep learning methods have been used in previous studies to classify brain glioma. The classification tasks can be implemented into two stages. The first stage is to classify the tumor into normal (benign) and abnormal (malignant) cases. The second stage is to further classify the malignant cases into the specific cancer type. The classification task is a difficult and challenging problem due to the inhomogeneous intensity, atypical shape of the tumors and tissue structures, and different types of noise in the MRI images. The designation of the cancer type is also equivalently difficult as the glioma tumors have similar patterns amongst each other.

Early detection and classification of brain tumors are critical in determining the most appropriate treatment for the patient. Computer aided diagnosis (CAD) based on deep learning is important in assisting the experts and clinicians with more detailed and consistent interpretation, as well as estimating the survival period of brain glioma patients [11] where the image features are extracted from the MRI image through a self-learning approach. Image processing for tumor identification begins with pre-planning, image division, extraction, and gathering. The image will then be automatically analyzed, which improves the identification of the brain structures [12].

The first deep learning algorithm used in image classification and segmentation is CNN. To improve the algorithms' accuracy, CNN has undergone several upgrades, including an adaptation of LeNet with changes to the arrangement and number of layers. Additionally, medial residual block (MidResBlock) classifier, which is based on a network of LeNet, is created by adding information about middle convolution layers to the output of each layer [13].

Another type of CNN is called Residual Neural Network (RNN) enables extremely deep training by using a shortcut

connection to skip one or more layers and by utilizing data augmentation to increase the dataset size and improve the accuracy [14]. The VGG16 high performance network is used in studies that use the fuzzy c-mean algorithm for segmentation before moving on to classify the segmented images. Each image is processed through the input layer, convolution layer, max pooling layer, drop out layer, fully connected layer, soft max function, and classification layer.

To capture more specific characteristics without incurring additional computational costs when using large filters in the convolution layer, dilation is implemented in CNN [15]. Another hyper parameter that is used along with this architecture is the dilation rate, which is higher on the outer layers for learning coarse features and lower on the inner layers for focusing on finer features to achieve the best results.

Deep CNN can extract multiple levels of features from image-based data and reveals distinguishing features between classes for classification problems [16,17]. While most studies focus on MRI or Computed Tomography (CT) images, histopathological images have a high potential in the background. To get the best results from this methodology, two factors must be considered: preprocessing and CNN architecture. Patch extraction, abnormal patch detection, intensity, contrast enhancement, and background removal are all part of the preprocessing process, while CNN models are trained to extract features from images, match features with clinical data, and predict relevant brain tumor types using linear or nonlinear methods [17]. Although CNN has been widely utilized to aid human inspection in tumor classification, it has its own limitations that must be addressed.

Capsule Network is proposed to alleviate the limitations of CNN, with the focus on maintaining image resolution and improving classification accuracy. It begins with segmented tumor regions as input and consists of a single CapsNet convolutional layer [18]. The Capsule Network demonstrates that it tends to account for everything in the input image, including the background, resulting in less detail in the image when compared to other methods [19]. The highest accuracy is achieved by reducing the number of feature maps from 256 to 64. Capsule Network sensitivity also supports the ability to access tumor surrounding tissues without distracting it from the main target [20].

Bayesian CapsNet [21] has been developed to handle uncertainty that tends to be incorrect and should be handed to a human expert for review and confirmation. Using Bayesian may result in lower accuracy and the need to filter out predictions with uncertainty over 0.15 and 0.1 to achieve more than 50% accuracy, in addition to higher computational costs. To better understand how the brain uses spiking signals and local learning rules, spiking networks on neuromorphic devices are also being conducted on CapsNet. Image-processing technology performs better because of increased use of neuromorphic computing and fresh learning algorithms [22].

DensNet201 and Inception V-3, two multiple-level feature extraction and concatenation methods, were employed for the diagnosis of brain tumors before features were sent to Softmax

for classification as in [23]. For DensNet 201 and Inception, features were taken out of DensNet blocks and the pre-trained Inception V3 model, respectively. This approach combined scratch-base models, more layers, and data augmentation techniques. A different screening and classification strategy that has been demonstrated to be effective and efficient in classifying gliomas, meningiomas, and metastatic brain tumors that is the combination of DenseNet-LSTM (CNN) and

holistic 3D MRI [24]. The YOLOV2-Inception V3 is an improved version of the pre-trained Inception V3 model, which segments data based on Kaput entropy after features are extracted from the mixed layer of the Inception V3 model [25]. Table I presents recent deep learning-based classification techniques for brain tumors based on CNN and Capsule network including several other published methods [26, 27, 28, 29, 30, 31, 32, 33, 34].

TABLE I. RECENT METHODOLOGIES FOR BRAIN GLIOMA CLASSIFICATION USING CNN AND CAPSULE NETWORK

Ref.	Methods	Accuracy
19	Capsule Network	86.56%
20	Capsule Network	90.89%
21	Bayesian CapNet	73.6% (0.15) 73.9% (0.1)
23	Inception V-3 DensNet 201	99.34% 99.15%
14	Residual Network	99% (image level) 97% (patient level)
11	Fuzzy C-Means and VGG16	96.70%
22	Capsule Network	89.3%
18	Capsule Network	95.54%
24	DenseNet-LSTM	92.13%
15	Dilated CNN	97%
13	CNN- LeNet CNN-MidResBlock	90% % 95%
25	YOLOV2- Inception V3	99%
16	Deep CNN	95%
26	6 DCNN	93%
27	GAN-ResNet50	88%
28	18 DCNN	99%
29	22 DCNN	96%
30	AlexNet-LSVM	63.1%
31	FLSCBN	89%
32	G-ResNet	95%
33	MDCNN	96%
34	Deep CNN	96%

B. Brain Glioma Segmentation

Machine learning algorithms play a critical role in enabling efficient and accurate image segmentation in the field of medical imaging. Classical learning models are less accurate, require more samples for efficient classification, and have a simpler structure when compared to deep learning techniques. There is significant evidence that accurate segmentation of various tumor subregions can lay the groundwork for quantitative image analysis to predict patient overall survival. Brain tumor segmentation is important not only in the diagnosis of brain glioma, radiation therapy, and clinical surgical planning, but it also increases the chances of a brain glioma patient's survival. Several methodologies have

been used to facilitate the segmentation of brain glioma, which include:

1) *Cascaded neural network*: Cascaded Dual-Scale LinkNet is used to improve segmentation precision. The architecture consists of two networks, the first of which learns features of specific areas of the brain tumor, while the second focuses on border areas and learns more detail about features [35]. The Deep Cascaded Neural Network system consists of two steps: (1) TLN subnet is used to localize the brain tumor and (2) ITCN subnet is used to classify the identified tumor regions into four tumor sub regions. This method can prevent pixel label imbalance where TLN merged different tumor

subregions into whole tumor and ITCN used the same number of image patches of each class to train [36].

2) *Convolutional neural network (CNN)*: CNN is the most widely used technique for image analysis and computer interventions, and its success rate is quite high when compared to other techniques. It is an improved version of the Artificial Neural Network (ANN) in terms of moving from manual decision and extraction to automated feature extraction from input data [37]. As a per-pixel classifier, multiple information from a small patch around each point is labelled, only standard intensity pre-processing is applied to the input data as scanner differences, and no post-processing is applied to the CNN output [38]. To overcome the lack of image specific adaptation and lack of generalizability to previously unseen object class, image-specific fine tuning is adapted in CNN model for unsupervised or supervised to significantly improve segmentation accuracy [39].

Features on the image will be extracted by building multiple layers of convolution layers in CNN. Smaller perceptive domain is used in front shallower convolution layer to allow learning some local features of image while larger perceptive domain in use in deeper convolution layer to learn more abstract features like size, location, and direction information. The CNN-based algorithm has three layers, namely convolution layer, pooling layer, and full connection layer. The preliminary product is obtained by multiplying the input image with the weights via the convolution layer. Filter matrix is chosen, for example by letting the step length to 1 and padding on the original image to obtain an input vector x . A pooling layer is being introduced to reduce abstract dimensionality and prevent over-fitting. The output is the classification probability gained from the regression in fully connected layers [39,40].

CNN has encoder and decoder parts, which allow discriminative comparison to be done easily and provide system extensibility [41]. The encoder part is responsible for extracting spatial information, after which the semantic map is inserted into the decoder part to produce the full-resolution probability map. UNet's modified architecture [42] includes three residual blocks in both the encoding and decoding phases, and each encoding employs batch normalization and the Parametric Rectified Linear Unit (PReLU). U-five Net's residual blocks are modified in both encoding and decoding while group normalization is used to add stability. DAU-Net with normalization is used to boost domain adaptation, which are then combined to produce U-Net (DAU-Net) [43].

3) *3D dense-UNets*: A triple segmentation network is created separately to predict the entire tumor, the tumor core, and the tumor enhancement. Three additional steps are included to maintain a high number of convolution layers while fitting into GPU memory. If a layer generated more feature maps than the initial number of convolution feature maps, the layer's total number of feature maps is reduced by one-fourth. A compression factor of 0.75 is applied to the total number of feature maps generated at the end of each dense

block. A bottle neck block is then utilized to connect the encoder and decoder parts of the network [44].

4) *Deep CNN*: Deep CNN consists of an architecture component for designing a network model and a learning algorithm to optimize the calculated parameters during the training phase. DCNN's architecture is made up of convolutional, pooling, Rectified Linear Unit (ReLU), and one fully connected layer. Each layer performs a simple computational operation, and each grid is only connected to a subset of the layers. Due to limited memory space and the number of parallel GPUs, higher volume data is difficult to handle [45].

Other architectures can be adapted to segment glioma, meningioma, and pituitary tumor tissues. DCNN ResNet50 employs skip connections to avoid gradient degradation when training a deeper network [14]. A combination of HCNN and CRF-RRNN models was proposed previously, in which the HCNN generates image slices at mixed scales to better leverage location while the CRF-RRNN takes the output and produces a segmentation based on the slides input into HCNN [45,46]. The FCNN is being improved by adding a Dense Micro-Block Difference feature to help with spatial consistency and utilizing Fisher vector encoding to texture rotation and scale [47], while 3D CNN focuses on intensity inhomogeneity by using N3T-spline to reduce noise and intensity in 3D scans and a T-spline for smoothing the output [48]. The Hourglass Network is an encoder-decoder with several residual blocks that is modified with five down sampling layers to perform better and be less computationally intensive [49]. Previous research has also proposed an ensemble network based on Model Cascade (MC) Net and One-Pass Multi-Task (OM) Net. Two MC-Nets are modified, one to improve encoder-decoder feature map coverage and the other to increase the effectively extract semantic feature at different resolutions. To create a more detailed model, an additional residual block is added to the original ON-Net [50].

5) *Region based CNN*: The region proposal network (RPN) is created by adding an extra convolutional layer that outputs the objectness score at image locations and the bounds of the region of interest (ROI). Using RCNN, it is possible to train a classifier on a smaller dataset and create bounding boxes with variable lengths. RPN will check the object's location, adjust to fit the dimension, use Region of Interest (ROI) pooling, and extract the features. The information is then used by RCNN to classify the content [51].

Before utilizing 3D volumetric CNN to fully leverage the 3D spatial contextual information of volumetric data, the RCNN model is applied to the largest area of tumor for tumor grading [52]. The 2D Mask R-CNN-based method yielded 0.935 (sensitivity), 0.972 (specificity), and 0.963 (accuracy), while the 3DConvNet method yielded 0.947 (sensitivity), 0.968 (specificity), and 0.971 (accuracy).

6) *Deep convolution neural network fusion support vector machine (DCNN-F-SVM)*: The traditional segmentation method involves training a suitable classifier on a training set before setting it up for verification. In contrast, this model comprises of three stages: (1) preprocessing, feature

extraction, and CNN and SVM training (2) running the final segmentation result through testing (3) utilizing the CNN-SVM cascade classifier [53]. These three stages are divided by DCNN-F-SVM. To obtain the mapping from the image space to the tumor label domain, DCNN is trained in the first stage. The second stage involves feeding the integrated SVM classifier with the test images along with the labelled output of the DCNN. In the third stage, a deep classifier that has more layers is trained by iteratively connecting the DCNN and the integrated SVM classifier.

7) *Capsule network (CapsNet)*: The concept of a capsule network was first proposed by Sabour et al. [54] to address a significant CNN shortcoming. The Capsule Network (CapsNet) extends the functionality of the conventional CNN by adding a new layer referred to as capsule layers, where each component is a capsule represented by a vector. It should represent a feature as well as the feature's characteristics, such as its location, texture, and deformation. The attitude matrix, W and the original input vector, U are processed in the capsule network to produce the final input vector, U . The final input vector, U is then multiplied by the appropriate weight, c and added to obtain the vector, s . Finally, a nonlinear function converts s into the final vector, v to enable the transition between the underlying and high-level feature. If the position of the high-level feature being pushed out by different underlying features points roughly in the same direction, the object has a high probability of existence [33].

One disadvantage of dynamic routing is that it can only be implemented in a fully connected manner. DeepCaps has been introduced to go deep into the architecture by skipping connections within a capsule cell that allow for good gradient flow in back propagation. Meanwhile, 3D convolution is used to generate votes from the capsule tensor to help route a localized group of capsules to a higher level. This combination allows architecture to go deeper while requiring less computational complexity [55]. According to the study, CapsNet performs better with a limited amount of training data and is suitable for detection or segmentation due to its high performance under class-imbalance for typical biomedical image database constraints [56].

8) *Segmentation capsule network (SegCaps)*: SegCaps is a modification to the original dynamic routing algorithm that makes it act locally when routing children capsules to parent capsules. It also allows the capsules within the same capsule type to share transformation matrices. This reduces the

memory and parameter burden and enables it to operate on large image sizes ranging from 32 x 32 pixels to 512 x 512 pixels. To compensate for information loss, the concept of deep convolutional-deconvolutional has been introduced for pixel level predictions of object labels. Finally, the masked reconstruction of the target class is extended as a regularization strategy for the segmentation problem [57,58,60]. The capsule segmentation task consists of [59]:

a) *Primary capsule*: Graphics inverse processes include convolution, reshape function or squash function and input image is fed into a couple of convolution layers.

b) *Higher layer capsule*: Due to the ability to trace the activation's path, the hierarchy of the parts can be easily sorted out. Additionally, it confirms the earlier prediction.

c) *Loss calculation*: Once the decision has been made, the classification takes place to determine whether the decision is correct or close to perfect.

Automatic glioma segmentation in brain MRI images had been conducted using CapsNet [61] by training the network into two steps (80% is used for training and 20% used as the validation data). Since capsule network has the capability to generalize novel viewpoint, it learns the spatial relationship between features using dynamic routing of capsules. The two-step training method resulted in about 3% improvement in dice score on validation and uses fewer data for training and contains 95.4% less parameter. From there it can be concluded that SegCaps can overcome the problem of data limitation. Results have also shown that SegCaps has been successful in segmenting the enhancing tumor core area. SegCaps only stores feature map indices and uses them in the decoder to achieve good performance. In comparison to other architectures, it is more efficient in terms of memory and computation [62]. According to previous research, SegCaps can capture the contours of the tumor core better than the contours of the entire tumor [63]. SegCaps can segment a small tumor region of the tumor core, but it tends to ignore the inner areas of the tumor core. When it comes to whole brain tumors, SegCaps is unable to capture fine-grained details, particularly when the region is exceptionally small, and the boundaries are coarse. Despite its capabilities, the algorithm is much slower, with higher computational complexity and a longer execution time. Table II shows some of the most recent research results on brain tumor segmentation based on CNN and capsule network specifically for the complete tumor (whole tumor), the tumor core that consists of enhancing, non-enhancing and necrotic parts and on the enhanced tumor only.

TABLE II. RECENT METHODOLOGIES FOR BRAIN GLIOMA SEGMENTATION USING CNN AND CAPSULE NETWORK

Ref.	Methods	Result (Dice)			Standard Mean IoU
		Complete	Core	Enhance	
35	Cascade Dual-Scale LinkNet		80.03%		90.73%
36	Deep Cascaded Neural Network	89%	77%	80%	
44	3D Dense-UNet	92% (cross validation) 90% (20 held out cases) 90% (Brats 2018) 90% (Brats 2017) 85% (Clinical)	84% (cross validation) 84% (20 held out cases) 82% (Brats 2018) 80% (Brats 2017) 80% (Clinical)	80% (cross validation) 80% (20 held out cases) 80% (Brats 2018) 78% (Brats 2017) 77% (Clinical)	
45	Deep CNN	90%	85%	84%	
51	Region based CNN		91.14%		
52	3D CNN-2D Context	91.8%	88.3%	85.4%	
53	DCNN-F-SVM		90.10%		
38	2D CNN	86%	82%	81%	
14	ResNet50		99		
46	HCNN; CRF-RRNN		96.6		
47	FCNN		91		
39	PC-Net		86.29		
48	FCNN		NA		
49	Hourglass Net		92		
43	U-Net (DAU-Net)	91%			
50	MC-Net – OM-Net	90%			
42	U-Net	86.8%			
61	SegCaps			85.56%	
63	SegCaps			89.21% (whole tumor) 82.44% (Tumor core)	
	U-Net			99.78% (whole tumor) 99.88% (Tumor core)	

C. Dataset for Brain Tumor Analysis

Deep learning tasks necessitate a large dataset for training and validation. For brain tumor analysis, various datasets are publicly offered like BraTS. The images are mostly available in NifTI, DICOM, JPG, and PNG formats, and they are described in native (T1) and post-contrast T1-weighted (T1Gd), T2-weighted (T2), and T2 Fluid Attenuated Inversion Recovery (T2-FLAIR). Since the dataset is derived from MRI patient cases, each one contains 154 - 155 sliced human brain images, implying that for each case of subset, there are 155 images in that subset. This preview most likely using Brain Tumor Segmentation Challenge (BraTS 2013-2018). The Brain Tumor Dataset (Public and Private) and The Cancer Genome Atlas are two other common sources of datasets used for brain tumor analysis (TCGA).

III. DISCUSSION

Since the early stages of deep learning development, Convolution Neural Networks (CNN) have been utilized to meet current requirements and improve classification and segmentation of brain glioma. CNN is made up of several layers; the convolution layer extracts image features, the pooling layer checks whether the features are present or not regardless of their position, and the max-pooling problem. Then, the full connection layer will predict the output of the learning features from the previous two layers.

However, CNN has limitations that lead to results that are not accurate enough, such as complex calculation if a standard multi-layer perceptron (all layers are fully connected) is used because the image dimensions are too large. Other disadvantages include loss of information in the pooling layer, such as spatial resolution, and the need to over-train all possible angles to overcome it, which consumes more time

and resources. It also demonstrates the inability to identify complex field-of-view images such as overlaps, mutual masking, and different backgrounds. Many modifications to the CNN algorithm have been made, such as DCNN, U-Net, RCNN, FCNN, and others while improving the performance of CNN.

Capsule Network (CapsNet) is a new deep learning methodology for brain tumor classification and segmentation. CapsNet is made up of four layers: a convolution layer for extracting features, a primary layer for storing feature vectors using eight convolution operations, a digital layer for storing higher level features in the form of vectors, and an auxiliary layer for replacing capsules with their lengths. CapsNet reduces the number of connections by using capsule group neurons and fewer parameters. Even though fewer parameters are used, CapsNet's viewpoint invariance allows it to recognize images and objects efficiently regardless of the viewpoint from which they are observed. Due to the typical dataset constraints of medical images, this also leads to object detection and image segmentation being more helpful.

CapsNet can solve the effect of max pooling in CNN, as this unit only examines whether attributes and features are present or not regardless of their position. CapsNet can resolve this issue by capturing part-whole relationships that consider both the existence of the features and their orientation. Furthermore, max pooling causes translational invariance, which results in incorrectly sized, positioned, or ordered components of the correct image. CapsNet uses equivariance by considering the position, proportional and translational invariances to tackle invariance problem.

However, there are several things that need to be improved in CapsNet for future research, such as the uncertainty in testing on larger images and the inner loop in dynamic routing by agreement algorithm, which causes the training to slow down. CapsNet may have the weakness to accurately identify interested region in images with a more varied background and very close identical objects due to noise and crowding. Since CapsNet is still in its early stages, more changes and improvements can be made to the algorithm, particularly to the dynamic routing and convolution strides.

Furthermore, there is substantial evidence that accurate segmentation of various tumor sub-regions can provide the foundation for quantitative image analysis to predict patient overall survival. In the future, it may help to lay a more precise and accurate foundation to understand unique tumoral characteristics, as well as better insights into the quantitative and qualitative aspects of a patient's disease care. Imaging analysis may improve glioma treatment and management by determining tumor heterogeneity, more comprehensive identification of tumor genotype, cases of progression and pseudo progression, tumor grading, and survival prediction.

In general, computer-assisted diagnosis makes a significant contribution in terms of speed. Deep learning approaches benefit from adapted automatic feature acquisitions to reduce time spent on manual practice. With the introduction of GPUs, computation processes become much faster, and more data can be trained. The amount of training

data increases computational performance and accuracy as well.

IV. CONCLUSION

This paper reviews the performance between two neural network architectures to classify and segment the brain glioma, namely CNN and CapsNet. It is concluded that CapsNet can overcome data loss and has a higher level of image classification and segmentation accuracy than CNN, but it involves more parameter training and requires more training time. The CNN model outperforms CapsNet in terms of less memory and training time. This review may serve as a guide in establishing future deep learning-based methods to further improve the brain glioma classification and segmentation performance.

ACKNOWLEDGMENT

The authors would like to thank the Ministry of Higher Education Malaysia and Universiti Kebangsaan Malaysia for the University Research Grant: GUP-2019-023 to support this work.

REFERENCES

- [1] J. Amin, M. Sharif, M. Yasmin, and S. L. Fernandes. *A distinctive approach in brain tumor detection and classification using MRI*. Pattern Recognition Letters, 2017.
- [2] V. Rajinikanth, S. C. Satapathy, S. L. Fernandes, and S. Nachiappan. *Entropy based segmentation of tumor from brain MR images—a study with teaching learning based optimization*. Pattern Recognition Letters, 94: 87-95, 2017.
- [3] V. Rajinikanth, S. L. Fernandes, B. Bhushan, and N. R. Sunder. *Segmentation and analysis of brain tumor using Tsallis entropy and regularised level set*, in Proceedings of 2nd International Conference on Micro-Electronics, Electromagnetics and Telecommunications, 313-321, 2018.
- [4] T. Saba, S. T. F. Bokhari, M. Sharif, M. Yasmin, and M. Raza. *Fundus image classification methods for the detection of glaucoma: A review*. Microscopy research and technique, 81:1105-1121, 2018.
- [5] U. R. Acharya, S. L. Fernandes, J. E. WeiKoh, E. J. Ciaccio, M. K. M. Fabell, U. J. Tanik, V. Rajinikanth, and C. H. Yeong. *Automated Detection of Alzheimer's Disease Using Brain MRI Images—A Study with Various Feature Extraction Techniques*. Journal of Medical Systems, 43:202, 2019.
- [6] S. L. Fernandes, U. J. Tanik, V. Rajinikanth, and K. A. Karthik. *A reliable framework for accurate brain image examination and treatment planning based on early diagnosis support for clinicians*. Neural Computing and Applications, 1-12, 2019.
- [7] National Cancer Registry, National Cancer Institute, Ministry of Health Malaysia. *Malaysian Study on Cancer Survival (MyScan)*, 2018.
- [8] Segato A, Marzullo A, Calimeri F, De Momi E. *Artificial intelligence for brain diseases: A systematic review*. APL bioengineering 4.4: 041503, 2020.
- [9] Madeleine M. Shaver, Paul A. Kohanteb, Catherine Chiou, Michelle D. Bardis, Chanon Chantaduly, Daniela Bota, Christopher G. Filippi, Brent Weinberg, Jack Grinband, Daniel S. Chow and Peter D. Chang. *Optimizing Neuro-Oncology Imaging: A Review of Deep Learning Approaches for Glioma Imaging*. 2019. Cancers 11.6: 829, 2019.
- [10] Evgin Goceri. *CapsNet topology to classify tumors from brain images and comparative evaluation*. IET Image Processing 14.5: 882-889, 2020.
- [11] V Ramakrishna Sajjaa and Hemantha Kumar Kalluri. *Classification of Brain Tumors using Fuzzy C-means and VGG16*. Turkish Journal of Computer and Mathematics Education (TURCOMAT) 12.9: 2103-2113, 2021.

- [12] Nandita Goyal and Dr. Bharti Sharma. *Image Processing Techniques for Brain Tumor Identification*. IOP Conference Series: Materials Science and Engineering. Vol. 1022. No. 1. IOP Publishing, 2021.
- [13] Zahra SobhaniNia, Nader Karimi, Pejman Khadivi, Roshank Roshandel, Shadrokh Samavi. *Brain Tumor Classification Using Medial Residual Encoder Layers*. arXiv preprint arXiv:2011.00628, 2020.
- [14] Sarah Ali Abdelaiz Ismael, Ammar Mohammed, Hesham Hefny, *An enhanced deep learning approach for brain cancer MRI images classification using residual networks*. Artificial intelligence in medicine 102: 101779, 2020.
- [15] Sanjiban Sekhar Roy, Nishant Rodrigues, and Y-h. Taguchi. *Incremental Dilations Using CNN for Brain*. Applied Sciences 10.14: 4915, 2020.
- [16] Gökalp Çınareri, Bülent Gürsel Emiroğlu, Recep Sinan Arslan, Ahmet Haşim Yurtakal. *Brain Tumor Classification Using Deep Neural Network*. Advances in Science, Technology and Engineering Systems Journal 5.5: 765-769, 2020.
- [17] Amin Zadeh Shirazi, Eric Fornaciari, Mark D. McDonnell, Mahdi Yaghoobi, Yesenia Cevallos, Luis Tello-Oquendo, Deysi Inca and Guillermo A. Gomez. *The Application of Deep Convolutional Neural Networks to Brain Cancer Images: A Survey*. Journal of personalized medicine 10.4: 224, 2020.
- [18] M. Waqas Nadeem, M. A. Al Ghamdi, Muzammil Hussain, M. Adnan Khan, Khalid Masood Khan, Sultan H. Almotiri and Suhail Ashfaq Butt. *Brain Tumor Analysis Empowered with Deep Learning: A Review, Taxonomy, and Future Challenges*. Brain sciences 10.2: 118, 2020.
- [19] Parnian Afshar, Arash Mohammadi, and Konstantinos N. Plataniotis. *Brain Tumor Type Classification Via Capsule Networks*. 25th IEEE international conference on image processing (ICIP). IEEE, 2018.
- [20] Parnian Afshar, Konstantinos N. Plataniotis, Arash Mohammadi. *Capsule Networks for Brain Tumor Classification based on MRI Image and Course Tumor Boundaries*. ICASSP 2019-2019 IEEE international conference on acoustics, speech, and signal processing (ICASSP). IEEE, 2019.
- [21] Parnian Afshar, Student Member, IEEE, Arash Mohammadi, Senior Member, IEEE, and Konstantinos N. Plataniotis. *BayesCap: A Bayesian Approach to Brain Tumor Classification Using Capsule Networks*. IEEE Signal Processing Letters 27: 2024-2028, 2020.
- [22] Getty, N., Brettin, T., Jin, D., Stevens, R., Xia. *Deep Medical Image Analysis with Representation Learning and Neuromorphic Computing*. Interface Focus 11.1: 20190122, 2021.
- [23] Neelum Noreen, Sellappan Palaniappan, A.Qayyum, Iftikhar Ahmad, M.Imran, M.Shoaib. *A Deep Learning Model Based on Concatenation Approach for the Diagnosis of Brain Tumor*. IEEE Access 8: 55135-55144, 2020.
- [24] Yufan Zhou, Zheshuo Li, Hong Zhu, Changyou Chen, Mingchen Gao, Kai Xu, and Jinhui Xu. *Holistic Brain Tumor Screening and Classification Based on DenseNet and Recurrent Neural Network*. 4th International Workshop, BrainLes, MICCAI 2019.
- [25] Muhammad Irfan Sharif, Jian Ping Li, Javeria Amin2, Abida Sharif. *An improved framework for brain tumor analysis using MRI based on YOLOv2 and convolutional neural network*. Complex & Intelligent Systems 7: 2023-2036, 2021.
- [26] Muhammad, K., Khan, S, Ser, J.D. de Albuquerque, V.H.C. *Deep learning for multigrade brain tumor classification in smart healthcare systems: A prospective survey*. IEEE Transactions on Neural Networks and Learning Systems 32.2: 507-522, 2020.
- [27] Liu, S., Shah, Z., Sav, A., Russo, C., Berkovsky, S., Qian, Y., Coiera, E., Di Ieva, A. *Isocitrate Dehydrogenase (Idh) Status Prediction in Histopathology images of gliomas using deep learning*. Scientific reports 10.1: 1-11, 2020.
- [28] Alqudah, A.M., Alquraan, H., Qasmieh, I.A., Alqudah, A., Al-Sharu, W.. *Brain Tumor Classification Using Deep Learning Technique—A Comparison between Cropped, Uncropped, and Segmented Lesion Images with Different Sizes*. arXiv preprint arXiv:2001.08844, 2020.
- [29] Badza, M.M., Barjaktarović, M.C.. *Classification of Brain Tumors from MRI Images Using a Convolutional Neural Network*. Applied Sciences 10.6, 2020.
- [30] Hoseini F, Shahbahrami A, Bayat P.. *An Efficient Implementation of Deep Convolutional Neural Networks for MRI Segmentation*. Journal of digital imaging 31: 738-747, 2018.
- [31] Kalaiselvi, T., Padmapriya, T., Sriramakrishnan, P., Priyadarshini, V.. *Development of Automatic Glioma Brain Tumor Detection System Using Deep Convolutional Neural Networks*. International Journal of Imaging Systems and Technology 30.4: 926-938, 2020.
- [32] Liu, D., Liu, Y., Dong, L., G-ResNet: *Improved ResNet for brain tumor classification*. In International Conference on Neural Information Processing. 2019.
- [33] Hemanth, D.J., Anitha, J., Naaji, A., Geman, O., Popescu, D.E., Son, L.H., Hoang, L.. *A modified deep convolutional neural network for abnormal brain image classification*. IEEE Access 7: 4275-4283, 2018.
- [34] Yonekura, A., Kawanaka, H., Prasath, V.S., Aronow, B.J., Takase, H.. *Automatic Disease Stage Classification of Glioblastoma Multiforme Histopathological Images Using Deep Convolutional Neural Network*. Biomedical engineering letters 8.3: 321-327, 2018.
- [35] Zahra Sobhaninia, Safiyah Rezaei, Nader Karimi, Ali Emami, Shadrokh Samavi. *Brain Tumor Segmentation by Cascaded Deep Neural Network using Multi-Resolution Image Scales*. 2020 28th Iranian Conference on Electrical Engineering (ICEE). IEEE, 2020.
- [36] Shaoguo Cui, Lei Mao, Jingfeng Jiang, Chang Liu and Shuyu Xiong. *Automatic Semantic Segmentation of Brain Gliomas from MRI Images Using a Deep Cascaded Neural Network*. Journal of healthcare engineering, 2018.
- [37] Muhammad Yaqub, Jinchao Feng, M. Sultan Zia, Kaleem Arshid, Kebin Jia, Zaka Ur Rehman and Atif Mehmood. *State-of-the-Art CNN Optimizer for Brain Tumor Segmentation in Magnetic Resonance Images*. Brain Sciences 10.7: 427, 2020.
- [38] Darko Zikic, Yani Ioannou, Matthew Brown, and Antonio Criminisi. *Segmentation of Brain Tumor Tissues with Convolutional Neural Networks*. Proceedings MICCAI-BRATS 36.2014: 36-39, 2014.
- [39] Guotai Wang, Wenqi Li, Maria A. Zuluaga, Rosalind Pratt, Premal A. Patel, Michael Aertsen, Tom Doel, Anna L. David, Jan Deprest, Sébastien Ourselin, Tom Vercauteren. *Interactive Medical Image Segmentation Using Deep Learning With Image-Specific Fine Tuning*. IEEE transactions on medical imaging 37.7: 1562-1573, 2018.
- [40] Xuefeng Jiang, Yikun Wang, Wenbo Liu, Shuying Li, and Junrui Liu. *CapsNet, CNN, FCN: Comparative Performance Evaluation for Image Classification*. Int. J. Mach. Learn. Comput 9.6: 840-848, 2019.
- [41] Jeffrey D. Rudie, David A.Weiss, Rachit Saluja, Andreas M. Rauschecker, Jiancong Wang, Leo Sugrue, Spyridon Bakas, and John B. Colby. *Multi-Disease Segmentation of Gliomas and White Matter Hyperintensities in the BraTS Data Using a 3D Convolutional Neural Network*. Frontiers in Computational Neuroscience 13: 84, 2019.
- [42] Kermi, A., Mahmoudi, I., Khadir, M.T. *Deep convolutional neural networks using U-Net for automatic brain tumor segmentation in multimodal MRI volumes*. 4th International Workshop, BrainLes 2018, MICCAI 2018.
- [43] Dai, L., Li, T., Shu, H., Zhong, L., Shen, H., Zhu, H. *Automatic brain tumor segmentation with domain adaptation*. In International MICCAI Brainlesion Workshop, MICCAI 2018.
- [44] Chandan Ganesh Bangalore Yogananda, Bhavya R. Shah, Maryam Vejdani-Jahromi, Sahil S. Nalawade, Gowtham K. Murugesan, Frank F. Yu, Marco C. Pinho, Benjamin C. Wagner, Kyrre E. Emblem, Atle Bjørnerud, Baowei Fei, Ananth J. Madhuranthakam, and Joseph A. Maldjian. *A Fully Automated Deep Learning Network for Brain Tumor Segmentation* Tomography 6.2: 186-193, 2020.
- [45] Farnaz Hoseini, Asadollah Shahbahrami, Peyman Bayat. *An Efficient Implementation of Deep Convolutional Neural Network for MRI Segmentation*. Journal of digital imaging 31, 738-747, 2018.
- [46] Deng, W., Shi, Q., Wang, M., Zheng, B., Ning, N. *Deep Learning-Based HCNN and CRF-RRNN Model for Brain Tumor Segmentation*. IEEE Access 8: 26665-26675, 2020.
- [47] Deng, W., Shi, Q., Luo, K., Yang, Y., Ning, N. *Brain Tumor Segmentation Based On Improved Convolutional Neural Network in Combination with Non-Quantifiable Local Texture Feature*. Journal of medical systems 43: 1-9, 2019.

- [48] Kumar, G.A., Sridevi, P. *Intensity Inhomogeneity Correction for Magnetic Resonance Imaging of Automatic Brain Tumor Segmentation*. Microelectronics, Electromagnetics and Telecommunications: Proceedings of the Fourth ICMEET 2018.
- [49] Benson, E., Pound, M.P., French, A.P., Jackson, A.S., Pridmore, T.P. *Deep hourglass for brain tumor segmentation*. 2018. 4th International Workshop, BrainLes 2018, MICCAI 2018.
- [50] Zhou, C., Chen, S., Ding, C., Tao, . *Deep Learning Contextual and Attentive Information for Brain Tumor Segmentation*. 4th International Workshop, BrainLes 2018, MICCAI 2018.
- [51] H.N.T.K.Kaldera, S.R.Gunasekara, M.B.Dissanayake. *MRI based Glioma Segmentation Using Deep Learning algorithms*. 2019 International research conference on smart computing and systems engineering (SCSE). IEEE, 2019.
- [52] Ying Zhuge, Holly Ning, Peter Mathen, Jason Y. Cheng, Andra V. Krauze, Kevin Camphausen, Robert W. Miller. *Automated Glioma Grading on Conventional MRI Images Using Deep Convolutional Neural Networks*. Medical physics 47.7: 3044-3053, 2020.
- [53] Wentao Wu, Daning Li, Jiaoyang Du, Xiangyu Gao, Wen Gu, Fanfan Zhao, Xiaojie Feng, and Hong Yan, *An Intelligent Diagnosis Method of Brain MRI Tumor Segmentation Using Deep Convolutional Neural Network and SVM Algorithm*. Computational and Mathematical Methods in Medicine, 2020.
- [54] Sara Sabour, Nicholas Frosst , Geoffrey E.Hinton. *Dynamic Routing Between Capsules*. Advances in neural information processing systems 30, 2017.
- [55] Jathushan Rajasegaran, Vinoj Jayasundara, Sandaru Jayasekara, Hirunima Jayasekara, Suranga Seneviratne, Ranga Rodrigo. *DeepCaps: Going Deeper with Capsule Networks*. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019.
- [56] Amelia Jim'enez-S'anchez, Shadi Albarqouni, Diana Mateus. *Capsule Networks against Medical Imaging Data Challenges*. 7th Joint International Workshop, CVII-STENT 2018 and Third International Workshop, LABELS 2018, MICCAI 2018.
- [57] Rodney Lalonde, Ulas Bagci. *Capsules for Object Segmentation*. arXiv preprint arXiv:1804.04241, 2018.
- [58] Rodney LaLonde, Ziyue Xub, Ismail Irmakcic, Sanjay Jaind, Ulas Bagcia. *Capsules for Biomedical Image Segmentation*. 2021. Medical image analysis 68: 101889, 2021.
- [59] Zhankun Luo, Andres Jara, Wen Ou. *Image Segmentation using U-Net, DenseNet and CapsuleNet*. Transition 40: 584, 2019.
- [60] Floris Van Beers. *Capsule Networks with Intersection over Union Loss for Binary Image Segmentation*. ICPRAM. 2021.
- [61] M. Jalili Aziz, A. Amiri Tehrani zade, P. Farnia, M. Alimohamadi, B. Makkiabadi, A. Ahmadian, J. Alirezaie. *Accurate Automatic Glioma Segmentation in Brain MRI images Based on CapsNet*. 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). IEEE, 2021.
- [62] Caterina Camborata. *Capsule Networks: a new approach for brain imaging*. 2018.
- [63] Sander Speet. *Brain Tumor Segmentation Using a Capsule Network*. Diss. Tilburg University. 2020.

Fuzzy Reasoning based Reliability Fault Prediction of CNC Machine Tools

Jie Yu, Tiebin Wang*, Weidong Wang, Gege Zhao, Yue Yao

School of Mechanical and Resource Engineering, Wuzhou University, Guangxi, China

Abstract—CNC machine tools are the infrastructure of the manufacturing industry, and many fields cannot do without them. This paper studies the fault data of a series of CNC machine tools, and predicts the fault level based on the activity parameters of Gutenberg Richter curve and fuzzy information theory. Apply the Gutenberg Richter curve model to the reliability analysis of CNC machine tools, and use this model to fit the curves separately. Fit the activity parameters of each stage with curves, and the results show that the b value can reflect the fault activity frequency of CNC machine tools. Due to the correlation and fuzziness between system faults, it is more appropriate to use a fuzzy neural network with strong adaptability and good learning ability, which can easily adjust parameters, and can express a more complex, high-dimensional nonlinear system through fewer conditions. The use of fuzzy reasoning can link the nonlinear relationship between fault level, b-value, and N-value. Analyze the error between the predicted fault level and the original level, and the small error indicates that the model has good predictive ability. Applying this predictive ability to the reliability research of CNC machine tools will yield good results.

Keywords—CNC machine tools; reliability; fuzzy inference; fault prediction

I. PREFACE

The failure of the CNC machine tool is recorded in real time, and the surrounding environment, the quality of its own components, and the operation of the controller all have a certain impact. Once a CNC machine tool has a high-impact failure it can bring great losses, not only affects the processing process, but also consumes personnel to detect and repair and time. In the reliability study of CNC machine tools, the quality of each component is tested and predicted to reduce the possibility of failure and reduce the loss caused by failure. If you can better predict the failure of CNC machine tools, it will reduce a lot of unnecessary trouble and loss. There will be some connection between some subsystem failures of CNC machine tools, and the failure of one system may cause the failure of another system as well, a phenomenon called system fault correlation [1]. The fault generation may be spontaneous or may be caused by other system correlations, and because of this uncertain relationship, fuzzy reasoning is used to study the method. Currently, there are many ways of fault prediction, statistical prediction techniques, mathematical prediction techniques, intelligent prediction techniques, and information fusion techniques, and it is very important to study fault prediction techniques to improve the maintenance and security of equipment [2].

Gutenberg and Richter proposed the famous earthquake magnitude relationship $\lg N = a - bM$ by studying the activity characteristics of the California earthquake. This relationship is one of the important properties of seismic activity research and has been widely used in the study of earthquake related issues. In this relation, M is the magnitude, and N represents the number of earthquakes in the area with a magnitude greater than or equal to M in a certain period of time. The values of a and b are constants, a-values reflect the average level of seismic activity in the area, and b-values reflect the proportional relationship between large and small earthquakes. Jie Yu proposed to use Gutenberg–Richter (G-R) curve analysis method to analyze the relationship between the failure level and the occurrence frequency of a series of CNC lathes. This is a useful attempt in the reliability analysis method research of CNC lathes, and some conclusions have been drawn.

Zheng et al. [3] and Wang et al. [4] proposed to use fuzzy information theory to analyze the relationship between the precursor anomalous elements and seismic elements of earthquakes, and to take the b value of 0.65 as the anomaly indicator based on the fuzzy matrix of b values calculated by different parameters with corresponding magnitude and the fuzzy information inference results. Wang et al. [4] proposed to use the beta value of the frequency of the magnitude of the gamma distribution also as an earthquake prediction indicator. Fuzzy theory is also used for CNC machine tool fault diagnosis, and the fuzzy diagnosis model is easy and simple to use and widely applied through its strong structural knowledge expression and ability to handle incomplete information. Fuzzy neural networks have been widely used in various fields and fuzzy systems have good superiority in dealing with complex nonlinear systems.

The whole paper is divided into four sections. The first section talks about the research background of the article, the second section talks about the research method fuzzy inference method used in this paper, the third section talks about the reliability fault prediction of CNC machine tools based on fuzzy inference, and the last section summarizes the paper, the shortcomings of the current research and the outlook for future research.

II. FUZZY INFERENCE

At the University of California, USA, Professor Zadeh introduced the concept of fuzzy sets in 1965 after years of research. Suppose the range U under study is the set, denoted

as {u}, and U is called the theoretical domain, and the elements u inside the domain are denoted. $\mu_F : U \rightarrow [0, 1]$

$$u \rightarrow \mu_F(u) \tag{1}$$

$$0 < \mu_F(u) < 1 \tag{2}$$

The mapping μ_F denotes the affiliation function of a fuzzy set U. $\mu_F(u)$ is called the affiliation of u to the fuzzy set U. The value interval of the subordination is [0,1], when the subordination function is $\mu_F(u)=1$, then the element u belongs to the domain U completely. When the subordination function is $\mu_F(u)=0$, then the element u does not belong to the domain U at all. When the subordination function is $0 < \mu_F(u) < 1$, then for the element u, it partially belongs to the theoretical domain U; the larger the value $\mu_F(u)$ of then, the greater the degree to which u belongs to the theoretical domain U; the smaller the value $\mu_F(u)$ of then, the smaller the degree to which u belongs to the theoretical domain U [5].

In classical set theory, the boundaries of the objects studied are clear; an element is affiliated with 1 if it belongs to a set, and 0 if it does not belong to a set.

$$f_A(x) = \begin{cases} 1, & x \in A \\ 0, & x \notin A \end{cases} \tag{3}$$

Reasoning is the process and way of thinking in which a new judgment is derived from a known judgment or judgments according to certain laws, and it is a thinking activity in which an unknown result is derived from known conditions [6]. Fuzzy reasoning is based on fuzzy set theory and fuzzy logic, and represents the probability that an event may occur by a likelihood value, which takes any number between 0 and 1 [7]. It is a process of converting input into output by fuzzy rules in an uncertainty inference method, and the result obtained is a fuzzy set or affiliation function. Fuzzy rules are the rules on which fuzzy inference is performed and can usually be expressed in natural language.

III. FUZZY INFERENCE-BASED RELIABILITY FAULT PREDICTION OF CNC MACHINE TOOLS

A. Fault Data Sample Collation

The data comes from collaborative teams. The cooperation team has established a long-term cooperative relationship with machine tool manufacturers.

The collected fault data were stage-wise analyzed and integrated into a total of 100 sets of sample data. 70 sets of data were used for training and 30 sets of data were used for detection through ANFIS. The inputs were two nodes: the N value and the b value, and the output was the predicted rank M. The data samples were listed in Table I.

TABLE I. TABLE OF OUTPUT PREDICTION LEVELS M

Serial number	N	b	M	Serial number	N	b	M
1	131	0.49	2	51	109	0.92	3
2	66	0.51	2.5	52	102	0.93	3
3	114	0.52	2	53	109	0.92	3
4	138	0.48	2	54	38	0.92	3.5
5	74	0.49	2.5	55	32	0.94	3.5
6	59	0.53	2.5	56	32	0.94	3.5
7	125	0.5	2	57	35	0.93	3.5
8	74	0.49	2.5	58	109	0.92	3
9	158	0.45	2	59	109	0.92	3
10	79	0.48	2.5	60	117	0.91	3
11	144	0.47	2	61	68	0.99	3.5
12	59	0.53	2.5	62	74	0.98	3.5
13	109	0.53	2	63	31	0.95	4
14	63	0.53	2.5	64	26	0.97	4
15	74	0.52	2.5	65	23	0.98	4
16	131	0.49	2	66	53	1.02	3.5
17	120	0.51	2	67	58	1.01	3.5
18	114	0.52	2	68	63	1	3.5
19	66	0.51	2.5	69	19	1	4
20	70	0.5	2.5	70	21	0.99	4
21	177	0.85	3	71	23	0.98	4
22	251	0.8	2.5	72	19	1	4
23	107	0.79	3	73	16	1.02	4
24	66	0.86	3	74	49	1.03	3.5
25	199	0.84	2.5	75	49	1.03	3.5
26	188	0.85	2.5	76	53	1.02	3.5
27	100	0.8	3	77	53	1.02	3.5
28	158	0.88	2.5	78	16	1.02	4
29	61	0.87	3	79	16	1.02	4
30	199	0.84	2.5	80	18	1.01	4
31	75	0.84	3	81	19	1	4
32	188	0.85	2.5	82	16	1.02	4
33	81	0.83	3	83	21	0.99	4
34	177	0.86	2.5	84	7	0.98	4.5
35	177	0.86	2.5	85	8	0.97	4.5
36	75	0.84	3	86	9	0.96	4
37	61	0.87	3	87	7	0.99	4.5
38	70	0.85	3	88	26	97	4
39	188	0.85	2.5	89	18	1.01	4
40	188	0.85	2.5	90	16	1.02	4
41	35	0.93	3.5	91	20	1	4
42	38	0.92	3.5	92	18	1.01	4
43	125	0.9	3	93	5	1.02	4.5
44	95	0.94	3	94	6	1.01	4.5
45	102	0.93	3	95	7	0.99	4.5
46	32	0.94	3.5	96	24	0.98	4
47	35	0.93	3.5	97	26	0.97	4
48	32	0.94	3.5	98	24	0.98	4
49	109	0.92	3	99	22	0.99	4
50	109	0.92	3	100	20	1	4

^a: N: fault numbers; M: fault levels.

^b: The b-value is the slope of a straight line, indicating the degree of activity of the fault occurrence.

B. Predictive Model Modeling

We applied the seismic G-R curve model to the CNC machine tool and analyzed the fault activity parameters of the CNC machine tool. Since the fault signs and faults of CNC machine tools are also fuzzy in nature, and fuzzy neural networks have good fuzzy reasoning and adaptive learning capabilities [8-11]. Nonlinear relationships can be well modeled.

The model has only two rules, as follows.

Rule 1:

$$\text{if } x \text{ is } A_1 \text{ and } y \text{ is } B_1 \text{ then } f^1 = p^1x + q^1y + r_1;$$

Rule 2:

$$\text{if } x \text{ is } A_2 \text{ and } y \text{ is } B_2 \text{ then } f^2 = p^2x + q^2y + r_2;$$

$$\text{Final output: } f = f_1 + f_2$$

For any input variable $[x, y]$, the output of this inference system, f is the weighted average of the two rule outputs, and then we get:

$$f = \frac{\omega_1 f_1 + \omega_2 f_2}{\omega_1 + \omega_2} = \bar{\omega}_1 f_1 + \bar{\omega}_2 f_2 \quad (4)$$

$$\bar{\omega}_1 = \frac{\omega_1}{\omega_1 + \omega_2} \quad (5)$$

$$\bar{\omega}_2 = \frac{\omega_2}{\omega_1 + \omega_2} \quad (6)$$

ω_1, ω_2 are the product of the values of the affiliation functions in general.

The network structure is shown in the figure below (see Fig. 1).

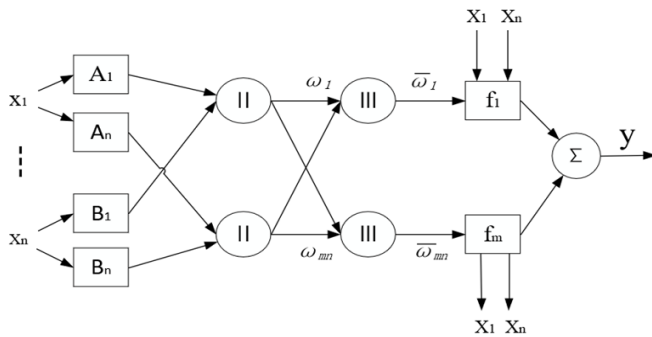


Fig. 1. Takagi-Sugeno fuzzy neural network model structure diagram.

Takagi-Sugeno fuzzy neural network model is divided into five layers: input layer, fuzzification layer, rule layer, defuzzification layer, and output layer. The fuzzy rule front piece is the first three layers, and the rule back piece is the last two layers. The square shape indicates the untrainable nodes and the circle indicates the trainable nodes.

Layer 1: Input layer. The n nodes in this layer are connected to the input vector $X = \{X_1, X_2, \dots, X_n\}$ and transmit the input value X to the next layer, A and B are the fuzzy linguistic variables of the input X .

Second layer: fuzzification layer. This layer has $n \times m$ nodes, m is the number of fuzzy sets of each variable, each node represents a linguistic variable value, the input of the i group of m nodes is X_i , the output is the affiliation function of

each input variable belongs to the fuzzy set of each linguistic variable value $\mu_i^j(X_i)$, $\mu_i^j(X_i)$ is the j fuzzy set of X_i .

Layer 3: Rule layer. Each node represents a fuzzy rule, and the specific role is to match each input variable to the corresponding fuzzy rule, and to calculate the fitness of the input value to each rule. The mathematical equation is calculated as follows.

$$\omega_j = \mu_1^i \mu_2^i \dots \mu_i^i \quad (7)$$

Layer 4: Normalization layer. This normalizes the incoming model data from the previous layer for calculation.

$$\bar{\omega}_j = \frac{\omega_j}{\sum_{i=1}^m \omega_j} \quad (8)$$

Layer 5: Output layer, output layer y is the weighted sum of the network results after the rule.

$$y = \sum_{j=1}^m \bar{\omega}_j y_j \quad (9)$$

y_j is the weighted sum of the posterior pieces of each rule.

The more the antecedents are satisfied, the stronger the rule is and the more it guides the output. The learning algorithm mainly adjusts the parameters to make the final output have better results.

The following figure shows the structure of the Sugeno fuzzy system model (see Fig. 2).

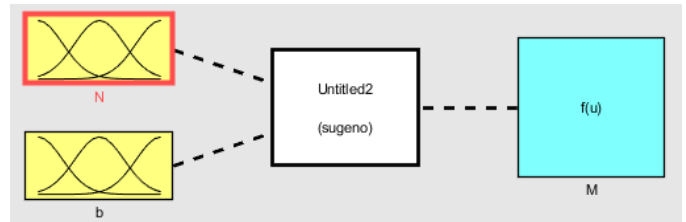


Fig. 2. Sugeno fuzzy system model.

The advantages of this model, which is a nonlinear model, is very adaptable, can easily adjust parameters, can express more complex, high-dimensional nonlinear systems with fewer conditions, and is insensitive to parameter changes [12-17]. Its output variables are constants or linear functions and the output is an exact quantity.

C. Model Matlab Simulation

The adaptive neuro-fuzzy inference system can be implemented in Matlab using a toolbox or an editor [17-21]. The goal is to make predictions about fault levels, based on the number of faults and the value of the active parameter b as input. 100 sets of data are input, 70 sets of data are used for training and 30 sets of data are used for detection. The inputs are two nodes: N values and b values, and the outputs are the predicted levels M . The initial FIS is formed with two variables as inputs, four fuzzy subsets for each input quantity, and four

regular outputs for each subset, for a total of 16, and finally these subsets are clarified to produce 1 output quantity. The network structure is illustrated below (Fig. 3).

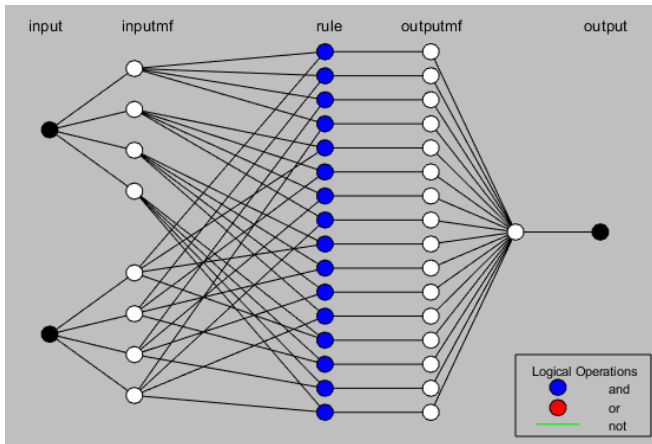


Fig. 3. Model system structure diagram.

Training on 70 sample data, the training error of the model can reach 0.142 after 9 training sessions, which can obtain good prediction results. The following figure shows the training error curve (from Fig. 4 to Fig. 8).

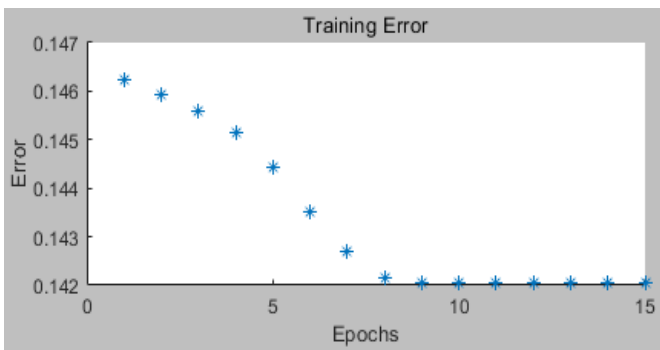


Fig. 4. Training error diagram.

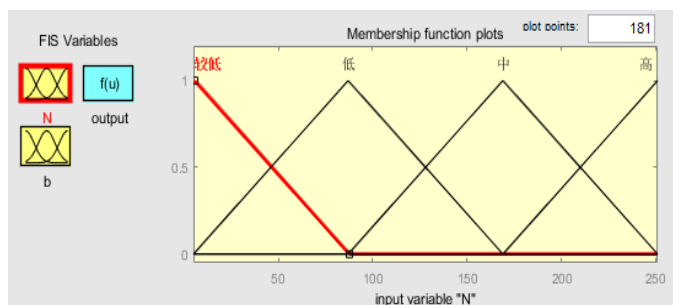


Fig. 5. Plot of the affiliation function of the input variable N.

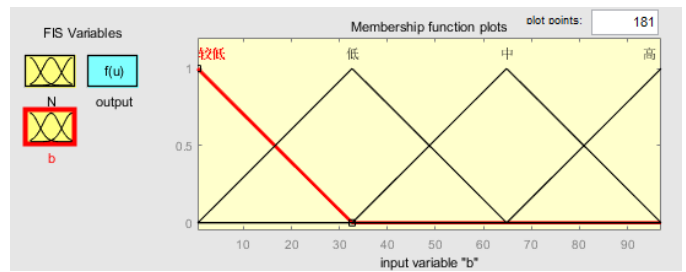


Fig. 6. Plot of the affiliation function of input variable b.

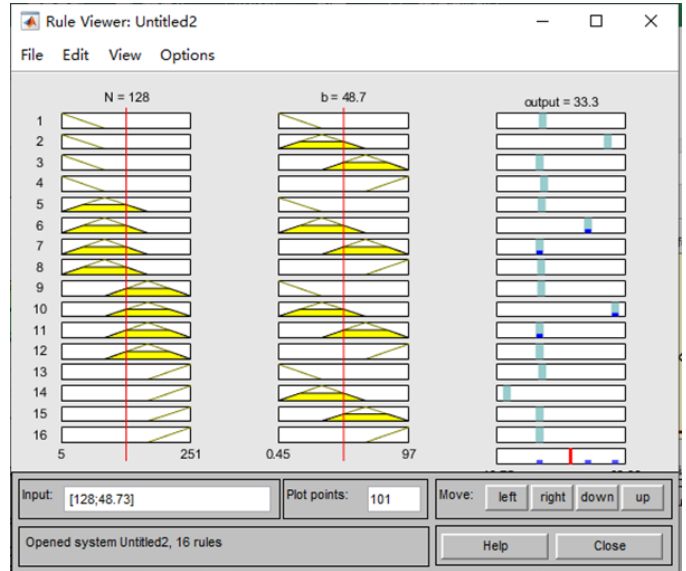


Fig. 7. Fuzzy rule observation chart.

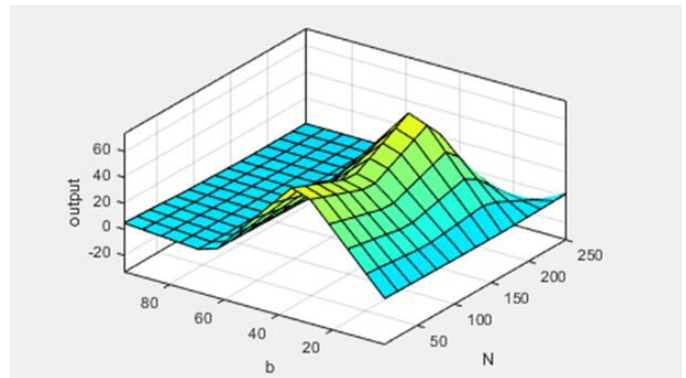


Fig. 8. 3D surface plot of input variables and output variables.

D. Model Validation and Analysis of Data

The output predicted data, original data and error data are organized in the following Table II.

TABLE II. PREDICTION GRADE ERROR TABLE

Original Grade	Prediction Level	Error	Original Grade	Prediction Level	Error
2	2.0	0.0	3	2.7	0.3
2.5	2.5	0.0	3	2.8	0.2
2	2.1	-0.1	3	2.7	0.3
2	2.0	0.0	3.5	3.6	-0.1
2.5	2.5	0.0	3.5	3.8	-0.3
2.5	2.6	-0.1	3.5	3.8	-0.3
2	2.0	0.0	3.5	3.7	-0.2
2.5	2.5	0.0	3	2.7	0.3
2	2.0	0.0	3	2.7	0.3
2.5	2.4	0.1	3	2.6	0.4
2	2.0	0.0	3.5	3.4	0.1
2.5	2.6	-0.1	3.5	3.3	0.2
2	2.2	-0.2	4	3.8	0.2
2.5	2.6	-0.1	4	3.9	0.1
2.5	2.5	0.0	4	4.0	0.0
2	2.0	0.0	3.5	3.6	-0.1
2	2.0	0.0	3.5	3.6	-0.1
2	2.1	-0.1	3.5	3.5	0.0
2.5	2.5	0.0	4	4.1	-0.1
2.5	2.5	0.0	4	4.0	0.0
3	2.5	0.5	4	4.0	0.0
2.5	2.6	-0.1	4	4.1	-0.1
3	2.6	0.4	4	4.1	-0.1
3	3.2	-0.2	3.5	3.7	-0.2
2.5	2.5	0.0	3.5	3.7	-0.2
2.5	2.5	0.0	3.5	3.6	-0.1
3	2.7	0.3	3.5	3.6	-0.1
2.5	2.5	0.0	4	4.1	-0.1
3	3.2	-0.2	4	4.1	-0.1
2.5	2.5	0.0	4	4.1	-0.1
3	3.0	0.0	4	4.1	-0.1
2.5	2.5	0.0	4	4.1	-0.1
3	2.9	0.1	4	4.0	0.0
2.5	2.5	0.0	4.5	4.2	0.3
2.5	2.5	0.0	4.5	4.1	0.4
3	3.0	0.0	4	4.1	-0.1
3	3.2	-0.2	4.5	4.2	0.3
3	3.1	-0.1	4	4.0	0.0
2.5	2.5	0.0	4	4.1	-0.1
2.5	2.5	0.0	4	4.1	-0.1
3.5	3.7	-0.2	4	4.0	0.0
3.5	3.6	-0.1	4	4.1	-0.1
3	2.5	0.5	4.5	4.3	0.2
3	2.9	0.1	4.5	4.3	0.2
3	2.8	0.2	4.5	4.2	0.3
3.5	3.8	-0.3	4	3.9	0.1
3.5	3.7	-0.2	4	3.9	0.1
3.5	3.8	-0.3	4	3.9	0.1
3	2.7	0.3	4	4.0	0.0
3	2.7	0.3	4	4.0	0.0

The fault level prediction curve has the same trend as the original curve (shown in Fig. 9), and the maximum error from the predicted fault level does not exceed 0.5 (shown in Fig. 10), indicating that this fuzzy model is suitable for predicting the fault level in this way.

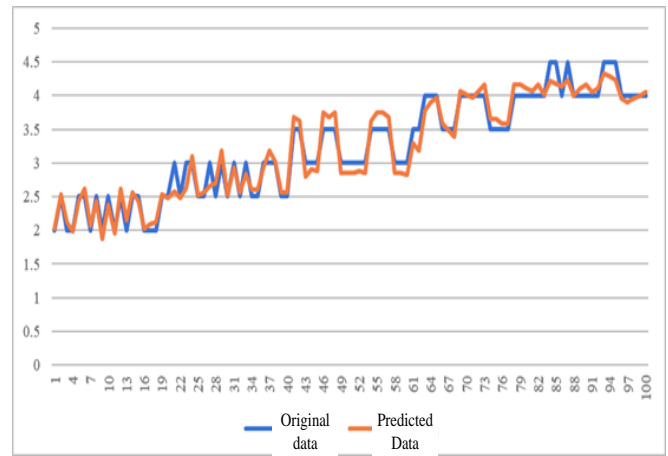


Fig. 9. Fault level prediction data and original data trend line graph.

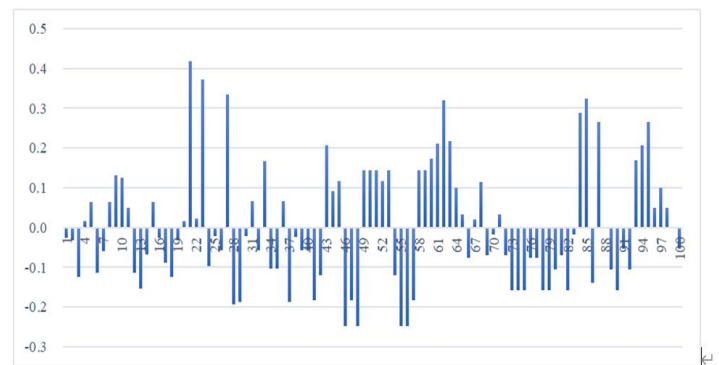


Fig. 10. Fault error cluster bar chart.

IV. CONCLUSION

The fault data of CNC machine tools have been collected and analyzed, and then the Gutenberg–Richter (G-R) curve relationship in the earthquake has been applied to CNC machine tools. The physical meaning of the parameters is studied to describe the failure grade of the CNC machine tools. By comparing the b-value and the ratio of large failure to small failure with time, it is proved that b-value can reflect the ratio of large failure to small fault over a period of time.

CNC machine tool faults are random and fuzzy in nature, but there may actually be close relationships linked internally, as fuzzy neural networks have good adaptability and can solve nonlinear fuzzy system capabilities. The non-linear relationship between fault level and b-value and N-value can be linked by using fuzzy inference. With 100 data samples, 70 samples were used for model training, and the training error did not exceed 0.15. 30 samples were then used for model testing, and the predicted fault levels were analyzed for error with the original levels, and the error was not significant indicating that the model has good prediction capability.

The results of this paper are more reasonable and will provide a basis for the future studies.

When using fuzzy reasoning method for fault prediction, the influence of various other factors on the prediction is not considered, and only the relationship between activity parameters and earthquake magnitude is predicted. Explore other influencing factors during later research.

ACKNOWLEDGMENT

This work was sponsored by Wuzhou University Research Foundation for Advanced Talents.

REFERENCES

- [1] D. Du, "Reliability design of dual power tool holder based on fault correlation analysis," Jilin University, 2016.
- [2] X. Zuo, J. Kang, H. Li, and L. Tang, "A review of failure prediction techniques," *Firepower and Command and Control*, vol. 35, no. 1, pp. 1-5, 2010.
- [3] Z. Zheng, D. Liu, X. Shen, and X. Wang, "Reliability of b-value full time scan results and correlation with earthquakes in North China," *Earthquakes*, no. 3, pp. 8-14, 2001.
- [4] W. Wang, Y. Zheng, and L. Bian, "Preliminary study of nonlinear magnitude frequency relationship applied to earthquake prediction," *International Earthquake Dynamics*, no. 8, pp. 100-101, 2019.
- [5] G. Zhang, C. Zhang, Z. Xie, "Application of T-S fuzzy neural network-based model in typhoon disaster prediction with the example of Hainan," *Disaster Science*, vol. 28, no. 2, pp. 86-89, 2013.
- [6] M. Cobaner, B. Unal, O. Kisi, "Suspended sediment concentration estimation by an adaptive neuro-fuzzy and neural network approaches using hydro-meteorological data," *Journal of Hydrology*, vol. 367, no. 12, pp. 52-61, 2009.
- [7] T. Xu, "Research application of fuzzy neural network in student performance prediction," Qingdao University, 2019.
- [8] X. Shi and Z. Hao, "Fuzzy control and its MATLAB simulation," Tsinghua University Press. Beijing. pp. 10-12, 2008.
- [9] S. Xiao, "Research on causal graph fault diagnosis based on binary decision diagram and fuzzy inference," Chongqing Normal University, 2019.
- [10] W. Cai, H. Li, H. Gong, J. Tuo, C. Liu, and Y. Jiao, "Evaluation method of smart energy meter suppliers based on hierarchical analysis method," *Electrical Measurement and Instrumentation*, vol. 56, no. 1, pp. 121-127+135, 2019.
- [11] H. Cai, "Construction of CNC machine tools," Beijing: Beijing Institute of Technology Press, 2016.
- [12] Y. Li, X. Hu, and A. Qiao, "An improved fuzzy hierarchical analysis method," *Journal of Northwestern University (Natural Science Edition)*, no. 1, pp. 11-1, 2005.
- [13] K. Chen, Y. Wang, and J. Liu, "Choice of cutting fluid for green manufacturing machine bed based on FAHP-GRA," *Modular Machine Tool and Automatic Manufacturing Technique*, vol. 564, no. 2, pp. 140-144, 2021.
- [14] J. Q. Wei, and C. D. Piao, "Reliability analysis and evaluation of foreign high-grade machining center," *Machine Tool and Hydraulics*, vol. 44, no. 1, pp. 194-197, 2016.
- [15] J. X. Ding, and J. Yu, "Analysis of fault G-R relationship of CNC Lathe," *Intern Comb Eng Part*, no. 14, pp. 179-180, 2019.
- [16] X. D. Qi, "Reliability comprehensive evaluation of manufacturing process for key components of CNC machine tools," Chongqing University, 2017.
- [17] G. B. Zhang, L. She, Y. Ran, D. M. Luo, "Study on extraction of key functional components in CNC machine tools' reliability tests," *China Mechanical Engineering*, vol. 27, no. 17, pp. 2372-2378, 2016.
- [18] Q. K. Bian, S. Z. Li, Z. W. Mao, S. L. Li, Q. C. Wang, "A study of the seismic activity anomaly before the ZhangBei MS 6.2 Earthquake and the characteristics of its sequence," *North China Earthquake Sciences*, no. 3, pp. 35-42, 1999.
- [19] J. N. Xue, X. Li, B. Zhang, Y. W. Wang, "Statistical analysis of B-values in Shandong area," *North China Earthquake*.
- [20] M. Mishra, Abhishek, R.B.S. Yadav, and M. Sandhu, "Probabilistic assessment of earthquake hazard in the Andaman-Nicobar-Sumatra region," *Natural Hazards*, vol. 105, no. 1, pp. 313-338, 2020.
- [21] V. M. Tiwari, "CSIR-national geophysical research institute—60 years of enduring scientific contributions," *Journal of the Geological Society of India*, vol. 96, no. 4, pp. 319-324, 2020.

Using Machine Learning Algorithm as a Method for Improving Stroke Prediction

Nojood Alageel, Rahaf Alharbi, Rehab Alharbi, Maryam Alsayil, Lubna A. Alharbi

Faculty of Computers and Information Technology
University of Tabuk
Tabuk, Saudi Arabia

Abstract—Having sudden strokes has had a very negative impact on all aspects in society to the point that it attracted efforts for better improvement and management of stroke diagnosis. Technological advancement also had an impact on the medical field such that nowadays caregivers have better options for taking care of their patients by mining and archiving their medical records for ease of retrieval. Furthermore, it is quite essential to understand the risk factors that make a patient more susceptible to strokes, thus there are some factors that make stroke prediction much easier. This research offers an analysis of the factors that enhance the stroke prediction process based on electronic health records. The most important factors for stroke prediction will be identified using statistical methods and Principal Component Analysis (PCA). It has been found that the most critical factors affecting stroke prediction are the age, average glucose level, heart disease, and hypertension. A balanced dataset is used for the model evaluation which was created by sub-sampling since the dataset for stroke occurrence is already highly imbalanced. In this study, seven different machine learning algorithms are implemented: Naïve Bayes, SVM, Random Forest, KNN, Decision Tree, Stacking, and majority voting to train on the Kaggle dataset to predict occurrence of stroke in patients. After preprocessing and splitting the dataset into training and testing sub-datasets, these proposed algorithms were evaluated according to accuracy, f1 score, recall value, and precision value. The NB classifier achieved the lowest accuracy level (86%), whereas the rest of the algorithms achieved similar accuracies 96%, f1 scores 0.98, precision 0.97, and recall 1.

Keywords—Stroke prediction; machine learning; PCA; decision tree; KNN; majority voting; Naïve Bayes

I. INTRODUCTION

Strokes or cerebrovascular accidents are considered among the top three causes of morbidity and mortality in many countries all over the world [1], such that it accounts for around 10% of the world-wide deaths which makes it the second leading cause of death. As an estimation, approximately 700,000 individuals suffer from strokes each year, and by the year 2030, it is expected that this number will be greatly increased and will cause a medical cost of 240 billion dollars in the US alone [2].

The world health organization WHO defines stroke as a brain-related illness such that it leads to the dysfunction of the brain, and it could be focal, acute, or diffuse. This dysfunction is mainly a result of vessel problems and it lasts for longer than 24 hours. Ultimately, there are many types of strokes

depending on the exact origin of the dysfunction, which defines four main types of strokes: ischemic stroke, subarachnoid hemorrhage, cerebral venous sinus thrombosis, and intra-cerebral hemorrhage [3].

In general, brain strokes can be classified as either ischemic or hemorrhagic. Ischemic strokes are the predominant type and they account for approximately 70% of the total stroke incidents [4]. Ischemic strokes occur as a result of clots in vessels, or hypotensive vasoconstriction, arterial tears, and sickle cell anemia [5]. On the other hand, hemorrhagic strokes account for approximately 15% of the total incidents, yet their effects are usually more detrimental as they often lead to serious morbidity and death [6]. Hemorrhagic strokes occur due to many causes among which are the vascular malfunction and uncontrolled hypertension [7].

When considering the risk factors or the reasons behind the occurrence of strokes, these can be divided into two types of factors depending on their origin, meaning that there are factors that can be changed or modified, and factors that cannot be modified [8]. Some of the modifiable (changeable) factors is hypercholesterolemia, diabetes, and hypertension. On the other hand, the non-modifiable factors include age, gender, and the genetic factors in play [9].

The traditional stroke identification methods are usually the magnetic resonance imaging MRI and Computed Tomography CT scans which are expensive and invasive [10]. However, since the stroke occurrence is a very time-sensitive issue, dealing with it in a timely efficient manner is very important because in most cases, death or permanent damage from stroke can be prevented if the diagnosis happens early on [11], [12]. Therefore, it is essential to develop medical tools and devices that allow physicians to diagnose a stroke without being invasive or uncomfortable, through relying on biomarkers for example or studying the risk factors. Machine learning poses as the perfect tool for predicting whether a stroke can occur or not based on different factors. Machine Learning is capable of diagnosing, treating, and predicting disease through analyzing clinical data.

In this research, the aim is to develop and implement a machine learning-based system for the accurate prediction of future occurrence of stroke in patients based on several features including age, gender, BMI, and medical history. The primary objective is to get this system to predict the occurrence of stroke by 100% accuracy so that lives can be saved. The

contributions that are provided in this report can be listed as follows:

- Predictive analytics approach to predict stroke recurrence is suggested.
- Machine learning and neural network algorithms are implemented.
- A publicly available dataset of electronic health records is used.
- The subsampling techniques for balancing the dataset is followed.
- Dimensionality reduction techniques are implemented in analyzing the attributes.
- The most impactful features for predicting strokes are picked out and shown.

Thus, after mentioning the contributions, it can be said that the added value of this paper lies in the fact that it uses simple algorithms to achieve high accuracies with explainable results, instead of using complex algorithms. More precisely, the majority of the chosen algorithms were able to score similarly high results.

The rest of the paper is distributed as follows: Section II is the literature review where some studies are mentioned with their relative results. Section III is for describing the details of the methodology followed in this study. Section IV shows the results that were obtained by the proposed model. Finally, the paper is concluded with Section V as a conclusion.

II. LITERATURE REVIEW

Since technologies like machine learning and deep learning can greatly benefit the medical sector by increasing the accuracy of stroke prediction, many studies were conducted to explore how exactly machine learning models can be used in predicting strokes. In this section, a group of similar studies that relied on freely available datasets such as Kaggle and datasets from local hospitals or labs were selected.

Dritsas and Trigka [13] gathered data from Kaggle such that the participants were 3254. The dataset consists of 10 independent features such as age, BMI value, glucose level, smoking status, hypertension, and whether the individual had contracted a stroke before. Data preprocessing was performed on the dataset, and class balancing was implemented through a resampling method known as SMOTE. Machine learning models namely Stacking, Decision Tree, Random Forest, Majority Voting, Naïve Bayes, Multilayer Perceptron, KNN, Stochastic Gradient Descent, and logistic regression were used for predicting stroke or no-stroke. It appears from the results that the stacking classifier performed best and achieves 0.989 AUC value, with 0.974 precision and 0.974 recall. The other high performing models were Random Forest, KNN, and Majority Voting.

Rakshit [14] also relied on the Kaggle dataset and some of the algorithms as [13] namely Decision Tree, Naïve bayes, Support Vector Machine, Random Forest, K-Nearest Neighbor and Logistic Regression. According to their results, the best

performance was recorded by Decision Tree followed by KNN (96.3%).

Using Kaggle dataset, Sailasya and [15] discussed the prediction of stroke based on machine learning algorithm namely Logistic Regression, K-Nearest Neighbour, Random Forest, Support Vector Machine, Naïve Bayes, and Decision Tree algorithms. Undersampling method was used to handle the imbalanced data. The results showed that among these algorithms, Naïve Bayes had the best performance with 82% overall accuracy compared to 80 % for both K-NN and support vector machine, and 78% for logistic regression.

Emon et al. [16] collected information for 5110 patients were taken from Bangladesh's medical clinic. Then, ten different machine learning classifiers, which are ANN, MLP, K Neighbours algorithm, SGD, QDA, AdaBoost, Gaussian, QDA, GBC, and XGB were used. The weighted voting classifier offered the highest accuracy of about 97%, GBC and XGB classifiers achieving 96% accuracy, right before AdaBoost classifier that scored 94% accuracy. On the other hand, the lowest accuracy was recorded by the SGD classifier with a value of 65%.

Shoily et al. [17] used KNN, Naïve Bayes, J48, and Random Forest classifiers. They gathered data from multiple sources to create their dataset of 1058 individuals overall and took a total of 28 features. The authors performed integer encoding to make the machine learning algorithms suitable for WEKA processing. After that, feature selection took place, and the models were trained and tested then evaluated according to f1 score, accuracy, precision, and recall. In terms of accuracy, Random Forest as well as KNN and J48 achieved the same results: 0.998 accuracy, 0.998 f1 score, 0.998 precision and 0.98 recall, whereas Naïve Bayes achieved 0.856 accuracy and 0.861 f1 score.

Abedi et al. [18] created a dataset termed "GNSIS", which is a collection of electronic health records from 2003 to 2019. Data preprocessing was performed, and the individuals within the dataset were classified into six groups totaling 2091 individuals, 1 group consists of those who didn't contract stroke in the last 5 years, and the other 5 groups are of stroke patients. After that, the dataset was split into training and testing by 80 to 20 ratio, where data imputation was also done. From the dataset, 53 features existed including BMI, diastolic blood pressure, creatinine, and smoking status. Then, four feature selection sets were created with exclusion of some features at times, and six machine algorithm models were used each in all of the 5 recurrence prediction window, which makes 24 models in total. For 1year prediction window, Random Forest achieved the better results with 90% accuracy, whereas the average accuracy of all models was 88%. The average accuracy achieved in the 5 years prediction window was 78%, thus the wider prediction window results in less accurate performances.

Relying on electronic health records, Nwosu et al. [19] used a dataset published by McKinsey & Company, containing 11 different attributes including body mass index, heart disease, marital status, age, average blood glucose, and smoking status. In the dataset, 548 patients suffered from stroke whereas 28524 patients didn't suffer from any previous strokes, thus the

dataset needed downsizing. In fact, 1000 downsizing experiments were done to avoid sampling bias. After that, 70% of the dataset was selected for training and 30% for testing. Over the 1000 experiments, the Neural Network model achieved the best accuracy of 75.02%, followed by Random Forest at 74.53% accuracy and Decision Tree at 74.31%.

In [13], the dataset was large and their study was able to score very similar results to ours, even though at times our metrics were better. However, they did not mention the scored accuracy. Similarly, our proposed model achieved better performances than [14].

It's noteworthy that the proposed method in this study achieved 96.7% accuracy, which is significantly higher than the accuracy of [15] (80%).

In [16] the authors chose complex algorithms such as ADABOOST and XGB and were only able to achieve similar results to ours, whereas we achieved the high performances using much simpler algorithms, which is more desirable.

In [17] the study relied on 28 inputs to predict stroke occurrence, which is usually difficult to obtain from patients for a quick prediction. Conversely, the proposed algorithms in the proposed system in this paper relied on 9 factors only as an input. In addition, [17] used a much smaller dataset.

Similarly, [18] used a very high number of input, which is not desirable for ML algorithms.

III. METHODOLOGY

Machine learning permits the advancement of a system by making it capable of learning and improving from past experiences without the need of constant continuous programming. These systems learn through machine learning how to analyze data to identify patterns that help them make decisions in the future without the help of humans.

The real influence of machine learning becomes crystal clear in the fields that deal with a huge amount of data such as retail, health, government, finance, and transportation. This is mainly due to the decision-making capabilities of machine learning since it can understand the data and fit them into the different models such that human can rely on them for decisions. Machine learning models are efficiently used for identifying diseases and computing risk satisfaction in the healthcare sector. The previous are only a few examples of the capabilities of machine learning.

Nonetheless, real-life data cannot be simply directly processed by the selected machine learning algorithms which is why data preprocessing is an essential step before applying the ML models. After that, the available dataset must be divided into training and testing datasets. The training step is performed in order to teach the algorithm about the data. In addition, unknown data can be predicted through ML algorithms, yet the prediction results are checked against each other.

This study is dedicated to implementing machine learning algorithms for stroke prediction, since it is a dangerous and common disease. Machine learning is often suitable for datasets due to its simplicity, structure, and compatibility with

a wide range of machine learning platforms and tools. For this reason, machine learning algorithms were chosen in this study. However, the limitation of this method is that it requires many inputs for the model to be able to make predictions. It is possible that when predicting a person's status, not all inputs are available, and then the model will not be able to predict. This issue was removed since the chosen dataset was large.

In general, a wide set of attributes are used to predict strokes such as gender, age, and blood pressure data among many others. Additionally, the performance of a number of machine learning algorithms was examined to see which one is best suited for predicting stroke incidence based on the dataset. Ultimately, the chosen ML algorithm must give the predictions with the highest accuracy.

A. Implementation

In this section, the machine learning algorithms that will be implemented and put to the test are presented and described.

1) *Naive bayes*: In the cases when the features are highly independent, the Naïve Bayes NB algorithm can lead to probability maximization [20]. There is a feature vector f_i for every subject i at that class c such that $P(c|f_1, \dots, f_n)$ is maximized. The formula that defines the conditional probability is as in (1):

$$P(c|f_1, \dots, f_n) = P(f_1, \dots, f_n|c) / P(c) P(f_1, \dots, f_n) \quad (1)$$

In (1), $P(f_1, \dots, f_n|c) = \prod_{j=1}^n P(f_{ij}|c)$ resembles the features probability given class, whereas the previous feature probability is resembled by $P(f_1, \dots, f_n)$, and previous class probability is resembled by $P(c)$. Through maximizing the numerator of 1, its number is also maximized, and the optimization becomes as in (2):

$$\hat{c} = \arg \max_c P(c) \prod_{j=1}^n P(f_{ij}|c) \quad (2)$$

2) *Random forest*: There are multiple decision trees in a Random Forest (RF) classifier [21]. When these independent trees are combined in an ensemble through resampling, the results become subsets of instances that are used for classification and regression. In a random forest, the final output is a result of majority voting, since each independent tree generates its own classification outcome.

3) *K-Nearest neighbors*: K-nearest neighbors (KNNs) classifier depends on Manhattan or Euclidean distances to evaluate similarities or differences between instances in the dataset [22]. More often than not, the Euclidean distance is the metric of choice in KNN classifiers. In stroke prediction, the features vector of the new samples would be f_{new} . The closest K vectors (neighbors) to f_{new} is determined through KNN. After that, the class where most neighbors belong is given the f_{new} value.

4) *Decision tree*: In the proposed Decision Tree model [23], J48 resembling the single classifier, and RepTree [24] resembling the base classifier were chosen. The classes are denoted by the leaf nodes, whereas the features are denoted by the internal nodes. The Gini index technique is employed by the J48 classifier in order to split a single feature

at each node. Gini index is a fast and simple decision learner that is capable of building a DT through the gained information as an impurity measure and pruning via reduced-error pruning.

5) *Majority voting*: Soft or hard voting is implemented through simple majority voting, assuming an ensemble of K basis models. This method allows the prediction of the class label associated to an instance [25]. The hard voting collects the votes related to each class label and chooses the one with most votes as an output, that is the candidate class. On the other hand, the predicted probabilities for every class label are collected by soft voting, and the class label with the largest probability is predicted. In the proposed model, the hard voting is adopted. Its general function of hard voting is represented by (3):

$$\max \sum_{k=1}^K 1 P_{k,c} \quad (3)$$

Such that $P_{k,c}$ is the prediction or probability of k-th model in class c, and $c = \{\text{Stroke, Non - Stroke}\}$.

6) *Stacking*: One of the ensemble learning techniques is the Stacking, where the predictions of multiple heterogeneous classifiers are integrated within a meta-classifier. Usually, the training set is used for training the base models whereas the outputs of the base models are used to train the meta-classifier. Here, J48, RF, NB, and RepTree were chosen to be included in the stacking ensemble classifier. The predictions of these collective classifiers are used for training a logistic regression meta-classifier.

The influence of machine learning parameters on the performance of a model can vary depending on the specific algorithm used, the dataset being analyzed, and the problem being solved. However, in general, adjusting the values of these parameters can have a significant impact on the accuracy and speed of a machine learning model. In this study, several parameters for the different algorithms were modified to make sure better results are achieved. The modifications to the parameters of each algorithm are shown in Table I.

TABLE I. THE CHANGED PARAMETERS FOR EACH ALGORITHM

Algorithm	Specific Parameters
<i>KNN</i>	- Number of neighbors (k): value is 6 - Distance metric: (Euclidean distance)
<i>SVM</i>	- Kernel type: default kernel is radial basis function (RBF) - Regularization parameter (C): default value is 1.0
<i>DT</i>	- Tree depth: 3
<i>NB</i>	No modifications
<i>RF</i>	- Number of trees in the forest: default value is 100 - Maximum depth of each tree: 9

B. Pre-Processing

1) *Dataset description*: For this study, the dataset of choice was adopted from Kaggle. The dataset comprises a large number of participants of which only those above 18 years old are chosen, making the total of the participants 3254.

The 9 input attributes (most of which are nominal) as well as the target class are briefly described in Table II.

TABLE II. DESCRIPTION OF THE ATTRIBUTES/FEATURES IN THE DATASET

Risk factor	Description	Details
<i>Age (year)</i>	The actual age of participants	All of the participants are older than 18
<i>Gender</i>	Whether the participants is male or female	In the dataset, 1260 participants are males, and 1994 participants are female
<i>Hypertension</i>	The participant suffering from hypertension or not	12.54% of the participants in the dataset are hypertensive
<i>Heart disease</i>	The participants suffering from heart diseases in general or not	6.33% of the participants in the dataset suffer from heart diseases
<i>Marital status</i>	The participant is married or not	In the dataset, 79.84% of the participants are married
<i>Work type</i>	The work status of the participants	65.02% of them work in the private sector, 19.21% are self-employed, 15.67% have a job, while 0.1% have never worked
<i>Residence type</i>	Whether the participant lives in an urban or rural place	51.14% of the participants in the dataset live in urban place whereas the rest live in rural places
<i>Avg glucose level (mg/dL)</i>	The average level of a participant's blood glucose	Numerical values for each patient
<i>BMI (Kg/m2)</i>	Participant's body mass index of the participants	Numerical values for each patient
<i>Smoking status</i>	Whether a participant currently smokes or not	22.37% of the participants smoke, 24.99% of them have smoked in the past, and 52.64% of them have never smoked
<i>Stroke history</i>	Whether the participant has had a stroke previously or not	5.53% of the participants have previously had a stroke

2) *Data pre-processing*: If the data were kept in their raw form, it might negatively affect the quality of the predictions, which is why data preprocessing is essential. In the raw data, there might be some missing values and redundancy as well as noisy data, so tasks like data discretization and reduction of redundant values are performed. Furthermore, one of the data pre-processing tasks is to balance the classes through selecting one of the available resampling techniques. In the proposed workflow, the SMOTE technique will be used so that the participants can be distributed over the stroke and non-stroke classes in a balanced way. In more details, the minor class which belongs to the stroke participants, oversampling was done to increase the number of participants in this class. In addition, there were not missing or null values, so neither dropping nor data imputation was applied.

C. Proposed Workflow

The details of the proposed approach and methodology can be summed up in a workflow chart presented in Fig. 1.

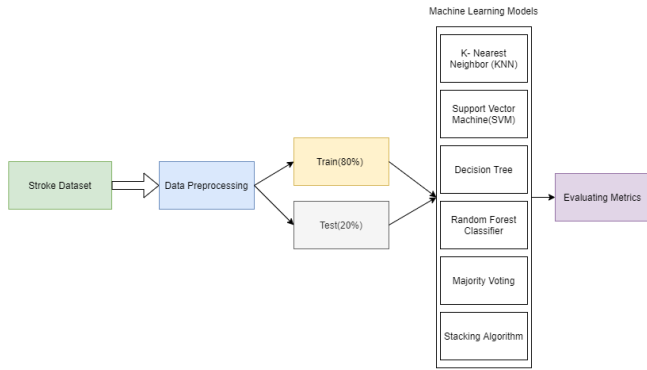


Fig. 1. Workflow of the proposed model.

Initially, the Kaggle dataset with 3254 participants is acquired. Then, the data is visualized to determine the specifics such as visualization of column and the relevant attributes. In this stage, the distribution of the participants can be visualized over the different features such as the age and gender distribution. After that, data preprocessing takes place where the data is being prepared through reduction of redundant information or resampling. In this approach, the SMOTE technique is selected. Later, the data is split into 80% for training and 20% for testing. Six different algorithms were selected to perform the predictions: Naive Bayes, Random Forest, K-Nearest Neighbors, Decision Tree, Majority Voting, and Stacking. These algorithms are then evaluated according to the evaluation metrics.

D. Evaluation

A group of performance metrics were chosen to evaluate the performance of the chosen machine learning methods. The most commonly used metrics in general will also be used in this study [see (4), (5), (6) and (7)]. Sensitivity, which is also termed Recall, represents the true positive results where participants who have had a stroke were successfully classified into the stroke class from the collective totality of the participants. Precision on the other hand specifies how many of those who had a stroke actually belong to this class. Whereas, Recall shows how many of those who had a stroke are correctly predicted. F-measure is the harmonic mean of the precision and recall and sums up the predictive performance of a model.

$$Precision = TP / (TP + FN) \quad (4)$$

$$Precision = TP / (TP + FP) \quad (5)$$

$$F\text{-Measure} = 2 \cdot Precision \cdot Recall / (Precision + Recall) \quad (6)$$

$$Accuracy = (TN + TP) / (TN + TP + FN + FP) \quad (7)$$

Where, true positive is designated by TP and false negative is designated by FN, false positive is designated by FP and true negative is designated by TN.

On the other hand, Area under curve (AUC) is also a beneficial metric, where the values must be between 0 and 1,

such that the higher the AUC value, the better the performance. If the model can discriminate between the instances of two classes perfectly, then AUC would be 1. Conversely, if the model fails to distinguish between any instances, the AUC would be 0.

IV. RESULTS

A. Data Visualization

The dataset can be visualized where each of the features or attributes are analyzed separately and against each other. Fig. 2, for instance, illustrates how the participants from the dataset are distributed according to age and gender. It can be seen that the patients have an average of 41 years old, and that there are slightly more females than males, specifically, 56% of the participants are female.

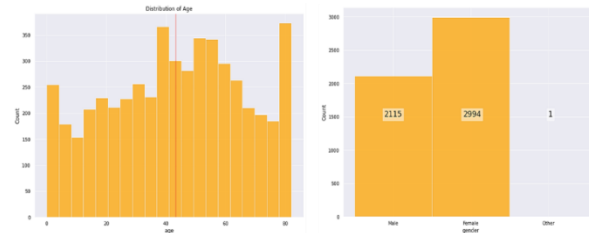


Fig. 2. Distribution of data by age and gender.

On the other hand, Fig. 3 shows how the patients who had suffered from a previous stroke are distributed according to age, where it becomes clear that approximately all of them were older than 40 years old, and the largest number of stroke patients was 80 years old. While the patients who didn't suffer from stroke were distributed among the different age groups.

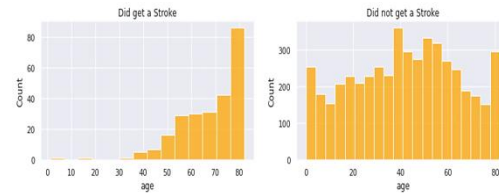


Fig. 3. Distribution of patients who suffered from a stroke and those who didn't according to age.

In addition, Fig. 4 shows that the majority of the participants didn't suffer from any heart diseases, nor did they suffer from Hypertension.

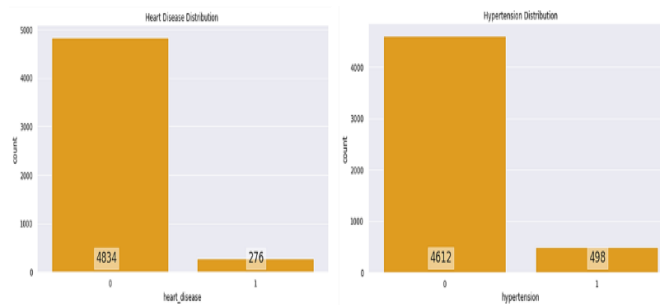


Fig. 4. Distribution of data over heart disease and hypertension cases.

Moreover, 25% of the patients were obese, and 18% of the participants were overweight according to Fig. 5.

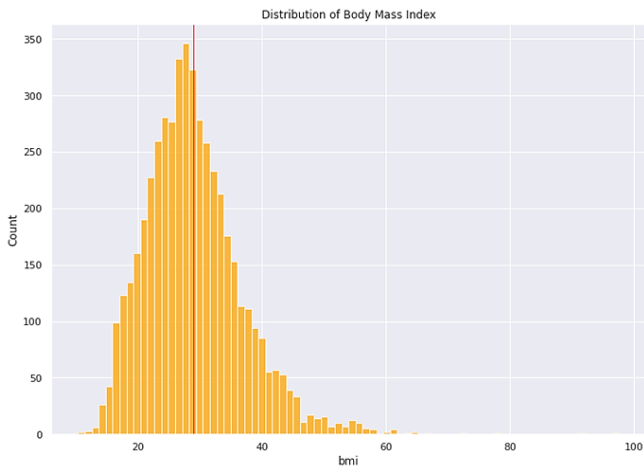


Fig. 5. Distribution of data according to BMI.

In Fig. 6 depicts that the majority of the patients were smokers, followed by a large group of participants with unknown smoking status (1544 participants).

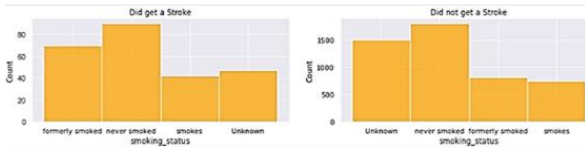
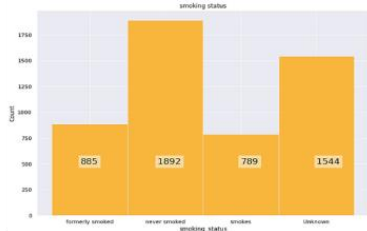


Fig. 6. Distribution of data according to smoking status and relation to stroke.

Additionally, the majority of the participants are employed in the private sector. Meanwhile, the data was almost equally distributed between living in rural and urban areas, as depicted in Fig. 7.

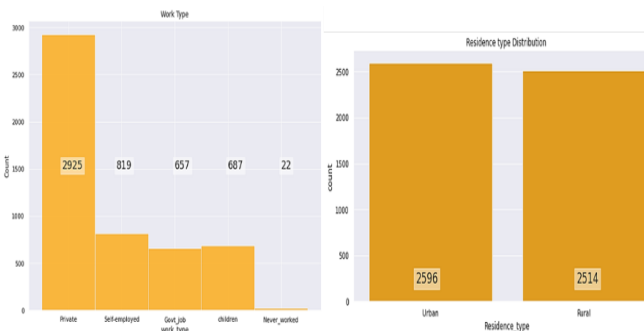


Fig. 7. Distribution of data according to work type and residency.

Furthermore, the participants in the dataset scored mostly healthy levels of blood glucose (below 100) as shown in Fig. 8.

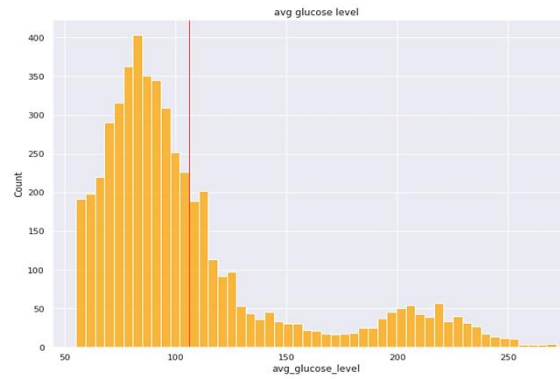


Fig. 8. Distribution of data according to average glucose level.

B. Model Evaluation

After acquiring the data, preprocessing it, and visualizing it, it was used to train and test several classifiers whose role was to predict whether a stroke occurs to a patient or not. The evaluation results for each classifier are presented in Table III.

TABLE III. EVALUATION OF THE DIFFERENT CLASSIFIERS IN TERMS OF ACCURACY, F1 SCORE, RECALL, AND PRECISION

Algorithm	Accuracy	F-1 Score	Recall	Precision
<i>KNN</i>	0.9633	0.98	1.00	0.97
<i>SVM</i>	0.9674	0.98	1.00	0.97
<i>Decision Tree</i>	0.9674	0.98	1.00	0.97
<i>Gaussian NB</i>	0.8655	0.93	0.89	0.97
<i>Random Forest</i>	0.96741	0.98	1.00	0.97
<i>Voting Classifier</i>	0.9674	0.98	1.00	0.97
<i>Stacking Classifier</i>	0.96741	0.98	1.00	0.97

In addition, these evaluation metrics can be seen in Fig. 9 which clearly illustrates that in fact, all of the proposed algorithms in this study have similar results in predicting the stroke occurrence in patients, except for Naïve Bayes which clearly has the worst performance among these classifiers.

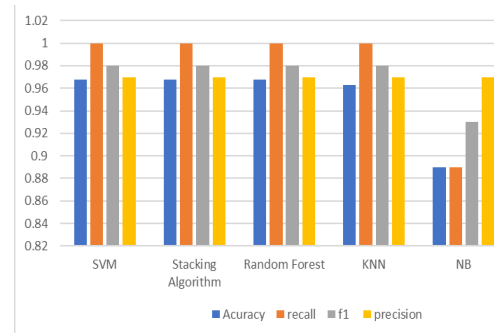


Fig. 9. Evaluation of different algorithms according to several evaluation metrics.

However, taking into consideration that the stacking classifier is an ensemble model, it can be said that choosing stacking algorithm might enhance the prediction results in case of stroke.

V. CONCLUSION

Stroke is among the top medical accidents that lead to death but even in the case of survival, stroke leaves serious implications on the lives of its patients. A patient who has previously suffered from brain stroke, shall he remain alive, might suffer the consequences in what seems like paralysis, among many other life-long complications. Since there are several risk factors that enhance the chances of strokes, its prediction beforehand is possible. Machine learning algorithms have been employed for this purpose promising fast and efficient prediction results.

In this study, the aim was to develop the optimal system that can predict stroke occurrence with high accuracy based on several risk factors collected about the patients. Here, multiple machine learning algorithms are implemented such as Naïve Bayes, SVM, Random Forest, KNN, Decision Tree, Stacking, and majority voting to check the results provided by each algorithm. After that, the choice of the optimal algorithm will be made depending on the evaluation results.

After appropriate preparation of the data, it was divided into training and testing parts such that all of the proposed algorithms are tested for their ability of predicting stroke occurrence. The evaluation metrics of choice were accuracy, f1 score, recall value, and precision value. Ultimately, the results showed that the selected algorithms perform quite well in predicting the strokes, such that SVM, DT, RF, KNN, Voting, and stacking classifier almost scored the same values. The algorithms scored 96% accuracy, 0.98 f1 score, 1 recall value, and 0.97 precision value. However, the achieved results suggest that the Naïve Bayes algorithm might not be the best choice for creating a stroke prediction model since it scored less accuracy levels (86%), less f1 score (0.93), less Recall (0.89), but the same precision value (0.97).

REFERENCES

- [1] V. L. Feigin, C. M. M. Lawes, D. A. Bennett, and C. S. Anderson, "Stroke epidemiology: A review of population-based studies of incidence, prevalence, and case-fatality in the late 20th century," *The Lancet Neurology*, vol. 2, pp. 43-53, 2003.
- [2] B. Ovbiagele, L. B. Goldstein, R. T. Higashida, V. J. Howard, S. C. Johnston, O. A. Khavjou, *et al.*, "Forecasting the future of stroke in the United States," *Stroke*, vol. 44, pp. 2361-2375, 2013.
- [3] World Health Organization. Noncommunicable Diseases and Mental Health Cluster, "WHO STEPS stroke manual : the WHO STEPwise approach to stroke surveillance," World Health Organization, Geneva, 2005.
- [4] B. C. V. Campbell, D. A. De Silva, M. R. Macleod, S. B. Coutts, L. H. Schwamm, S. M. Davis, *et al.*, "Ischaemic stroke," *Nature Reviews Disease Primers*, vol. 5, p. 70, 2019.
- [5] R. V. Krishnamurthi, V. L. Feigin, M. H. Forouzanfar, G. A. Mensah, M. Connor, D. A. Bennett, *et al.*, "Global and regional burden of first-ever ischaemic and haemorrhagic stroke during 1990–2010: findings from the Global Burden of Disease Study 2010," *The Lancet Global Health*, vol. 1, pp. e259-e281, 2013.
- [6] S. Kamalakannan, A. S. V. Gudlavalleti, V. S. M. Gudlavalleti, S. Goenka, and H. Kuper, "Incidence & prevalence of stroke in India: A systematic review," *Indian Journal of Medical Research*, vol. 146, pp. 175-185, 2017.
- [7] E. S. Donkor, "Stroke in the 21st century: A snapshot of the burden, epidemiology, and quality of life," *Stroke Research and Treatment*, vol. 2018, p. 3238165, 2018.
- [8] V. L. Feigin, B. Norrving, and G. A. Mensah, "Global burden of stroke," *Circulation Research*, vol. 120, pp. 439-448, 2017.
- [9] M. J. O'Donnell, D. Xavier, L. Liu, H. Zhang, S. L. Chin, P. Rao-Melacini, *et al.*, "Risk factors for ischaemic and intracerebral haemorrhagic stroke in 22 countries (the INTERSTROKE study): A case-control study," *The Lancet*, vol. 376, pp. 112-123, 2010.
- [10] M. Kaur, S. R. Sakhare, K. Wanjale, and F. Akter, "Early stroke prediction methods for prevention of strokes," *Behavioural Neurology*, vol. 2022, p. 7725597, 2022.
- [11] M. Lee, J. Ryu, and D.-H. Kim, "Automated epileptic seizure waveform detection method based on the feature of the mean slope of wavelet coefficient counts using a hidden Markov model and EEG signals," *ETRI Journal*, vol. 42, pp. 217-229, 2020.
- [12] B. Kim, N. Schweighofer, J. P. Haldar, R. M. Leahy, and C. J. Winstein, "Corticospinal tract microstructure predicts distal arm motor improvements in chronic stroke," *Journal of Neurologic Physical Therapy*, vol. 45, pp. 273-281, 2021.
- [13] E. Dritsas and M. Trigka, "Stroke risk prediction with machine learning techniques," *Sensors*, vol. 22, p. 4670, 2022.
- [14] T. Rakshit and A. Shrestha, "Comparative analysis and implementation of heart stroke prediction using various machine learning techniques," *International Journal of Engineering Research & Technology*, vol. 10, pp. 886-890, 2021.
- [15] G. Sailasya and G. L. A. Kumari, "Analyzing the performance of stroke prediction using ML classification algorithms," *International Journal of Advanced Computer Science and Applications*, vol. 12, pp. 539-545, 2021.
- [16] M. U. Emon, M. S. Keya, T. I. Meghla, M. M. Rahman, M. S. A. Mamun, and M. S. Kaiser, "Performance analysis of machine learning approaches in stroke prediction," in *2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, Coimbatore, India, 2020, pp. 1464-1469.
- [17] T. I. Shoily, T. Islam, S. Jannat, S. A. Tanna, T. M. Alif, and R. R. Ema, "Detection of stroke disease using machine learning algorithms," in *2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, Kanpur, India, 2019, pp. 1-6.
- [18] V. Abedi, V. Avula, D. Chaudhary, S. Shahjouei, A. Khan, C. J. Griessenauer, *et al.*, "Prediction of long-term stroke recurrence using machine learning models," *Journal of Clinical Medicine*, vol. 10, p. 1286, 2021.
- [19] C. S. Nwosu, S. Dev, P. Bhardwaj, B. Veeravalli, and D. John, "Predicting stroke from electronic health records," in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Berlin, Germany, 2019, pp. 5704-5707.
- [20] D. Berrar, "Bayes' theorem and naive bayes classifier," in *Encyclopedia of Bioinformatics and Computational Biology*, S. Ranganathan, M. Gribskov, K. Nakai, and C. Schönbach, Eds. ed Oxford: Academic Press, 2019, pp. 403-412.
- [21] S. Alexiou, E. Dritsas, O. Kocsis, K. Moustakas, and N. Fakotakis, "An approach for personalized continuous glucose prediction with regression trees," in *2021 6th South-East Europe Design Automation, Computer Engineering, Computer Networks and Social Media Conference (SEEDA-CECNSM)*, Preveza, Greece, 2021, pp. 1-6.
- [22] P. Cunningham and S. J. Delany, "K-Nearest neighbour classifiers - a tutorial," *ACM Computing Surveys*, vol. 54, p. Article 128, 2021.
- [23] M. B. A. Snousy, H. M. El-Deeb, K. Badran, and I. A. A. Khilil, "Suite of decision tree-based classification algorithms on cancer gene expression data," *Egyptian Informatics Journal*, vol. 12, pp. 73-82, 2011.
- [24] K. G. Dinesh, K. Arumugaraj, K. D. Santhosh, and V. Mareeswari, "Prediction of cardiovascular disease using machine learning algorithms," in *2018 International Conference on Current Trends towards Converging Technologies (ICCTCT)*, Coimbatore, India, 2018, pp. 1-7.
- [25] A. Dogan and D. Birant, "A weighted majority voting ensemble approach for classification," in *2019 4th International Conference on Computer Science and Engineering (UBMK)*, Samsun, Turkey, 2019, pp. 1-6.

Deep Learning Localization Algorithm Integrating Attention Mechanism in Database Information Query

Yang Li, Xianghui Hui, Xiaolei Wang, Fei Yin*

College of Information and Management Science, Henan Agricultural University, Zhengzhou 450046, China

Abstract—This study aims to solve the problems of traditional indoor car search positioning technology in terms of positioning accuracy and functionality. Based on database technology and deep learning technology, an LSTM model with attention mechanism was established. This model can simultaneously extract temporal and spatial features, and use attention mechanism for feature importance recognition. The entire positioning model has been designed as a triple functional entrance that includes positioning, car storage, and reverse car search, enhancing the user's coherent experience. The data results show that the root mean square error of the LSTM (Attention) model designed in the study is 0.216, and the variance is 0.092. Among similar positioning models, the index value is the smallest, while the CDF line rises the fastest and the maximum value is the highest. The research conclusion indicates that the LSTM (Attention) indoor positioning model designed in this study has better computational performance and can help users achieve more accurate positioning and vehicle navigation.

Keywords—LSTM; attention mechanism; positioning; database

I. INTRODUCTION

Indoor vehicle positioning has always been a problem that many people need to solve in real life, and existing positioning technologies often have shortcomings such as significant positioning error and single positioning function. With the continuous maturity of the Internet of Things and intelligent computing technology, its application in the field of smart cities is also deepening. This comprehensive application based on life network interaction can be found in various aspects. Among them, deep learning technology, as an intelligent computing technology, has been widely applied and widely used [1-3]. Deep learning technology relies on database technology and modern network technology, utilizing network technology to achieve interaction between people and things, and utilizing database technology to store, manage, and update information [4-6]. As an intelligent computing technology based on human vision, attention mechanism can extract highly important target regions in global scenes, filter out impurities for the current task, and retain more important regions [7-9]. The main goal of this study is to use deep learning technology to solve the problem of vehicle positioning in the city, improve the accuracy and stability of positioning. This technology can not only be applied to indoor vehicle search, but also to many other indoor positioning problems, making it a positioning research of certain importance. In the future, this indoor vehicle positioning method based on database and deep learning technology can help people find parking spaces more conveniently, improving the efficiency and convenience of urban transportation. To achieve the research objectives, a

LSTM model with attention mechanism was established based on database technology and deep learning technology. This model can simultaneously extract temporal and spatial features. On this basis, the entire positioning model has been designed as a triple functional entrance that includes positioning, car storage, and reverse car search, enhancing the user's coherent experience.

II. RELATED WORKS

With the maturity of intelligent computing technique, the utility of deep learning algorithms in various fields has gradually deepened. Bédard et al. [10] applied deep learning algorithms to the medical field, and constructed a mouse pattern of sodium dextran sulfate colitis through convolutional neural networks. The study mainly used medical slide pictures as the main samples to establish training dataset, and based on this network pattern is trained. The pattern can improve overall efficiency of medical compound screening. Yu et al. [11] applied the deep learning algorithm to the prediction of the operating load of the power system, and selected the short-term load prediction suitable for the machine learning scheme as the main research direction, and designed a deep learning pattern for the nonlinear problem. The pattern can significantly improve the prediction efficiency. Shen et al. [12] applied the convolutional neural network algorithm to the short-term prediction problem of subway traffic passenger flow, and combined the auto-encoder to perform convolutional coding summation to solve the passenger flow matrix problem. At the same time, the back-propagation strategy was used to solve the optimization problem of matrix weight parameters. It shows that the technique can predict more accurately than the traditional pattern, and the pattern has a significant influence on the forecast prediction of passenger flow between stations. Tang et al. [13] proposed a low-cost image acquisition strategy that combines UAVs with intelligent algorithms, through which researchers can conduct more efficient and low-cost ground image acquisition. The strategy mainly uses a convolutional neural network with 12 layers to layer and recognize images. The research results show that this strategy is less time-consuming and more accurate than other strategies, and is more suitable for ground sampling. Wang et al. [14] proposed a fault localization pattern based on multi-class features, which constructs a ranking pattern from the perspectives of depth pattern and breadth pattern, and explores the relationship between different features. It means pattern designed in the study has a greater performance advantage in early troubleshooting than the traditional pattern. This research also adopts the deep learning method to design the localization

pattern. The research also integrates attention mechanism in the pattern design.

Attention mechanisms are used in different types of pattern designs. Zheng et al. [15] combined attention mechanism with 3D-network. By embedding the attention mechanism into the first level of the convolutional neural network, the band attention vector in the band selection process was learned, and then the plant Prediction of chlorophyll content. The research results show the pattern has higher classification accuracy than other types of patterns such as genetic patterns. Lai et al. [16] designed an attention application mechanism based on convolutional neural network that is, designing an excitation module after the convolutional layer of the convolutional neural network, and then forming an attention convolutional network pattern to emphasize important features. Research results show that the pattern can minimize the number of redundant candidates, thereby improving the efficiency of the pattern. Xiao et al. [17] combined an attention mechanism learning pattern with a bilinear efficiency network for the identification and classification of radar jamming signals. It shows that the bilinear network pattern based on the attention mechanism can extract signal features more efficiently and achieve the effect of automatic signal interference type classification. Wang et al. [18] proposed a temporal convolutional neural network pattern that combines soft threshold technology and attention mechanism, and applied it to the prediction of machine remaining life. The pattern does not need to perform overall feature preprocessing, but directly Multi-channel sensors for data input and retain valid features during feature extraction. It shows that the prediction performance of this pattern is more outstanding better than similar patterns. Shobana and Murali [19] designed an automatic review pattern based on attention mechanism for the difficult problem of manual review of online customer reviews. The pattern can perform abstract review through encoding-decoding and solve the problem of manual repetitive labor. The review efficiency of this pattern is more outstanding than that of similar patterns. This research designs a database-based indoor positioning pattern design. The pattern integrates the characteristics of the attention mechanism, making the indoor positioning more accurate and more practical.

III. ATTENTION-LSTM LOCALIZATION PATTERN DESIGN

A. Attention-LSTM Car-Finding and Positioning Pattern Framework Design

The indoor positioning pattern researched and designed is mainly based on the database-based position fingerprint indoor vehicle positioning pattern. The pattern mainly searches the data stored in the database and the position fingerprint in the positioning area through the fingerprint positioning strategy, and finally achieves indoor positioning. In order to adapt to the search strategy of fingerprint positioning, the pattern must have a certain intelligent learning ability, so as to ensure the accuracy and efficiency of the query in the query process. The research mainly combines the LSTM (Long Short-Term Memory) temporal network with attention mechanism to design the pattern for deep learning, and the pattern framework is mainly designed in combination with the main functional requirements of indoor car-finding and positioning. The

Attention-LSTM car-finding and positioning pattern designed by the research can be mainly divided into three functional parts, namely, the car-finding navigation module, the vehicle temporary storage module and the reverse car-finding module formed according to the user's car-finding and positioning requirements. The specific functional flow of the car-finding navigation module and the vehicle temporary storage module is shown in Fig. 1.

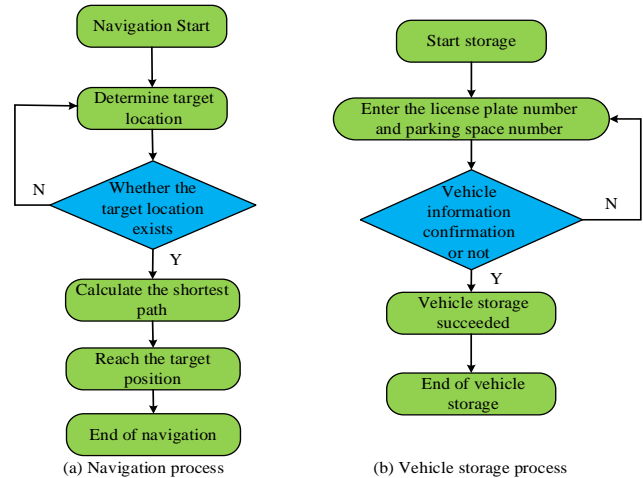


Fig. 1. Car search navigation module and vehicle temporary storage module.

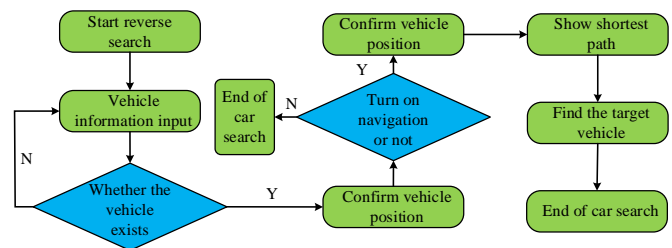


Fig. 2. Reverse car search module.

The sub-figure (a) in Fig. 1 describes the car-finding navigation module. The car-finding navigation module mainly serves to receive the user's query target signal, and to determine whether the query target position actually exists. Once the module determines that there is a target position, the A* (A-Star) algorithm can be used for shortest path navigation. The A* algorithm is a search algorithm for the shortest path. The algorithm converts the path planning into a graph search problem, and converts the overall search area into a set of two-dimensional map arrays to complete the transformation and search of elements in the area, and then converts the entire search area into a set of two-dimensional map arrays. After the search is completed, the shortest distance from the target starting point to the target destination will be displayed according to the distribution of obstacles in the map. The sub-picture (b) in Fig. 1 is the vehicle temporary storage module, which is mainly responsible for meeting the needs of users in the vehicle storage process. The corresponding location fingerprint can be formed, which is convenient for the user to perform subsequent reverse car search. The operation mode of the reverse car search module is shown in Fig. 2.

The reverse car-finding module is a module for car-finding and navigation based on the location fingerprint. In the pattern, the location fingerprint of the vehicle is mainly composed of two parts of information, namely vehicle information and positioning information. The search path associated with the location. Therefore, in the reverse car storage module, the pattern will first confirm the vehicle information to confirm whether the vehicle actually exists. After confirming the real existence of the vehicle, the pattern will further determine the vehicle location according to the search path. Once the vehicle position is determined successfully, the navigation part can be entered, and the shortest path navigation can be carried out through the A* algorithm. Since the indoor positioning of the pattern mainly adopts the indoor Bluetooth technology, it is necessary to arrange the Bluetooth beacons according to the indoor environment and the characteristics of the positioning algorithm. However, in this process, the strength evaluation of the received signal needs to be solved. The research mainly uses the logarithmic normal path pattern for evaluation. The evaluation of signal propagation distance and received signal strength variation, the specific calculation process is shown in Equation (1):

$$RSSI(d) = RSSI(d_0) - 10 * n * \lg \frac{d}{d_0} + N_{\sigma} \quad (1)$$

In Equation (1), the *RSSI* received signal strength is *d* expressed, and the distance between the signal transmitter and the receiver is expressed. This distance is often expressed by the closest distance of signal propagation, which is the *d₀* reference distance, the *n* attenuation factor, and *N_σ* the random noise.

B. Attention-LSTM Positioning Mechanism Design

Indoor positioning strategies based on deep learning patterns often use neural networks as a feature extraction tool, through which effective features of received signals are extracted, and location estimation is performed based on the features. However, when most neural network patterns perform feature extraction and position calculation, the calculation of position information lacks time, that is, the temporal continuity of signals in the positioning process cannot be fully calculated [20-21]. Therefore, when designing the indoor vehicle positioning mechanism, the research mainly selects the LSTM pattern as the basic positioning pattern. Unlike other neural networks, the LSTM pattern itself has a strong processing ability for time series. In the case of introducing the attention mechanism, its consideration Feature selection with screening is performed simultaneously with the dual features of spatial and temporal features, as shown in Fig. 3.

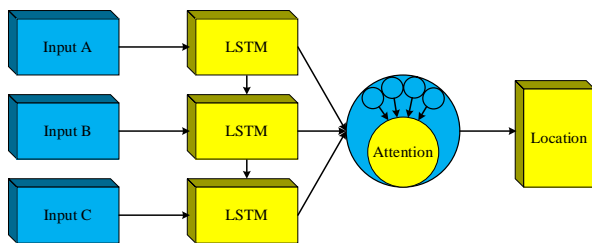


Fig. 3. Basic framework of the pattern.

Fig. 3 describes how the attention mechanism and the LSTM algorithm are combined in the overall pattern. The pattern will input information A, information B, and information C into the LSTM pattern respectively after receiving information A, information B, and information C. These three signals contain their own spatial characteristics, and they also have time domain continuous characteristics between each other. After the LSTM pattern extracts its features, it will be introduced into the attention mechanism module. The attention mechanism module will target the importance of the signal, and finally form an effective feature unit with temporal and spatial features. The feature unit is the main basis for calculating the position coordinates. The LSTM pattern itself is evolved based on the network pattern. The network pattern has specialized processing capabilities for sequence data. The calculation equation for single data is shown in Equation (2):

$$y = f(Wx + b) \quad (2)$$

In Equation (2), the cyclic network *x* performs a linear transformation on a single input value, and the linear transformation *Wx + b* is expressed *y* as *f(·)* When facing a series of data, the calculation method of the network pattern state at the current moment is as shown in Equation (3).

$$s_t = f(U \cdot x_t + W \cdot s_{t-1}) \quad (3)$$

In Equation (3), *U* and *W* are all pattern parameters that need to be learned, *x_t* representing the input value at a certain moment, *s_{t-1}* representing the network state at a certain moment in the past, which *t* is a time expression. The calculation method of the pattern output value at the current moment is shown in Equation (4).

$$o_t = g(V \cdot s_t) \quad (4)$$

Equation (4) is *V* the pattern parameter that needs to be learned. The LSTM pattern adds a transfer state to the RNN pattern, as shown in Fig. 4.

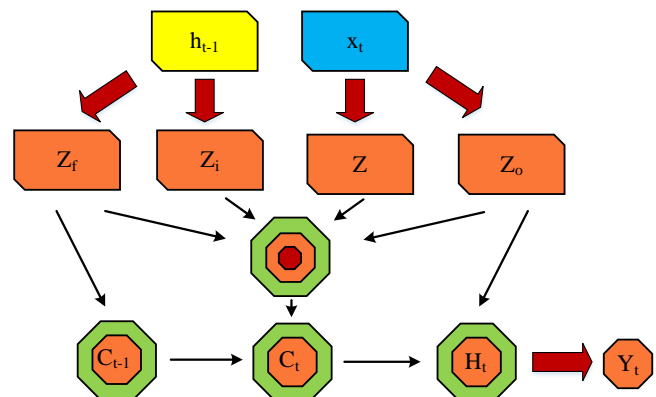


Fig. 4. LSTM pattern construction.

In Fig. 4, the input variable is composed of two different element states. The element states are combined to form a

splicing vector, and the splicing vector forms four different types of outputs after four different operations, which are Z_f , Z_i , Z , Z_o , where Z_f , Z_i , Z_o are all three gated states, and Z represent the updated state value after the unit is updated. Z The calculation method is shown in Equation (5).

$$Z = \tan h(w[x_t, h_{t-1}]) \quad (5)$$

Weight parameter h_{t-1} is represented in Equation (5), and w , x_t respectively represent two different information states of the input variable. The calculation method of the forget gate Z_f is shown in Equation (6).

$$Z_f = \sigma(w_f[x_t, h_{t-1}]) \quad (6)$$

In Equation (6), the σ activation function is expressed. The input gate Z_i is calculated as shown in Equation (7).

$$Z_i = \sigma(w_i[x_t, h_{t-1}]) \quad (7)$$

Output gate Z_o is calculated as shown in Equation (8).

$$Z_o = \sigma(w_o[x_t, h_{t-1}]) \quad (8)$$

Under the attention mechanism, the pattern can extract information that is more critical to the current target from the global information. In a time-series-related environment, the influence of input elements at different times on output elements. At the same time, the attention mechanism change applies more attention to the input features with high influence. The flow of the LSTM (attention) localization pattern formed under the attention mechanism is shown in Fig. 5.

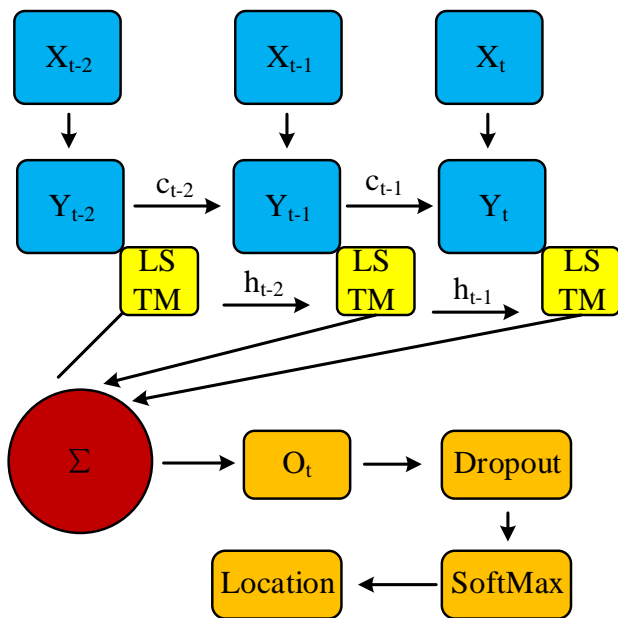


Fig. 5. Attention LSTM positioning pattern process.

Assuming that the input element values in Fig. 5 come from a given input sequence, and the vectors in the sequence are all one-dimensional vectors, the sequence is expressed as shown in Equation (9).

$$\begin{cases} X = [x_{t-2}, x_{t-1}, x_t] \\ x_i = [r_1, r_2, r_3, \dots, r_k] \end{cases} \quad (9)$$

In Equation (9), x_{t-2} , x_{t-1} and x_t respectively represent the input signal at three consecutive time points, and t represent a certain moment, i represent the general name of different moments, and represent x_i a one-dimensional vector in the signal. After the NN structure operation of LSTM, the pattern will obtain the output vector Y , which is expressed as Equation (10).

$$\begin{cases} Y = [y_{t-2}, y_{t-1}, y_t] \\ y_i = [v_1, v_2, v_3, \dots, v_n] \end{cases} \quad (10)$$

A one-dimensional vector is represented in Equation (10). y_i Under the action of the attention mechanism, the pattern will obtain the importance vector in this part $\hat{\delta}$, and the specific equation is shown in Equation (11).

$$\begin{cases} \hat{\delta} = [\alpha_{t-2}, \alpha_{t-1}, \alpha_t] \\ \alpha_i = [u_1, u_2, \dots, u_n] \end{cases} \quad (11)$$

A one-dimensional vector is represented in Equation (11). α_i The final calculation equation of spatiotemporal features is shown in Equation (12).

$$o_t = \sum_{i=0}^2 \alpha_{t-i} \square y_{t-1} \quad (12)$$

The output network of the pattern is mainly composed of Droup level and SoftMax level. The Droup level is mainly used to solve the overfitting problem of the pattern, and the Droup leve randomly disconnects part of the network connection to prevent the occurrence of overfitting. The SoftMax layer is mainly used for multi-classification of the positioning problem, and the final position probability can be written in the form of Equation (13).

$$\sum_{i=0}^n P_i = 1 \quad (13)$$

In Equation (13), i represents the position, and P_i represents the probability of being located at the position. The final position of the pattern output can be expressed in two ways. The first way is the maximum probability way, as shown in Equation (14).

$$\begin{cases} p_t = \max(p) \\ position_x = x_t \\ position_y = y_t \end{cases} \quad (14)$$

The position in Equation (14) is $position = (position_x, position_y)$ expressed in terms of positioning coordinates. The second representation method is the weighted average representation method, as shown in Equation (15).

$$\begin{cases} position_x = \sum_{i=0}^n p_i \cdot x_i \\ position_y = \sum_{i=0}^n p_i \cdot y_i \end{cases} \quad (15)$$

IV. LSTM-CNN POSITIONING PATTERN POSITIONING EFFECT ANALYSIS

A. Pattern Performance Analysis

When analyzing the localization effect of the LSTM-CNN localization pattern, the research will analyze from the two perspectives of the LSTM-CNN pattern's operational performance and the pattern's localization effect. In the data analysis, the research mainly uses the underground garage on the first floor below ground as the main data collection site. In the study, the beacons were evenly arranged in the indoor environment, and the coordinate positions of the signals were recorded. After the data is obtained, 80% of the experimental data is training set, and the other 20% is test set. Since the LSTM pattern requires information sequence input, the research assumes that the order size is 3 when selecting the data set, so the available data sets are combined as shown in Table I.

Based on the data set setting, the research will explore whether adding attention mechanism to the pattern will affect the performance of the pattern. In the comparison process, the research uses four patterns: LSTM pattern, LSTM (Mean pooling) pattern, LSTM (Max pooling) pattern, and LSTM-attention pattern, the variance of four indicators to analyze, the specific data results are shown in Fig. 6.

TABLE I. DATA SET COMBINATION

Order size 1		Order size 2		Order size 3	
Signal order	Position representation n	Signal order	Position representation n	Signal order	Position representation n
x_0	y_0	$[x_0, x_1]$	y_1	$[x_0, x_1, x_2]$	y_2
x_1	y_1	$[x_1, x_2]$	y_2	$[x_1, x_2, x_3]$	y_3
x_2	y_2	$[x_2, x_3]$	y_3	$[x_2, x_3, x_4]$	y_4
...

x_{n-1}	y_{n-1}	$[x_{n-2}, x_{n-1}]$	y_{n-1}	$[x_{n-3}, x_{n-2}, x_{n-1}]$	the_{n-1}
x_n	the_n	$[x_{n-1}, x_n]$	the_n	$[x_{n-2}, x_{n-1}, x_n]$	the_n

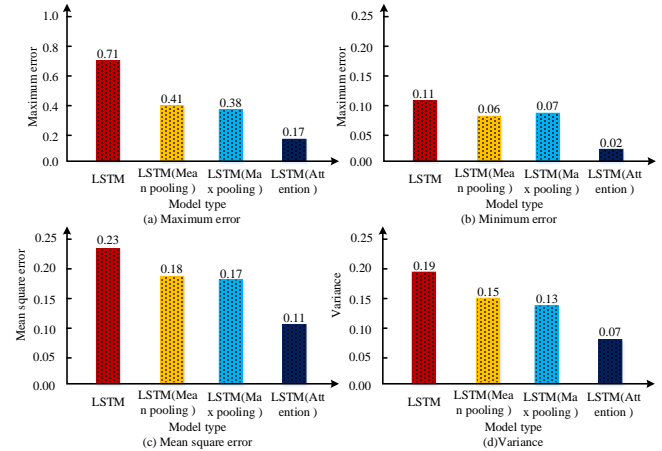


Fig. 6. Performance index comparison.

In Fig. 6, according to maximum error, the maximum error value of the data obtained by the LSTM pattern is 0.71, the max-error of data obtained by the LSTM (Mean pooling) pattern is 0.41, and the max-error of data obtained by the LSTM (Max pooling) pattern is 0.41. The maximum error value of the data obtained is 0.38, while the maximum error value of the data obtained by the LSTM-attention pattern designed by the research is 0.17. The max-error of the data obtained by the LSTM-attention pattern is the smallest, much less than The other three patterns; in terms of minimum error, the maximum error value of the data obtained by the LSTM pattern is 0.11, the maximum error value of the data obtained by the LSTM (Mean pooling) pattern is 0.06, and the data obtained by the LSTM (Max pooling) pattern. The maximum error value of the LSTM-attention pattern is 0.07, and the maximum error value of the data obtained by the LSTM-attention pattern designed by the research is 0.02. The min-error of the data obtained by the LSTM-attention pattern is also the smallest, much smaller than the other three. According to mean square error, the mean square error of the data obtained by the LSTM pattern is 0.23, the mean square error of the data obtained by the LSTM (Mean pooling) pattern is 0.18, and the mean square error of the data obtained by the LSTM (Max pooling) pattern. The mean square error is 0.17, and the mean square error of the data obtained by the LSTM-attention pattern designed by the research is 0.11. The mean square error of the data obtained by the LSTM-attention pattern is the minimum value, which is much smaller than the other three. In terms of variance, the variance difference of the data obtained by the LSTM pattern is 0.19, the variance of the data obtained by the LSTM (Mean pooling) pattern is 0.15, and the variance of the data obtained by the LSTM (Max pooling) pattern is 0.13, while The variance of the data obtained by the LSTM-attention pattern designed in the research is 0.07. It can be seen that the variance of the data obtained by the LSTM-attention pattern is the minimum value, which is much smaller than the other three

patterns. In summary, the LSTM-attention pattern designed by the research has the smallest values in the four indicators of maximum error, minimum error, mean square error, and variance. LSTM pattern shows stronger capability in both accurate and stable performance. In addition, the research starts from the perspective of different order sizes of data input, analyzes the accuracy changes of the patterns under different order sizes, and uses the CDF (Cumulative Distribution Function) to compare the accuracy. The specific data differences are shown in Fig. 7.

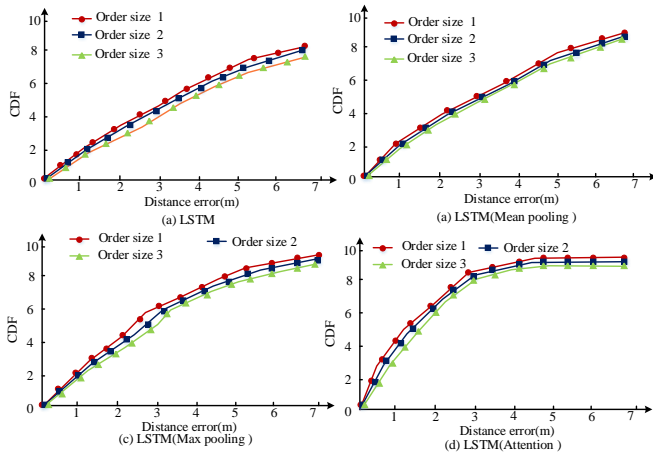


Fig. 7. Cumulative probability distribution function.

In Fig. 7, the CDF values of the order size 1, order size 2, and order size 3 of the LSTM pattern all show a uniform upward trend. When the distance error reaches 7, order size 1, order size 2. The CDF value of order size 3 reaches around 80%. It can be seen that the rising trend of CDF value is slow, and the highest CDF value does not reach 90%; order size 1, order size 2, order size of LSTM (Mean pooling) pattern 3 The CDF values of the three data order sizes also show an upward trend. Although the upward trend is also uniform, when the distance error reaches 7, the CDF values of order size 1, order size 2, and order size 3 reach more than 90%. It can be seen that The accuracy of the LSTM (Max pooling) pattern is better than that of the LSTM pattern; in addition, the CDF values of the order size 1, order size 2, and order size 3 of the LSTM (Max pooling) pattern also show an upward trend. The upward trend is faster than that of the LSTM (Mean pooling) pattern in the first half of the rise, and the speed of reaching the high CDF value area is faster. When the distance error reaches 7, the CDF values of order size 1, order size 2, and order size 3 are the same. When it reaches more than 90%, it can be seen that the accuracy of the LSTM (Max pooling) pattern is more outstanding than that of the LSTM (Mean pooling) pattern; the LSTM-attention pattern designed by the study has three order sizes: 1, order size 2, and order size 3. The CDF values of the data order size all show an upward trend. It can be seen that the first half of the data line of the LSTM-attention pattern has the fastest upward trend, while the second half of the line has a relatively stable trend. When the distance error reaches 3, the order size is 1, the sequence The CDF values of length 2 and order size 3 both reach more than 90%, and are close to 100% when the distance error reaches 5. The overall precision of

LSTM-attention pattern designed by the research is higher and the performance is better, indicating that the attention mechanism can indeed significantly improve the computing function of the LSTM pattern.

B. Quantitative Analysis of Positioning Effect

In the analysis of the positioning effect of the overall pattern, the research mainly adopts the method of comparing the positioning effects of different positioning patterns. First, the positioning accuracy and stability of different patterns are compared, and then the positioning and navigation routes are compared. The comparison of positioning accuracy and stability data is shown in Fig. 8.

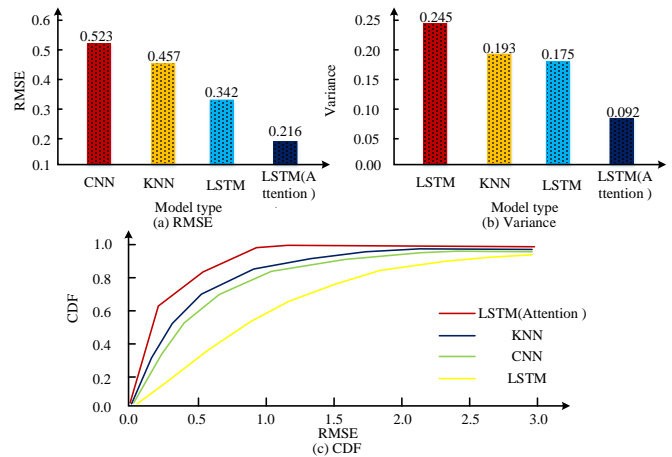


Fig. 8. Pattern comparison.

As can be seen from Fig. 8, according to root mean square error, the root mean square error of the CNN pattern is 0.523, the root mean square error of the KNN pattern is 0.457, the root mean square error of the LSTM pattern is 0.342, and the LSTM-attention pattern The root mean square error is 0.216. From the comparison, the root mean square error of the LSTM-attention pattern is the smallest; in terms of variance value, the variance value of the CNN pattern is 0.245, the variance value of the KNN pattern is 0.193, and the variance of the LSTM pattern is 0.175. The figure is 0.175, and the variance figure of LSTM-attention pattern is 0.092. From the comparison point of view, the variance figure of LSTM-attention pattern is the smallest; in terms of CDF value changes, it can be seen that among the four patterns, the fastest rising speed is LSTM. (Attention) pattern, and in the final accuracy effect presentation, the LSTM-attention pattern that shows the highest CDF value is also the LSTM-attention pattern, followed by the KNN pattern, the CNN pattern again, and the LSTM pattern. Although the KNN pattern and the CNN pattern show a higher rising speed and a better rising trend, they still show certain disadvantages compared with the LSTM-attention pattern designed by the research. Compared with else patterns, the LSTM-attention indoor positioning pattern designed by the research has higher positioning accuracy and positioning stability, and can reflect better results in practical applications. The research also compares the positioning and navigation routes of the four patterns, and the comparison results is Fig. 9.

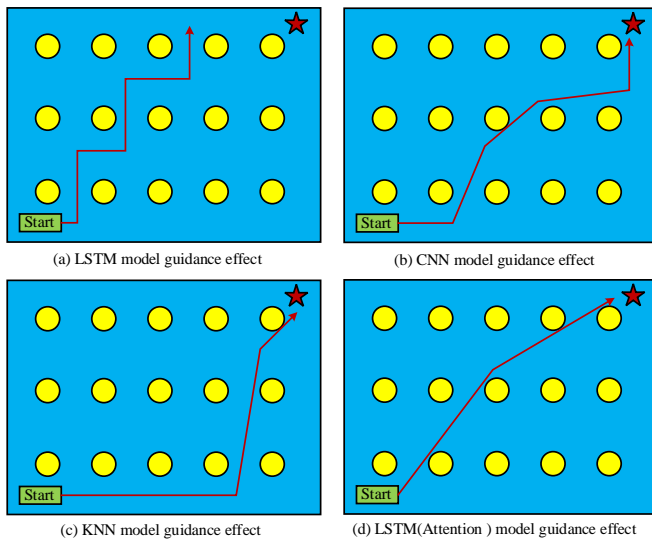


Fig. 9. Positioning and navigation effect.

In Fig. 9, the navigation route of LSTM pattern shows the shape of multiple right-angle corners, the overall navigation route is relatively complex, and the final navigation target point deviates and fails to reach the target point accurately. Both the positioning ability and the navigation ability are insufficient; the navigation route of the CNN pattern shows the form of right-angle corners and oblique interspersed routes. The overall navigation route is also more complicated, but the pattern locates the target point more accurately, and the final navigation route is completed. more precise guidance. However, it can be seen from the route planning that the route planning is redundant and not simple and direct; the navigation route of the CNN pattern also shows the form of right-angle corners and obliquely interspersed routes, and the overall route planning is simpler and more direct. It shows relatively accurate target point positioning ability; the navigation route of the LSTM-attention indoor positioning pattern designed by the research shows an accurate oblique interspersed route form. Among the four patterns, the navigation route form is the most concise, and finally reaches the target point accurately. It can be seen that compared with the other three positioning patterns, the LSTM-attention indoor positioning pattern designed in the study has the best positioning effect and navigation effect, and can obtain the best effect in actual indoor positioning and navigation.

V. CONCLUSION

To solve the issue of insufficient accuracy of indoor spatial positioning in traditional methods, an indoor positioning pattern with temporal and spatial characteristics is established based on LSTM pattern, and on this basis, the attention mechanism is integrated into the pattern, so that the pattern can Features are effectively screened, and features with higher importance are retained for more precise positioning and navigation. The research shows the LSTM-attention pattern designed in the study has higher index values of 0.17, 0.02, 0.11, and 0.07, in line with maximum error, minimum error, mean square error, and variance, respectively, than other LSTM patterns that do not incorporate the attention

mechanism. The global operation effect of the LSTM-attention pattern is more outstanding. At the same time, the CDF data trend that the LSTM-attention pattern rises faster and the highest value is the highest. In addition, in terms of positioning and navigation effect, the root mean square error and variance of the LSTM-attention pattern designed by the study are 0.216 and 0.092 respectively, the index value is the smallest, the CDF curve rises the fastest, and the maximum value is the highest. The navigation route of the LSTM-attention pattern is the most direct and accurate. The LSTM-attention pattern is more stable and accurate than similar patterns, with better positioning and navigation effects.

DATA AVAILABILITY STATEMENT

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

CONFLICTS OF INTEREST

It is declared by the authors that this article is free of conflict of interest.

REFERENCES

- [1] Ramot Y, Deshpande A, Morello V, et al. "Microscope-based automated quantification of liver fibrosis in mice using a deep learning algorithm," *Toxicologic Pathology*. 2021; 49(5): 1126-1133.
- [2] Yu W, Rao E, Chin C, et al. "A preliminary study of deep-learning algorithm for analyzing multiplex immunofluorescence biomarkers in body fluid cytology specimens," *Acta Cytologica*. 2021; 65(4): 348-353.
- [3] Park HJ, Jang I, Ko KE. "Meal intention recognition system based on gaze direction estimation using deep learning for the user of meal assistant robot," *Journal of Institute of Control*. 2021; 27(5): 334-341.
- [4] Roberts A. "Caffeine: An evaluation of the safety database," *Nutraceuticals (Second Edition)*. 2021; 501-518.
- [5] Arputharaj V, Chezian RM. "Data mining with human genetics to enhance gene based algorithm and DNA database security," *International Journal of Computer Engineering & Technology*. 2020; 4(3): 176-181.
- [6] Kopaczka M, Kolk R, Schock J, et al. "A thermal infrared face database with facial landmarks and emotion labels," *IEEE Transactions on Instrumentation & Measurement*. 2019; 68(5): 1389-1401.
- [7] Hu H, Lin Z, Hu Q, et al. "Attention mechanism with spatial-temporal joint model for traffic flow speed prediction," *IEEE Transactions on Intelligent Transportation Systems*. 2022; 23(9):16612-16621.
- [8] Sun M, Lyu C, Han Y, et al. "Weakly supervised surface defect detection based on attention mechanism," *Journal of Computer-Aided Design & Computer Graphics*. 2021; 33(6): 920-928.
- [9] Sangeroki BA, Cenggoro TW. "A fast and accurate pattern of thoracic disease detection by integrating attention mechanism to a lightweight convolutional neural network," *Procedia Computer Science*. 2021; 179(11): 112-118.
- [10] Bédard A, Westerling-Bui T, Zuraw A. "Proof of concept for a deep learning algorithm for identification and quantification of key microscopic features in the murine pattern of DSS-induced colitis," *Toxicologic Pathology*. 2021; 49(4): 897-904.
- [11] Yu W, Rao E, Chin C, et al. "A preliminary study of deep-learning algorithm for analyzing multiplex immunofluorescence biomarkers in body fluid cytology specimens," *Acta Cytologica*. 2021; 65(4): 348-353.
- [12] Shen L, Shao Z, Yu Y, et al. "Hybrid approach combining modified gravity pattern and deep learning for short-term forecasting of metro transit passenger flows," *Transportation Research Record*. 2021; 2675(1): 25-38.
- [13] Tang F, Zhang D, Zhao X. "Efficiently deep learning for monitoring *Ipomoea cairica* (L.) sweets in the wild. *Mathematical Biosciences and Engineering*," 2021; 18(2): 1121-1135.

- [14] Wang TT, Yu HL, Wang KC, et al. "Fault localization based on wide & deep learning pattern by mining software behavior," *Future Generation Computer Systems.* 2022; 127: 309-319.
- [15] Zheng Z, Liu Y, He M, et al. "Effective band selection of hyperspectral image by an attention mechanism-based convolutional network," *RSC Advances.* 2022; 12(14): 8750-8759.
- [16] Lai KD, Nguyen TT, Le TH. "Detection of lung nodules on CT images based on the convolutional neural network with attention mechanism," *Annals of Emerging Technologies in Computing.* 2021; 5(2): 78-89.
- [17] Xiao Y, Zhou J, Yu Y, et al. "Active jamming recognition based on bilinear EfficientNet and attention mechanism," *IET Radar, Sonar & Navigation.* 2021; 15(9): 957-968.
- [18] Wang Y, Deng L, Zheng L, et al. "Temporal convolutional network with soft thresholding and attention mechanism for machinery prognostics," *Journal of Manufacturing Systems.* 2021; 60: 512-526.
- [19] Shobana J, Murali M. "Abstractive review summarization based on improved attention mechanism with pointer generator network pattern," *Webology.* 2021; 18(1): 77-91.
- [20] Rui Z, Yan R, Chen Z, et al. "Deep learning and its applications to machine health monitoring," *Mechanical Systems and Signal Processing.* 2019; 115: 213-237.
- [21] Allam JP, Samantray S, Behara C, et al. "Customized deep learning algorithm for drowsiness detection using single-channel EEG signal," *Artificial Intelligence-Based Brain-Computer Interface.* 2022: 189-201.

Develop an Olive-based Grading Algorithm using Image Processing

Dongliang Jin

School of International Business, Zhejiang Guangsha Vocational and Technical University of Construction,
Dongyang 322100, China

Abstract—Olives come in a number of external and internal varieties. The Shengeh kind, which is available in three colours—green, brown, and black—was chosen at random by the researchers to ensure that the sample was diverse. To avoid discoloration throughout the experiment, 150 healthy olives were harvested and stored correctly. These olives had not been subjected to any external harm, such as crushing or milling. The particular kind of olives were kept chilled at 2°C and preserved in water. This study investigates the possibility of grading Shengeh cultivars from olives that have different uses, based on color using image processing. After preparing images of olives using MATLAB software and image processing techniques, olives are graded based on their color in three categories: immature with green, semi-ripe with brown, and ripe with black. The results showed that image processing technology can be used to grade olives of the Shengeh type in terms of their ripeness as a single-color grain with acceptable accuracy. The HSV color space is one of the best color spaces to separate the colors of the olive cultivar. The accuracy of the software for detecting olives with the mentioned degrees is 98%, 96%, and 100%, respectively.

Keywords—Image processing; grading; color; olive; MATLAB; HSV

I. INTRODUCTION

Olives are native to the Mediterranean and sub-Mediterranean climates, and altitude has a definite effect on their composition. Olives grow well at altitudes of 400 to 2000 meters depending on the climate [1,2]. *Olea europaea* is almost a small tree species of the Oleaceae family and is widely distributed from the Mediterranean, North Africa, South Asia, North to South China, Scotland, and Eastern Australia. The olive tree is evergreen and has small, solid leaves that are stacked on top of each other [3]. The fruit of the olive plant is of the shaft type. The most famous species of this plant is called *Olea europaea*, which used since ancient times to prepare olive oil and to eat its fruit. Olive trees grow very slowly [4], even when they grow freely without pruning, but when they are able to grow normally over several years, sometimes their trunk diameters increase dramatically.

One of the salient features of olives is their ability to adapt to different geographical areas. In modern olive cultivation, in order to select suitable cultivars, a set of traits such as fertility, resistance to pests and unfavorable natural conditions, high oil content, good oil quality, suitable size and shape, and high ratio of meat to the kernel are important. For example, in Iran, a number of cultivars have survived under local conditions, but on the other hand, due to being confined to a specific area, there has not been much genetic diversity [5]. Most of the olive

groves in Iran are composed of local yellow and oily cultivars and Fish mi, Shengeh, and Mari cultivars are also observed at limited levels.

Today, the quality of appearance of fruits and agricultural products cannot be evaluated using color methods using traditional methods. Olive is a valuable product that has nutritional and medicinal value and has different cultivars in different regions with different climates that are offered to the consumer market both as canned products and as its products (Eaton oil). Grading of olive products, in order to be used in the production of various products, is always considered by food industry factories. Most existing traditional olive grading systems have the disadvantages of reduced efficiency, low grading speed, and high cost [6,7]. Therefore, it is important to increase the speed and reduce the cost in the grading and recognition system of size, color, shape, and defects of olives; one of the procedures used for this purpose is the use of machine vision and image processing.

Image processing, computer image analysis, or more precisely, the ability of a machine to perceive what it sees around it simultaneously without the need of human observation. This technology is actually one of the branches of electronic engineering that aims to achieve a kind of robotic eye. This technology is used in various fields such as agriculture, medicine, automated industrial production, remote sensing of machine guides, robots, and of course the food industry. One of the important features of the car's visual system, in addition to its high speed and accuracy, and economic efficiency, is the possibility of examining materials in all types of visible and invisible light (infrared, ultraviolet, etc.). Exploration and development of machine vision methods for grading and fruit separation operations accelerate new techniques for estimating the quality characteristics of agricultural products [8,9], due to the importance of olives as valuable products with high nutritional, medicinal, and cost values. Traditional systems use machine vision and image processing in isolation, grading, and other applications that are cost-effective and have higher speeds.

The machine vision system is a non-destructive and scientific method for measuring the spectrum of colors at non-uniform levels. Usually, the color factor is examined to evaluate the quality of agricultural products by image processing. Among the physical characteristics of agricultural products, color is expressed as the most important appearance properties of products [10]. Color is the most important property of an image; in fact, everything in an image is a component of the color stored in the pixels of the image [11-

13]; based on the fact that each color can be reconstructed based on a combination of three main colors. Placing the image in different color spaces and calculating the mean and standard deviation of color intensity in image pixels, in different color spaces, the color information of the image can be extracted.

Traditional olive grading methods have many uses for a variety of purposes, but they are time-consuming and costly. In addition, the performance of these methods is not guaranteed. These factors make it possible to develop practical methods of grading olives. Olive color image processing is one of these methods, which is non-destructive, efficient, and cost-effective through the methods of computer vision systems, and also provides more stable results. In general, computer vision systems are based on olive imagery; even while crossing the production line and then processing the image and analyzing the image's work. Computer vision systems not only recognize the size, shape, color, and texture of olives but also determine the numerical properties of the olive with the shooting scene. The use of machine vision systems in olive processing factories in addition to the separation and grading of high quality and low-quality olives in terms of color, size leads to the development and design of new olive grading systems in olive product production lines for various uses that can be mentioned as one of the applications of researches [19-25].

This study investigates the possibility of grading Shengeh cultivars from olives that have different uses, based on color using image processing, and the question that image processing is the ability to provide an algorithm for grading cultivars, canned and oily olives for increasing the speed of grading and reducing waste in each of the production lines is responsible. The main objectives of this paper are to investigate the possibility of presenting the Schengen diet olive grading algorithm based on color using image processing. In addition, the possibility of developing an olive grading system using artificial intelligence (artificial neural networks or emphysema) is being investigated.

In summary, this paper has the following contributions:

- 1) Proposing an image processing algorithm to olives' grading of the Shengeh type in terms of their ripeness using intensity-level based on HSV color feature analysis.
- 2) Developing an automated olive grading system in order to tackle the interior and exterior feature analysis of the different types of olives.
- 3) Conducting extensive experiments and performance analysis to evaluate and validate the efficacy of the proposed algorithm.

The remainder of this paper is consists of Section II which presents the related works, Section III that describes materials and methods, Section IV presents results and discussion. Finally, conclusion is presented in Section V.

II. RELATED WORKS

Pronce et al. [22] used a computer vision-based system to automatically count and calculate the quantity and size of individual olive fruits. The suggested method separates the olives from the background using image processing techniques, and then utilises machine learning algorithms to calculate an

estimate of the olives' mass and size based on attributes taken from the segmented pictures. The findings demonstrate the efficacy and precision of the suggested approach, with a counting accuracy of 97.5% and a size and mass estimation error of less than 5%. The efficiency and precision of the procedures involved in olive harvesting, grading, and quality control might all be enhanced by using this technique. However, this work is conducted under controlled laboratory conditions.

This research [23] suggested a 3D imaging-based vision-based system for automatically evaluating mangoes according to their volume. The system takes several 2D pictures of the mangoes from various perspectives, and then it uses stereo vision and form from shading algorithms to create a 3D model. The 3D model is then used to determine the mangoes' volume, and a grading system is created using the volume measurements. The outcomes show how well the suggested strategy performs when it comes to precisely grading mangoes according to their volume. A drawback of the study is that it only analyses the system on a small sample size; bigger datasets and testing in more scenarios are required to determine the system's generalizability and scalability.

The authors in [24] proposed a method for categorising various types of olive fruits based on their visual traits using image processing and convolutional neural networks. The method employs a convolutional neural network to categorise the types based on the derived characteristics after extracting features from the photos of olive fruit using image processing techniques. The outcomes show that the suggested approach can categorise different types of olive fruit with an accuracy of up to 97.5%. The system's generalizability and scalability must be tested further on bigger datasets, which is one of the study's limitations because it only analyses the system on a very small dataset.

III. MATERIALS AND METHODS

The method of conducting the research is after photographing the object or the desired fruit (olive of Shengeh cultivar) which is inside a box with lighting. Image storage eliminates factors such as image noise, image background, as well as reduced light reflection in images for detailed examination, and then image processing by MATLAB software to identify and grade olives based on perceptible changes in the color spectrum. The tree of this cultivar is olive with medium growth power, with a tall crown, and thin and medium distance between nodes. Medium leaf size, oval bayonet leaf shape, medium leaf size with leaf length 6.8cm, leaf width 1.1cm. The surface of the leaf is glossy, smooth with dark green top color and the back color of the leaf is gray-green, as shown in Fig. 1.

The flowering of this cultivar of olives is such that it flowers very early but the browned petals fall later than other cultivars. Medium inflorescence length (26 mm); the number of flowers in the average inflorescence is 12 flowers. The inflorescence has many branches and lateral flowers [14,15]. The average fruit weight (average 5.2 mg), average fruit size, and ovoid shape (fruit length is 2.6 mm, and fruit diameter is 1.9 mm). Lentil or low, the ratio of flesh to the medium core (8.6), the shape of the fruit in position A is almost asymmetric,

in the position of the largest diameter in the center of the fruit, the tip of the fruit is round, early ripening, the color of the fruit when ripe is black. The percentage of oil in the dry matter is medium (50%) as shown in Fig. 2.

As you can see in Fig. 3, the closed weight of this cultivar of olives is on average 0.6 g, with a core length of 46.5 mm. In unfavorable dry climatic conditions, Shengeh olives that are sensitive to the cold, shrink and fall off the tree. In the tropics, the skin of the trunk and main branches suffers from sunburn [16-18].

After preparing the desired photos of olives, MATLAB software will be used to process the images. Content software has different toolboxes that students and engineers in each field can use the toolbox to suit their problems. One of these tools is image processing, which will be used in the present study. After saving the images, the olives are transferred to MATLAB software for detailed examination. First, actions such as noise cancellation, light reflection, image background, etc. are performed on the images. Then, by placing images in color spaces, determining different color parameters of HSV and RGB, drawing histogram diagrams of color parameters, and determining the noticeable color spectral changes of three random samples in color spaces, grading olives samples based on color characteristics (transparency and opacity). Canned and oily types and the separation of waste samples of olives are discussed.



Fig. 1. View of the surface and back of the leaves of the olive tree of the shengeh cultivar.

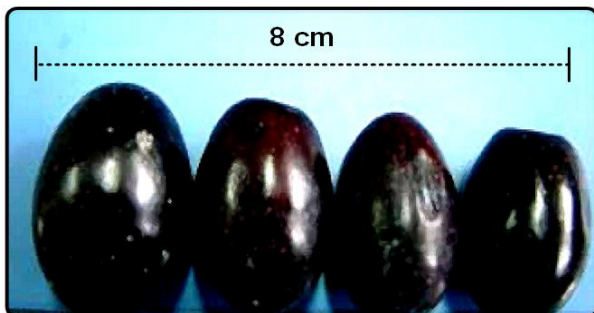


Fig. 2. Shengeh olives are ripe in different weights.

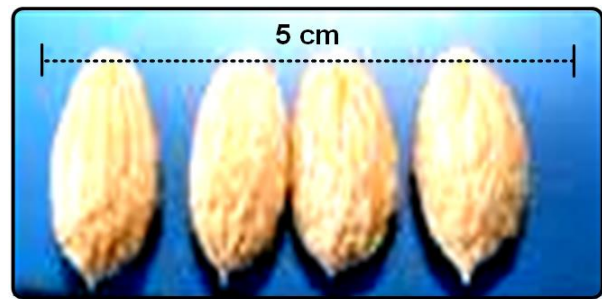


Fig. 3. Shengeh olive kernel.

IV. RESULTS AND DISCUSSION

Olive is one of the fruits that is consumed orally and its oil is used for medicinal, food, and industrial purposes. Today, with the increase in production of this product, the need to produce mechanized equipment with high accuracy in measurement and separation, to increase the efficiency of harvesting and processing of the product is essential, and in recent years the use of non-destructive and fast sight machine with computers and cameras Digital has become widespread and popular. Among the applications of machine vision in the agricultural sector is the classification of products. Traditional sensory evaluation methods have applications in determining food quality, but these methods are time-consuming and costly. In addition, the performance of this method is either not guaranteed and uniform and stable quality control of food products with this method are either not possible. These factors motivated the development of alternative methods such as image processing. In traditional methods, computer vision systems are efficient and cost-effective, and also provide more stable results.

In this study, an olive-based grading algorithm has been developed using image processing. Due to the fact that olives have a variety of domestic and foreign cultivars, Shengeh cultivar olives with three colors of black, brown, and green are randomly selected, of which 150 cultivars are processed as a research sample. Shengeh cultivar olives are prepared after a random collection of 150 healthy and without external damages such as crushing for olives (black, brown, green) in two different stages of harvest (late summer and mid-autumn). The olives are stored in water and refrigerated at 5°C for proper storage and to prevent discoloration in the experimental process. After preparing the images of the olives, they are manually classified by the author into three categories according to color: immature (green), semi-ripe (brownish), and ripe (black). Then, for color grading of olives, HSV color space is used in MATLAB software and the accuracy of the software for grading olives is evaluated.

The $L^*A^*B^*$ color space is used less frequently than the HSV color space in other works. However, due to the advantages of this color space, it has been considered in this study. For example, to filter green in an image, this range includes a spectrum that is dark green on one side and light green on the other. To separate it in RGB color space, it is not possible to select linearly, i.e. each channel with a desired color range condition. Because of such problems, the image is transferred to the HSV color space, which is composed of Hue,

Saturation, and Value components. Olive is also suitable for HSV color space due to its green color spectrum.

In the first step, a graphic interface is designed in Fig. 4. By pressing the loading button, you can select the type of olive that is displayed in the first image box. By pressing the HSV Show button, first, the image goes from RGB color space to HSV color space and then three layers of this color space are displayed in order. In the end, by pressing the Grade button, the type of olive is determined in terms of immature, semi-ripe, and ripe.

Fig. 5 shows the graphical interface execution view. In designing the graphical interface, various elements such as buttons or other elements are used so that the user can use the program in a simpler and more understandable way.

In Fig. 5, by clicking the loading button, the desired image is selected from a database of 150 olive varieties, of which 50 are immature olives, 50 are semi-ripe olives, and the last 50 are related to ripe olives. In the image below, an example of a ripe olive is selected, which is black. Fig. 6 shows the selected olive image by the program and displays it.

At this stage, according to Fig. 7, the olive type selected in the previous step, which is in RGB color space, is converted to HSV color space, which can be seen as HSV image in the image below, and then three layers of this color space include hue, saturation, and value that the appearance of olives in these three layers can also be seen in the image. Fig. 8 shows the determined degree of the olive by pressing the Grade button.

In the final step, by pressing the Grade button of the olive grade, according to its color in the HSV color space, it is determined that it is immature, semi-ripe, or ripe.

In the following, the results obtained from the software and its accuracy in diagnosis are examined, which can be seen in Table I and Fig. 9.

The table and diagram above show information about Shengeh cultivar olives and indicate that in total the number of samples is 150, of which 50 olives are in the category of immature with green color, 50 are semi-ripe with a brownish color. And the last 50 numbers belong to the mature category, the color of which is black. According to the male output, the number of correct diagnoses of the olive grade of Shengeh cultivar for the immature category is 49%, which is 98%, for the semi-mature category, 48, which is equivalent to 96%, and for the mature category, this number is equal to 50, which is equal to 100%, which is the highest. The diagnosis belongs to the degree of ripe olives and the lowest diagnosis belongs to the semi-ripe olives.

In order to classify using a neural network, among the extractive properties, different sets are selected as input and transferred to the neural network as input. Among 150 samples, the percentage of samples (105 samples) is used randomly for neural network training and validation and the remaining 30% (45 samples) are used for neural network testing. The structure of the multilayer perceptron neural network is shown in the Fig. 9.

Fig. 10 shows the structure of the multilayer perceptron neural network, which includes input, layers 1 and 2 as input and output layers, and finally network output. Table II contains the values of the artificial neural network parameters of the multilayer perceptron, including the number of layers and neurons in these layers, the transmission function used, the reverse propagation network training function, and the reverse propagation weight/bias learning function. The transfer function is selected by default.

In artificial intelligence topics, this table is used to determine the value of evaluation indicators such as accuracy. Accuracy is how many of the selected samples are correct. True depends on how many of the available correct samples are selected. Table II shows the results of test data classification using a multilayer perceptron artificial neural network classifier.

According to the values in the table above (perturbation matrix), the accuracy of the algorithm is expressed in the grading of olives of Shengeh cultivar, the accuracy of the algorithm is 98% for immature with 49 correct diagnoses, 96% for semi-ripe with 48, and 50 for mature olives. It is exactly equal to 100%. It is clear that the numbers on the main diameter of the matrix represent the number of correct classifications. Therefore, if all the numbers that are not on the original diameter are zero, the algorithm has maximum accuracy. To obtain the function of a classifier, it is sufficient to divide the sum of the elements of the original diameter by the sum of the total elements of the matrix. The efficiency for classifying the above algorithm is 98. This is divided by the division of the elements of the original diameter, which is equal to 147, which is equal to the total number of samples, which is 150.

Table III shows the performance criteria of the perturbation matrix for the classification of the artificial neural network, which indicates that the sensitivity and strength of the olive grading algorithm of the immature Schengen cultivar is 96, semi-ripe is 94, and ripe is 98. On average, the accuracy of the algorithm for grading olives of Shengeh cultivar is 98%.

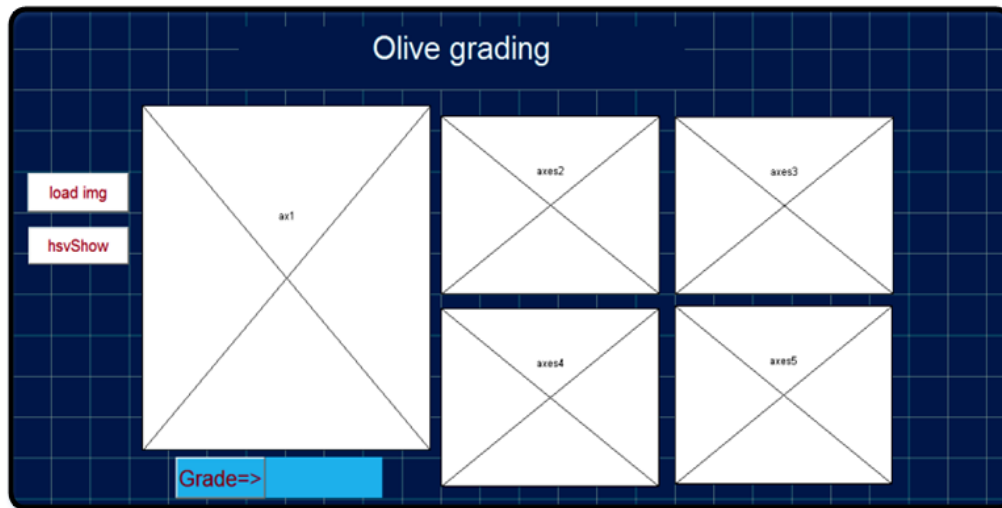


Fig. 4. Overview of graphical interface designed for grading olives.



Fig. 5. Graphic interface execution view.

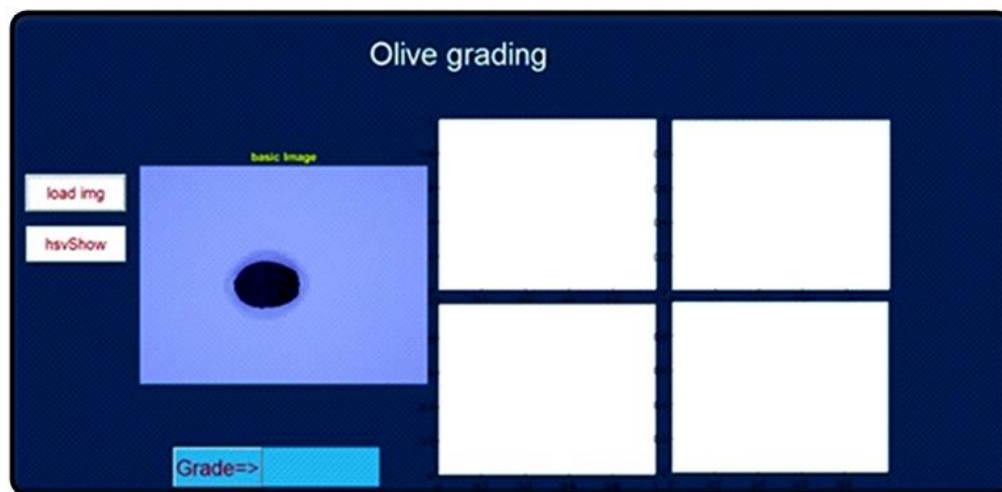


Fig. 6. Read the selected olive image by the program and display it.

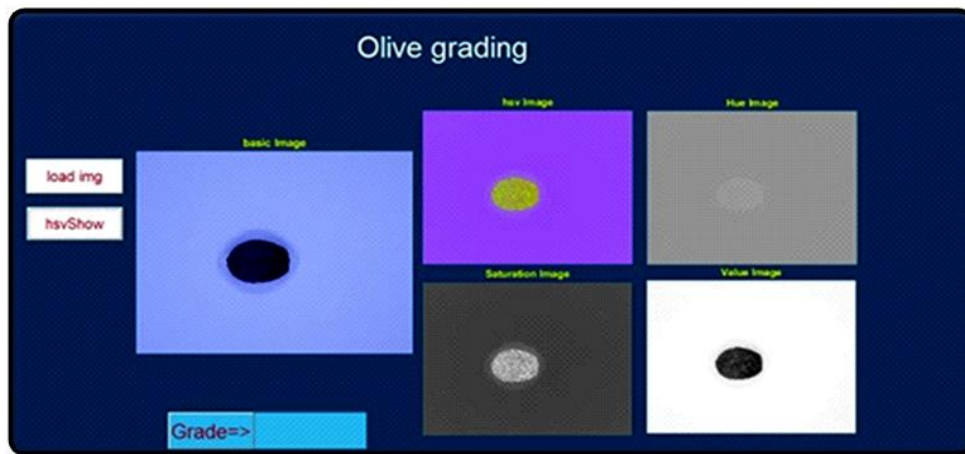


Fig. 7. Shows the olive image in the HSV color space and its components.

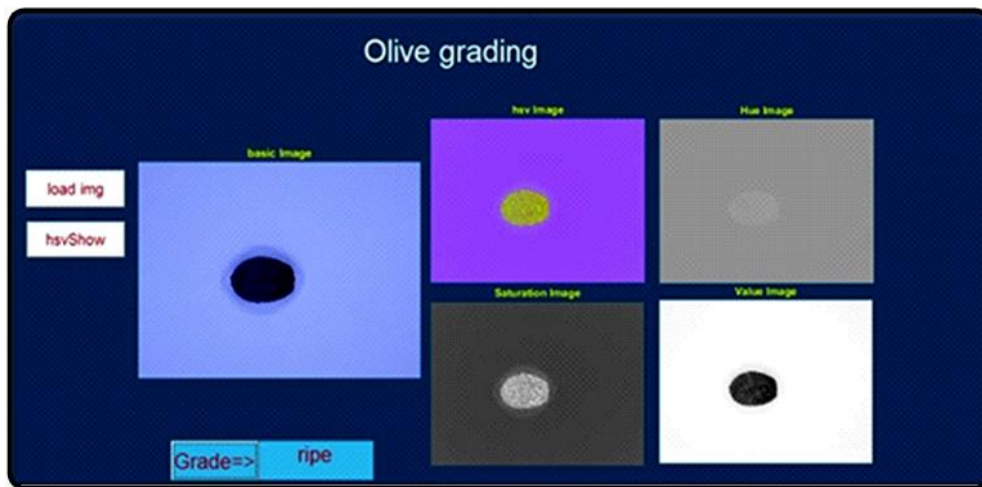


Fig. 8. Determine the degree of the olive by pressing the Grade button.

TABLE I. THE NUMBER OF SHENGEH OLIVES AND THE PERCENTAGE OF CORRECT AND INCORRECT DETECTION BY SOFTWARE

Olive Type	Number of samples	Number of Correct Detections	Correct Detection	Number of Incorrect Detection	Number of Incorrect Detections
Immature	50	49%	98	1	2
Semi-Ripe	50	48%	96	2	4
Ripe	50	50%	100	0	0

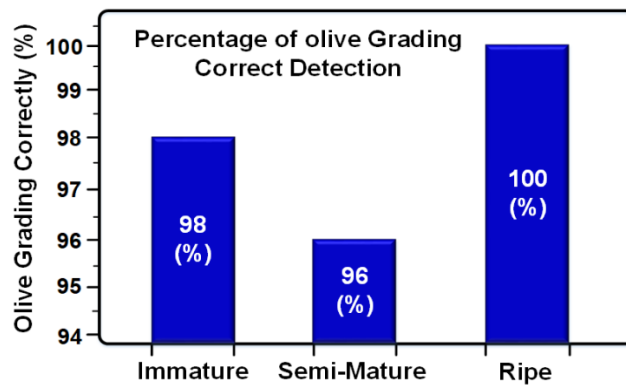


Fig. 9. Percentage of correct detection of degree of ripeness of Shengeh olive cultivar by software.

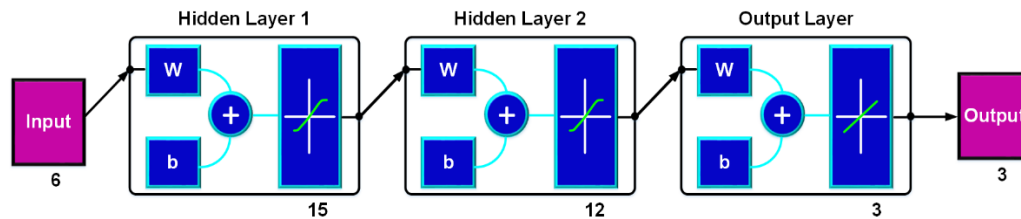


Fig. 10. The structure of the multilayer perceptron neural network.

TABLE II. COMPARISON OF CLASSIFICATION RESULTS OF IMMATURE, SEMI-MATURE, RIPE CLASSES

Classes	Immature	Semi-Mature	Ripe	Total Data	Wrong Classification	Total Correct Classification
Immature	49	1	0	50	2%	98%
Semi-Ripe	1	48	1	50	4%	96%
Ripe	0	0	50	50	0%	100%

TABLE III. THE SENSITIVITY, ACCURACY, AND PERFORMANCE VALUES FOR IMMATURE, SEMI-MATURE, AND MATURE CLASSIFICATIONS

Artificial neural network			
Classes	Sensitivity	Accuracy	Performance
Immature	98	96%	98
Semi-Ripe	96	94%	
Ripe	100	98%	

V. CONCLUSION

This paper presents an image processing-based method to grade the type of olive in terms of its ripeness as a single grain of color. In this method, the color-based feature using HSV color space as one of the best color spaces is used to separate the colors of the cultivar. Using this feature, the olive regions can be extracted. Then using analysis of extracted region, the grading task is performed. As experimental results showed, the accuracy of the method for detecting olives is 98% in the immature group, 96% in the semi-ripe group, and 100% in the ripe group. The average accuracy of the software is generally equal to 98%. However, the limitation of this method is sensitivity to light condition especially in olive region extraction. For direction of future study, deep learning-based methods can be explored to improve the proposed method.

REFERENCE

- [1] Kailis S, Harris D. Producing Table Olives. Landlinks. Press, pp. 82-84. (2007).
- [2] Radha T, Mathew L. Fruit crops. New India Publishing.; pp. 256-257. (2007).
- [3] Acar-Tek, Nilüfer, and Duygu Ağagündüz. "Olive leaf (*Olea europaea* L. folium): potential effects on Glycemia and Lipidemia." *Annals of Nutrition and Metabolism* 76, pp 10-15. (2020).
- [4] Charoenprasert, Suthawan, and Alyson Mitchell. "Factors influencing phenolic compounds in table olives (*Olea europaea*)." *Journal of Agricultural and Food Chemistry* 60, pp 7081-7095. (2012).
- [5] Sadeghi, H. Olive production and management. Agricultural education. Press, pp 121-122 (2003).
- [6] Sonka, Milan, Vaclav Hlavac, and Roger Boyle. Image processing, analysis, and machine vision. Cengage Learning, (2014).
- [7] Guyer, D. Eetal, G. E. Miles, M. M. Schreiber, O. R. Mitchell, and V. C. Vanderbilt. "Machine vision and image processing for plant identification." *Transactions of the ASAE* 29, pp 1500-1507. (1986).
- [8] Mohebbi, Mohebbat, Mohammad-R. Akbarzadeh-T, Fakhri Shahidi, Mahmoud Moussavi, and Hamid-B. Ghoddusi. "Computer vision systems (CVS) for moisture content estimation in dehydrated shrimp." *Computers and electronics in agriculture* 69 , pp 128-134. (2009).
- [9] Nadian, M.H., Rafiee, S., Aghbashlo, M., Hosseinpour, S. and Mohtasebi, S.S., Continuous real-time monitoring and neural network modeling of apple slices color changes during hot air drying. *Food and bioproducts processing*, 94, pp.263-274. (2015).
- [10] Chen, Jin, Yi Lian, and Yaoming Li. "Real-time grain impurity sensing for rice combine harvesters using image processing and decision-tree algorithm." *Computers and Electronics in Agriculture* 175, 105591. (2020).
- [11] Dingle Robertson, L., Davidson, A., McNairn, H., Hosseini, M., Mitchell, S., De Abelleira, D., & Cosh, M. H. Synthetic Aperture Radar (SAR) image processing for operational space-based agriculture mapping. *International Journal of Remote Sensing*, 41(18) , pp 7112-7144. (2020).
- [12] Jahanbakhshi, A., Momeny, M., Mahmoudi, M., & Zhang, Y. D. Classification of sour lemons based on apparent defects using stochastic pooling mechanism in deep convolutional neural networks. *Scientia Horticulturae*, 263, pp 109133. (2020).
- [13] Du, C. J., & Sun, D. W. Recent developments in the applications of image processing techniques for food quality evaluation. *Trends in food science & technology*, 15(5) , pp 230-249. (2004).
- [14] Wu, S. B., Collins, G., & Sedgley, M. Sexual compatibility within and between olive cultivars. *The Journal of Horticultural Science and Biotechnology*, 77(6) , pp 665-673. (2002).
- [15] Seifi, E., Guerin, J., Kaiser, B., & Sedgley, M. Sexual compatibility and floral biology of some olive cultivars. *New Zealand Journal of Crop and Horticultural Science*, 39(2) , pp 141-151. (2011).
- [16] Kobusingye, O. C., Hyder, A. A., Bishai, D., Hicks, E. R., Mock, C., & Josphura, M. Emergency medical systems in low-and middle-income countries: recommendations for action. *Bulletin of the World Health Organization*, 83, pp 626-631.
- [17] Kobusingye, O. C., Hyder, A. A., Bishai, D., Josphura, M., Hicks, E. R., & Mock, C. Emergency medical services. *Disease Control Priorities in Developing Countries*. 2nd edition. (2006).

- [18] Berner, K., Tawa, N., & Louw, Q. Multimorbidity Patterns and Function Among Adults in Low-and Middle-income Countries: A Scoping Review Protocol. (2021).
- [19] Ince, F. B. K., Tasdemir, S., & Ozkan, İ. A. Dimension and Color Classification of Olive Fruit with Image Processing Techniques. Selcuk University Journal of Engineering Sciences, 19(4) , pp 156-167. (2020).
- [20] Afkari-Sayyah, A. H., Azarndel, H., Rasekh, M., & Mesri-Gundoshmian, T. Discriminating defected and sound fruits of olive according to external damage area using image processing techniques. International Journal of Farming and Allied Sciences, 3(6) , pp 647-652. (2014).
- [21] Sabanci, K., & Aydin, C. Using Image Processing and Artificial Neural Networks to Determine Classification Parameters of Olives. Tarım Makinaları Bilimi Dergisi, 10(3) , pp 243-246. (2014).
- [22] Ponce, J.M., Aquino, A., Millan, B. and Andujar, J.M., 2019. Automatic counting and individual size and mass estimation of olive-fruits through computer vision techniques. IEEE Access, 7, pp.59451-59465.
- [23] Mon, T. and ZarAung, N., 2020. Vision based volume estimation method for automatic mango grading system. Biosystems Engineering, 198, pp.338-349.
- [24] Ponce, J.M., Aquino, A. and Andujar, J.M., 2019. Olive-fruit variety classification by means of image processing and convolutional neural networks. IEEE Access, 7, pp.147629-147641.
- [25] Azarndel, H., Afkari-Sayyah, A. H., Ghaffari, H., & Alipasandi, A. Olive classification according to RGB, HSV, and L* a* b* color parameters using Image processing. (2014).

Intelligent Abnormal Residents' Behavior Detection in Smart Homes for Risk Management using Fuzzy Logic Algorithm

Bo Feng¹, Lili Miao^{2*}, HuiXiang Liu³

Shanxi Railway Vocational and Technical College, Taiyuan 030013, ShanXi, China¹
Taiyuan University of Technology University, Taiyuan 030008, ShanXi, China²
Shanxi Railway Vocational and Technical College, Taiyuan 030013, ShanXi³

Abstract—In recent years, the population of sick and elderly people who are alone and need care has increased. This issue increases the need to have a smart home to be aware of the patient's condition. Identifying the patient's activity using sensors embedded in the environment is the first step to reach a smart home where the people around the patient can leave the patient alone at home with less worry. In literature, a variety of methods for detecting the performance of users in the smart home are discussed. In this study, a method for abnormal behavior detection and identifying the level of risk is proposed, in which fuzzy logic is used in cases such as when the activity start. Experimental results demonstrate that the proposed method achieved satisfied performance with 90% accuracy rate that presented better results compared to other existing methods.

Keywords—Smart home; abnormal detection; behavior analysis; activity recognition; elderly people; fuzzy logic

I. INTRODUCTION

Over the past few years, the global elderly population has been steadily increasing, according to statistics from the World Health Organization (WHO). By 2363, the number of people aged 56 and above is projected to reach two billion, accounting for 22% of the elderly population [9]. The first cause of dependence and need in the elderly is considered and there are not enough people to take care of this population. With paying attention to the increase in the number of elderlies compared to people, the incidence of this disease has increased and the issue of taking care of them is also more important now [2,3].

Designing a smart home can help these people in a very effective way. Smart home uses user behavior among the sensors embedded in the environment. It has been investigated that he plays the role of a caretaker and only if he feels the need Musk, he informs it to one of the dependents of the patient - for further investigation; informs one of the problems that can be for an elderly person. The phenomenon of falling is usually accompanied by complications and physical disabilities and even death. So according to a report, death in people over 65 years of age 2007 to 2016 has increased by 31% [4,5].

In the phase of examining abnormal behavior, it should be noted whether the detection of abnormal behavior is for the purpose of detecting cyber-attacks on the smart home system or if it is meant for abnormal activities that occur by the patient

himself. In relation to the first case, various articles [23-25] have been proposed to identify attacks by examining the sequence of various events performed by the user. But in this article, abnormal behavior refers to the second state and the abnormality of the patient's own activity is investigated.

In fact, there are two methods for detecting the abnormality of behavior, in the first method, normal behavior is modeled, and any new input that does not match this model is considered as abnormality. But in the second method, a pattern is found for the anomalies by using the background data, and with the entry of new data, if they match this pattern, they are identified as anomalies. The first strategy seems to be more efficient and realistic, because abnormal data is rarely seen in real life. As shown in Table I [10, 26], Forkan and colleagues have pointed out two defects of the anomaly detection system. The first one is the inability to predict the leading anomalies, and the second one is that the use of a context for decision-making has caused more false alarms [20].

All these issues increase the importance of designing a system to determine the unusual activity of users. In this research, firstly, user activity identification is presented in different ways. Secondly, determining the degree of abnormality of activity in different articles is discussed and in the third part, the model is presented for determination of the unusual level of user activity in smart homes and it is reviewed. Fourthly, the experimental results of the proposed method are presented and the method is compared to other methods. Finally, the paper concludes in the last section.

The research gap in this study is the lack of a method for detecting abnormal behavior and identifying the level of risk in smart homes for sick and elderly people who are alone and in need of care. While various methods have been proposed to detect users' performance in a smart home setting, there is still a need for a more comprehensive approach that can monitor and detect abnormal behavior in real-time and provide a level of risk assessment. This study aims to fill this gap by proposing a method that uses fuzzy logic to detect abnormal behavior and assess the level of risk, particularly when an activity starts. This will enable caregivers to monitor the patient's condition remotely and take appropriate actions when necessary, thereby providing better care and support for vulnerable individuals in smart homes.

*Corresponding author

TABLE I. ANALYSIS OF CURRENT ANOMALY DETECTION METHODS

Method	Advantages	Disadvantages
Classification (supervised)	Measure the accuracy of the model using the confusion matrix	1. It is needed huge data for the training phase 2. Manual labeling by humans
Clustering and statistics-based (unsupervised)	Not required to sample data	1. Efficiency depending on the type of data distribution 2. Not optimized for anomaly detection

II. RELATED WORKS

Numerous studies have been conducted on the security, energy management, and detection of user behavior—just a few of the features of the smart home [10-8]. In the behavior identification section, some researches identify unusual user behavior by using fall detection and some by using different data mining and analysis algorithms. The user's prior actions have generated funds for a variety of categories of individuals, including those with partial paralysis, Alzheimer's illness, Parkinson's disease, heart disease, children, and others [11].

In the user activity detection research area, different algorithms are used. In [92, 93] is used Flocking Algorithm and stated to detect user behavior. Comparing this algorithm to other approaches has the benefit that there is no need to determine the initial cluster number, but this system has not done any work in the field of assessing the risk for the elderly.

In [14], a hybrid algorithm called CBC-AR has been used. This approach uses the PCA technique to choose a set of features initially; then, the K-means algorithm is used to cluster the data. The NN-K algorithm is then used to classify each cluster individually in order to increase accuracy. However, this technique has a very low sensitivity to specific actions, such as leaving home and doing the dishes. It also needs better sensitivity to determine the risk factor.

According to the author in [9], modeling user behavior is inappropriate since people conduct a hierarchy of actions for a task, grouping into Alone. Clustering is employed in this article as part of a brand-new hybrid approach dubbed pattern-K. The FP Growth algorithm is used in this approach to find recurring patterns in the primary raw data after some pre-processing. The actions are clustered using the patterns that the algorithm has discovered in the following phase, and the last step employs the neural network algorithm to forecast the user's future behavior. As their results showed, the advantage of this method is to detect simultaneous and separate activities and also it has also been able to analyze the activities in different periods of time with different interpretations, however, no report of the results and final accuracy of the tests has been presented.

In the article [15], the SVM algorithm, which is effective for data classification, has been used in such a way that a series of features are first selected using the PCA algorithm, and after normalizing the data, the aforementioned algorithm is applied to the data.

In the article [6, 7] a hierarchical method for diagnosis of their activities and their abnormality have been used. The hierarchical framework is designed in such a way that the factors with higher priority are in lower layers, and each layer is executed if the lower layer reports abnormality of activity. In each layer, one of the artificial intelligence algorithms called MLN has been used to get its output. The algorithm's benefit is

that in each layer, depending on the structure defined for it, both soft rules (rules obtained by the system from the routine in the data) and hard rules (rules known by people) can be used. However, this method is not enabled to identify the complex activities.

In the article [18], ESN is used to predict the activity values of a person. The benefit of this strategy over the others is that input data may be fed into the system at any moment, whereas with the earlier ways, the data had to be entered into the system all at once.

In the article [19], it has been investigated that it is possible for more than one person to live in a house, and to solve these problems. They used video cameras in three dimensions, and by processing these videos, the speed, type of behavior and other characteristics of each user can be determined. Different Bayesian networks characterized by these features are obtained and the best model is extracted using algorithms such as the hill climbing algorithm.

The accuracy of forecasting the following event (activity) using probabilistic approaches and LSTM networks was compared in the article [20], utilizing binary sensors, and this method reached 83% accuracy.

In the article [21], machine learning algorithms and ultrasonic sensors were used to predict the activity of one or two users present at home. The suggested system is able to identify the activities and residents in the house, which, of course, is less accurate in some activities when both people were at home.

In the article [22], in order to predict the next activity of the user, the activities are placed in separate nodes and by using the definition of fuzzy rules and FSM structure, it jumps from each node to another node, but in this article, a solution is also proposed to determine the level of risk. It has not been done and the accuracy is between 86 to 99% for different activities.

According to investigation of previous studies, although various methods have been proposed to deal with detecting abnormal behavior and identifying the level of risk in smart homes for sick and elderly people, still it is required to improve the accuracy rate in abnormal behaviors detection. This study aims to fill this gap by proposing a method that uses fuzzy logic to detect abnormal behavior and assess the level of risk, particularly when an activity starts. This will enable caregivers to monitor the patient's condition remotely and take appropriate actions, when necessary, thereby providing better care and support for vulnerable individuals in smart homes.

In the following, a new method for identifying and determining the amount of risk that arises in the occurrence of each activity is discussed. In this method, one of the important things in Alzheimer's patients, i.e., "starting time of each

activity" is also investigated, from which one can understand the abnormality of the patient's behavior.

III. METHODOLOGY

One of the most important steps in detecting the abnormal state of users in smart homes is to determine the risk level of each activity, which causes each activity to be categorized as absolute, normal or abnormal. To take advantage of this feature, fuzzy logic has been used at this stage. The three characteristics of the activity start time, the duration of the sensor being on and the duration of the sensor being off are factors that are effective in determining the level of risk caused by the activity of a user living at home, which will be explained in detail below.

A. Activity Start Time

One of the important and influential factors in detecting the level of abnormality of an activity is the start time of that activity, which has been investigated using fuzzy logic. As shown in Fig. 1, this system gives the activity start time and the type of detected activity to the fuzzy logic to check the level of abnormality of the detected activity and the level of abnormality of the activity with three levels of warning: safe, moderate and high. If this warning is high, the result is sent to the output and does not go to the next layers, but if the warning is moderate, the warning level is sent to the next layer for further checks.

In the phase of examining the activity start time, it should be noted that Alzheimer's patients usually suffer from depression at sunset, and this state sometimes continues until the end of the night, and this state is called sunset syndrome [37, 36]. Therefore, this period of time should also be considered in dividing the day and night, and the patient's behavior during these times should be examined more carefully. After considering these things, a day with 24 hours is divided into five parts, which can be seen in Fig. 2:

- Midnight, which is from 12 o'clock to about 6 am
- Day, which is evening, from morning to noon
- Sunset, this is the time of sunset when the patient collapses due to various reasons such as depression,
- 4) Night, at this time until the end of the night, there is a possibility of depression, but this level of depression and anxiety is minor than at sunset and has fewer effect, but it is still important
- 5) Before sleep, this time is around 22:00 to 20:00.

The next input is the type of activity that is being performed, since according to the research done on this group of patients, indiscriminate reminders to Alzheimer's patients cause their disease to worsen, in order to prevent this issue, only a series of activities are considered that they are performed outside of normal hours. These activities are dinner, sleeping, resting (things like watching TV), preparing food, washing dishes, cleaning the house, and leaving the house.

Rules are defined once the fuzzy inputs are defined based on these inputs and taking into account the daily habits of the Alzheimer's patient, for example, if the patient washes dishes

or prepares food at 2 in the morning, this is abnormal and by giving a lot of warning to the patient's caregiver informs him that the patient is probably not in a suitable mental condition and examines his condition. A diagram of the defined rules can be seen in Fig. 3. To define the rules, we consider the usual behavior of most people in the society (which was also obtained by examining the behavior of several elderly people) and assume that this person only sleeps and goes to the bathroom at night, and as a result, does things like cleaning the house and watching TV. It is unusual for this person if he does start preparing food from his sleep

B. Sensor ON Duration

Another key element in evaluating the degree of the unusualness of action is the length of time that each sensor is active. The patient may attempt to alert others to his health worsening and rescue himself if a sensor is on excessively. For this reason, a sensor is left on more than usual. It is also crucial to consider the nature of the activity. For instance, a patient may be cleaning a room in the house, which would be okay if the sensor did not switch off for roughly 30 minutes. With these considerations in mind, the third layer, which receives three entries, then assesses the unusualness of the activity. As shown in Fig. 4, these three inputs are,

- the duration of the sensor being on
- the type of sensor
- the type of activity that are given to the fuzzy logic and using defined rules. Outputs the degree of abnormality

A series of activities such as cleaning the house are activities that because the person is moving, the location sensor does not turn off even for half an hour, and this is normal. Therefore, the three activities of washing dishes, preparing food and cleaning the house are considered. Different sensors interpret the length of time as indicating something different. For instance, if a person is awake and sitting on the sofa, likely watching TV, it makes sense that the sensor in that location will not switch off. Or the sensors that have the ability to cover a space, the patient may go to any part of that space (where this sensor is in that space), stay on and stay on for a longer time or the door sensors that detect the opening and closing of the door can be on for up to 20 minutes. In this research, we divide the sensors into 6 groups. This division is comprehensive, as a result of being off during that period of time the next layer is also included, for example, if a person goes to the toilet or bathroom and according to the structure of the existing smart home, if there is no sensor in the toilet or bathroom, it is normal that no movement is reported for 20 minutes.

Based on this, the sensors are divided into the following six types:

- Sleep sensors, where the patient sleeps and as a result, it is natural for them to be off for a long time.
- Sensors that are included in the rest. They may have a high rate of coverage and remain active for a considerable amount of time (but not to the extent of sleep sensors).

- Blind spot sensors, which are situated in a location that is not, for a long distance, covered by any other sensor.
- Door sensors that record when a door opens and closes.
- In front of the toilet and bathroom doors are toilet sensors.
- Other sensors, which include sensors that do not fit into any of the aforementioned categories.

C. Sensor OFF Duration

As it was stated before, falling in the elderly is very dangerous and brings problems for them. The elderly will find it challenging, and a person with Alzheimer's may forget to wear the sensor after taking a bath or changing into new clothes if wearable sensors are employed to monitor the recurrence of this issue. This layer determines whether it is normal or abnormal if there is no indication of the user's presence in other areas and the user is in one location for an extended time. It also measures the time between a sensor turning off and the first sensor turning on after that. If it is marked and in a static state, it is dangerous and there is a possibility of the patient falling or getting worse. This time is different based on the sensor's type. For instance, if the bed sensor goes off for a few hours, it is normal because the person may have fallen asleep, but if the sensor in front of the toilet goes off for 30 minutes, there is a possibility of danger because

the patient may be ill and as a result, he stayed in the bathroom for an unusual amount of time.

As shown in Fig. 5, the fuzzy logic of this layer has two inputs: the time the sensor is off and the type of sensor, which is divided into 6 groups as described in the previous section. And the duration of being off is divided into the same four categories in which the intervals are longer in this case.

The rules of the fourth layer are such that if the type of sensor is "toilet" and the patient enters the bathroom for more than 40 minutes, no sensor is turned on and there is no information about the patient, the patient is probably sick or something bad happened in the bathroom. For example, rule number 2 states that if the sensor is off for a moderate duration (about 6 to 40 minutes) and the type of sensor should be sleep sensor. The fact that the individual may have slept off on his bed makes this behavior reasonable, or in the case of the predefined rule, it is stated that if the duration is too long and more than two hours. In the case of any type of sensor except the sleep sensor, this is abnormal, and the longer this time is more than two hours, the greater the risk, because according to the behavior of the person under investigation and the research that has been conducted on several elderly people, sitting for more than two hours in the bathroom or toilet, as well as for resting and watching TV and for crossing blind spots, is considered unnatural.

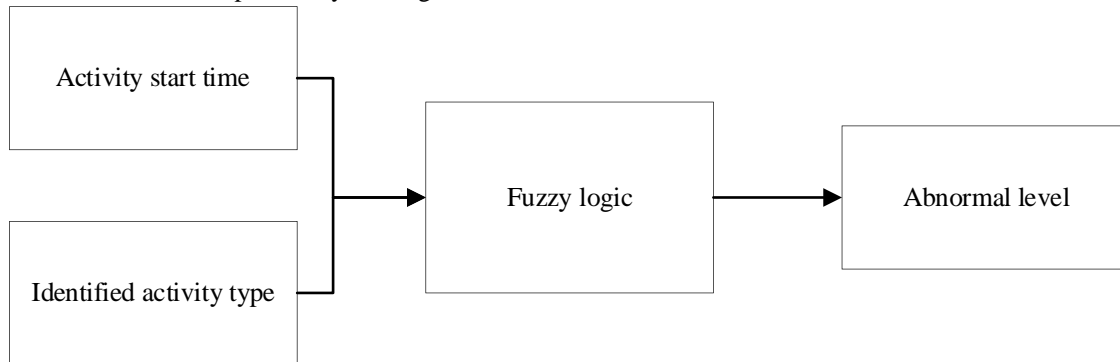


Fig. 1. Fuzzy logic inputs for activity start time.



Fig. 2. Dividing a day into 5 categories.

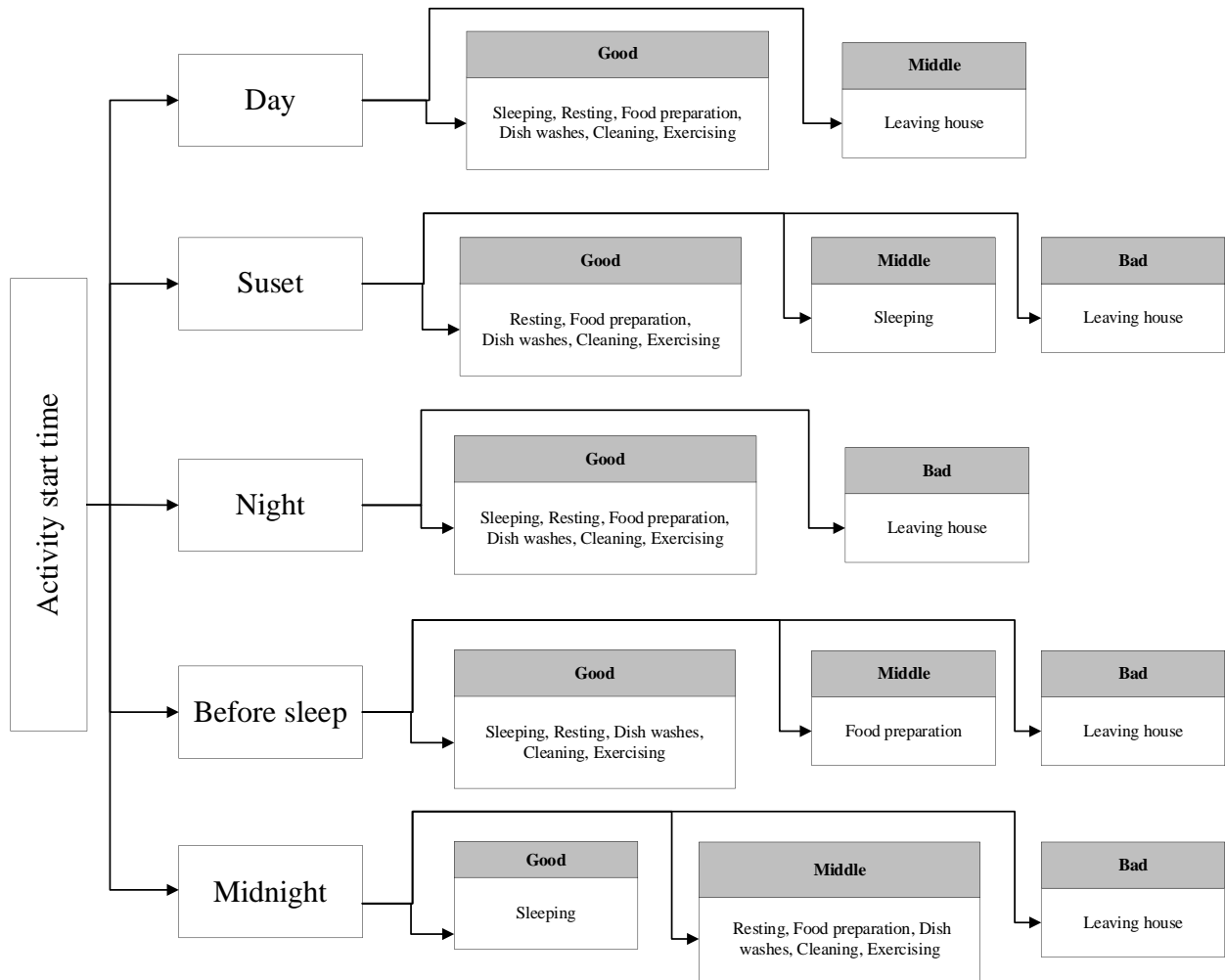


Fig. 3. Defined rules for activity start time.

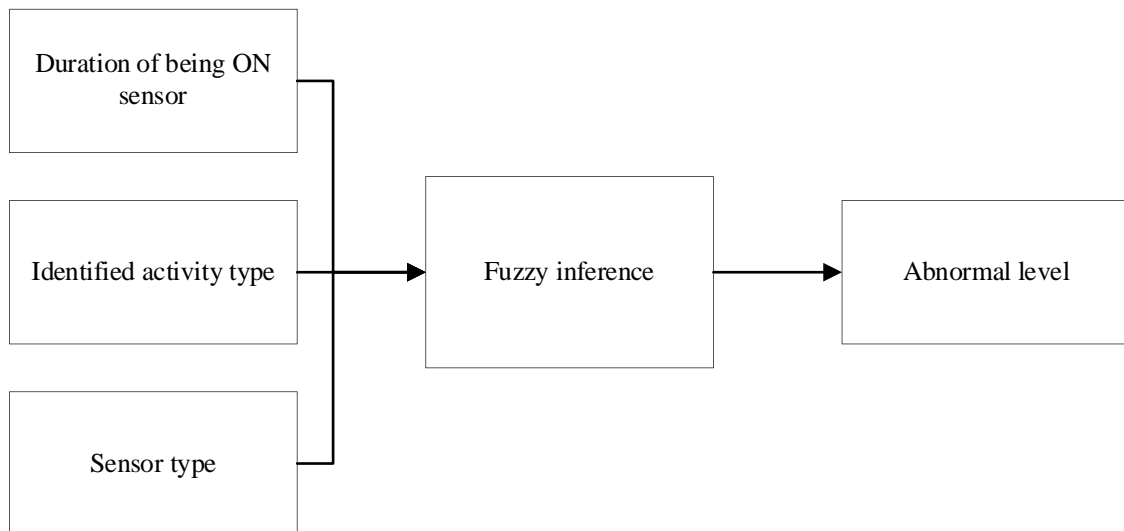


Fig. 4. Fuzzy logic inputs for the duration of the sensor being ON.

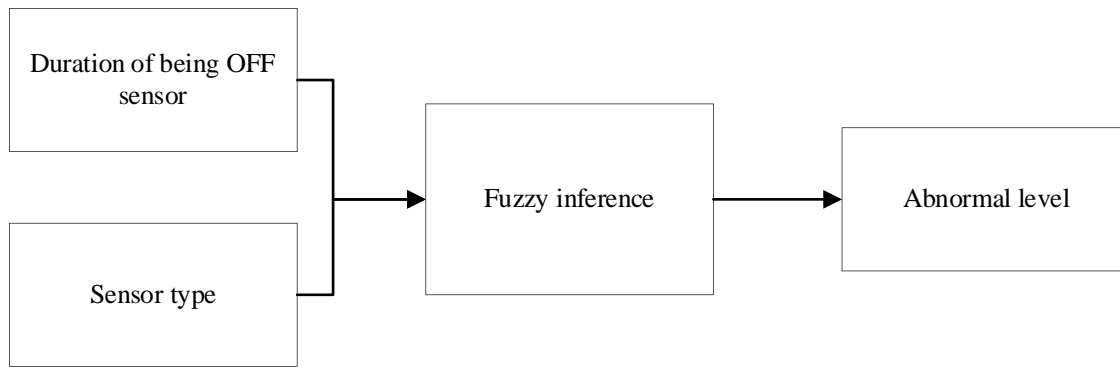


Fig. 5. Fuzzy logic inputs during sensor OFF time.

IV. RESULTS AND DISCUSSION

To check whether our fuzzy logic is working properly, we could not find a dataset that includes checking the degree of abnormality of the activity. For this reason, we randomly selected a number of transactions of the Aruba dataset from the collection of CASAS datasets [38], the number of which was 48 transactions, and after obtaining the results of the system, according to the form prepared by a doctor and a person familiar with the behavior Elderly people and Alzheimer's were filled, a comparison was made between the results, which is given below as an example of the steps of the work method.

In order to check that "if the patient washes the dishes at 10 o'clock in the morning," in the first step, after the type of activity was detected using one of the algorithms of the activity type determination section, washing the dishes, this activity along with its start time was classified into the next step. The second one is sent and there, according to the rules defined for each person, the level of abnormality of that activity is determined in three stages. These scenarios were given to two experts in the same way and their average opinions about each scenario were calculated and compared with the specific results from the system.

In order to validate the system, we collected data for 20 individuals which 4 individuals exhibit abnormal behavior (i.e., positive) and 16 individuals exhibit normal behavior (i.e., negative). As experimental results shown, our algorithm correctly identifies 3 individuals with abnormal behavior and misclassifies 2 individuals with normal behavior as abnormal. In this case, we can calculate the performance metrics as shown in Table II,

TABLE II. OBTAINED VALUSE FOR PERFORMANCE METRICS

Performance Metric	Value
True Positive (TP)	16
False Positive (FP)	1
True Negative (TN)	2
False Negative (FN)	1

Based on TP, FP, TN and FN, True Positive Rate (TPR) and False Positive Rate (FPR) are calculated as shown in Table III.

TABLE III. OVERALL PERFORMANCE

Performance Metric	Formula	Value
True Positive Rate (TPR)	$TP / (TP + FN)$	94%
False Positive Rate (FPR)	$FP / (FP + TN)$	33%
Accuracy	$(TP + TN) / (TP + FP + TN + FN)$	90%

In order to compare the proposed method with other existing methods, we experimented the Farhad et al. [14] and Casagrande et al. [20] on these 20 individuals. This experiment is conducted to prepare a fair comparison based on the same dataset. It is not tailored for a specific individual and is developed based on the demands and routines of old and Alzheimer's patients in general. Despite this, it offers decent accuracy. It is anticipated that if the patient's unique habits learned from his caregiver were taken into account, the accuracy would be better than it is now. Table IV presents the performance comparison between the proposed method and other existing methods.

TABLE IV. PERFORMANCE COMPARISON BETWEEN PROPOSED METHOD AND OTHERS

Method	Accuracy rate
Casagrande et al [20]	84%
Fahad et al [14]	88.6 %
The proposed method	90 %

As experimental results and overall performance comparison which obtained by TPR, FPR and accuracy rate, the proposed method achieved better accuracy rate in abnormal behavior detection compared to other existing methods.

V. CONCLUSION

In this research, firstly, a review of the works done in detecting the user's performance in the smart home, which includes two subcategories of fall detection and activity detection, was done. In the following, fuzzy logic was used to determine the level of risk of the patient's activity so that the habits of each patient can be obtained in the form of rules in accordance with his special conditions, and these rules can be used to determine the level of abnormality of the activities. In order to reach the defined goal, three options were used and according to them, the level of abnormality was reported. In

order to reach the defined goal, three options were used and according to them, the level of abnormality was reported. These options include (1) the start time of the activity, (2) the duration of the sensor being on, and (3) the duration of the sensor being off and the patient actually being motionless. Another noteworthy point is that in the proposed method, due to the implementation of the method for people with special conditions, the wearable sensors were not used. Because in the case of using such sensors to check the occurrence of problems, it is difficult for the elderly, and especially for an Alzheimer's person, it may be associated with forgetting to wear the sensors after certain conditions such as going to the bathroom or changing clothes. Anyway, according to these mentioned conditions, the implementation of this layer also produces good results with 90% accuracy.

For future works, it is suggested to use a method to individualize this system and detect the presence of several users at home and implement the system for heart patients, children, etc. Furthermore, the hardware implementation of this model can also be of great help to make it more accurate.

REFERENCES

- [1] World Health Organization. 10 facts on ageing and health.; Available from: <http://www.who.int/features/factfiles/ageing/en/>.(2017)
- [2] Alzheimer's Association Report, Alzheimer's disease facts and figures, Alzheimer's & dementia, 9(2):pp208-245, (2013).
- [3] World Health Organization, 10 facts on ageing and the life course.; Available from: http://www.who.int/features/factfiles/ageing/ageing_facts/en/index.html.(2015)
- [4] K Allen., Deaths from Falls by Older Adults Sharply Increase.; Available from: <https://www.aarp.org/health/conditions-treatments/info-2018/falling-deaths-surge-for-elderly.html>. (2018)
- [5] S Shakeri, "A Smartphone-based Fall Detection System using Accelerometer and Microphone", Iranian Journal of Biomedical Engineering, 9(4):399-410, 2016. [6]
- [6] K Gayathri.,S Elias., and B Ravindran., "Hierarchical activity recognition for dementia care using Markov Logic Network", Personal and Ubiquitous Computing, 19(2):pp271-285, (2015)
- [7] K Gayathri. and K Easwarakumar., "Intelligent decision support system for dementia care through smart home", Procedia Computer Science, 93:pp 947- 955, (2016).
- [8] M M Hossain.,M Fotouhi., and R Hasan., "Towards an Analysis of Security Issues, Challenges, and Open Problems in the Internet of Things", in 2015 IEEE World Congress on Services. 2015. New York, NY, USA: IEEE.
- [9] P Sukanya. and K S Gayathri ., "An Unsupervised Pattern Clustering Approach for Identifying Abnormal User Behaviors in Smart Homes", IJCSN, 2(3), (2013).
- [10] U.A.B.U.A Bakar., H Ghayvat., S F Hasanm., and S C Mukhopadhyay., "Activity and anomaly detection in smart home: A survey", in Next Generation Sensors and Systems, Springer. pp. 191-220, (2016).
- [11] R Damaševičius, M Vasiljevas ., J Šalkevičius., and M Woźniak., "Human activity recognition in AAL environments using random projections", Computational and mathematical methods in medicine,(2016)
- [12] J Lapalu., K Bouchard., A Bouzouane.,B Bouchard., and S Giroux., "Unsupervised Mining of Activities for Smart Home Prediction", Procedia Computer Science, 19:pp503-510, (2013).
- [13] K Bouchard., "Unsupervised spatial data mining for human activity recognition based on objects movement and emergent behaviors", Université du Québec à Chicoutimi, (2014).
- [14] L G Fahad., S F Tahir., and M Rajarajan., "Activity Recognition in Smart Homes Using Clustering Based Classification", 22nd International Conference on Pattern Recognition. (2014).
- [15] A Fleury., M Vacher., and N Noury., "SVM-based multimodal classification of activities of daily living in health smart homes: sensors, algorithms, and first experimental results", IEEE transactions on information technology in biomedicine, 14(2):pp 274-283, (2010).
- [16] H Medjahed., D Istrate., J Boudy., and B Dorizzi., "Human activities of daily living recognition using fuzzy logic for elderly home monitoring", in Fuzzy Systems, FUZZ-IEEE 2009. IEEE International Conference on. IEEE, (2009).
- [17] E Kim., S Helal ., and D Cook ., "Human activity recognition and pattern discovery", IEEE Pervasive Comput, 9(1):pp 48-53, (2010).
- [18] A Lotfi.,C S Langensiepen.,S M Mahmoud., and M J Akhlaghinia., "Smart homes for the elderly dementia sufferers: identification and prediction of abnormal behaviour", J. Ambient Intell. Humaniz. Comput. 3(3): pp205-218, (2012).
- [19] Y L Hsueh., N H Lin., C C Chang., O. T.-C Chen., andW N Lie., "Abnormal event detection using Bayesian networks at a smart home", in Ubi-Media Computing (UMEDIA), 2015 8th International Conference on. IEEE, (2015).
- [20]] F D Casagrande. And E Zouganeli ., "Activity Recognition and Prediction in Real Homes", in Symposium of the Norwegian AI Society. Springer,(2019).
- [21] V S Kashyap., "Activity recognition and resident identification in smart home environment", Unitech Institute of Technology, Auckland, New Zealand, (2020).
- [22] G Mohmed., A Lotfi., and A Pourabdollah., "Human activities recognition based on neurofuzzy finite state machine", Technologies, 6(4):p110, (2018).
- [23] S Ramapatruni.,S N Narayanan.,S Mittal.,A Joshi., and Joshi K. P., "Anomaly detection models for smart home security", in IEEE 5th Intl Conference on Big Data Security on Cloud (BigDataSecurity), IEEE Intl Conference on High Performance and Smart Computing (HPSC) and IEEE Intl Conference on Intelligent Data and Security (IDS) (2019).
- [24] M Yamauchi.,Y Ohsita.,M Murata.,K Ueda., andY Kato., "Anomaly detection for smart home based on user behavior", in IEEE International Conference on Consumer Electronics (ICCE). (2019).
- [25] M Yamauchi.,Y Ohsita.,M Murata.,K Ueda., and Y Kato., "Anomaly Detection in Smart Home Operation from User Behaviors and Home Conditions", IEEE Transactions on Consumer Electronics, 66(2):pp183-192, (2020).
- [26] F Cardinaux.,S Brownsell.,M Hawley., and D Bradley., "Modelling of behavioural patterns for abnormality detection in the context of lifestyle reassurance", springer, Iberoamerican Congress on Pattern Recognition, pp. 243-251, (2008).
- [27] A.R.M Forkan.,I Khalil.,Z Tari.,S Foufou., and A Bouras., "A context-aware approach for longterm behavioural change detection and abnormality prediction in ambient assisted living", Pattern Recognition, 48(3):pp628-641, (2015).
- [28] W.A Khan., M Hussain., M Afzal., M. B Amin., and S Lee. "Healthcare standards based sensory data exchange for Home Healthcare Monitoring System", in Annual International Conference of the IEEE Engineering in Medicine and Biology Society. (2012).
- [29] M.S Lee., J.G Lim., K.R Park., and D.S Kwon., "Unsupervised clustering for abnormality detection based on the tri-axial accelerometer", in ICCASSICE. pp. 134-137, (2009).
- [30] Y S Lee. and W Y Chung., "Automated abnormal behavior detection for ubiquitous healthcare application in daytime and nighttime", in Proceedings of 2012 IEEE-EMBS International Conference on Biomedical and Health Informatics. (2012).
- [31] G Virone.,M Alwan.,S Dalal., S.W Kell., B Turner., J.A Stankovic., and R.A Felder., "Behavioral patterns of older-adults in assisted living", IEEE Trans Inf Technol Biomed, 12(3):pp387-98, (2008).
- [32] T.H.T Hou., W.L Liu., and L Lin., "Intelligent remote monitoring and diagnosis of manufacturing processes using an integrated approach of neural networks and rough sets", Journal of Intelligent Manufacturing, 14(2):pp239-253, (2003).
- [33] P Barsocchi., M.G Cimino., E Ferro., A Lazzeri .,F Palumbo., and G Vaglini., "Monitoring elderly behavior via indoor position-based stigmergy", Pervasive and Mobile Computing, 23:pp26-42,(2015).

- [34] Rashidi P., Cook D.J., and Holder L.B., “Schmitter-Edgecombe M., Discovering activities to recognize and track in a smart environment”, *IEEE transactions on knowledge and data engineering*, 23(4):527-539, 2010.
- [35] QYan.,S Xia., and Y Shi., “An anomaly detection approach based on symbolic similarity”, in *Chinese Control and Decision Conference*. (2010).
- [36] E Roth., 7 Tips for Reducing Sundowning.; Available from: <https://www.healthline.com/health/dementiasundowning#take-care-ofyourself.> (2016)
- [37] G Mokhtari.,Q Zhang., and A Fazlollahi., “Nonwearable UWB sensor to detect falls in smart home environment”, in *Pervasive Computing and Communications Workshops (PerCom Workshops)*, pp. 274-278. IEEE, 2017.
- [38] CASAS, t.W.C.s.h. project, Editor. (2011)

BREPubSub: A Secure Publish-Subscribe Model using Blockchain and Re-encryption for IoT Data Sharing Management

Hoang-Anh Pham

Ho Chi Minh City University of Technology (HCMUT), 268 Ly Thuong Kiet St., District 10, Ho Chi Minh City, Vietnam
Vietnam National University Ho Chi Minh City (VNU-HCM), Linh Trung Ward, Ho Chi Minh City, Vietnam

Abstract—As a result of the incredible growth and diversity of IoT systems and applications over the past several years, an enormous amount of sensing data has been generated, which is critical for developing IoT-based intelligent systems. So far, it has taken a significant amount of time and money to collect sufficient sensing data for these smart systems leading to demands of sharing or exchanging available and valuable data to reduce the time and money spent on the data collection process. However, ensuring the data sharing process's integrity, security, and fairness is fraught with challenges. This paper proposes a Blockchain-based model that supports a secure publish-subscribe protocol for data sharing management by addressing three criteria such as confidentiality, integrity, and availability. In addition, the proposed model adopts a re-encryption technique to optimize shared data storage with multiple users and enhance the security of the data exchange process in a transparent and public environment like Blockchain. We have developed a DApp to demonstrate the feasibility of our design and evaluate its performance.

Keywords—Publish-subscribe; blockchain; re-encryption; IoT data sharing

I. INTRODUCTION

The Internet of Things (IoT) is a term that refers to the fusion of sensor technologies, big data, artificial intelligence, and network infrastructure. Sensors operate as senses in IoT devices, collecting data about their surrounding environment. These data can be analyzed locally on IoT end-devices or transferred to remote servers for advanced analysis. IoT devices can take actions based on the analytical results via actuators. Currently, IoT is one of the core technologies in the industry 4.0 era and plays a crucial role in many aspects of our lives, from inside to outside of society. IoT is ubiquitous, yet it is not always visible. IoT is renovating physical objects into an ecosystem of information that can be shared between implanted, portable, and even wearable devices. This will enrich our lives with both data and technology. However, building a sufficient database system to manage and store those IoT data spends a lot of time, money and is inefficient. Poor management might result in the loss of sensitive data such as health status, lifestyle patterns, and device control.

Traditional IoT access management programs are mainly built on popular access management models such as discretionary access control (DAC) [1], attribute-based access control (ABAC) [2], role-based access control (RBAC) [3], or capability-based access control (CapBAC) [4]. It's worth mentioning that in the aforementioned management methods, a centralized entity confirms object access permissions. As a

result, it can lead to a single point of failure. To address this issue, distributed CapBAC models have been developed [5][6], in which IoT devices themselves, rather than a centralized entity, authenticate access. However, IoT devices are small, low-power, and have limited computing capacity, thus they can't serve as an access authorization entity.

In the meanwhile, the decentralized architecture enables Blockchain to work efficiently without a central authority. It allows participants to conduct transactions securely, even though the fact that they may not trust one another in a trustless network. Various studies presented systematic investigations and reviews of the potential use cases of Blockchain beyond cryptocurrencies, focusing on how Blockchain may mitigate certain problems in IoT, such as access control [7][8], security and privacy [9][10]. Combining Blockchain and IoT is more applicable since smart contracts can automatically execute agreement terms and conditions after being digitalized, built, and stored on Blockchain. In addition, the execution of smart contracts on Blockchain is precise and transparent to all parties in the ecosystem. Therefore, Blockchain is a viable option for IoT access control management.

Many works have been studied to demonstrate the potential of incorporating Blockchain into IoT under various application scenarios such as shared economy [11], data trading management [12], authorization [13], smart-home [14], healthcare [15], access control [16][17], and IoT cloud [18]. These works inspired us to contribute a study on applying Blockchain to an IoT application scenario, such as access control management. Regarding IoT access management, it can take several forms, such as device ownership management, data disclosure management, and data sharing management. Within the scope of this study, we focus on a data management system that can manage and distribute data between people who own the equipment that generates the data and those who need to access and utilize those IoT data under a publish-subscribe protocol. As above-mentioned, Blockchain combined with IoT will be an efficient solution to solve the problems encountered in centralized architecture when managing IoT data. Additionally, we adopt re-encryption methods to minimize the storage but still secure shared data. All data sharing transactions between system participants are kept transparent and immutable. Consequently, tracking transactions and settling inter-party disputes are simple, and non-repudiation is assured. Some factors like smart contract security and transaction costs are taken into account while evaluating the system. The main contributions of our work are summarized as follows.

- Propose a Blockchain-based model for IoT data sharing management under publish-subscribe fashion.
- Utilize proxy re-encryption to enhance the security of the data exchange process and optimize the storage space for sharing data with multiple users.
- Develop a DApp to demonstrate the feasibility of the proposed model and evaluate the performance in terms of transaction's cost.
- Conduct additional tests to validate security issues related to smart contracts in Blockchain, such as re-entrancy attack.

The rest of this paper is organized as follows. Section 2 summarizes related works to clarify the scope of our study. Then, Section 3 describes the architecture of our proposed system with explanation in detail. The implementation and evaluation will be discussed in Section 4, and Section 5 provides the final concluding remarks and future works.

II. RELATED WORKS

The industrial revolution 4.0 brings tremendous global changes, particularly in the Internet of Things, machine learning, and big data. The amount of data created by sensors is enormous and sensitive, posing challenges for access management in the IoT that may be handled in various methods. However, as above-mentioned, our study only focuses on the issue of IoT data storage and sharing management. The proposed system, unlike traditional models, eliminates the role of the third-party service during data exchange between a *Data Owner* DO, who has data to share or sell, and a *Data User* DU, who wishes to purchase the data. Our proposed model employs Blockchain, a decentralized architecture to strengthen the security of the data exchange process, in which all actions are executed via smart contracts, like other existing works.

Both DO and DU are identified using the Blockchain's key system. Everyone on the Blockchain network will own a key pair, including a private key (secret key) and a public key generated from the private key. Each user in the system will be identified by the address associated with the public key. Key pairs are used for encryption during data storage and exchange and for creating IDs for network users. The system will use the Blockchain network environment (e.g., Ethereum) to allow the DO and DU to exchange data. Instead of relying on a third-party intermediary for data sharing, the system performs data exchange using sets of rules and management rules built on Smart contracts.

Unlike conventional methods without Blockchain, data sharing is transparent and public via smart contracts. Every share transaction is recorded on the decentralized network, making changing the transaction history incredibly impossible. At the same time, utilizing the Blockchain network platform's public and private keys combination helps secure data integrity and security, avoiding intermediaries stealing, decrypting, and exploiting data (i.e., a man-in-the-middle attack).

Our proposed solution is comparable to the work given in [19], which largely solves the problem of sharing data that the Aggregator wants to share their collected data and does not have to expose the secret key of subscribers to decrypt the data

packet. Each data slot, as well as each subscriber, will have a unique re-encryption key, which will prohibit subscribers from reading all of the data in the system. Furthermore, because both the Aggregator and the Publisher interact with Blockchain, there is no single point of failure, and all actions are recorded in Blockchain as verifiable and trustable transactions. When Aggregator publishes data, they must sign it with their private key, which helps to prove who owns that data packet. Furthermore, data is fragmented into chunks and kept across several storage nodes when utilizing distributed hash-tables, making it challenging to assemble data. However, the work in [19] has yet to present detailed implementation and performance evaluation.

Another similar work [20] primarily comprises a gateway that receives and processes data from IoT devices before re-encrypting and sending it to DU. The authors have successfully developed a PoC system to demonstrate the proposed idea. However, its current model still has some security and performance issues.

- First, since the gateway listens to all events from smart contracts, bottleneck concerns may arise when the number of queries is significant enough. For example, if multiple DUs try to request data simultaneously, the gateway will get overloaded, resulting in a single-point failure. Although, the authors also attempted to tackle the problem by limiting access to a certain number of times. However, it is not practicable and creates an overhead for managing access in the smart contract.
- Second, when receiving a data request event, the gateway will retrieve data from cloud storage, encrypt it with DO's private and DU's public keys, and then transfer data packets to DU accordingly. To do the re-encryption, the gateway must store DO's private key, which can be readily compromised on the gateway. As a result, if hostile parties attack the gateway, they can obtain the DO's private key and data packets, allowing information to leak.
- Third, the work in [20] did not utilize a proxy re-encryption; instead, as demonstrated in Fig. 1, whenever a DU requests data, DO will decrypt the encrypted data with DO's secret key sk_{Pub} (private key), and then encrypt data again with DU's public key pk_{Sub} before uploading data back to the Storage for sharing to DU who will download shared data and decrypt it with DU's secret key sk_{Sub} . This procedure will be repeated for different DUs, which will consume time (due to multiple encryption and decryption) and storage space (due to duplicates of the same data with different encryption). Leveraging proxy re-encryption can overcome the above issues. As demonstrated in Fig. 2, DO does not need decrypt the data, instead, DO will generate and share the re-encryption key rk_{Sub} for each DU. Then, DU can use this re-encryption key to decrypt the shared data.

III. THE PROPOSED APPROACH

A. System Design

Fig. 3 illustrates the architecture of our proposed model, in which data can be gathered from IoT end-devices and

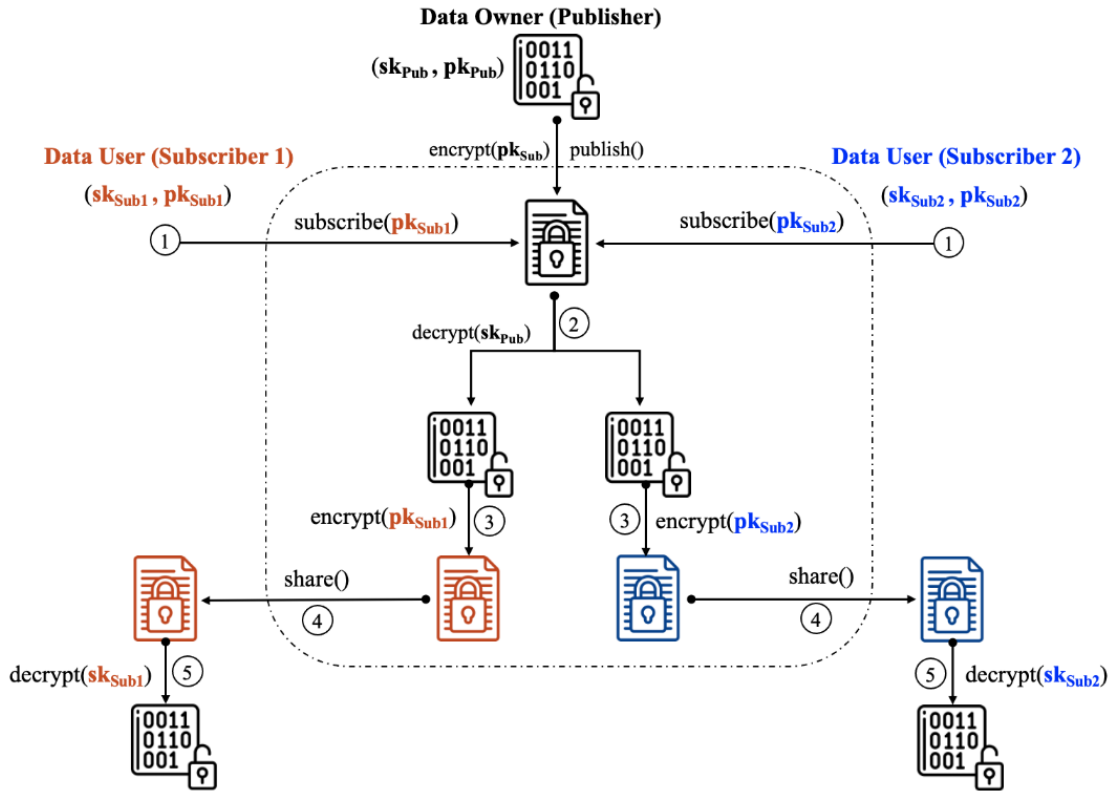


Fig. 1. Repeated encryptions/decryptions and multiple copies of shared data in the sharing process without using re-encryption.

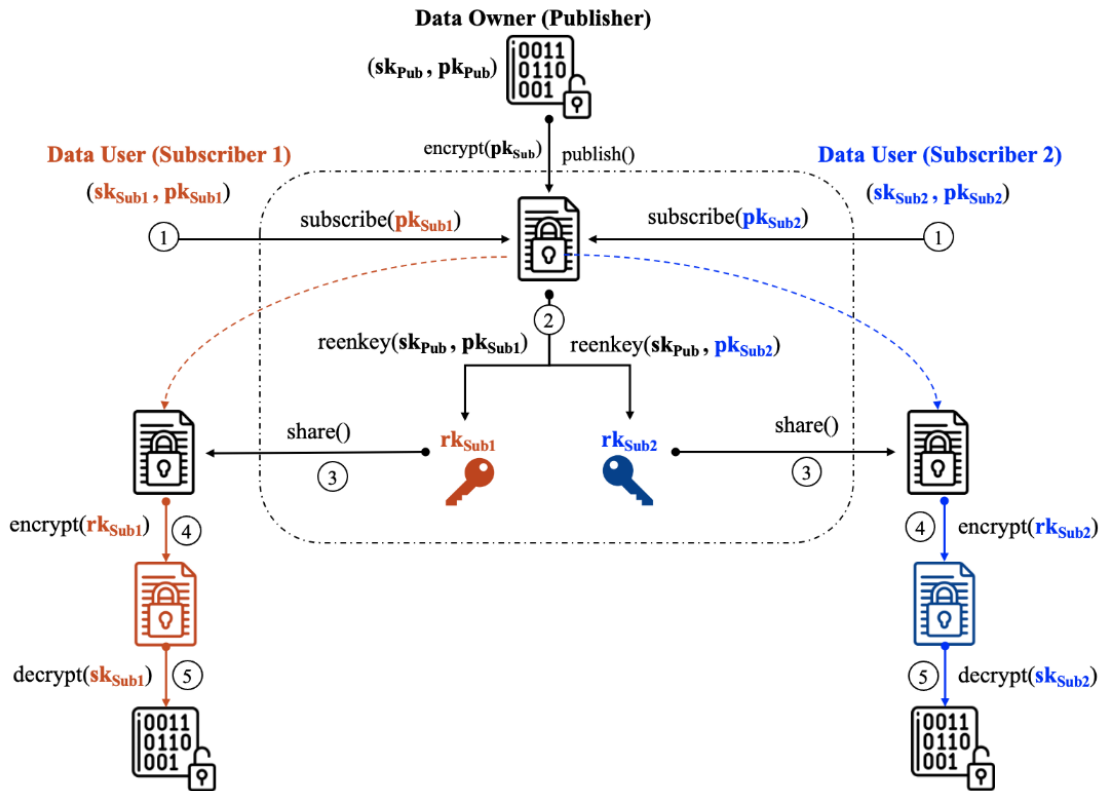


Fig. 2. Optimize shared data storage by using re-encryption.

subsequently transferred to the *Off-chain Storage* via the *Data Management System* (DMS). The proposed model adopts IPFS - a distributed system, as the Off-chain Storage for storing and gaining access to data and files via content identifiers (CID). The detail of the publish-subscribe process between DO and DU can be summarized in eight steps as follows.

- 1) The DO register IoT devices information to smart contract, including device's name, description, and price per day that will be used as a parameter to calculate the amount of money the subscriber has to pay.
- 2) When the DU subscribes to a device, he must send the subscription fee to the smart contract that corresponds to the device and the time period for which he wishes to get shared data. The fee will be retained in the smart contract until it is confirmed by the DU after successfully receiving data.
- 3) The DO will receive an event once a DU subscribes to data from the smart contract.
- 4) As shown in Fig. 3, DO collects the data from IoT devices and sends these data to the DMS, which will encrypt data by DO's public key before uploading it to IPFS to get the CID. Then, DMS will create the re-encryption key from the corresponding DU's public and DO's secret keys. Finally, DO publishes the CID of encrypted data and the list of re-encryption keys to the smart contract for sharing data with DUs.
- 5) The DU listens and receives the published data or updates key-event in order to get the CID of encrypted data and list of re-encryption key or new list of re-encryption key from the smart contract.
- 6) The DU goes to the IPFS to get both the encrypted-data and corresponding re-encryption keys.
- 7) The DU extracts the corresponding re-encryption key. In order to get the raw data, the DU will perform two steps of decryption sequentially. First, the DU will re-encrypt the encrypted-data with re-encryption key, then continue to decrypt re-encrypted data with DU's private key.
- 8) If the data is valid, the DU will certify on the smart contract that the data was received accurately. Once the DU confirms successful data reception, the subscription fee pre-paid by the DU will be transferred to the DO via the smart contract.

B. Implementation

We use many technologies, including Blockchain platform, programming languages, development framework, and libraries as summarized in Table I, to develop a prototype to demonstrate and evaluate the proposed model. Meanwhile, Fig. 4 depicts whole sequence diagram of the data sharing management process between DO and DU that correctly executes eight steps described in the proposed model as shown in Fig. 3. In addition, Fig. 5 shows a snapshot of the WebApp that displays all DO's registered devices that are ready to share and be subscribed by users.

Since IoT data collection is not the main objective in our current study, we use virtual devices to generate data for evaluating the sharing management in the proposed model on Blockchain.

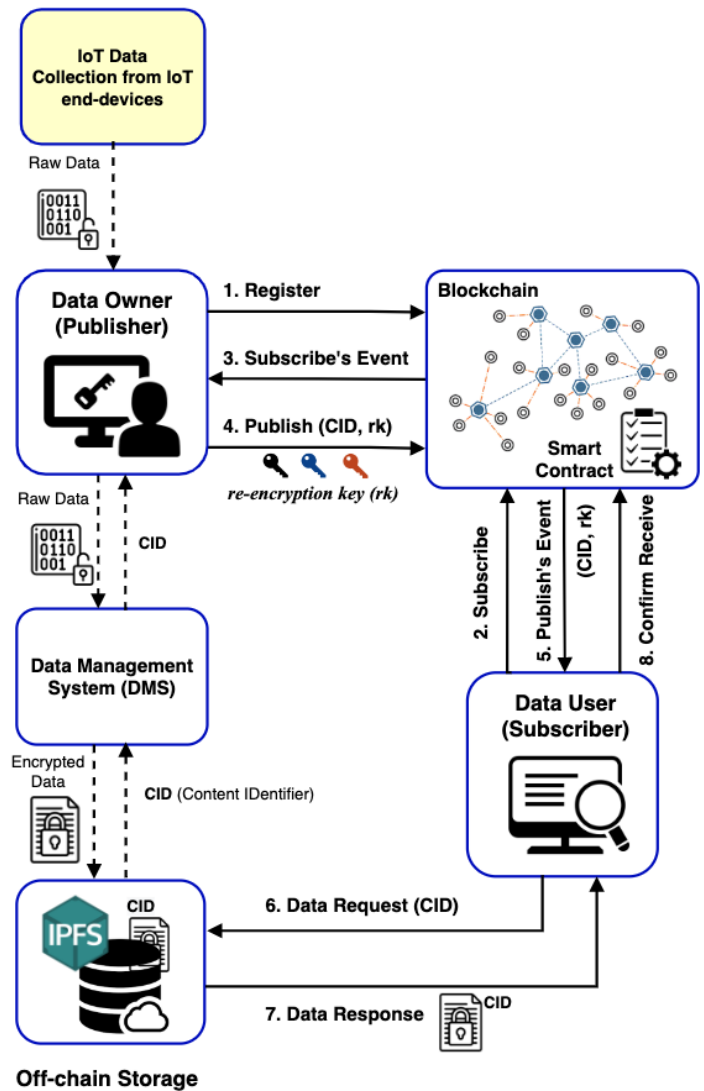


Fig. 3. System architecture of the proposed model.

TABLE I. IMPLEMENTATION DESCRIPTION

Modules	Implementation descriptions
Blockchain platform	Ethereum
Smart Contract	Solididy
Data Management System (DMS)	Nodejs, Express
Communication between DMS and IPFS	ipfs-http-client
Proxy Re-encryption	recryptjs
DApp testing	Truffle, Ganache

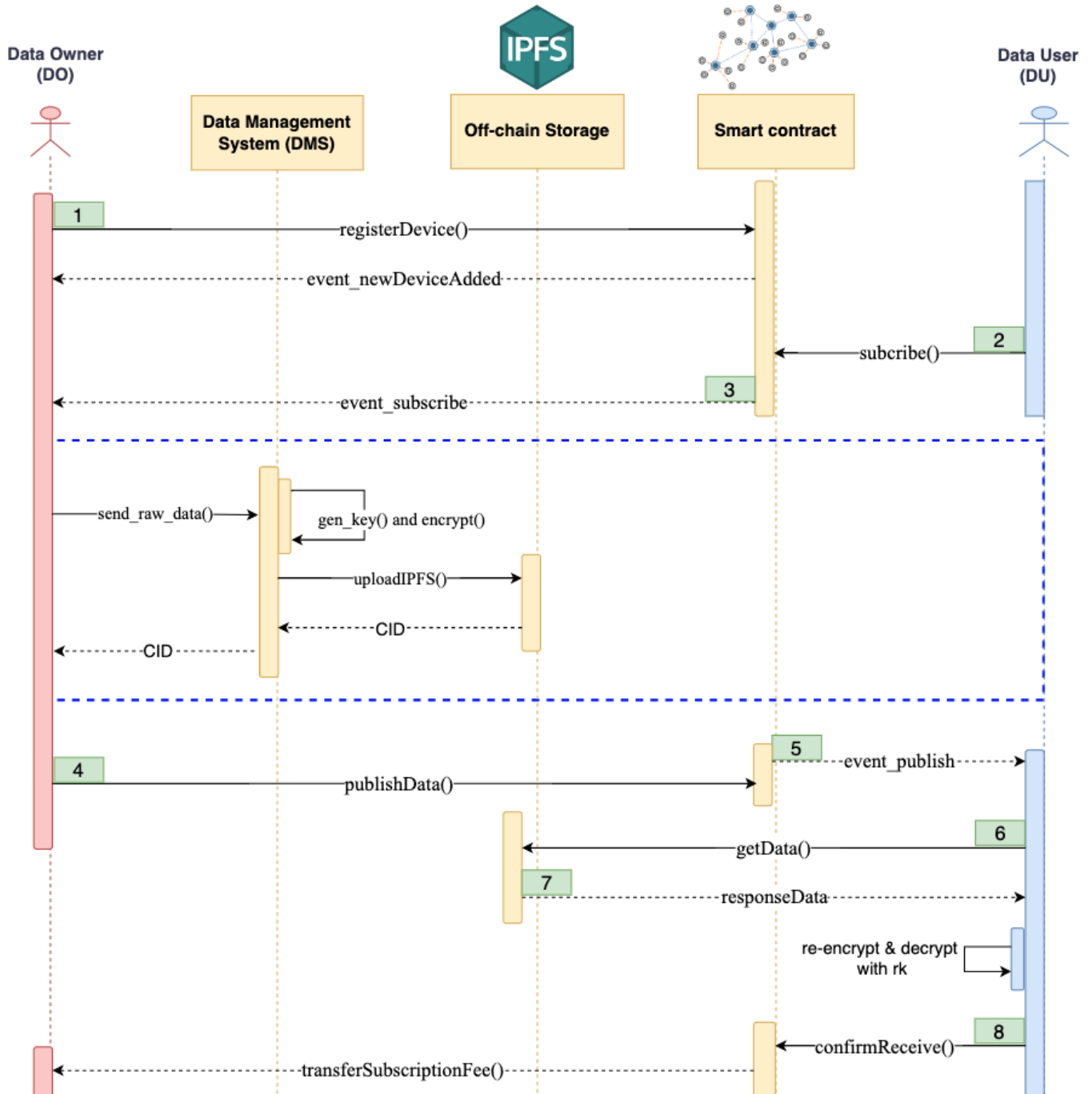


Fig. 4. Sequence diagram of the publish-subscribe procedure in the proposed model.

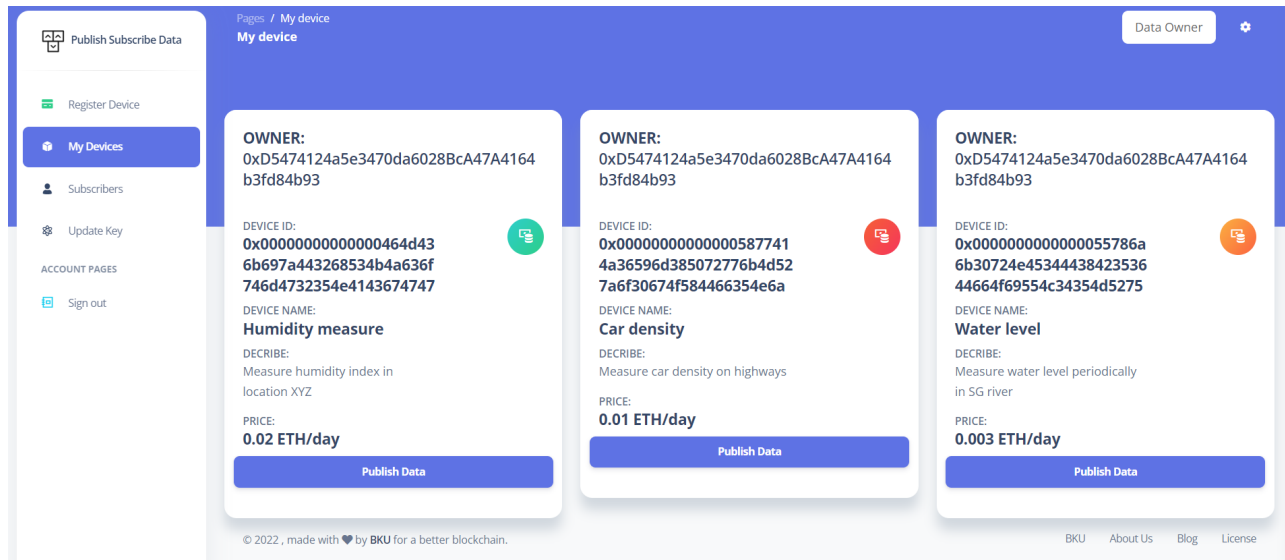


Fig. 5. A snapshot of our WebApp displaying all registered devices that are ready to publish data for sharing.

IV. EXPERIMENTS AND EVALUATION

A. Smart Contract Testing

Since all functions related to publish-subscribe process are executed via smart contracts, it is critical to validate the correction of smart contract's execution. We use Truffle, which is a framework for developing dApps on Ethereum to perform testing on smart contracts written in Javascript. In addition, we utilize Ganache, toolkit in Truffle framework, to build a virtual Ethereum environment on a personal workstation for testing smart contracts before real deployment. We develop numerous testing scripts written in JavaScript that might occur in our proposed model and ensure that all scenarios are passed. Table II summarizes descriptions of 10 testing scenarios conducted in our study.

In addition, smart contracts can hold a lot of values, so they can become the subjects for hackers, such as Re-entrancy attack, that happens when a function in a smart contract calling to a function in an external malicious smart contract. Then, the malicious smart contract can call back to the original one. By that way, attackers can drain the fund of the smart contract. As illustrated in Fig. 6, when we call transfer function, smart contract will check balance of the caller. If the condition is satisfied, it begins to transfer the cryptocurrency to receiver before updating the balance. However, the victim smart contract will call the malicious smart control in order to transfer cryptocurrency. Then, the malicious smart contract will call transfer again since the balance is still not updated. Therefore, attackers can drain all the money inside the victim's smart contract. In our implementation, we adopt checks-effects-interactions pattern in smart contracts to prevent this attack.

B. Security Analysis

Data security of the proposed model is discussed according to four main criteria as follows:

- **Confidentiality:** Combining asymmetric encryption and re-encryption based on public and private key

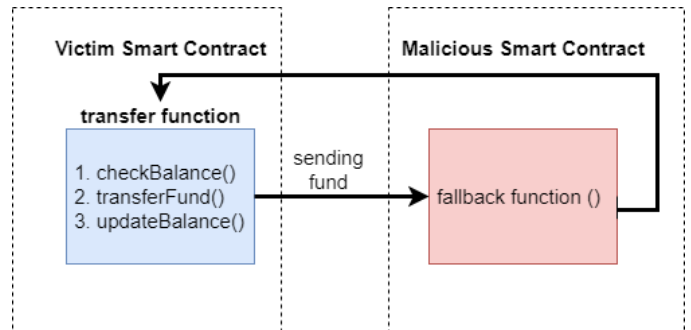


Fig. 6. An illustration of re-entrancy attack.

pairs allows only the DO and authorized DU to access and read the shared data. All the stored secret keys in the database are encrypted with a secret key phrase. Therefore, all the key is confidential.

- **Integrity:** Since IPFS is a distributed content-addressing database, which means that the CID of a file is like the hash of a file. Each content will generate a unique CID; therefore, with a given CID, the integrity of the data is guaranteed. Furthermore, Blockchain assures that transactions are not tampered with and that the entire transaction process is transparent and traceable.
- **Availability:** The decentralized architecture of Blockchain networks and file storage eliminates single points of failure and protects smart contract services from DDoS attacks, hence enhancing service availability.
- **Immutability:** DU can trust that once DU received the data and re-key CID. No one can revoke the privilege including DO thanks to the combination of SC on blockchain and decentralized database (IPFS). Data once publish will be record permanently and DU

TABLE II. SCENARIOS FOR TESTING SMART CONTRACTS

No.	Scenario Name	Description	Expected Result
1	SC-Deploy	Check the deployment of a smart contract.	Successful deployment
2	Dev-Register	DO registers a new device with non-existing ID.	Be able to retrieve device information via device ID.
3	Dev-Register-E	Check if a device is registered with existing ID.	Raise an error message "duplicate device ID".
4	Subscribe	Check if a DU is able to subscribe a registered device.	Subscribe device successfully.
5	Subscribe-E1	DU can not subscribe a registered device due to payment fee.	Raise an error message "not enough money to subscribe".
6	Subscribe-E2	DU can not subscribe a unregistered device	Raise an error message "non-existing device".
7	Publish	Check if a DO is able to publish data belonging to his registered device.	Data can be published.
8	Publish-E1	DO can not publish data because of data duplication.	Raise an error message "data duplication".
9	Publish-E2	DO can not publish data belonging to not-owned devices.	Raise an error message "non-self-owned device".
10	Confirmation	Check if money is transferred to DO after DU confirmed successful data reception.	Payment is successful.

always can retrieve the data.

C. Gas Fee Estimation

Besides security threats, gas fee optimization is another parameter that should be considered. The gas fee will affect a lot on the topic's practicality. One of the ways that we can reduce the cost fee on each function call without changing the structure of the code is to put the appropriate data location of the variable and function argument. There are three locations to store our variables, including *storage*, *memory*, and *calldata*. The fees for those three locations are different, and we can decide where to store them depending on specific purposes.

- The *storage* is the most expensive one since the lifetime of the variable is limited to the contract's lifetime.
- The less expensive one is *memory*, which is mostly used for function arguments since it only lasts in the function call, you can read and modify the variable in memory.
- The least one is *calldata*, which behaves mostly like memory but is not able to modify the variable. In our smart contracts, we mostly chose *calldata* for our complex variable location which will reduce the gas fee for each function call.

Regardless of the content, the length of each data CID is fixed at 46 characters (base58), which means that the CID size is the same for all data. Therefore, the gas price of a smart contract in our model will not depend on the data packet length because we only store the link of data (IPFS CID) on the Blockchain. After performing multiple tests, gas fees of main actions in the proposed model via smart contracts are approximated in Table III.

The cost for each transaction corresponding to each above action via smart contracts will be computed by a product of gas fee and gas price. While the gas fee will not change for the same workload, the gas price will depend on the specific Blockchain network used. Each user can offer their own gas price; the higher the gas price, the more likely their transaction will get picked. The transaction will never be picked if the gas price is too low. Based on the gas fee estimated in Table III, we estimate the transaction's costs in two popular Blockchain

TABLE III. GAS FEE ESTIMATION FOR EACH ACTION VIA SMART CONTRACTS IN OUR PROPOSED MODEL

Action	Gas
Deploy smart contract	2606585
Register devices (data collectors)	174668
Publish data	216969
Subscribe devices (data)	70763
Confirm (two data packets)	58359

networks such as Ethereum and Binance Smart Chain (BSC), as presented in Tables IV and V, respectively.

TABLE IV. ESTIMATED COSTS IN ETHEREUM

Action	Gas	Cost (ETH)	Cost (USD)
Deploy smart contract	2606585	0.0417	69.89
Register devices (data collectors)	174668	0.0028	4.69
Publish data	216969	0.0034	5.70
Subscribe devices (data)	70763	0.0011	1.84
Confirm (two data packets)	58359	0.0009	1.51

*The gas price is 16Gwei and ETH price is 1,676\$ collected on 14-March-2023 from <https://etherscan.io/gastracker>

TABLE V. ESTIMATED COSTS IN BINANCE SMART CHAIN (BSC)

Action	Gas	Cost (BSC)	Cost (USD)
Deploy smart contract	2606585	0.0013	3.98
Register devices (data collectors)	174668	0.0009	0.27
Publish data	216969	0.0011	0.33
Subscribe devices (data)	70763	0.0004	0.11
Confirm (two data packets)	58359	0.0003	0.09

*The gas price is 5Gwei and BSC price is 305\$ collected on 14-March-2023 from <https://bscscan.com/gastracker>

V. CONCLUSION

We presented a Blockchain-based method to resolve a critical issue in today's IoT, data sharing management. The

proposed model allows Data Owners to securely share their data with Data Users in a publish-subscribe fashion via smart contracts on Blockchain. Like other existing works, adopting smart contracts on a decentralized architecture enables a trustless data sharing mechanism without a third party, increasing service availability. Moreover, the proposed method also inherits the characteristics of Blockchain in enhancing data security in terms of confidentiality, integrity, and availability. However, Blockchain and smart contracts still have security threats themselves. In our implementation, we utilized the checks-effects-interactions pattern to prevent the re-entrancy attack on smart contracts. Besides, we also conducted testing scenarios to guarantee the correct execution of smart contracts before actual deployment.

Compared to previous works, the proposed model employs a proxy re-encryption to minimize shared data storage and enhance data security and privacy in a transparent environment like Blockchain. In addition, we successfully developed a DApp to demonstrate and evaluate how the proposed model works.

ACKNOWLEDGMENT

The authors acknowledge Ho Chi Minh City University of Technology (HCMUT), VNU-HCM for supporting this study. The authors also sincerely thank Mr. Long P. Duong and Mr. Phuc M. Nguyen for their assistance in implementing and conducting experiments.

REFERENCES

- [1] Moffett, J., Sloman, M., and Twidle, K. *Specifying discretionary access control policy for distributed systems*. Computer Communications, vol. 13, no. 9, pp. 571-580, 1990. DOI:10.1016/0140-3664(90)90008-5.
- [2] Vincent C. Hu, D. Richard Kuhn, David F. Ferraiolo, and Jeffrey Voas, *Attribute-Based Access Control*, IEEE Computer, vol. 48, no. 2, pp. 85-88, Feb. 2015, DOI:10.1109/MC.2015.33.
- [3] R. S. Sandhu, E. J. Coyne, H. L. Feinstein and C. E. Youman, *Role-Based Access Control Models*, IEEE Computer, vol. 29, no. 2, pp. 38-47, Feb. 1996, DOI:10.1109/2.485845.
- [4] Domenico Rotondi and Salvatore Piccione, *Managing Access Control for Things: a Capability Based Approach*, in: Proceedings of Workshop on Security Tools and Techniques for Internet of Things (SeTTIT), 2012, DOI:10.4108/icst.bodynets.2012.250234.
- [5] Ramos, J.L., Jara, A.J., Marín, L., and Gómez-Skarmeta, A.F. *Distributed Capability-based Access Control for the Internet of Things*, J. Internet Serv. Inf. Secur., vol. 3, no. 34, pp. 1-16, 1993. DOI:10.22667/IJISIS.2013.11.31.001.
- [6] Shorouq Alansari, Federica Paci, and Vladimiro Sassone, *A Distributed Access Control System for Cloud Federations*, in: Proceedings of IEEE 37th International Conference on Distributed Computing Systems (ICDCS), pp. 2131-2136, 2017. DOI:10.1109/ICDCS.2017.241.
- [7] M. Conoscenti, A. Vetr'o, and J. C. De Martin, *Blockchain for the Internet of Things: A Systematic Literature Review*, in: Proceedings of the IEEE/ACS 13th International Conference of Computer Systems and Applications, pp. 1-6, 2016. DOI:10.1109/AICCSA.2016.7945805.
- [8] S. Pal, A. Dorri, and R. Jurdak, *Blockchain for IoT access control: Recent trends and future research directions*. Journal of Network and Computer Applications, Volume 203, 103371, 2022. DOI: 10.1016/j.jnca.2022.103371.
- [9] Khan, M.A., and Salah, K. *IoT security: Review, Blockchain Solutions, and Open Challenges*. Future Gener. Comput. Syst., vol. 82, pp. 395-411, 2018. DOI:10.1016/j.future.2017.11.022.
- [10] P. Patil, M. Sangeetha, and V. Bhaskar, *Blockchain for IoT Access Control, Security and Privacy: A Review*. Wireless Personal Communications, vol. 117, pp. 1815-1834, 2021. DOI:10.1007/s11277-020-07947-2.
- [11] Huckle, S., Bhattacharya, R., White, M., Beloff, N., *Internet of Things, Blockchain and Shared Economy Applications*, Procedia Comput. Sci. 98(C), pp. 461-466, 2016. DOI:10.1016/J.PROCS.2016.09.074.
- [12] D. Worner and T. von Bomhard, *When your sensor earns money: Exchanging data for cash with bitcoin*, in: Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication, pp. 295-298, 2014. DOI:10.1145/2638728.2638786.
- [13] Otto Julio Ahlert Pinno, Andre Ricardo Abed Gregio, and Luis CE De Bona. *Controlchain: Blockchain as a Central Enabler for Access Control Authorizations in the IoT*, in: Proceedings of IEEE Global Communications Conference, pp. 1-6, 2017. DOI:10.1109/GLOCOM.2017.8254521.
- [14] M. J. Baucas, S. A. Gadsden and P. Spachos, *IoT-Based Smart Home Device Monitor Using Private Blockchain Technology and Localization*, IEEE Networking Letters, vol. 3, no. 2, pp. 52-55, June 2021, DOI:10.1109/LNET.2021.3070270.
- [15] M. R. Bataineh, W. Mardini, Y. M. Khamayseh and M. M. B. Yassein, *Novel and Secure Blockchain Framework for Health Applications in IoT*, IEEE Access, vol. 10, pp. 14914-14926, 2022, DOI:10.1109/ACCESS.2022.3147795.
- [16] Sara Rouhani and Ralph Deters, *Blockchain based access control systems: State of the art and challenges*. in: Proceedings of IEEE/WIC/ACM International Conference on Web Intelligence, pp. 423-428, 2019.
- [17] C. Yang, L. Tan, N. Shi, B. Xu, Y. Cao and K. Yu, *AuthPrivacy-Chain: A Blockchain-Based Access Control Framework With Privacy Protection in Cloud*, IEEE Access, vol. 8, pp. 70604-70615, 2020. DOI:10.1109/ACCESS.2020.2985762.
- [18] Lihua Song et al., *A Novel Access Control for Internet of Things Based on Blockchain Smart Contract*, in: Proceedings of IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), Vol. 5, pp. 111-117, 2021. DOI:10.1109/IAEAC50856.2021.9390662.
- [19] Nguyen, T.D.T., Pham, HA., and Thai, M.T., *Leveraging Blockchain to Enhance Data Privacy in IoT-Based Applications*, in: Proceedings of International Conference on Computational Social Networks (CSoNet), LNTCS 11280, pp. 211-221, 2018. DOI:10.1007/978-3-030-04648-4_18.
- [20] Pham, HA, Le, T.-K, Pham, T.N.M, Nguyen, H.Q.T, and Le, T.V. *Enhanced Security of IoT Data Sharing Management by Smart Contracts and Blockchain*, in: Proceeding of 19th International Symposium on Communications and Information Technologies (ISCIT), pp. 398-403, 2019. DOI:10.1109/ISCIT.2019.8905219.

A Review of Trending Crowdsourcing Topics in Software Engineering Highlighting Mobile Crowdsourcing and AI Utilization

Mohammed Alghasham, Mousa Alzakan, Mohammed Al-Hagery

Department of Computer Science, College of Computer, Qassim University, Buraydah 51452, Saudi Arabia

Abstract—Today's modern technologies and requirements make the utilization of crowdsourcing more viable and applicable. It is one of the problem-solving models that can be used in various domains to reduce costs and time. It is also an excellent way to find new and different ideas and solutions. This paper studies the use of crowdsourcing in software engineering and reveals adequate details to highlight its significance. A few recent literature reviews have been published to address specific topics or study general attributes of papers in crowdsourced software engineering. This paper, however, explores all recent publications related to software and crowdsourcing to find the trends and highlight mobile and AI usage in software crowdsourcing. The findings of this paper show that most research papers are in the areas of software management and software verification and validation. The results also reveal that machine learning and data mining techniques are predominant in software management crowdsourcing and software verification and validation. Furthermore, this study shows that the methods and techniques used in general crowdsourcing apply to mobile crowdsourcing except in mobile testing, where there is a need for clustering and prioritization of test reports.

Keywords—Software engineering; crowdsourcing; mobile crowdsourcing; software management; software verification and validation

I. INTRODUCTION

The word engineering in the field of software was inspired by the field of architecture engineering, where the design and building go through defined steps, even though the software has different characteristics. For example, a step in designing a building, for example, should take many considerations, such as budget, before sketching the design of the building. Similarly, when building software, the first step should not be coding the software, especially in large software and systems. The globalization of the current world forced both fields to adopt and use outsourcing, dispatch part of the software or building steps/processes to another company, to compete and evolve, especially when they lack time, workers, expertise, or other reasons. Consequently, the availability of the Internet to a tremendous number of users with various backgrounds and expertise adds more opportunities and challenges to the current working process, which leads to the use of heterogeneous users of the Internet in the working process for both fields. This new methodology was later called crowdsourced and defined by [1] in 2006, and also, different terminologies could refer to the same methodology as provided by [2].

In the crowdsourcing era, in its early days, it has been preliminarily defined as a problem-solving model [3], and

recently, it has been reviewed in a wide range of domains [4]; software engineering, in between, has gained a considerable share in this emerging field [5], [6]. The notion of a problem-solving model for crowdsourcing is suggested by [3] for various applications while providing model examples and denoting contentious points such as crowds diversity, crowds exploitations, and intellectual labor, as well as other topics. A recent comprehensive literature review of all various fields where crowdsourcing has been utilized is authored by [4], which is the first literature to investigate all possible domains related to crowdsourcing. As crowdsourcing has been adopted in various domains, it has become an important topic. One of these domains is software engineering, which started with two publications in 2008 and a total of 509 publications at the end of 2020 [6]. For example, one publication [5] shows the possibility of engaging in empirical studies with crowdsourcing and presents the lesson learned to others for a successful one. Nevertheless, crowdsourcing is a promising technique employed in uncountable areas of SE regardless of other fields, and it is still ongoing research.

A number of recent literature reviews have been published that address crowdsourcing in software engineering. The study of the relationships for both co-authors and citations through social network analysis without considering the contextual content of papers is presented by [6]. Despite that, the research provides cohesive and extensive relationships and connectivity regarding basic publication attributes, such as authors' locations. The paper researches and reviews all of the publications until 2020. Another review [7] provides a baseline understanding of microtasks and explores previous research within crowdsourcing while listing and verifying microtask activities and their categories, which could be helpful for researchers and platforms. For integrating agile development methodology with crowdsourcing, a literature review is carried out to specify challenges and summarize them into five categories [8]. An overview of key processes and platforms, as well as other matters, to facilitate the adaptation of CSSD by organizations and highlight obstacles that make organizations reluctant to recognize CSSD is conducted by [9].

This paper is structured as follows: in this section, an introduction and motivation to crowdsourced software engineering is provided. Section II discusses the research methodology, research questions, and how the research is conducted. Then, Section III describes the literature outcomes, which are divided into areas related to software engineering. In Section IV, the research questions are answered by listing and discussing the findings. The conclusion of this literature review is in Section

V. Finally, suggestions for future work are in Section VI.

II. RESEARCH METHODOLOGY

The research methodology is an essential and well-defined step in literature review papers. Hence, this section presents the research questions and discusses the process for searching and selecting research papers and extracting and approaching the data.

A. Research Questions

This section provides answers to the following three listed questions:

- **RQ1: What are the directions and trends in crowd-sourced software engineering?** The aim is to investigate and analyze the recent topics in crowdsourced software engineering.
- **RQ2: Did the papers focus on mobile crowdsourcing? Can general crowdsourced software engineering methods be used in mobile crowdsourcing?** The aim is to examine mobile crowdsourcing and conventional methods within mobile software engineering crowdsourcing.
- **RQ3: Did the papers in the review use AI? What type of AI did the papers use? In what areas did the papers use AI?** The aim is to explore the algorithms that the literature uses and in which areas of software engineering they are employed.

B. Conducting the Research

The first step in each research topic is to choose relevant keywords to find all relevant papers. Unrelated keywords could lead to small numbers of papers without any further hope of obtaining additional suitable research papers. Subsequently, gradually, more keywords are added, and the list of all keywords that are used within advanced search queries is as follows:

- (“Crowdsourcing” OR “crowdsourced” OR “crowd” OR “crowdsource”) as (CrowdKeywords)

The first query, (CrowdKeywords), is used in conjunction with the following ones:

- (CrowdKeywords) AND (“Software Engineering”)
- (CrowdKeywords) AND (“Software Development”)
- (CrowdKeywords) AND (“Software Design”)
- (CrowdKeywords) AND (“Requirements” or “crowdRE”)
- (CrowdKeywords) AND (“Software” or “testing” or “test” or “defect”)

As shown in Fig. 1, the search is first established through scholarly search engines online, without engaging in manual search activities, such as printed journals, and excluding books and thesis. There are a large number of databases. The papers are collected from IEEE, ACM, Science Direct, SpringerLink,

Wiley Online Library, MDPI, AIMS Sciences, and Airiti. Additionally, this review focuses solely on recent papers and contributions, so any research paper published before 2022 is filtered out. After the preliminary collection of more than 100 publications, stored in a reference manager software, by reading just titles and keywords, all the collection papers are validated against the following criteria:

- 1) Papers must be very recent, and there are no duplications.
- 2) For quality control, remove preprinted or publications that are not peer-reviewed.
- 3) To ensure the relevance of collected publications, at least two authors review each paper’s abstract to check whether the paper is related to the scope of the research or not.

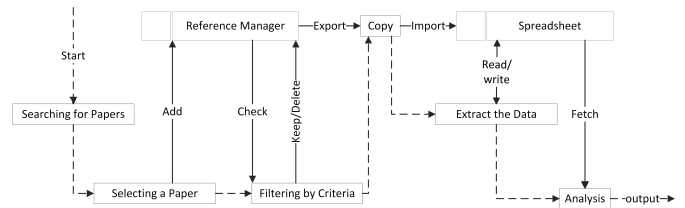


Fig. 1. Research methodology.

Next, the remaining 41 papers are exported from the reference manager to a spreadsheet. The spreadsheet contains the main attributes of each research paper and other details related to this research, such as the paper title, type, library, authors, date, the software engineering area, subarea, keywords, aim, objectives, main points, and summary. When the paper is selected to be included in the spreadsheet file, all details need to be obtained and filled out in the spreadsheet. With the help of the reference manager software, the keywords are also exported instead of manually filling them out. Still, several papers’ keywords are manually extracted, which indicates inconsistent standardization between databases. These keywords are processed differently, and a program is developed that is executed inside the spreadsheet to handle the data. The program simply fetched all keywords with their corresponding paper and then clustered them based on each keyword.

III. LITERATURE OUTCOMES

In this section, the papers selected for review are discussed. This literature review is divided into five subsections based on the problems the research papers solve. The subsections are software management, software specification, software development, software verification and validation, and software evolution. Fig. 2 depicts these subsections. Furthermore, two subsections: software management and software verification and validation, are further divided into subsubsections.

A. Software Management

Since software management is an extensive topic, its topics are further divided into subsections according to what is discussed in the selected papers. These subsections are productivity and motivation, task and crowd worker recommendation, trust issues, task pricing, and project documentation. Fig. 3 shows the subsections of software management.

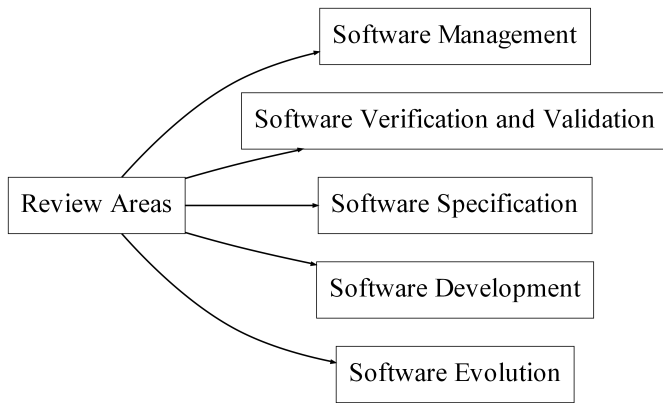


Fig. 2. Literature outcome main areas.

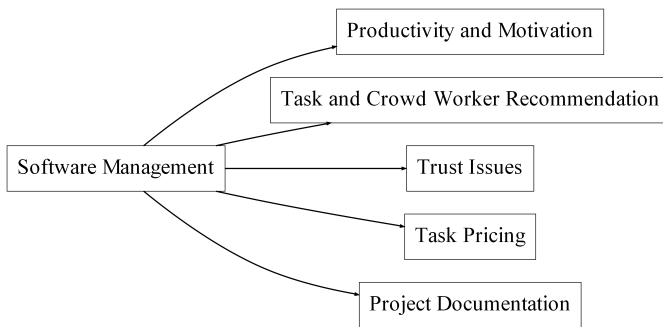


Fig. 3. Software management subareas.

1) *Productivity and motivation*: The authors of [10] investigate the collaboration of crowd workers in transient teams versus solo developers using a data-driven empirical study. The results show that the experience of teamwork affects the teams' performance in both the long and short term. In particular, the results show that individuals in each team can learn in the short term by getting support and sharing ideas with other members. Furthermore, in the long term, the members learn from working with other experienced team members. The paper contributes strategies for collaborative contest designers, platform operators, and crowd workers. These strategies include increasing the complexity of the contests, emphasizing virtual teams in contests, or reducing the number of total medals.

While [10] examines the effects of working solo as a developer or in teams on performance and learning, the authors of [11] study the effects of game elements that exist in several crowdsourcing platforms on individual contributors. Multiple crowdsourcing platforms use game elements such as contests, leaderboards, and rankings with the intent to motivate crowd workers. The results show that there is a different effect on the performance and effort of workers depending on their abilities. Data from TopCoder shows that feedback has a positive effect on high-scoring workers and a negative effect on low-scoring individuals.

Unlike [10], [11], the writers of [12] define, explain, and justify the area and use of the microtasks programming approach, regarding functions programming, on the factors

such as team size, the time needed for new developers, and the velocity of the whole project. Moreover, they have conducted an experiment to study the positive and negative sides without focusing on areas related to the design and maintenance of microtasks. The experiment has shown several advantages of adopting microtasks, especially for the short project schedule. Overall, this study establishes a path for software corporations and encourages them to use microtasks, which could lead to the involvement of external developers, crowdsourcing in particular.

2) *Task and crowd worker recommendation*: The authors of [13] state that crowd workers often choose their testing tasks from whatever is immediately available. This abrupt choice could lead to wasted time and effort for both the tester and the requester. Furthermore, it could lead the tester to the inability to discover and find the bugs in the assigned task. Thus, the authors of this paper propose a context-aware personalized crowdsourced software testing method for task recommendation named PTRec. This method uses contextual, historical data and the preferences of the tester to recommend the appropriate task. PTRec consists of two models that are able to extract 60 features automatically to help the tester choose a suitable task. The goal is to reduce wasted efforts and the number of unpaid tasks. PTRec also uses the random forest learner technique to find the proper testing task which matches the workers' interests and expertise. Tests on 2404 crowd workers and 636 tasks reveal that this approach has 82% precision and saves the efforts of exploring tasks by 81%.

In addition to [13], another research [14] has proposed a new recommendation model which considers users' preferences. This method is a capability-corrected long- and short-term attention network (CLEAN). It outperforms the existing traditional models. The model considers the gradual interest changes of workers and the constraints and skills needed to perform the tasks while incorporating the contextual data of tasks besides common attributes.

According to the authors of [15], existing studies use one-time recommendations based on the knowledge of the worker at the start of a new task. Furthermore, they assert that this recommendation has a popularity bias. In other words, the methods in these studies recommend almost all of the tasks to users with the highest experience. To remedy these issues: this study proposes iRec2.0, which is a context- and fairness-aware in-process crowd worker recommendation method. iRec2.0 achieves its stated goals by modeling the dynamic testing context, using a learning-based technique, and applying a multi-objective optimization component. The outcomes of the evaluations show that this method has the potential to divide the tasks fairly among the users, decrease the testing process time, and save costs.

In [16], the authors assert that the quantity and quality of the completed tasks are directly affected by task recommendation. The authors also state that previous task recommendation modules focus on one user only per task. Unlike [13], [15], the authors of [16] present a method to recommend tasks and coworkers for these tasks. This method operates on the user's performance history on previous tasks in combination with a social network component for the coworker recommendation. This method checks and informs users if it is best to complete this task independently or if they should seek help from a

friend in their social network. This method factorizes a user-task rating matrix to get the latent matrix. Then it applies a greedy method to select tasks for each user. The method calculates the intimacy data and the extroversion data of users with their friends in the social network to recommend coworkers. The results by the authors show that it outperforms the current existing algorithms.

Crowdsource contribution is open to every qualified person. This openness could pose a problem of malicious workers and malicious task submissions. To tackle these problems, the authors of [17] propose a method named Outlier Detection for Streaming Task Assignment. Its goal is to detect malicious crowd workers. This technique uses an evolving time series to model the arrival of workers and their task submissions. This framework has a novel method based on Generative Adversarial Network (GAN), which is socially aware and can work with time series. The authors also propose a novel method to train the loss functions of GANs with social awareness. It also has the capability to assign tasks to users similar to [13]–[16]. This method uses a greedy algorithm to improve the efficiency of the process of task assignment.

Whereas [13]–[17] focus on task assignment, [18] focuses on the reliability of recruits. In [18], the authors investigate four different crowdsourcing platforms and a computer science (CS) mailing list to determine the reliability of their recruits for empirical software development studies. The crowdsourcing platforms are Prolific, Appen, Clickworker, and Mechanical Turk. The authors' criteria of reliability are programming skills, privacy, and security attitudes. For the university CS students, the authors also consider self-efficacy. The results show that while university CS students rated themselves lower than other crowdsourcing participants in secure development and self-efficacy, 89% of them answered all programming skills correctly. Furthermore, the study shows that university CS students are the most cost-effective recruits.

Out of all the studies on task management in this paper, only one study highlights the need for coordination in the platform for doing tasks related to software design between designers and clients. Therefore, it identifies all potential coordination limitations encountered in the process and promotes coordination propositions with the help of a questionnaire. The feedback from participations through the questionnaire verifies the limitations and welcomes the purpose solution to alleviate the lack of coordination in platforms [19]. Moreover, another study proposes four steps to task flow from the beginning of constructing the task until the aggregation of the results, and this solution is evaluated in two ways. Also, the crowdsourcing platforms have the advantage of incorporating and implementing proposed approaches [20].

3) *Task pricing*: Is giving crowd workers the ability to choose their preferred incentive will result in better performance and solution quality? Is one type of incentive for all participants an optimal motivator? In [21], the authors of this paper empirically investigate the effects of giving the choice of reward to participants on the quality of the solution submitted. The results of this work show that when participants can choose their preferred incentive, they will spend more time on their assigned tasks and produce better-quality solutions in contrast to participants who offered one type of incentive. The results show the importance of having a flexible reward

structure and allowing participants to select what matches their motives.

Both [21], [22] state that personalized pricing can yield better resulting tasks than common pricing, which is pricing with no personalization. The authors of [22], though, remark that personalized pricing is arduous to incorporate into some systems due to its complexity. Therefore, [22] investigate two schemes: personalized pricing per worker and common pricing with bonus payments after task completion to explore their impacts. The results show that with the proper bonuses, common pricing is close to an approximation of optimal personalized pricing.

The authors of [23] argue that instead of having a fixed price for tasks, there should be a dynamic system for pricing tasks to incentivize crowd workers. Accordingly, The authors of this paper propose a system called CrowdPricer, which gives, in addition to the base payment for the accomplished tasks, bonuses for completing tasks which is the recommended method by [22]. In addition, CrowdPricer increases the utility expectation of the requester to guarantee profits. To achieve its stated goals, CrowdPricer learns the effect of bonuses on the quality of the delivered tasks using deep time sequence modeling. The authors' experimentations using a crowdsourcing platform and simulations show that using CrowdPricer results in higher-quality task solutions and maximizes the utility of the requester.

4) *Trust issues*: Several critical and ordinary trust issues are raised in the crowdsourcing implementation. The authors [24] present these issues by conducting a survey with practitioners, listing nine critical issues as “deficient assistance to best practices”, “malicious code”, “lack of licensed software utilization”, “loss of data”, “network security risks”, “quality of workers”, “social attacks”, “crowd legal action”, and “loss of intellectual property”. Additionally, the study results are validated via a focus group of four experts in academia. One of the issues of intellectual property is investigated in the context of testing reports by one paper [25]. It proposes a system for intellectual rights confirmation with the integration of blockchain and the implementation of other methods. This system would overcome problems, such as code plagiarism, and prevent unwanted modifications of data with the help of the blockchain decentralized methodology [25]. Another paper [26] indicates the need for implementing blockchain to diminish unwelcome behavior raised by centralized systems. However, it showed that the quality of previous works regarding keeping traceability or increasing privacy is less than required. Therefore, it proposes STPChain based on blockchain, which better preserves traceability and improves privacy, including preventing wrong actions.

5) *Project documentation*: The paper [27] studies the effect of using various types of media as a documentation type and focuses solely on instructional screencast documentation. As the popularity of this type has recently increased, especially for crowd-based content and new developers, the paper suggests a platform for this content. In addition, the platform could have essential functionalities such as making the content searchable and linking its content to other artifacts.

B. Software Specification

As there is a vague overlap between crowd-based requirements engineering (CrowdRE) and market research (MR) primarily caused by the incremental use of automation, one study explains and identifies them. Furthermore, after providing various scenarios and equivalent implementation of both CrowdRE and MR, the study implies the overall benefits of CrowdRE, which could be sufficient [28].

Specialized public web forums and general user stories format are a great way to elicit crowd opinions and experiences on many topics. The authors of [29] use Reddit forums to collect requirement engineering data by analyzing the discussion in the forums. In particular, the authors propose crowdsourcing requirement engineering by valuation argumentation (CrowdRE-VArg) to identify and prioritize issues, design changes, and new features and decide on the appropriate requirements. CrowdRE-Varg uses machine learning and natural language processing to analyze end-users supporting and attacking arguments in discussions from users' posts on Reddit. Their results show the validity of the approach for using Reddit as a platform for rational mining and the eliciting of opinions. Another paper [30] analyzes user stories and proposes CREUS, Crowd-based Requirements Elicitation with User Stories, as an iterative process practical design for conducting pull feedback after engaging in three case studies. It provides qualitative analysis of user stories or feedback as the main contribution besides quantitative results as usual case studies.

Software requirement engineering (RE) is a challenging process, and it requires the constant availability of the stakeholders, which is not guaranteed. To handle the problems of RE, the author of [31] proposes a conceptual framework that combines the crowdsource software development (CSSD) approach with the SCRUM software development approach. The framework collects the data from the crowd at large, which increases efficiency and reduces costs. This framework consists of four main layers designed to use the features of both approaches. The layers deal with document preparation, prioritization of tasks, planning, design, and retrospective meeting. On the other hand, [32] present specific challenges regarding requirements for a specific area. The authors show and discuss the needs of older well-being adults for intelligent assistance systems through crowd-based requirements engineering (crowd-RE). Also, it demonstrates the crowd-RE process and some challenges in this area.

One research [33] studies prototype validation. It develops a platform that uses the crowd to obtain feedback and validate the prototype iteratively before actual development. First, it conducts a design science study to address the vague of applying crowd-workers and prototype validation in platforms. Then, through the formed knowledge and implementation, it develops a platform that tackles the difficulties. Moreover, the study is valuable for building new or enhancing current mechanisms with the crowd-validation process.

C. Software Development

In [34], the authors aim to identify the percentages of vulnerabilities in code submitted by participants in code in competitive programming (CP) platforms. This paper focuses on data, 6.1 million submissions to be exact, from the CodeChef

CP platform. The results show that 34.2% of submissions have software vulnerabilities. The authors did not find conclusive evidence to correlate the number of vulnerabilities with the leaderboard position of the participant. Furthermore, the study shows that participants do not follow secure coding practices, and even when a participant with perfect scores reattempts the task, the study shows there are no security improvements in the new submission. One way to mitigate these issues is to use crowdsourcing. The authors of [20] suggest that instead of using automated or manual anti-pattern detection methods, which consume time and lack certainty, the use of crowdsourcing and propose four steps for task flow.

D. Software Verification and Validation

On the subject of software verification and validation, the selected papers solve problems in software testing and usability, test report clustering and prioritization, and quality of defects reports. Fig. 4 shows the subareas of software verification and validation.

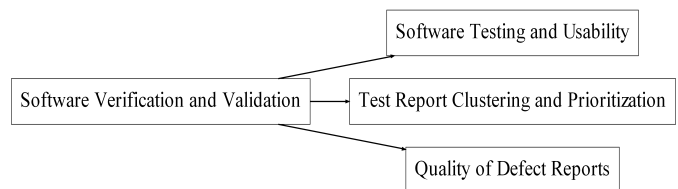


Fig. 4. Software verification and validation subareas.

1) *Software testing and usability*: After analyzing the problems that developers face in crowdsourced software development (CSD), the authors of [35] concluded that crowdsourcing is more fitting for software testing than software development. The authors state that the main advantages of crowdsourced software testing (CST) are reducing the time and cost of software testing. Furthermore, the authors express that better tools make developers work efficiently. Hence, the paper proposes a new testing program that incorporates crowdsourcing and open-source sharing techniques. An example of using crowdsourcing in testing is [36]. The authors of [36] did a comprehensive study to compare the cost and time of using novice crowd and expert heuristics in usability inspection. A single expert's heuristic usability inspection leads the novice crowd. The results show that, on average, both methods detect the same usability issues. However, the novice crowd method takes less time to identify the problems and costs less than the expert heuristic usability inspection.

To overcome the challenges of testing human-AI interactions and collaborations, the authors of [37] propose a Human-AI Intergation Testing framework (HINT). HINT is a crowd-based framework that uses a humans-in-the-loop workflow to test AI-based experiences. This framework aims to solve the drawbacks of existing methods by simulating AI experiences that evolve over time, allowing rapid testing, providing early feedback during the development phase, and evaluating crowd workers and AI in offline testing. In addition, to overcome problems that exist in current testing implementation for IoTs, another research [38] develops a new crowdsourcing test system oriented toward the Internet of things with the integration of blockchain technologies as a potential solution. The system

consists of two modes, online and offline testing. The device or devices in the online mode, which are the main focus of interest, are real devices. In this case, testers engage and test the devices online with multi-thread technology implemented to allow concurrent testers, which is helpful, especially for limited hardware resources. The offline mode is to test the devices physically. Besides all security that the system provides, it contributes to a dependable online testing system for IoTs, especially with the lack of online testing for IoTs.

Regarding fault localization and exploiting the power of crowdsourcing, only one research [39] has investigated this area. It has a distinguished and unfamiliar approach for automated fault localization (AFL) in crowdsourcing software engineering by exploiting the solutions of all works and making them one set of referenced solutions. The main point is that when encountering fault statements, each statement in buggy programs is referenced and evaluated with an equivalent statement from other solutions.

2) *Test report clustering and prioritization:* The authors of [40] assert that an issue of the previous research papers dealing with the clustering of test reports is that the papers do not take into account the semantic connection between the screenshots and text in the analyzed test reports, which results in suboptimal results, especially in the deduplication of test reports. Therefore, This paper proposes a method using semi-supervised clustering using deep image understanding to analyze crowdsourced mobile application test reports. This method is SemCluster. SemCluster creates semantic binding rules from the semantic connection between screenshots and text descriptions in test reports. The results of this paper show that SemCluster outperforms the state-of-the-art method in six metrics of clustering results. Another approach [41] uses a fused features approach after obtaining text and screenshots features and then using common classification algorithms.

Liu et al. in [42] state that in addition to the large number of test reports produced by crowdsourcing that needs inspecting, one specific issue of mobile application test reports is that they have more screenshots than text descriptions of the tests. In addition to this paper, [40], [41], [43], [44] deal with these issues as well. To solve the issues of the number of reports and screenshots, the authors of [42] propose a novel method to understand text and images to cluster test reports. This method uses natural language processing to calculate the distance between reports. It also uses Spatial Pyramid Matching (SPM) to compute the similarity of the screenshots in the reports. The authors tested the method on 1400 screenshots and more than 1600 test reports from six industrial crowdsourced projects. Tests show that this method results in up to 37% improvement over the baseline in the average percentage of faults detected (APFD). Moreover, only the following paper [41] points out that existing automatic test report classification techniques are incompatible with crowdsourced mobile test reports as they contain incomplete texts, as well as the previously mentioned screenshots.

In [43], the authors propose and evaluate a method adapted from the prioritization of test reports in regression testing. This method sorts test reports in two phases. First, to process the text of the reports, this method uses natural language processing and word segmentation. Second, to prioritize the test reports, the paper uses a combination of a genetic algo-

rithm, two greedy algorithms, and an adaptive random test case prioritization algorithm. It aims to make it easy and efficient for developers to check reports according to their priority. The results show that this method has promising performance in prioritizing the test reports, with an average percentage of faults detected (APFD) of more than 0.8.

The authors of [44] devise a new method called DivClass to prioritize test reports. DivClass combines diversity and classification strategies to order the reports for inspection. A feature of this method is that it handles duplicate test reports similar to the method proposed in [40], [42]–[44] in which they use natural language processing to analyze the test reports in one of their method steps. The next step in [44] is to build a similarity matrix using an asymmetric computation strategy. The final step consists of the previous two steps to prioritize the reports. The authors state that it reduces the number of tests to inspect, reducing the inspection cost. It also improves DivRisk, the state-of-the-art method, by 14.12% on average and has 0.8887 APFD.

3) *Quality of defect reports:* Three studies [45]–[47] provide ways to enhance the quality of the defect reports generated by usually non-expert crowds. The first research [45] studies the reports, which contain a good and bad description of bugs, submitted by crowdsourcing, usually non-professional testers. After it shows possible quality indicators, the paper proposes CTRQS as a framework to qualify test reports using analytical indicators based on dependency parsing. Therefore, in the end, just the reports that describe the defect better should be processed for localization and fixing. The second paper [46] attempts to generate more promising defect reports results by adding more than one tester participating together to find and report defects instead of one tester working alone. The result shows an enhancement regarding the quality of the final report by decreasing the invalid reported defects and increasing the report of difficulty defects. On the other hand, the third research [47] states the need for improving the crowd-workers instead of only establishing techniques for the quality of crowdsourced testing reports. Also, it indicates that enhancing the knowledge and abilities of testers from the beginning will provide better-quality reports. Therefore, this study proposes an assistant approach to suggest, guide, and educate crowds by exploiting Android's automated testing results.

In a different area of software quality, one research [48] suggests using crowdsourced workers, who meet minimum quality requirements, as a third party to evaluate and certify software meant for public users based on the available software documents and templates. Then, a quantitative quality score is assigned to the software-specific version, and the evaluation process is repeated after each considerable change in the current software version. Another area of research [49] claims that almost all previous techniques study the duplicated defects without false defects. Hence, this paper contributes by using duplicated defects to build a model which can provide a high estimation accuracy for valid defects. Furthermore, the model accuracy increases when the approach is applied to crowdsourced testing.

E. Software Evolution

In [50], Reis et al. use supervised machine learning methods with crowdsourced data collected over three years to identify code smells. The authors focus on Java code and three types of code smell, which are long methods, god classes, and feature envy. The data is collected from about a hundred teams, each team with an average of three members. The results of the papers prove the feasibility of crowdsemling using supervised machine learning techniques applied to data collected from software developers (wisdom of the crowd). However, the authors state that further studies are needed to cover other types of code smells. The authors are currently developing an Eclipse IDE plugin which should simplify the crowdsourcing process. This plugin collects data about the code, identifies code smells, and gets the developer’s opinion regarding the data detected by the plugin.

IV. RESULTS AND DISCUSSION

RQ1: What are the directions and trends in crowd-sourced software engineering? After collecting and analyzing the 41 papers published in 2022, This research reveals that the collected papers discuss the following five areas: software management, software specification, software development, software verification and validation, and software evolution. The most discussed area in crowdsourced software engineering is the area of software management. Out of the 41 papers, 17 papers deal with various software management problems. This review also determines that the least papers, one paper, are in the area of software evolution, even though it is an essential and costly activity in software engineering. Table I shows the areas, the count of the papers, and the paper selected in these areas.

TABLE I. AREAS OF THE SELECTED PAPERS

Area	Selected	Papers
Software Management	17	[10], [11], [13]–[27]
Software Specification	6	[28]–[33]
Software Design & Implementation	2	[20], [34]
Software Verification & Validation	15	[35]–[49]
Software Evolution	1	[50]

Continuing the discussion on the directions and trends of the research in crowdsourced software engineering, in this review, both software management and software verification and validation are divided into subareas based on the collected papers. Table II shows the subareas of software management, the collected and selected papers, and the number of papers in each of the subareas. Software management has five subareas shown in Table II. A look at the areas of software management shows that most research papers in this literature review, which are 7 out of 17, are on task and crowd worker recommendations. There are three papers in each of the following fields productivity and motivation, task pricing, and trust issues in software. The least number of papers which is one study is on project documentation.

In the subarea of software verification and validation, there are 15 papers spanning three subareas. There exist five studies

TABLE II. SUBAREAS OF SOFTWARE MANAGEMENT

Subareas	Selected	Papers
Productivity & Motivation	3	[10]–[12]
Task & Worker Recommendation	7	[13]–[19]
Task Pricing	3	[21]–[23]
Trust Issues	3	[24]–[26]
Project Documentation	1	[27]

in each one of the three subareas. Table III shows the details of the subareas and papers of software verification and validation.

TABLE III. SUBAREAS OF SOFTWARE VERIFICATION AND VALIDATION

Subareas	Selected	Papers
Testing & Usability	5	[35]–[39]
Test Report Clustering & Prioritization	5	[40]–[44]
Quality of Defect Reports	5	[45]–[49]

The results of the Tables I, II, and III lead us to conclude that the most trending topics in crowdsourced software engineering are two. First, the topics that deal with software management especially managing tasks, workers, and motivation. Second, all the subareas of software verification and validation. These subareas are software testing and usability, test report clustering and prioritization, and quality of defect reports.

Furthermore, an analysis of all keywords in the papers was conducted to investigate the trending topics from different viewpoints. The keywords are also used to see how diverse the papers are in the context of subtopics. As a result, the number of different keywords is more than 180 words from 41 papers. Table IV shows the keywords that appear in three papers or more, only the top eight keywords are listed, and the common keywords to the primary research topic are excluded, such as crowdsourcing, software engineering, and software. The results agree with the previous conclusion that software management and testing are the most dominant topics. Moreover, the selected papers in this review are more diverse, covering various subtopics, as there are 162 keywords that are unique and used only in one paper.

TABLE IV. MOST COMMON KEYWORDS THAT APPEAR IN PAPERS

Keywords	Selected	Papers
Crowdsourced testing	9	[13], [15], [25], [41], [42], [44], [45], [47], [49]
Task Analysis	7	[13], [19], [42]–[44], [47], [49]
Testing	5	[13], [37], [42], [45], [47]
Computer Bugs	5	[13], [42], [44], [45], [47]
Software Testing	4	[43], [44], [46], [49]
Requirements Engineering	3	[28], [31], [32]
Mobile Applications	3	[42], [45], [47]
Software Quality	3	[20], [26], [50]

RQ2: Did the papers focus on mobile crowdsourcing? Can general crowdsourced software engineering methods be used in mobile crowdsourcing? Mobile development comes with its own set of problems, especially in software testing [40]–[44] describe these issues. In this review, there are ten papers that focus exclusively on mobile crowdsourcing, specifically mobile testing. The remaining papers, 31 papers,

focus on general software engineering crowdsourcing, which is applicable to mobile crowdsourcing as well. Table V lists the papers with their focus area. These results lead us to deduce that the crowdsourcing methods in general software engineering crowdsourcing are suitable for mobile crowdsourcing, except for mobile testing methods highlighted by [40]–[44].

TABLE V. PAPERS FOCUS

Focus	Selected	Papers
Mobile Crowdsourcing	10	[13], [15], [16], [40]–[45], [47]
General Crowdsourcing	31	[10]–[12], [14], [17]–[39], [46], [48]–[50]

RQ3: Did the papers in the review use AI? What type of AI did the papers use? In what areas did the papers use AI?

TABLE VI. AI IN PAPERS

AI Methods	Selected	Papers
ML & DM	14	[13]–[17], [23], [29], [30], [40]–[44], [50]
None	27	[10]–[12], [18]–[22], [24]–[28], [31]–[39], [45]–[49]

Yes, 14 papers use AI techniques, while 27 do not. Specifically, the papers use either Machine Learning (ML) or Data Mining (DM) methods. Table VI list the papers that use the AI methods. However, not all AI is applied to all the areas of crowdsourced software engineering. Table VII shows the areas and papers in which AI methods are employed. In Table VII, the results show that test report clustering and prioritization and task and worker recommendation use AI the most, with five papers each. Two papers in software specification use AI. Both task pricing and software evolution have one paper each that uses AI techniques.

TABLE VII. AI AREAS IN THE SELECTED PAPERS

Area	Selected	Papers
Test Report Clustering & Prioritization	5	[40]–[44]
Task & Worker Recommendation	5	[13]–[17]
Software Specification	2	[29], [30]
Task Pricing	1	[23]
Software Evolution	1	[50]

V. CONCLUSION

This literature review examines and studies the latest papers in crowdsourced software engineering to find the current trends and directions of the research literature. This paper focuses exclusively on all the publications of 2022 to get a clear and accurate picture of the crowdsourcing landscape. This review also answers a number of relevant and current questions. It answers whether the papers focus solely on mobile crowdsourcing and whether general crowdsourced software engineering methods apply to mobile crowdsourcing. Furthermore, it discusses the question of AI usage in the papers. In particular, this research checks whether the selected papers incorporate machine learning or data mining techniques into their proposed crowdsourcing solutions. The results of this literature review show that the largest number of contemporary research focuses on

software management and software verification and validation. In mobile crowdsourcing, the results show that while general crowdsourcing methods work for most mobile crowdsourcing activities, mobile testing requires specific techniques to deal with the large number and the nature of tests. The results also show the papers that use machine learning and data mining methods to tackle specific crowdsourced software engineering areas.

VI. FUTURE WORK

One of the least discussed topics in software engineering crowdsourcing is secure software development. Crowd workers come from diverse places and have different programming and security backgrounds. Therefore, they will have various goals to achieve and different experiences. In this research, paper [34] shows the percentage of vulnerabilities in competitive programming platform submissions. The study did not find a connection between the leaderboard position of the participants and the number of vulnerabilities in their submitted code. Furthermore, even the resubmissions of full-scoring tasks did not have security improvements. These results beg the following research questions: Is it possible to have a crowdsourcing platform for secure software development? How to incentivize crowd workers to submit secure code? Can the incentivization techniques in [21]–[23] be used to encourage secure coding practices? Will it affect software verification and validation? These are all questions that need further research.

There are many possible areas of improvement in the quality of crowdsourcing that can be investigated. One of them is the minimum number of crowd workers engaging in one task. Each type of task requires a different number of workers, and in this way, the power and quality of crowd wisdom can be exploited in a cost-effective manner. Moreover, the crowds must be certified for the required type of tasks before joining crowdsourcing platforms. Certification will ensure higher quality workers. Alternatively, to encourage worker certification, certified workers can be paid more than non-certified ones. There could be several types of certifications depending on the type of platform or task. In addition, to our knowledge, there are no established crowdsourcing standards of best practices that ensure continuity and portability for both tasks and processes.

ACKNOWLEDGMENT

The researchers would like to thank the Deanship of Scientific Research, Qassim University, for funding the publication of this project.

REFERENCES

- [1] J. Howe, “The rise of crowdsourcing,” *Wired magazine*, vol. 14, no. 6, pp. 1–4, 2006, number: 6.
- [2] A. Doan, R. Ramakrishnan, and A. Y. Halevy, “Crowdsourcing systems on the World-Wide Web,” *Communications of the ACM*, vol. 54, no. 4, pp. 86–96, Apr. 2011. [Online]. Available: <https://doi.org/10.1145/1924421.1924442>
- [3] D. C. Brabham, “Crowdsourcing as a Model for Problem Solving: An Introduction and Cases,” *Convergence*, vol. 14, no. 1, pp. 75–90, Feb. 2008, publisher: SAGE Publications Ltd. [Online]. Available: <https://doi.org/10.1177/1354856507084420>

- [4] N. Kasturi, S. G. Totad, and G. Ghosh, "Analysis on Potential Use of Crowdsourcing in Different Domain Using Metasynthesis," in *Emerging Technologies in Data Mining and Information Security*, ser. Lecture Notes in Networks and Systems, P. Dutta, S. Chakrabarti, A. Bhat-tacharya, S. Dutta, and V. Piuri, Eds. Singapore: Springer Nature, 2023, pp. 747–756.
- [5] K. T. Stolee and S. Elbaum, "Exploring the use of crowdsourcing to support empirical studies in software engineering," in *Proceedings of the 2010 ACM-IEEE International Symposium on Empirical Software Engineering and Measurement*, ser. ESEM '10. New York, NY, USA: Association for Computing Machinery, Sep. 2010, pp. 1–4. [Online]. Available: <https://doi.org/10.1145/1852786.1852832>
- [6] A. Alabduljabbar and S. Alyahya, "Leveraging Social Network Analysis for Crowdsourced Software Engineering Research," *Applied Sciences*, vol. 12, no. 3, p. 1715, Jan. 2022, number: 3 Publisher: Multidisciplinary Digital Publishing Institute. [Online]. Available: <https://www.mdpi.com/2076-3417/12/3/1715>
- [7] M. Zulfiqar, M. N. Malik, and H. H. Khan, "Microtasking Activities in Crowdsourced Software Development: A Systematic Literature Review," *IEEE Access*, vol. 10, pp. 24 721–24 737, 2022, conference Name: IEEE Access.
- [8] S. Qayyum, S. Imtiaz, and H. H. Khan, "Challenges of Agile–Crowd Software Development: A Systematic Literature Review," *Journal of Circuits, Systems and Computers*, p. 2330001, 2022, publisher: World Scientific.
- [9] D. d. C. Candria and R. M. d. Araujo, "Crowdsourcing Software Development - a possible path?" in *XVIII Brazilian Symposium on Information Systems*, ser. SBSI. New York, NY, USA: Association for Computing Machinery, Jun. 2022, pp. 1–8. [Online]. Available: <https://doi.org/10.1145/3535511.3535532>
- [10] K. Huang, J. Zhou, and S. Chen, "Being a Solo Endeavor or Team Worker in Crowdsourcing Contests? It is a Long-term Decision You Need to Make," *Proceedings of the ACM on Human-Computer Interaction*, vol. 6, no. CSCW2, pp. 494:1–494:32, Nov. 2022, number: CSCW2. [Online]. Available: <http://doi.org/10.1145/3555595>
- [11] M. Tsvetkova, S. Müller, O. Vuculescu, H. Ham, and R. A. Sergeev, "Relative Feedback Increases Disparities in Effort and Performance in Crowdsourcing Contests: Evidence from a Quasi-Experiment on Topcoder," *Proceedings of the ACM on Human-Computer Interaction*, vol. 6, no. CSCW2, pp. 536:1–536:27, Nov. 2022, number: CSCW2. [Online]. Available: <http://doi.org/10.1145/3555649>
- [12] E. Aghayi and T. D. LaToza, "A controlled experiment on the impact of microtasking on programming," *Empirical Software Engineering*, vol. 28, no. 1, p. 10, Nov. 2022. [Online]. Available: <https://doi.org/10.1007/s10664-022-10226-2>
- [13] J. Wang, Y. Yang, S. Wang, C. Chen, D. Wang, and Q. Wang, "Context-Aware Personalized Crowdttesting Task Recommendation," *IEEE Transactions on Software Engineering*, vol. 48, no. 8, pp. 3131–3144, Aug. 2022, number: 8 Conference Name: IEEE Transactions on Software Engineering.
- [14] Z. Peng, D. Wan, A. Wang, X. Lu, and P. M. Pardalos, "Deep learning-based recommendation method for top-K tasks in software crowdsourcing systems," *Journal of Industrial and Management Optimization*, pp. 0–0, Nov. 2022, publisher: Journal of Industrial and Management Optimization. [Online]. Available: <https://www.aims sciences.org/en/article/doi/10.3934/jimo.2022223>
- [15] J. Wang, Y. Yang, S. Wang, J. Hu, and Q. Wang, "Context-and Fairness-Aware In-Process Crowdworker Recommendation," *ACM Transactions on Software Engineering and Methodology*, vol. 31, no. 3, pp. 35:1–35:31, Mar. 2022, number: 3. [Online]. Available: <http://doi.org/10.1145/3487571>
- [16] S. Chen, X. Zhao, J. Liu, G. Gao, and Y. Du, "Social-Network-Assisted Task Recommendation Algorithm in Mobile Crowd Sensing," in *Proceedings of the 7th International Conference on Information and Education Innovations*, ser. ICIEI '22. New York, NY, USA: Association for Computing Machinery, Sep. 2022, pp. 136–142. [Online]. Available: <http://doi.org/10.1145/3535735.3535751>
- [17] Y. Zhao, X. Chen, L. Deng, T. Kieu, C. Guo, B. Yang, K. Zheng, and C. S. Jensen, "Outlier Detection for Streaming Task Assignment in Crowdsourcing," in *Proceedings of the ACM Web Conference 2022*, ser. WWW '22. New York, NY, USA: Association for Computing Machinery, Apr. 2022, pp. 1933–1943. [Online]. Available: <http://doi.org/10.1145/3485447.3512067>
- [18] M. Tahaei and K. Vaniea, "Recruiting Participants With Programming Skills: A Comparison of Four Crowdsourcing Platforms and a CS Student Mailing List," in *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, ser. CHI '22. New York, NY, USA: Association for Computing Machinery, Apr. 2022, pp. 1–15. [Online]. Available: <http://doi.org/10.1145/3491102.3501957>
- [19] O. A. Haqbani and S. Alyahya, "Supporting Coordination among Participants in Crowdsourcing Software Design," in *2022 IEEE/ACIS 20th International Conference on Software Engineering Research, Management and Applications (SERA)*, May 2022, pp. 132–139, iSSN: 2770-8209.
- [20] R. Esmailyfard, "Improving detection of web service antipatterns using crowdsourcing," *The Journal of Supercomputing*, vol. 78, no. 5, pp. 6340–6370, Apr. 2022, number: 5. [Online]. Available: <https://doi.org/10.1007/s11227-021-04134-3>
- [21] E. N. Moghaddam, A. Aliahmadi, M. Bagherzadeh, S. Markovic, M. Micevski, and F. Saghafi, "Let me choose what I want: The influence of incentive choice flexibility on the quality of crowdsourcing solutions to innovation problems," *Technovation*, p. 102679, Dec. 2022. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0166497222002309>
- [22] S. Shin, H. Choi, Y. Yi, and J. Ok, "Power of Bonus in Pricing for Crowdsourcing," in *Abstract Proceedings of the 2022 ACM SIGMETRICS/IFIP PERFORMANCE Joint International Conference on Measurement and Modeling of Computer Systems*, ser. SIGMETRICS/PERFORMANCE '22. New York, NY, USA: Association for Computing Machinery, Jun. 2022, pp. 43–44. [Online]. Available: <http://doi.org/10.1145/3489048.3522633>
- [23] X. Miao, H. Peng, Y. Gao, Z. Zhang, and J. Yin, "On Dynamically Pricing Crowdsourcing Tasks," *ACM Transactions on Knowledge Discovery from Data*, Jun. 2022, just Accepted. [Online]. Available: <http://doi.org/10.1145/3544018>
- [24] H. H. Khan, M. N. Malik, and Y. Alotaibi, "Trust Issues in Crowd-sourced Software Engineering: An Empirical Study," *Journal of Information Science & Engineering*, vol. 38, no. 4, 2022, number: 4 ISBN: 1016-2364.
- [25] S. Huang, Z. Yang, C. Zheng, Y. Wang, J. Du, Y. Ding, and J. Wan, "Intellectual Property Right Confirmation System Oriented to Crowdsourced Testing Services," in *2022 International Conference on Blockchain Technology and Information Security (ICBTIS)*, Jul. 2022, pp. 64–68.
- [26] M. Li, L. Yang, Q. Xia, M. Fang, G. Liang, and C. Zuo, "STPChain: a Crowdsourced Software Engineering Method for Software Traceability and Fine-grained Privacy Based on Blockchain," in *2022 IEEE 46th Annual Computers, Software, and Applications Conference (COMPSAC)*, Jun. 2022, pp. 849–859, iSSN: 0730-3157.
- [27] P. Moslehi, J. Rilling, and B. Adams, "A user survey on the adoption of crowd-based software engineering instructional screencasts by the new generation of software developers," *Journal of Systems and Software*, vol. 185, p. 111144, Mar. 2022. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0164121221002405>
- [28] E. C. Groen, "Where Does Crowd-based Requirements Engineering End and Market Research Begin?" in *2022 IEEE 30th International Requirements Engineering Conference Workshops (REW)*, Aug. 2022, pp. 136–138, iSSN: 2770-6834.
- [29] J. A. Khan, A. Yasin, R. Fatima, D. Vasan, A. A. Khan, and A. W. Khan, "Valuating requirements arguments in the online user's forum for requirements decision-making: The CrowdRE-VArg framework," *Software: Practice and Experience*, vol. 52, no. 12, pp. 2537–2573, 2022, number: 12 eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/spe.3137>. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/spe.3137>
- [30] J. Wouters, A. Menkveld, S. Brinkkemper, and F. Dalpiaz, "Crowd-based requirements elicitation via pull feedback: method and case studies," *Requirements Engineering*, vol. 27, no. 4, pp. 429–455, Dec. 2022, number: 4. [Online]. Available: <https://doi.org/10.1007/s00766-022-00384-6>
- [31] M. N. Alatawi, "A conceptual framework for crowdsourcing requirements engineering in SCRUM-based environment," *IET Software*, 2022, publisher: Wiley Online Library.

- [32] L. Radeck, B. Paech, F. Kramer-Gmeiner, M. Wettstein, H.-W. Wahl, A.-L. Schubert, and U. Sperling, "Understanding IT-related Well-being, Aging and Health Needs of Older Adults with Crowd-Requirements Engineering," in *2022 IEEE 30th International Requirements Engineering Conference Workshops (REW)*, Aug. 2022, pp. 57–64, iSSN: 2770-6834.
- [33] S. Gottschalk, S. Parvez, E. Yigitbas, and G. Engels, "Designing Platforms for Crowd-Based Software Prototype Validation: A Design Science Study," in *Product-Focused Software Process Improvement*, ser. Lecture Notes in Computer Science, D. Taibi, M. Kuhrmann, T. Mikkonen, J. Klünder, and P. Abrahamsson, Eds. Cham: Springer International Publishing, 2022, pp. 334–350.
- [34] D. Das, N. S. Mathews, and S. Chimalakonda, "Exploring Security Vulnerabilities in Competitive Programming: An Empirical Study," in *Proceedings of the International Conference on Evaluation and Assessment in Software Engineering 2022*, ser. EASE '22. New York, NY, USA: Association for Computing Machinery, Jun. 2022, pp. 110–119. [Online]. Available: <http://doi.org/10.1145/3530019.3530031>
- [35] W.-T. Tsai, L. Zhang, and S. Hu, "From Crowdsourced Software Development to Crowdtesting," in *5th International Conference on Crowd Science and Engineering*, ser. ICCSE '21. New York, NY, USA: Association for Computing Machinery, Mar. 2022, pp. 18–23. [Online]. Available: <http://doi.org/10.1145/3503181.3503185>
- [36] M. Nasir, N. Ikram, and Z. Jalil, "Usability inspection: Novice crowd inspectors versus expert," *Journal of Systems and Software*, vol. 183, p. 111122, Jan. 2022. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0164121221002193>
- [37] Q. Z. Chen, T. Schnabel, B. Nushi, and S. Amershi, "HINT: Integration Testing for AI-based features with Humans in the Loop," in *27th International Conference on Intelligent User Interfaces*, ser. IUI '22. New York, NY, USA: Association for Computing Machinery, Mar. 2022, pp. 549–565. [Online]. Available: <http://doi.org/10.1145/3490099.3511141>
- [38] Y. Lin, Z. Li, W. Yue, and J. Wen, "CrowdIoT: The Crowd-Sourcing Test System for IoT Devices Based on Blockchain," *Advances in Internet of Things*, vol. 12, no. 2, pp. 19–34, 2022, number: 2 Publisher: Scientific Research Publishing.
- [39] LI Le-Ping, ZHANG Yu-Xia, and LIU Hui, "Crowdsourcing Software Development Oriented Fault Localization," *Journal of Software*, pp. 1–18, Nov. 2022. [Online]. Available: <https://www.jos.org.cn/josen/article/abstract/6498>
- [40] M. Du, S. Yu, C. Fang, T. Li, H. Zhang, and Z. Chen, "SemCluster: a semi-supervised clustering tool for crowdsourced test reports with deep image understanding," in *Proceedings of the 30th ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering*, ser. ESEC/FSE 2022. New York, NY, USA: Association for Computing Machinery, Nov. 2022, pp. 1756–1759. [Online]. Available: <http://doi.org/10.1145/3540250.3558933>
- [41] Y. Li, Y. Feng, R. Hao, D. Liu, C. Fang, Z. Chen, and B. Xu, "Classifying crowdsourced mobile test reports with image features: An empirical study," *Journal of Systems and Software*, vol. 184, p. 111121, Feb. 2022. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0164121221002181>
- [42] D. Liu, Y. Feng, X. Zhang, J. A. Jones, and Z. Chen, "Clustering Crowdsourced Test Reports of Mobile Applications Using Image Understanding," *IEEE Transactions on Software Engineering*, vol. 48, no. 4, pp. 1290–1308, Apr. 2022, number: 4 Conference Name: IEEE Transactions on Software Engineering.
- [43] P. Zhu, Y. Li, T. Li, H. Ren, and X. Sun, "Advanced Crowdsourced Test Report Prioritization Based on Adaptive Strategy," *IEEE Access*, vol. 10, pp. 53 522–53 532, 2022, conference Name: IEEE Access.
- [44] Y. Yang and X. Chen, "Crowdsourced Test Report Prioritization Based on Text Classification," *IEEE Access*, vol. 10, pp. 92 692–92 705, 2022, conference Name: IEEE Access.
- [45] H. Zhang, Y. Zhao, S. Yu, and Z. Chen, "Automated Quality Assessment for Crowdsourced Test Reports Based on Dependency Parsing," in *2022 9th International Conference on Dependable Systems and Their Applications (DSA)*, Aug. 2022, pp. 34–41, iSSN: 2767-6684.
- [46] S. Alyahya, "Collaborative Crowdsourced Software Testing," *Electronics*, vol. 11, no. 20, p. 3340, Jan. 2022, number: 20 Publisher: Multidisciplinary Digital Publishing Institute. [Online]. Available: <https://www.mdpi.com/2079-9292/11/20/3340>
- [47] X. Ge, S. Yu, C. Fang, Q. Zhu, and Z. Zhao, "Leveraging Android Automated Testing to Assist Crowdsourced Testing," *IEEE Transactions on Software Engineering*, pp. 1–18, 2022, conference Name: IEEE Transactions on Software Engineering.
- [48] R. Nandakumar, "Quantitative Quality Score for Software," in *15th Innovations in Software Engineering Conference*, ser. ISEC 2022. New York, NY, USA: Association for Computing Machinery, Feb. 2022, pp. 1–5. [Online]. Available: <https://doi.org/10.1145/3511430.3511457>
- [49] K. Wu, S. Huang, Y. Shi, J. Zhu, and S. Tang, "Estimate the Precision of Defects Based on Reports Duplication in Crowdsourced Testing," *IEEE Access*, vol. 10, pp. 130 415–130 423, 2022, conference Name: IEEE Access.
- [50] J. P. d. Reis, F. B. e. Abreu, and G. d. F. Carneiro, "Crowdsmelling: A preliminary study on using collective knowledge in code smells detection," *Empirical Software Engineering*, vol. 27, no. 3, p. 69, Mar. 2022, number: 3. [Online]. Available: <https://doi.org/10.1007/s10664-021-10110-5>

Evaluation of Wood Species Identification Using CNN-Based Networks at Different Magnification Levels

Khanh Nguyen-Trong*

Posts and Telecommunications Institute of Technology, Hanoi, Vietnam*

Abstract—Wood species identification (WoodID) is a crucial task in many industries, including forestry, construction, and furniture manufacturing. However, this process currently requires highly trained individuals and is time-consuming. With the recent advances in machine learning and computer vision techniques, automatic WoodID using macro-images of cross-section wood has gained attention. Nevertheless, existing works have been evaluated on ad-hoc datasets with pre-fixed magnification levels. To address this issue, this paper proposes an evaluation of deep learning-based methods for WoodID on multiple datasets with varying magnification levels. Several popular Convolutional Neural Networks, including DenseNet, ResNet50, and MobileNet, were examined to identify the best network and magnification levels. The experiments were conducted on five datasets with different magnifications, including a self-collected dataset and four existing ones. The results demonstrate that the DenseNet121 network achieved superior accuracy and F1-Score on the 20X dataset. The findings of this study provide useful insights into the development of automatic WoodID systems for practical applications.

Keywords—Wood species identification; convolutional neural network; ResNet50; DensNet

I. INTRODUCTION

Wood species identification (WoodID) is a critical aspect in various industries, including forestry, construction, and furniture manufacturing [1], [2]. The identification process involves determining the type of wood based on its unique physical and anatomical characteristics such as growth rings, knots, ray patterns, and texture [3]. In many countries, this process is currently performed manually by forestry experts. The manual process not only poses a challenge in terms of cost and time, but also limits the ability to scale up the identification process to meet the demands of the growing timber industry. This highlights the need for a more automated and accessible approach to WoodID, which could have a significant impact on the efficiency and effectiveness of the forestry industry.

Recently, with the development of machine learning and computer vision techniques, automatic WoodID has received increasing attention. Numerous deep-learning-based methods have been proposed in the literature, which rely on the unique visual characteristics of wood cross-sections [4]. Owing to the widespread availability of low-cost portable digital cameras and stereo microscopes, most of these works used macroscopic images of transverse sections of wood. For example, these studies have been applied to automatic identification of North American hardwood species [5], European tree species [6], Chinese wood species [7], Japanese wood species [8], or

Indonesian commercial wood [9]. They have demonstrated their high accuracy, robustness, and efficiency compared to traditional methods that involve manual feature extraction.

Among the deep learning methods applied for WoodID, Convolutional Neural Network (CNN) based approaches have gained widespread use. Examples of these networks include VGG16 [10], ResNet50 [11], DenseNet [12], and MobileNet [13]. The advantage of using CNNs is that they can automatically identify relevant features in the input images, eliminating the need for manual feature extraction. The availability of large annotated datasets and advancements in hardware and software have further driven the application of CNNs in this field.

These networks have proven their ability to effectively handle the complex macroscopic characteristics of wood and accurately recognize species, as demonstrated, for example, in studies focusing on Pacific and Colombian Amazon wood species [14] and Brazilian flora species [15]. This makes them suitable for practical application in countries where illegal logging is a prevalent issue, such as Vietnam.

Although of the high performance in WoodID, there is no concrete recommendation as to which magnification level is most suitable. These studies applied various magnification levels with different wood species datasets. Besides, the application of these advanced techniques is still limited in many underdevelopment countries, such as Vietnam, where access to resources and data might be scarce. Therefore, there is a need for suitable solutions that consider the best network architecture and magnification level of macroscopic images of wood cross-sections. This would not only aid in better forestry management, but also support efforts to curb illegal logging.

In this study, we evaluated the performance of different convolutional neural networks (CNNs) for WoodID using macro-images of wood cross-sections at various magnification levels. Our objective was to propose a practical and effective CNN-based method for this task. To the best of our knowledge, this is the first attempt to evaluate WoodID on different magnification levels, which has led to the development of a more accurate and practical method not only in Vietnam but also in other countries.

We examined the proposed method's performance on five datasets with different magnifications, including a self-collected dataset of popular imported wood species in Vietnam and four existing ones. The study's results provided insights into the effectiveness of different networks and magnification levels for WoodID.

The remainder of this paper is structured as follows. Section II discusses relevant previous studies. Section III provides the details regarding the evaluated networks. The experimental evaluation is presented in Section IV, and finally, some concluding remarks and a brief discussion are provided in Section V.

II. RELATED WORKS

Convolutional Neural Networks (CNNs) have been widely used in previous studies for WoodID using images of transverse cross-sections of wood. Popular CNN architectures, such as VGG16, ResNet, DenseNet, and MobileNet, have been proposed and trained on various datasets with different magnification levels. In this study, we focus on the methods that employ two main image levels, namely microscopic and macroscopic transverse cross-sections of wood, to investigate the effectiveness of CNNs for WoodID.

For the first level, previous studies have typically utilized dedicated microscopic devices to capture wood cross-section images [16]. For example, Silva *et al.* [17] presented a method to automatically classify wood species using microscopic images of 77 commercial Central-African timber captured at 25x magnification with an Olympus BX60 microscope and an Olympus UC30 digital camera. Geus *et al.* [18] introduced a wood image dataset consisting of 281 species, with three sets of 20 samples each, corresponding to transverse, radial, and tangential sections, taken on an Olympus SZX7 stereo microscope at 20x magnification. V. Stagno *et al.* [19] conducted a study using NMR to capture samples, including one softwood and four hardwood species. The samples were boiled in distilled water until saturation, and microscopy images were captured using a Zeiss EVO LS10 Environmental Scanning Electron Microscope (ESEM) with EDS. Low vacuum mode was chosen to obtain cross-sectional images of the wood, with a working distance ranging from 4.5 to 5.0 mm and an electron high tension of 20.00 kV. Magnifications of 300x and 1000x were used. The data were acquired using a Bruker Avance-400 spectrometer operating at 9.4 T with a 10 mm micro-imaging probe and XWINNMR and ParaVision 3.2 software.

For macroscopic level studies, digital magnifying glasses connected to smartphones or computers are often used to capture images of wood cross-sections [20]–[22]. These images are taken at different levels of magnification, typically ranging from 10-50X. For example, Saenz *et al.* [23] captured images of 11 forest species using a smartphone with a CMOS sensor and a magnification of 3.9 microns per pixel (about 20x of magnification). The area of interest was 2.5mm x 1.9mm, and lighting was provided by the device.

Lee *et al.* [24] proposed a dataset of 25 species from Yunnan Province, China. The woodblocks were first collected and cut by an electric moto saw into 1 cm³ pieces and flattened with sandpaper (400 grit, 800 grit, 1000 grit). The images were taken using a mobile phone (OnePlus 3, China) and a 20X magnifying glass, with the original image size being 2048 × 1024 pixels. The central 300 × 300 pixels were selected as the experimental material, as they were clearer and less fuzzy than the other areas. A total of 3000 images were obtained, with which each species including 120 images.

In another study, Filho *et al.* [25] used a Sony DSC T20 camera to capture images of 46 species with a resolution of 150 dpi using artificial lighting. Meanwhile, Sun *et al.* [7] used a mobile phone and a 20X magnifying glass to capture images of 25 wood species with a resolution of 2048x1024 pixels. The images were polished and the middle 300x300 pixels were cropped for clearer results. The dataset was divided into two parts for training and testing, with 2498 images for training and 502 images for testing.

In a study conducted by Souza *et al.* [15] in 2020, 46 Brazilian wood species were collected from the Wood Anatomy and Quality Laboratory (LANAQM) at Federal University of Paraná (UFPR) in Curitiba, Paraná. The transversal surfaces of the samples were sanded with a 120 sandpaper, and macroscopic images were taken using a Zeiss Discovery V 12 stereo microscopes with a resolution of 2080 × 1540 pixels and a 10× magnification. This resulted in a total of 1,901 images.

Additionally, in 2021, de Geus *et al.* [14] introduced a new dataset of 11 Brazilian wood species with high commercial value, with all images being taken from the transverse section using a low-cost portable microscope connected to a smartphone with a resolution of 640 × 480 pixels. The dataset consists of 440 images, with 40 images for each of the 11 species.

Various methods have been proposed in the literature, including traditional machine learning (i.e., k-NN, SVM, ANN) and deep learning (i.e., VGG16, ResNet50, SqueezeNet, DenseNet). Traditional methods require fewer data but rely on handcrafted features, while deep learning models demand more data, but perform automatic feature extraction, resulting in higher accuracy. Among these methods, CNN-based networks have been the most successful for WoodID, with several studies achieving high accuracy rates using different networks such as VGG16, ResNet50, SqueezeNet, and DenseNet. For example, de Geus *et al.* [14] obtained an accuracy of 98.13% with DenseNet, while Lee *et al.* [24] achieved an accuracy of 99.6% with ResNet50. However, each proposed method has specific parameters, such as network architecture and magnification levels, and no concrete recommendation has been made regarding the optimal magnification level for CNN-based WoodID. Therefore, this study aims to evaluate the performance of common CNN-based networks for WoodID using macroscopic images of wood cross-sections at different magnification levels, with a focus on evaluating their practicality for use in Vietnam.

III. MATERIALS AND METHODS

This section provides a comprehensive overview of the data collection process and the Convolutional Neural Network (CNN) based methods applied in the recognition of wood species. To provide a comparative and objective measure of different wood species, we collected five datasets from different regions, including Vietnam, Pacific and Colombian Amazon, and China. Three popular CNN architectures were examined and evaluated on these datasets, including ResNet50, MobileNetV2, and DenseNet121.

A. Data Preparation and Collection

The study utilized five datasets, including (i) VN_26, a self-collected dataset, (ii) WRD_21, a Southeast Asia wood

TABLE I. THE VN_26 DATASET (10X, 20X AND 50X)

ID	Family	Scientific Name	ID	Family	Scientific Name
1	Fabaceae	Afzelia africana Smith	14	Fabaceae	Guibourtia coleosperma (Benth.) Leonard
2	Fabaceae	Afzelia pachyloba Harms	15	Fabaceae	Guibourtia demeusei J.Leon.
3	Sapotaceae	Autranella congolensis A.Chev.	16	Fabaceae	Julbernardia pellegriniana Troupin
4	Fabaceae	Berlinia bracteosa Benth.	17	Moraceae	Milicia excelsa (Welw.) C.C.Berg
5	Fabaceae	Brachystegia laurentii (De Wild.) Hoyle	18	Fabaceae	Millettia laurentii De Wild.
6	Fabaceae	Cylicodiscus gabunensis Harms	19	Fabaceae	Monopetalanthus coriaceus Mor.
7	Fabaceae	Dalbergia melanoxylon Guill. et Perr.	20	Rubiaceae	Nauclea diderrichii Merr.
8	Fabaceae	Daniellia thurifera Benn.	21	Fabaceae	Pachyelasma tessmannii (Harms) Harms.
9	Fabaceae	Detarium macrocarpum Harms	22	Fabaceae	Piptadeniastrum africanum Brenan
10	Fabaceae	Distemonanthus benthamianus Baill.	23	Fabaceae	Pterocarpus angolensis DC.
11	Meliaceae	Entandrophragma cylindricum Sprague	24	Fabaceae	Pterocarpus Jacq.
12	Fabaceae	Erythrophleum suaveolens Brenan	25	Fabaceae	Pterocarpus soyauxii Taub.
13	Meliaceae	Guarea cedrata Pellegr.	26	Sapotaceae	Tieghemella africana Pierre

TABLE II. THE WRD_21 DATASET (20X)

ID	Family	Scientific Name	Number of images	ID	Family	Scientific Name	Number of images
1	Fabaceae	Cassia siamea	28	12	Calophyllaceae	Mesua ferrea	36
2	Lauraceae	Cinnamomum camphora	45	13	Fabaceae	Pterocarpus erinaceus Poir	32
3	Fabaceae	Dalbergia bariensis	96	14	Fabaceae	Pterocarpus indicus	24
4	Fabaceae	Dalbergia cochinchinensis	50	15	Fabaceae	Pterocarpus macrocarpus	91
5	Fabaceae	Dalbergia fusca	59	16	Fabaceae	Pterocarpus santalinus	39
6	Fabaceae	Dalbergia latifolia	36	17	Cupressaceae	Taiwania flousiana	42
7	Fabaceae	Dalbergia odorifera	25	18	Lamiaceae	Tectona grandis	74
8	Fabaceae	Dalbergia oliveri	145	19	Combretaceae	Terminalia myriocarpa	63
9	Ebenaceae	Diospyros ebum	43	20	Combretaceae	Terminalia tomentosa	52
10	Malvaceae	Excentrodendron hsienmu	42	21	Fabaceae	Xylia dolabriformis	54
11	Fabaceae	Intsia spp	50				

dataset published by Sun [26], (iii) BFS_46, the Brazilian flora species dataset collected in 2020 [15], (iv) BD_11, the Brazil dataset collected in 2021 [14], and (v) PCA_11, a dataset containing wood from Pacific and Colombian Amazon region [23]. The datasets include macro-images of wood cross-sections captured by different devices at various magnifications and resolutions. Wood samples were sometimes sanded and polished to improve image quality.

The VN_26 dataset contains 26 wood species imported to Vietnam, as presented in Table I, with samples collected from different locations on the wood cross-section. We captured the images at three magnification levels (10x, 20x, and 50x) with three different types of microscopes. Each image set has a different resolution and image size. The surfaces of the samples were treated by sanding and polishing before being captured.

The WRD_21 dataset includes 1,126 macro-images of 21 Southeast Asia wood species [26], as presented in Table II. To create the dataset, the authors polished the cross-section of each wood block using 200 grits and 400 grits sandpaper in sequence and then cleared any remaining dust with a toothbrush. Next, they marked a circular area on the wood block to serve as the fingerprint area, which ensured that the same location for subsequent captures could be chosen. They then used a 20X magnifying glass to acquire an image of the marked area. Finally, the macro-images were captured by a Huawei Honor 8 cellphone.

The BFS_46 dataset, as presented in Table III, was col-

lected in 2020 and contains 1,901 images of 46 Brazilian flora species with two different sizes (640x480 pixels and 2080x1540 pixels). The wood samples were obtained from the Wood Anatomy and Quality Laboratory (LANAQM) at the Federal University of Paraná (UFPR) in Curitiba, Paraná. They were then sanded with 120 grit sandpaper. The authors used a Zeiss Discovery V 12 stereo-microscope at 10x magnification to capture the macro-images with a resolution of 150 dpi.

The BD_11 is another Brazilian dataset collected in 2021, consisting of 11 high-commercial-value wood species, as shown in Table IV. The timber samples were not polished; instead, the authors used a pocket knife to cut them to expose the anatomical characteristics. The samples were captured using a low-cost portable microscope connected to a smartphone with a resolution of 640x480 pixels. To account for variance in anatomical characteristics, each species comprises 40 images, with four image samples extracted from 10 different specimens.

The PCA_11 dataset contains 10,792 images of 11 wood species captured using a digital microscope device at 20x magnification, as presented in Table V. First, timber samples from the Pacific and Colombian Amazon region were aggregated and moistened to increase contrast. Next, they were captured with a digital microscope device at a resolution of 640x480 pixels.

In general, these datasets were collected by capturing images of wood samples using either microscopes or digital

TABLE III. THE BFS_46 DATASET (10X)

ID	Family	Scientific Name	Number of images	ID	Family	Scientific Name	Number of images
1	Fabaceae	Acrocarpus fraxinifolius Arn.	17	24	Fabaceae	Hymenaea sp. L.	32
2	Araucariaceae	Araucaria angustifolia (Bertol.) Kuntze	55	25	Fabaceae	Hymenolobium petraeum Ducke	28
3	Apocynaceae	Aspidosperma polyneuron Mull. Arg.	20	26	Fabaceae	Hymenolobium sp. Benth.	28
4	Apocynaceae	Aspidosperma Mart. & Zucc.	41	27	Fabaceae	Inga vera Willd.	40
5	Moraceae	Bagassa guianensis Aubl.	52	28	Lauraceae	Laurus nobilis L.	36
6	Rutaceae	Balfourodendron riedelianum (Engl.) Engl.	61	29	Fabaceae	Machaerium paraguayense Hassl.	37
7	Lecythidaceae	Bertholletia excelsa Bonpl.	35	30	Fabaceae	Machaerium sp. Pers.	15
8	Fabaceae	Bowdichia sp. Kunth	68	31	Sapotaceae	Manilkara elata (Allemão ex Miq.) Monach.	39
9	Moraceae	Brosimum parinarioides Ducke	25	32	Meliaceae	Melia azedarach L.	47
10	Meliaceae	Carapa guianensis Aubl.	21	33	Lauraceae	Mezilaurus itauba (Meisn.) Taub. ex Mez	83
11	Lecythidaceae	Cariniana estrellensis (Raddi) Kuntze	36	34	Sapotaceae	Micropholis venulosa (Mart. & Eichler) Pierre	71
12	Meliaceae	Cedrela fissilis Vell.	22	35	Fabaceae	Mimosa scabrella Benth.	30
13	Fabaceae	Cedrelinga cateniformis (Ducke) Ducke	65	36	Fabaceae	Muelleria campestris (Mart. ex Benth.) M.J. Silva & A.M.G. Azevedo	39
14	Boraginaceae	Cordia goeldiana Huber	36	37	Fabaceae	Myroxylon balsamum (L.) Harms	53
15	Lecythidaceae	Couratari sp. Aubl.	41	38	Lauraceae	Nectandra megapotamica (Spreng.) Mez	28
16	Fabaceae	Dipteryx sp. Schreb.	27	39	Lauraceae	Ocotea indecora (Schott) Mez	36
17	Vochysiaceae	Erisma uncinatum Warm.	58	40	Lauraceae	Ocotea porosa (Nees & Mart.) Barroso	46
18	Myrtaceae	Eucalyptus sp. L'Hér.	27	41	Fabaceae	Peltogyne sp. Vogel	60
19	Myrtaceae	Eugenia pyriformis Cambess.	35	42	Pinaceae	Pinus sp. L.	42
20	Rutaceae	Euxylophora paraensis Huber	66	43	Sapotaceae	Pouteria pachycarpa Pires	47
21	Goupiaceae	Goupia glabra Aubl.	32	44	Simaroubaceae	Simarouba amara Aubl.	30
22	Proteaceae	Grevillea robusta A. Cunn. ex R. Br.	48	45	Meliaceae	Swietenia macrophylla King	70
23	Bignoniaceae	Handroanthus sp. Mattos	33	46	Vochysiaceae	Vochysia sp. Aubl	43

TABLE IV. THE BD_11 DATASET (50X)

ID	Family	Scientific Name	Number of images
1	Lecythidaceae	Allantoma decandra	40
2	Calophyllaceae	Caraipa densifolia	40
3	Lecythidaceae	Cariniana micrantha	40
4	Caryocaraceae	Caryocar villosum	40
5	Moraceae	Clarisia racemosa	40
6	Fabaceae	Dipteryx odorata	40
7	Goupiaceae	Goupia glabra	40
8	Bignoniaceae	Handroanthus incanus	40
9	Malvaceae	Lueheopsis duckeana	40
10	Myristicaceae	Osteophloeum platyspermum	40
11	Sapotaceae	Pouteria caimito	40

TABLE V. THE PCA_11 DATASET (20X)

ID	Family	Scientific Name	Number of images
1	Anacardiaceae	Camposperma panamensis	823
2	Meliaceae	Cedrela odorata	1128
3	Fabaceae	Cedrelinga cateniformis	1189
4	Boraginaceae	Cordia alliodora	929
5	Myristicaceae	Dialyanthera gracilipes	1100
6	Myrtaceae	Eucalyptus globulus	1105
7	Bignoniaceae	Handroanthus chrysanthus	1106
8	Humiriaceae	Humiriastrum procerum	1001
9	Oleaceae	Fraxinus uhdei	1025
10	Cupressaceae	Cupressus lusitanica	815
11	Pinaceae	Pinus patula	571

cameras. In some cases, the samples were treated by sanding and polishing to improve the image quality. The resolution and size of the images varied among the datasets.

B. Convolutional Neural Network Architecture

In this study, we focus on deep learning techniques to recognize wood species based on macro-images of wood cross-sections. To achieve this, we utilized several well-known Convolutional Neural Network (CNN) architectures, namely

ResNet [11], DenseNet [12], and Mobilenet [13], which have been widely applied in the field. Each of these models has its strengths and weaknesses. We compared their performance to determine the most effective architecture for WoodID on the studied datasets. In this section, we provide an overview of each network and its specific characteristics.

TABLE VI. RESNET50 ARCHITECTURE

Layer	Type	Output shape	No parameters
1	Input layer	(224,224,3)	-
2	Convolutional	(112,112,64)	1792
3	Max pooling	(56,56,64)	-
4. Res block	Convolutional	(56,56,256)	89600
	Batch normalization	(56,56,256)	1024
	Identity mapping	(56,56,256)	-
5. Res block	Convolutional	(56,56,512)	354944
	Batch normalization	(56,56,512)	2048
	Identity mapping	(56,56,512)	-
6. Res block	Convolutional	(28,28,1024)	1407584
	Batch normalization	(28,28,1024)	4096
	Identity mapping	(28,28,1024)	-
7. Res block	Convolutional	(14,14,2048)	5621248
	Batch normalization	(14,14,2048)	8192
	Identity mapping	(14,14,2048)	-
Average pool	Average pooling	(2048)	-
FC	Fully connected	(1000)	2097000

1) *ResNet50*: ResNet50 is a deep convolutional neural network architecture that was introduced by Microsoft researchers in 2015 [11]. It is part of a family of ResNet models that were designed to address the problem of vanishing gradients in deep neural networks. The authors introduce shortcut connections that allow the gradient signal to bypass some layers during training.

ResNet50 has been widely used in various computer vision tasks, including object detection, image segmentation, and image classification. It has achieved state-of-the-art performance on several benchmark datasets, including ImageNet, which contains over one million images from a thousand different classes. ResNet50 has shown remarkable accuracy and efficiency in recognizing different types of objects, including

wood species, making it a popular choice for many practical applications.

The network consists of a series of convolutional and pooling layers, followed by multiple residual blocks, and finally a global average pooling layer and a fully connected layer, as shown in Table VI.

2) *Densenet*: DenseNet121 is a deep convolutional neural network that was introduced in 2017. It is part of the DenseNet family of models, which is based on the idea of densely connecting each layer to every other layer in a feed-forward fashion [12]. This architecture allows for a deeper and more efficient network that can achieve higher accuracy with fewer parameters compared to traditional architectures like VGG or ResNet.

DenseNet121 consists of 121 layers, with a total of 8.06 million parameters. It has a similar structure to other DenseNet models, where each layer is densely connected to every other layer in a feed-forward fashion. The difference with DenseNet121 is that it uses smaller filters and less dense connections between layers, which allows for better memory usage and faster training. The model also includes skip connections, which help to mitigate the vanishing gradient problem and improve the flow of information throughout the network. Overall, DenseNet121 is a powerful deep learning model that has shown excellent performance in a variety of computer vision tasks, including WoodID.

3) *MobileNet*: MobileNet is a deep learning architecture designed for mobile and embedded devices with limited computational resources. It was introduced by Google in 2017 and has gained significant popularity due to its efficient and lightweight design [13]. The network uses depthwise separable convolutions that separate the spatial and channel-wise convolutions in a standard convolutional layer, as presented in Table VII. This design reduces the number of computations required while maintaining a high level of accuracy.

MobileNet has several variants, including MobileNet v1, v2, and v3. MobileNetV1 was the first version of the architecture and was introduced in 2017. It uses depthwise separable convolutions with a width multiplier, which reduces the number of channels in the network, making it more lightweight. MobileNetV2, introduced in 2018, builds upon the original architecture and introduces several improvements, including linear bottleneck layers and inverted residuals. These changes improve the accuracy and reduce the number of computations required. MobileNetV3, introduced in 2019, focuses on improving the speed and accuracy of the architecture further. It uses a combination of channel and spatial attention modules to improve the network's performance while maintaining its lightweight design. In this study, we evaluated the performance of MobileNetV2.

IV. EXPERIMENTS

A. Dataset

In this study, we utilized a total of five datasets, of which the VN_26 dataset was self-collected while the remaining four were obtained from existing sources. For the VN_26 dataset, wood samples measuring 1 inch (2.54 cm) in size were collected from different locations on the cross-section

TABLE VII. MOBILENET ARCHITECTURE

Layer	Output Shape	Kernel size	Stride
Input	224 × 224 × 3	-	-
Convolution	112 × 112 × 32	3 × 3	2
Depthwise Convolution	112 × 112 × 32	3 × 3	1
Pointwise Convolution	112 × 112 × 64	1 × 1	1
Depthwise Convolution	56 × 56 × 64	3 × 3	2
Pointwise Convolution	56 × 56 × 128	1 × 1	1
Depthwise Convolution	28 × 28 × 128	3 × 3	2
Pointwise Convolution	28 × 28 × 256	1 × 1	1
Depthwise Convolution	14 × 14 × 256	3 × 3	2
Pointwise Convolution	14 × 14 × 512	1 × 1	1
Depthwise Convolution	7 × 7 × 512	3 × 3	2
Pointwise Convolution	7 × 7 × 1024	1 × 1	1
Global Average Pooling	1 × 1 × 1024	-	-
Fully Connected	1000	-	-

of the wood. The collected samples were then processed by surface sanding using sandpaper and polished with 600-grit sandpaper to ensure that the surface texture of each sample was uniform. A small blade was used then to cut this surface. Fig. 1 presents the steps to prepare and collect this dataset. This standardized process of sample preparation allowed us to maintain consistency across the dataset and ensured that the resulting images were of high quality, making them ideal for use.

The dataset was captured using three different types of microscopes at three magnification levels (10x, 20x, and 50x) to obtain a comprehensive set of images. For the 10x dataset, a PCE microscope equipped with a 0.3-megapixel CMOS sensor was used to capture images with a magnification of 10x and a resolution of 640x480 pixels. The 20x dataset was captured using a handheld Dino-Lite electronic microscope, which had a resolution of 1.3 megapixels, a magnification of 20x, and resulted in images of size 1280x1024 pixels. To achieve focus, a distance of 5 cm was maintained between the lens and the sample. Lastly, the 50x dataset was captured using a Wi-Fi Microscope with a magnification ratio of 50x-1000x and resulting in images of size 640x480 pixels.

By using different types of microscopes, we were able to capture images with varying levels of detail, which helped in creating a diverse dataset. The VN_26 dataset thus obtained serves as a valuable resource for studying wood characteristics. The comprehensive set of images obtained at different magnifications provides ample opportunities to evaluate WoodID at different levels of microscopy.

Table VIII presents an overview of five used datasets. The datasets differ in the number of classes, the total number of images, and the magnification levels, as well as in image resolutions available. For example, the VN_26 dataset has the highest number of classes and images, with images available at 640x480 and 1280x1024 pixel resolutions. In contrast, the WRD_21 dataset has fewer classes and images, with images available at a lower resolution of 300x300 pixels. The BFS_46 dataset has a high number of classes but a relatively low number of images, with images available at two different resolutions of 640 x 480 and 2080 x 1540 pixels. Finally, the BD_11 and PCA_11 datasets have similar characteristics, with a lower number of classes but a high number of images available at a resolution of 640 x 480 pixels. To ensure the robustness and generalizability of the models, each dataset was

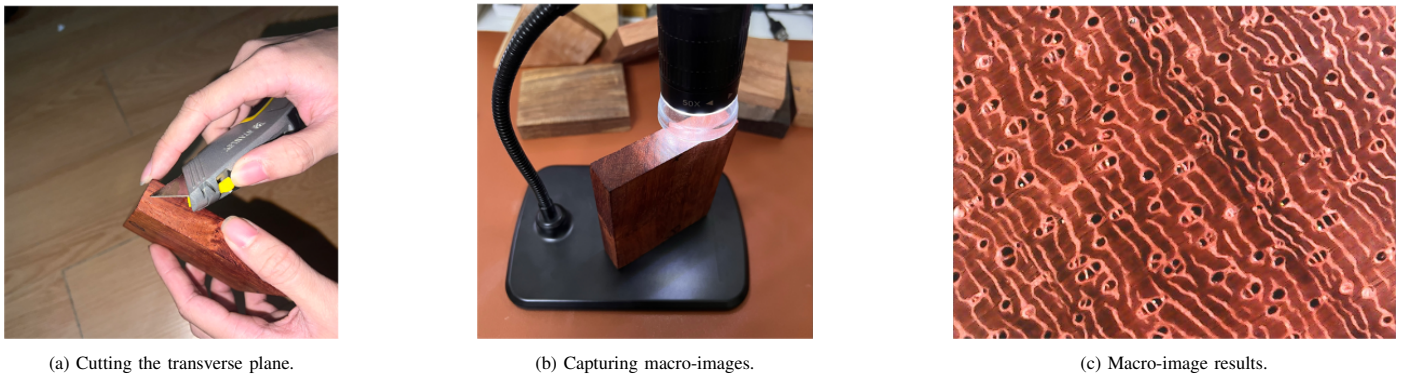


Fig. 1. The VN_26 dataset preparation and collection.

split into three subsets with a ratio of 70:15:15 for training, validation, and testing.

TABLE VIII. FIVE DATASETS USED IN THIS STUDY

Dataset	Number of classes	Total images	10X	20X	50X	Resolution (pixel)
BFS_46	46	1,901	X			640x480 & 2080x1540
WRD_21	21	1,126		X		300x300
PCA_11	11	10,792		X		640x480
BD_11	11	440			X	640x480
VN_26	26	7,800	X	X	X	640x480 & 1280x1024

B. Preprocessing

Before training and evaluating the models, we performed several pre-processing steps on the collected dataset. Firstly, the wood identification experts filtered out the poor-quality images, such as blurry and skewed images. The dataset was then re-collected to ensure an adequate number of images for each species (VN_26). Next, all images were normalized to ensure that the pixel values were within the range of [0, 255]. We then applied techniques to resize them to avoid losing important information.

To enrich the training dataset, we applied various data augmentation techniques such as randomly rotating images within a range of 10 degrees, flipping images randomly horizontally and vertically. These techniques increased the diversity of the training dataset and helped to prevent overfitting of the models.

Finally, we saved the images and their corresponding labels in HDF5 format. HDF5, short for Hierarchical Data Format version 5, is a data model, library, and file format for storing and managing large and complex data. HDF5 provides a flexible and efficient way to store and retrieve large numerical arrays, metadata, and other types of data that are common in scientific computing and data analysis.

Overall, these pre-processing steps played a crucial role in improving the performance of the models and ensuring accurate WoodID. The resulting dataset was of high quality and was well-suited for training deep learning models.

TABLE IX. EXPERIMENTAL SCENARIOS

	Purpose	Network	Dataset
1	Data augmentation evaluation	DenseNet121	PCA_11
2	CNN evaluation on existing datasets	DenseNet121, ResNet50, MobileNetV2	WRD_21, BD_11, BFS_46, PCA_11
3	CNN evaluation on dataset mixed magnifications	DenseNet121, ResNet50, MobileNetV2	VN_26
4	CNN evaluation on dataset with different magnifications	DenseNet121, ResNet50, MobileNetV2	10X, 20X, 50X subsets

C. Experiment Scenarios

To evaluate the performance of studied deep learning models on various datasets and configurations, we conducted four experiment scenarios, as presented in Table IX. In the first scenario, we evaluated the effect of data augmentation on the PCA_11 dataset. We trained the DenseNet121 model with an input image size of 224x224 pixels, and a batch size of 32 for 500 epochs using the Adam optimizer with a learning rate of 0.001 and the category cross-entropy loss function. We stopped training if there was no improvement in the validation loss after 200 epochs. Regarding the second scenario, we trained DenseNet121, ResNet50, and MobileNetV2 on four datasets: WRD_21, BFS_46, BD_11, and PCA_11. The input image size was 224x224 pixels, and we performed training with a batch size of 64 using the Adam optimizer with a learning rate of 0.001. The category cross-entropy loss function was used, and training was stopped if there was no improvement in the model after 200 epochs.

For the third scenario, we trained the same three models on the self-collected dataset, VN_26, which includes three levels of image magnification. The input image size and other parameters were identical to those used in the first scenario. In the fourth scenario, these four models were separately trained on three subsets of VN_26, corresponding to the three magnification levels. The same parameters as the previous scenarios were applied, but we stopped training if there was no improvement after 300 epochs.

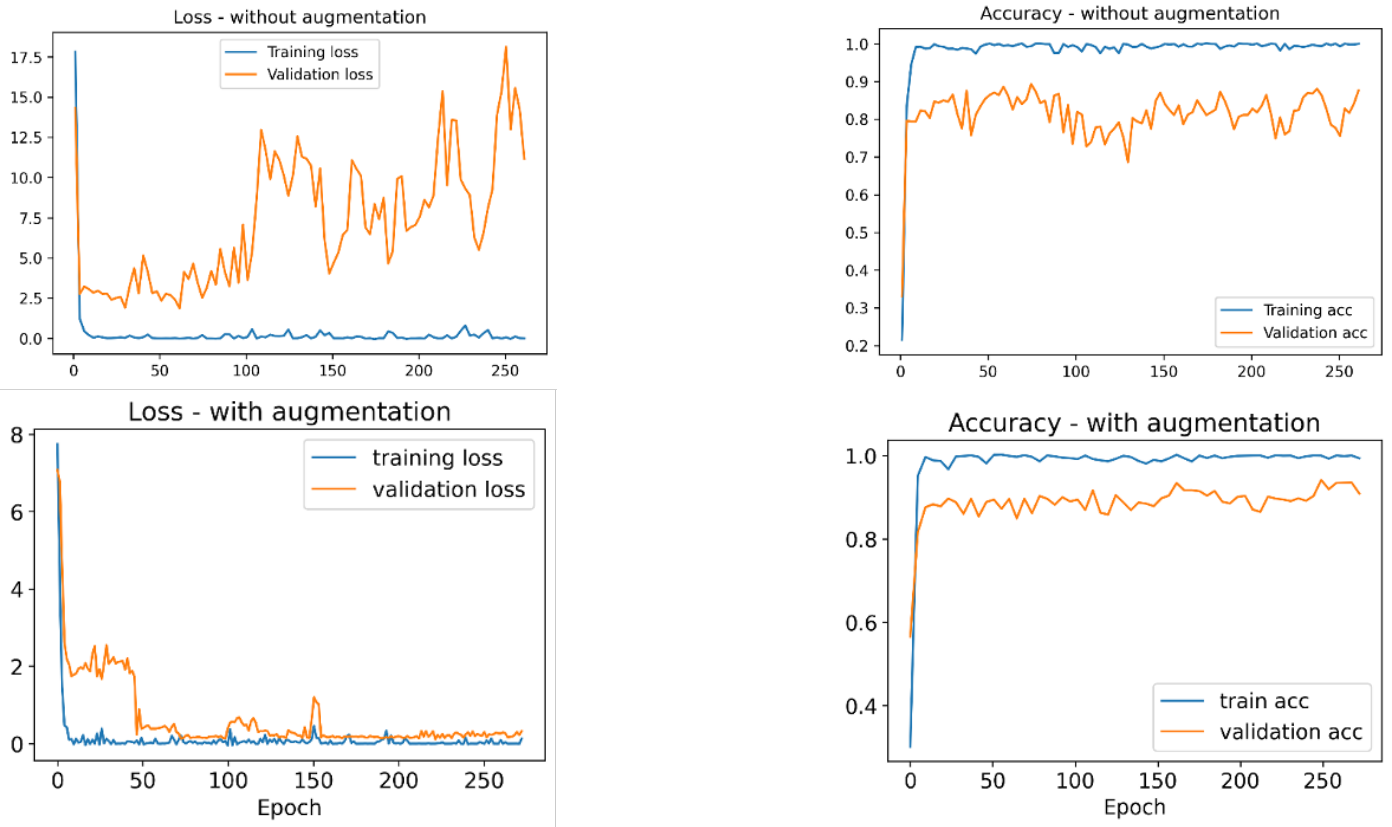


Fig. 2. Experiment 1 – Training loss and accuracy progresses with and without data augmentation.

D. Results and Discussion

The obtained results from the first experiment are depicted in Fig. 2, showing the progress of loss and accuracy during the training process on both the non-augmented (the upper plots) and augmented (the lower plots) datasets. It was observed that without data augmentation, the gap between the training and validation accuracy remained large and constant, while the gap between the two loss progresses tended to increase, indicating an overfitting trend. However, with data augmentation, the overfitting was mitigated, as the gap between the two loss progresses was small, and the accuracy gap was reduced. Consequently, the model achieved a higher accuracy of over 95.9% on the validation dataset.

TABLE X. EXPERIMENT 2 - PERFORMANCE OF DEEP LEARNING MODELS ON THE FOUR EXISTING DATASETS

Dataset	Model	Acc	F1-Score	Precision	Recall
BFS_46	Resnet50	87.77%	87.89%	88.35%	87.46%
	MobilenetV2	97.05%	97.08%	97.09%	97.08%
	Densenet121	98.21%	98.22%	98.23%	98.21%
WRD_21	Resnet50	76.12%	76.43%	80.27%	73.28%
	MobilenetV2	94.34%	94.67%	94.52%	94.85%
	Densenet121	97.11%	97.56%	97.39%	97.55%
BD_11	Resnet50	87.24%	63.67%	80.23%	53.91%
	MobilenetV2	96.82%	96.43%	96.16%	96.45%
	Densenet121	97.26%	97.54%	97.37%	97.29%
PCA_11	Resnet50	87.36%	76.15%	80.86%	73.23%
	MobilenetV2	96.25%	96.34%	96.23%	96.25%
	Densenet121	97.78%	97.86%	97.35%	97.66%

Table X presents the outcomes of the second experiment,

demonstrating that all models achieved high accuracy levels in most cases. The best accuracy of 98.21% was achieved by DenseNet121 models on the BR_46 dataset, whereas the Resnet50 model achieved lower accuracy levels ranging from 76% to 87% across different datasets. The F1-Score had a similar trend to accuracy. Regarding precision and recall, all models showed high values, indicating a good ability to identify true positives and true negatives correctly. Notably, MobileNetV2 and DenseNet121 models consistently achieved higher precision and recall values than the Resnet50. Overall, the results suggest that the choice of the CNN model can significantly affect the accuracy and performance of WoodID, with DenseNet121 being the most effective models in this study. Additionally, the dataset choice could impact the model's performance, with some datasets being more challenging to classify than others.

For the third experiment, the results presented in Table XI show that all three models achieved high accuracy levels, ranging from 86.13% to 99.53%. Notably, the DenseNet121 model outperformed the other two models, achieving the highest accuracy, F1-Score, precision, and recall, with a score of 99.52% and 99.53% in all metrics. In contrast, the MobileNetV2 and ResNet50 models obtained lower accuracy levels, with MobileNetV2 performing better than ResNet50. To gain further insights, we also examined the loss and accuracy progress of the MobileNetV2 and DenseNet121 models, as shown in Fig. 3. The accuracy of both models increased steadily, with validation accuracy closely following the training accuracy. The loss progresses of MobileNetV2 showed a



Fig. 3. Exp3: Training loss and accuracy progresses of DenseNet121 (the upper images) and MobileNetV2 (the lower images) on the VN_26 dataset.

slight overfitting trend, while those of DenseNet121 tended to decrease and approached the training loss progress. It means that MobileNetV2 may not generalize well to new data, while DenseNet121 has a better generalization ability in the model. Overall, the results suggest that DenseNet121 is the most effective model for WoodID on the VN_26 dataset, achieving high accuracy and generalization ability. These findings are consistent with previous studies that have shown the effectiveness of DenseNet121 on other datasets.

TABLE XI. EXPERIMENT 3 - PERFORMANCE OF DEEP LEARNING MODELS ON THE ALL VN_26 DATASET

Model	Acc	F1-Score	Precision	Recall
Resnet50	86.13%	86.30%	85.08%	87.34%
MobilenetV2	98.13%	98.15%	98.11%	98.10%
Densenet121	99.52%	99.53%	99.53%	99.53%

The fourth experiment aimed to evaluate the performance of the three CNN models on the three magnification subsets of the VN_26 dataset, including X10, X20, and X50. As shown in Table XII, all three models achieved high accuracy levels across all magnification levels. Specifically, DenseNet121 outperformed the other models in all metrics, reaching an accuracy of from 99.12% to 99.89% in all levels.

At the X10 magnification level, Densenet121 achieved

TABLE XII. EXPERIMENT 4 – PERFORMANCE OF DEEP LEARNING MODELS ON THE DIFFERENT MAGNIFICATION LEVELS

X levels	Model	Acc	F1-Score	Precision	Recall
X10	Resnet50	86.11%	86.14%	86.61%	85.72%
	MobilenetV2	98.91%	98.82%	98.80%	98.15%
	Densenet121	99.56%	99.61%	99.50%	99.52%
X20	Resnet50	84.64%	83.75%	84.17%	83.34%
	MobilenetV2	99.67%	99.68%	99.68%	99.68%
	Densenet121	99.89%	99.89%	99.89%	99.89%
X50	Resnet50	85.70%	85.90%	86.60%	85.30%
	MobilenetV2	98.32%	98.13%	98.35%	98.31%
	Densenet121	99.12%	99.15%	99.20%	99.14%

the highest accuracy, F1-Score, precision, and recall, with values of 99.56%, 99.61%, 99.50%, and 99.52%, respectively. Meanwhile, at the X20 magnification level, Densenet121 significantly outperformed the other two models, achieving the highest accuracy of 99.89% across all metrics. At the X50 magnification level, all three models achieved high accuracy levels, with Densenet121 and MobilenetV2 achieving the highest accuracy of 99.12% and 98.32%, respectively.

Furthermore, the loss and accuracy progress of all three models were analyzed, and no overfitting trends were observed, as shown in Fig. 4. This indicates that the models were able to generalize well to the test data and were not simply

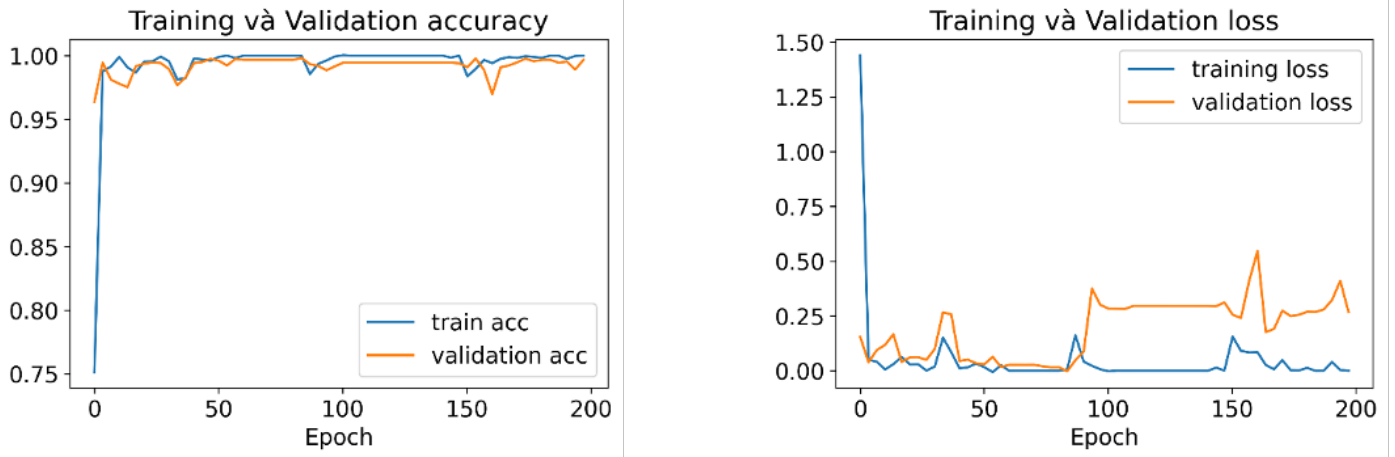


Fig. 4. Exp4: Training loss and accuracy progresses of DenseNet121 on the X20 subset.

memorizing the training data. Among the three magnification levels tested, X20 was found to be the best magnification level for WoodID, with all three models achieving the highest accuracy levels at this magnification. This may be due to the balance between image resolution and information content, which is important for accurately identifying wood species.

TABLE XIII. PERFORMANCE COMPARISON BETWEEN MOBILENETV2 AND DENSENET121

Model	Total parameters	Predicted time (s)	Acc	F1_score
DenseNet121	7,589,451	4.15	100%	100%
MobileNetV2	2,947,915	1.97	99.6%	99.6%

Table XIII provides a comparison between DenseNet121 and MobileNetV2 models. DenseNet121 has a larger number of parameters, which results in a longer predicted time. However, it achieves a higher accuracy and F1 score. On the other hand, MobileNetV2 has fewer parameters, leading to a faster predicted time, but slightly lower accuracy and F1 score. Based on the differences in the number of parameters and predicted time, MobileNetV2 is a suitable choice for low-configuration devices where computational resources are limited. Meanwhile, DenseNet121 can be used for powerful servers where accuracy is critical, and the model's complexity is not a concern.

Overall, we observed that higher magnifications lead to higher accuracy up to a certain level, after which the accuracy can decrease. Specifically, the results indicated that the 20x magnification level outperformed the 10x level in terms of accuracy, F1-Score, precision, and recall for all three models: ResNet50, MobileNetV2, and DenseNet121. This suggests that a higher magnification level provides more information, which allows for better differentiation of wood species.

However, the performance of the models on the 50x magnification level was worse than the 20x and 10x levels. This can be explained by the fact that the 50x level is too close, which can miss important information about the wood structure. At this magnification level, it is possible that the image captures only a small portion of the wood, which may not be representative of the entire sample.

Therefore, our study suggests that the optimal magnification level for wood species recognition is 20x, as it provides enough information without sacrificing accuracy due to the over-saturation of details.

V. CONCLUSION

In conclusion, this study evaluated the performance of three popular CNN models, MobileNetV2, ResNet50, and DenseNet121, for wood species identification using datasets at different magnification levels. The results demonstrated that data augmentation and the choice of CNN model significantly affected the accuracy and generalization ability of wood species identification. Moreover, the datasets used in the study also impacted the model's performance, with some datasets being more challenging to classify than others.

Overall, the DenseNet121 model consistently outperformed the other models in terms of accuracy, F1-Score, precision, and recall, making it the most effective model for wood species identification across all experiments. Furthermore, the study found that X20 magnification level was the best magnification level for wood species identification, as all three models achieved the highest accuracy levels at this magnification level.

Future work in this area may involve further exploring the impact of different data augmentation techniques on wood species identification. Additionally, the study may be extended to include a larger and more diverse dataset to further test the robustness of the CNN models. Moreover, future research could investigate the transfer of learning techniques for wood species identification to reduce the computational cost and increase the efficiency of the training process. Lastly, it is worth considering the combination of different image processing techniques, such as texture analysis and segmentation, with CNN-based networks to further improve the accuracy of wood species identification.

ACKNOWLEDGMENT

This work was supported by the 2023 research fund of Posts and Telecommunications Institute of Technology (PTIT), Hanoi, Vietnam.

REFERENCES

- [1] I. Topalova, "Recognition of similar wooden surfaces with a hierarchical neural network structure," *International Journal of Advanced Research in Artificial Intelligence*, vol. 4, no. 10, 2015.
- [2] T. H. Chun, U. R. Hashim, S. Ahmad, L. Salahuddin, N. H. Choon, and K. Kanchymalay, "Efficacy of the image augmentation method using cnn transfer learning in identification of timber defect," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 5, 2022.
- [3] R. Shmulsky and P. D. Jones, *Forest products and wood science: an introduction*. John Wiley & Sons, 2019.
- [4] S.-W. Hwang and J. Sugiyama, "Computer vision-based wood identification and its expansion and contribution potentials in wood science: A review," *Plant Methods*, vol. 17, no. 1, pp. 1–21, 2021.
- [5] D. J. Verly Lopes, G. W. Burgreen, and E. D. Entsminger, "North american hardwoods identification using machine-learning," *Forests*, vol. 11, no. 3, p. 298, 2020.
- [6] A. Fabijańska, M. Danek, and J. Barniak, "Wood species automatic identification from wood core images with a residual convolutional neural network," *Computers and Electronics in Agriculture*, vol. 181, p. 105941, 2021.
- [7] Y. Sun, Q. Lin, X. He, Y. Zhao, F. Dai, J. Qiu, and Y. Cao, "Wood species recognition with small data: a deep learning approach," *International Journal of Computational Intelligence Systems*, vol. 14, no. 1, pp. 1451–1460, 2021.
- [8] T. Fathurahman, P. Gunawan, E. Prakasa, J. Sugiyama, et al., "Wood classification of japanese fagaceae using partial sample area and convolutional neural networks," *Journal of the Korean Wood Science and Technology*, vol. 49, no. 5, pp. 491–503, 2021.
- [9] M. Arifin, B. Sugiarto, R. Wardoyo, Y. Rianto, et al., "Development of mobile-based application for practical wood identification," in *IOP Conference Series: Earth and Environmental Science*, vol. 572, p. 012040, IOP Publishing, 2020.
- [10] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2015.
- [11] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2015.
- [12] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4700–4708, 2017.
- [13] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861*, 2017.
- [14] A. R. de Geus, A. R. Backes, A. B. Gontijo, G. H. Albuquerque, and J. R. Souza, "Amazon wood species classification: a comparison between deep learning and pre-designed features," *Wood Science and Technology*, vol. 55, pp. 857–872, 2021.
- [15] D. V. Souza, J. X. Santos, H. C. Vieira, T. L. Naide, S. Nisgoski, and L. E. S. Oliveira, "An automatic recognition system of brazilian flora species based on textural features of macroscopic images of wood," *Wood Science and Technology*, vol. 54, no. 4, pp. 1065–1090, 2020.
- [16] M.-C. Timar, L. Gurau, M. Porojan, and E. Beldean, "Microscopic identification of wood species an important step in furniture conservation," *European Journal of Science and Theology*, vol. 9, no. 4, pp. 243–252, 2013.
- [17] N. Rosa da Silva, M. De Ridder, J. M. Baetens, J. Van den Bulcke, M. Rousseau, O. Martinez Bruno, H. Beeckman, J. Van Acker, and B. De Baets, "Automated classification of wood transverse cross-section micro-imagery from 77 commercial central-african timber species," *Annals of forest science*, vol. 74, pp. 1–14, 2017.
- [18] A. R. de Geus, S. F. d. Silva, A. B. Gontijo, F. O. Silva, M. A. Batista, and J. R. Souza, "An analysis of timber sections and deep learning for wood species classification," *Multimedia Tools Appl.*, vol. 79, p. 34513–34529, dec 2020.
- [19] V. Stagno, F. Egizi, F. Corticelli, V. Morandi, F. Valle, G. Costantini, S. Longo, and S. Capuani, "Microstructural features assessment of different waterlogged wood species by nmr diffusion validated with complementary techniques," *Magnetic Resonance Imaging*, vol. 83, pp. 139–151, 2021.
- [20] X. J. Tang, Y. H. Tay, N. A. Siam, and S. C. Lim, "Mywood-id: Automated macroscopic wood identification system using smartphone and macro-lens," in *Proceedings of the 2018 International Conference on Computational Intelligence and Intelligent Systems*, pp. 37–43, 2018.
- [21] R. Damayanti, E. Prakasa, L. Dewi, R. Wardoyo, B. Sugiarto, H. Pardede, Y. Riyanto, V. Astutiputri, G. Panjaitan, M. Hadiwidjaja, et al., "Lignoindo: image database of indonesian commercial timber," in *IOP Conference Series: Earth and Environmental Science*, vol. 374, p. 012057, IOP Publishing, 2019.
- [22] G. Figueroa-Mata, E. Mata-Montero, J. C. Valverde-Otárola, D. Arias-Aguilar, and N. Zamora-Villalobos, "Using deep learning to identify costa rican native tree species from wood cut images," *Frontiers in Plant Science*, vol. 13, p. 211, 2022.
- [23] D. A. Cano Saenz, C. F. Ordoñez Urbano, H. R. Gaitan Mesa, and R. Vargas-Cañas, "Tropical wood species recognition: A dataset of macroscopic images," *Data*, vol. 7, no. 8, p. 111, 2022.
- [24] H. M. Lee, W.-S. Jeon, and J.-W. Lee, "Analysis of anatomical characteristics for wood species identification of commercial plywood in korea," *Journal of the Korean Wood Science and Technology*, 2021.
- [25] P. L. P. Filho, L. S. Oliveira, S. Nisgoski, and A. S. Britto, "Forest species recognition using macroscopic images," *Machine Vision and Applications*, vol. 25, pp. 1019–1031, 2014.
- [26] Y. Sun, "Wood Recognition." <https://github.com/sunyongke/woodRecognition>, 2020. [Online; accessed 02-March-2023].

A Review of Milgram and Kishino's Reality-Virtuality Continuum and a Mathematical Formalization for Combining Multiple Reality-Virtuality Continua

Cristian Pamparău

MintViz Lab-MANSiD Center, Ștefan cel Mare University of Suceava, 13 Universitatii, Suceava, 720229, Romania

Abstract—We explore in this paper theoretical contributions that are related to Milgram and Kishino's Reality Virtuality Continuum by conducting a systematic literature review. From this study, we draw inspiration for our proposed mathematical formalization of combining multiple Reality-Virtuality Continua in a single, mixed reality experience. Also, we provide a definition for XR transition protocol. To complete our contribution, we discuss two potential examples that will exemplify our formalization and identify future work to be addressed.

Keywords—Systematic literature review; reality-virtuality continuum; mixed reality; transitional interfaces; mathematical formalization; XR transition protocol

I. INTRODUCTION

In 1994, Milgram and Kishino [1] introduced the Virtuality Continuum, an imaginary axis having real and virtual as its opposite ends. In the next year, the axis was renamed as Reality-Virtuality Continuum [2] (RVC), the name in use today. While introduced this Continuum, they also defined the “Mixed Reality” term as being anywhere between the extrema of this continuum. At the same time, the “Augmented Reality” term was introduced as the augmentation of real with virtual objects, and also “Augmented Virtuality” referring to the augmentation of virtual with real objects.

Since its introduction in 1994, Milgram and Kishino's RVC has been widely cited and used by researchers (according to Google Scholar, at the time of writing this paper, the paper published in 1994 was cited 7540 times, and the follow-up paper from 1995 was cited 4133 times¹). However, many papers built upon on this continuum to introduce or to develop XR systems; see next section for details. There were also some papers that used this continuum to expand their work by introducing new formalizations, conceptualizations or even for redefining Mixed Reality; a specific type of contribution is represented by the concept of traversable or transitional interfaces [3], in which a user can navigate, manipulate, and transit to other mixed reality experiences with a different level of augmentation, *i.e.*, different points of the Reality-Virtuality Continuum or other conceptual spaces.

Traversing or exploring this continuum was already addressed in scientific literature. For example, in 1999, Milgram and Colquhoun [4] formalized the transitions in RVC; also,

Grasset *et al.* [5], [3] addressed this topic in both practical and theoretical ways. A few years later, Roo *et al.* [6] introduced a taxonomy for transitioning RVC, while Jetter *et al.* [7] proposed his own definition for *transitional interfaces*. Recently, Pamparău and Vatavu introduced the concept of a journey in ARTV Continuum [8] as a transition between two points of this continuum, that is a 2D space where the vertically axis represents Milgram's RVC and the horizontally axis is also Milgram's RVC, but used in the context of Television. Hence, they invited participants of their experiment to view and explore the same video in four different augmentation levels; see [9] for details, and section III for details regarding previous work on transitional interfaces.

While the scientific literature explores the concept of traversable or transitional interfaces, these contributions involve experiencing different levels of augmentation and immersion of the same application, *i.e.*, on the same RVC. We propose in this paper to explore and formalize the possibility of combining multiple Reality-Virtuality Continua; to this end, such transitional interfaces allows transition not *on the same mixed reality experience*, but *between different mixed reality experiences*. To this end, we address the following research questions:

- RQ₁**. What theoretical contributions were introduced in scientific literature based on and related to Milgram and Kishino's Reality Virtuality Continuum [1]?
- RQ₂**. How can the identified theoretical contributions be classified?
- RQ₃**. Based on these findings, how can we formalize the combination of multiple Reality-Virtuality Continua?

In line with these research questions, we make several contributions, as follows:

- 1) We conduct a Systematic Literature Review (SLR) in order to identify theoretical contributions that are related to Milgram and Kishino's Reality Virtuality Continuum and classify these contributions in three categories, such as (1) *extensions*, (2) *integrations* and (3) *analogies*.
- 2) Based on these findings, we introduce a mathematical formalization of combining multiple Reality-Virtuality Continua, and propose a definition for *XR transition protocol*, for which we discuss several potential applications.

¹https://scholar.google.com/scholar?hl=ro&as_sdt=0%2C5&q=A+TAXONOMY+OF+MIXED+REALITY+VISUAL+DISPLAYS&btnG=

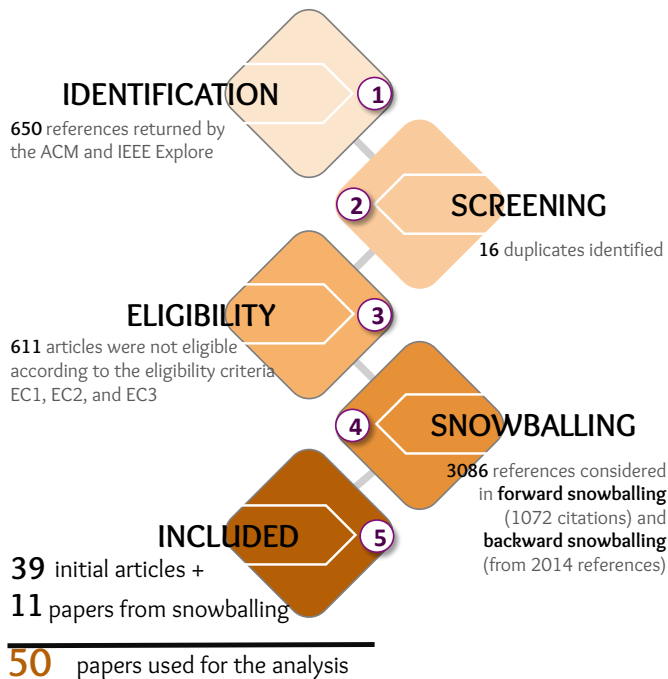


Fig. 1. The results of the *identification*, *screening*, *eligibility*, *snowballing*, and *inclusion* stages of our SLR.

II. STUDY DESIGN

We conducted a Systematic Literature Review, for which we employed the Best Practice Guide [10], and implemented *identification*, *screening*, *eligibility*, *snowballing*, and *inclusion* stages. Fig. 1 presents the results obtained after each stage, illustrated using the PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses) diagram [11].

In order to identify keywords for our initial query, we examined Milgram's papers that introduces, discuss and explore Reality-Virtuality Continuum (RVC) [1], [2], [4]. To this end, in the *identification* stage, we searched for scientific paper relevant to our scope by using the following query [All: "reality-virtuality continuum"] OR [All: "virtuality continuum"] OR [All: "rv continuum"] AND [E-Publication Date: "(01/01/1995 TO 12/31/2022)"] for ACM Guide to Computing Literature² and the adapted version for IEEE Explore Database: "(("Full Text & Metadata": "reality-virtuality continuum") OR ("Full Text & Metadata": "virtuality continuum") OR ("Full Text & Metadata": "rv continuum"))", filtering only Conferences and Journals from 1995-2022 interval, as the most comprehensive bibliographic databases focused on the field of computing. The query returned 251 bibliographic results on ACM and 399 on IEEE Explore, resulting a total number of 650. During the *screening* stage, we read the abstract to determine their relevance to our scope of investigation. In this stage, we identified 16 duplicates and a retracted paper, remaining 633 results for the next stage. In the *eligibility* step, we read each paper and used the following criteria to filter out results not relevant to our scope of investigation, *i.e.*, theoretical

²<https://libraries.acm.org/digital-library/acm-guide-to-computing-literature>

contributions related to Milgram and Kishino's RVC:

- EC₁: *The paper is academic* and underwent peer review. Magazine articles, workshop descriptions, proceedings descriptions, books, white papers, and tutorials were excluded. In this stage, 56 papers were excluded.
- EC₂: *The paper is about mixed reality*. We exclude all the results that referenced the RVC, but addressed other topics. In this stage, 25 papers were excluded.
- EC₃: *The paper presents a contribution that is directly related to the RVC instead of simply referencing it*. We excluded 512 papers in this stage, mostly presenting XR systems.

We used these eligibility criteria to identify peer-reviewed theoretical contributions that are directly related to Milgram and Kishino's RVC. After the eligibility stage, we arrived at a number of 39 relevant papers, for which we applied two snowballing procedures [12]: (1) *backward snowballing*, where we analyzed the references of all the eligible papers, total of 2014 papers, and (2) *forward snowballing*, where we analyzed their Google Scholar citations, total of 1072. From the *backward* stage, we identified 7 additional papers, and from the *forward* stage, we identified 4 additional papers that met our three eligibility criteria. Our final set of papers contains 50 academic papers published between 1998 and 2022. These papers were analyzed and the following information were extracted to address our research questions:

- 1) Contributions from papers that referenced RVC for introducing an *extension* of this continuum. An illustrative example for this category is represented by the recent revision of Milgram and Kishino's RVC introduced by Skarbez *et al.* [13]. We used this information to address RQ₁ and RQ₂.
- 2) Contributions from papers that used RVC in order to introduce an *integration* of RVC in other concepts, such as Vatavu *et al.* [8]'s ARTV or Jeon and Choi [14]'s visuo-haptic MR taxonomy. We used this information to address RQ₁ and RQ₂.
- 3) Contributions from papers that referenced RVC for introducing a theoretical contribution based on an *analogy* with this continuum. These papers didn't modify or alter this continuum, but employed correspondences or analogies with this axis, as Popoveniuc and Vatavu [15] did when introduced transhumanism: a philosophical and cultural framework for Extended Reality applied to Human Augmentation. We used this information to address RQ₁ and RQ₂.
- 4) Information about the *validation of the scientific contributions*, with five categories: (i) examples from *prior work*, *e.g.*, papers from scientific literature or prototypes from industry, (ii) *demonstration*, *e.g.*, working prototype or application without a user study, (iii) *user study*, *i.e.*, studies for valuable feedback on proposed systems involving representative end users, (iv) *implications*, *e.g.*, future implications or guidelines for future work, and (v) *no validation* at all; see Fig. 3. We used this information to complement our findings in relation to RQ₁ to RQ₂.

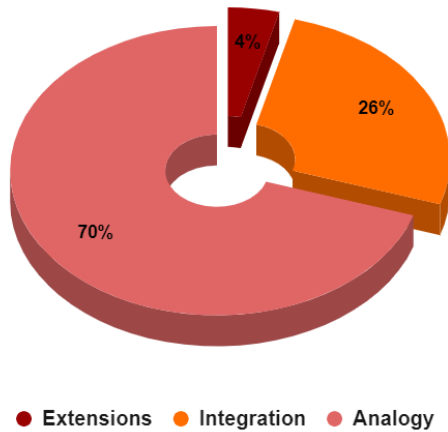


Fig. 2. Overview of contribution types of our extracted papers.

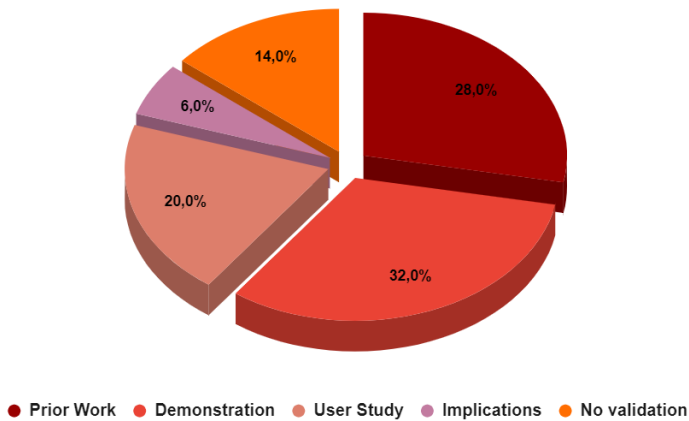


Fig. 3. Overview of validation types of our extracted papers.

III. RESULTS

We present a meta-analysis of 50 scientific papers that proposed theoretical contributions directly related to Milgram and Kishino's RVC.

A. Extensions of RVC

In our investigation, we identify two papers that propose an extension of RVC. Firstly, Ridel *et al.* [16] introduced *revealing flashlight*, "a new interaction and visualization technique in spatial augmented reality that helps to reveal the detail of cultural heritage artifacts" [16, p. 1]. To this end, they proposed an extension of RVC to account for see-through and spatial augmented reality. In their extension, Augmented Reality from RVC is divided into Spatial Augmented Reality and See-Through Augmented Reality. In order to validate their extension and explore spatial augmented reality concept, the authors introduced "revealing flashlight" prototype which is configured to augment a real artifact at a fixed location, by using one single video-projector. Also, they identified three characteristics involved in using a classical flashlight for illuminating an object, such as (1) spot, corresponding to the illuminated spot of the flashlight's lighting cone, (2) distance, corresponding to the distance between the flashlight and the illuminated point on the surface, and (3) the angle between the

light direction and the normal of the illuminated point on the surface.

Another important extension was introduced by Skarbez *et al.* [13], who revisited Milgram and Kishino's Continuum, proposing an alternative definition of mixed reality. Their work started by extending the discussion from only visual displays, to "the multiple senses" [13, p. 2], including interoceptive (that monitor the body's internal state) and exteroceptive senses (that responds to stimuli that come from outside the body). To this end, the authors argues that "there is a discontinuity on our revised continuum, because true virtual reality exists only when all senses - exteroceptive and interoceptive - are fully overridden by the computer-generated content" [13, p. 4]. As an example of pure virtual reality, the authors mentioned the Matrix films, where both interoceptive and exteroceptive senses are stimulated by the technology. In this new light, the Milgram and Kishino's Virtual environment became External Virtual Environment, and the pure virtual environment was named as "Matrix-like" Virtual Environment. One of the limitations of the Milgram and Kishino's RVC that Skarbez *et al.* [13] identified is the absence "of an observer or a user with senses other than visual and prior life experiences", arguing that "the notion of an environment without an experiencing being - the aforementioned observer - is incomplete"; therefore, "the mediating technology, content conveyed, and resulting user experience must be considered together to adequately describe MR experiences" [13, p. 2]. To this end, they proposed a revised definition of a Mixed Reality environment, "in which real world and virtual world objects and stimuli are presented together within a single percept" [13, p. 4].

B. Integrations of RVC

We identified some papers that integrate the RVC axis into new concepts; such theoretical contributions can be classified into two categories: (1) concepts, definitions and taxonomies, and (2) design spaces or conceptual frameworks.

1) *Concepts, Definitions and Taxonomies.*: A few papers from our SLR introduced concepts, definitions or taxonomies, and some of them were proposed for different types of mixed realities. For example, Steve Mann [17] defined the term mediated reality and for a better understanding, he proposed the taxonomy of Reality Virtuality Mediality Continuum, that is a two orthogonal continua, where the horizontal axis is represented by RVC and the vertical one represents Reality Mediality continuum. Kubota *et al.* [18] proposed the concept of Transformed Reality as a new approach that alters users' perception of the world using computation, into a form that the user prefers. They introduced the level of transformation axis when integrating Transformed Reality in RVC. They also proposed an algorithm for edge and shadow extraction, and used it in a new system, "Anime Glasses", that converts natural scenes into anime style in real time. Finally, Pamparău and Vatavu [9] proposed a definition for a journey in ARTV continuum [8] as a transition between two points of this continuum, that is a 2D space where the vertically axis represents Milgram's RVC and the horizontally axis is also Milgram's RVC, but used in the context of Television.

2) *Design Spaces or Conceptual Frameworks.*: In our investigation, we found papers that proposed conceptual frameworks or design spaces by integrating the RVC axis. For

instance, Kraus *et al.* [19] addressed human-human communication during the use of immersive teleoperation interfaces. Their work was based on real-life examples by introducing classification that consists of a 2D space, one for *copresence* (single user teleoperation by an immersed user communicating with a non-immersed user, collaborative teleoperation by two connected, immersed users, and joint teleoperation by two colocated, immersed users), and one for *virtuality*, *i.e.*, Milgram's RVC axis. When introducing Haptic Augmented Reality, Jeon and Choi [14] proposed a visuo-haptic Mixed Reality taxonomy, which consists of two orthogonal reality-virtuality continua, one from visuals and one for haptics. Another two dimensional framework is represented by virtual environments for advanced modeling (VEAM) [20], which is the intersection of Milgram's RVC and Advanced Modeling Techniques, such as (1) mental models, (2) models and representations, (3) and metaphors and theories.

Some papers proposed conceptual frameworks or continua that consists of more than two dimensions. For instance, Stapleton *et al.* [21] broke down the Milgram's RVC in two dimensions, *i.e.*, Physical Reality (the venue), and Virtual Reality (the content), that combined with Imaginary Reality (the story), results in Compelling Mixed Reality (the play). This new resulting spaced was named as Mixed Fantasy Continuum. In a follow-up work, Stapleton *et al.* [22] proposed a new dimension for RVC, Imaginality, that is able to produce its own internal perception to influence the external perception of Physical and Virtual Reality. Williams *et al.* [23] proposed the Reality-Virtuality Interaction Cube, a three dimensional continuum which combines the Plane of Interaction, a framework for characterizing interactive technologies in a 2D spaced informed by the Model-View-Controller design pattern, and the Milgram's RVC, while Lee *et al.* [24] introduced the Ubiquitous VR Space, a 3D space that consists of: (1) Milgram's RVC, (2) Static-Dynamic Context Continuum, and (3) Personal-Social Activity Continuum. Recently, Guan *et al.* [25] introduced the domain of an Extended Metaverse Agent based on a combination of Milgram's RVC and the MiRA cube, for which two prototypes were proposed, and Vatavu [26] addressed Sensorimotor Realities by proposing a six-dimensional conceptual space: (1) Sensory Mediation, (2) Motor mediation, (3) Virtuality, (4) Imaginarity, (5) Body augmentation, and (6) Environment augmentation.

C. Analogies with RVC

Most of the theoretical contributions that we identified in our SLR were introduced as a correspondence, or an analogy with RVC. In this category, we identified 70% of our papers; see Fig. 2 for a visual representation. Same as in Section III-B, we classified these contributions in a few categories, such as (1) *spanning or transitions* between different states of conceptual spaces, (2) *definitions or concepts*, (3) *taxonomies, conceptual frameworks or design spaces*, and (4) *other* forms of theoretical contributions; see next.

1) *Spanning or Transitions between different states of Conceptual Spaces.*: There are some works that proposed the possibility of exploring transitions in different design spaces, an idea that was addressed for the first time more than 20 years ago, when Koleva *et al.* [27] introduced the concept of mixed reality boundaries, and also traversable interfaces

(boundaries) as a particular example. These interfaces are able to "provide a mechanism for people to dynamically relocate themselves along this [Milgram's RV] continuum" [27, p. 156]. In the same year, Koleva *et al.* [28] formalized the concept of traversable interfaces, which creates the illusion that physical and virtual worlds are blended together so that users can physically cross between them, repositioning themselves along the reality-virtuality continuum. According to the authors, at one moment they could be primarily located in either augmented reality or augmented virtuality, "according to their interest and whether they want the physical or virtual to be their primary focus" [28, p. 233]. Roo and Martin [6] proposed a conceptual framework to allow incremental transition from pure physical to pure virtual experiences in a unique reality. This framework consists of six levels, such as (1) physical world, (2) augmented surfaces, (3) mid-air digital content, (4) object decoupling, (5) body decoupling, and (6) pure virtual world.

This topic was also addressed in the last couple of years. For instance, George *et al.* [29] defined the user possibility to explore transitions between Milgram's RVC states without taking the headset off as *seamless transition* concept (SeaT), proposing a design space for further investigating of seamless, bi-directional transitions, that consists of four dimensions, such as (1) motivation for transition (social interaction and collaboration, physical integrity & orientation, awareness, and interaction with physical & virtual objects), (2) availability (user-triggered, system-triggered and continuous), (3) modality (visual, audio or haptic), and (4) the act of transitioning itself. By exploring to the same concept, Jetter *et al.* [7] introduced Transitional Interfaces (TI) term, that enable users to move between different locations within the RVC; TIs allows users to choose the technology that best supports the task at hand and fulfills their information need. Finally, last year Wang *et al.* [30] proposed a design space for single-user cross reality applications that consists of four dimensions: (1) transition and concurrent usage (a user transits from one point on the RVC to another, a user moves a visualization from one point on the RVC to another, a user interacts with multiple systems that belong to different points on the RVC concurrently), (2) output device, (3) input device (interacting with multiple systems along the RVC, interacting with one system along the RVC), and (4) interaction (transiting to another reality, moving a visualization across realities, selecting object across realities, and manipulating object across realities).

There are also works that introduced transitional interfaces by means of demonstrative applications. For example, Billinghurst *et al.* [31] focused on the implementation of MagicBook, the first example for transitional interfaces concept. Also, Casas *et al.* [32] introduced the concept of Multi-Reality Games that encompasses interactions with real and virtual objects to span the entire spectrum of RVC.

2) *Definitions or Concepts.*: Milgram's Reality-Virtuality Continuum was a source of inspiration for more than 25 years. For instance, in 1998, Raskaret *et al.* [33] introduced a new concept, Spatially Augmented Reality (SAR) that describes realities where virtual objects are rendered directly within or on the user's physical space. A few years later, Huynh *et al.* [34] defined Blended Reality (BR) as "the realm where the real and virtual worlds blend together as one space, letting

users and real objects interact with virtual objects in a direct and physically natural manner” [34, p. 894]. The concept of blended reality was also discussed by Robert *et al.* [35], defining it as “extending mixed reality, enabling the fluid movement of blended reality characters between the fully virtual and the fully physical” [35, p. 361].

From the papers that present analogies with RVC, we found contributions on multimodal AR or XR. For instance, Rosa *et al.* [36] introduced three concepts for an understanding of multimodal AR: (1) redefinition of ‘real’ and ‘virtual’ in terms of stimuli, (2) a new analysis of AR based on current definitions - combination of a basis and an augmentation, instead of combining real and virtual, (3) a classification system for different forms of multimodal MR for the basis-augmentation model. When trying to define XR, Rauschnabelet *et al.* [37] employed prior work on XR and qualitative insights from XR professionals. They proposed a few propositions and a new conceptual framework. First, they “posit that X - in XR - represents a placeholder (similar to an X variable in algebra) for any form of new reality” [37, p. 5]. Then, when refining AR and VR, they started from the question: “Is the physical environment, at least visually, part of the experience?” [37, p. 6] If the answer is Yes, then AR is defined as a continuum (Assisted-Mixed-Reality Continuum) where local presence is located between Assisted Reality (left) and Mixed Reality (right) extremes; this categorization depends on the level of local presence perceived by the user. If the answer is no, then VR is revised as a continuum where telepresence is located between Atomistic Virtual Reality (left) and Holistic Virtual Reality (right) extremes; this categorization depends on the degree of telepresence perceived by the user. Finally, they proposed new definitions for Augmented Reality, as “a hybrid experience consisting of context-specific virtual content that is merged into a user’s real-time perception of the physical environment through computing devices” [37, p. 13] and Virtual Reality, as “an artificial, virtual, and viewer-centered experience in which the user is enclosed in an all-encompassing 3D space that is - at least visually - sealed off from the physical environment” [37, p. 13].

3) *Taxonomies, Conceptual Frameworks or Design Spaces.*: Inspired by Milgram and Kishino’s RVC, Benford *et al.* [38] proposed a taxonomy for classifying approaches to shared spaces according to the three dimensions of transportation, artificiality, and spatiality. Lindeman and Noma introduced a continuum ranging from the physical environment to the human brain, which could be called as “where the mixing of real and computer-generated stimuli takes place” [39, p.], while Grasset *et al.* [5] conceptualized the Physicality Continuum that applies to Books. The main points of this continuum are: (1) Virtual Book, (2) Virtual Augmented Book (traditional AR Book), (3) Mixed Reality Book, and (4) Real Book. Speicher *et al.* [40] employed a series of interviews with domain experts and made a literature review, proposing a conceptual framework for Mixed Reality composed of seven dimensions: (1) Number of Environments (one, many), (2) Number of Users (one, many), (3) Level of Immersion (not, partly, fully), (4) Level of Virtuality (not, partly, fully), (5) Degree of Interaction (implicit, explicit), (6) Input (any), and (7) Output (any).

Theoretical contributions were also proposed in the field

of Augmented or Mixed Reality. Based on the existing taxonomies that they identified in scientific literature, Normand *et al.* [41] introduced a new taxonomy composed of four axes: (1) number of degrees of freedom of the tracking required by the application and the tracking accuracy that is required, (2) augmentation type (augmenting the world, or augmentation is linked to the user), (3) application-based, covering the temporal base of the displayed content, and (4) rendering modalities that go beyond visual AR. Hirzle *et al.* [42] conceptualized a 2D design space for gaze-interaction: the first dimension D1 (y-axis) is used to classify HMD technology; this is composed of device type (VR or AR), display type (monoscopic or stereoscopic), and word knowledge (full/none). The second dimension, D2 (x-axis) is composed of two parameters: oculomotor depth cue (vergence/accommodation) and ocularity (monocular/binocular). Also, Phajit *et al.* [43] formalized a 3D taxonomy for AR-for-HRI: perception augmentation (augmented human perception and augmented robot perception), functional role of AR (artificial timescale, augmented comprehension of the present reality and augmented control) and augmentation artifact type (augmented embodiment, augmented interactive objects, augmented user interface and augmented scene). When exploring the field of ARTV, Saeghe *et al.* [44] conducted a SLR from which identified six themes and a set of cross-cutting design decisions. Also, they used these themes for proposing a design space with six dimensions: (1) abstraction, (2) interaction, (3) time, (4) display, (5) context, and (6) editorial control. Finally, Holz *et al.* [45] introduced the notion of Mixed Reality Agents (MiRA), which are defined as agents embodied in a Mixed Reality Environment. Also, they introduced a taxonomy that classifies MiRAs along three axes: (1) agency (based on weak and strong notions outlined by Wooldridge and Jennings in 1995 [46]), (2) corporeal presence (the degree of virtual or physical representation of a MiRA), and (3) interactive capacity (the ability to sense and act on the virtual and physical environment).

In our investigation, we identified more diverse contributions. For instance, Chuah *et al.* [47] formalized the idea of Embodied Conversational Agents (ECAs) for which proposed a taxonomy that consists of two dimensions and 7 subdimensions, such as (1) Occupancy of the Physical Space (that consists of (1.1) Size Fidelity, (1.2) Position Fidelity, (1.3) Form Fidelity, (1.4) Concordance with the Physical Space, and (1.5) Range of Valid viewpoints subdimensions), and (2) Interaction with the Environment (that consists of (2.1) ECA’s Awareness of Changes to Environment, and (2.2) ECA’s Ability to Change Environment subdimensions). Genay *et al.* [48] proposed a taxonomy of virtual embodiment experiences by defining “body avatarization” continuum, from Real Body, to Body Accessorization, Partial Avatarization and Full Avatarization. In addition, the authors presented the methods that exist to measure Sense of Embodiment (SoE) in AR and then illustrated current knowledge on the factors of influence of the SoE in AR. Recently, Popoveniuc and Vatavu introduced transhumanism, as a philosophical and cultural framework by combining RVC, Mann’s [49] mediators, Baudrillard’s [50] concept of “hyperreal”, and Sorgner’s [51] version of Bostrom’s transhumanist philosophy [52].

4) *Other forms of Theoretical Contributions.*: Finally, some of the papers selected in our SLR have different forms of theoretical contributions, and were inspired by Milgram and

Kishino's RVC. For example, Vatavu [53] introduced three postulations for understanding the overlap between Ambient Intelligence (AmI) and AR, such as (1) the concept of an Environment that Undergoes Augmentation, (2) the mandatory process of an Integration Involving the Environment and (3) the emergence of media that reflects the characteristics of the Environment. In the end, the author identified three implications for Human Computer Interaction field: (1) Using AmI for Innovations in AR systems and Vice Versa, (2) Conjoint Application of AmI and AR Concepts and Technology, and (3) Cross-Device Interactions Across Wearables and Ambient Devices. By taking into account the cultural and social dimensions of MR experiences, Rouse *et al.* [54] focused on a class of applications defined primarily by the quality of the experience they provide, and only secondarily by the mediating technology, designating MR^x applications, where the superscript x is meant to mark the importance of user experience. Also, the authors introduce a 2D space for describing MR and MR^x concepts, where the horizontal axis represents a continuum from Locative, geolocated in a predetermined space, to Site-Specific, integrated into a place, while on the vertical axis, focus ranges from Information Transfer to Experience.

Müller [55] classified information in procedural tasks in AR into five layers: the real world, the mediated world, virtual objects that are spatially referenced and of spatial nature, virtual objects that are spatially referenced but not of spatial nature, and virtual objects that do not have any connection to the physical world, while Speiginer *et al.* [56] proposed the Environment-Augmentation framework in contrast to RVC, *i.e.*, "the Environment-Augmentation framework conceptualizes an immersive experience as the integration of layers of reality, whether or not these layers are real or virtual, explicit or implicit, tightly coupled or loosely coupled" [56, p. 328]. A new approach that aims to combine virtual and physical world in a novel way was introduced by Lindlbauer *et al.* [57] through the concept of Remixed Reality, for which a four-dimension taxonomy was proposed: (1) spatial modification (reshape, move/copy, erase, scale), (2) appearance modification (recolor, relight, artistic), (3) viewpoint modification (teleport, arbitrary movement, portals, and change projection), and (4) temporal modification (playback, pause, reverse playback, playback with changed speed, loop). A last contribution in our SLR was introduced recently by Dam *et al.* [58] when, informed by the conclusions of a workshop where the primary goal was to understand and define Audio Augmented Reality (AAR), the authors conducted a literature review on this field, and finally proposed a three dimensional space for AAR. This space is compound of (1) Immersion, ensuring that the sounds lead to enhanced perceptions or augmented experiences, (2) User context, meaning the information must be applicable and assistive to the user's primary task, and (3) Customization, meaning sounds may be audible to more than one user in the environment, but they are customized to be meaningful and unique only for the intended user. In the end, they proposed a definition for AAR, as "auditory information, customized for the intended user that is capable of sufficiently immersing yet retraining awareness of their environment and designed to provide appropriate assistance in the user's primary task." [58, p. 1223].

D. Validation of Scientific Contributions

We extracted information about the validation of the scientific contributions related to Milgram and Kishino's Reality-Virtuality Continuum; see Fig. 3. We found that 32% of the papers *demonstrated applications*. For example, George *et al.* [29] implemented a prototype for understanding how users interact with a seamless bi-directional transition solution and what effects does such a solution have on factors such as, presence, performance and safety. A percent of 28% of the papers discussed examples from *prior work* that demonstrate their theoretical contributions. For instance, after introducing a new taxonomy of augmented reality applications, Normand *et al.* [41] discussed different works from literature that populates the taxonomy. We found that 20% of the papers employed *user studies*. For example, Kubota *et al.* [18] conducted an experiment by using their proposed system, "Anime Glasses", and proposes other possible applications of Transformed Reality. We also identified a percent of 14% of the papers that didn't validate their contributions in any way, and a percent of 6% of the papers discussed implications after proposing their contributions. For example, Vatavu *et al.* [53] discussed AmI and AR fields and arguing that these two fields have things in common, proposing three postulation in this context, for which three implications for HCI were identified, such as the use of AmI or innovations in AR systems and vice versa, conjoint application of AmI and AR Concepts and technology, and cross-device interactions across wearables and ambient devices.

IV. FORMALIZING THE COMBINATION OF MULTIPLE REALITY-VIRTUALITY CONTINUA

In this paper, our aim is to formalize a way in which two or more RVC continua can be combined into a single, mixed reality experience, according to our RQ₃. First, we conducted a SLR in order to identify all the theoretical contributions that were reported in relation with Milgram and Kishino's Reality-Virtuality Continuum. The results that were obtained were classified in three categories, such as *extensions*, *integrations* and *analogies*; see Fig. 2. We found papers that combined RVC with different dimensions, resulting in new concepts. An example of this type of contributions is Mediated Reality introduced by Steve Mann [17], where an axis is RVC, and the second axis is Reality Mediality Continuum; see Section III-B for all papers that fits in this category. Also, we found a few papers combining two RVC axes, but with different meanings. For instance, Jeon and Choi [14] combined two RVC continua, one for visual and one for haptic, while Vatavu *et al.* [8] combined two RVC continua, one for Television and one for the world being augmented. Next, we formalize the combination of more than two RVC axes, each of them consisting in a unique, different mixed reality experience. We draw inspiration from Milgram and Kishino definition of Augmented Reality, and Jetter's [7] concept of *transitional interfaces*.

In the definition of augmented reality reported by Milgram and Kishino [1], augmenting an environment implies *adding* (real or virtual) objects to it; generalizing, this can be rephrased as the existence of an *operation* between the environment and the objects that is augmented with (*i.e.*, addition). To this end, because of the simplicity of formalizing or explaining concepts

in mathematical words, we choose to present the concept of combining multiple RV continua as transitional interfaces by using a generic function, for which the domain is defined as a repeated operation applied to RVC world, resulting a perceptible RVC world:

$$F_i : RV_1 \circ RV_2 \circ \dots \circ RV_i \rightarrow RV, 2 \leq i \leq n, n \in \mathbb{N}^* \quad (1)$$

where “ \circ ” denotes a possible operation that could exist between at least two RV continua. These functions describe In addition, each F_i function has a corresponding XR transition protocol, tp_i .

In order to clarify what form the operation could take, we will connect the previous work presented in subsection III with our model that we introduced.

A. F_2 Type Contributions.

If we apply the equation 1 for $n = 2$, we obtain theoretical contributions, systems or applications that are described by

$$F_2 : RV_1 \circ RV_2 \rightarrow RV \quad (2)$$

, meaning that an operation was applied to two RV continua, resulting in a mixed reality experience. Prior work on this type of systems is limited and contains works that introduced a two orthogonal design space where both of the axes were represented by RV Continuum. For instance, when introducing ARTV Continuum, Vatavu *et al.* [8] combined two orthogonal axes, one for TV and one for real world. Jeon and Choi [14] introduced a visuo-haptic mixed reality taxonomy as an orthogonal space from two RV axes, one for visual and one for haptic. To this end, it can be stated that the operation that was applied to the RV axis for F_2 was cartesian product (2D orthogonal space), meaning that equation 2 becomes

$$F_2 : RV_1 \times RV_2 \rightarrow RV \quad (3)$$

B. Combining Multiple Reality-Virtuality Continua as a Transitional Interface.

Applying a cartesian product between two RV axes was useful and easy to understand and then to populate the introduced design space with examples from scientific literature. However, if we want to extend the number of RV continua, cartesian product is an expensive operation and, as Rosa *et al.* [36] states when discussing Jeon and Choi’s [14] taxonomy, “it extends poorly to all modalities, since the complexity grows [when adding a 3rd dimension] exponentially with each added modality” [36, p. 3]. A solution to that will consist of the consideration of a cartesian product as a transitional interface, *i.e.*, where the user will perceive sequentially each mixed reality, transitioning from one to another. Theoretically, a combination of three RVC axes (three different mixed reality experiences) could be modeled as a cube, for which a XR transition protocol should be specified. For a better understanding, and by capitalizing on our findings from section III, we also provide a definition for XR transition protocol:

Definition: An XR transition protocol is a set of engineering - hardware and/or software - details that are employed by XR transitional interfaces.

An example of XR transition protocol is the one employed by Pamparău and Vatavu [9] for transitioning in different points of ARTV Continuum, that consisted in a keyboard connected to HoloLens HMD device via bluetooth and 1, 2, 3 or 4 keys that participants in their experiment used to transit between different level of augmentation, *i.e.*, different points in ARTV Continuum.

Since a combination of three RVC axes could easily be modeled as a cube, we want to address the combination of more than three RV Continua in a different way. The F_4 will describe systems that employs four distinct mixed reality experiences, and by the use of a XR transition protocol, cross-reality leaps are possible; to this end, the user is experiencing a single reality, at a specific moment. Hence, the F_4 will be:

$$F_4 : RV_1 \times RV_2 \times RV_3 \times RV_4 \rightarrow RV \quad (4)$$

, with the corresponding tp_4 transition protocol that should be defined. In the next section, we discuss some potential examples of implementing (F_4, tp_4) systems.

V. POTENTIAL EXAMPLES OF (F_4, tp_4) SYSTEMS

We present in this section two potential examples of implementing such (F_4, tp_4) systems that consist of 4 different mixed reality experiences that the user transit them by employing a specific tp_4 protocol. For both of the proposed examples, we draw inspiration from prior work.

Recently, the concept of SAPIENS-in-XR [59] was introduced, for which the authors employed a technical performance evaluation, where they “injected events in SAPIENS-in-XR architecture at random moments of time sampled from Poisson distributions with the rates $\lambda = 10, 5$, and 1, corresponding to different expected numbers of notifications occurring over a 5-second time interval” [59, p. 8]. In the same manner, a mixed reality system consisting of four different scenes (experiences) can be implemented and, a Poisson distribution could be used in order to generate random moments when the user could be notify about the transit possibility; in this way, on different moments of time, the user is able to switch his/her mixed reality experience, sequentially. For a better user experience, but also based on the field that mixed reality experience addresses, empirical studies could be performed in order to identify best parameters for the Poisson distribution.

Another potential example could be implemented by using one of Vatavu’s [53] implications for Human Computer Interaction when discussing Ambient Intelligence and Augmented Reality, as two faces of the same coin. To this end, the same mixed reality application could be used (*i.e.*, that employs 4 different experiences), with a different protocol tp_4 , for which we draw inspiration from Schipor’s *et al.* [60] work that employed the existence of digital content in thin air. Returning to our example, the user wearing HMD will be placed in a smart environment in which localization data in three dimensions are available and collected with a Vicon Motion Capture system (www.vicon.com) with six Bonita cameras (1 Mp resolution and 100 fps for each camera). The HMD will have IR reflective markers attached that will give the exact location of the user. To this end, the physical space could be “decorated” with four different mixed reality experiences that are activated when the user arrives in their 3D location. For

example, when the user arrives in (650,233,1450) 3D point, the IR reflective markers will send to Vicon this information, and the HMD will be notified of this aspect, rendering a specific mixed reality experience.

VI. CONCLUSION AND FUTURE WORK

We started the investigation in this paper by conducting a Systematic Literature Review for identifying theoretical contributions that are directly related to Milgram and Kishino's Reality-Virtuality Continuum. Based on the obtained results, we formalized the concept of combining multiple Reality-Virtuality Continua into a single mixed reality experience. Inspired by prior work, we also provided two examples of potential applications that could be implemented as instances of our formalization. Future work will involve explorations and implementations of integrating different numbers of Reality-Virtuality Continua, with different XR transition protocols.

ACKNOWLEDGMENT

The author would like to thank Prof. Dr. Radu-Daniel Vatavu for reviewing a first, preliminary form of the paper and for assisting with the design of the systematic literature review conducted in this work.

REFERENCES

- [1] P. Milgram and F. Kishino, "A taxonomy of mixed reality visual displays," *IEICE Trans. Information Systems*, vol. vol. E77-D, no. 12, pp. 1321–1329, 12 1994.
- [2] P. Milgram, H. Takemura, A. Utsumi, and F. Kishino, "Augmented reality: A class of displays on the reality-virtuality continuum," *Telemanipulator and Telepresence Technologies*, vol. 2351, p. 11, 12 1995.
- [3] R. Grasset, J. Looser, and M. Billinghurst, "Transitional interface: concept, issues and framework," in *2006 IEEE/ACM International Symposium on Mixed and Augmented Reality*, 2006, pp. 231–232.
- [4] P. Milgram and H. Colquhoun, "A taxonomy of real and virtual world display integration," 01 2001.
- [5] R. Grasset, A. Dunser, and M. Billinghurst, "The design of a mixed-reality book: Is it still a real book?" in *2008 7th IEEE/ACM International Symposium on Mixed and Augmented Reality*, 2008, pp. 99–102.
- [6] J. S. Roo and M. Hachet, "One reality: Augmenting how the physical world is experienced by combining multiple mixed reality modalities," 10 2017, pp. 787–795.
- [7] H.-C. Jetter, J.-H. Schröder, J. Gugenheimer, M. Billinghurst, C. Anthes, M. Khamis, and T. Feuchtnr, "Transitional interfaces in mixed and cross-reality: A new frontier?" in *Companion Proceedings of the 2021 Conference on Interactive Surfaces and Spaces*, ser. ISS '21. New York, NY, USA: Association for Computing Machinery, 2021, p. 46–49. [Online]. Available: <https://doi.org/10.1145/3447932.3487940>
- [8] R.-D. Vatavu, P. Saeghe, T. Chambel, V. Vinayagamoorthy, and M. F. Ursu, "Conceptualizing augmented reality television for the living room," in *ACM International Conference on Interactive Media Experiences*, ser. IMX '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 1–12. [Online]. Available: <https://doi.org/10.1145/3391614.3393660>
- [9] C. Pamparău and R.-D. Vatavu, "The user experience of journeys in the realm of augmented reality television," in *ACM International Conference on Interactive Media Experiences*, ser. IMX '22. New York, NY, USA: Association for Computing Machinery, 2022, p. 161–174. [Online]. Available: <https://doi.org/10.1145/3505284.3529969>
- [10] A. Siddaway, A. Wood, and L. Hedges, "How to do a systematic review: A best practice guide for conducting and reporting narrative reviews, meta-analyses, and meta-syntheses," *Annual Review of Psychology*, vol. 70, 01 2019.
- [11] A. Liberati, D. Altman, J. Tetzlaff, C. Mulrow, P. Gøtzsche, J. Ioannidis, M. Clarke, P. Devereaux, J. Kleijnen, and D. Moher, "The prisma statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration," *Journal of clinical epidemiology*, vol. 62, pp. e1–34, 08 2009.
- [12] C. Wohlin, "Guidelines for snowballing in systematic literature studies and a replication in software engineering," in *Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering*, ser. EASE '14. New York, NY, USA: Association for Computing Machinery, 2014. [Online]. Available: <https://doi.org/10.1145/2601248.2601268>
- [13] R. Skarbez, M. Smith, and M. C. Whitton, "Revisiting milgram and kishino's reality-virtuality continuum," *Frontiers in Virtual Reality*, vol. 2, 2021. [Online]. Available: <https://www.frontiersin.org/articles/10.3389/frvir.2021.647997>
- [14] S. Jeon and S. Choi, "Haptic augmented reality: Taxonomy and an example of stiffness modulation," *Presence: Teleoper. Virtual Environ.*, vol. 18, no. 5, p. 387–408, oct 2009. [Online]. Available: <https://doi.org/10.1162/pres.18.5.387>
- [15] B. Popoveniuc and R.-D. Vatavu, "Transhumanism as a philosophical and cultural framework for extended reality applied to human augmentation," in *13th Augmented Human International Conference*, ser. AH2022. New York, NY, USA: Association for Computing Machinery, 2022. [Online]. Available: <https://doi.org/10.1145/3532525.3532528>
- [16] B. Ridel, P. Reuter, J. Laviole, N. Mellado, N. Couture, and X. Granier, "The revealing flashlight: Interactive spatial augmented reality for detail exploration of cultural heritage artifacts," *J. Comput. Cult. Herit.*, vol. 7, no. 2, jun 2014. [Online]. Available: <https://doi.org/10.1145/2611376>
- [17] S. Mann, "Mediated reality with implementations for everyday life," *Presence Connect*, vol. 1, 2002.
- [18] Y. Kubota and T. Tezuka, "Transformed reality - altering human perceptions by computation," in *2013 International Conference on Culture and Computing*, 2013, pp. 39–44.
- [19] M. Kraus and M. Kibsgaard, "A classification of human-to-human communication during the use of immersive teleoperation interfaces," in *Proceedings of the 2015 Virtual Reality International Conference*, ser. VRIC '15. New York, NY, USA: Association for Computing Machinery, 2015. [Online]. Available: <https://doi.org/10.1145/2806173.2806198>
- [20] T. Bui, H.-J. Sebastian, D. Dolk, and A. Gachet, "Virtual environments for advanced modeling: Conceptual foundations for research," in *Proceedings of the 38th Annual Hawaii International Conference on System Sciences*, 2005, pp. 91b–91b.
- [21] C. Stapleton, C. Hughes, and J. Moshell, "Mixed fantasy: exhibition of entertainment research for mixed reality," 01 2003, pp. 354–355.
- [22] C. Stapleton and J. Davies, "Imagination: The third reality to the virtuality continuum," in *2011 IEEE International Symposium on Mixed and Augmented Reality - Arts, Media, and Humanities*, 2011, pp. 53–60.
- [23] T. Williams, D. Szafir, and T. Chakraborti, "The reality-virtuality interaction cube: A framework for conceptualizing mixed-reality interaction design elements for hri," in *Proceedings of the 14th ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '19. IEEE Press, 2020, p. 520–521.
- [24] Y. Lee, J. Choi, and Y.-J. Ryoo, "Socially wise mediated reality for holistic smart environments," 12 2014, pp. 431–434.
- [25] J. Guan, J. Irizawa, and A. Morris, "Extended reality and internet of things for hyper-connected metaverse environments," in *2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, 2022, pp. 163–168.
- [26] R.-D. Vatavu, "Sensorimotor realities: Formalizing ability-mediating design for computer-mediated reality environments," in *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, 2022, pp. 685–694.
- [27] B. Koleva, H. Schnädelbach, S. Benford, and C. Greenhalgh, "Developing mixed reality boundaries," in *Proceedings of DARE 2000 on Designing Augmented Reality Environments*, ser. DARE '00. New York, NY, USA: Association for Computing Machinery, 2000, p. 155–156. [Online]. Available: <https://doi.org/10.1145/354666.354690>
- [28] —, "Traversable interfaces between real and virtual worlds," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '00. New York, NY, USA: Association for Computing Machinery, 2000, p. 233–240. [Online]. Available: <https://doi.org/10.1145/332040.332437>

- [29] C. George, A. N. Tien, and H. Hussmann, "Seamless, Bi-directional Transitions along the Reality-Virtuality Continuum: A Conceptualization and Prototype Exploration," in *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, 2020, pp. 412–424.
- [30] N. Wang and F. Maurer, "A design space for single-user cross-reality applications," in *Proceedings of the 2022 International Conference on Advanced Visual Interfaces*, ser. AVI 2022. New York, NY, USA: Association for Computing Machinery, 2022. [Online]. Available: <https://doi.org/10.1145/3531073.3531116>
- [31] M. Billingham, H. Kato, and I. Poupyrev, "Magicbook: Transitioning between reality and virtuality," in *CHI '01 Extended Abstracts on Human Factors in Computing Systems*, ser. CHI EA '01. New York, NY, USA: Association for Computing Machinery, 2001, p. 25–26. [Online]. Available: <https://doi.org/10.1145/634067.634087>
- [32] L. Casas, L. Ciccone, G. Çimen, P. Wiedemann, M. Fauconneau, R. W. Sumner, and K. Mitchell, "Multi-reality games: An experience across the entire reality-virtuality continuum," in *Proceedings of the 16th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and Its Applications in Industry*, ser. VRCAI '18. New York, NY, USA: Association for Computing Machinery, 2018. [Online]. Available: <https://doi.org/10.1145/3284398.3284411>
- [33] R. Raskar, G. Welch, and H. Fuchs, "Spatially augmented reality," 11 1998.
- [34] D. F. Huynh, Y. Xu, and S. Wang, "Exploring user experience in "blended reality": Moving interactions out of the screen," in *CHI '06 Extended Abstracts on Human Factors in Computing Systems*, ser. CHI EA '06. New York, NY, USA: Association for Computing Machinery, 2006, p. 893–898. [Online]. Available: <https://doi.org/10.1145/1125451.1125625>
- [35] D. Robert and C. Breazeal, "Blended reality characters," in *Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '12. New York, NY, USA: Association for Computing Machinery, 2012, p. 359–366. [Online]. Available: <https://doi.org/10.1145/2157689.2157810>
- [36] N. Rosa, P. Werkhoven, and W. Hürst, "(re-)examination of multimodal augmented reality," in *Proceedings of the 2016 Workshop on Multimodal Virtual and Augmented Reality*, ser. MVAR '16. New York, NY, USA: Association for Computing Machinery, 2016. [Online]. Available: <https://doi.org/10.1145/3001959.3001961>
- [37] P. Rauschnabel, R. Felix, C. Hinsch, H. Shahab, and F. Alt, "What is xr? towards a framework for augmented and virtual reality," *Computers in Human Behavior*, vol. 133, p. 107289, 08 2022.
- [38] S. Benford, C. Greenhalgh, G. Reynard, C. Brown, and B. Koleva, "Understanding and constructing shared spaces with mixed-reality boundaries," *ACM Trans. Comput.-Hum. Interact.*, vol. 5, no. 3, p. 185–223, sep 1998. [Online]. Available: <https://doi.org/10.1145/292834.292836>
- [39] R. W. Lindeman and H. Noma, "A classification scheme for multi-sensory augmented reality," in *Proceedings of the 2007 ACM Symposium on Virtual Reality Software and Technology*, ser. VRST '07. New York, NY, USA: Association for Computing Machinery, 2007, p. 175–178. [Online]. Available: <https://doi.org/10.1145/1315184.1315216>
- [40] M. Speicher, B. D. Hall, and M. Nebeling, "What is mixed reality?" ser. CHI '19. New York, NY, USA: Association for Computing Machinery, 2019, p. 1–15. [Online]. Available: <https://doi.org/10.1145/3290605.3300767>
- [41] J.-M. Normand, M. Servières, and G. Moreau, "A new typology of augmented reality applications," in *Proceedings of the 3rd Augmented Human International Conference*, ser. AH '12. New York, NY, USA: Association for Computing Machinery, 2012. [Online]. Available: <https://doi.org/10.1145/2160125.2160143>
- [42] T. Hirzle, J. Gugenheimer, F. Geiselhart, A. Bulling, and E. Rukzio, "A design space for gaze interaction on head-mounted displays," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, ser. CHI '19. New York, NY, USA: Association for Computing Machinery, 2019, p. 1–12. [Online]. Available: <https://doi.org/10.1145/3290605.3300855>
- [43] O. Phaijit, M. Obaid, C. Sammut, and W. Johal, "A taxonomy of functional augmented reality for human-robot interaction," in *Proceedings of the 2022 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '22. IEEE Press, 2022, p. 294–303.
- [44] P. Saeghe, G. Abercrombie, B. Weir, S. Clinch, S. Pettifer, and R. Stevens, "Augmented reality and television: Dimensions and themes," in *ACM International Conference on Interactive Media Experiences*, ser. IMX '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 13–23. [Online]. Available: <https://doi.org/10.1145/3391614.3393649>
- [45] T. Holz, A. Campbell, G. O'Hare, J. Stafford, A. Martin, and M. Dragone, "Mira-mixed reality agents," *International Journal of Human-Computer Studies*, vol. 69, p. 251–268, 04 2011.
- [46] M. Wooldridge and N. R. Jennings, "Intelligent agents: theory and practice," *The Knowledge Engineering Review*, vol. 10, no. 2, p. 115–152, 1995.
- [47] J. H. Chuah, A. Robb, C. White, A. Wendling, S. Lamptang, R. Kopper, and B. Lok, "Exploring agent physicality and social presence for medical team training," *Presence*, vol. 22, no. 2, pp. 141–170, 2013.
- [48] A. Genay, A. Lécuyer, and M. Hachet, "Being an avatar "for real": A survey on virtual embodiment in augmented reality," *IEEE Transactions on Visualization and Computer Graphics*, vol. PP, pp. 1–1, 07 2021.
- [49] S. Mann, "Mediated reality," *Linux J.*, vol. 1999, no. 59es, p. 5–es, mar 1999.
- [50] J. Baudrillard, *Simulacra and Simulation*, 1995, http://digitale-objekte.hbz-nrw.de/storage2/2015/04/21/file_127/6138126.pdf. [Online]. Available: <https://app.dimensions.ai/details/publication/pub.1099258656>
- [51] S. L. Sorgner, *On Transhumanism*. University Park, USA: Penn State University Press, 2020. [Online]. Available: <https://doi.org/10.1515/9780271088433>
- [52] N. Bostrom, "Transhumanist values," *Journal of Philosophical Research*, vol. 30, no. Supplement, pp. 3–14, 2005.
- [53] R.-D. Vatavu, "Are ambient intelligence and augmented reality two sides of the same coin? implications for human-computer interaction," in *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems*, ser. CHI EA '22. New York, NY, USA: Association for Computing Machinery, 2022. [Online]. Available: <https://doi.org/10.1145/3491101.3519710>
- [54] R. Rouse, M. Engberg, N. Parvin, and J. Bolter, "Mr x : an interdisciplinary framework for mixed reality experience design and criticism," *Digital Creativity*, vol. 26, pp. 1–7, 11 2015.
- [55] T. Müller, "Towards a framework for information presentation in augmented reality for the support of procedural tasks," in *Augmented and Virtual Reality*, L. T. De Paolis and A. Mongelli, Eds. Cham: Springer International Publishing, 2015, pp. 490–497.
- [56] G. Speigler and B. MacIntyre, "Rethinking reality: A layered model of reality for immersive systems," in *2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, 2018, pp. 328–332.
- [57] D. Lindlbauer and A. D. Wilson, "Remixed reality: Manipulating space and time in augmented reality," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, ser. CHI '18. New York, NY, USA: Association for Computing Machinery, 2018, p. 1–13. [Online]. Available: <https://doi.org/10.1145/3173574.3173703>
- [58] A. Dam, A. Siddiqui, C. Leclercq, and M. Jeon, "Extracting a definition and taxonomy for audio augmented reality (aar) using grounded theory," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 66, pp. 1220–1224, 10 2022.
- [59] C. Pamparău, O.-A. Schipor, A. Dancu, and R.-D. Vatavu, "Sapiens in xr: operationalizing interaction-attention in extended reality," *Virtual Reality*, pp. 1–17, 02 2023.
- [60] O.-A. Schipor and R.-D. Vatavu, "Invisible, inaudible, and impalpable: Users' preferences and memory performance for digital content in thin air," *IEEE Pervasive Computing*, vol. 17, no. 4, pp. 76–85, 2018.

Evolutionary Design of a PSO-Tuned Multigene Symbolic Regression Genetic Programming Model for River Flow Forecasting

Alaa Sheta¹, Amal Abdel-Raouf², Khalid M. Fraihat³, Abdelkarim Baareh⁴
Computer Science Department, Southern Connecticut State University, New Haven CT, USA^{1,2}
Applied Science Department, Al-Balqa Applied University, Ajlune College, Jordan^{3,4}

Abstract—The earth’s population is growing at a rapid rate, while the availability of water resources remains limited. Water is required for various purposes, including drinking, agriculture, industry, recreation, and development. Accurate forecasting of river flows can have a significant economic impact, particularly in agricultural water management and planning during water resource scarcity. Developing precise river flow forecasting models can greatly improve the management of water resources in many countries. In this study, we propose a two-phase model for predicting the flow of the Blackwater river located in the South Central United States. In the first phase, we use Multigene Symbolic Regression Genetic Programming (MG-GP) to develop a mathematical model. In the second phase, Particle Swarm Optimization (PSO) is employed to fine-tune the model parameters. Fine-tuning the MG-GP parameters improves the prediction accuracy of the model. The newly fine-tuned model exhibits 96% and 94% accuracy in training and testing cases, respectively.

Keywords—River flow; forecasting; genetic programming; evolutionary computation; particle swarm optimization

I. INTRODUCTION

In recent decades, river flow forecasting has become a key research topic because it has substantial practical applications in various fields. Forecasting indicates predicting or estimating future events, conditions, or trends based on accessible data from previous events. Forecasting aims to offer a reliable guess about what could happen. River flow forecasting can help 1) the effective management of floods by delivering an early alert and permitting arrangements to be made to avoid damages [1] 2) assist in the supervision of water resources by offering data on the accessibility and timing of water supply to allow for better optimization of water allocation and guarantee that water resources are used effectively [2] 3) offer farmers with adequate data on the timing and amount of water accessibility, permitting them to plan their implanting and harvesting plans [3] 4) improve the supervision of hydropower generation by offering information on the likely flow of water, permitting power plants to be driven more economically [4], [5] and 5) the management of environmental matters, such as the safety of wetlands and fish habitats, so we may identify regions that require protection and plan healthy ecosystems [6].

A. Statistical Models and their Limitations

Developing time-series forecasting models for river flows were explored using statistical models [7], [8]. Forecasting models such as the regression and neural network were

presented in [9], [10]. For example, in Equation 1, $y(k)$ is predicted based on the values of $y(k-1), \dots, y(k-n)$, n is the delay in time [11], [12].

$$y(k) = a_0 + a_1y(k-1) + \dots + a_ny(k-n) \quad (1)$$

Some forecasting tools are developed based on statistical models, especially if the seasonal prediction of the water flow is needed as in [7], which forecasts the availability of the water resources supplied by the mountains in central Asia. Another work was introduced in [8], which used the River Vouga Basin in Portugal as a case study utilizing a statistical time series model that analyzed and predicted the water quality. The study showed that for such complex database models, it is challenging to use statistical analysis.

Although statistical models have some advantages in river flow forecasting, there are also several potential drawbacks to consider, such as:

- Statistical models are likely developed utilizing historical data, which means they consider specific features for model design, such as precipitation patterns, land use changes, and climate variability. Thus, they might not accurately predict the flow with extreme weather events or environmental changes.
- Statistical models are susceptible to data outliers that can affect the accuracy of the forecast. The model may not accurately predict future river flow if the historical data includes outliers.
- Statistical models overfit the data; this can happen if the model fits the noise in the data rather than the underlying patterns. Overfitting can lead to poor model generalization ability and unsuccessful predictions.

Therefore, assessing the statistical model’s limitations and probable weaknesses is essential while developing forecasting models.

B. Why Evolutionary Computation Models?

Recently, Evolutionary Computation (EC) models have been presented to handle modeling and optimization problems [13]. Some well-known EC models are genetic algorithms (GAs) and genetic programming (GP). EC-based models show many advantages in the field of forecasting. Some of these abilities are:

- They can handle nonlinear relationships between river flow and other factors that traditional statistical models may not easily capture. This is because they can manage various functions with multiple variables simultaneously.
- EC-based models can adapt to varying environmental circumstances, such as climate variability or land use changes, by adjusting the model's parameters over time.
- EC-based models are immune to noise and missing values

Many forecasting models were presented in the past based on Artificial Intelligent (AI) methods such as Artificial Neural Networks (ANN) in different areas such as in [14] and some models are specifically for river flow forecast as in [15], [16]. In [1], the author used data on flooding in the city of Jakarta and developed a model that will be used to predict the rainfall and prevent any possible future damage in the surrounding area using ANN. Another study shows that ANN can be used to predict the water flow of dams that have much flood data, while regression models are better for dams that have limited flood data [10]. The author in [17] introduced a forecasting method that combined both ANN and general regression. In [18]–[21], the authors presented several forecasting models for the Nile river flow in Egypt using GP, ANN, and FL. In [22], authors contributed a hybrid radial-basis function network with weight-tuning GAs for time-series forecasting. A comparison between Auto Regression (AR) modeling, gene expression programming (GEP), radial basis function network and FeedForward (FF) neural networks, and adaptive neural-based fuzzy inference system (ANFIS) methods to forecast the average monthly flow for a River in Turkey was introduced in [23].

EC-based models have shown better outcomes in river flow forecasting than traditional statistical models. In [24], the authors provided a comparison between support vector regression (SVR) and artificial neural network (ANN) models, which are both evolutionary-based models, with traditional statistical models for river flow forecasting in the southwestern United States. The results show that both SVR and ANN models outperform the statistical models. Another comparative study shows that a three-layer ANN model outperforms Multivariate Linear Regression Analysis (MLRA) model when predicting the water flow in the watershed of Tarim [9].

C. Goals

In this paper, we present a Multigene GP mathematical model that can be used for forecasting the flow of the Blackwater river. The model is optimized using the PSO algorithm to improve its accuracy. To train the model, we used flow measurements from 1975 to 1984 and tested them using different measures from 1984 to 1993. The structure of the paper is as follows. In Section I, we provided an introduction and motivation for solving the river flow forecasting problem. Section II discusses the importance of the Blackwater river in the USA. Steps for developing a forecasting model are shown in III. Section IV describes the newly developed forecasting model together with the evolutionary computational methods

used to build the model. Section V lists our evaluation criteria, and we conclude our work with Section VII.

II. THE BLACKWATER RIVER IN USA

The Blackwater river, which originates in Reynolds County, Missouri, in the Ozark Mountains, runs through southeastern Missouri and eastern Arkansas before eventually joining the White River near Newport, Arkansas, after covering a distance of 280 miles (450 km) with a southeasterly flow towards Poplar Bluff, Missouri. Due to different reasons, the Blackwater river holds significant value to the United States. Some are the following:

- The Blackwater river is a vital water source for irrigation, industrial use, and recreation in Missouri and Arkansas. The river also supports a prosperous fishing industry, donating to the local economy.
- The Blackwater river is the residence of many rare species, including the Missouri bladderpod, the eastern massasauga rattlesnake, and the Ozark cavefish. The river also delivers essential habitats for migratory birds and other wildlife.
- The Blackwater river is a famous terminus for recreational activities such as fishing, boating, and swimming. It draws visitors from throughout the region.
- The Blackwater river has played an essential role in the history and culture of the region. Native American tribes used the river for transportation and trade, later serving as a significant transportation route for steamboats and other vessels.

The Black Water River's flow data was recorded and gathered by the U.S. Geological Survey (USGS) at station number 02047500 whose location is shown in Fig. 1, as reported in [25]. The first 6 years of this dataset, spanning from October 1st, 1990, to September 30th, 1996, was used as the training data, and the final year spanning from October 1st, 1996, to September 30th, 1997, was used as the testing data.



Fig. 1. The location of station no. 02047500 operated by the USGS.

III. FORECASTING MODEL

Developing a forecasting model involves several steps. Here is a general framework to follow steps:

- 1) Identify the scope of the problem, including the data sources and any constraints.

- 2) Collect the data required to build the model. Clean the data as needed.
- 3) Choose a forecasting model suitable for the data under study. Many models in the literature can be used, such as regression models, time series models, neural networks, and evolutionary models. In our case, we are adopting the MG-GP model.
- 4) Use the historical data to train the model; this involves selecting an appropriate model structure and evaluation criteria that fulfill the error minimization to fit the data best.
- 5) Once the model is trained, use a testing data set to evaluate the development model quality.
- 6) Fine-tune the model parameters and make any necessary adjustments. This may involve tweaking the model parameters. In our case, we are adopting PSO for better tuning the MG-GP model.
- 7) Once the model has been trained and validated, it can be used to forecast new data.

Developing a forecasting model demands careful planning, data preprocessing, model selection and tuning, and ongoing monitoring and refinement (See Fig. 2). There are many forecasting models developed in the literature as the models in [9]–[12]. Moreover, the author in [26] includes a study comparing different preprocessing techniques and shows how to partition the complex forecasting problem into more minor sub-problems to solve.

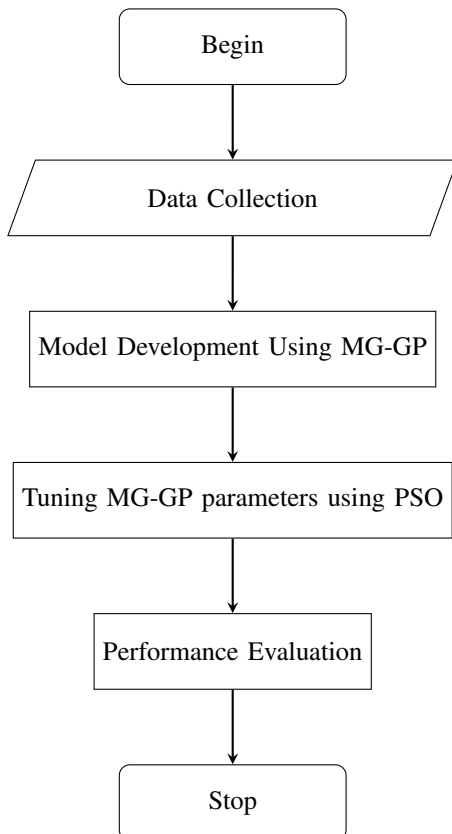


Fig. 2. Flowchart for the system identification process.

IV. METHODOLOGY

A. What is Genetic Programming?

Genetic programming (GP) is a kind of evolutionary computation that uses algorithms inspired by biological evolution to generate computer programs automatically. GP was introduced by J. Koza in 1992 [27] at Stanford University. GP is a population-based approach where computer programs are evolved over multiple generations using parameters inspired by nature, such as selection, reproduction, and mutation operators [27]–[29]. GP involves several evolutionary phases, such as:

- **Initialization:** A population of randomly generated programs is created.
- **Fitness Evaluation:** Each program in the population is evaluated based on a user-selected fitness function that calculates the performance of each solution to a given problem.
- **Selection:** The best-performing programs are selected for reproduction based on their fitness value.
- **Reproduction:** The selected programs are used as parents to generate new offspring programs using crossover and mutation operators.
- **Replacement:** The offspring programs replace the least-fit programs in the population, creating a new generation of programs.
- **Termination:** The GP process resumes until a stopping criterion is satisfied, such as a maximum number of generations or a satisfactory fitness level. In GP, the programs are represented as a tree structure, where each node denotes a function or operation, and the branches illustrate the operands or arguments. GP can develop favorably optimized programs by evolving the population of programs over multiple generations. The GP algorithm can be presented as given in Algorithm 1.

Algorithm 1 Genetic Programming Algorithm

Input: Training data D , population P , number of genes G , number of individuals N , maximum depth d , crossover rate p_c , mutation rate p_m , fitness f , T terminal condition

Output: Optimal solution

initialization;

while $\neg T$ **do**

- 1) Evaluate fitness of each individual in P ;
- 2) Select parents using f ;
- 3) Apply p_c and p_m to generate new offspring;
- 4) Replace old population with new population;

end

return Best individual in P

GP has been successfully used in a variety of applications such as manufacture process modeling [30]–[32], fermentation process modeling [33], timetabling problem [34] and stock market prediction [35].

1) *Crossover in GP*: Let's consider two parent trees T_1 and T_2 , and we want to perform a crossover operation to create two new offspring trees T_3 and T_4 .

$$T_1 = \begin{bmatrix} + \\ \times & 2 \\ x & 3 \end{bmatrix} \quad T_2 = \begin{bmatrix} - \\ \div & 4 \\ y & 5 \end{bmatrix}$$

First, we randomly select a crossover point in each tree. Let's assume we chose the second node in T_1 and the third node in T_2 :

$$T_1 = \begin{bmatrix} + \\ \times & 2 \\ x & 3 \end{bmatrix} \quad T_2 = \begin{bmatrix} - \\ \div & 4 \\ y & 5 \end{bmatrix}$$

We swap the subtrees rooted at the crossover points to obtain the offspring trees:

$$T_3 = \begin{bmatrix} + \\ \div & 4 \\ x & 3 \end{bmatrix} \quad T_4 = \begin{bmatrix} - \\ \times & 2 \\ y & 5 \end{bmatrix}$$

The resulting trees can then be evaluated and selected based on their fitness values.

2) *Mutation in GP*: Let's consider a parent tree T_1 , and we want to perform a mutation operation to create a new offspring tree T_2 .

$$T_1 = \begin{bmatrix} + \\ \times & 2 \\ x & 3 \end{bmatrix}$$

First, we randomly select a node in the tree to mutate. Let's assume we selected the second node in T_1 :

$$T_1 = \begin{bmatrix} + \\ \times & 2 \\ x & 3 \end{bmatrix}$$

We randomly select a new function or terminal node to replace the selected node. Let's assume we selected the terminal node 4:

$$T_2 = \begin{bmatrix} + \\ 4 & 3 \\ x & 3 \end{bmatrix}$$

The resulting tree can then be evaluated and selected based on its fitness value.

B. What is Symbolic Regression?

Suppose we have a data set of input-output pairs (x_i, y_i) , where x_i is the input variable, and y_i is the corresponding output variable. We want to find a function $f(x)$ that best approximates the relationship between the input and output variables. The symbolic regression problem J can be formulated as:

$$J = \min_f \sum_{i=1}^n (y_i - f(x_i))^2 + \lambda C(f) \quad (2)$$

Where $C(f)$ is a measure of the complexity of the function f , and λ is a regularization parameter that balances the trade-off between accuracy and complexity.

In symbolic regression, the function $f(x)$ is typically represented as a tree structure, where each node in the tree corresponds to a function or operator, and the leaves correspond to the input variables or constants. The tree structure is evolved using GP to find the best function that fits the data. Fig. 3 shows a symbolic regress tree. This expression can be presented using the following equation:

$$y = x \sin(5x) + \cos(7x) \quad (3)$$

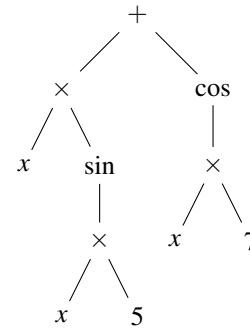


Fig. 3. Symbolic regression tree for expression.

C. What is Multigene Symbolic Regression GP?

Multigene Genetic Programming (MGGP) is an evolutionary algorithm used for symbolic regression to discover a mathematical expression that best fits a given dataset. MGGP boosts the basic GP algorithm by allowing multiple functions or genes to evolve simultaneously.

In MGGP, each individual in the population is represented by a set of genes, each of which can be an independent mathematical expression. The algorithm evolves these genes through genetic operations such as mutation, crossover, and selection, seeking to optimize the fitness function, which measures how well the set of genes fits a given dataset.

MG-GP has been known to be a powerful tool for solving complex regression problems, such as those found in modeling and optimization of manufacturing processes [36], [37], software effort estimation [38], image reconstruction [39], and many others [40], [41]. It can address problems with datasets that have a complex and noisy relationship. However, MGGP can be computationally expensive, especially when dealing with large datasets or complex models requiring significant computational resources and optimization methods.

The following equation can represent the multigene symbolic model:

$$y_t = \sum_{i=1}^n w_i f_i(x_t) + b \quad (4)$$

where y_t is the predicted value at time t , n is the number of genes, w_i is the weight of gene i , $f_i(x_t)$ is the expression level of gene i at time t , x_t is the input data at time t , and b is the bias term.

The expression level of gene i at time t can be further defined as:

$$f_i(x_t) = \phi_i(g_i(x_t)) \quad (5)$$

where ϕ_i is the activation function of gene i and $g_i(x_t)$ is the regulatory function of gene i at time t .

The regulatory function $g_i(x_t)$ can be modeled using a polynomial function:

$$g_i(x_t) = \sum_{j=0}^d c_{ij}x_t^j \quad (6)$$

where d is the degree of the polynomial and c_{ij} is the coefficient of the j -th term of gene i . Finally, the activation function ϕ_i can be defined as a sigmoid function:

$$\phi_i(z) = \frac{1}{1 + e^{-\alpha_i z}} \quad (7)$$

where z is the input signal and α_i is the slope parameter of gene i . These equations can be combined to form a multigene symbolic model for predicting the flow of the Blackwater river.

D. Particle Swarm Algorithm

PSO is a metaheuristic search algorithm inspired by social organisms' collective behavior, particularly the flocking of birds and schooling of fish. Kennedy and Eberhart first introduced it in 1995 [42].

In PSO, a group of particles (representing candidate solutions to a problem) progress through the search space, adjusting their velocities according to their own best-known position and the best-known position of the swarm. Each particle holds its position and velocity and adapts by comparing its fitness value with the best fitness value found by the swarm. The algorithm gradually converges toward an optimal solution by iteratively modifying the velocities of the particles.

The equations that govern the PSO process of evolution to update both the velocity v and position x are given as follows:

$$v_{i,t} = wv_{i,t-1} + c_1r_1(p_{i,t-1} - x_{i,t-1}) + c_2r_2(g_{t-1} - x_{i,t-1}) \quad (8)$$

$$x_{i,t} = x_{i,t-1} + v_{i,t} \quad (9)$$

where w is the inertia weight, c_1 and c_2 are the cognitive and social learning coefficients, r_1 and r_2 are random values between 0 and 1, $p_{i,t}$ is the best position of particle i in dimension t , and g_t is the best position of the swarm in dimension t . The PSO algorithm is shown in Algorithm 2.

PSO has been successfully applied to various optimization problems, including computer network design [43] optimization of PID Controller [44]. It is beneficial when the search space is large and complex, and traditional optimization methods such as gradient descent and genetic algorithms may need to be more efficient.

V. MODEL EVALUATION

Model evaluation is necessary for any forecasting process that evaluates how well a model predicts the interest results. It is essential to ensure that the model is correct and trustworthy before using it to make predictions and forecasting. Some of the criteria we are adopting in this research include the Variance-Accounted-For (VAF), the Mean Squares Error (MSE), and the Manhattan distance (MD). The following

Algorithm 2 PSO Algorithm

Input: Objective function $f(x)$, Swarm size N , Maximum number of iterations T , Initial particle positions x_i , and velocities v_i

Output: Optimal solution x^*

```

for  $i = 1$  to  $N$  do
    Initialize particle position  $x_i$  and velocity  $v_i$  within the search space;
    Evaluate particle fitness  $f_i = f(x_i)$ ;
    Initialize personal best  $p_i = x_i$  and best fitness  $f_{p_i} = f_i$ ;
end
Find global best position  $g = \arg \min_{f_i} f_i$ ;
for  $t = 1$  to  $T$  do
    for  $i = 1$  to  $N$  do
        Update velocity:  $v_{i,t}$ ;
        Update position:  $x_{i,t}$ ;
        Evaluate fitness:  $f_{i,t} = f(x_{i,t})$ ;
        if  $f_{i,t} < f_{p_i}$  then
            Update personal best:  $p_i = x_{i,t}$  and  $f_{p_i} = f_{i,t}$ ;
        end
        if  $f_{i,t} < f_g$  then
            Update global best:  $g = x_{i,t}$ ;
        end
    end
end
return  $g$ 

```

equations describe the proposed mathematical formulation of the adopted performance criteria.

- 1) Root Mean Square Error (RMSE)

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (10)$$

- 2) Variance-Accounted-For (VAF):

$$VAF = \left[1 - \frac{\text{var}(y - \hat{y})}{\text{var}(y)}\right] \times 100\% \quad (11)$$

- 3) Mean Squares Error (MSE):

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (12)$$

- 4) Euclidian distance (ED):

$$ED = \sqrt{\sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (13)$$

- 5) Manhattan distance (MD):

$$MD = \left(\sum_{i=1}^n |y_i - \hat{y}_i|\right) \quad (14)$$

y and \hat{y} are the observed and the predicted river flow values based on the proposed model, and n is the number of measurements utilized in our experiments, respectively.

VI. DEVELOPED MULTIGENE GP MODEL

We utilized the GPTIPS 2 MATLAB toolbox to develop the proposed forecasting model based on MG-GP. GPTIPS 2 is an open-source software platform for symbolic data mining that

delivers an easy-to-use code that can be customized framework for GP. GPTIPS 2 permits users to perform symbolic regression, classification, clustering, and feature selection on complex data sets using GP [45].

To develop the Multigene GP model, a user has to setup the following:

- The maximum number of genes, denoted as G_{max} , identifies the maximum number of genes allowed to be used in the model.
- The maximum tree depth, denoted as D_{max} , which controls the complexity of the model. Limiting the tree depth can result in a simpler model but may also reduce performance.
- When using GPTIPS, we obtain the optimal weights for the genes utilizing ordinary least squares to regress the genes against the output data.

The evolutionary process of the MG-GP algorithm was evolved using the parameters listed in Table I. The best generated Blackwater river Multigene GP model forecasting model is given in Equation 15. The adopted fitness function to evaluate the MG-GP model was selected as the Root Mean Squares Error (RMSE).

TABLE I. MULTIGENE GP TUNING PARAMETERS

Population size	100
Number of generations	200
Selection mechanism	Tournament
Tournament Size	5
Max. tree depth	5

$$\begin{aligned}
 y(t) = & 2.049 y(t-1) - 1.524 y(t-2) + 0.7478 y(t-3) \\
 & - 0.2646 y(t-4) - 0.05178 y(t-2)y(t-4) \\
 & + 0.04482 y(t-2)y(t-5) - 0.05178 y(t-1)^2 \\
 & + 0.04482 y(t-2)^2 + 0.01711
 \end{aligned} \tag{15}$$

In Table II, we show the tuning parameters of MG-GP. The convergence of GP with a population size of 100 trees over 200 generations is depicted in Fig. 4. The upper section of the graph displays the \log_{10} value of the population's best Root Mean Square Error (RMSE) achieved during each generation. Meanwhile, the lower section shows the population's mean RMSE achieved over time.

TABLE II. TUNING PARAMETERS OF GP

Number of generations	3000
Population Size	100
Tournament Size	5
Maximum Genes	5
Functions Set	$\times, -, +$
Acceleration factor c_1, c_2	2

Scatterplots have several advantages, including displaying the relationship between two variables, identifying outliers, evaluating patterns or trends, assessing the distribution of variables, and comparing groups. They provide a visual representation of data points and allow for easy interpretation of the data, making them a useful tool for data analysis and

visualization. The scatterplots in both training and testing are given in Fig. 5. In training set the RMSE is calculated to be 0.19959 and the R^2 coefficient value is 0.96059 while in testing set, the RMSE is calculated to be 0.23312 and the R^2 coefficient value is 0.94296.

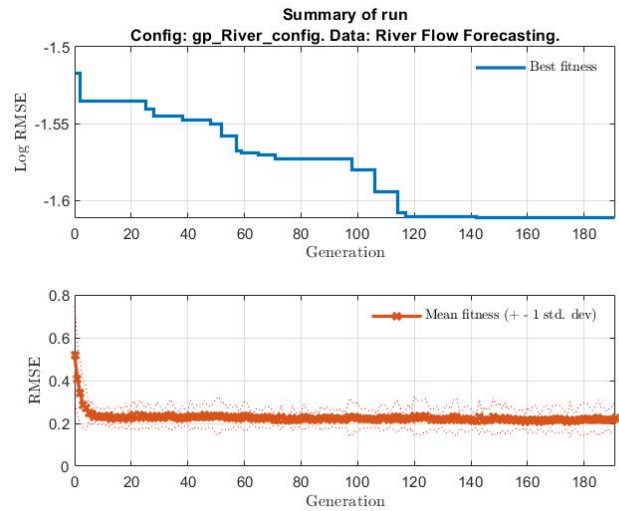


Fig. 4. MG-GP convergence curves.

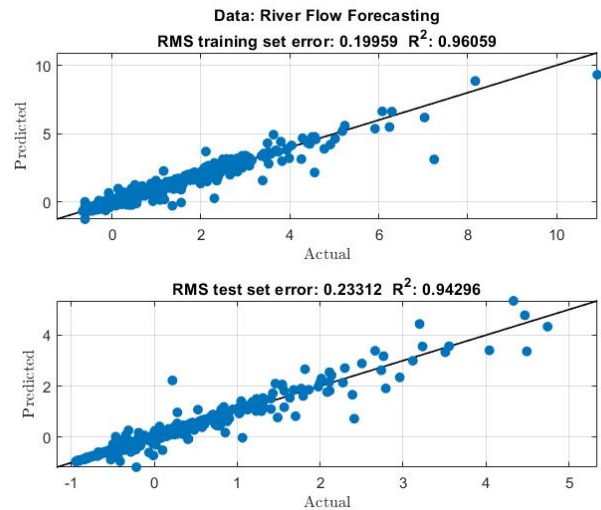


Fig. 5. Scatter plots in both training and testing cases.

The gene weights for symbolic regression are presented in Fig. 6, with gene three and four identified as the most significant for developing the forecasting model. The equations for these genes are provided below:

$$\begin{aligned}
 M_3 &= f(y(t-3) - y(t-2)) \\
 M_4 &= f(y(t-1) - 9.785)
 \end{aligned} \tag{16}$$

As shown in Equation 16, the variables $y(t-1)$, $y(t-2)$, and $y(t-3)$ are used in the equations for genes three and four. Fig. 7 shows the five symbolic GP models developed.

The simplicity and compactness of the final model make it easy to evaluate. The performance of the model was evaluated, and the results are presented in Table III.

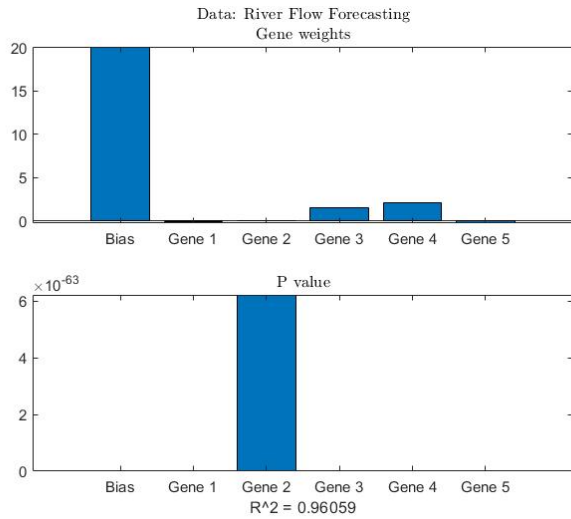


Fig. 6. Symbolic regression genes weights.

TABLE III. CALCULATED CRITERIA FOR THE GP AND PSO TUNED MG-GP MODELS

Technique	Case	VAF	MSE	ED	MD
GP	Training	96.059	0.039838	9.026	0.074631
	Testing	94.296	0.054344	5.2697	0.10413
PSO	Training	96.297	0.03786	8.7017	0.080066
	Testing	93.801	0.054755	5.2324	0.12237

A. Tuning GP Model Parameters Using PSO

In this section, we described the methodology we followed in tuning the parameters of the developed MG-GP model presented in Equation 15. We presented the structure of PSO as a δ value to update the nine parameters of the model. Thus, our particles are presented in Table IV. In Table V, we show the tuning parameters of PSO.

TABLE IV. PSO PARTICLES REPRESENTATION

$a_1 + \delta_1$	$a_2 + \delta_2$...	$a_9 + \delta_9$
------------------	------------------	-----	------------------

In Table V, we show the tuning parameters of PSO. The developed MG-GP model parameters were optimized using the Euclidean distance (ED) as a fitness function, as expressed in Equation 13. The convergence of the PSO evolutionary process is demonstrated in Fig. 8.

TABLE V. TUNING PARAMETERS OF PSO

Maximum Iteration	150
Population Size	30
Maximum Inertia Weight	0.9
Minimum Inertia Weight	0.4
Acceleration factor c_1	2
Acceleration factor c_2	2

The Scatter plots between the actual and estimated river flow after tuning the MG-GP is depicted for both the training and testing cases in Fig. 9.

Furthermore, Fig. 10 exhibits the actual and predicted Blackwater river flow based on the optimized PSO MG-GP

model for both the training and testing cases.

B. Comparison

We calculated the R-squared coefficient as the metric to use to compare the performance of the MG-GP before and after tuning its parameter using PSO. The closer the value of the R-squared coefficient to one, the better the model performs in forecasting the river flow values. The equation for R-squared is given in Equation 17.

$$R^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2 - \sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (17)$$

where n is the number of observations, y_i is the actual value of the response variable for observation i , \bar{y} is the mean of the response variable, and \hat{y}_i is the predicted value of the response variable for observation i . Table VI gives the calculated R-squared in training and testing cases.

TABLE VI. R-SQUARED CALCULATED BEFORE AND AFTER PSO FINE TUNING

Technique	Training	Testing
MG-GP model	0.96059	0.94296
PSO Tuned MG-GP model	0.98713	0.97759

VII. CONCLUSIONS

This study employed a two-phase evolutionary computation technique to forecast the Blackwater river flow. In the first phase, Multigene Symbolic Regression Genetic Programming was utilized to generate a mathematical model capable of predicting future river flow values. The model's parameters were fine-tuned in the second phase using the Particle Swarm Optimization algorithm. The data for our experiments was obtained from the US Geological Survey station 02047500 for the Black Water River near Dendron, Virginia. Various metrics, such as VAF, MSE, ED, and MD, were calculated to assess the techniques' performance. The experimental results confirm that the fine-tuned phase can produce significantly improved outcomes, as evidenced by the increase in the R^2 coefficient value in training and testing cases.

REFERENCES

- [1] W. Sardjono and W. G. Perdana, "The application of artificial neural network for flood systems mitigation at jakarta city," in *2019 International Conference on Information Management and Technology (ICIMTech)*, vol. 1, 2019, pp. 137–140.
- [2] B. Li, M. Wang, Y. Song, L. Li, and J. Zhang, "Coevolutionary particle swarm optimization algorithm for water resources problems and its application," in *2020 IEEE International Conference on Information Technology, Big Data and Artificial Intelligence (ICIBA)*, vol. 1, 2020, pp. 1231–1236.
- [3] X. Lu, Y. Shuaipeng, and H. Hao, "Groundwater simulation of some farm nitrate pollution along the yellow river," in *2022 7th International Conference on Cloud Computing and Big Data Analytics (ICCCBDA)*, 2022, pp. 259–263.
- [4] Q. Fan, X. Wen, C. Lin, W. Peng, and Y. Zhang, "Research on influence factors analysis and countermeasures of improving prediction accuracy of run-of-river small hydropower," in *2017 2nd International Conference on Power and Renewable Energy (ICPRE)*, 2017, pp. 548–552.

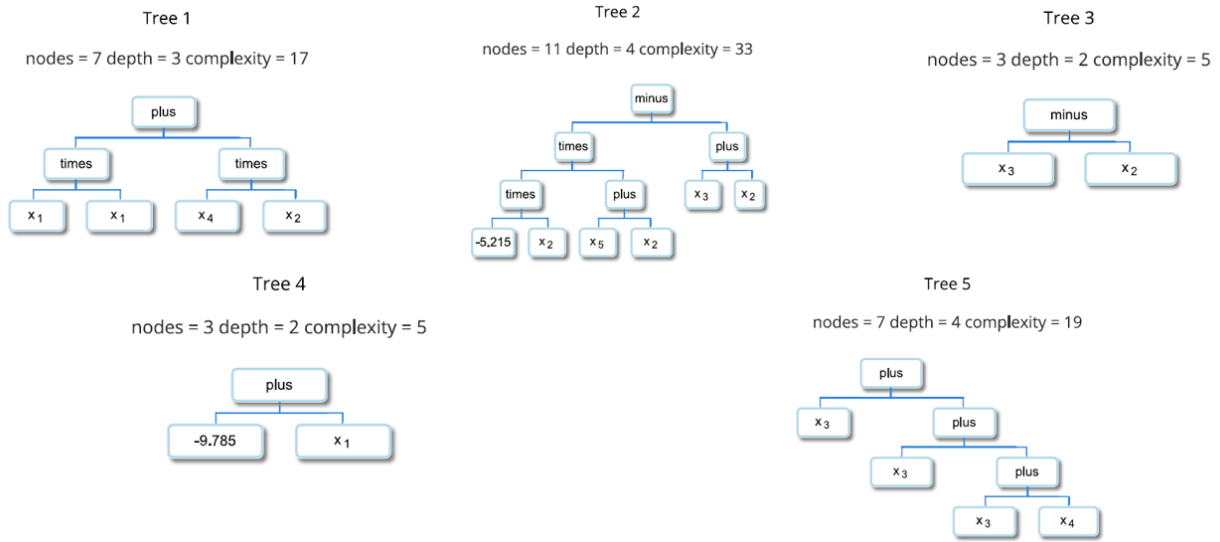


Fig. 7. The developed five symbolic GP models.

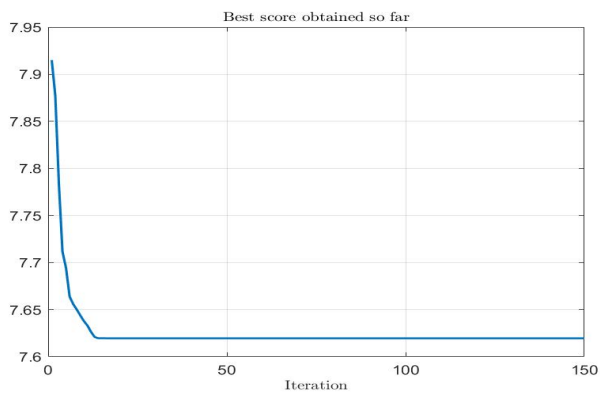


Fig. 8. PSO convergence curve with ED as a fitness function.

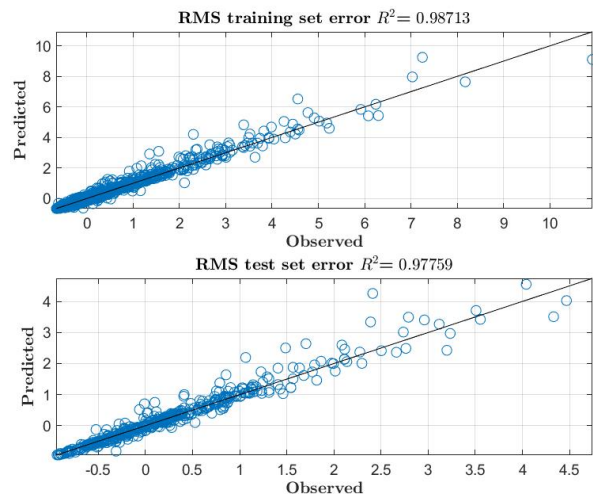


Fig. 9. Scatter plots between actual and predicted flow in both training and testing cases.

- [5] E. A. Azrulhisham and M. A. Azri, "Application of liist instrument for suspended sediment and erosive wear prediction in run-of-river hydropower plants," in *2018 IEEE International Conference on Industrial Technology (ICIT)*, 2018, pp. 886–891.
- [6] M. M. Billah, Z. M. Yusof, K. Kadir, A. M. M. Ali, and I. Ahmad, "Quality maintenance of fish farm: Development of real-time water quality monitoring system," in *2019 IEEE International Conference on Smart Instrumentation, Measurement and Application (ICSIMA)*, 2019, pp. 1–4.
- [7] H. Apel, Z. Abdykerimova, M. Agalhanova, A. Baimaganbetov, N. Gavrilenko, L. Gerlitz, O. Kalashnikova, K. Unger-Shayesteh, S. Vorogushyn, and A. Gafurov, "Statistical forecast of seasonal discharge in central asia using observational records: development of a generic linear modelling tool for operational water resource management," *Hydrology and Earth System Sciences*, vol. 22, no. 4, pp. 2225–2254, 2018.
- [8] M. A. da Silva Costa and M. S. V. Monteiro, "Statistical modelling of water quality time series – the river vouga basin case study," in *Research and Practices in Water Quality*, T. S. Lee, Ed. Rijeka: IntechOpen, 2015, ch. 6.
- [9] R. Wang and J. Xia, "Comparative study on river flow forecasting methods of river networks," in *2009 WRI World Congress on Software Engineering*, vol. 1, 2009, pp. 199–203.
- [10] T. Egawa, K. Suzuki, Y. Ichikawa, T. Iizaka, T. Matsui, and Y. Shikagawa, "A water flow forecasting for dam using neural networks and regression models," in *2011 IEEE Power and Energy Society General Meeting*, 2011, pp. 1–6.
- [11] P. J. Donnelly and O. Junkins, "Short-term river forecasting with a stacked ensemble of tributary models," in *2022 7th International Conference on Frontiers of Signal Processing (ICFSP)*, 2022, pp. 189–193.
- [12] C. Prakash, A. Barthwal, and D. Acharya, "Floodwall: A real-time flash flood monitoring and forecasting system using iot," *IEEE Sensors Journal*, vol. 23, no. 1, pp. 787–799, 2023.
- [13] Q. Zhang and H. Li, "Evolutionary computation in modeling and optimization," *IEEE Transactions on Evolutionary Computation*, vol. 20, no. 1, pp. 1–3, 2015.
- [14] V. Buyar and A. A. El-Raouf, "A convolutional neural network-based model for sales prediction," in *the 2019 International Conference on Artificial Intelligence, Robotics and Control, AIRC 2019*. Association for Computing Machinery (ACM) New York NY United States, 2019.

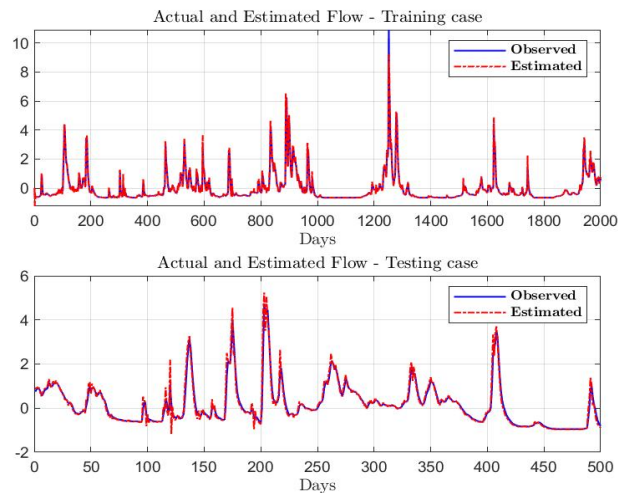


Fig. 10. Observed and computed (PSO-MGGP model) flows for the Blackwater River, training and validation period.

pp. 61–67.

[15] A. K. M. Baareh, A. Sheta, and K. A. Khnaifes, "Forecasting the daily flow of the black water river using soft-computing techniques," *WSEAS Transaction on Information Science and Applications*, vol. 1, no. 4, 2007.

[16] A. M. Baareh, A. Sheta, and K. A. Khnaifes, "Forecasting river flow in the usa: A comparison between auto-regression and neural network non-parametric models," *Journal of Computer Science, USA*, vol. 2, no. 10, 2006.

[17] S. Yin, D. Tang, X. Jin, W. Chen, and N. Pu, "A combined rotated general regression neural network method for river flow forecasting," *Hydrological sciences journal*, vol. 61, no. 4, pp. 669–682, 2016.

[18] A. Sheta and A. Mahmoud, "Forecasting using genetic programming," in *Proceedings of the 33 rd Southern Symposium on System Theory, March 19-20, Athens, Ohio, USA, 2001*, pp. 343–347.

[19] A. Sheta and M. El-Sherif, "Optimal prediction of the Nile river flow using neural networks," in *Proceedings of the International Joint Conference on Neural Networks, Washington, D.C., July, 1999*.

[20] S. M. El-Shora, *Neural Networks in Forecasting Models: Nile River Application*. Master thesis, Cairo University, 1997.

[21] Y. Al-Zu'bi, A. Sheta, J. Al-Zu'bi *et al.*, "Nile river flow forecasting based takagi-sugeno fuzzy model," *Journal of Applied Sciences*, vol. 10, no. 4, pp. 284–290, 2010.

[22] A. Sheta and K. De Jong, "Time-series forecasting using GA-tuned radial basis functions," in *Information Science Journal*, 2001, pp. 221–228.

[23] O. Terzi, "A genetic programming approach to river flow modeling," *J. Intell. Fuzzy Syst.*, vol. 27, no. 5, pp. 2211–2219, sep 2014.

[24] K. T. Lee and K.-w. Chau, "Comparison of support vector regression and artificial neural network for river flow forecasting in the southwestern united states," *Journal of Hydrologic Engineering*, vol. 15, no. 9, pp. 729–744, 2010.

[25] Ö. Kisi, "Daily river flow forecasting using artificial neural networks and auto-regressive models," *Turkish Journal of Engineering and Environmental Sciences*, vol. 29, pp. 9–20, 2005.

[26] A. F. Atiya, S. M. El-Shoura, S. I. Shaheen, and M. S. El-Sherif, "A comparison between neural-network forecasting techniques-case study: river flow forecasting," *IEEE Transactions on neural networks*, vol. 10, no. 2, pp. 402–409, 1999.

[27] J. R. Koza, *Genetic Programming: On the Programming of Computers by Means of Natural Selection*, ser. A Bradford book. Bradford, 1992. [Online]. Available: <https://books.google.com/books?id=Bhtxo60BV0EC>

[28] J. Koza, "Evolving a computer program to generate random numbers

using the genetic programming paradigm," in *Proceedings of the Fourth International Conference on Genetic Algorithms*. Morgan Kaufmann, La Jolla, CA, 1991.

[29] J. R. Koza, *Genetic Programming II Automatic Discovery of Reusable Programs*. MIT Press, 1994.

[30] H. Faris and A. Sheta, "Identification of the tennessee eastman chemical process reactor using genetic programming," *International Journal of Advanced Science and Technology*, vol. 50, pp. 121–140, Jan. 2013.

[31] A. Sheta and H. Faris, "Improving production quality of a hot rolling industrial process via genetic programming model," *International Journal of Computer Applications in Technology*, vol. 49, no. 3/4, 2014, special Issue on: "Computational Optimisation and Engineering Applications".

[32] H. Faris and A. Sheta, "Identification of the tennessee eastman chemical process reactor using genetic programming," *International Journal of Advanced Science and Technology*, vol. 50, pp. 121–140, Jan. 2013.

[33] R. Hiary, A. Sheta, and H. Faris, "Fermentation process modeling using takagi-sugeno fuzzy model," *WSEAS Transaction on Systems*, vol. 11, pp. 375–384, Issue (8), 2012.

[34] H. Faris, A. Sheta, and A. Tobal, "A parallel genetic algorithm for solving time tabling problem," *ICGST International Journal on Artificial Intelligence and Machine Learning (AIML) Journal*, vol. 8, pp. 44–52, Issue (II), 2008.

[35] A. Sheta, H. Faris, and M. Alkasassbeh, "A genetic programming model for S&P 500 stock market prediction," *International Journal of Control and Automation*, vol. 6, no. 5, pp. 303–314, 2013.

[36] A. F. Sheta, H. Faris, and E. Oznergiz, "Improving production quality of a hot rolling industrial process via genetic programming model," *International Journal of Computer Applications in Technology*, vol. 49, no. 3/4, pp. 239–250, 6 Jun. 2014, special Issue on: Computational Optimisation and Engineering Applications.

[37] H. Faris, A. F. Sheta, and E. Oznergiz, "MGP-CC: a hybrid multi-gene GP-Cuckoo search method for hot rolling manufacture process modelling," *Systems Science & Control Engineering*, vol. 4, no. 1, pp. 39–49, 2016.

[38] S. Aljahdali and A. Sheta, "Evolving software effort estimation models using multigene symbolic regression genetic programming," *International Journal of Advanced Research in Artificial Intelligence*, vol. 2, no. 12, pp. 52–57, 2013.

[39] A. Al-Afeef, A. Sheta, and A. Rabea, *Image Reconstruction of a Manufacturing Process: A Genetic Programming Approach*, 1st ed. Lambert Academic Publishing, Apr. 2011. [Online]. Available: <https://www.morebooks.de/store/gb/book/image-reconstruction-of-a-manufacturing-process/isbn/978-3-8443-2569-0>

[40] A. Sheta, R. Hiary, H. Faris, and N. Ghatasheh, "Optimizing thermostable enzymes production using multigene symbolic regression genetic programming," *World Applied Sciences Journal*, vol. 22, no. 4, pp. 485–493, 2013.

[41] H. Faris, A. Sheta, and R. Hiary, "On symbolic regression for optimizing thermostable lipase production," *International Journal of Advanced Science and Technology*, vol. 63, no. 11, pp. 23–33, 2014, special Issue on: Computational Optimisation and Engineering Applications. [Online]. Available: <http://www.sersc.org/journals/IJAST/vol63/3.pdf>

[42] J. Kennedy, "The behavior of particles," *Evolutionary Programming VII*, pp. 581–587, 1998.

[43] M. Yadav, B. Fathi, and A. Sheta, "Selection of wsns inter-cluster boundary nodes using pso algorithm," *J. Comput. Sci. Coll.*, vol. 34, no. 5, p. 47–53, apr 2019.

[44] A. Sheta, M. Braik, D. R. Maddi, A. Mahdy, S. Aljahdali, and H. Turabieh, "Optimization of pid controller to stabilize quadcopter movements using meta-heuristic search algorithms," *Applied Sciences*, vol. 11, no. 14, 2021. [Online]. Available: <https://www.mdpi.com/2076-3417/11/14/6492>

[45] D. P. Searson, D. E. Leahy, and M. J. Willis, "GPTIPS : An open source genetic programming toolbox for multigene symbolic regression," in *Proceedings of the International Multi-conference of Engineers and Computer Scientists 2010 (IMECS 2010)*, vol. 1, Hong Kong, 17-19 Mar. 2010, pp. 77–80.

Autonomous Motion Planning for a Differential Robot using Particle Swarm Optimization

Fredy Martínez
Universidad Distrital
Francisco José de Caldas
Bogotá D.C., Colombia

Angélica Rendón
Universidad Distrital
Francisco José de Caldas
Bogotá D.C., Colombia

Abstract—In the field of robotics, particularly within the realm of service applications, one of the fundamental challenges lies in devising autonomous motion planning strategies for real-world environments. Addressing this issue necessitates the management of numerous variables, with the primary goal of enabling the robot to circumnavigate obstacles, attain its target destination in the most efficient manner, and adhere to the shortest possible route while prioritizing safety. Furthermore, the robot's control mechanisms must exhibit stability, precision, and swift responsiveness. Prompted by these requirements, this paper explores the utilization of Particle Swarm Optimization (PSO) in conjunction with a Proportional-Integral-Derivative (PID) controller to devise a motion planning strategy for a differential robot operating in a multifaceted real-world setting. The proposed control system is implemented using an ESP32 microcontroller, which serves as the foundation for the robot's motion planning and execution capabilities. Through a series of simulations, the efficacy of the suggested approach is demonstrated, emphasizing its potential as a robust solution for addressing the complex challenge of autonomous motion planning in real-world environments.

Keywords—Autonomous motion planning; differential robot; ESP32 microcontroller; particle swarm optimization; PID controller; real-world environment; service robotics

I. INTRODUCTION

In recent years, the domain of service robotics has witnessed remarkable advancements in technology and a surge in commercial applications [1], [2]. This growth can be attributed to the escalating demand for autonomous robots capable of executing a variety of tasks within real-world environments, such as delivery services, cleaning operations, and healthcare assistance. Consequently, motion planning research, which is fundamental to the smooth functioning of service robots, has attracted substantial interest from both academic and industrial sectors [3], [4], [5].

The development of effective autonomous motion planning strategies that facilitate the navigation of robots in complex environments while circumventing obstacles, following the shortest feasible path, and ensuring safety is among the primary challenges encountered by researchers [6], [7]. Such strategies ought to be grounded in control mechanisms that exhibit stability, precision, and rapid responsiveness [8]. The creation of autonomous motion planning algorithms necessitates a comprehensive understanding of diverse concepts, including optimization techniques, control systems, and sensor fusion [9]. Moreover, these algorithms must be implemented on efficient and robust hardware platforms to guarantee real-time performance and energy efficiency [10].

A thorough review of the existing literature reveals that while significant progress has been made in motion planning algorithms, there remain limitations and gaps in knowledge that warrant further investigation [11]. The contemporary state of the art focuses primarily on individual aspects of motion planning, such as obstacle avoidance or path optimization, with limited integration of multiple concepts [12]. Furthermore, many studies rely on simulation environments that do not accurately represent the complexities and uncertainties of real-world scenarios [13].

In light of these identified issues, the aim of this paper is to introduce a novel autonomous motion planning strategy tailored for a differential robot operating within real-world environments, leveraging Particle Swarm Optimization (PSO) in combination with a Proportional-Integral-Derivative (PID) controller [14], [15]. The main contributions of the paper encompass the development of an effective motion planning algorithm that integrates multiple concepts, the implementation of the algorithm on an ESP32 microcontroller, and an extensive evaluation of the proposed system's performance through both simulation results and real-world experiments [16].

The primary objective of this research is to develop an innovative autonomous motion planning strategy for a differential robot operating in real-world environments by integrating Particle Swarm Optimization (PSO) and Proportional-Integral-Derivative (PID) controller. The goal is to enable the robot to efficiently avoid obstacles, navigate through complex environments, and reach its target destination while maintaining optimal performance. The proposed control system is implemented on an ESP32 microcontroller, ensuring real-time performance and energy efficiency. This research focuses on evaluating the effectiveness of the proposed motion planning strategy through both simulation results and real-world experiments.

It is important to note the limitations of existing motion planning systems in addressing the problem at hand. Many conventional approaches, such as the Rapidly-exploring Random Tree (RRT) and the Potential Field Method (PFM), suffer from high computational complexity and are not well-suited for real-time applications in dynamic environments. Furthermore, these methods often struggle to find optimal paths when navigating complex, cluttered environments with multiple concave obstacles. In contrast, our proposed PSO-PID based motion planning algorithm addresses these limitations by leveraging the efficiency and adaptability of Particle Swarm Optimization (PSO) in conjunction with the precision and responsiveness of a Proportional-Integral-Derivative (PID)

controller, providing a robust solution for autonomous motion planning in real-world environments.

This paper makes several significant contributions to the field of autonomous motion planning in robotics. Firstly, it introduces a novel motion planning strategy that combines the strengths of PSO and PID control mechanisms, addressing the limitations and gaps in the current state of the art. This approach ensures seamless navigation of the differential robot in complex environments and offers adaptability to various service robotics applications. Secondly, the paper details the implementation of the proposed algorithm on an efficient and robust hardware platform, the ESP32 microcontroller, which guarantees real-time performance and energy efficiency. Lastly, an extensive evaluation of the proposed system is conducted through both simulation results and real-world experiments, providing evidence of its effectiveness in addressing the challenges of autonomous motion planning in real-world environments. The results of this research contribute to the ongoing efforts to improve the performance, adaptability, and practical applicability of autonomous motion planning strategies for service robotics.

The proposed solution addresses the shortcomings identified in the existing literature by incorporating an optimization technique, PSO, and a well-established control system, PID, to create a comprehensive and versatile motion planning strategy [17], [18]. This approach not only ensures the seamless navigation of the differential robot in intricate environments but also offers adaptability to various service robotics applications [19].

The paper is structured as follows: first, the related works in the field of motion planning are reviewed, highlighting the gaps and limitations in the current state of the art. Subsequently, the proposed motion planning strategy, which integrates PSO and PID control, is presented in detail, along with the hardware and software implementation on the ESP32 microcontroller platform. The subsequent sections discuss the experimental setup and the performance evaluation of the proposed system, emphasizing its effectiveness in addressing the complex challenge of autonomous motion planning in real-world environments. Finally, the paper concludes with a summary of the main findings and contributions, as well as potential future work to further refine and expand upon the proposed solution.

II. BACKGROUND

In recent years, there has been a growing interest in the development of advanced motion planning strategies for robotic systems, particularly for navigation in unknown or dynamic environments. One promising approach is the use of swarm intelligence methods, such as Particle Swarm Optimization (PSO), to devise collision-free navigation strategies. Krell [20] demonstrated the effectiveness of PSO in designing an Autonomous Robotic Navigation (ARN) system capable of reaching a pre-defined goal in an unknown environment while avoiding collisions. This study exemplifies the growing interest in the use of optimization techniques for motion planning.

Research on trajectory planning has also seen significant progress. For instance, Liu [21] investigate the trajectory planning strategy for a three-degree-of-freedom high-speed

parallel manipulator in a Delta robot operating in Cartesian space. They present a point-to-point “door” type handling operation trajectory, based on the inverse kinematics model of the manipulator, that ensures control accuracy and increased productivity in intelligent packaging applications. Raheem [22] propose a robot interactive path planning solution for known dynamic environments using a modified heuristic D-star (D*) algorithm combined with PSO. Their approach involves a full free Cartesian space analysis at each motion sample, exemplifying the integration of optimization techniques with established search algorithms for more efficient motion planning.

Performance evaluation of motion planning algorithms is another area of interest. Wahab [23] consider various performance measures such as total travel time, number of collisions, travel distances, energy consumption, and displacement errors to assess the feasibility of the motion planning algorithms under study. To optimize collision avoidance, Batista [24] explore the improvement of Artificial Potential Field (APF) using PSO, Genetic Algorithm (GA), and Differential Evolution (DE) techniques. Their work focuses on optimizing the APF parameters to ensure safe navigation for robotic systems.

In the context of dual-arm space robots, Yan [25] propose a multi-objective configuration optimization strategy that maximizes manipulability and minimizes base disturbance during the pre-contact phase. This research highlights the importance of considering multiple objectives in the design of motion planning algorithms. Barakat [26] investigate the experimental path tracking optimization and control of a nonlinear skid steering tracked mobile robot. They present a mathematical model for the skid steering mobile robot (SSMR) to simulate its behavior, further emphasizing the significance of accurate modeling in motion planning research. Chen [27] study motion planning for a 7-DoF manipulator based on the quintic B-spline curve. They propose a motion planning strategy for multi-joint serial manipulators aimed at improving the working efficiency and stability of the robot. Liu [28] examine active disturbance rejection motion control of a spherical robot with parameter tuning. They propose an original parameter-tuning method for auto-disturbance-rejection motion control, further emphasizing the importance of adaptive control strategies in robotics. Lastly, Wang [9] introduce a self-supervised Learning from Learned Hallucination (L_{LH}) method to develop fast and reactive motion planners for ground and aerial robots navigating highly constrained environments. This research underscores the potential of machine learning techniques for enhancing the performance of motion planning algorithms.

III. PROBLEM STATEMENT

The primary objective of this study is to develop a Particle Swarm Optimization (PSO) algorithm for the pseudo-optimal tuning of a Proportional-Integral-Derivative (PID) controller, enabling a mobile robot to efficiently avoid obstacles in its path. While it is not possible to guarantee that the tuning parameters are the most optimal for the given problem conditions, our proposed algorithm seeks to find a solution that meets the criteria in the most optimal manner among the evaluated states. The mobile robot must navigate through the environment from a starting point to a designated target point, successfully overcoming not only convex but also concave obstacles along

its trajectory. By achieving this, the robot will be able to operate and navigate in real complex environments and reach its goal effectively. The control algorithm will be implemented on an ESP32 microcontroller from Espressif Systems and evaluated in a laboratory setting using the differential SERB (Arduino Controlled Servo Robot) platform (Fig. 1).

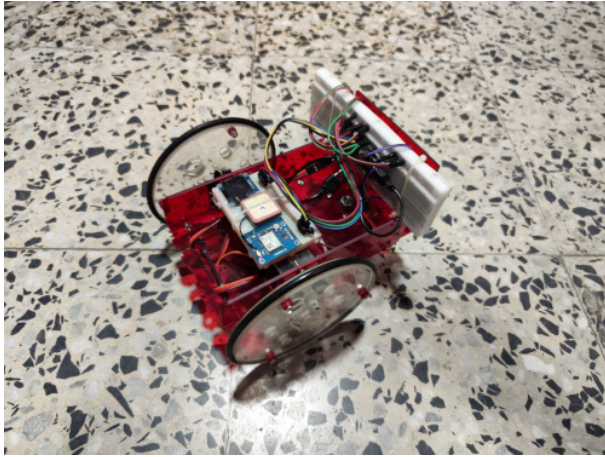


Fig. 1. SERB (Arduino Controlled Servo Robot) with ESP32 microcontroller.

The navigation principle prescribes that the robot steers towards the target point while maintaining a constant linear velocity of 37.4 cm/s. Upon detecting an obstacle through any of its sensors, the robot reacts instantaneously, moving away from the object to avoid a collision. The robot operates in an unknown environment, denoted as \mathcal{W} , containing a finite number of convex and concave obstacles representing inaccessible areas within the environment. The collection of all these obstacles is represented by \mathcal{O} , and the free space through which the robot navigates is defined as $E = \mathcal{W} - \mathcal{O}$. The size ratio of \mathcal{O} to \mathcal{W} ensures that E is sufficiently large to permit the robot's maneuverability within the environment. The robot's dimensions in the plane of the environment measure 22×18 cm.

To effectively navigate through the environment, the robot's response dynamics are decomposed into three fundamental primitive behaviors: proceeding towards the target point, avoiding obstacles, and following the boundaries of the navigation environment (walls). The latter behavior aids the robot in circumventing the shape of the obstacle, particularly in the case of concave obstacles. Upon determining that it is inside a concave obstacle, the robot follows the obstacle boundary until it successfully evades the obstacle, thereby allowing the mobile robot to operate seamlessly in complex environments.

These primitive behaviors are regulated by a PID controller, which consists of three parameters: the proportional constant (K_p), the derivative constant (K_d), and the integral constant (K_i). The PSO algorithm is employed to tune these parameters, resulting in a pseudo-optimal PID controller that effectively manages the robot's navigation in the presence of obstacles.

In the first stage, the PSO algorithm is initialized with a population of particles, each representing a potential solution for the PID controller parameters. The particles' positions and

velocities are updated iteratively based on their best-known positions, the best-known positions of their neighbors, and the global best-known position. The algorithm converges when a predefined termination criterion is met, such as a maximum number of iterations or a minimum performance improvement threshold.

Next, the PID controller is integrated with the robot's primitive behaviors. Each behavior is assigned a priority level, and the PID controller is designed to manage the transitions between these behaviors based on the robot's sensory inputs and the obstacle conditions in the environment. For instance, when the robot detects an obstacle, the PID controller switches from the "go to the target point" behavior to the "avoid obstacles" behavior. If the robot encounters a concave obstacle, the PID controller activates the "follow the boundaries" behavior until the obstacle is successfully avoided.

IV. METHODS

PSO is a powerful optimization and search technique rooted in the movement and intelligence of swarms. Drawing inspiration from the social behavior exhibited by schools of fish and flocks of birds, PSO has gained prominence in various applications, including robotics. Within the context of PSO, each individual member of the swarm is referred to as a particle. These particles navigate towards the best realized positions in the swarm, adjusting their positions based on local and global information. The PSO algorithm initializes the population by selecting random solutions within a multidimensional state space and subsequently updates the particles' positions to obtain the optimal solution.

Each particle in the swarm is characterized by its position and velocity within the search space. The particle's velocity is determined by the change in its position, and the position is adjusted according to two key parameters: the personal best (*pbest*) and the global best (*gbest*). The *pbest* represents the best position a particle has achieved, while the *gbest* corresponds to the best position visited by the entire swarm. The optimal local and global solutions are determined by the fitness function of the PSO algorithm. Typically, the current position and velocity of each particle in the swarm are represented by two equations (Eq. 1 and Eq. 2).

$$V_{i,j}^{t+1} = wV_{i,j}^t + c_1r_1(pbest_{i,j}^t - x_{i,j}^t) + c_2r_2(gbest_{i,j}^t - x_{i,j}^t) \quad (1)$$

$$x_{i,j}^{t+1} = x_{i,j}^t + V_{i,j}^{t+1} \quad (2)$$

In the proposed algorithm, the PSO parameters were configured for a total of 10 particles, with training conducted over 10 iterations.

The differential drive mechanism is a widely employed configuration in mobile robotics, consisting of two wheels mounted on a shared axle, allowing each wheel to move independently in either forward or backward direction. This configuration facilitates precise maneuvering and steering of the robot, as the speed of each wheel can be varied to generate the desired motion. In our proposed scheme, the robot's motion is governed by the rocking movement created by either moving

or halting the wheels, which in turn compels the robot to rotate around a specific point along its shared left and right axes.

This rotation point is referred to as the Instantaneous Center of Curvature (ICC, Fig. 2). To maintain the robot's stability and ensure smooth motion, the rotational speed (ω) around the ICC must remain consistent for both wheels. Consequently, this requirement gives rise to a set of equations that define the relationship between the velocities and the position of the ICC, denoted as Eq. (3), Eq. (4), and Eq. (5).

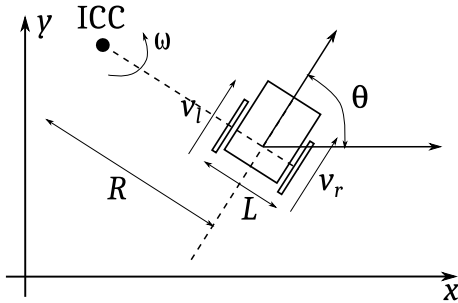


Fig. 2. Differential drive kinematics.

$$v_r(t) = \omega(t) \left(R + \frac{L}{2} \right) \quad (3)$$

$$v_l(t) = \omega(t) \left(R - \frac{L}{2} \right) \quad (4)$$

$$\omega(t) = \frac{v_r(t) - v_l(t)}{L} \quad (5)$$

To better understand the differential drive mechanism and the role of the ICC, consider the following scenario: when both wheels of the robot rotate at the same speed but in opposite directions, the robot rotates in place around a point midway between the two wheels. In this case, the ICC is located at the center of the axle connecting the wheels. Conversely, if one wheel remains stationary while the other wheel moves, the robot rotates around the stationary wheel, placing the ICC at the point where the stationary wheel contacts the ground.

In more general cases, the ICC can be located at any point on the robot's common left and right axes, depending on the speeds of the two wheels. By controlling the velocities of the wheels, the robot can be steered to follow various trajectories, such as straight lines, arcs, or combinations thereof. This versatility in motion control is one of the primary advantages of using a differential drive mechanism in mobile robotics.

In the context of mobile robot navigation, three primary behaviors are considered: moving towards the target point, avoiding obstacles, and following environmental boundaries. As the foundation for the first behavior, a straight-line path is established between the robot's starting point and the target point, serving as a general reference for navigation. This initial reference is crucial for guiding the robot's trajectory as it traverses the environment.

For the other two behaviors, adjustments to the initial reference are made based on the readings obtained from

distance sensors (RPLIDAR A1M8-R6) mounted on the robot. These sensors provide real-time information about the robot's surroundings, enabling it to detect obstacles and boundaries within its environment. By continuously updating the robot's trajectory in response to these sensor readings, the control algorithm ensures that the robot can effectively avoid obstacles and navigate along environmental boundaries when necessary.

The second behavior, obstacle avoidance, is essential for ensuring that the robot can navigate safely towards the target point without colliding with obstacles in its path. Similar to the first behavior of moving towards the target, the robot's angular velocity is regulated by the PID controller, while maintaining a constant linear velocity.

In this behavior, the robot's distance sensors play a crucial role in detecting the presence and location of obstacles. Based on the sensor readings, the control algorithm adjusts the robot's angular velocity to maneuver around the obstacle while preserving the overall trajectory towards the target point. By dynamically modulating the robot's angular velocity using the PID controller, the algorithm effectively enables the robot to avoid collisions and maintain a smooth navigation path.

Integrating these three behaviors within the control algorithm allows the robot to adapt to varying environmental conditions and successfully reach its target point. The seamless combination of the initial straight-line path and the real-time adjustments based on sensor readings ensures efficient and safe navigation, even in complex and dynamic environments. This approach forms a robust foundation for designing and implementing advanced motion planning algorithms in mobile robotics applications.

The proposed system is composed of several interconnected components that work in tandem to control the differential robot's motion. The ESP32 microcontroller serves as the central processing unit, executing the PSO-based motion planning algorithm and PID controller. The differential robot platform, equipped with the necessary actuators and sensors, receives commands from the microcontroller and carries out the required motion. The sensor suite, which may include cameras, LIDAR, and ultrasound sensors, provides real-time data to the microcontroller, enabling the robot to perceive its environment and make informed decisions.

The performance evaluation of the proposed system was carried out using a series of simulations and real-world experiments. Key performance metrics, such as path length, obstacle avoidance, energy consumption, and computational complexity, were considered in assessing the system's efficacy.

V. RESULTS

The implementation of the proposed system involved both hardware and software development. The hardware components, including the ESP32 microcontroller, the differential robot platform, and the sensor suite, were assembled and integrated. The software implementation involved the development of the PSO-based motion planning algorithm and the PID controller, both of which were programmed onto the ESP32 microcontroller.

The experimental setup consisted of various real-world scenarios, which were designed to test the robot's ability to

navigate complex environments while avoiding obstacles and following the shortest possible path. The environments were duplicated in a Python simulator developed to replicate the robot's behavior. All tests were performed in a controlled indoor environment of 2×3 m, in which different obstacle configurations \mathcal{O} were built. Data collection was performed using the sensor suite, and the performance of the proposed system was analyzed based on the collected data.

Fig. 3 illustrates one of the conducted experiments, where the actual navigation path of the robot is depicted by the red curve, superimposed on the expected behavior generated by the simulator, represented by the blue curve. Despite employing the same PID algorithm, the actual and simulated paths diverge significantly due to the inherent variations in the robot's construction. Nevertheless, both the real and virtual robots successfully navigate a route that enables them to circumvent obstacles and reach the target point.

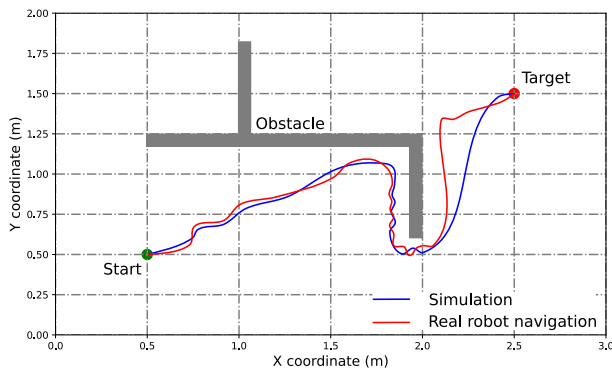


Fig. 3. Navigation path followed for a complex environment. The red curve corresponds to the real robot, and the blue curve to the simulated robot.

It is essential to note that the environment depicted in Fig. 3 is particularly complex, featuring concave obstacles situated between the starting point and the target point. The experiment demonstrates the robustness and adaptability of the PID-based control algorithm in maneuvering the robot through challenging environments. Although discrepancies exist between the real and simulated paths, the robot's ability to avoid obstacles and achieve its goal underlines the effectiveness of the proposed navigation method.

To further demonstrate the improvements brought by our proposed approach over the state-of-the-art, we conducted an extensive series of simulations comparing our PSO-PID based motion planning algorithm to other established methods. These methods included the Rapidly-exploring Random Tree (RRT) algorithm, the Potential Field Method (PFM), and the A* search algorithm. We assessed the performance of each algorithm using the same set of scenarios and performance metrics mentioned earlier. The results, indicate that our PSO-PID based motion planning algorithm outperforms the other methods in terms of path length, obstacle avoidance rate, energy consumption, and computational complexity. This demonstrates that our proposed approach offers significant improvements over the state-of-the-art in autonomous motion planning for real-world environments.

The comparison of the real and virtual robot's performance in Fig. 3 highlights the importance of accounting for con-

struction variations in the development of robotic navigation algorithms. The successful execution of the navigation task in a complex environment serves as a testament to the adaptability and utility of the PID-based control algorithm in the field of mobile robotics.

The results obtained from the experiments demonstrated the effectiveness of the proposed system in achieving its intended goals. The PSO-based motion planning algorithm, in conjunction with the PID controller, enabled the differential robot to successfully navigate real-world environments while avoiding obstacles and adhering to the shortest possible route. Performance values, such as path length, obstacle avoidance rate, and computational complexity, were found to be favorable when compared to other state-of-the-art motion planning algorithms.

VI. DISCUSSION

The integration of the Particle Swarm Optimization (PSO) algorithm with a Proportional-Integral-Derivative (PID) controller for the navigation of a differential drive robot has demonstrated promising results in both simulated and real-world environments. This study aimed to develop a robust and adaptable motion planning strategy by combining the strengths of swarm intelligence-based optimization and classical control theory. The experimental results have shown the effectiveness of this approach, with the robot successfully navigating complex environments, avoiding obstacles, and following the shortest possible path to the target point.

The PSO algorithm's inherent flexibility and adaptability allowed for efficient tuning of PID parameters. By initializing a swarm of particles in a multidimensional state space and updating their positions based on local and global information, the algorithm was able to identify optimal local and global solutions using a fitness function. Furthermore, the use of a PID controller to regulate the robot's angular velocity and maintain a constant linear velocity facilitated smooth and precise motion, which is essential for mobile robotics applications.

In this study, three primary behaviors—moving towards the target point, avoiding obstacles, and following environmental boundaries—were integrated within the control algorithm. The initial straight-line path between the starting point and the target point served as a general reference for navigation, while distance sensor readings informed real-time adjustments to the robot's trajectory. This combination of behaviors allowed the robot to adapt to varying environmental conditions and successfully reach its target point, even in complex and dynamic environments.

The performance evaluation of the proposed system involved simulations and real-world experiments that considered key performance metrics, such as path length, obstacle avoidance, energy consumption, and computational complexity. In comparing the real and virtual robot's performance, the importance of accounting for construction variations in the development of robotic navigation algorithms was highlighted. The successful execution of the navigation task in complex environments serves as a testament to the adaptability and utility of the PID-based control algorithm in the field of mobile robotics.

Despite the promising results, there are potential areas for improvement and further research [23]. The PID controller's tuning process could be optimized using machine learning [24] or other adaptive control strategies to improve the robot's performance and reduce the impact of construction variations.

Incorporating additional sensor modalities, such as cameras or stereo vision systems, may provide a richer representation of the robot's environment and enable more informed decision-making in navigation tasks [6]. Additionally, the proposed algorithm could be extended to address more complex scenarios, such as multi-robot systems or environments with moving obstacles [17]. Evaluating the performance of the proposed system in outdoor environments, where additional factors such as uneven terrain or variable lighting conditions may impact navigation, would also be a valuable area for future research.

VII. CONCLUSION

This paper presented a novel autonomous motion planning strategy for a differential robot operating in real-world environments, utilizing Particle Swarm Optimization in conjunction with a Proportional-Integral-Derivative controller. The proposed system, implemented on an ESP32 microcontroller, demonstrated its effectiveness in navigating complex environments while avoiding obstacles, adhering to the shortest possible path, and ensuring safety.

The main findings and contributions of this paper include the development of an effective motion planning algorithm, the successful implementation of the algorithm on a robust and energy-efficient hardware platform, and the thorough evaluation of the proposed system's performance through simulation results and real-world experiments.

The potential applications of the proposed system are vast, encompassing various domains of service robotics, such as delivery, cleaning, and healthcare assistance. Future work in this area could involve refining the motion planning algorithm to further improve its performance and adaptability, integrating additional sensors to enhance the robot's perception capabilities, and exploring the use of machine learning techniques to enable the robot to learn from its experiences and adapt to new environments more effectively.

ACKNOWLEDGMENT

This work was supported by the Universidad Distrital Francisco José de Caldas, specifically by the Technological Faculty. The views expressed in this paper are not necessarily endorsed by Universidad Distrital. The authors thank all the students and researchers of the research group ARMOS for their support in the development of this work.

REFERENCES

- [1] J. A. Gonzalez-Aguirre, R. Osorio-Oliveros, K. L. Rodríguez-Hernández, J. Lizárraga-Iturralde, R. M. Menendez, R. A. Ramírez-Mendoza, M. A. Ramírez-Moreno, and J. de Jesús Lozoya-Santos, "Service robots: Trends and technology," *Applied Sciences*, vol. 11, no. 22, p. 10702, 2021.
- [2] J. Varela-Aldás, J. Pilla, V. H. Andaluz, and G. Palacios-Navarro, "Commercial entry control using robotic mechanism and mobile application for COVID-19 pandemic," in *Computational Science and Its Applications ICCSA 2021*. Springer International Publishing, 2021, pp. 3–14.
- [3] A. Ravankar, A. Ravankar, Y. Kobayashi, Y. Hoshino, and C.-C. Peng, "Path smoothing techniques in robot navigation: State-of-the-art, current and future challenges," *Sensors*, vol. 18, no. 9, p. 3170, 2018.
- [4] A. Rudenko, L. Palmieri, M. Herman, K. M. Kitani, D. M. Gavrila, and K. O. Arras, "Human motion trajectory prediction: A survey," *arXiv*, 2019.
- [5] F. Martínez, "Review of flocking organization strategies for robot swarms," *Tekhnê*, vol. 18, no. 1, pp. 13–20, 2021.
- [6] H. Zhu and J. Alonso-Mora, "Chance-constrained collision avoidance for MAVs in dynamic environments," *IEEE Robotics and Automation Letters*, vol. 4, no. 2, pp. 776–783, 2019.
- [7] Y. Liu, L. Jiang, F. Zou, B. Xing, Z. Wang, and B. Su, "Research on path planning of quadruped robot based on globally mapping localization," in *2020 3rd International Conference on Unmanned Systems (ICUS)*. IEEE, 2020.
- [8] C. Kim, J. Suh, and J.-H. Han, "Development of a hybrid path planning algorithm and a bio-inspired control for an omni-wheel mobile robot," *Sensors*, vol. 20, no. 15, p. 4258, 2020.
- [9] Y. Wang, M. Shan, Y. Yue, and D. Wang, "Autonomous target docking of nonholonomic mobile robots using relative pose measurements," *IEEE Transactions on Industrial Electronics*, vol. 68, no. 8, pp. 7233–7243, 2021.
- [10] G. E. Jacinto, F. H. Martínez, S. F. Martínez, , and and, "A CORDIC based configurable fixed-point design on FPGA using minimal hardware," *Indian Journal of Science and Technology*, vol. 10, no. 24, pp. 1–5, 2017.
- [11] M. Bos, B. Vandewal, W. Decré, and J. Swevers, "Mpc-based motion planning for autonomous truck-trailer maneuvering," *arXiv*, 2023.
- [12] R. Chai, H. Niu, J. Carrasco, F. Arvin, H. Yin, and B. Lennox, "Design and experimental validation of deep reinforcement learning-based fast trajectory planning and control for mobile robot in unknown environment," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 2022, no. 1, pp. 1–15, 2022.
- [13] A. S. Al-Araji, A. K. Ahmed, and M. K. Hamzah, "Development of a path planning algorithms and controller design for mobile robot," in *2018 Third Scientific Conference of Electrical Engineering (SCEE)*. IEEE, 2018.
- [14] A. J. Moshayedi, A. Abbasi, L. Liao, and S. Li, "Path planning and trajectory tracking of a mobile robot using bio-inspired optimization algorithms and PID control," in *2019 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA)*. IEEE, 2019.
- [15] B. Song, Z. Wang, L. Zou, L. Xu, and F. E. Alsaadi, "A new approach to smooth global path planning of mobile robots with kinematic constraints," *International Journal of Machine Learning and Cybernetics*, vol. 10, no. 1, pp. 107–119, 2017.
- [16] W. Hao, W. Jun, L. Wei, and Y. Zhen-tao, "Structure design and control of mass of fog monitoring robot installed on the highway guardrail based on ESP32," in *2020 39th Chinese Control Conference (CCC)*. IEEE, 2020.
- [17] T. Xiao-Ling, "Trajectory stability control of the lidar tracking robot based on improved particle swarm filter algorithm," *Journal of Physics: Conference Series*, vol. 1168, p. 022107, 2019.
- [18] B. Song, Z. Wang, and L. Zou, "An improved PSO algorithm for smooth path planning of mobile robots using continuous high-degree bezier curve," *Applied Soft Computing*, vol. 100, no. 1, p. 106960, 2021.
- [19] F. Martínez, F. Martínez, and H. Montiel, "Hybrid free-obstacle path planning algorithm using image processing and geometric techniques," *ARPN Journal of Engineering and Applied Sciences*, vol. 14, no. 18, pp. 3135–3139, 2019.
- [20] E. Krell, A. Sheta, A. P. R. Balasubramanian, and S. A. King, "Collision-free autonomous robot navigation in unknown environments utilizing PSO for path planning," *Journal of Artificial Intelligence and Soft Computing Research*, vol. 9, no. 4, pp. 267–282, 2019.
- [21] C. LIU, G.-H. CAO, Y.-Y. QU, and Y.-M. CHENG, "An improved PSO algorithm for time-optimal trajectory planning of delta robot in intelligent packaging," *The International Journal of Advanced Manufacturing Technology*, vol. 107, no. 3–4, pp. 1091–1099, 2019.

- [22] F. A. Raheem and U. I. Hameed, "Interactive heuristic d* path planning solution based on PSO for two-link robotic arm in dynamic environment," *World Journal of Engineering and Technology*, vol. 07, no. 01, pp. 80–99, 2019.
- [23] M. N. A. Wahab, S. Nefti-Meziani, and A. Atyabi, "A comparative review on mobile robot path planning: Classical or meta-heuristic methods?" *Annual Reviews in Control*, vol. 50, no. 1, pp. 233–252, 2020.
- [24] J. Batista, D. Souza, J. Silva, K. Ramos, J. Costa, L. dos Reis, and A. Braga, "Trajectory planning using artificial potential fields with metaheuristics," *IEEE Latin America Transactions*, vol. 18, no. 05, pp. 914–922, 2020.
- [25] L. Yan, W. Xu, Z. Hu, and B. Liang, "Multi-objective configuration optimization for coordinated capture of dual-arm space robot," *Acta Astronautica*, vol. 167, no. 1, pp. 189–200, 2020.
- [26] M. H. Barakat, H. H. Ammar, and M. Elsamanty, "Experimental path tracking optimization and control of a nonlinear skid steering tracked mobile robot," in *2020 2nd Novel Intelligent and Leading Emerging Sciences Conference (NILES)*. IEEE, 2020.
- [27] Z. Chen, X. Xu, W. Zha, A. Rodic, and P. B. Petrovic, "Motion planning of 7-DOF manipulator based on quintic b-spline curve," in *2021 6th IEEE International Conference on Advanced Robotics and Mechatronics (ICARM)*. IEEE, 2021.
- [28] M. Liu, R. Lin, M. Yang, A. V. Nazarova, and J. Huo, "Active disturbance rejection motion control of spherical robot with parameter tuning," *Industrial Robot: the international journal of robotics research and application*, vol. 49, no. 2, pp. 332–343, 2021.

Software Effort Estimation using Machine Learning Technique

Mizanur Rahman¹

Faculty of Computing

Universiti Malaysia Pahang

26600, Pekan, Pahang, Malaysia

Partha Protim Roy²

Research Assistant

United International University

Dhaka, Bangladesh

Mohammad Ali³

Developer

Edusoft Consultants Ltd

Dhaka, Bangladesh

Teresa Gonçalves⁴

Associate Professor

Department of Computer Science

University of Évora, Portugal

Hasan Sarwar⁵

Professor

Department of Computer Science and Engineering

United International University

Satarkul, Badda, Dhaka, Bangladesh

Abstract—Software engineering effort estimation plays a significant role in managing project cost, quality, and time and creating software. Researchers have been paying close attention to software estimation during the past few decades, and a great amount of work has been done utilizing a variety of machine-learning techniques and algorithms. In order to better effectively evaluate predictions, this study recommends various machine learning algorithms for estimating, including k-nearest neighbor regression, support vector regression, and decision trees. These methods are now used by the software development industry for software estimating with the goal of overcoming the limitations of parametric and conventional estimation techniques and advancing projects. Our dataset, which was created by a software company called Edusoft Consulted LTD, was used to assess the effectiveness of the established method. The three commonly used performance evaluation measures, mean absolute error (MAE), mean squared error (MSE), and R square error, represent the base for these. Comparative experimental results demonstrate that decision trees perform better at predicting effort than other techniques.

Keywords—Software effort estimation; K-nearest neighbor regression; machine learning; decision tree; support vector regression

LIST OF ABBREVIATION

ML	Machine Learning	KNN	K-nearest neighbour
SVR	Support Vector Regression	DT	Decision Tree
ANN	Artificial Neural Network	RF	Random Forest
CBR	Case Base Reasoning	LR	Linear regression
MSE	Mean Square Error	MAE	Mean Absolute Error
RMSE	Root Mean Square Error	SEE	Software Effort estimation
COCOMO	Constructive Cost Model	LASSO	Least Absolute Shrinkage and Selection

I. INTRODUCTION

Software teams and businesses have long had substantial difficulty with software effort estimation, which should be taken into account at the outset of a software project[1]. For software project success and risk reduction, accurate software work estimation is essential. The practice of estimating the amount of time and money it will take to produce a software process or product is known as effort estimation. In order to effectively budget, plan, control, and manage the project, proper estimating may require accurately projecting software

costs. In order to allocate resources effectively and prepare for the development of software, precise cost and time estimation are crucial. Project planning determines whether a project will succeed or fail because during this stage, the time and financial restrictions necessary to finish the project successfully are estimated [2]. Since the 1940s, when the computer industry first started to take off, the idea of software effort estimation has gained popularity. Research in this area is still ongoing [3].

Numerous estimation methods are categorized into three broad groups in the literature on software effort estimation: algorithmic, non-algorithmic, and machine learning [4]. Algorithmic approaches utilize statistical and mathematical concepts for software project estimation. COCOMO-II, Putnam Software Life cycle Management (SLIM), SEER-SEM, and True Planning are some examples of estimate techniques. The fundamental input to these models is the size of the software being evaluated, It is usually quantified in a metric like function points, source lines of code, or use case points. Models that don't rely on algorithms rely on subjective evaluations and interpretations of data. Data from prior projects are analyzed by these models. Expert judgment, planning poker, wide-band Delphi, and the work breakdown structure (WBS) are all examples of non-algorithmic approaches. Machine learning is an alternative to algorithmic model building. Artificial neural networks (ANN), case-based reasoning (CBR), support vector regression (SVR), decision trees (DT), fuzzy models, Bayesian networks, and genetic algorithms are all examples of machine learning estimation approaches [5].

Several datasets have been proposed that are used to measure the effort estimation of software development. These datasets were proposed quite a long time ago. At present the effort engagement in software development has changed a lot. One perspective is that in this post-COVID era, many companies have moved to a hybrid development approach. In hybrid mode, developers are not bound to come to the office physically every day. They can work from home. The office visit is expected only if there is a need for so. Another perspective is that an extreme level of change requirements

is needed due to the dynamic nature of clients' businesses, updates to new technology, and extreme competition in the business. Considering these two scenarios, In this study, we have developed a new dataset and we have successfully shown that this new dataset is capable of providing future insight into software efforts. We have used three machine learning models, namely, SVM, KNN, and DT. We have shown that both these models are able to predict the software effort for the future. We claim that this is the first instance of using such a dataset in the prediction of effort in software development. We have also shown that these new parameters are perfectly able to predict the future.

Here is how the article progresses for the remainder. In section II, we review the many methods of effort estimation that have been published by scholars. The section III explains how we conducted our research. In section IV, we detail the performance evaluation of the several competing ML approaches we employed based on the trial outcomes. Thereafter, the essay finishes with Section VI.

II. RELATED WORK

The majority of software effort estimates are made using a variety of approaches that have been put out by scholars over the past few decades. The first three methods to calculate the effort required to create widely used software were the function point-based model, the constructive cost model (COCOMO), and the Putnam program life cycle model (SLIM), and they each provided a specific formula for calculating the effort required from historical data[6][7]. Machine learning (ML) algorithms have been widely utilized to estimate the software development effort [8]. These ML Algorithms allow professionals to devote more time to other client-pleasing aspects of software systems and less time to analyzing new projects. Many scholars have utilized machine learning to estimate effort over the past ten years.

A systematic study to investigate the use of ML models in predicting Software Development Life Cycle (SDLC) effort reports CBR – Case Based Reasoning, Neural Network, and Decision Tree as the most frequently used algorithms [9]. In another report, the datasets that were considered are Albrecht, China, Desharnais, Kemerer, Kitchenham, Maxwell, and Cocomo81. Here various stacking models were used. These are stacking using generalized linear models (S-GLM), stacking using S-DT, stacking using SVM, and stacking using RF [10]. A survey of 35 selected studies on effort estimation accuracy implemented on both public and non-public domain datasets suggests ensemble effort estimation as a better technique than solo techniques [5]. The major goal of this research [11] is to empirically compare the performance of several Machine Learning (ML) algorithms in order to identify a performance model for evaluating the software effort. Seven datasets have been used for Effort Estimation, and various ML approaches have been applied. The LASSO approach with the China dataset gave the best performance when compared to the other algorithms, according to the results and trials with several ML algorithms for software effort estimation.

To determine the precise software effort, Abdelali et al. [12] constructed an RF model and experimentally adjusted the effectiveness by altering the important parameters. Specifically,

ISBSG, Tukutuku, and COCOMO datasets were employed. The 30% hold-out validation approach was used to manage the evaluation. Three performance indicators, such as the MdmRE, MMRE, and Pred (0.25), are used to assess and identify the well-performed technique. When the generated RF model was compared to the traditional regression tree, it was clear that the upgraded RF model outperformed it. Nassif et al. [13] experiment with fuzzy models to estimate software effort. In order to compare and implement three fuzzy logic models, namely Sugeno, Mamdani with constant output, and Sugeno with linear output, regression analysis was done. These models were used to estimate the program effort. The ISBSG dataset was used for model training and testing, and the Scott-Knott, mean inverted balance relative error, and other performance measures were used to assess the models' effectiveness. The Sugeno fuzzy model with linear output performed better when compared to other fuzzy models created to help in regression and analysis.

The soft computing techniques of linear regression (LR), SVR, ANN, RF, DT, and bagging methodology were used by Sharma and Vijayvargiya [14] to predict the time and resources required for software projects using the benchmark datasets. It was decided that the results from the RF and choice tree were the most helpful. The cost-benefit analysis led to enhanced cognitive performance. This strategy, however, was not evaluated using a deep learning classifier.

A gradient boosting regressor model was suggested by [15]. The stochastic gradient descent, KNN, DT, bagging, RF, AdaBoost, and gradient-boosting neighbors are all compared to the model. MSE, root mean square error (RMSE), and R2 were used to evaluate the model by authors. They displayed the outcomes using the China and Cocomo81 datasets.

The goal of this [16] work is to give an approach for accurately estimating the time and resources required to complete a software development project using only a subset of that organization's past project data. Two techniques, the correlation matrix, and the decision tree were utilized to determine the optimal prediction parameters. Each test's results were double-checked by using two different methods. The outcomes of the two analyses were identical, and the same three parameters were chosen for prediction. For certain variables, multiple prediction models were constructed and trained. According to the findings of the tests, Evolutionary SVM is the most accurate predictor.

The AdaBoost ensemble learning method and RF are used in this [17] study, while the Bayesian optimization method is used to calculate the model's hyperparameters. The SEE model was built using the PROMISE repository and the ISBSG dataset. Under 3-fold cross-validation, the created model was thoroughly compared with four machine learning approaches. The RF method based on AdaBoost ensemble learning and Bayesian optimization clearly outperforms this approach. Furthermore, the AdaBoost-based model assigns a feature relevance rating, making it a viable tool for predicting software effort.

III. METHODOLOGY

There has been extensive study into the use of machine learning (ML) based prediction methods for software devel-

opment effort estimation to improve predictions. The goal of this machine learning method is to minimize the loss function while simultaneously optimizing the support vector boundaries by transferring non-linear separable patterns in the input into higher feature space. Fig. 1 shows the methodology of software effort estimation. Three common machine-learning techniques are described below.

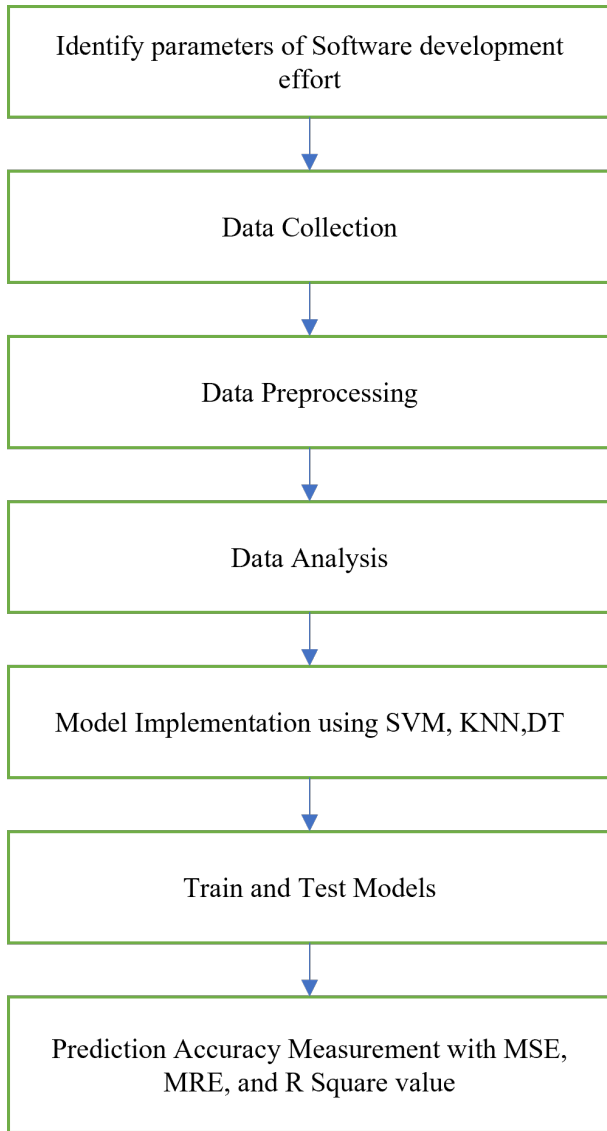


Fig. 1. Methodology of SEE.

A. Support Vector Regression (SVR)

A support vector machine (SVM) handles classification and regression issues. In machine learning, a classifier is a model built to make inferences about a class from additional features. The term “classification” refers to the act of labeling an unlabeled record with a unique value [18]. An SVM variant is an SVR. The regression issues are transformed into classification issues. A linear decision surface (known as a hyperplane) divides two segments of vectors in the SVM training process. The margin between the LDS and the vectors that are nearest

to it is maximum in this separation. They are known as support vectors. The input vectors are nonlinearly mapped into a high-dimensional space when the LDS is supposed to be non-linear. In this feature space, the LDS is built using characteristics that provide maximum margin and, thus, a low generalization error for the machine. This LDS has the best generalization of all speculative hyperplanes. Using this ideal hyperplane, classification or regression models’ predictive ability is improved.

$$f(\mathbf{x}) = \omega^T \phi(\mathbf{x}) + b \quad (1)$$

where \mathbf{x} is the input vector, $\phi(\mathbf{x})$ is the feature mapping function, ω is the weight vector, and b is the bias term.

B. K-Nearest Neighbor Regression (KNN)

KNN, one of the most straightforward estimation techniques, was chosen for this study because of its perceived resemblance to human-based expert opinion[19]. Technically, KNN does not train a model[20] but rather uses Euclidean distance [21] to calculate distances between locations. The algorithm makes a prediction about the class to which a point belongs by gathering the nearest samples. Regression involves taking the average of the closest samples to a location to determine its value. The effort of the target project is then estimated by averaging the efforts of the k projects that are the closest analogs.

$$\text{Euclidean} = \sqrt{\sum_{i=1}^k (x_i - y_i)^2} \quad (2)$$

Where k is the user-defined constant, i is the number of instances x and y are the vectors of each instance

C. Decision Tree (DT)

To obtain insightful information that will help it achieve its objectives, data mining uses DT. DT is an intelligent model that looks like a binary tree with the root at the top that has been turned on its side. The decision tree model is used to turn the data into a tree structure to help with the machine learning challenge. DT serves as an example of forecasting a dependent variable using a set of predictor variables. The decision-making process in this model is analogous to other models, which facilitates understanding. Because a picture is worth a thousand words, this approach makes it simple for anyone to comprehend the essence of a complex problem by simply looking at its schematic. The method used by DT is comparable to how people make decisions. DT does have certain disadvantages, though. Contrasted with other machine learning models, the accuracy of the dataset predictions is lower. Because DTs provide a collection of if-then-else rules, they are easier to grasp and analyze when compared to other machine learning techniques like neural networks and Bayesian networks [10].

$$f(\mathbf{x}) = \sum_{j=1}^J c_j I(\mathbf{x} \in R_j) \quad (3)$$

where \mathbf{x} is the input vector, $f(\mathbf{x})$ is the predicted output label or value, J is the number of leaf nodes in the tree, R_j

TABLE I. SETTING PARAMETERS FOR ALL TECHNIQUES TAKEN INTO CONSIDERATION

Used method	Values for the parameters
SVR	kernel= 'rbf', degree=3, gamma='scale', cache_size=200
KNN	N_neighbors=5, weights='uniform', leaf_size=30,metric='minkowski'
DT	Criterion='gini' splitter='best', min_sample_split=2, min_samples_leaf=1

TABLE II. PROPERTIES OF DIFFERENT DATASETS

Dataset name	Source	No of attributes	Output attributes
Albrecht	PROMISE	8	Person-months
China	PROMISE	16	Person-hours
Desharnais	GITHUB	12	Person-hours
Kemerer	GITHUB	7	Person-months
Maxwell	PROMISE	27	Person-hours
Kitchenham	GITHUB	9	Person-hours
Cocomo81	GITHUB	17	Person-months
UCP	GITHUB	23	Person-months
ISBSG10	GITHUB	10	Person-months
Our dataset	GITHUB	7	Person-hours

is the region of the input space that is assigned to the j th leaf node, c_j is the output value or class label assigned to the j th leaf node, and $I(\cdot)$ is the indicator function that returns 1 if the input condition is true and 0 otherwise

IV. EXPERIMENTAL DESIGN

A. Dataset

In this research, we use a real-world dataset compiled by Edusoft Consultant Ltd. to conduct an empirical evaluation of the presented models for estimating the time and effort required for software development[22]. It's worth noting that this dataset includes quite a variety of characteristics, including task history ID, project ID, Client ID, task types, task priority, task overall state, total working time in hours, etc. our dataset consist of 2000 real-time data samples.

For estimating effort, a number of databases are utilized, including China, Kemerer, Cocomo81[23], Albrecht[24], Maxwell[25], Desharnais, and Kitchenham, Nasa93, ISBSF10. Along with our dataset, Table II shows repository information for other datasets, such as the number of characteristics, source, and an output unit for each dataset[26]. We won't consider all datasets in this research for performance evaluation, but we will evaluate the performance of our dataset.

B. Dataset Pre-Processing

The data preprocessing methodology is an efficient method for estimating the amount of work that will be required [27], and it is an essential step in the process of enhancing the performance of machine learning [28]. The first stage is to eliminate irrelevant features from the dataset as stated in [29]. If unnecessary features are taken out of machine learning algorithms, they will function better. Categorical data undergoes further processing to become numerical. Each category is given its own numeric code in ordinal coding, which has the benefit of not expanding the problem space by displaying each category as a separate input [30].

After the data collection, the dataset has been pre-processed to eliminate inconsistencies, duplicates in data, or missing values, which can otherwise negatively affect a model's accuracy. The researchers have used different methods to achieve the qualitative dataset [31]. The collected dataset also had some missing values, and some missing data were eliminated by inspecting the dataset, and features were chosen according to the degree of correlation between each dataset's values. Furthermore, some of the missing information has been filled in using the moving median method with a window of the length of 10.

C. Performance Evaluation

Research has demonstrated the capabilities of software development effort estimation models using a wide range of performance indicators. Different aspects of performance are being measured and/or represented by these different metrics. Performance evaluation measures are crucial to the accuracy of performance measurements [32]. However, no single metric has gained widespread acceptance for use across all software development effort estimation model comparisons without some form of criticism. we employ more generic evaluation metrics such as Mean Absolute Error (MAE) (2), Mean Square Error (MSE) (1), and R Squared (3).

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (4)$$

where y_i is the true value of the i th data point, \hat{y}_i is the predicted value for the i th data point, n is the total number of data points, and $(\cdot)^2$ denotes squaring. The MSE measures the average of the squared differences between the predicted values and the true values, which gives greater weight to larger differences than smaller ones. This means that the MSE is more sensitive to large errors than to small errors. A lower MSE indicates that the model is better at predicting the true values, while a higher MSE indicates that the model is less accurate.

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (5)$$

where y_i is the true value of the i th data point, \hat{y}_i is the predicted value for the i th data point, n is the total number of data points, and $|\cdot|$ denotes the absolute value. The MAE is a useful metric for evaluating regression models, as it gives an idea of how far off the predictions are on average. A lower MAE indicates that the model is better at predicting the true values, while a higher MAE indicates that the model is less accurate

$$R^2 = 1 - \frac{SS_{\text{Regression}}}{SS_{\text{Total}}} = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2} \quad (6)$$

where y_i is the true value of the i th data point, \hat{y}_i is the predicted value for the i th data point, \bar{y} is the mean of the true values, and n is the total number of data points. A higher R^2 value indicates that the model explains more of the variance in the dependent variable. However, a high R^2 value does not necessarily indicate that the model is a good

fit for the data, as it may be overfitting the data or failing to capture important relationships between the independent and dependent variables.

D. Parameter Setting

In Table I, a comprehensive list of all parameter values was examined for each SDEE technique in this study.

V. EXPERIMENTAL RESULTS AND DISCUSSION

In this study, machine learning algorithms have been conducted with our proposed dataset. Firstly, the preprocessed dataset has been separated into a training and a testing set. The training set has been used to train the models, and the testing set has been used to test the performance of the models with our dataset. In this experiment, we have used a total of 80% dataset for training, and the rest 20% dataset for testing the models. The parameter values that have been examined for each SDEE method in this study are all listed in Table I.

Table III shows the performance metrics (MAE, MSE, and R-Squared) for ML algorithms in predicting our dataset’s data. Table III illustrates that the Decision Tree Algorithm yields more accurate outcomes with smaller MAE and MSE values. Besides, based on the MSE value second best result was given by the KNN algorithm. Finally, the third position is held by SVR based on the performance evaluation matrix. Fig. 2 shows the actual vs predicted result using SVM, Fig. 3 shows the actual vs predicted result using DT, and finally, Fig. 4 show the predicted results plot with respect to the actual data using KNN. As we proposed a new dataset in this work so we did not compare it with other datasets, we will compare the result of our dataset with other existing datasets using various ML algorithms in our next paper.

TABLE III. COMPARISON OF ALL MODELS’ EFFECTIVENESS IN TERMS OF MSE, MAE AND R SQUARE

Model	MSE	MAE	R Square
Decision Tree	20.54043	1.649842	-0.008089
SVM	148.4454	4.63788	-0.059185
KNN	135.3942	5.42981	0.033978

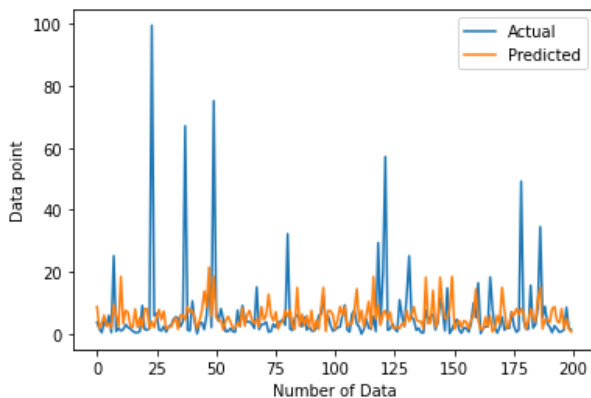


Fig. 2. SVM models’ actual effort and predicted effort.

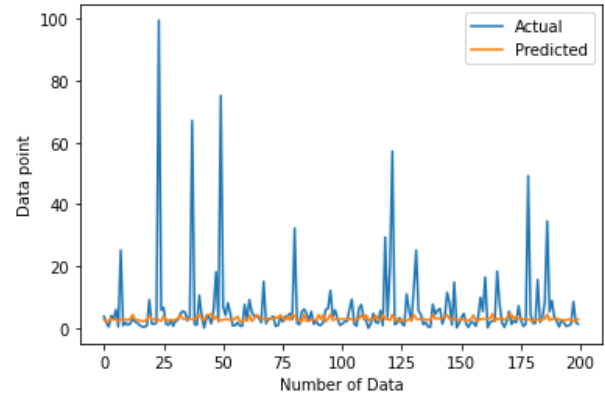


Fig. 3. DT models’ actual effort and predicted effort.

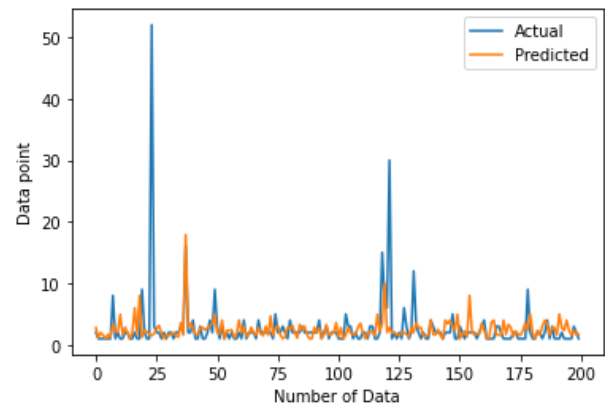


Fig. 4. KNN models’ actual effort and predicted effort.

VI. CONCLUSION

At the beginning of software development, the project manager must take care of an extremely important step called effort estimation. To estimate effort, we used a real dataset created by Edusoft Consultant Ltd. In this study, we implemented the decision tree regressor, random forest, and KNN machine learning models. Comparative experimental results demonstrate that decision trees perform better at predicting effort than other techniques. The MAE, MSE, and R-Squared values are evaluated as evaluation metrics. In the future, we will compare the result of our dataset with other existing datasets using various ML algorithms.

ACKNOWLEDGMENT

The work reported in this paper is funded by the Institute for Advanced Research (IAR), United International University (UIU), Bangladesh titled: Implementation of 3D model to assess the performance improvement of a software company UIU/IAR/02/2019-20/SE/06.

REFERENCES

[1] Z. R. Mohsin, “Comparative study for software effort estimation by soft computing models,” *Journal of Education for Pure Science-University of Thi-Qar*, vol. 11, no. 2, pp. 108–120, 2021.

- [2] A. Idri and I. Abnane, "Fuzzy analogy based effort estimation: An empirical comparative study," in *2017 IEEE International Conference on Computer and Information Technology (CIT)*. IEEE, 2017, pp. 114–121.
- [3] A. Zaid, M. H. Selamat, A. Ghani, R. Atan, and K. Wei, "Issues in software cost estimation," *International Journal of Computer Science and Network Security*, vol. 8, no. 11, pp. 350–356, 2008.
- [4] T. Vera, S. F. Ochoa, and D. Perovich, "Survey of software development effort estimation taxonomies," *Computer Science Department, University of Chile: Santiago, Chile*, 2017.
- [5] Y. Mahmood, N. Kama, A. Azmi, A. S. Khan, and M. Ali, "Software effort estimation accuracy prediction of machine learning techniques: A systematic performance evaluation," *Software: Practice and Experience*, vol. 52, no. 1, pp. 39–65, 2022.
- [6] A. Najm, A. Zakrani, and A. Marzak, "Systematic review study of decision trees based software development effort estimation," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 7, 2020.
- [7] A. Ali and C. Gravino, "A systematic literature review of software effort prediction using machine learning methods," *Journal of software: evolution and process*, vol. 31, no. 10, p. e2211, 2019.
- [8] Y. Garg *et al.*, "Comparative analysis of machine learning techniques in effort estimation," in *2022 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COM-IT-CON)*, vol. 1. IEEE, 2022, pp. 401–405.
- [9] S. S. Gautam and V. Singh, "Adaptive discretization using golden section to aid outlier detection for software development effort estimation," *IEEE Access*, vol. 10, pp. 90 369–90 387, 2022.
- [10] P. V. AG and V. Varadarajan, "Estimating software development efforts using a random forest-based stacked ensemble approach," *Electronics*, vol. 10, no. 10, p. 1195, 2021.
- [11] F. B. Ahmad and L. M. Ibrahim, "Software development effort estimation techniques: A survey," *development*, vol. 2, p. 23.
- [12] H. Mustapha, N. Abdelwahed *et al.*, "Investigating the use of random forest in software effort estimation," *Procedia computer science*, vol. 148, pp. 343–352, 2019.
- [13] A. B. Nassif, M. Azzeh, A. Idri, and A. Abran, "Software development effort estimation using regression fuzzy models," *Computational intelligence and neuroscience*, vol. 2019, 2019.
- [14] S. Sharma and S. Vijayvargiya, "Applying soft computing techniques for software project effort estimation modelling," in *Nanoelectronics, Circuits and Communication Systems: Proceeding of NCCS 2019*. Springer, 2021, pp. 211–227.
- [15] P. Suresh Kumar, H. Behera, J. Nayak, and B. Naik, "A pragmatic ensemble learning approach for effective software effort estimation," *Innovations in Systems and Software Engineering*, vol. 18, no. 2, pp. 283–299, 2022.
- [16] T. Mahboob, S. Gull, S. Ehsan, and B. Sikandar, "Predictive approach towards software effort estimation using evolutionary support vector machine," *International journal of advanced computer science and applications*, vol. 8, no. 5, 2017.
- [17] R. Marco, S. S. S. Ahmad, and S. Ahmad, "Bayesian hyperparameter optimization and ensemble learning for machine learning models on software effort estimation," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 3, 2022.
- [18] O. A. Montesinos López, A. Montesinos López, and J. Crossa, "Support vector machines and support vector regression," in *Multivariate Statistical Machine Learning Methods for Genomic Prediction*. Springer, 2022, pp. 337–378.
- [19] V. Resmi and S. Vijayalakshmi, "Analogy-based approaches to improve software project effort estimation accuracy," *Journal of Intelligent Systems*, vol. 29, no. 1, pp. 1468–1479, 2019.
- [20] C. Albon, *Machine learning with python cookbook: Practical solutions from preprocessing to deep learning*. O'Reilly Media, Inc., 2018.
- [21] A. C. Faul, *A concise introduction to machine learning*. CRC Press, 2019.
- [22] E. C. Ltd, "Software effort estimation," https://github.com/edusoftresearch/SEE_Data, 2023.
- [23] B. Boehm, *Software Engineering Economics*. Prentice Hall, 1981.
- [24] Albrecht, "Software function, source lines of code, and development effort prediction: A software engineering," <https://zenodo.org/record/268467#.ZB5-BHZBxPY>, 1983.
- [25] K. D. Maxwell, "Maxwell dataset," <https://zenodo.org/record/268461#.ZBBPX3ZBzIU>, 2002.
- [26] J. Sayyad Shirabad and T. Menzies, "The PROMISE Repository of Software Engineering Databases." School of Information Technology and Engineering, University of Ottawa, Canada, 2005. [Online]. Available: <http://promise.site.uottawa.ca/SERepository>
- [27] J. Huang, Y.-F. Li, J. W. Keung, Y. T. Yu, and W. Chan, "An empirical analysis of three-stage data-preprocessing for analogy-based software effort estimation on the isbgs data," in *2017 IEEE International Conference on Software Quality, Reliability and Security (QRS)*. IEEE, 2017, pp. 442–449.
- [28] J. Huang, Y.-F. Li, and M. Xie, "An empirical analysis of data preprocessing for machine learning-based software cost estimation," *Information and software Technology*, vol. 67, pp. 108–127, 2015.
- [29] E. T. Bekar, P. Nyqvist, and A. Skoogh, "An intelligent approach for data pre-processing and analysis in predictive maintenance with an industrial case study," *Advances in Mechanical Engineering*, vol. 12, no. 5, p. 1687814020919207, 2020.
- [30] E. Fitkov-Norris, S. Vahid, and C. Hand, "Evaluating the impact of categorical data encoding and scaling on neural network classification performance: The case of repeat consumption of identical cultural goods," in *Engineering Applications of Neural Networks: 13th International Conference, EANN 2012, London, UK, September 20-23, 2012. Proceedings 13*. Springer, 2012, pp. 343–352.
- [31] A. Ahmed, S. Elkhatatny, A. Ali, M. Abughaban, and A. Abdulaheem, "Application of artificial intelligence techniques in predicting the lost circulation zones using drilling sensors," *Journal of Sensors*, vol. 2020, 2020.
- [32] A. Jadhav, M. Kaur, and F. Akter, "Evolution of software development effort and cost estimation techniques: five decades study using automated text mining approach," *Mathematical Problems in Engineering*, vol. 2022, pp. 1–17, 2022.

An Approach to Hyperparameter Tuning in Transfer Learning for Driver Drowsiness Detection Based on Bayesian Optimization and Random Search

Hoang-Tu Vo, Hoang Tran Ngoc, Luyi-Da Quach
Software Engineering Department
FPT University, Cantho City, Vietnam

Abstract—Driver drowsiness is a critical factor in road safety, and developing accurate models for detecting it is essential. Transfer learning has been shown to be an effective technique for driver drowsiness detection, as it enables models to leverage large, pre-existing datasets. However, the optimization of hyperparameters in transfer learning models can be challenging, as it involves a large search space. The core purpose of this research is on presenting an approach to hyperparameter tuning in transfer learning for driving fatigue detection based on Bayesian optimization and Random search algorithms. We examine the efficiency of our approach on a publicly available dataset using transfer learning models with the MobileNetV2, Xception, and VGG19 architectures. We explore the impact of hyperparameters such as dropout rate, activation function, the number of units (the number of dense nodes), optimizer, and learning rate on the transfer learning models' overall performance. Our experiments show that our approach improves the performance of the transfer learning models, obtaining cutting-edge results on the dataset for all three architectures. We also compare the efficiency of Bayesian optimization and Random search algorithms in terms of their ability to find optimal hyperparameters and indicate that Bayesian optimization is more efficient in finding optimal hyperparameters than Random search. The results of our study provide insights into the importance of hyperparameter tuning for transfer learning-based driver drowsiness detection using different transfer learning models and can guide the selection of hyperparameters and models for future studies in this field. Our proposed approach can be applied to other transfer learning tasks, making it a valuable contribution to the field of ML.

Keywords—Hyperparameter tuning; driver drowsiness detection; transfer learning; Bayesian optimization; Random search

I. INTRODUCTION

ML models for detecting driver fatigue have shown promise in accurately identifying and alerting drowsy drivers. Transfer learning is a popular approach in developing such models, as it allows leveraging pre-existing datasets to improve model performance. However, optimizing hyperparameters for transfer learning models can be a daunting task given the vast search space involved. In fact, compared to its default value, the model performs far better when the appropriate hyperparameter is used. The hyperparameter changes based on the data collection. The study of hyperparameter optimization for well-known ML models is presented in article [30]. In order to build an effective drowsiness detection system, it is crucial to choose appropriate hyperparameters for the ML model used for this task. Some critical hyperparameters for the drowsiness

detection system include dropout rate, activation functions, units (the number of dense nodes), optimizer, and learning rate.

Dropout [12]: In order to prevent overfitting, Dropout is a regularization method that functions by randomly dropping a certain proportion of neurons while training. This reduces overfitting and enhances the model's capacity to make accurate predictions on new data.

Activation function [15]: The activation function is used to provide non-linearity features into the model. Non-linear activation functions are necessary for deep learning (DL) models as they allow the model to learn complex relationships between the input and output data. Common activation functions include Adam, RMSprop, ReLU and tanh. Selecting an activation function will depend on the particular task and the kind of data in use.

Units: The number of dense nodes in a neural network layer is a hyperparameter. Increasing the number of units in a layer increases the model's capacity to learn but also raises the possibility of overfitting. A careful balance must be struck between the number of units and the regularization techniques used to prevent overfitting.

Optimizer [25][34][26]: By utilizing optimization algorithms, researchers can improve the effectiveness of a CNN model for a specific task. The optimizer is used to update the model's weights during training. Common optimizers include SGD, Adam, and RMSprop. When selecting an optimizer, it is important to consider both the task to be performed and the nature of the data.

Learning rate [18]: The learning rate is a crucial factor in determining the size of the weight updates that take place during training. A high learning rate will result in rapid convergence but may also result in the optimizer overshooting the optimal solution. A low learning rate will converge slowly but is less likely to overshoot the optimal solution. The learning rate must be carefully chosen to ensure that the optimizer converges to an optimal solution.

The main contribution of the study is to introduce a method for optimizing hyperparameters in transfer learning for detecting driving fatigue. The approach utilizes both Bayesian optimization and Random search algorithms. The publicly available dataset was used to assess the efficacy of the

suggested approach, which was applied to different transfer learning models.

II. RELATED WORKS

Driving when fatigued is a serious safety risk, with potentially disastrous consequences. Various methods have been developed for driver drowsiness detection to prevent accidents caused by sleepy or fatigued drivers. One such method is the use of sensors that can detect changes in Physiological-based measures, Physiological-based measures refer to the use of physical signals, to detect changes that indicate drowsiness, such as steering wheel movements [3]. Study [14] presents a low-cost ECG sensor designed for drowsiness detection in drivers. The sensor provides good results in extracting ECG parameters and is used in combination with facial recognition for improved detection in unfavorable conditions. [24] In this study, ECG signals were employed to detect and analyze a driver's condition, with 13 features extracted and trained through SVM, KNN, and Ensemble classifiers. The findings demonstrate high accuracy levels (ranging from 93.1% to 100%) in two-class identification, but a reduced accuracy rate of 58.3% in five-class detection. The study [16] evaluated the efficiency of in-ear EEG in detecting alertness-drowsiness in drivers and compared it with three peripheral signals - ECG, PPG, and GSR. A review article explores various methods that use Electroencephalogram (EEG) signals for detecting and managing driver drowsiness [22]. The paper [23] describes a wearable EEG device that is placed on the driver's forehead and can automatically detect the driver's mental state. [5] This paper proposes a new DL architecture that automatically detects sleepiness from single-channel EEG data using a CNN. Behavioral-based measures refer to the use of observable behaviors and actions to detect drowsiness, such as, The study [29] proposes a method to detect drowsiness in real-time security camera footage by analyzing if someone's eyes are open or closed. This involves identifying the person's face and eyes in the image and applying an extended Sobel operator to detect the shape of the eyelids' curvature. Based on the concavity of the curves, the technique determines whether an individual's eyes are open or closed. The paper proposes a method for detecting driver drowsiness based on eye blink detection. The system detects the eyes using facial landmark extraction and measures the Eye Aspect Ratio to detect blinks. The total number of blinks per minute is compared to a standard value to identify whether the driver is drowsy [17]. A new method [2] for detecting mental fatigue and drowsiness in drivers by using the XSENS motion capture system to analyze head posture movements. A deep neural network using LSTM architecture is utilized to classify driver states based on three dimensions time-series head angular acceleration information, achieving high accuracy with an overall training accuracy of 99.2%. A new system is proposed in [32] for detecting the condition of a driver's eyes and mouth using Retinaface and the residual channel attention network. Results showed high accuracy of 98.962% for eye state classification and 98.561% for mouth state classification on a dataset developed for this study. [4] Describes a drowsiness identification model that integrates face and head pose detection using Dlib models. Recognition of driver fatigue through detecting yawns [20][19][31]. In this work [1][10] a driver drowsiness detection system in real-time that uses facial landmarks and dlib to detect eye

movement and yawning. [21] The ratios for eye aspect, mouth opening, and nose length are determined through a process of adaptive thresholding. The study [13] proposes a drowsiness detection method that combines yawning and closing of the eyes using mouth and eye aspect ratios and head pose analysis through optical flow. Two classification techniques, multilayer perception, and K-nearest neighbors were investigated for the prediction of drowsiness. The project [7] detects facial features from a live stream of a driver and then localizes the eyes that are either open or closed. In [31] uses two modules to detect driver fatigue by identifying the condition of the eyes and mouth. The mouth area is identified through depth information while a semi-VGG architecture using CNN networks is employed to detect open or closed eyes. The findings from both detections are integrated and achieved an accuracy of around 90%. The study [11] developed an image-based approach using YOLOv3 CNN and LSTM neural network for detecting driver drowsiness, which showed effectiveness in detecting yawning and blinking time periods with real-time experiments.

III. TRANSFER LEARNING MODELS

Transfer learning is a machine learning (ML) technique to enhance the learning process by utilizing a pre-trained model for a new task. Transfer learning models are neural networks that have been trained on a large dataset, typically for image recognition or natural language processing tasks, and then adapted for a new task with a smaller dataset. The basic idea behind transfer learning is that a model that has learned to recognize certain features in one domain can be re-purposed to recognize similar features in another domain. There are several approaches to implementing transfer learning models, but they generally involve modifying the pretrained model to suit the new task. This might involve adding new layers to the model, changing the total of neurons in the existing layers, or fine-tuning the weights of the existing layers. One of the most evident advantages of transfer learning is that it can considerably minimize the volume of data needed to train a fresh model since many of the relevant characteristics have already been learned by the pre-existing model. This can be particularly useful for tasks where there is limited labeled data available. Transfer learning can be advantageous in that it helps to enhance the performance of a model since the pretrained model has already learned to recognize a variety of crucial features in the input data. Examples of transfer learning in various fields include Computer Vision: Pretrained models can be fine-tuned on a smaller dataset of medical images to detect specific types of tumors in X-ray images. Natural Language Processing: Pretrained models can be further adjusted on a smaller dataset of news articles to detect fake news. Speech Recognition: Pretrained models for speech recognition can be fine-tuned on a smaller dataset of spoken commands to control a smart home device.

The study aims to use is to use Bayesian Optimization and Random Search to find hyperparameters for the Transfer Learning model, including VGG19 [28], MobileNetV2 [27] and Xception [8]. The experiments carried out for this research are listed in Table I and the search space for hyperparameters is given in Table II.

TABLE I. THE TECHNIQUE INVOLVES LEVERAGING A PRE-TRAINED NETWORK FOR FEATURE EXTRACTION AND THEN TUNING THE MODEL THROUGH FINE-TUNING

Models	VGG19	Xception	MobileNetV2
Usage of pre-trained	The technique involves leveraging a pre-trained network for feature extraction and then tuning the model through fine-tuning to improve its effectiveness		
Description	Remove the classification layer at the top of the pre-trained model. A new model head is constructed on top of the base model. This includes an AveragePooling2D layer. A Flatten layer to transform the output into a vector. A layer that is Dense and specifies the number of units and activation function. A Dropout layer to prevent overfitting. A Dense layer which is typically used in the final output layer of a neural network with a softmax activation function is added to output the probability distribution over four classes. During the training process, the weights of the base model layers are frozen to prevent updating, while only the weights of the new head model are updated		
Hyperparameter	Dropout rate, Activation function, Number of units, Optimizer, and Learning rate		

TABLE II. THE SEARCH SPACE FOR HYPERPARAMETERS

Hyper parameter	Search space
Dropout rate	values=[0.0, 0.1, 0.2, 0.3, 0.4, 0.5]
Activation function	values=['softplus', 'softmax', 'softsign', 'relu', 'hard_sigmoid', 'tanh', 'sigmoid']
Number of units	values=[32, 64, 128, 256, 512, 1024, 2048]
Optimizer	values=['Adadelta', 'Adagrad', 'Adam', 'RMSprop', 'SGD', 'Adamax', 'Nadam']
Learning rate	values=[1e-1, 1e-2, 1e-3, 1e-4, 1e-5]

IV. HYPERPARAMETER OPTIMIZATION

The main goal of hyperparameter optimization is to automate the process of hyperparameter tuning and make it more efficient. The overall objective of a hyper-parameter optimization problem is to achieve [30].

$$x^* = \arg \min_{x \in X} f(x) \quad (1)$$

Where $f(x)$ is the objective function; x^* is the configuration of hyperparameters that generates the optimal value for $f(x)$; the hyperparameter x can assume any element in the search space X .

A. Grid Search

Grid search [30] is a popular method used for exploring the configuration space of hyperparameters. This method involves evaluating all possible combinations of hyperparameters. Grid search is a straightforward approach that can be easily implemented and parallelized. However, it becomes less efficient when the number of hyperparameters increases, as the number of evaluations required to search the entire configuration space grows exponentially.

B. Random Search

Random search [30] [6] [33] is a simple and straightforward hyperparameter optimization method that is commonly used in ML. The approach involves randomly sampling hyperparameters from a specified search space, and evaluating the resulting models on a validation set to determine the performance of each set of hyperparameters.

One of the main advantages of Random search is its simplicity and ease of implementation. It requires minimal tuning and can be easily parallelized, allowing for efficient use of computational resources. Additionally, Random search can often outperform more sophisticated optimization methods for low-dimensional hyperparameter spaces. A significant

drawback of both Grid search and Random search is that each evaluation made throughout the search process is independent of any prior evaluations. Therefore, while selecting the next set of hyperparameters to assess, these approaches do not consider the outcomes of earlier evaluations.

C. Bayesian Optimization

Bayesian optimization [30][9] is a probabilistic approach to global optimization that is commonly used in ML, optimization, and experimental design. The goal of Bayesian optimization is to find the optimal set of hyperparameters that maximizes a given objective function, such as the accuracy of DL and ML models or the efficiency of a manufacturing process.

Several models, such as the Gaussian Process (GP), Random Forest (RF), and Tree-structured Parzen Estimators (TPE) models, can be used as the surrogate function in Bayesian optimization to simulate the distribution of the objective function, to approximate the unknown objective function. Iteratively updating the surrogate model with new observations of the objective function as the optimization process progresses, allowing for more accurate predictions of the optimal set of hyperparameters.

One of the main advantages of Bayesian optimization is its ability to efficiently optimize complex, high-dimensional, and non-convex functions with noisy and expensive evaluations. This makes it particularly useful for optimizing the hyperparameters of DL and ML models, which often involve complex and high-dimensional spaces. Furthermore, Bayesian optimization is able to handle constraints and multi-objective optimization problems.

V. METHODOLOGY

A. Dataset and Data Preparation

The dataset used in this study consists of a total of 3064 images, which were collected for the purpose of driver drowsiness

detection. The dataset is composed of four categories: eyes closed, eyes open, yawning, and no yawning. There are 726 images of eyes closed, which were obtained by capturing the eye region of subjects when their eyes were closed. Similarly, there are 726 images of Eye Open, which were obtained when subjects had their eyes open. The third category, Yawn, consists of 820 images, which were captured when subjects were yawning, and the fourth category, No Yawn, consists of 812 images, which were captured when subjects were not yawning. The dataset was partitioned into three distinct sets: training set, testing set, and validation set - in a random manner. The distribution of the dataset can be found in Fig. 1 and Fig. 2.

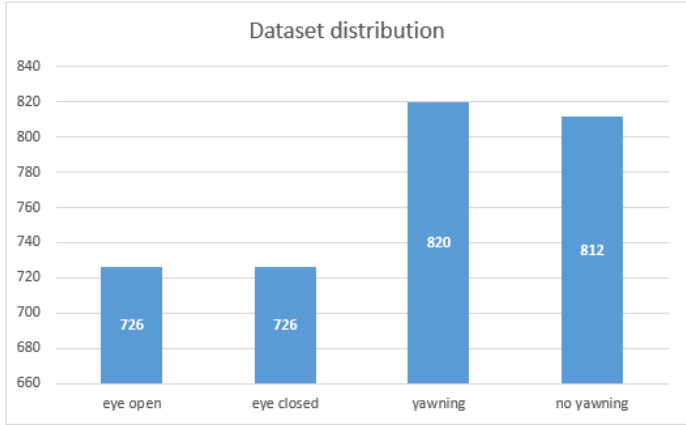


Fig. 1. A dataset distribution.

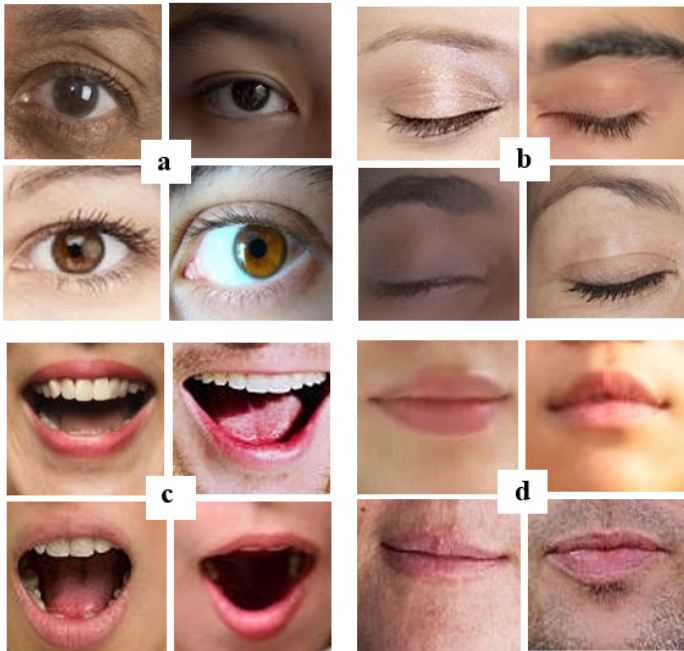


Fig. 2. Samples of drowsiness disease images. (a) Eye open, (b) Eye closed, (c) Yawn (d) No yawn.

B. Model Evaluation Metrics

In this research, the effectiveness of the DL models was examined using a variety of metrics. These metrics include

Precision, Recall, F1-score, and Accuracy. Accuracy was used to measure the overall effectiveness of the models in predicting the target variable. Precision was used to measure the proportion of true positive results among all the positive predictions is being calculated, while recall was used to measure the ratio of correctly predicted positive instances out of all actual positive instances in the dataset is being evaluated. F1-score, which combines both precision and recall, was used to provide a balanced view of the model's performance, especially in cases where the classes are imbalanced. By utilizing multiple evaluation metrics, were able to achieve an in-depth comprehension of the model's performance.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

$$F_1 - Score = \frac{Precision * Recall}{Precision + Recall} \quad (5)$$

In which, TP: True Positive, TN: True Negative, FP: False Positive, FN: False Negative.

C. Results

In order to train the DL model, the images were altered in size to either 224 x 224 for the VGG19, MobileNetV2, and Xception architectures. Removing the original fully connected Layer in MobileNetV2, VGG199, and Xception base network and changing them with new fully-connected Layers, then fine-tuning to a dataset for detecting drowsiness. The models of architecture utilized in this research are displayed in Fig. 3. During our experiments, we employed the Bayesian optimization and Random search techniques to adjust a set of five parameters that are fine-tuned within the neural networks. Specifically, we fine-tuned the dropout rate, activation function, units (number of dense nodes), optimizer, and learning rate.

Three models, VGG19, Xception, and MobileNetV2, were trained and tested on a given dataset. Bayesian optimization and Random search were employed to fine-tune the models. VGG19 achieved an accuracy of 0.97039 with Bayesian optimization, and 0.96875 with Random search. Xception obtained an accuracy of 0.98355 with both Bayesian optimization and Random search, while MobileNetV2 achieved the highest accuracy of 0.98684 with Bayesian optimization and 0.98190 with Random search. Overall, the results show that MobileNetV2 outperformed the other models with the highest accuracy in Bayesian optimization. The optimized hyperparameters found using Bayesian Optimization and Random Search for VGG19, Xception, and MobileNetV2 models are shown in Table III and is used to conduct various test scenarios described in Table I.

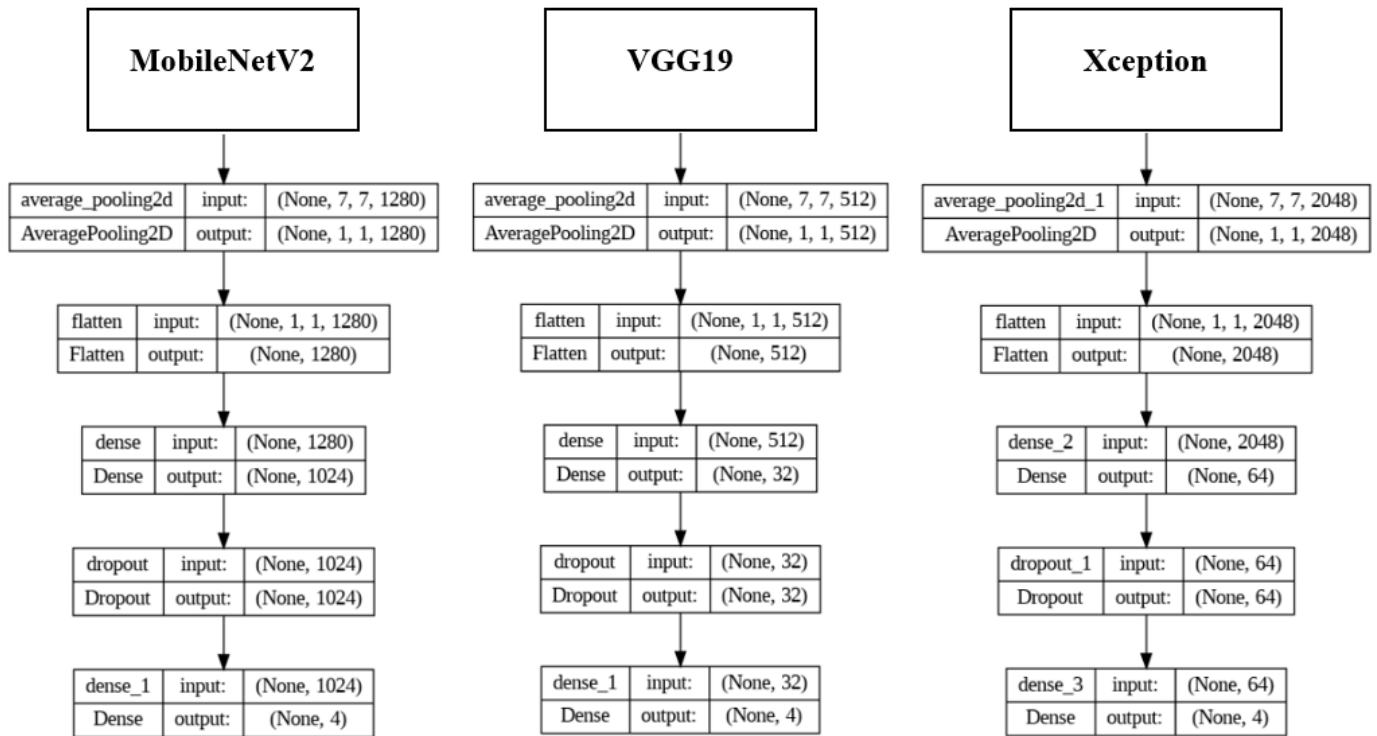


Fig. 3. Removing the original fully connected Layer in MobileNetV2, VGG199, and Xception base network and changing them with new fully connected Layers.

1) *VGG19*: Fig. 4 displays the optimized confusion matrices for the VGG19 model that have been optimized using Bayesian optimization and Random search techniques. The VGG19 model’s training and validation performance metrics including loss and accuracy are shown in Fig. 5. To get these results, Bayesian optimization and Random search strategies were also used. The precision, recall, F1-score, and accuracy were achieved during the evaluation of this model on each class using the best hyperparameters determined by Bayesian optimization and Random search presented in Table IV.

2) *Xception*: The optimized confusion matrices for the Xception model are shown in Fig. 6, which were produced using both Bayesian optimization and Random search methods. Additionally, using the above optimization techniques, Fig. 7 shows the Xception model’s performance during training and validation. The precision, recall, F1-score, and accuracy achieved during the evaluation of this model on each class using the best hyperparameters determined by Bayesian optimization and Random search are presented in Table V.

3) *MobileNetV2*: Both Bayesian optimization and Random search methods were utilized to create the optimized confusion matrices for the MobileNetV2 model, as presented in Fig. 8. Additionally, Fig. 9 showcases the performance during training and validation of the MobileNetV2 model, obtained through the application of the aforementioned optimization methods. The precision, recall, F1-score, and accuracy were achieved during the evaluation of this model on each class using the best hyperparameters determined by Bayesian optimization and Random search presented in Table VI.

The classification performance of the three optimized mod-

els is displayed in Fig. 10. In terms of precision, the MobileNetV2 model performs better than all the others. Compared to Random search, Bayesian optimization is more effective at obtaining optimal hyperparameters.

VI. CONCLUSION

In this study, we proposed an approach for hyperparameter tuning in transfer learning for driver drowsiness detection using Bayesian optimization and Random search algorithms. We evaluated our approach on a publicly available dataset using transfer learning models with the MobileNetV2, Xception, and VGG19 architectures. We explored the impact of various hyperparameters including dropout rate, activation function, number of units, optimizer, and learning rate on the performance of the transfer learning models.

Our experiments show that our approach improves the effectiveness of the transfer learning models, obtaining cutting-edge results on the dataset for all three architectures. We also compared the performances of Bayesian optimization and Random search algorithms in terms of their ability to find optimal hyperparameters and demonstrated that Bayesian optimization is more efficient in finding optimal hyperparameters than Random search.

Our findings highlight the importance of hyperparameter tuning in transfer learning-based driver drowsiness detection and provide insights into the impact of different hyperparameters on the effectiveness of the transfer learning models. Furthermore, our proposed approach can be applied to other transfer learning tasks, making it a valuable contribution to the field of ML.

TABLE III. HYPERPARAMETER VALUES FOUND USING BAYESIAN OPTIMIZATION AND RANDOM SEARCH

Models	Dropout rate	Activation function	Number of units	Optimizer	Learning rate	Accuracy
VGG19						
Bayesian Optimization	0.2	Elu	32	Adam	0.001	0.97039
Random Search	0.1	Tanh	1024	RMSprop	0.001	0.96875
Xception						
Bayesian Optimization	0.1	Elu	64	RMSprop	0.01	0.98355
Random Search	0.5	Tanh	128	Adam	0.001	0.98355
MobileNetV2						
Bayesian Optimization	0.1	Tanh	1024	RMSprop	0.001	0.98684
Random Search	0.5	Elu	1024	Adagrad	0.1	0.98190

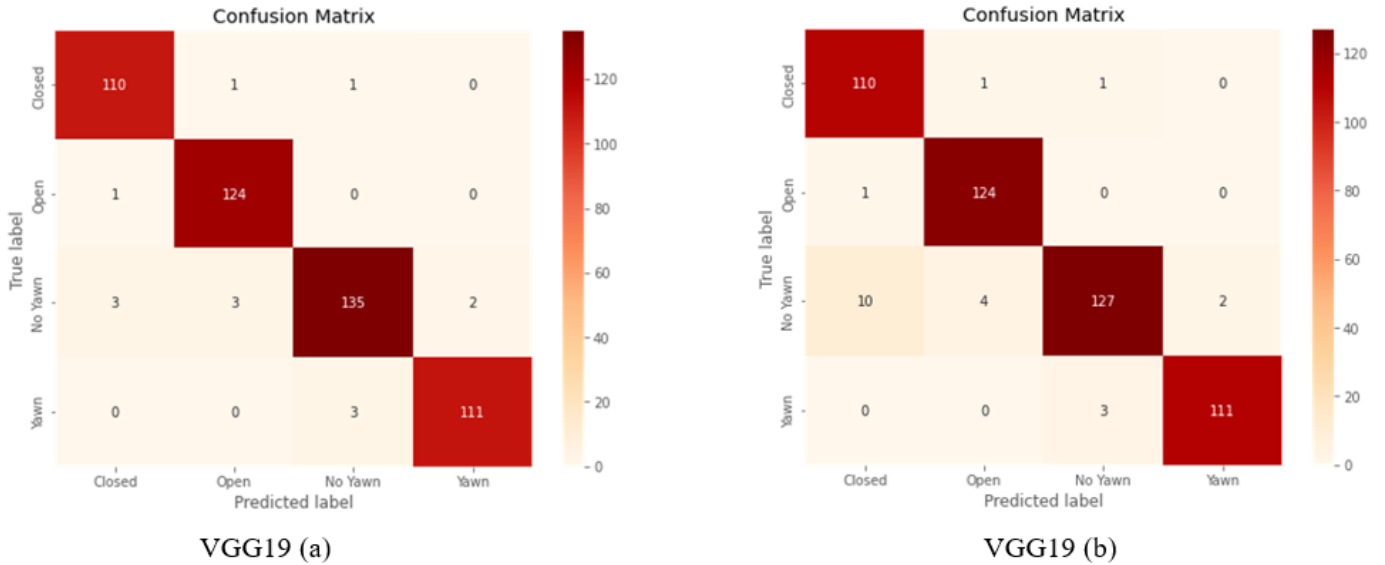


Fig. 4. Confusion matrix of VGG19. VGG19 (a) The ideal hyperparameter values are chosen through the utilization of Bayesian optimization. VGG19 (b) The ideal hyperparameter values are chosen through the utilization of Random search.

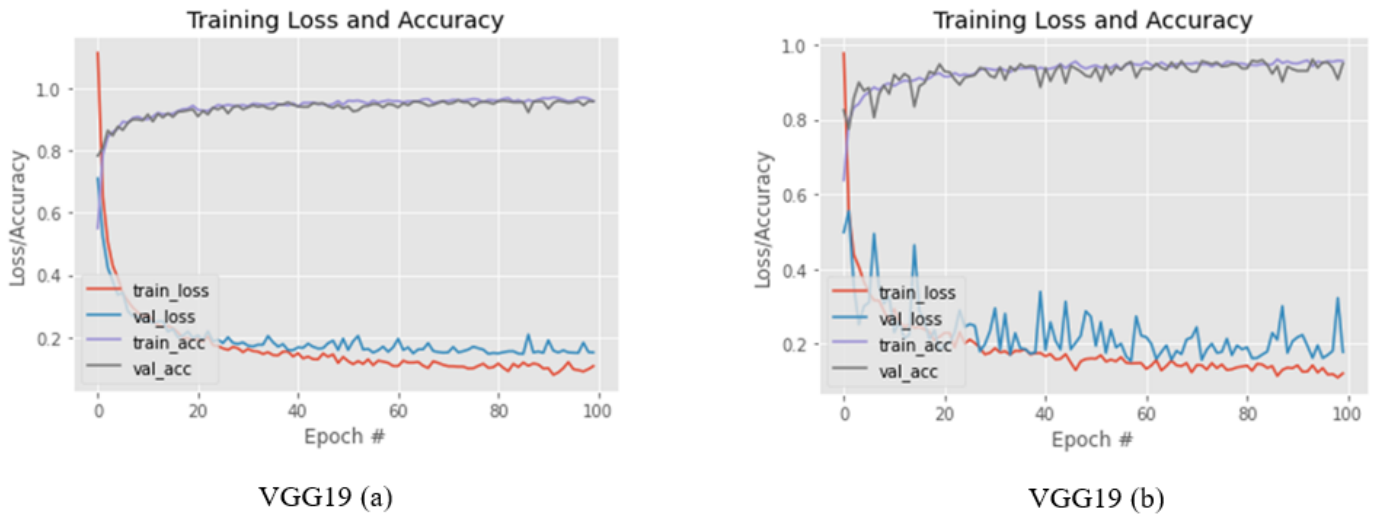


Fig. 5. Loss and accuracy plots of VGG19. VGG19 (a) The ideal hyperparameter values are chosen through the utilization of Bayesian optimization. VGG19 (b) The ideal hyperparameter values are chosen through the utilization of Random search.

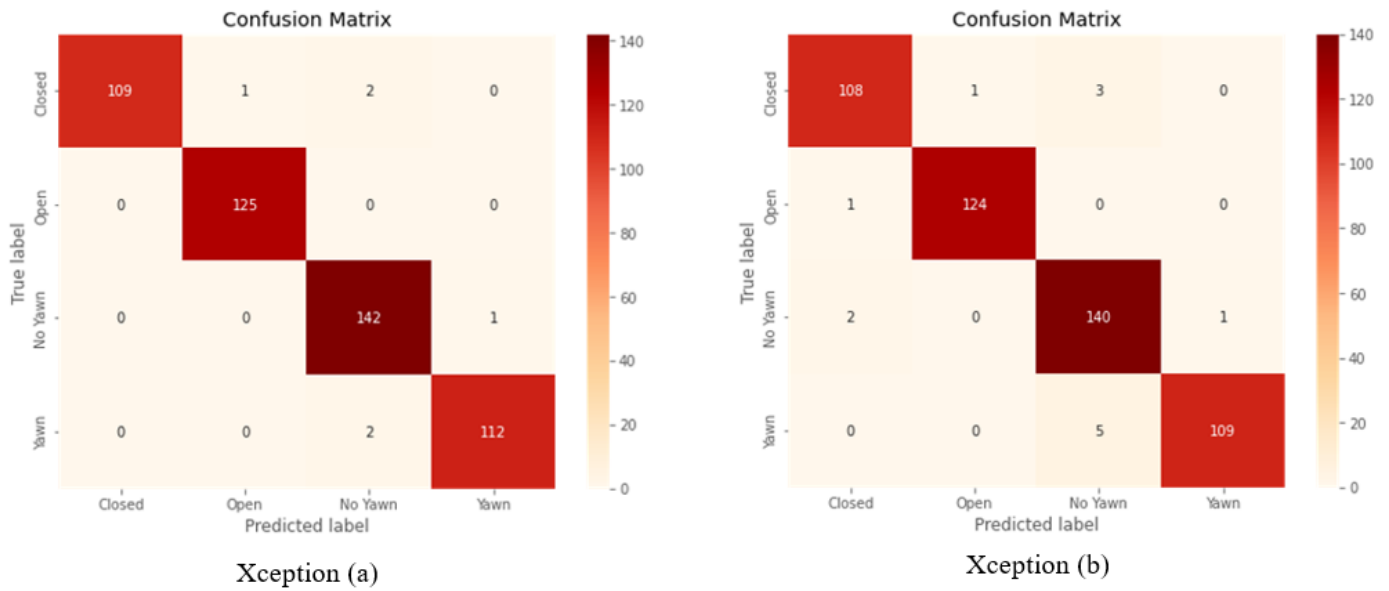


Fig. 6. Confusion matrix of Xception. Xception (a) The ideal hyperparameter values are chosen through the utilization of Bayesian optimization. Xception (b) The ideal hyperparameter values are chosen through the utilization of Random search.

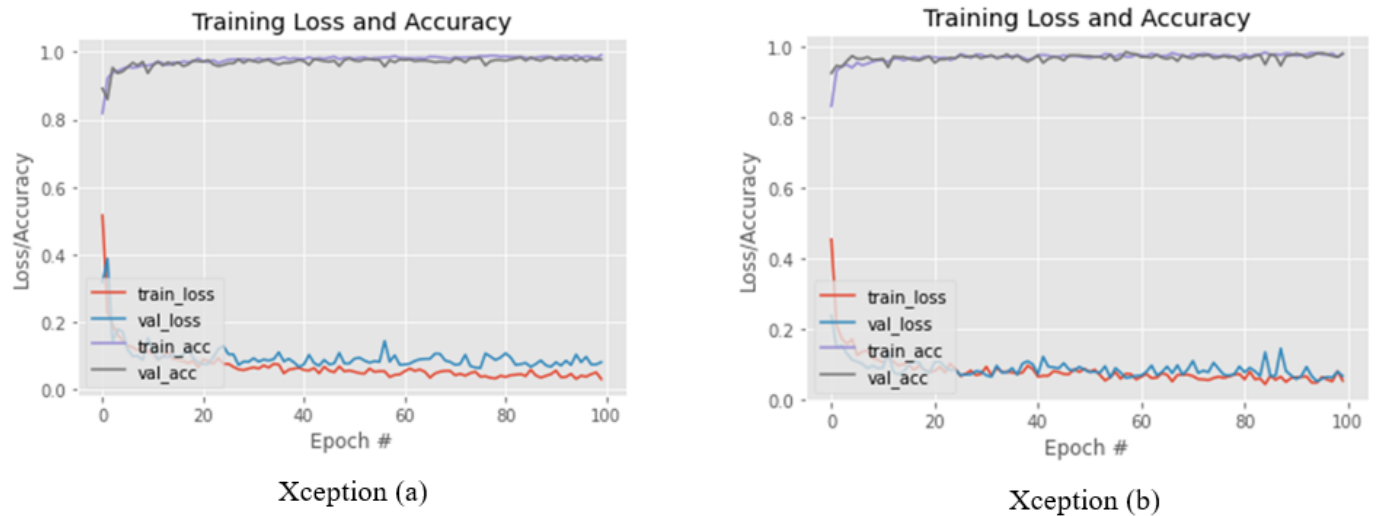


Fig. 7. Loss and accuracy plots of Xception. Xception (a) The ideal hyperparameter values are chosen through the utilization of Bayesian optimization. Xception (b) The ideal hyperparameter values are chosen through the utilization of Random search.

Overall, our study emphasizes the significance of selecting appropriate hyperparameters to achieve optimal performance in transfer learning models for detecting driver sleepiness and provides a framework for doing so. The results of this study can guide the selection of hyperparameters and models for future research in this area.

FUTURE WORKS

Further investigations can be carried out to explore the impact of other hyperparameters, such as weight decay and batch size, on the effectiveness of transfer learning models for detecting driver sleepiness. Furthermore, the combination of multiple optimization techniques, such as Bayesian opti-

mization, Random search, and Hyperband can be researched to further improve the efficiency of hyperparameter tuning. Finally, the proposed approach can be extended to real-time sleepiness detection for drivers using mobile devices. The implementation of optimized transfer learning models on such devices can enable real-time monitoring of driver drowsiness, thus enhancing road safety.

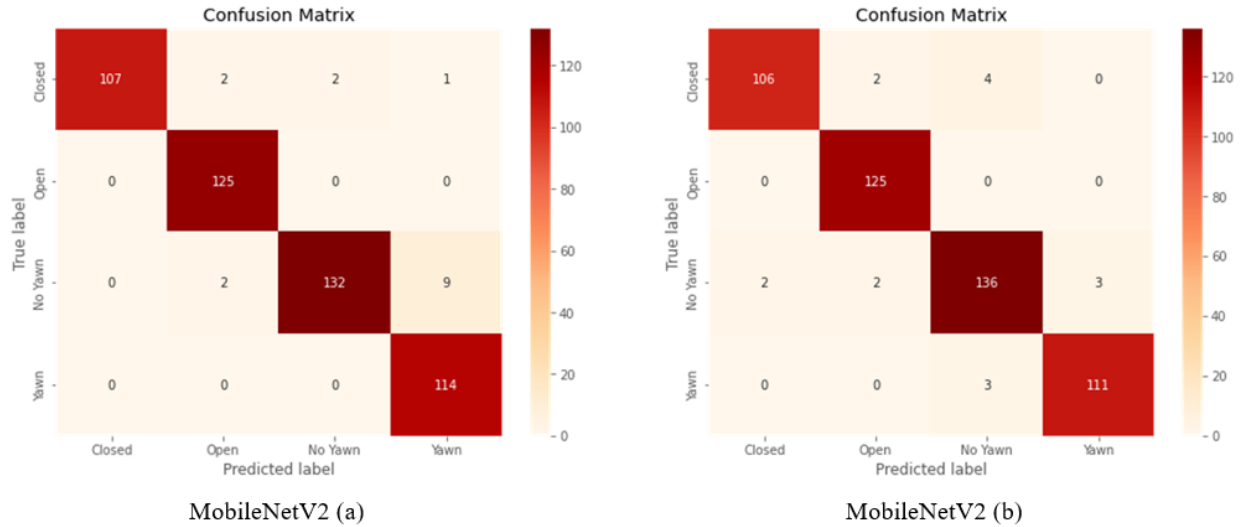


Fig. 8. Confusion matrix of MobileNetV2. MobileNetV2 (a) The ideal hyperparameter values are chosen through the utilization of Bayesian optimization. MobileNetV2 (b) The ideal hyperparameter values are chosen through the utilization of Random search.

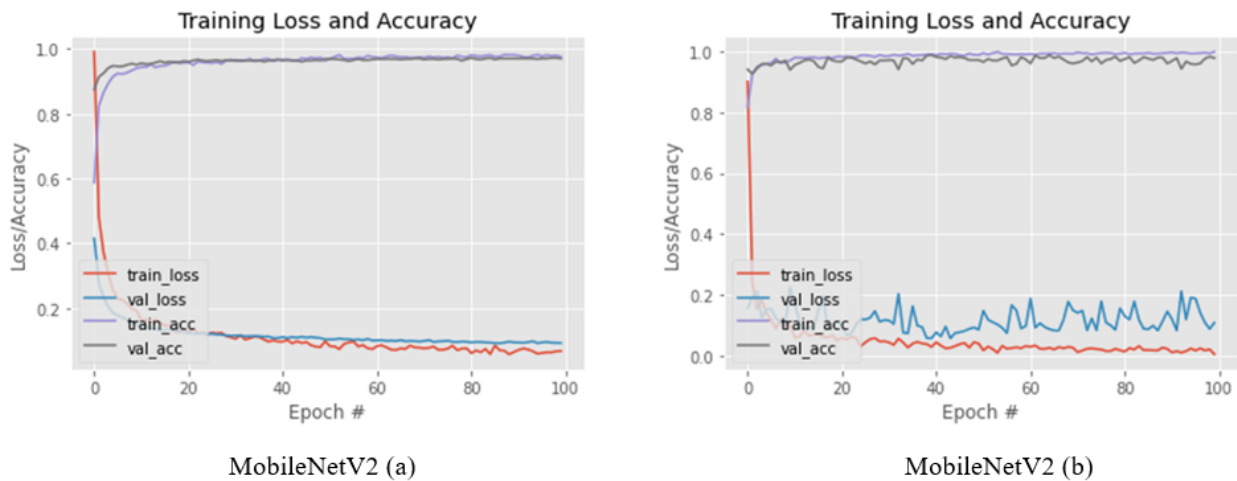


Fig. 9. Loss and accuracy plots of MobileNetV2. MobileNetV2 (a) The ideal hyperparameter values are chosen through the utilization of Bayesian optimization. MobileNetV2 (b) The ideal hyperparameter values are chosen through the utilization of Random search.

TABLE IV. THE PRECISION, RECALL, F1-SCORE, AND ACCURACY ACHIEVED DURING THE EVALUATION OF THE VGG-19 MODEL USING THE BEST HYPERPARAMETERS DETERMINED BY BAYESIAN OPTIMIZATION AND RANDOM SEARCH

VGG19	Precision	Recall	F1-score	Accuracy
<i>Bayesian Optimization</i>				
Closed	0.9649	0.9821	0.9734	0.9821
Open	0.9687	0.9920	0.9802	0.9920
No Yawn	0.9712	0.9440	0.9574	0.9440
Yawn	0.9823	0.9736	0.9779	0.9736
<i>Random Search</i>				
Closed	0.9090	0.9821	0.9442	0.9821
Open	0.9612	0.9920	0.9763	0.9920
No Yawn	0.9694	0.8881	0.9270	0.8881
Yawn	0.9823	0.9736	0.9779	0.9736

TABLE V. THE PRECISION, RECALL, F1-SCORE, AND ACCURACY ACHIEVED DURING THE EVALUATION OF THE XCEPTION MODEL USING THE BEST HYPERPARAMETERS DETERMINED BY BAYESIAN OPTIMIZATION AND RANDOM SEARCH

Xception	Precision	Recall	F1-score	Accuracy
<i>Bayesian Optimization</i>				
Closed	1.0000	0.9732	0.9864	0.9732
Open	0.9920	1.0000	0.9960	1.0000
No Yawn	0.9726	0.9930	0.9826	0.9930
Yawn	0.9911	0.9824	0.9867	0.9824
<i>Random Search</i>				
Closed	0.9729	0.9642	0.9686	0.9642
Open	0.9920	0.9920	0.9920	0.9920
No Yawn	0.9459	0.9790	0.9621	0.9790
Yawn	0.9909	0.9561	0.9732	0.9561

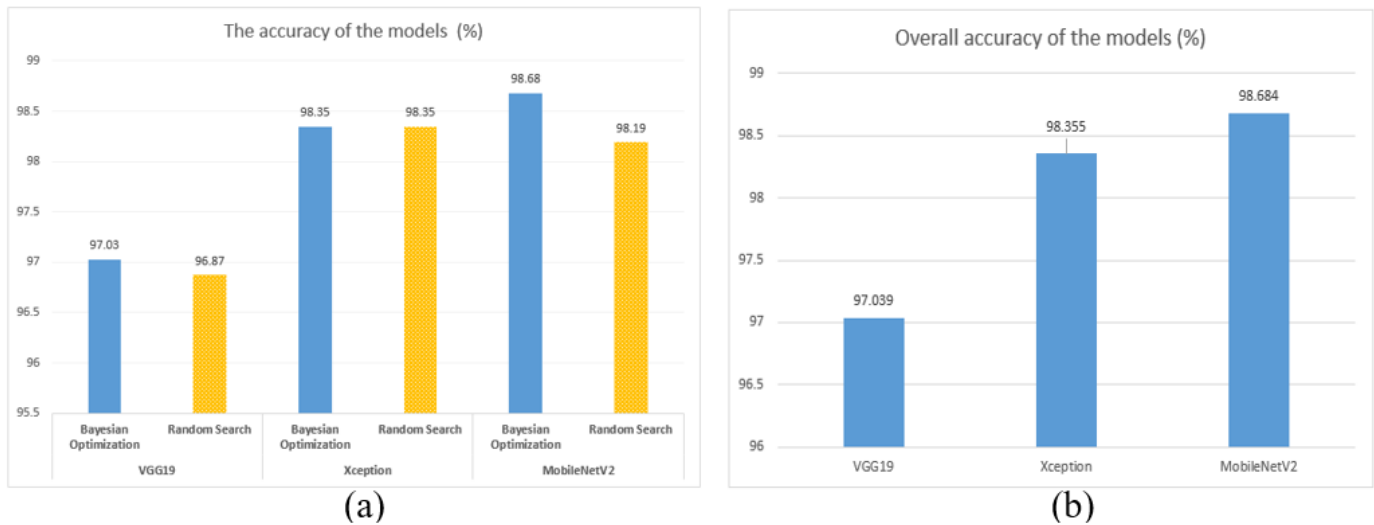


Fig. 10. (a) The accuracy of the models based on hyperparameters found by two methods: Bayesian Optimization and Random Search; (b) Overall accuracy of the models.

TABLE VI. THE PRECISION, RECALL, F1-SCORE, AND ACCURACY ACHIEVED DURING THE EVALUATION OF THE MOBILENETV2 MODEL USING THE BEST HYPERPARAMETERS DETERMINED BY BAYESIAN OPTIMIZATION AND RANDOM SEARCH

MobileNetV2	Precision	Recall	F1-score	Accuracy
<i>Bayesian Optimization</i>				
Closed	1.0000	0.9553	0.9771	0.9553
Open	0.9689	1.0000	0.9842	1.0000
No Yawn	0.9850	0.9230	0.9530	0.9230
Yawn	0.9193	1.0000	0.9579	1.0000
<i>Random Search</i>				
Closed	0.9814	0.9464	0.9636	0.9464
Open	0.9689	1.0000	0.9842	1.0000
No Yawn	0.9510	0.9510	0.9510	0.9510
Yawn	0.9736	0.9736	0.9736	0.9736

REFERENCES

- [1] Ali Mansour Al-Madani, Ashok T Gaikwad, Vivek Mahale, Zeyad AT Ahmed, and Ahmed Abdullah A Shareef. Real-time driver drowsiness detection based on eye movement and yawning using facial landmark. In *2021 International Conference on Computer Communication and Informatics (ICCCI)*, pages 1–4. IEEE, 2021.
- [2] Shahzeb Ansari, Fazel Naghdy, Haiping Du, and Yasmeen Naz Pahnwar. Driver mental fatigue detection based on head posture using new modified relu-bilstm deep neural network. *IEEE Transactions on Intelligent Transportation Systems*, 23(8):10957–10969, 2021.
- [3] Sadegh Arefnezhad, Sajjad Samiee, Arno Eichberger, and Ali Nahvi. Driver drowsiness detection based on steering wheel data applying adaptive neuro-fuzzy feature selection. *Sensors*, 19:943, 02 2019.
- [4] Athira Babu, Shruti Nair, and K Sreekumar. Driver’s drowsiness detection system using dlib hog. In *Ubiquitous Intelligent Systems: Proceedings of ICUIS 2021*, pages 219–229. Springer, 2022.
- [5] Venkata Phanikrishna Balam, Venkata Udaya Sameer, and Suchismitha Chinara. Automated classification system for drowsiness detection using convolutional neural network and electroencephalogram. *IET Intelligent Transport Systems*, 15(4):514–524, 2021.
- [6] James Bergstra and Yoshua Bengio. Random search for hyper-parameter optimization. *Journal of machine learning research*, 13(2), 2012.
- [7] Varun Chaudhary, Ziyad Dalwai, and Vikram Kulkarni. Intelligent distraction and drowsiness detection system for automobiles. In *2021 International Conference on Intelligent Technologies (CONIT)*, pages 1–4. IEEE, 2021.
- [8] François Chollet. Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1251–1258, 2017.
- [9] Nando de Freitas. A tutorial on bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning. 2009.
- [10] Wanghua Deng and Ruoxue Wu. Real-time driver-drowsiness detection system using facial features. *Ieee Access*, 7:118727–118738, 2019.
- [11] Farnoosh Faraji, Faraz Lotfi, Javad Khorramdel, Ali Najafi, and Ali Ghaffari. Drowsiness detection based on driver temporal behavior using a new developed dataset. *arXiv preprint arXiv:2104.00125*, 2021.
- [12] Christian Garbin, Xingquan Zhu, and Oge Marques. Dropout vs. batch normalization: an empirical study of their impact to deep learning. *Multimedia Tools and Applications*, 79:12777–12815, 2020.
- [13] Aicha Ghourabi, Haythem Ghazouani, and Walid Barhoumi. Driver drowsiness detection based on joint monitoring of yawning, blinking and nodding. In *2020 IEEE 16th International Conference on Intelligent Computer Communication and Processing (ICCP)*, pages 407–414. IEEE, 2020.
- [14] Markus Gromer, David Salb, Thomas Walzer, Natividad Martínez Madrid, and Ralf Seepold. Ecg sensor for detection of driver’s drowsiness. *Procedia Computer Science*, 159:1938–1946, 2019.
- [15] Soufiane Hayou, Arnaud Doucet, and Judith Rousseau. On the impact of the activation function on deep neural networks training. In *International conference on machine learning*, pages 2672–2680. PMLR, 2019.
- [16] Taeho Hwang, Miyoung Kim, Seunghyeok Hong, and Kwang Suk Park. Driver drowsiness detection using the in-ear eeg. In *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 4646–4649. IEEE, 2016.
- [17] Arafat Islam, Naimur Rahaman, and Md Atiqur Rahman Ahad. A study on tiredness assessment by using eye blink detection. *Jurnal Kejuruteraan*, 31(2):209–214, 2019.
- [18] Jennifer Jepkoech, David Muchangi Mugo, Benson K Kenduiwo, and Edna Chebet Too. The effect of adaptive learning rate on the accuracy of neural networks. 2021.
- [19] Zhuoni Jie, Marwa Mahmoud, Quentin Stafford-Fraser, Peter Robinson, Eduardo Dias, and Lee Skrypchuk. Analysis of yawning behaviour in spontaneous expressions of drowsy drivers. In *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, pages 571–576. IEEE, 2018.

- [20] Mateusz Knapik and Bogusław Cyganek. Driver's fatigue recognition based on yawn detection in thermal images. *Neurocomputing*, 338:274–292, 2019.
- [21] Ashish Kumar and Rusha Patra. Driver drowsiness monitoring system using visual behaviour and machine learning. In *2018 IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE)*, pages 339–344. IEEE, 2018.
- [22] Gang Li and Wan-Young Chung. Electroencephalogram-based approaches for driver drowsiness detection and management: a review. *Sensors*, 22(3):1100, 2022.
- [23] Jennalyn N Mindoro, Cherry D Casuat, Alvin Sarraga Alon, Mon Arjay F Malbog, and Julie Ann B Susa. Drowsy or not? early drowsiness detection utilizing arduino based on electroencephalogram (eeg) neuro-signal. *International Journal*, 9(2), 2020.
- [24] Suganiya Murugan, Jerritta Selvaraj, and Arun Sahayadhas. Detection and analysis: driver state with electrocardiogram (ecg). *Physical and engineering sciences in medicine*, 43(2):525–537, 2020.
- [25] Luyi-Da Quach, Nguyen Quoc Khang, Anh Nguyen Quynh, and Tran Ngoc Hoang. Evaluation of the efficiency of the optimization algorithms for transfer learning on the rice leaf disease dataset. *International Journal of Advanced Computer Science and Applications*, 13(10), 2022.
- [26] Luyi-Da Quach, Anh Nguyen Quynh, Khang Nguyen Quoc, and Nghe Nguyen Thai. Using optimization algorithm to improve the accuracy of the cnn model on the rice leaf disease dataset. In *Information Systems for Intelligent Systems: Proceedings of ISBM 2022*, pages 535–544. Springer, 2023.
- [27] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4510–4520, 2018.
- [28] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.
- [29] Muhammad Tayab Khan, Hafeez Anwar, Farman Ullah, Ata Ur Rehman, Rehmat Ullah, Asif Iqbal, Bok-Hee Lee, and Kyung Sup Kwak. Smart real-time video surveillance platform for drowsiness detection based on eyelid closure. *Wireless communications and mobile computing*, 2019, 2019.
- [30] Li Yang and Abdallah Shami. On hyperparameter optimization of machine learning algorithms: Theory and practice. *Neurocomputing*, 415:295–316, 2020.
- [31] Mina Zohoorian Jafari Yazdi and Mohsen Soryani. Driver drowsiness detection by yawn identification based on depth information and active contour model. In *2019 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICT)*, volume 1, pages 1522–1526. IEEE, 2019.
- [32] Mu Ye, Weiwei Zhang, Pengcheng Cao, and Kangan Liu. Driver fatigue detection based on residual channel attention network and head pose estimation. *Applied Sciences*, 11(19):9195, 2021.
- [33] Zelda B Zabinsky et al. Random search algorithms. *Department of Industrial and Systems Engineering, University of Washington, USA*, 2009.
- [34] Raniah Zaheer and Humera Shaziya. A study of the optimization algorithms in deep learning. In *2019 third international conference on inventive systems and control (ICISC)*, pages 536–539. IEEE, 2019.

Discovering COVID-19 Death Patterns from Deceased Patients: A Case Study in Saudi Arabia

Abdulrahman Alomary¹ Tarik Alafif² Abdulmohsen Almalawi³ Anas Hadi⁴
Computer Science Department Computer Science Department Computer Science Department Computer Science Department
Jamoum University College Jamoum University College King Abdulaziz University Arab Open University
Umm Al-Qura University Umm Al-Qura University Jeddah, Saudi Arabia Riyadh, Saudi Arabia
Jamoum, Saudi Arabia Jamoum, Saudi Arabia

Faris Alkhilawi⁵
Natural Products and Alternative Medicine Department
King Abdulaziz University
Jeddah, Saudi Arabia

Yasser Alatawi⁶
Pharmacy Practice Department
University of Tabuk
Tabuk, Saudi Arabia

Abstract—COVID-19 is a serious infection that cause severe injuries and deaths worldwide. The COVID-19 virus can infect people of all ages, especially the elderly. Furthermore, elderly who have co-morbid conditions (e.g., chronic conditions) are at an increased risk of death. At the present time, no approach exists that can facilitate the characterization of patterns of COVID-19 death efficiently and systematically is applied by adapting the Apriori algorithm. Validation and evaluation of the proposed approach are based on a robust and reliable dataset collected from Health Affairs in the Makkah region of Saudi Arabia. The study results show that there are strong associations between hypertension, diabetes, cardiovascular disease, and kidney disease and death among COVID-19 deceased patients.

Keywords—COVID-19; association rules; Apriori algorithm; patterns; death; chronic diseases

I. INTRODUCTION

COVID-19 first appeared in the Chinese city of Wuhan in December 2019, and it was the starting point of the disease spreading to the world. It was officially announced by the World Health Organization (WHO) on March 11, 2020 a pandemic [1]. Many attempts were involved to engage latest technologies for diagnosing and treating COVID-19 patients [2].

The Kingdom of Saudi Arabia (KSA) was not immune from that virus, as the virus spread rapidly. The first confirmed recorded case appeared in the KSA on March 2, 2020 [3]. There were several factors that contributed to the spread of the virus in the KSA, including its geographic location, trade exchanges with China, and its religious and recreational tourism industries [4], [5]. The KSA had previous experiences with epidemics. Viruses such as the Middle East Respiratory Syndrome (MERS-CoV), helped to control and deal well with the pandemic [6]. Since the start of the pandemic until 10 Nov 2022, there have been 284,273 cases and 9,426 deaths of COVID-19 in the KSA [7].

As a precautionary measure, the KSA had implemented strict curfews, shut down all non-essential services, required

everyone to wear masks, and closed its air, sea, and land borders. [8], [9]. The government had also transferred education to distance education in all public and private schools, universities, institutes and training centers [10]. Furthermore, Saudi Arabia intensified health care and established field hospitals to cope with the likely increase in COVID-19 patients, and relied on WHO and Ministry of Health protocols when treating COVID-19 patients [11], which significantly decreased the rate of death and hospitalization [12]. In the latest informed statistics during the writing of the last chapter of this research on December 29, 2022, the confirmed cases exceeded 816,470 cases, and 9,610 deaths of COVID-19 patients [13].

Most of the previous studies showed that the large number of people who died from COVID-19 had one or more chronic diseases in a remarkable way. Hence, the need to study those patterns of death from chronic diseases that are more frequent. Understanding these patterns helps to identify the factors that increase the risk of death in this group of people, and thus can help in developing better strategies to control the pandemic and provide a priority care. The issue of obtaining the database of the deceased patients from COVID-19 is difficult due to confidentiality, privacy, and sensitivity of medical data. In this proposed work, the medical data is officially obtained through health protocols from the Health Affairs in Makkah Al-Mukarramah region in the Kingdom of Saudi Arabia.

Our work is motivated by Apriori Association Rules (ARA) algorithm that was introduced in [14], which extracts the most frequent and appearing patterns from a specific database according to the degree of support and confidence [15]. Our work is similar to the work presented in [16]. However, we, in this study, discover COVID-19 death patterns from deceased patients rather than discovering COVID-19 infection patterns.

The remainder of the paper is organized as follows: In Section II, we present related works. In Section III, we describe and analyze the dataset and discuss our proposed approach for detecting death patterns among COVID-19 patients. In Section IV, we describe the experimental procedure based on the algorithm used. Section V discusses the Weka program while Section VI details implementation of Weka program in

our study. Section VII presents the results and discussion. In Section VIII, we conclude our study and present our future work.

II. RELATED WORK

Several research works thoroughly studied the COVID-19 patterns. In this section, we review the most recent related works.

Robert et al. [17] introduced the present COVID-19 situation among diabetics, newly diagnosed diabetics, diabetic ketoacidosis, and programmatic initiatives including immunizations. The study method involved performing a literature study through the use of PubMed, Google, and Scopus Up until July 15, 2021. The study results had shown that most research conducted in the KSA had shown diabetes as one of the most common comorbidities among COVID-19 patients. There had been few works conducted in the KSA on COVID-19-induced diabetic ketoacidosis and newly diagnosed diabetes. The Saudi ministry had implemented a number of steps, including thorough recommendations and prioritized immunizations, to reduce the impact of COVID-19 among individuals with diabetes. Telehealth services were heavily utilized in Saudi Arabian diabetes clinics during the COVID-19 pandemic.

Geng et al. [18] attempted to comprehend the risk variables for symptom deterioration and death in COVID-19 patients. By establishing the predictive value of chronic diseases for COVID-19 severity and mortality, this systematic review intended to fill the gap. The study results had revealed that hypertension was a fairly prevalent illness among COVID-19 patients, linked to mortality, admission to the intensive care unit (ICU), acute respiratory distress syndrome, and higher severity. Asthma was linked to a lower likelihood of COVID-19 death, while chronic obstructive pulmonary disease was the best predictor of COVID-19 severity, ICU admission, and fatality. Instead of mortality, the fat patients were more likely to develop severe COVID-19 symptoms. The patients were more likely to develop severe COVID-19 instances and die if they had cancer, chronic liver disease, chronic renal illness, or cerebrovascular disease.

Wang et al. [19] conducted searches in several databases for articles published until April 6, 2020, and after selecting a very large number of published articles. Only 34 articles were reached after sorting. Among the pre-existing chronic diseases were having high blood pressure, cardiovascular disease, kidney disease and diabetes associated with the risk of infection with the virus. The results confirmed that people with pre-existing chronic diseases had an increased risk of developing more serious complications in COVID-19 patients, and severe organ damage or dysfunction was linked to an increased death rate. The study showed that acute kidney disease and heart injury were closely associated with a 3-4-fold higher risk of death associated with COVID-19. The study suggested that healthcare providers need to put those who have a history of high blood pressure and cardiovascular diseases under observation and be subject to continuous follow-up.

Al Mutair et al. [20] collected information from several private hospitals in the KSA between April 2020 to June 2020. The study was based on descriptive and inferential

analysis of the results and on the data analysis of COVID-19 patients' information. The study was classified into two categories: survivors (recovered) and non-survivors (deceased). The inferential reading showed that 31.8% of the survivors were in the age group of 30-40 years, and 24% were in the age group of 21-30 years, and among the non-survivors, the study showed that 66% of the non-survivors over 50 years old, and that 86% of the non-survivors were males. The results showed that 63.8% had no history of chronic diseases Hypertension (HTN), that 19.8% had one of HTN or diabetes mellitus (DM), and that 16.4% had a history of chronic diseases HTN and DM together. The study indicated a strong relationship between differences in age, gender, chronic diseases and deaths associated with COVID-19 in general, and that advanced age and males and the presence of chronic diseases such as high blood pressure and diabetes among COVID-19 patients are prevalent among the deceased patients. The study indicated that there were poor predictive factors that led to the possibility of an increase in the death rate that may reach 9 times among males over the age of 50 years and who have a history of chronic diseases. There is an association between HTN and mortality, and there is an increasing number of deaths among males.

Similar to our work, Alafif et al. [16] explored the most common infection patterns for patients who had recovered from COVID-19, and they were able to collect 131 records of people who recovered from COVID-19 using the survey method through a questionnaire by communicating with them or with their relatives by direct contact. The ARA algorithm was used on the manually collected data to explore the most common infection patterns. The study concluded that there were strong association rules with high confidence scores among males, weight above 70 kg, height above 160 cm, and fever patterns. Also, Alafif et al. [21] predicted the status of COVID-19 patients using their patterns as COVID-19 treatments and diagnosis were reviewed in [22].

From the previous studies, we find most of these studies confirm that the age factor, males, non-Saudis, and those with chronic diseases are the most at risk of death among people infected with the COVID-19, with differences in ratios, study sample and place of study. It was also found that the most prevalent chronic diseases among those infected and deceased from the COVID-19 virus are hypertension, diabetes, cardiovascular disease and kidney disease.

We did not find any previous study that dealt with the same topic and finding of the most frequent chronic diseases found among the deceased due to the COVID-19 death pattern, as we are in this study to discover the death patterns of chronic diseases that were among the deceased and the most frequent.

We discuss the results of our study and a comparison with previous studies in the discussion section, clarifying the compatibility with previous studies and the contradiction of their results.

III. METHODOLOGY

Fig. 1 shows the course of the study stages, from defining the study sample, obtaining the database, the stages of cleaning, coordinating and arranging the database, and analyzing the data, then the stage of selecting the study tools, which

is the ARA algorithm, followed by the use of two methods to apply the algorithm through manual calculations and the Weka program to verify the conformity of the outputs and finally the stage of writing the results.

A. Association Rules Algorithms

In our modern era, the uses of artificial intelligence are abundant to utilize as much as possible. These can be employed in several fields, including data science, data mining, data analysis, and serial cases that in the past took effort a long and hard time. The most classic algorithm is the ARA algorithm, which is considered the originator of the recursive element mining algorithm [23].

The amount of information or data being maintained in databases has dramatically increased in recent years. Latent knowledge mining is crucial to enhance decision-making when the size of available databases grows exponentially. The critical phase of the knowledge discovery process is data mining. Predictive and descriptive tasks are the two broad categories into which data mining's primary tasks are typically categorized. While the descriptive tasks aim to extract previously undiscovered and significant information from massive databases, such as patterns, associations, changes, anomalies, and substantial structures, the predictive tasks aim to forecast the value of a given attribute based on the standards of other characteristics. These goals of data mining can be met in many ways [24].

There has been more than one algorithm whose works are based on association rules, such as the ARA algorithm, the ECLAT algorithm, and the FP Growth algorithm, but the ARA algorithm is the most famous. The ARA algorithm appeared by Agrawal et al. in 1993 as one of the tools for mining databases of association data. The ARA algorithm extracts recurring patterns are based on the strength of the degree of support and confidence and provide services through their analysis to planning and marketing managers to facilitate the administrative and marketing process and help make appropriate decisions [25].

The main task of the ARA algorithm is to extract the duplicate elements with iterative operations by scanning the database more than once according to the minimum support limit, excluding other data and filtering [26].

Association rules that explore recurring, relational, or causal patterns between groups of itemset emerged from the database. The rules of association were used through the exploration and analysis of shopping cart data at the beginning, and their use has evolved in several areas due to their efficiency and effectiveness in results. The science of data mining is of great importance to obtain information from large or small sources of various categories and to present it for optimal use. The rules of the association are one of the most important methods of data mining. There are several measures by which the strength of the degree of correlation of the algorithm outputs is measured, and we have relied on two measures support and confidence. The method of knowledge discovery in database, is called Data mining. It is based on extracting important patterns from large databases and then analyzing that data to give beneficial information, ideas, and solutions. Through what the discovery of knowledge in databases provides in terms of methods and methods for extracting interesting remarkable

patterns in the shortest possible and appropriate period and taking paths at the intersection point of data stores and machine learning, and statistics systems. Data mining is known as applying algorithms to desired patterns found in large databases [27].

The ARA algorithm is an intriguing way to determine what we need to buy or get suggestions for what we need. We are all aware that the e-commerce platform has a variety of approaches. Flipkart, Amazon, Snapdeal, and other online retailers, are what it exactly is. The application proposes items to buy when we attempt to purchase in the online store. It anticipates additional clients who usually make purchases together. This method also enables us to understand how various things are predicted [28].

B. Apriori Algorithm

The Apriori algorithm is an analogue algorithm for association base mining (association base mining) and ranks among the top ten data mining algorithms. Association rule mining is a significant research direction in data mining. It is also a long-standing topic whose main task is to find the inner connection between things [29]. The association rules approach can be implemented using various methods to draw out relevant links between variables or items in a large database by identifying more frequent item combinations. As long as those itemsets appear in the dataset sufficiently and frequently, it recognizes the frequently occurring items in the database and expands them to larger and larger itemsets. If an itemset is frequent, then all of its subgroups' items will also be frequent, according to the Apriori algorithm. In other words, any subset of a frequent itemset must also be frequent [30].

C. History of the ARA

The first algorithm to be suggested for frequent itemset mining was the ARA algorithm. Later, it was enhanced by R Agrawal and R Srikant [14]. The result was known as Apriori. The join and prune steps of this algorithm are used to condense the search space. Finding the most common itemset is done iteratively. The term ARA refers to the algorithm that determines the rules of association between items. It refers to the relationship between two or more objects [15]. To put it another way, we can say that the ARA algorithm is a leaning association rule that examines whether customers of product (A) also purchase product (B) [31].

D. The ARA Algorithm Template

To find the huge l-items, the algorithm initially counts the occurrences of each item. These stages make up the second step. The candidate itemsets C_k are first created using the large itemsets L_{k-1} . After that, the database is scanned to determine the support of candidates in C_k , which involves increasing the count of all candidates included in a certain transaction t . After that, $supp_min$ is used to compare the support of each candidate k -itemset. Only when the support is more than the minimal support given by the user, $supp_min$, is the k -itemset frequent. The algorithm then deletes all rules that do not satisfy the criterion of $supp_conf$. The next stage is to identify those rules with confidence greater than or equal to the user-specified $supp_conf$ [32].

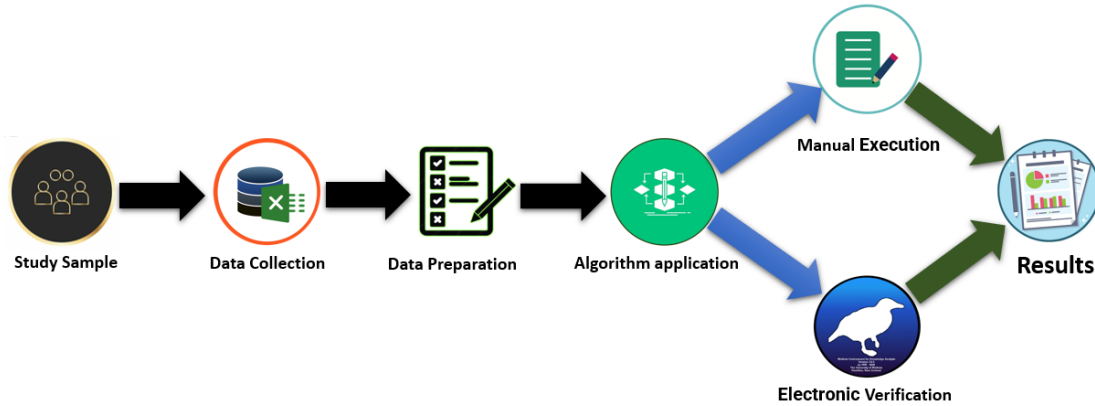


Fig. 1. The pipeline methodology of the proposed approach.

E. Components of the ARA Algorithm

The ARA algorithm consists of more than one component, there are several measures by which the strength of the degree of correlation of the algorithm outputs is measured [33], and we have relied on two criteria; support and confidence. The term support refers to the frequency of itemset in the dataset on all parameters and is defined as follows:

$$Support(X \Rightarrow Y) = \frac{x \cup y}{TotalTransactions} \quad (1)$$

The term (confidence) refers to the frequency of the dataset of utmost Y in transactions containing X , which are defined as:

$$Confidence(X \Rightarrow Y) = \frac{supp(x \cup y)}{supp(x)} \quad (2)$$

We employed the ARA algorithm to compute the probability of the appearance of the most frequent chronic diseases along with other chronic diseases in the database of the deceased patients from COVID-19 to provide the results to medical providers to take the necessary precautions or provide appropriate health care. In an iterative process, the data is filtered to extract the itemset that achieves the minimum support, and then it is re-read to extract a new itemset, considering $k+1$. As appears in Fig. 2.

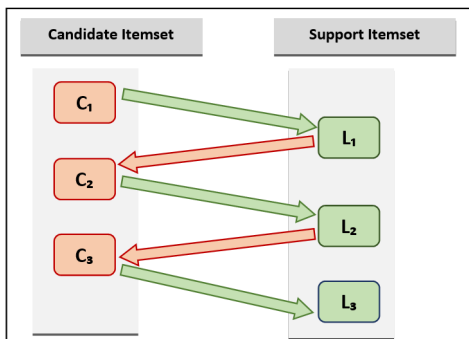


Fig. 2. Filtering iterations.

Fig. 3 shows a simple example of the work of the ARA algorithm in analysing the shopping cart, where the database is completely read, and then each item and its number of repetitions is separately determined. Then, the itemset is filtered according to the minimum support specified in advance by the user, which we call $C1$ to produce a grouping of the dataset that meets the minimum support as a single item 1 -Itemset with the exclusion or deletion of any item that does not meet the threshold of the minimum support and this is what we call $L1$.

With $L1$ we iterated the process but with two items together $K+1$ or 2 -Itemset which is the process of filtering the groups of itemset $C2$ comparing the number of iterations with the minimum support, and then trimmed the elements that did not meet the threshold of the minimum support. The process was repeated according to the repetition of the itemset with the increase of $K+1$ each time, meaning that in the third time, the itemset was three with each other 1 -Itemset and compared the number of iterations with the threshold of the minimum support until we reached that it was not possible to create a new group of itemset because the minimum support was not achieved. There was a logical consequence that a non-frequent set cannot be in a set of frequent itemsets. Next, we computed the specific confidence value for those potential itemsets that were most frequent, Accordingly, we saw the strength of that confidence and the correlation of the group of elements whether appearing next to each other or not.

F. Dataset

Our study is based on an Excel database of those who died due to COVID-19 and had a disease or chronic diseases that contributed to the death. After taking approval, the database was obtained from the information centre at Al-Noor Specialist Hospital in Makkah. We processed, audited, revised, and corrected the database, as it initially contained 18,785 records, most of which were duplicates, incomplete, or invalid. The reason for its large size is due to the placement of more than one record for the same patient, where each one contains its patient's daily and weekly condition development independently, but neither an update nor a placement of a cell for the new case, as well as the patient's data, such as medical

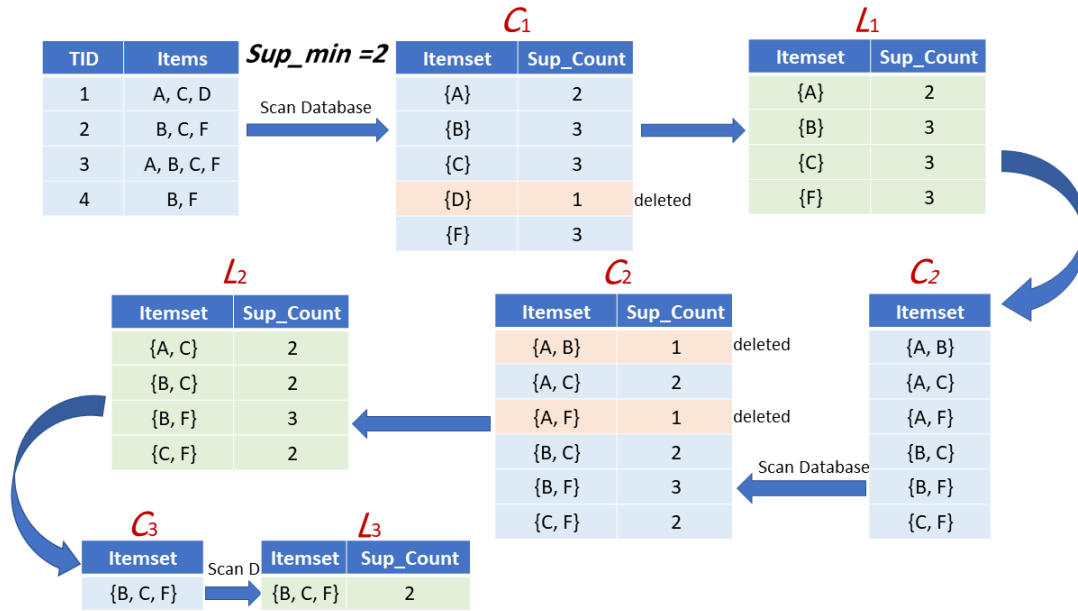


Fig. 3. A simple example of the workflow of the ARA algorithm.

number, age, gender, and nationality. Some patients had more than 43 records for each of them individually. This is a large number and major reason for the records to search 18,785. This is if we know that every patient data has a recurrence. At the end of the corrections and audits, the number of valid records for the study was only 349 for those who died due to COVID-19 and were accompanied by one or more chronic diseases such as diabetes, hypertension, cardiovascular, kidney diseases, immunodeficiency, blood diseases, cancer, immunodeficiency, cardiovascular, and obesity. The dataset is publicly available on¹.

G. Database Analysis

The database included 349 records of those who died of COVID-19 and had one or more chronic diseases, and the number of registered diseases amounted to 10 different ones, as shown in Fig. 4.

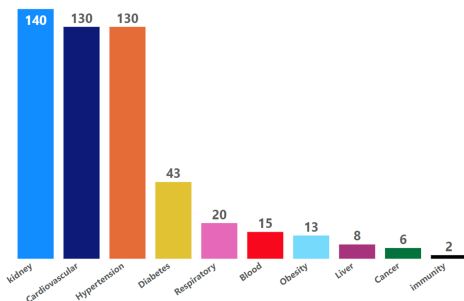


Fig. 4. Chronic diseases and how often they appear in the database in the study population.

The percentage of the non-Saudis was more than of the Saudis, according to the results of the database analysis of the deceased patients. The non-Saudi number was 217 and the Saudis was 132, as shown in Fig. 5.

The number of non-Saudi constitutes from 20 different countries. The countries are Jordan, Afghanistan, Indonesia, Pakistan, Burma, Bangladesh, Thailand, Chad, Tunisia, Syria, Sudan, Philippines, Palestine, Malaysia, Egypt, Morocco, Mauritania, Nigeria, India, and Yemen.

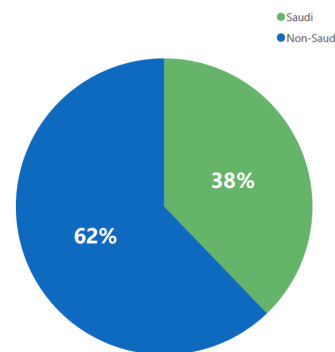


Fig. 5. The ratio of Saudi to non-Saudi COVID-19 patients in the study population.

Fig. 6 shows that the analysis results of the database also showed that the number of males exceeds the number of females in deaths. According to the previous studies we reviewed, they were fully compatible with the fact that the number of males exceeded the number of females since the onset of the disease until the date of writing the study.

¹<http://www.kaggle.com/datasets/abduhrahmanalomaly/discovering-covid-19-death-patterns>

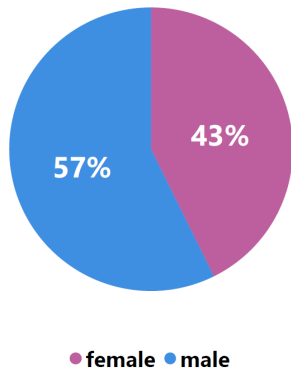


Fig. 6. The ratio of males to females in the study population.

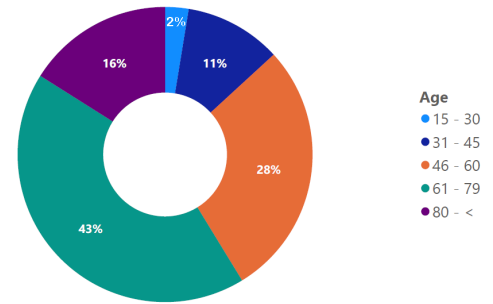


Fig. 8. Distribution of age groups in the study population.

Also, the database records of the deceased patients also show a diversity of blood groups. The blood group O+ is more prominent, followed by A+, then B+, and then the rest of the blood groups, as shown in Fig. 7. However, there are many records in the database of unknown blood types (not recorded) from the source and written in the field (unknown). Fortunately, we do not rely on blood type to explore the most frequent patterns of the cause of death, as the type of chronic diseases and their nomenclature are the study subjects for the implementation of treatment operations.

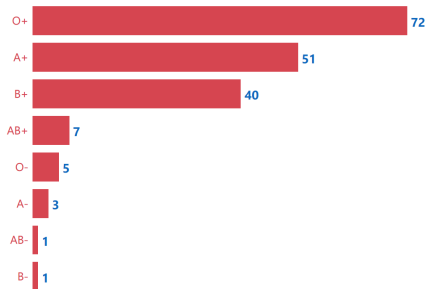


Fig. 7. Types of blood groups in the study population.

The strength of youth appears in all aspects, as the nature of youth is strength, vitality, activity, and endurance of hardships, as well as the case when diseases are infected, we find the strength of immunity or resistance, which is different from the elderly.

The study showed the disparity in the ages of the deceased patients with COVID-19. They had chronic diseases, as the ages of young people age range from 15 to 45 which represented only 13% of the total of deceased patients according to the study's database.

On the other hand, we find that the elderly, whose age ranges from 46 to over 80, are the highest percentage of deaths, reaching 87%. Fig. 8 shows the distribution of ages by age groups into groups.

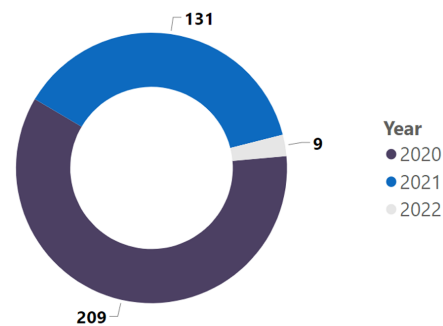


Fig. 9. Total deaths by years in the study population.

COVID-19 appeared at the end of the year 2019 and spread in the year 2020, the disease spread largely and frighteningly, as well as the number of deaths, then the percentage decreased in the following year 2021 until the percentage decreased significantly in the year 2022. Fig. 9 shows the number of deaths during the three consecutive years 2020, 2021 and 2022, and the significant difference between them.

IV. IMPLEMENTATION OF ALGORITHM

In our study, we relied on the ARA algorithm to discover (patterns of death) according to strong association rules through the database of deceased COVID-19 patients. In this section, we manually applied the ARA algorithm at first, the database file (Excel) was manually emptied and written as in Table IV. The medical file number for each deceased patient has been replaced by serial numbers. The names of chronic diseases (patterns) were abbreviated with corresponding letters for ease of writing. We had 10 different types of chronic diseases (patterns) in the database, and their names have been abbreviated with Latin letters for ease of reading, writing, and dealing with them during the manual analysis, as explained in Table I.

TABLE I. LIST OF TYPES THE DISEASES AND THEIR ABBREVIATIONS

Disease Name	Abbreviation
Kidney Diseases →	K
Hypertension →	H
Cardiovascular →	C
Diabetes →	D
Respiratory →	R
Blood Diseases →	B
Obesity →	O
Liver Diseases →	L
Cancer →	N
Immunodeficiency →	I

The minimum support was determined according to our reading and estimation of the size of the records and the presence and frequency of patterns in the database. The minimum support is 7.

TABLE II. RESULTS OF THE INITIAL FILTERING PROCESS FOR THE DATASET (L1)

Patterns	Supp_Count
{B}	15
{C}	130
{D}	43
{K}	140
{L}	8
{O}	13
{R}	20
{H}	130

We first performed the process of reading all the patterns, an element of an element from the database (1- frequent), and the number of supporting each element. This is called a filtering process C1 as in Table III. Depending on the predetermined minimum support, the patterns were pruned (deleted) and excluded for the items that did not meet the minimum support. Table II embodies L1.

TABLE III. SCANNING THE DATABASE AND READING ALL PATTERNS AND THEIR FREQUENCY (C1)

Patterns	Supp_Count	
{B}	15	
{C}	130	
{D}	43	
{I}	2	deleted
{K}	140	
{L}	8	
{N}	6	deleted
{O}	13	
{R}	20	
{H}	130	

According to the results of the first filter C1 in Table III, it is clear that the pattern set I and N did not achieve the minimum support 7, and pruning (deletion) was implemented. Table V shows the second filtering process C2 which was performed by reading two 2-frequent items for all patterns from Table II (L1).

In the next step, the filter results L2 were extracted for every two items (2-Itemset) according to the minimum support from reading Table V (C2), and the patterns that did not meet the minimum support were excluded. Many patterns for which the number of supporting was below the threshold were excluded. 7 frequent patterns 2-itemset were obtained as it appears in Table VI.

TABLE VI. RESULTS OF THE SECOND FILTRATION PROCESS (L2) 2-ITEMSET

Patterns	Supp_Count
{B, H}	7
{C, D}	9
{C, K}	30
{C, H}	31
{D, K}	15
{D, H}	28
{K, H}	38

After knowing the number of support and obtaining a successful reading for each of the two elements (L2), and since there is a repetition that fulfilled the minimum condition for support, we moved to the next step through which all the three elements (patterns) read 3-Itemset, i.e., forming new triple patterns based on the results of Table VI (L2).

In Table VII, (C3), the data was filtered with a 3-Itemset for every three items. The number of repetitions was computed based on the results of Table V.

Through the results of the 3-Itemset reading of the three repetitive patterns, only two groups that met the threshold of support are shown C, K, H and D, K, H. The first group of elements was repeated 13 times, and the second 10 times, as shown in Table VIII.

We continued filtering the data to generate a new set of items (Patterns) for every four 4- Itemset, where there were two sets, and the number of supporting for each set was greater than the threshold minimum support.

In Table IX (L4), 4- Itemset data is filtered for every four items, and the number of repetitions is computed based on the results of Table VIII (L3), which achieved the threshold minimum support. The pattern set shows C, D, K, and H.

TABLE IV. LIST OF TYPES (CHRONIC DISEASES) DATABASE RECORDS OF COVID-19 DEATHS

ID	Patterns (Diseases)	ID	Patterns (Diseases)	ID	Patterns (Diseases)	ID	Patterns (Diseases)
1	H,D	89	H,R	177	K,H	265	C
2	K	90	C	178	H	266	B
3	K	91	K,I	179	H	267	L
4	C	92	H,D,R	180	C	268	K
5	K	93	C	181	H	269	O
6	H	94	H	182	H	270	K
7	K,H,C	95	K	183	H,C	271	K,C
8	K,C,R	96	L	184	H	272	C
9	K	97	C	185	C	273	H,D
10	C	98	H,D	186	H,C	274	K
11	H,D	99	K	187	C	275	C
12	H,D	100	K	188	K	276	B
13	H	101	K,H,C,D,R	189	H	277	K,C,N
14	D	102	C	190	H	278	H
15	D	103	K,H,D	191	H	279	C
16	C	104	K,H	192	H,C	280	C
17	K	105	H,C,R	193	K	281	K
18	B	106	C	194	K,D	282	D
19	K	107	K,C	195	H	283	C
20	C	108	K,H	196	K	284	C
21	C	109	H	197	K,H,C	285	H,C,B
22	L	110	H	198	I	286	K
23	K,C	111	H,C	199	H	287	K,D
24	C	112	H,D	200	O	288	K
25	K,H	113	K,N	201	H	289	K
26	C,B	114	K	202	H	290	H
27	C	115	H	203	N	291	H
28	K,H,C,D ,B	116	C	204	K	292	K,H,C
29	R	117	N	205	C	293	K,H,C
30	K,C	118	C	206	H	294	K
31	H	119	C,D	207	C	295	H
32	H	120	H	208	L	296	B
33	H	121	H,D	209	H	297	K,H
34	D	122	K	210	H	298	H,C
35	K,H,D	123	K	211	K,C,O	299	H
36	R	124	R	212	H,B	300	K
37	K,H	125	K	213	K	301	C
38	C	126	K	214	R	302	K
39	C	127	K,R	215	K,D	303	H
40	C	128	K,O	216	H	304	K,N
41	K,C	129	D	217	K	305	D
42	O	130	H	218	K	306	K
43	C,R	131	H,C	219	K,C	307	H
44	C	132	H,R	220	K,H	308	H,D
45	K	133	K,C	221	C,O	309	K
46	K,C	134	C	222	C	310	K,H
47	K,C	135	N	223	K,H,C	311	K,H,C,B
48	K	136	B	224	K	312	H,C
49	O	137	K	225	K,O	313	K,C
50	C	138	K	226	C	314	K
51	K	139	H,D	227	K,H,L	315	K,H,D
52	H	140	K,R	228	K	316	H,D
53	K	141	K	229	B	317	K,H,B
54	K	142	K,H,C	230	K,H	318	D
55	K,C	143	H	231	K,C	319	K,C
56	O	144	K	232	K	320	H,C
57	K	145	H,C	233	K	321	K,C
58	C,D	146	C	234	C	322	H,D
59	C	147	K	235	K	323	R
60	C	148	K	236	C	324	H,D
61	C	149	K	237	K,D	325	K,H,D
62	L	150	K	238	H	326	H,D
63	H	151	O	239	H,C	327	C
64	K	152	K	240	H,C	328	C
65	K,H,R	153	K,H	241	H	329	K,H,D,C
66	K,H,B	154	H	242	C	330	H
67	K	155	K	243	C	331	K
68	H	156	K,C	244	C	332	K
69	C	157	H	245	C	333	K,H
70	K	158	H,C,D	246	H,C	334	C,O
71	K,H	159	H	247	H,C	335	C
72	H	160	H,D	248	C	336	K,C
73	K,C	161	C	249	R	337	K,B
74	C,O	162	K,D	250	C	338	K
75	K	163	H,C	251	C	339	C
76	K	164	H	252	H	340	K,H,C,D,O
77	K	165	C	253	C	341	K,C
78	C	166	H,D	254	H,C	342	C
79	K	167	H	255	R	343	H,D
80	C	168	H	256	R	344	K
81	K,H,C,D,B	169	K,H	257	C	345	K,H,D
82	C	170	C,D	258	R	346	H
83	K	171	K	259	L	347	K,H
84	C	172	K,H	260	C	348	R
85	K,H,C	173	H	261	C	349	C
86	C	174	K,H	262	C,L		
87	H	175	H	263	C		
88	C	176	H	264	C		

TABLE V. READING 2-ITEMSET FROM THE DATASET (C2)

Patterns	Supp_Count		Patterns	Supp_Count	
{B, C}	5	deleted	{D, L}	0	deleted
{B, D}	2	deleted	{D, O}	1	deleted
{B, K}	6	deleted	{D, R}	2	deleted
{B, L}	0	deleted	{D, H}	28	
{B, O}	0	deleted	{K, L}	1	deleted
{B, R}	0	deleted	{K, O}	4	deleted
{B, H}	7		{K, R}	5	deleted
{C, D}	9		{K, H}	38	
{C, K}	30		{L, O}	0	deleted
{C, L}	1	deleted	{L, R}	0	deleted
{C, O}	5	deleted	{L, H}	1	deleted
{C, R}	4	deleted	{O, R}	0	deleted
{C, H}	31		{O, H}	1	deleted
{D, K}	15		{R, H}	6	deleted

TABLE VII. READING 3-ITEMSET (C3)

Patterns	Supp_Count	
{B, C, D}	2	deleted
{B, C, K}	2	deleted
{B, C, H}	4	deleted
{B, D, K}	2	deleted
{B, D, H}	2	deleted
{B, K, H}	4	deleted
{C, D, K}	5	deleted
{C, D, H}	6	deleted
{C, K, H}	13	
{D, K, H}	10	

TABLE VIII. FILTERING THE RESULTS FOR 3-ITEMSET (L3)

Patterns	Supp_Count
{C, K, H}	13
{D, K, H}	10

V. USING WEKA PROGRAM

Weka is an acronym for Waikato Environment for Knowledge Analysis and is developed by the University of Waikato in New Zealand [34]. Weka is an open-source software that contains a set of algorithms and graphics for data analysis and predictive modeling, which is easy to access and use with a graphical interface. The third version of WEKA was fully developed through the Java language in 1997. It is used in applying several tools and is freely available according to the GNU General License. The program can run on any operating system because it is implemented on the Java platform to data pre-processing, classification, regression, clustering, correlation base mining, visualization, and modeling [35], [36].

TABLE IX. THE RESULTS OF THE FILTRATION PROCESS (L4)

Patterns	Supp_Count
{C, D, K, H}	13

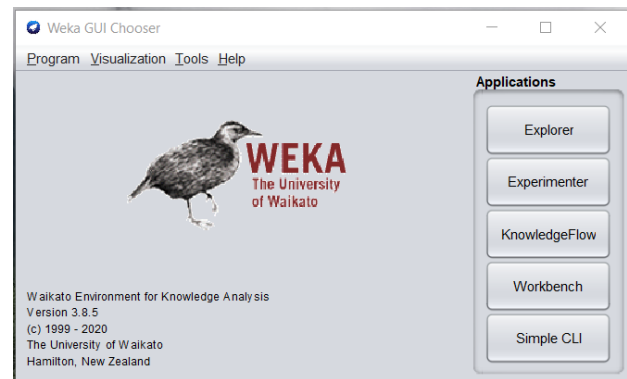


Fig. 10. Weka main interface [31].

VI. DATA PREPARATION AND IMPLEMENTATION USING WEKA

To use the Weka program, there is a command that must be modified on the database, a copy of the database (Excel) to convert it to a system (YES - NO) with the extension (.csv) with values separated by commas so that the program read it and deals with it, according to the program settings and apply the ARA algorithm. After opening the file in the program, it is converted and saved with the program extension (.arff) to apply the algorithm and extract the results.

Fig. 10 shows the program's main graphic interface, and in Fig. 12 shows the window for selecting working files and selecting the classification type and algorithm. In Fig. 11, we can see the settings for the algorithm, such as the amount of support and confidence.

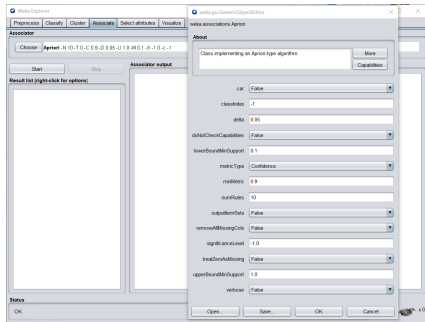


Fig. 11. Apriori algorithm properties settings window for support and confidence.

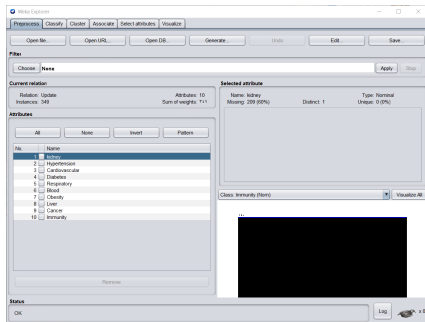


Fig. 12. The window for selecting rules and categories (Associate).

VII. RESULTS AND DISCUSSION

A. Manual Results

Through our manual calculations of the ARA algorithm, we explore a set of 4-itemset recurring patterns, which are the most frequent patterns and have a strong correlation base for the probability of appearing next to each other, which are C, D, K, H.

Each pattern has a certain frequency with one or more other patterns explored C, D, K, H, where we find more than one association rule that we read with differences in confidence. According to the rule of Apriori algorithm that states, if an itemset is frequent, then all of its subgroups' items will also be frequent. In other words, any subset of a frequent itemset must also be frequent. Here is a reading of more than one association rule for the set of detected patterns that have high confidence:

- {K, D} \Rightarrow {H} - {D} \Rightarrow {K}
- {D} \Rightarrow {H} - {K, H} \Rightarrow {C}
- {H, C} \Rightarrow {K} - {H} \Rightarrow {K}
- {H, D} \Rightarrow {K} - {K} \Rightarrow {H}
- {K, C} \Rightarrow {H} - {C} \Rightarrow {K}

Table X shows the number of appeared of the most frequent patterns (support) in the 349 records of the database, where we counted each pattern individually. Then, every two patterns were counted together, counted the number of recurrences for every three patterns, and finally, every four patterns were computed together in the database.

TABLE X. SUPPORT COMPUTATIONS FREQUENCY FOR ALL MOST FREQUENT PATTERNS IN DATABASE

Patterns	Appearance (frequency)	Patterns	Appearance (frequency)
{C}	130	{D, K}	15
{D}	43	{D, H}	28
{K}	140	{K, H}	39
{H}	131	{C, K, H}	13
{C, D}	9	{D, K, H}	10
{C, K}	34	{C, D, K, H}	13
{C, H}	31	—	—

TABLE XI. COMPUTATION OF SUPPORT FOR THE MOST FREQUENT PATTERNS

Association Rule	Compute of Support
{Kidney, Diabetes} \Rightarrow {Hypertension}	146/349 = 0.418 = (41.8%)
{Diabetes} \Rightarrow {Hypertension}	174/349 = 0.498 = (49.8%)
{Hypertension, Cardiovascular} \Rightarrow {Kidney}	171/349 = 0.489 = (48.9%)
{Hypertension, Diabetes} \Rightarrow {Kidney}	168/349 = 0.532 = (53.2%)
{Kidney, Cardiovascular} \Rightarrow {Hypertension}	165/349 = 0.472 = (47.2%)
{Diabetes} \Rightarrow {kidney}	183/349 = 0.524 = (52.4%)
{Kidney, Hypertension} \Rightarrow {Cardiovascular}	169/349 = 0.484 = (48.4%)

Before we compute the degree of confidence for the results, we need to compute the support for all groups of itemsets frequent in the results according to the predefined confidence rule.

Table XI shows the computation of the support for all the patterns explored by applying the ARA algorithm, which appeared as the highest results with a strong association rule. In Table XII, we computed the confidence for these results and verified them according to the confidence rule of the ARA algorithm. The results showed a strong association rule between the patterns.

Table XII above reveals that we established seven clear rules and discovered that two of the four patterns were the most frequent among patients who died from the Covid-19 virus, as the confidence rate exceeded 60%:

- {kidney, Diabetes} \Rightarrow {Hypertension} (66.6%).
- {Diabetes} \Rightarrow {Hypertension} (65.1%).

TABLE XII. COMPUTATION OF CONFIDENCE FOR THE MOST FREQUENT PATTERNS

Association Rule	Compute of Confidence
{Kidney, Diabetes} \implies {Hypertension}	10/15 =0.651 = (66.6%)
{Diabetes} \implies {Hypertension}	28/43 =0.651 = (65.1%)
{Hypertension, Cardiovascular} \implies {Kidney}	13/31 =0.419 = (41.9%)
{Hypertension, Diabetes} \implies {Kidney}	10/29 =0.357 = (35.7%)
{Kidney, Cardiovascular} \implies {Hypertension}	13/34 =0.382 = (38.2%)
{Diabetes} \implies {kidney}	15/43 =0.348 = (34.8%)
{Kidney, Hypertension} \implies {Cardiovascular}	13/39 =0.333 = (33.3%)

Through the previous two patterns with a high degree of confidence, we noted that hypertension and diabetes were common and present in both patterns. In the first reading of these results among the most frequent patterns, we found that the deceased patients who had chronic diseases such as kidney disease and diabetes also had hypertension with a high confidence level of 66.6%. We also found that the deceased patients who had diabetes also had hypertension with a high confidence level of 65.1%.

This confirms that chronic diseases have significantly impacted the incidence of death among those infected with COVID-19, especially when we found that there were deceased patients who had more than one chronic disease or more. In this case, we found that there was a strong correlation between the diseases of the kidney and diabetes and the emergence of hypertension with a confidence degree of 66.6%.

In the second degree, diabetes led to the emergence of hypertension disease with a confidence level of 65.1%. In other words, the deceased patients who were suffering from kidney disease and diabetes together were also likely to have high blood pressure. In degrees less than 50% confidence, we found that the emergence of hypertension and cardiovascular diseases led to the emergence of kidney disease.

B. Weka Results

After the database has been converted and modified to suit the settings of the Weka program. The program was implemented through the (Associate) tab, and some modifications were made to the graphical algorithm options by opening the algorithm options window (Apriori) to implement and show the algorithm results in the program in Fig. 13, The following is a display of the top three results:

- kidney=**K** Diabetes=**D** \implies Hypertension=**H** (0.67).
- Diabetes=**D** \implies Hypertension=**H** (0.65).
- Hypertension=**H** Cardiovascular=**C** \implies kidney=**K** (0.42).
- kidney=**K** Cardiovascular=**C** \implies Hypertension=**H** (0.38).
- Hypertension=**H** Diabetes=**D** \implies kidney=**K** (0.36).
- Diabetes=**D** \implies kidney=**K** (0.35).

- kidney=**K** Hypertension=**H** \implies Cardiovascular=**C** ((0.33).

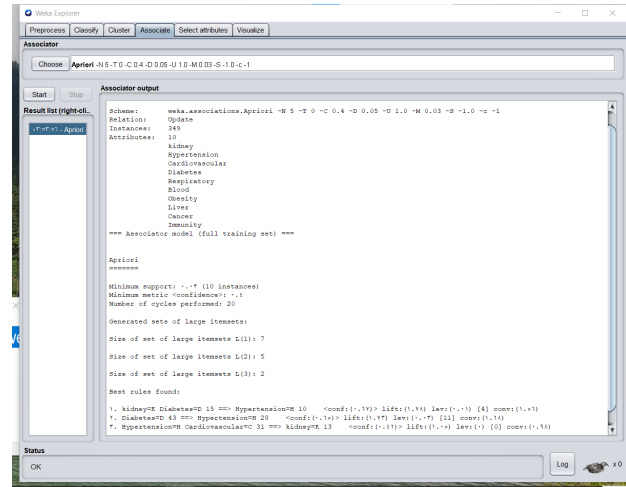


Fig. 13. Program outputs of Weka.

It is noted in the Weka results that it rounds the results to the nearest number, and thus there is a very slight difference in some numbers. They indicate that the pattern set C, D, H, K is the most frequent pattern in the database of COVID-19 deaths. They are cardiovascular - disease - diabetes - hypertension - kidney disease

The occurrences of each of the two patterns were binarily repeated according to the following numbers:

- Cardiovascular with diabetes 9 times.
- Cardiovascular with kidney disease 30 times.
- Cardiovascular with hypertension 31 times.
- Diabetes with kidney disease 15 times.
- Diabetes with hypertension 28 times.
- Kidney disease with hypertension 38 times.

We conclude that those who had chronic diseases (kidney disease and diabetes) often had a hypertension with a high confidence level of 67%. This is if we know that the number of occurrences of kidney diseases and diabetes together in the database of the study is 15 times and that of hypertension is 10 times next to (kidney diseases and diabetes).

Also, those who had diabetes often had hypertension with a high confidence degree of 65%. Where diseases appeared and repeated 43 times, hypertension appeared 131 times, and the two diseases together diabetes appeared next to hypertension in the database 29 times. There were who carried hypertension and cardiovascular diseases and often had kidney disease, but this confidence level was less than 50%.

VIII. CONCLUSION

Discovering death patterns contributes to enhancing the capabilities of healthcare decision-makers in order to know the most frequent and prevalent chronic diseases among the

deceased from COVID-19. In this research, we proposed to use the ARA algorithm to detect death patterns from the data of deceased COVID-19 patients. A non-clinical COVID-19 database, consisting of COVID-19 death patterns was presented and analyzed. The results of manual calculations and our experimental results showed strong association rules with high confidence scores between hypertension, diabetes, cardiovascular, and kidney disease. The results are confirmed by a study of Al Mutair et al. [20] that there is an increase in deaths due to high blood pressure and diabetes. The results are also confirmed by the study of Geng et al. [18], which indicated that cardiovascular diseases and kidney diseases were frequent among the deaths from COVID-19 patients at a high rate. The study of Wang et al. [19] matches with our results and stated that there is a link between high blood pressure and heart disease among patients who died of COVID-19.

Studying the death patterns of people with COVID-19 who have chronic diseases contributes to the medical knowledge in the medical domain, as it helps to understand the most factors that affect the severity of the disease to identify patients who need more special care. Understanding these patterns helps to develop better strategies to control the epidemic by giving more attention and care and priority care to COVID-19 patients. However, this study is only limited to a small number of deceased COVID-19 patients due to challenges of obtaining their medical data from one medical institution. However, in the future, more data will be collected from different medical institutions. Also, this study can be extended to include the analysis of environmental genetic factors that may affect the risk of death with COVID-19.

REFERENCES

- [1] WHO. Coronavirus disease (covid-19) reports (2020). [Online]. Available: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports>
- [2] T. Alafif, A. M. Tehame, S. Bajaba, A. Barnawi, and S. Zia, "Machine and deep learning towards covid-19 diagnosis and treatment: survey, challenges, and future directions," *International journal of environmental research and public health*, vol. 18, no. 3, p. 1117, 2021.
- [3] MOH. Ministry of health, saudi moh reports first case of infection 2021. [Online]. Available: <https://www.moh.gov.sa/en/Ministry/MediaCenter/News/Pages/News-2020-03-02-002.aspx>
- [4] S. Yezli and A. Khan, "Covid-19 social distancing in the kingdom of saudi arabia: Bold measures in the face of political, economic, social and religious challenges," *Travel medicine and infectious disease*, vol. 37, p. 101692, 2020.
- [5] A. A. Alahmari, A. A. Khan, A. Elganainy, E. L. Almohammadi, A. M. Hakawi, A. M. Assiri, and H. A. Jokhdar, "Epidemiological and clinical features of covid-19 patients in saudi arabia," *Journal of Infection and Public Health*, vol. 14, no. 4, pp. 437–443, 2021.
- [6] A. M. Al-Khani, M. A. Khalifa, A. Almazrou, and N. Saquib, "The sars-cov-2 pandemic course in saudi arabia: A dynamic epidemiological model," *Infectious Disease Modelling*, vol. 5, pp. 766–771, 2020.
- [7] OWD. Covid-19 data explorer." our world in data. [Online]. Available: <https://ourworldindata.org/coronavirus/country/saudi-arabia>
- [8] J. A. Al-Tawfiq and Z. A. Memish, "Covid-19 in the eastern mediterranean region and saudi arabia: prevention and therapeutic strategies," *International Journal of Antimicrobial Agents*, vol. 55, no. 5, p. 105968, 2020.
- [9] S. Ebrahim and Z. Memish, "Saudi arabia's measures to curb the covid-19 outbreak: temporary suspension of the umrah pilgrimage, 2020," *J Travel Med*.
- [10] M. Hassounah, H. Raheel, and M. Alhefzi, "Digital response during the covid-19 pandemic in saudi arabia," *Journal of medical Internet research*, vol. 22, no. 9, p. e19338, 2020.
- [11] M. Barry, M. Al Amri, and Z. A. Memish, "Covid-19 in the shadows of mers-cov in the kingdom of saudi arabia," *Journal of epidemiology and global health*, vol. 10, no. 1, p. 1, 2020.
- [12] MOH. Ministry of health saudi moh records 4,211 new covid-19 cases and 5,162 new recoveries, [cited 31 january 2022]. [Online]. Available: <https://www.moh.gov.sa/en/Ministry/MediaCenter/News/Pages/News-2022-01-31-006.aspx>.
- [13] MoH. Kingdom of saudi arabia ministry of health. covid19 command and control center ccc. the national health emergency operation center december,2022. [Online]. Available: <https://COVID19.moh.gov.sa>
- [14] R. Agrawal, T. Imieliński, and A. Swami, "Mining association rules between sets of items in large databases," in *Proceedings of the 1993 ACM SIGMOD international conference on Management of data*, 1993, pp. 207–216.
- [15] K. E. Heraguemi, N. Kamel, and H. Drias, "Association rule mining based on bat algorithm," in *Bio-Inspired Computing-Theories and Applications*. Springer, 2014, pp. 182–186.
- [16] T. Alafif, A. Etaiwi, Y. Hawsawi, A. Alrefaei, A. Albassam, and H. Althobaiti, "Discovid: discovering patterns of covid-19 infection from recovered patients: a case study in saudi arabia," *International Journal of Information Technology*, vol. 14, no. 6, pp. 2825–2838, 2022.
- [17] A. A. Robert, A. Al Saeed, and M. A. Al Dawish, "Covid-19 among people with diabetes mellitus in saudi arabia: Current situation and new perspectives," *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, vol. 15, no. 5, p. 102231, 2021.
- [18] J. Geng, X. Yu, H. Bao, Z. Feng, X. Yuan, J. Zhang, X. Chen, Y. Chen, C. Li, and H. Yu, "Chronic diseases as a predictor for severity and mortality of covid-19: a systematic review with cumulative meta-analysis," *Frontiers in Medicine*, p. 1442, 2021.
- [19] X. Wang, X. Fang, Z. Cai, X. Wu, X. Gao, J. Min, and F. Wang, "Comorbid chronic diseases and acute organ injuries are strongly correlated with disease severity and mortality among covid-19 patients: a systemic review and meta-analysis," *Research*, vol. 2020, 2020.
- [20] A. Al Mutair, A. Al Mutairi, S. Alhumaid, S. Maaz Abdullah, A. R. Zia Zaidi, A. A. Rabaan, and A. Al-Omari, "Examining and investigating the impact of demographic characteristics and chronic diseases on mortality of covid-19: Retrospective study," *PLoS one*, vol. 16, no. 9, p. e0257131, 2021.
- [21] T. Alafif, R. Alotaibi, A. Albassam, and A. Almudhayyani, "On the prediction of isolation, release, and decease states for covid-19 patients: A case study in south korea," *ISA transactions*, vol. 124, pp. 191–196, 2022.
- [22] T. Alafif, A. M. Tehame, S. Bajaba, A. Barnawi, and S. Zia, "Machine and deep learning towards covid-19 diagnosis and treatment: survey, challenges, and future directions," *International journal of environmental research and public health*, vol. 18, no. 3, p. 1117, 2021.
- [23] H. Wu, Z. Lu, L. Pan, R. Xu, and W. Jiang, "An improved apriori-based algorithm for association rules mining," in *2009 Sixth International Conference on Fuzzy Systems and Knowledge Discovery*, vol. 2, 2009, pp. 51–55.
- [24] F. H. AL-Zawaidah, Y. H. Jbara, and A. Marwan, "An improved algorithm for mining association rules in large databases," *World of Computer science and information technology journal*, vol. 1, no. 7, pp. 311–316, 2011.
- [25] M. Kaur, U. Garg, and S. Kaur, "Advanced eclat algorithm for frequent itemsets generation," *International Journal of Applied Engineering Research*, vol. 10, no. 9, pp. 23 263–23 279, 2015.
- [26] R. Zhong and H. Wang, "Research of commonly used association rules mining algorithm in data mining," in *2011 International Conference on Internet Computing and Information Services*. IEEE, 2011, pp. 219–222.
- [27] S. Rao and P. Gupta, "Implementing improved algorithm over apriori data mining association rule algorithm 1," 2012.
- [28] J. Singh, H. Ram, and D. J. Sodhi, "Improving efficiency of apriori algorithm using transaction reduction," *International Journal of Scientific and Research Publications*, vol. 3, no. 1, pp. 1–4, 2013.
- [29] J. Yabing, "Research of an improved apriori algorithm in data mining association rules," *International Journal of Computer and Communication Engineering*, vol. 2, no. 1, p. 25, 2013.

- [30] T. A. Kumbhare and S. V. Chobe, "An overview of association rule mining algorithms," *International Journal of Computer Science and Information Technologies*, vol. 5, no. 1, pp. 927–930, 2014.
- [31] K. Vanitha, "Using hash based apriori algorithm to reduce the candidate 2-itemsets for mining association rule," *Journal of Global Research in computer science*, vol. 2, no. 5, pp. 78–80, 2011.
- [32] K. Dahdouh, A. Dakkak, L. Oughdir, and A. Ibriz, "Association rules mining method of big data for e-learning recommendation engine," in *International Conference on Advanced Intelligent Systems for Sustainable Development*. Springer, 2018, pp. 477–491.
- [33] C. Li, N. Ding, G. Zhang, and L. Li, "Association analysis of serial cases based on apriori algorithm," pp. 137–140, 2019.
- [34] Waikato. University of waikato, new zealand, web site:. [Online]. Available: <https://www.cs.waikato.ac.nz/ml/weka/index.html>
- [35] I. H. Witten and E. Frank, "Data mining: practical machine learning tools and techniques with java implementations," *Acm Sigmod Record*, vol. 31, no. 1, pp. 76–77, 2002.
- [36] G. Holmes, A. Donkin, and I. H. Witten, "Weka: A machine learning workbench," in *Proceedings of ANZIIS'94-Australian New Zealand Intelligent Information Systems Conference*. IEEE, 1994, pp. 357–361.

Highly Accurate Deep Learning Model for Olive Leaf Disease Classification: A Study in Tacna-Perú

Erbert F. Osco-Mamani¹, Israel N. Chaparro-Cruz²
Department of Computer Science and Systems Engineering^{1,2}
Universidad Nacional Jorge Basadre Grohmann, Tacna, Perú^{1,2}

Abstract—Deep learning applied to computer vision has different applications in agriculture, medicine, marketing, meteorology, etc. In agriculture, plant diseases can cause significant yield and quality losses. The treatment of these diseases depends on accurate and rapid classification. Olive leaf diseases are a problem that threatens the crop quality of olive growers. The objective of this work was to classify olive leaf diseases with Deep Learning in olive crops of the La Yarada-Los Palos area in the Tacna region, Peru. Disease classification is a critical task, nevertheless, for the most common diseases in the region: virosis, fumagina, and nutritional deficiencies, there is no dataset to train deep learning models. Due to the latter, a novel dataset of RGB olive leaf images is elaborated and published. Then, an extensive comparative experimental study was conducted using all possible configurations of Learning from Scratch, Transfer Learning, Fine-Tuning, and Data Augmentation state-of-the-art methods to train a modified VGG16 architecture for the classification of Olive Leaf Diseases. It was demonstrated experimentally: (i) The ineffectiveness of Data Augmentation when the model Learning from Scratch, (ii) A high improvement by using Transfer Learning vs Learning from Scratch, (iii) Similar performance using Transfer Learning vs Transfer Learning + Fine-Tuning vs Transfer Learning + Data Augmentation, and (iv) Very high improvement using Transfer Learning + Fine-Tuning + Data Augmentation. This led us to a Deep Learning Model with an accuracy of 100%, 99.93%, and 100% in the training, validation, and test sets and F1-Score on the validation set of 1, 0.9901, and 0.9899 in the Nutritional Deficiencies, Fumagina, and Virosis olive leaf diseases respectively. Replication of the results is ensured by publishing the novel dataset and the final model on GitHub.

Keywords—Olive; leaf diseases; disease classification; deep learning; data augmentation; transfer learning; fine-tuning; VGG16

I. INTRODUCTION

Tacna is the leading nationally producer of olives [1]. Peru is the second-largest exporter of olives and third-largest exporter of olive oil in South America [1]. With over 22,897 hectares under olive cultivation and an average yield of 7,995 kg/ha, olive production and processing is critical to the local and regional economy [2]. Tacna accounts for 81.4% of the olive area that exists in Peru [3]. Despite its importance, olive cultivation is still managed traditionally and the use of data is incipient. Pests and diseases such as *Orthezia Olivicola* and fruit borer can significantly reduce the number of fruits per harvest [4]. This and other problems will be exacerbated by impending climate change [5].

The problems derived from pests and diseases affect the production and the production cycle as they affect leaves, flowers, and fruits and can initiate neighboring cycles [6]. In addition, the scarcity of diagnostic tools in underdeveloped

countries has a devastating impact on their development and quality of life. Therefore, there is an urgent need to automatically classify plant diseases with affordable and easy-to-use solutions.

To reduce olive harvesting diseases, it is necessary to create and modernize technologies for efficient productivity. Adequate and fast classification of olive leaf diseases could prevent quality and crop losses. There is research related to the classification of olive leaf diseases [7] with datasets collected in countries such as Saudi Arabia [8], [9] or Turkey [10], [11], [12]. However, for the most common diseases in the region of Tacna-Perú: virosis, fumagina, and nutritional deficiencies, there is no dataset to train computer vision models for classification.

Deep Learning has become the artificial intelligence method by excellence for solving a variety of problems, including those related to computer vision [13]. The area of computer vision encompasses a large number of tasks such as segmentation, detection, and classification, among which are those related to diseases. In disease problems, Deep Learning has been shown to be state-of-the-art in both agricultural and medical applications using CNN architectures [14], [15], [16]. Deep Learning models learn directly from data and require large datasets to obtain good accuracies. To avoid the latter, some techniques have been well proven to obtain models with better results, such as Data Augmentation [17], [18], [19], Transfer Learning [20], [21], [22], and Fine-Tuning [23], [24]. But there is no study on the impact of each and combination of these techniques.

In this context, CNNs such as the well-known VGG16 architecture [25] could be trained to classify olive leaf diseases. VGG16 model was the winner of the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) [26], which consists of classifying 1.2 million labeled images into a hundred classes using a 133GB dataset. The model already trained on this gigantic dataset is available, and through techniques such as Transfer Learning and Fine-Tuning, one can take advantage of feature extraction to train another classifier, if the new task and dataset have similarity to the initial task and dataset.

In this paper, our main contributions are: (i) A novel and public olive leaf disease classification dataset for the classification of virosis, fumagina, and nutritional deficiencies diseases that affect olive harvests in Tacna-Perú. (ii) Conduct an extensive comparative experimental study using all possible configurations of Data Augmentation, Transfer Learning, and Fine-Tuning techniques to train a modified VGG16 architecture for olive leaf disease classification. (iii) Obtain and publish a Highly Accurate Deep Learning model trained using

Data Augmentation, Transfer Learning, and Fine-Tuning that solves the olive leaf disease classification task.

The rest of this paper is organized as follows: Section II provides the necessary background to understand the work. Section III presents the novel olive leaf disease classification dataset. Section IV defines the materials and methods used to carry out the experiments. Section V sets the experimental setup and pipeline of experiments. Section VI presents the experimental results and performance analysis of them. Section VII presents the discussion. Finally, Section VIII provides the conclusion and further work.

II. BACKGROUND

A. Deep Learning

Deep Learning is a sub-field of machine learning that consists of using multiple layers of non-linear processing to learn data representations with multiple levels of abstraction [27].

B. CNNs

Convolutional Neural Networks (CNNs) are designed to process multiple types of data, especially two-dimensional images, and are directly inspired by the visual cortex of the brain. In the visual cortex, there is a hierarchy of basic cells: simple and complex [28]. A CNN can extract the image features without the need for this process to be performed by hand as before 2011, finally, adding a neural network at the end allows the classification task to be performed.

Fig. 1 shows a typical CNN architecture. Feature extraction consists of convolution layers (C1, C2, C3) and pooling layers (P1, P2). The classification consists of fully connected layers (FC) and an output layer [29].

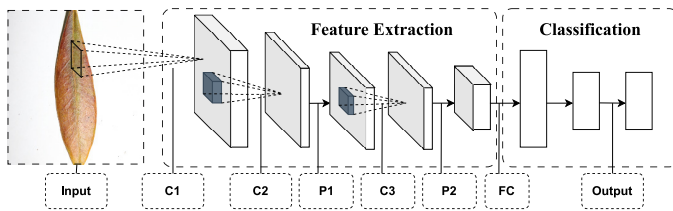


Fig. 1. Typical CNN architecture, based on [29].

C. VGG16 Architecture

The VGG16 architecture [25] was proposed by Simonyan and Sisserman in 2014, they developed the convolutional neural network in the Oxford Visual Geometry Group (VGG16). This architecture is constituted of 13 convolutional layers, each group of convolution layers is followed by a max pooling layer composing the feature extraction or base model, which is followed by three fully connected layers as classification or top model, hence the name includes 16. Finally, a softmax layer is added to the classifier.

VGG16 is a classification model which is able to classify 14M images of 1000 different categories (Imagenet dataset - ILSVRC 2014) with 92.7% accuracy. Despite the existence of more recent models, the simplicity of this architecture makes

it perfect for applying techniques such as Transfer Learning and Fine-Tuning. The network (model and trained weights) is available in Keras. Fig. 2 shows the architecture of VGG16.

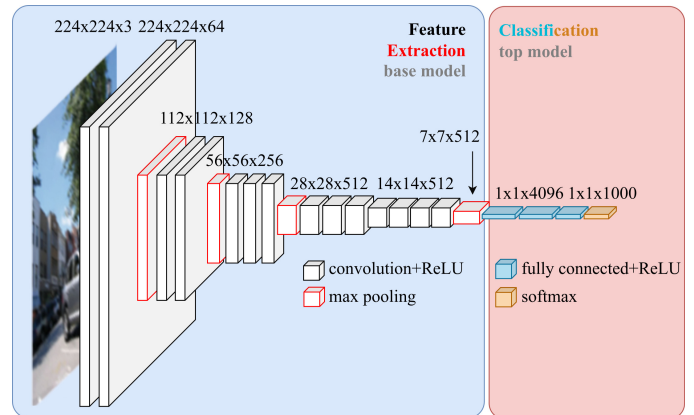


Fig. 2. VGG-16 architecture overview, based on [30].

D. Underfitting and Overfitting

Underfitting is a scenario where the model is not adjusted to the training dataset, causing a high error rate. In training across epochs, this happens when the model is not sufficiently capable of modeling the relationship between input and output.

When a model performs well with training data, it is necessary to consider a control group to ensure that the model performs well with data with which it has not been trained. That control group is better known as the validation set. It is possible for the model to perform very well on the training data but poorly on the validation data, which is known as overfitting (lack of generalization) [13].

E. Transfer Learning

Transfer Learning is a machine learning technique that seeks to leverage an already trained model for one task (pre-trained) into another [13], based on the assumption that the data with which the model has been initially trained are in the same feature space and have the same distribution as the new data. In cases where the latter is true, one could use the freeze feature extraction base of a pre-trained model and link it to a new classification network (with weights randomly initialized) to train only the latter and obtain good results.

This process results in a fast convergence of the model, either due to the constraints of feature extraction that have been frozen or due to the good compatibility of the dataset domains. When this is not sufficient, it is also necessary to train the feature extraction base, through a technique called Fine-Tuning.

F. Fine-Tuning

Fine-tuning is a technique that is applied over Transfer Learning to finish adjusting the model to the new dataset [13], for that purpose the feature extraction layers are unfrozen and the complete model is trained. It is important that this process is carried out after having trained the model classification layers and selected the best epoch for Fine-Tuning.

Fine-tuning allows the model to improve the feature extraction to obtain a better result in the classification task. The need to use Fine-Tuning, indicating that the feature space and distribution of the datasets (pre-trained and new) are not exactly the same, the distance between the feature space and distribution of datasets can be reflected in the model improvement.

G. Data Augmentation

Data augmentation is the generation of synthetic data by perturbing the original data [13], this artificially allows to have a larger amount of data and avoids biases related to the perturbations performed. The use of this technique reduces the need for large amounts of data to train Deep Learning models. In addition, when data are difficult to obtain, it is possible to use Data Augmentation as a regularization technique to generalize better.

Data augmentation also prevents model overfitting by adding diversity and some randomness to the dataset, which allows models to be trained over more epochs. The most used deformations or transformations are related to rotation, zoom, width and height shift, and horizontal and vertical flips. It is very important to use these transformations taking into consideration the problem domain in order not to obtain undesired results.

H. Learning from Scratch

Learning from Scratch is the process of training a model with randomly initialized parameters or weights. This is the usual process when pre-trained models are not used. When pre-trained models are used, the weights are not randomly initialized, but rather those that have been fitted to a pre-training dataset are loaded. This process may be necessary in cases where a pre-trained model compatible with the dataset being worked with is not encountered.

I. Confusion Matrix

A confusion Matrix is an NxN table that summarizes the number of correct and incorrect predictions that a classification model made, where N is the number of classes. The following values are calculated: *TP* are True Positive, *FP* False Positive, *FN* False Negative, and *TN* are True Negative. Taking these values into account, different classification metrics can be calculated.

J. Classification Metrics

1) *Accuracy*: Number of correct predictions over the total number of predictions.

2) *Loss*: It is a measure of how much error is being made in the prediction. Specifically, this value will be 0 when the prediction is equal to the desired output. For multiclass classification, the categorical cross-entropy loss is often used, the definition of which is as follows:

$$L_{CE} = - \sum_{i=1}^n t_i \log(p_i), \text{ for } n \text{ classes} \quad (1)$$

where t_i is the truth label and p_i is the Softmax probability for the i^{th} class.

3) *Precision*: Try to answer the following question: What proportion of positive identifications was correct? In the case of multi-class classification, one metric per class is obtained; is defined as follows:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2)$$

4) *Recall*: Try to answer the following question: What proportion of real positives was correctly identified? In the case of multi-class classification, one metric per class is obtained; is defined as follows:

$$\text{Recall} = \frac{TP}{TP + FN} \quad (3)$$

5) *F1-Score*: It is a single score defined as the harmonic mean of precision and recall. In the case of multi-class classification, one metric per class is obtained; is defined as follows:

$$F_1 = 2 \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} = \frac{2TP}{2TP + FP + FN} \quad (4)$$

III. NOVEL OLIVE LEAF DISEASE CLASSIFICATION DATASET

A. Data Acquisition and Processing

Images of olive leaves found in crops at La Yarada-Los Palos in the Tacna region of Peru were collected in August of 2019 with the next most common diseases in the region.

1) *Virosis*: The name is descriptive of this disease of olive foliage due to the curved or sickle shape shown by the leaf affected with this disease. The disease is common in all commercial varieties of olives grown in South America [31].

2) *Fumagina*: It is a species of fungi, blackish in color, that covers plant tissues as a layer of soot, which hinders photosynthesis so that the affected olive trees see their productive capacity diminished. This layer comes off when passing the finger over the affected parts [32].

3) *Nutritional Deficiency*: Leaves with abnormal colorations at the terminal end of the leaves such as phosphorus and potassium [33].

The RGB images of 3984 x 2656 px were taken using a Canon EOS Rebel T6i 24.2MPX camera and fixed configurations (55 mm focal distance, 1/200s aperture, ISO-400). To highlight the characteristics of the leaves and keep the image focused and stable, a device was used to stabilize and focus the mobile device at a height of 30 cm. Each leaf was placed showing its frontal area on a blank sheet of paper.

One way to distinguish between diseased leaves is to focus on the level of green, brown, and yellow color. However, this is a criterion that does not apply to all cases, and an agronomic expert is usually required. Fig. 3 shows samples of the images taken.

The procedure for the construction of the olive leaf diseases dataset includes three phases. (i) First, images of olive trees



Fig. 3. Samples of images and classes of dataset.

are captured, selecting three diseases according to the study. (ii) Second, after the images are collected, they are given to an agronomist specialist expert in olive disease images. (iii) Finally, the specialist filtered the images and manually labeled them according to their characteristics.

The total images acquired resulted in 773 useful images, of which 258 are nutritional deficiency disease, 257 are fumagina, and 258 are virosis. In order to have class balance, a 77:19:3 split was performed to form the training, validation, and test sets. Table I specifies the final number of images per class and dataset.

TABLE I. NOVEL OLIVE LEAF DISEASE CLASSIFICATION DATASET SIZE

Category	Training Set	Validation Set	Test Set	Total
Deficiencies	200	50	8	258
Fumagina	200	50	7	257
Virosis	200	50	8	258
Total	600	150	23	773

Finally, the novel olive leaf disease dataset is made available in our GitHub for further use and evaluation.

IV. MATERIALS AND METHOD

A. Experimental Platform

For the design of the experiments and their execution, the following computer hardware was made available to us through a Google Colab Pro subscription: GPU 1x Nvidia Tesla T4 with 15GB of RAM and CPU 1x Xeon Processors @2.3Ghz with 12.7 GB of RAM.

Keras, one of the main Deep Learning APIs written in Python, was used in this study. It is compatible with multiple back-end neural network computation engines and runs on the TensorFlow machine learning platform. Versions were: Keras 2.11.0 and Tensorflow 2.11.0 as the backend in a Python version 3.8 environment. Each epoch of training takes 12 seconds and the total amount of time for 10 runs of all experiments was 33.3 hours.

B. Deep Learning Architecture

The selected Deep Learning architecture was VGG16 [25], chosen because it extracts features at a low level by using a smaller kernel size, demonstrating a good feature extraction capability for image classification. As a Deep Learning classification architecture, it is composed of 2 sub-models: feature extraction (base) and classification (top). Feature extraction was carried out by a convolutional neural network. Table II

TABLE II. VGG16 BASE MODEL ARCHITECTURE

Layer (type)	Output Shape	Number of Parameters
input_1 (InputLayer)	224 x 224 x 3	0
block1_conv1 (Conv2D)	224 x 224 x 64	1,792
block1_conv2 (Conv2D)	224 x 224 x 64	36,928
block1_pool (MaxPooling2D)	112 x 112 x 64	0
block2_conv1 (Conv2D)	112 x 112 x 128	73,856
block2_conv2 (Conv2D)	112 x 112 x 128	147,584
block2_pool (MaxPooling2D)	56 x 56 x 128	0
block3_conv1 (Conv2D)	56 x 56 x 256	295,168
block3_conv2 (Conv2D)	56 x 56 x 256	590,080
block3_conv3 (Conv2D)	56 x 56 x 256	590,080
block3_pool (MaxPooling2D)	28 x 28 x 256	0
block4_conv1 (Conv2D)	28 x 28 x 512	1,180,160
block4_conv2 (Conv2D)	28 x 28 x 512	2,359,808
block4_conv3 (Conv2D)	28 x 28 x 512	2,359,808
block4_pool (MaxPooling2D)	14 x 14 x 512	0
block5_conv1 (Conv2D)	14 x 14 x 512	2,359,808
block5_conv2 (Conv2D)	14 x 14 x 512	2,359,808
block5_conv3 (Conv2D)	14 x 14 x 512	2,359,808
block5_pool (MaxPooling2D)	7 x 7 x 512	0
Total		14,714,688

presents the architecture of the feature extraction basis of VGG16 and the number of parameters.

Classification was carried out by a dense neural network. Original VGG16 architecture [25] uses three sequential dense layers as classifier (top-model), that allow mapping the model input to 1000 classes (because it was conceived for Imagenet). The architecture was modified by replacing the top-model with a max-pooling layer (to reduce the dimensionality of the extracted features) connected to a dense layer that maps the model input to 3 classes. Table III presents our modified VGG16 architecture and the number of parameters.

It was hypothesized that only one dense layer is sufficient to map features to classes, due to the complexity of our dataset compared with ImageNet, and perform the problem task. This hypothesis is proven through our experiments.

TABLE III. OUR MODIFIED VGG16 ARCHITECTURE

Layer (component)	Output Shape	Number of Parameters
input_1 (Input Layer)	224 x 224 x 3	0
vgg16_base_model	7 x 7 x 512	14,714,688
Base Model (Feature Extraction)		14,714,688
GlobalAveragePooling2D (new)	512	0
Dense (new)	3	1,539
Top Model (Classification)		1,539
Total		14,716,227

V. EXPERIMENTAL SETUP

Fig. 4 presents the complete pipeline of every experiment carried out which is detailed below.

The following comparative objectives were considered in order to achieve the objective of the study:

- 1) To determine the effect of Data Augmentation.

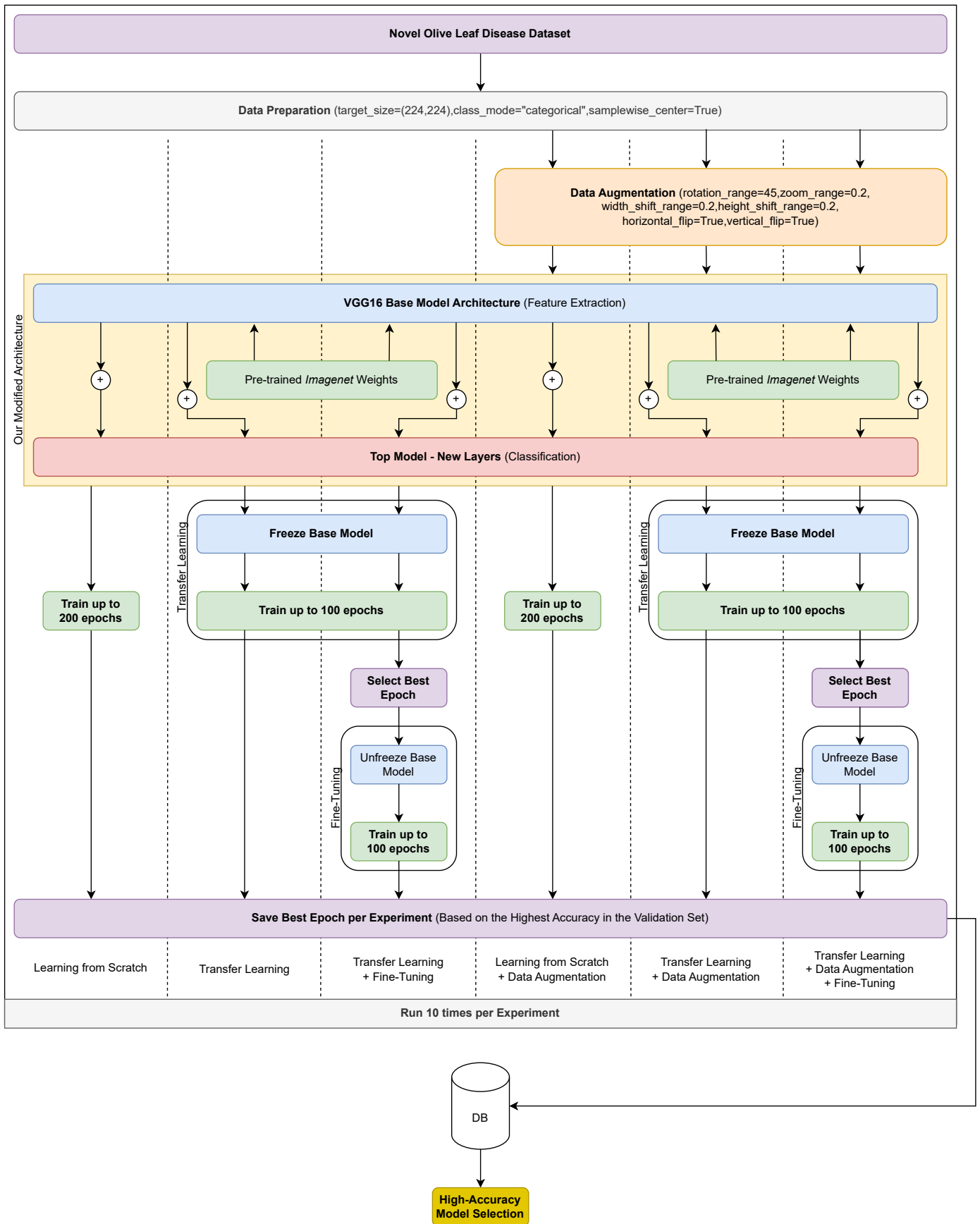


Fig. 4. Pipeline of experiments.

- 2) To determine the effect of Transfer Learning.
- 3) To determine the effect of Fine-Tuning.

For data preparation, the RGB images in the dataset were resized to the input size of the selected architecture (224,224) and each sample was scaled to mean 0. In addition, the labels of each of the images are categorized as a dummy variable to match the output size of the modified architecture. This was done in all experiments.

For Data Augmentation, which is used in some experiments, the images sent to the model for training are artificially enhanced through the following random transformations:

- 1) 0-45 degree random rotation.
- 2) 0.8-1.2 zoom.
- 3) 0-0.2 of total width shift.
- 4) 0-0.2 of total height shift.
- 5) Horizontal flip.
- 6) Vertical flip.

We describe this configuration as “Aggressive Data Augmentation” because it achieves a transformation that rotates the image from 0° to 359° with a variety of zooms, shifts, and flips.

Finally, for training models, the loss was Categorical Cross Entropy and the optimizer was ADAM [34].

In order to obtain more trustworthy results, each of the following experiments of all possible setups of Transfer Learning, Fine-Tuning, and Data Augmentation was run 10 times:

A. Learning from Scratch

For this experiment, the modified architecture was trained up to 200 epochs and the best-performing epoch is selected.

B. Learning from Scratch + Data Augmentation

For this experiment, the images sent to the model for training were artificially enhanced through Data Augmentation. Then, the modified architecture was trained up to 200 epochs and the best-performing epoch was selected.

C. Transfer Learning

For this experiment, pre-trained Imagenet weights were loaded into the base model of our modified architecture. Then, the base model was frozen and only the top model was trained up to 100 epochs.

D. Transfer Learning + Fine-Tuning

For this experiment, the best epoch of the Transfer Learning experiment is taken to unfreeze the base model and train the whole architecture up to 100 epochs.

E. Transfer Learning + Data Augmentation

For this experiment, the images sent to the model for training were artificially enhanced through Data Augmentation, and pre-trained Imagenet weights were loaded into the base model of our modified architecture. Then, the base model was frozen and only the top model was trained up to 100 epochs.

F. Transfer Learning + Data Augmentation + Finetuning

For this experiment, the best epoch of the Transfer Learning + Data Augmentation experiment is taken to unfreeze the base model and train the whole architecture up to 100 epochs.

For all experiments, the evaluation for best epoch selection was made taking into account the accuracy in the validation set. Because the experiment was run 10 times, each selection of the best model was stored in a database in order to finally select the highest accurate model.

All numbers of epochs frame the underfitting and overfitting processes of model training, so in none of the cases was the number of training epochs insufficient.

VI. EXPERIMENTAL RESULTS

A. Impact of Data Augmentation

From Fig. 5, the accuracy in the training set and validation set over 200 epochs is presented with an interquartile range on a model trained Learning from Scratch. The best validation accuracy was 0.86 with a validation loss of 0.558051.

From Fig. 6, the accuracy in the training set and validation set over 200 epochs is presented with an interquartile range on a model trained Learning from Scratch. The best validation accuracy was 0.8 with a validation loss of 0.641917.

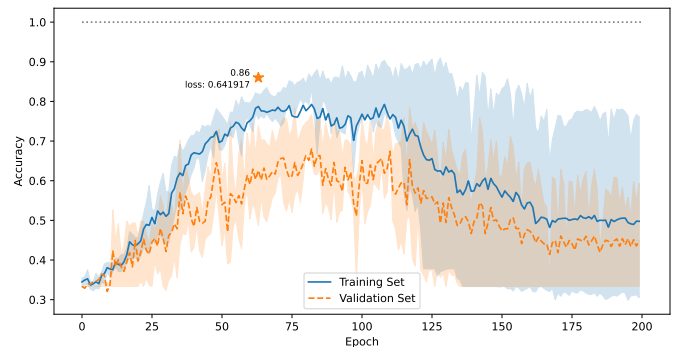


Fig. 5. Accuracy of learning from scratch.

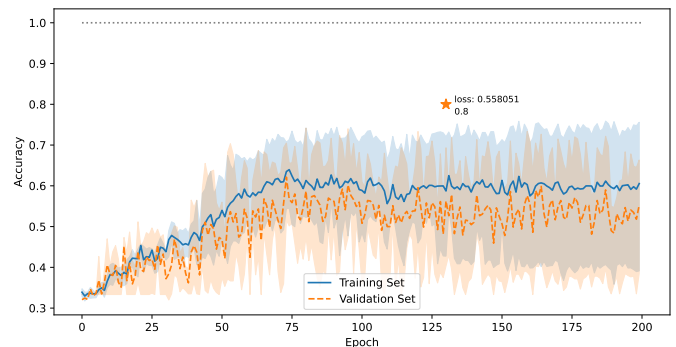


Fig. 6. Accuracy of learning from scratch + data augmentation.

Although the accuracy was higher without using data augmentation, the loss was lower when using data augmentation, which means that without data augmentation the model was

less sure of its prediction. On the other hand, without the use of data augmentation, the best validation accuracy was reached in 63 epochs, with the use of Data Augmentation it was reached in 130 epochs, which shows that Data Augmentation allows for avoiding overfitting in training. Moreover, both models achieve similar performances.

It is worth mentioning that, while the model without data augmentation remains stationary, the model trained without data augmentation decreases in validation value. This denotes that Learning from Scratch was not able to update the model in such a way that it adjusts to the variability of the data caused by our aggressive data augmentation.

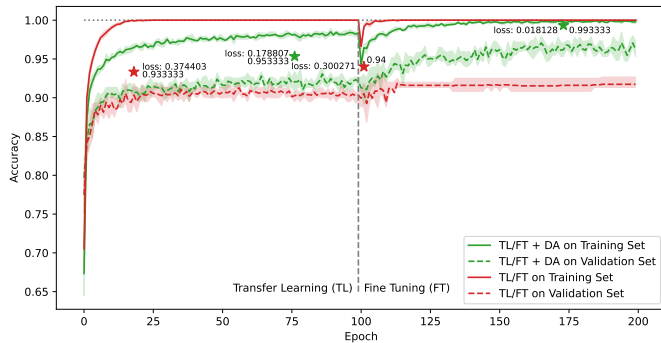


Fig. 7. Accuracy of transfer learning vs transfer learning with data augmentation.

From Fig. 7, the accuracy in the training set and validation set over 200 epochs is presented with an interquartile range on a model trained Transfer Learning (left side), Transfer Learning + Data Augmentation (left side), Transfer Learning + Fine-Tuning (right side), and Transfer Learning + Fine-Tuning + Data Augmentation (right side).

For Transfer Learning the best validation accuracy was 0.93333 with a validation loss of 0.374403 and was reached at epoch 18, for Transfer Learning + Data Augmentation the best validation accuracy was 0.953333 with a validation loss of 0.178807 and was reached at epoch 76, for Transfer Learning + Fine-Tuning the best validation accuracy was 0.94 with a validation loss of 0.300271 and was reached at epoch 101, and for Transfer Learning + Fine-Tuning + Data Augmentation the best validation accuracy was 0.993333 with a validation loss of 0.018128 and was reached at epoch 176.

The accuracy was higher in the validation set using Data Augmentation in Transfer Learning and Transfer Learning + Fine Tuning. On the other hand, without the use of Data Augmentation, the best validation accuracy was reached before vs with the use of Data Augmentation. Moreover, both models achieve similar performances.

Data Augmentation when applied improves accuracy by 2% in Transfer Learning and 5% in Transfer Learning + Fine Tuning. This difference is consistent with the fact that Data Augmentation generates a greater variety of input data and Fine-Tuning is able to take advantage of this variety because it has unfrozen feature extraction layers.

From Fig. 8, the loss in the training set and validation set over 200 epochs is presented with an interquartile range on a

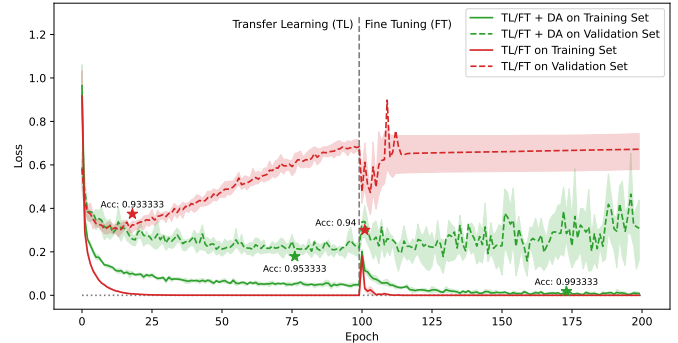


Fig. 8. Loss of transfer learning vs transfer learning with data augmentation.

model trained Transfer Learning (left side), Transfer Learning + Data Augmentation (left side), Transfer Learning + Fine-Tuning (right side), and Transfer Learning + Fine-Tuning + Data Augmentation (right side).

It can be observed that training using Data Augmentation avoids the rapid overfitting to which the model is prone due to the low variability of the data.

Furthermore, the use of Data Augmentation in Fine-Tuning did not show an improvement over 1% when Data Augmentation was not applied, where the model tends even more quickly to fall into overfitting than in Transfer Learning.

B. Impact of Transfer Learning

Fig. 9 shows the average accuracy over all experiments. The difference between using any configuration of Learning from Scratch vs any configuration using Transfer Learning was noticeable in the graph so that the lines do not overlap at any point. The use of transfer learning, regardless of also using data augmentation, improved the accuracy by at least 10% on average.

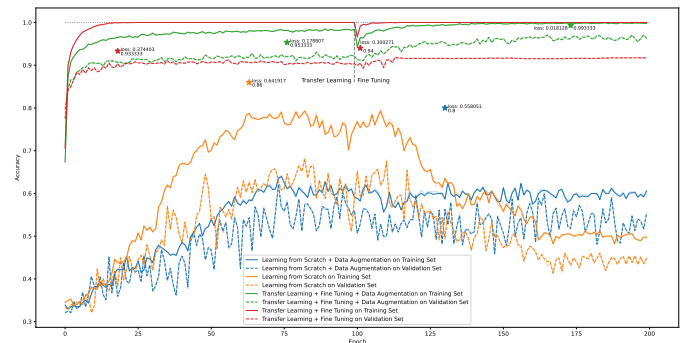


Fig. 9. Average accuracy of all experiments.

Fig. 11 shows the run with the best validation accuracy over all experiments. Again, the difference between using any configuration with Learning from Scratch vs any configuration using Transfer Learning is noticeable in the graph so that the lines do not overlap at any point. The use of transfer learning, regardless of also using data augmentation, improved the accuracy by at least 7.33333% on the best experiment run.

C. Impact of Fine-Tuning

From Fig. 7 (described above), for Transfer Learning (with or without Data Augmentation) the best validation accuracy was 0.953333 with a validation loss of 0.178807 and was reached at epoch 76, for Transfer Learning + Fine-Tuning the best validation accuracy was 0.993333 with a validation loss of 0.018128 and was reached at epoch 173.

Fine-Tuning when applied improves accuracy by less than 1% in Transfer Learning and 4% in Transfer Learning + Data Augmentation. This difference is consistent with the fact that some specific-domain problems need an adjustment in the feature extraction layers of the model to improve the performance.

Additionally, without Data Augmentation the difference between Transfer Learning and Transfer Learning + Fine-Tuning was similar.

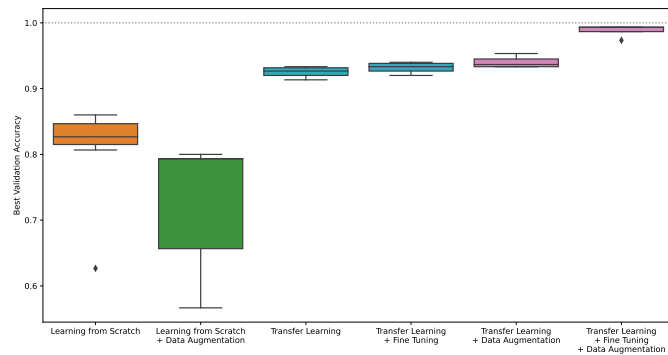


Fig. 10. Boxplot of validation accuracy.

From Fig. 10, all experiments are presented in a boxplot that resumes the best validation accuracies over 10 runs of each experiment. It can be observed that the use of Fine-Tuning over Transfer Learning without Data Augmentation gave similar results. Additionally, there was a clear improvement and separation between the values with respect to the combined use of Transfer Learning + Fine-Tuning + Data Augmentation. Furthermore, the range of values was lowest than in any other experiment.

D. Highly Accurate Deep Learning Model

From Fig. 11, the best run of all experiments is presented. The criterion used was to choose the model in whose: best epoch with the highest accuracy and lowest loss in the validation set has been obtained. Therefore, the model of run 0, epoch 173, loss 0.000008, accuracy 1.0, validation loss 0.018128, validation accuracy 0.993333 that made use of Transfer Learning + Fine-Tuning + Data Augmentation was selected as a Highly Accurate Deep Learning Model.

From Table IV, metrics of all experiments are presented, and the best values are highlighted in bold. In the validation set, the best result was obtained for all metrics in the combined use of Transfer Learning + Fine-Tuning + Data Augmentation, except for the interquartile range. However, taking into consideration the Q4-Q1 range this result was still the best. This demonstrates, due to the execution of 10 runs, that the result

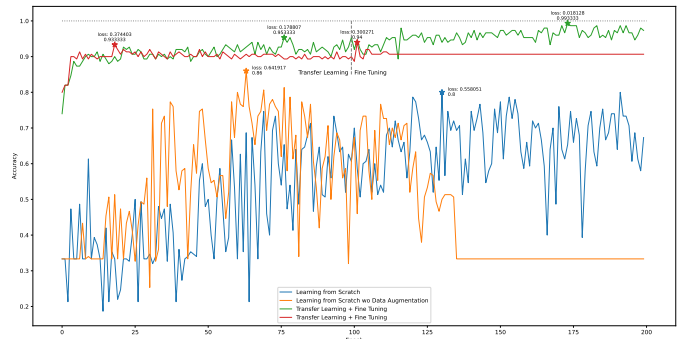


Fig. 11. Best run accuracy of all experiments.

is the product of a consistent improvement provided by the experiment setup.

Once the model has been selected, is submitted to the test set to corroborate the performance. Table V shows the confusion matrix of the selected model over training, validation, and test sets. In overall sets, only in the validation set, there is one sample that belongs to the virosis class but is classified as fumagina by the model.

Table VI shows additional metrics for each disease to be classified. The average F1-Score over all diseases was 0.9933 on the validation set. The value coincides with the average accuracy of the model in the validation set due to the fact that the dataset is balanced with respect to its classes.

Finally, the highly accurate model is made available on our GitHub for further use and evaluation.

VII. DISCUSSIONS

Regarding other studies related to the detection of olive leaf diseases [7], [8], [9], [10], [11], [12], it is not possible to make a direct comparison because those studies and the present are case studies of different olive leaf diseases in different countries. However, we agree with them that it is possible to classify olive leaf diseases using Deep Learning.

The results obtained reaffirm the findings regarding the use of Data Augmentation [17], [18], [19], Transfer Learning [20], [21], [22], and Fine-Tuning [23], [24]. In addition, our extensive experimental comparative study introduces new findings on the combination of these techniques, which are presented in the conclusions.

VIII. CONCLUSIONS

A novel and public olive leaf dataset for the classification of Virosis, Fumagina, and nutritional deficiencies diseases that affect olive harvests at La Yarada-Los Palos in Tacna Region, Perú was presented.

In addition, extensive comparative experimental studies were conducted using all possible configurations of Data Augmentation, Transfer Learning, and Fine-Tuning with the next conclusions: (i) Ineffectiveness of Data Augmentation when the model Learning from Scratch, (ii) High improvement by using Transfer Learning vs Learning from Scratch, (iii) Similar performance using Transfer Learning vs Transfer Learning

TABLE IV. METRICS OF ALL EXPERIMENTS

Set	Metric	Config:	LFS	LFS+DA	TL	TL+DA	TL+FT	TL+FT+DA
Training Set	Loss	min ↓	0.40205	0.64825	0.00002	0.03743	0.00016	0.00001
		max ↓	4.36961	2.36000	0.05210	0.07445	0.08969	0.03955
		median ↓	0.65133	0.92559	0.00060	0.05794	0.01349	0.00242
		mean ↓	1.20267	1.13607	0.00710	0.05766	0.02150	0.00854
		iqr ↓	0.63291	0.45144	0.00356	0.01647	0.01919	0.01357
	Accuracy	min ↑	0.36833	0.42000	0.98167	0.97000	0.98167	0.99500
		max ↑	0.94500	0.77000	1.00000	0.98500	1.00000	1.00000
		median ↑	0.83250	0.69750	1.00000	0.97750	0.99667	0.99833
		mean ↑	0.80983	0.64067	0.99800	0.97800	0.99450	0.99850
		iqr ↓	0.08542	0.21583	0.00000	0.00958	0.00500	0.00167
Validation Set	Loss	min ↓	0.37960	0.51447	0.30650	0.15781	0.30027	0.01813
		max ↓	1.50668	3.66409	0.51401	0.29471	1.06716	0.23049
		median ↓	0.75769	0.66646	0.39528	0.19313	0.50248	0.07109
		mean ↓	0.79619	0.99644	0.40222	0.20341	0.53828	0.07605
		iqr ↓	0.42499	0.32405	0.08068	0.03825	0.15090	0.04152
	Accuracy	min ↑	0.62667	0.56667	0.91333	0.93333	0.92000	0.97333
		max ↑	0.86000	0.80000	0.93333	0.95333	0.94000	0.99333
		median ↑	0.82667	0.79333	0.92667	0.93667	0.93333	0.99333
		mean ↑	0.81267	0.73333	0.92467	0.93933	0.93200	0.98933
		iqr ↓	0.03167	0.13667	0.01167	0.01167	0.01167	0.00667

TABLE V. CONFUSION MATRIX

		Training Set		
		Class	Deficiency	Fumagina
True Class	Deficiency	200	0	0
	Fumagina	0	200	0
	Virosis	0	0	200
	Validation Set			
	Class	Deficiency	Fumagina	Virosis
	Deficiency	50	0	0
	Fumagina	0	50	0
	Virosis	0	1	49
	Test Set			
	Class	Deficiency	Fumagina	Virosis
	Deficiency	8	0	0
	Fumagina	0	7	0
Virosis	0	0	8	
		Predicted Class		

TABLE VI. PRECISION, RECALL, AND F1-SCORE PER DISEASE

		Training Set		
		Disease	Precision	Recall
Deficiencies		1.0000	1.0000	1.0000
Fumagina		1.0000	1.0000	1.0000
Virosis		1.0000	1.0000	1.0000
		Validation Set		
Disease	Precision	Recall	F1-Score	
Deficiencies	1.0000	1.0000	1.0000	
Fumagina	0.9804	1.0000	0.9901	
Virosis	1.0000	0.9800	0.9899	
		Test Set		
Disease	Precision	Recall	F1-Score	
Deficiencies	1.0000	1.0000	1.0000	
Fumagina	1.0000	1.0000	1.0000	
Virosis	1.0000	1.0000	1.0000	

+ Fine-Tuning vs Transfer Learning + Data Augmentation, and (iv) Very high improvement using Transfer Learning + Fine-Tuning + Data Augmentation.

Finally, a highly accurate Deep Learning model (100%, 99.33%, and 100% of accuracy in the training, validation, and test set respectively) based on modified VGG16 architecture using Data Augmentation, Transfer Learning, and Fine-Tuning to solve the olive leaf disease classification task was obtained and published. F1-Score was 1 for the diseases in training and test sets, and 1, 0.9901, and 0.9899 in the Nutritional Deficiencies, Fumagina, and Virosis diseases.

Making the dataset and selected model public allows for the reproducibility of the results. Future works could deploy this trained model to a mobile or edge device for validation and agricultural use, experiment with different Data Augmentation

configurations, and compare VGG16 with other models.

ACKNOWLEDGMENT

We would like to express our appreciation and thanks to the NVIDIA Deep Learning Institute (DLI) for providing the Free Faculty Development Workshop “Fundamentals of Deep Learning” on February 18, 2022.

This research was supported by Canon, Sobrecanon, and Mining Royalties of the Universidad Nacional Jorge Basadre Grohmann - Tacna - Perú.

REFERENCES

[1] MINCETUR. Perfil de mercado y competitividad exportadora de la aceituna. <https://www.mincetur.gob.pe/wp-content/uploads/>

- documentos/comercio_exterior/plan_exportador/publicaciones/Aceituna.pdf, 2022.
- [2] DIRESA. Anuario estadístico agrario 2020 región tacna. www.agritacna.gob.pe, 2020.
- [3] H Baumann. Agraria.pe. exportaciones de la aceituna Perú. <https://agraria.pe/buscar?q=exportaciones+de+aceituna+peru>. Accessed: 2022-09-30.
- [4] SENASA. Tacna promueve control de plagas. <https://www.senasa.gob.pe/senasacontigo/tacna-senasa-promueve-control-de-plagas-del-olivo-con-metodos-no-contaminantes/>, 2021.
- [5] FAO. El trabajo de la fao sobre el cambio climático conferencia de las naciones unidas sobre el cambio climático 2019. <http://www.fao.org/3/a-i8037s.pdf>.
- [6] Martín Eloy Casilla García. Análisis de los factores que influyen en la vejería del olivo (olea europea l.) en la región tacna. 2011.
- [7] Mohamed Lachgar, Hamid Hrimech, Ali Kartit, et al. Optimization techniques in deep convolutional neuronal networks applied to olive diseases classification. *Artificial Intelligence in Agriculture*, 6:77–89, 2022.
- [8] Madallah Alruwaili, Saad Alanazi, Sameh Abd El-Ghany, and Abdulaziz Shehab. An efficient deep learning model for olive diseases detection. *International Journal of Advanced Computer Science and Applications*, 10(8), 2019.
- [9] Amel Ksibi, Manel Ayadi, Ben Othman Soufiene, Mona M Jamjoom, and Zahid Ullah. Mobires-net: A hybrid deep learning model for detecting and classifying olive leaf diseases. *Applied Sciences*, 12(20):10278, 2022.
- [10] Sinan Uğuz and Nese Uysal. Classification of olive leaf diseases using deep convolutional neural networks. *Neural computing and applications*, 33(9):4133–4149, 2021.
- [11] Hamoud H Alshammari, Ahmed I Taloba, and Osama R Shahin. Identification of olive leaf disease through optimized deep learning approach. *Alexandria Engineering Journal*, 72:213–224, 2023.
- [12] Hamoud Alshammari, Karim Gasmi, Ibtihel Ben Ltaifa, Moez Krichen, Lassaad Ben Ammar, and Mahmood A Mahmood. Olive disease classification based on vision transformer and cnn models. *Computational Intelligence and Neuroscience*, 2022, 2022.
- [13] Magnus Ekman. *Learning Deep Learning: Theory and Practice of Neural Networks, Computer Vision, NLP, and Transformers Using TensorFlow*. Addison-Wesley Professional, 2021.
- [14] Jiraporn Thomkaew and Sarun Intakosum. Improvement classification approach in tomato leaf disease using modified visual geometry group (vgg)-inceptionv3. *International Journal of Advanced Computer Science and Applications*, 13(12), 2022.
- [15] Orlando Iparraguirre-Villanueva, Victor Guevara-Ponce, Ofelia Roque Paredes, Fernando Sierra-Liñan, Joselyn Zapata-Paulini, and Michael Cabanillas-Carbonell. Convolutional neural networks with transfer learning for pneumonia detection. 2022.
- [16] Mimoun Yandouzi, Mounir Grari, Idriss Idrissi, Mohammed Boukabous, Omar Moussaoui, Mostafa Azizi, Kamal Ghoumid, and Aissa Kerkour Elmiad. Forest fires detection using deep transfer learning. *International Journal of Advanced Computer Science and Applications*, 13(8), 2022.
- [17] Hong Chun Teo, Umami Rabaah Hashim, Sabrina Ahmad, Lizawati Salahuddin, Hea Choon Ngo, and Kasturi Kanchymalay. Efficacy of the image augmentation method using cnn transfer learning in identification of timber defect. *International Journal of Advanced Computer Science and Applications*, 13(5), 2022.
- [18] Walid Al-Dhabyani, Mohammed Goma, Hussien Khaled, and Fahmy Aly. Deep learning approaches for data augmentation and classification of breast masses using ultrasound images. *Int. J. Adv. Comput. Sci. Appl*, 10(5):1–11, 2019.
- [19] Sudeepthi Govathoti, A Mallikarjuna Reddy, Deepthi Kamidi, G Balakrishna, Sri Silpa Padmanabhuni, and Pradeepini Gera. Data augmentation techniques on chilly plants to classify healthy and bacterial blight disease leaves. *International Journal of Advanced Computer Science and Applications*, 13(6), 2022.
- [20] Priyanka Jaiswal, Vijay Katkar, and SG Bhirud. Multi oral disease classification from panoramic radiograph using transfer learning and xgboost. *International Journal of Advanced Computer Science and Applications*, 13(12), 2022.
- [21] Farhana Alam, Farhana Chowdhury Tisha, Sara Anisa Rahman, Samia Sultana, Md Ahied Mahi Chowdhury, Wasif Reza Ahmed, and Mohammad Shamsul Arefin. Automated brain disease classification using transfer learning based deep learning models. *International Journal of Advanced Computer Science and Applications*, 13(9), 2022.
- [22] Jayapalan Senthil Kumar, Syahid Anuar, and Noor Hafizah Hassan. Transfer learning based performance comparison of the pre-trained deep neural networks. *International Journal of Advanced Computer Science and Applications*, 13(1), 2022.
- [23] Usman Bello Abubakar, Moussa Mahamat Boukar, and Steve Adeshina. Evaluation of parameter fine-tuning with transfer learning for osteoporosis classification in knee radiograph. *International Journal of Advanced Computer Science and Applications*, 13(8), 2022.
- [24] Fadi Alharbi, Murtada K Elbashir, Mohamad Mohammed, and Mohamed Elhafiz Mustafa. Fine-tuning pre-trained convolutional neural networks for women common cancer classification using rna-seq gene expression. *International Journal of Advanced Computer Science and Applications*, 11(11), 2020.
- [25] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.
- [26] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.
- [27] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning*. MIT press, 2016.
- [28] Seonwoo Min, Byunghan Lee, and Sungroh Yoon. Deep learning in bioinformatics. *Briefings in bioinformatics*, 18(5):851–869, 2017.
- [29] Shanwen Zhang, Wenzhun Huang, and Chuanlei Zhang. Three-channel convolutional neural networks for vegetable leaf disease recognition. *Cognitive Systems Research*, 53:31–41, 2019.
- [30] D Frossard. Vgg in tensorflow: Model and pre-trained parameters for vgg16 in tensorflow. *Department of Computer Science, University of Toronto, Toronto, ON*, 2016.
- [31] René Chávez Alfaro, Eloy Casilla Garcia, Luis Salazar, Germán Sepulveda, Alfredo Huarachi, and Ida Bartolini. Avances en la investigación colaborativa y control integrado de la “hoja de hoz” en los cultivos de olivo de tacna y arica. *Ciencia & Desarrollo*, (7):7–12, 2003.
- [32] JL Molina de la Rosa, B Jiménez Herrera, F Ruiz Coletto, F García Zamorano, J Cano Rodríguez, and J Pérez García. Técnicas de cultivo: Plagas y enfermedades del olivo. *Consejería de Agricultura, Pesca y Desarrollo Rural: Instituto de Investigación y Formación Agraria y Pesquera. Recuperado de https://www.juntadeandalucia.es/export/drupaljda/publicacion/17/07/1.%20Plagas%20y%20enferm_olivo_2017%20BAJA.pdf*, 2017.
- [33] R Fernández-Escobar, M Guerreiro, M Benlloch, and M Benlloch-González. Symptoms of nutrient deficiencies in young olive trees and leaf nutrient concentration at which such symptoms appear. *Scientia Horticulturae*, 209:279–285, 2016.
- [34] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Enhanced MQTT Architecture for Smart Supply Chain

Raouya AKNIN[✉], Youssef Bentaleb

Ibn Tofail University, Engineering Sciences Laboratory, Kenitra, Morocco

Abstract—In industry 4.0, the use of smart supply chains has become necessary in order to overcome the shortcomings of traditional supply chains, such as overstocking, delivery delays, and stock out. However, the use of smart supply chains has introduced new security challenges because of the internet of things (IOT) constraint nature. Thus, the problem raised is ensuring the supply chain security requirements while taking into consideration the properties of the constraint environment. For this purpose, this paper aims to strengthen the authentication and data transmitting processes of the Message Queuing Telemetry Transport (MQTT) protocol, as the most used protocol for communication in the IOT environment, using blockchain and smart contracts. The new MQTT architecture allows to avoid a single point of failure, to ensure data immutability and to automatize the authentication mechanism as well as the publishing and the subscribing processes. In addition, the use of a one-time password (OTP) instead of a permanent one is another security measure used to protect the architecture from identity spoofing. The new architecture comprises three phases: Registration, Connection, and Publishing. Each phase is automatically controlled by a smart contract. For attack simulation tests, the smart contracts are implemented in a remix environment. The results of the simulation tests show that the new architecture is robust and resistant to different attacks.

Keywords—Smart supply chain; internet of things; MQTT protocol; blockchain; smart contracts; Ethereum; solidity; one-time password

I. INTRODUCTION

Industry 4.0 has proven its effectiveness in managing and optimizing the value chain. It thus enables better production with minimal cost and increased accuracy. The use of various technologies such as internet of things (IOT), cloud computing, artificial intelligence, and big data in order to interconnect different production and control units, collect data and analyze it has offered better visibility, quick decision-making, and intervention in the best time frame. The successful transformation of industries towards 4.0 requires first a review of supply chain management as a crucial element in managing the relationships between the different stakeholders involved in the product life cycle from raw materials acquisition until the delivery of the final product to the customer. Indeed, traditional supply chains have created several challenges, such as overstocking, delivery delays, and stockouts. Although the integration of IOT in supply chain management has solved the above challenges, security issues have arisen and have become more and more worrisome. Indeed, smart supply chains are subjects of several attacks that threaten data confidentiality, integrity, and availability due to security policies that have been breached to meet the constrained environment's requirements. Research works addressing the supply chain security challenges have either used the IOT protocols' native security mechanisms or have

proposed a robust solution without considering the constrained environment nature. Hence, the challenge is to fit the supply chains security requirements while taking into consideration the constrained environment properties. This paper will propose a security enhancement of the Message Queuing Telemetry Transport (MQTT) protocol, one of the most used protocol in the IOT environment, thanks to its lightness in terms of resources and bandwidth, without affecting the protocol's overall operation and its performances. The new architecture is based on smart contracts and consortium blockchain in order to automatize access control, publishing, and subscribing processes of devices involved in different stages of the product life cycle as well as to ensure the security of shared data between the supply chain's stakeholders. It is composed of three phases: Registration, Connecting, and Publishing phases. Each phase is controlled by a smart contract. In the registration phase, stakeholders' trusted administrators register the devices in the blockchain, define the topics that they have the rights to publish and subscribe on and recover the necessary keys for connection and transmitting data. Before, publishing or subscribing, the devices must first be connected to the broker network using one-time password (OTP) for authentication. Once the authentication data are approved, they can publish or register in a topic according to the rights granted by their administrators. To further strengthen the architecture, this paper will propose some security measures to protect the supply chain from different attacks. Indeed, it uses a broker network instead of one to avoid a denial of service. It uses the OTP for connection to the broker network in order to protect the device from identity spoofing. It also automatically verifies the packet number before any transaction to protect the architecture from reply attacks. The simulation tests attack will show that the new architecture is resilient to these attacks. The remainder of this paper is structured as follows: Section II will discuss research works that address the smart supply chains security challenges. Section III will report the main smart supply chain management challenges. Section IV will review some preliminary concepts needed for the proposal. Section V will describe the proposed MQTT architecture for the supply chain. Section VI will implement the smart contracts used for the registration, the connection, and the publishing phases using remix IDE. Section VII will show the results of attack simulation tests. Section VIII will describe the contribution of this paper in addressing smart supply chains challenges. Section IX will summarize the main ideas discussed in this article.

II. RELATED WORKS

In this section, we will introduce research works addressing the security challenges in supply chain management.

Article[23] proposed an architecture and an implementation of automated supply chain management for position and shipment tracking using global positioning system (GPS) and Radio Frequency Identification (RFID) technologies respectively. The proposed architecture has used MQTT protocol for communication, however, it used native security mechanisms which are not suitable for transmitting a critical data. Article[22] proposed a conceptual framework for supply chain management using blockchain and smart contracts in order to reduce the involvement of third parties and improve data security. However, it focused only on product purchasing and the agreement between buyer and seller while the supply chain management requires the involvement of different stakeholders participating in the product life cycle. Article[20] proposed an implementation of a food tracing system based on permissioned blockchain for the food supply chain. The security supply chain requirements regarding data privacy, confidentiality and access control are fully respected. However, because the supply chain is composed of different stockholders, it will be preferable to use a consortium blockchain instead of a permissioned blockchain. Moreover, the solution used a blockchain in constrained environment which can cause problems of resources and bandwidth. Article [21] applies Cyber Threat Intelligence with Machine Learning techniques to analyze and predict threats against supply chains. However, it focused only on analyzing and predicting the threats without offering the necessary countermeasures. Hence, the proposed solution will be efficient for an already secured architecture. After analyzing the aforementioned research works, it turns out that the papers either use the IOT protocols native security mechanisms for supply chain management which does not fit the security requirements of supply chains, or propose robust security solutions without considering the constrained environment properties. The architecture proposed in this paper aims to improve the security of the MQTT protocol in order to fit the security supply chain's requirements without affecting the overall operations neither the lightness of the protocol. The solution is based on consortium blockchain to allow the interaction between different stakeholders and smart contracts to automatize the authentication, the publication, and the subscription processes as well as to ensure data confidentiality, integrity, and availability. The advantage of this architecture is that only the broker's network interacts with the blockchain and the smart contracts which eliminate the use of the blockchain in the constrained environment. To avoid using the Transport Layer Security (TLS) protocol in this environment, the parameters for calculating the OTP are exchanged on different channels, making it impossible for a hacker to retrieve it. Verification of the packet numbers by smart contracts is another security measure used to avoid replay attacks.

III. SMART SUPPLY CHAIN ISSUE

The goal behind supply chain management is to supervise the product life cycle from the purchase of the raw material until the product's deliverance to a customer. The supply chain is composed of many independent entities that must share the data with each other. The use of IOT devices in the supply chain management such as sensors, Radio Frequency Identification (RFID), and Global Positioning System (GPS) improves the performances as well as the transparency in the

whole process product life cycle. The gathered data in each step must be shared with other entities. Hence, the MQTT protocol is the fittest protocol thanks to its publish-subscribe model. Indeed, the IOT components can share the data in a specific topic and the subscribers in this topic can receive the data thanks to the intermediate server called a broker. However, the native security measures are insufficient to ensure the security supply chain requirements, and it can make it a subject of many attacks, namely Man in the middle attack, Denial of service, reply attacks, identity spoofing, Information Disclosure, Privilege Escalation, Tampering Data, and so on [15]. In fact, MQTT Messages are not natively encrypted and the data is exchanged in plaintext. Hence any network sniffer can acquire valuable information such as : IP broker, credentials, Name of a topic, Data payload, MQTT port number, and so on. By having this information, several attacks can be carried out. First, a hacker can steal the credentials during the connection establishment phase and then publishes the wrong data on behalf of the legitimate publisher. Moreover, since the credentials are permanent, the hacker can use them forever until the client changes them. Data integrity is also a challenge for the current MQTT architecture since the hacker can modify the MQTT messages content including the topic name that the legitimate publisher had published in. Another attack scenario targeting data integrity can happen when a hacker sends malicious firmware to subscribers in order to transform them into botnets [16] [17] [18]. The proposed solution for this challenge is the use of the Transport Layer Security (TLS) protocol. However, it doesn't seem to be a good alternative since it increases the computational overhead on resource-constrained devices. Another attack type that MQTT is facing is a Denial of service (DoS). Since the Broker presents a single point of failure in MQTT architecture, it can be a target of many DOS attacks. Hence, when a broker is broken down, communication between publisher and subscriber is no longer possible. The Slow DoS against Internet of Things Environments (SlowITe) attack is one among other DOS attacks that target the MQTT protocol. Indeed, it tries to saturate the broker with the maximum possible connections in order to deprive the legitimate clients to connect to the broker [19]. In order to overcome these challenges, the proposed architecture will use a blockchain and smart contract in order to automatize and strengthen the authentication process as well as to ensure data confidentiality and integrity. It will base on a consortium blockchain since the supply chain entities are independent . Indeed, each entity which composed the supply chain has a trusted node called device administrator who registers the IOT devices in the blockchain as well as the topics that has the right to publish or subscribe on a. For instance, the transportation company register the GPS device in the blockchain in order to publish in the Geolocalisation topic. The entities concerned by product tracking such as the customer, the Retailer and the manufactory, are registered in the blockchain as subscribers on this topic.

IV. PRELIMINARIES

This section is a reminder of the main concepts and the technologies used in this proposal.

A. Supply Chain

Supply Chain can be defined as a connected set of resources and processes that are involved in providing goods to customers. It establishes a multistakeholder collaboration environment between different entities that are involved in the product life cycle (manufacturers, suppliers, distributors, retailers and transportation, information and other logistics management service providers) from the raw materials sourcing until the deliverance of finished goods to the end consumer [1][2]. In industry 4.0, the term “smart supply chain” or “digital supply chain” is used to describe the adoption of innovative technologies across all supply chain stages in order to increase the performances and improve customer service [3][4].

B. MQTT Protocol

Message Queuing Telemetry Transport (MQTT) is a publish-subscribe protocol used as an alternative to Hypertext Transfer Protocol (HTTP) in the constraint environment. As depicted in Fig. 1, The MQTT architecture is composed of three components: the publisher that sends the data related to a specific topic, the subscriber that is registered in a specific topic in order to receive a notification when this topic is updated and the broker is an intermediary server that gets the data from different publishers and sends them to subscribers that are already registered in this topic. MQTT protocol used a Transmission Control Protocol (TCP) as an underlying transport protocol [5] [6]. It uses port 1883 for unencrypted messages and port 8883 for encrypted ones. Before publishing or subscribing to any topic, the clients must be connected to a broker. The native authentication method used by the MQTT protocol is the login and password transited in plaintext format [6]. MQTT protocol offers to publishers and subscribers a feature of choosing quality of service (QOS) levels depending on the network condition, the device characteristics and the application criticality: in QOS 0, the message is delivered at most once without any acknowledgment; in QOS1, the message is delivered at least one time. Hence, the sender keeps the message stored until the reception of an acknowledgment; in QOS 2, the message is delivered exactly one time without any duplication [7].

C. Blockchain

Blockchain is a distributed database composed of a list of ordered blocks and it runs on a Peer to peer network [9]. Each block is structured as follows: A header that contains information related to the block namely Block version, Timestamp, Merkle tree root hash, Parent Block hash and nonce, which is a random number for verifying the hash. This information varies based on the blockchain network provider. A body that contains transactions. Each block is linked to the previous one thanks to the Parent block hash field which makes the blockchain immutable from frauds. Special nodes in the network, called Miners, hold the responsibility of adding a block to a blockchain. For this purpose, a consensus algorithm is used in order to reach a common agreement between untrusted nodes and define the winner miner. Proof of work and Proof of stake are the most used consensus algorithms [9] [6].

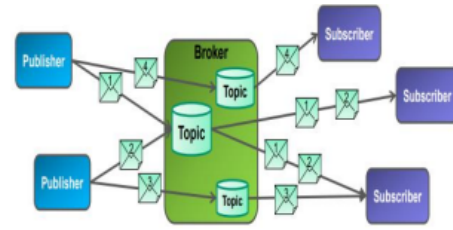


Fig. 1. The MQTT architecture [8].

D. Smart Contract

Smart contract was first introduced by Szabo in the mid-1990s in order to minimize contracting cost between transacting parties and to avoid accidental exceptions or malicious actions during contract performance. He suggested translating a contract into code that will be self-executed when predetermined conditions are met. The advent of the blockchain technology has made the implementation of smart contract possible [10]. Nowadays, the term of smart contract is popularly used to refer [9] to code scripts that run synchronously on multiple nodes of a distributed ledger namely blockchain. Ethereum is the most popular platform to implement a smart contract.

E. Ethereum

Ethereum is an open-source blockchain platform that enables the development of smart contracts [12] using a Turing-complete programming language called “solidity” [11]. Then a solidity compiler transforms a source code into bytecode in order to be interpreted by Ethereum Virtual Machine (EVM). The interaction between Ethereum smart contracts and the users is through transactions. Indeed, the Ethereum platform supports two types of accounts: user accounts and smart contract accounts. This latter is assigned to the contract, once it is deployed. User account is assigned to the Ethereum users in order to deploy contracts and to interact with them. Ether is the cryptocurrency used in the Ethereum platform. The transaction cost is defined according to two parameters: Gas limit and Gas price. Gas limit is the maximum amount of gas that the user can pay and the Gas price is the amount of Ether that the user is willing to pay for one unit of gas [13].

V. THE ENHANCED MQTT ARCHITECTURE

The architecture presented in this section is a security enhancement of the current MQTT protocol architecture. The goal is to strengthen the protocol’s security to meet supply chain requirements without disrupting its normal operation. This architecture is based on blockchain and smart contracts to automatize authentication, publication, and subscription. It aims also to ensure data confidentiality and integrity. As depicted in Fig. 2, it is composed of clients (Publisher and subscriber) which represent the IOT devices used by each entity in supply chain and the broker network which execute automatically smart contracts. The choice of consortium blockchain allows each supply chain entity to have a trusted administrator who registers the devices in the blockchain as well as the topics which can publish or subscribe on. The communication between the components is divided into three phases: The

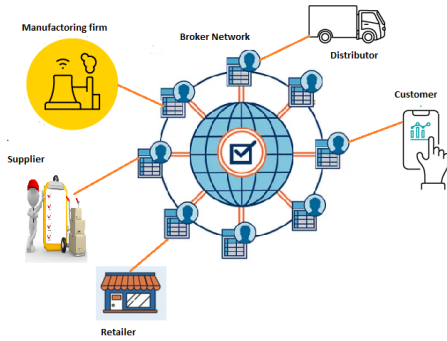


Fig. 2. The enhanced MQTT architecture.

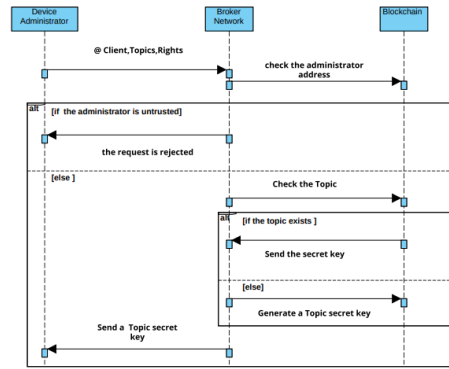


Fig. 3. Registration sequence diagram.

registration phase, the Connecting phase and the Publishing phase. Each phase is automatically controlled by a smart contract. To address the drawbacks of permanent password, the new architecture is based on OTP (One Time password) for authentication. This password that becomes invalid in few minutes protect the architecture from the identity spoofing.

A. Registration Phase

During this phase, the device administrator (Trusted node) calls the registration smart contract in order to register his devices in the blockchain and to recover the necessary keys for OTP calculating and for messages encryption. These keys are communicated to the devices in an out-of-band mode. Since the exchanged information is critical and the administrator has no computation and memory constraints, the TLS protocol can be used in this phase. After this transaction, for each device, a token containing the device address as well a list of topics that the device has the right to publish or subscribe on is generated and stored in the blockchain. Also, for each new topic a key is automatically generated and stored in the blockchain. The main steps of the registration phase are shown in Fig. 3.

B. Connecting Phase

As the current MQTT architecture, before publishing or subscribing in the topic the client must first connect to the broker. In this architecture, the device calls the connecting smart contract to connect to the broker network. Before allowing the communication, first the smart contract verifies the device registration and the Packet Number in order to avoid a Reply attack then it sends a challenge necessary for the one-time password (OTP) calculating. The device calculates the OTP and sends the a hash OTP. The smart contract verifies the hash OTP and then allows the connection to the broker network. The main steps of the connecting phase are shown in Fig. 4.

1) *OTP calculating*: The authentication process is based on one-time password (OTP) in order to avoid identity spoofing. The OTP calculation is based on or function. The inputs of this function are respectively :the list of publication's key Topics, the list of subscription's key Topics and the challenge. The detail of OTP Calculating is depicted in the equation (1):

$$OTP = F(LKP, LKS, challenge) \quad (1)$$

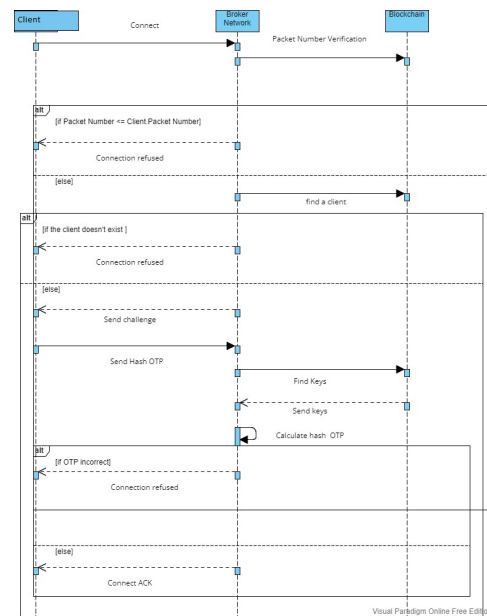


Fig. 4. Connection sequence diagram.

- F :is OR function.
- LKP: the list of publication's key Topics.
- LKS: the list of subscription 's key Topics.

N.B List of Keys and function are communicated to the device in outband mode.

C. Publishing Phase

As depicted in Fig. 5, the publishing process is the same as the current MQTT architecture in overall. However, we have added some security measures in order to strengthen the architecture. Indeed, in this phase, the device first calls the publishing smart contract. This latter checks its rights on publishing in this topic before allowing it to publish. Then, it notifies all the subscribers in that topic. In order to ensure data confidentiality, the published messages are encrypted with the topic key.

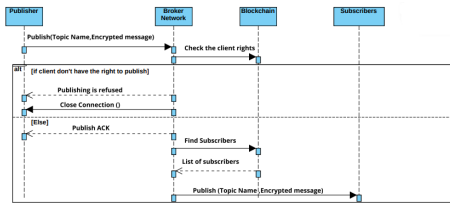


Fig. 5. Publishing sequence diagram.

VI. SMART CONTRACTS IMPLEMENTATION

The authentication mechanism as well as the publication and the subscription processes are implemented in smart contracts for automatization. They are implemented in consortium blockchain and executed by broker network. The smart contracts proposed in this paper are developed using solidity language in a remix environment.

A. Registration Smart Contract

In order to register his device, the Device administrator calls the `regisrationMain` function depicted in Fig. 6, This function allows the device administrator to interact with the registration smart contract. The administrator communicates to the function the following information: client address as well as the topics that the device has the right to publish or subscribe. Before allowing device registration, the smart contract first verifies the administrator privileges. Then it stores in the blockchain the required device information for authentication, publication and subscription. It also creates a topic if it doesn't exist. Finally, it sends to the administrator the necessary keys for OTP calculating as well as for messages encryption.

B. Connecting Smart Contract

The functions used to interact with the connecting smart contract are: `connectionMAin` in Fig. 7 and `OTPVerification` in Fig. 8. When a device sends a connecting message to the broker network, the `connectionMAin` function verifies the packet number in order to avoid a reply attack and then, it checks the existence of the device in the blockchain. After the previous verifications, the function sends a challenge to a device. In its turn, the device calculates the Hash OTP using the method detailed in the connecting phase section and calls the `OTPVerification` function. This latter verifies the hash OTP to allow or deny the connection.

C. Publishing Smart Contract

The interaction between the device and the publishing smart contract is through the `publish` function in Fig. 9. Indeed, the device calls this function for a publishing request. The function verifies the client's rights. Once the publishing is allowed, it sends a notification to all subscribers on this topic.

VII. SIMULATION TESTS

A. Case Study

In this section, we will perform simulation tests attack through a case study. Considering the simplified supply chain

```

function regisrationMain(address client_address,Topic memory client_right)
public returns (uint){
bool check_administrator=checkingAdministrator(msg.sender);
if (check_administrator==false){
emit ConnectionMessage("You don't have a privilege to add devices!The
connection is closed");
return (0);
}
else {
adddevice(client_address, client_right);
bool ccheck=CheckTopic(client_right.name);
if (ccheck==false){
uint key_Topic= addingTopic(client_right.name);
emit ConnectionMessage("The client is added to the blockchain and
the key is generated");
return(key_Topic);
}
else
for (uint h=0;h<i;i++){
bool
comparaison=stringComparison(Keys[h].name,client_right.name);
if (comparaison==true){
emit ConnectionMessage("The client is added to the blockchain and
the key is sent");
return(Keys[h].key);
}
}
}
}
    
```

Fig. 6. Resgistration main function.

```

function connectionMain(uint Packet_Number) public returns (uint) { @infinite gas
bool check_packet=Packet_verification(Packet_Number);
if (check_packet==false){
emit ConnectionMessage("Reply attack is detected and the connection is refused");
return(0);
}
else {
uint i=Clients[msg.sender].IntPacketNumber=Packet_Number;
if (Clients[msg.sender].taille_mapping==0){
emit ConnectionMessage("connection is unauthorized for unregistered device");
return(1);
}
else {
emit ConnectionMessage("The challenge is sent");
challenge =calc(challenge);
return (challenge);
}
}
}
    
```

Fig. 7. connection main function.

```

function OTP_Verification(uint hashOTP_Client) public returns (bool) { @infinite gas
uint i=Clients[msg.sender].taille_mapping;
for (uint j=0;j<i;j++){
string memory Topic_name= Clients[msg.sender].Topics[j].name;
bool Topic_read=Clients[msg.sender].Topics[j].read;
bool Topic_write=Clients[msg.sender].Topics[j].write;
uint Key=recupererCcle (Topic_name);
if (Topic_read==true){keysPublish.push(Key);}
if (Topic_write==true){keysSubscribe.push(Key);}
}
uint keyPublish=0;uint KeySubscribe=0;uint otp;
for (uint h=0;h<keysPublish.length;h++){ keyPublish=keysPublish[h]+keyPublish;
for (uint m=0;m<keysSubscribe.length;h++){ KeySubscribe=keysSubscribe[m]+keySubscribe;
}
otp=keyPublish+keySubscribe+challenge;
uint hash_OTP = uint(keccak256(abi.encodePacked(otp)));
if (hash_OTP==hashOTP_Client)
{ emit ConnectionMessage("OTP is correct and the connection is accepted");
return (true);}
else {emit ConnectionMessage("OTP is incorrect and the connection is refused");
return (false);}
}
    
```

Fig. 8. OTP verification function.

```

function publish(string memory Topic_name) public { @infinite gas
uint p;
uint i;
for (p=0;p<Clients[msg.sender].taille_mapping;++){
bool comparaison=stringComparison(Clients[msg.sender].Topics[p].name,Topic_name);
if (comparaison==true){
if (Clients[msg.sender].Topics[p].write==false){
emit ConnectionMessage("You don't have a right to publish and connection is closed!");
break;
}
else{ emit ConnectionMessage("Your message will be sent to all subscribers");
break;
}
}
}
if (p==Clients[msg.sender].taille_mapping)
emit ConnectionMessage("The topic doesn't exist in your list and the connection is closed");
else {for (uint t=0;t<taille_mapping_clients;++){
address o= Client[t];
for (y=0;y<Clients[o].taille_mapping;++){
bool comparaison1=stringComparison(Clients[o].Topics[y].name,Topic_name);
if (comparaison1==true && Clients[o].Topics[y].read==true )
Subscribers.push(o);
}
}
Display(i);
}
}
    
```

Fig. 9. Publish function.

depicted in Fig. 10 and supposing that the products are sensitive to temperature and humidity. Hence, it is important to monitor these parameters in the whole product life cycle through respectively temperature and humidity sensors. The Manufactory adds product information such as Fabrication date, end date and product ingredients in RFID Tag. This information will be published through a RFID reader. In the distribution phase a manufacturer, a retailer as well as the customer need to know the product location that is why a GPS device is used in this phase. Each entity in the supply chain has an end



Fig. 10. Simplified supply chain [14].

device in order to monitor the product manufacture. Device Administrator of each entity ensures the configuration and the maintenance of devices belonging to it. In order to collect customers’ feedback, the trusted retailer administrator adds the end device customer in the blockchain to publish his feedback later . The topics are structured in hierarchical way. Tables I, II, III, IV, V, VI, VII, VIII, IX,X, XI, XII, XIII, XIV, XV bellow summarizes the topics and the devices using in this case as well as the Devices’ rights for each topic.

B. Attack Simulation Tests

This section describes, for each phase, the nominal scenario as well as scenarios of possible attacks.

1) Registration phase:

- 1) Scenario 1: The nominal scenario
 - A trusted Administrator (supplier Administrator for example) calls the registrationMain function to register his device (Supplier Temperature sensor for example). The device has the following address (0x4B20993Bc481177ec7E8f571ceCaE8A9e22C02db). Table I depicts the device’s rights (Fig. 11).
 - The smart contract verifies the administrator address, then it registers the device in the blockchain and sends the secret topic key to the administrator (Fig. 12).
- 2) Scenario 2: Device is added by an untrusted administrator:
 - An unknown administrator tries to add a new Device. As depicted in Fig. 13, an error message is displayed.

2) Connecting phase:

- 1) Scenario 3: The nominal scenario
 - A registered Device (Supplier Temperature sensor for example) calls connectionMain function to connect to the broker network.
 - The Smart contract verifies that the device is already registered, checks the packet number to avoid a reply attack and then sends a challenge Fig. 14.
 - After receiving the challenge, the device calculates the OTP and sends the hash OTP Fig. 15.
 - The OTPVerification function verifies that the sent OTP is correct, then it allows the connection to the broker network Fig.16.
- 2) Scenario 4: The Device sends a wrong Hash OTP
 - A registered device (Supplier Temperature sensor for example) calls connectionMain function to connect to the broker network.

TABLE I. TOPIC 1

Topic Name	Device	Publish	Subscribe
Temperature /raw Material	Supplier Temperature sensor	X	-
	End device Manufactory firm	-	X
	End device Retailer	-	X
	End device customer	-	X

TABLE II. TOPIC 2

Topic Name	Device	Publish	Subscribe
Humidity /raw Material	Supplier humidity sensor	X	-
	End device Manufactory firm	-	X
	End device Retailer	-	X
	End device customer	-	X

TABLE III. TOPIC 3

Topic Name	Device	Publish	Subscribe
Product information	RFID Reader	X	-
	End device Retailer	-	X
	End device customer	-	X

TABLE IV. TOPIC 4

Topic Name	Device	Publish	Subscribe
Temperature / warehouse	Manufactory Temperature sensor	X	-
	End device Manufactory firm	-	X
	End device Retailer	-	X
	End device customer	-	X

TABLE V. TOPIC 5

Topic Name	Device	Publish	Subscribe
humidity /warehouse	Manufactory humidity sensor	X	-
	End device Manufactory firm	-	X
	End device Retailer	-	X
	End device customer	-	X

TABLE VI. TOPIC 6

Topic Name	Device	Publish	Subscribe
Geolocation	GPS	X	-
	End device Manufactory firm	-	X
	End device Retailer	-	X
	End device customer	-	X

TABLE VII. TOPIC 7

Topic Name	Device	Publish	Subscribe
Temperature /Distribution	Distribution Temperature sensor	X	-
	End device Manufactory firm	-	X
	End device Retailer	-	X
	End device customer	-	X

- The Smart contract verifies that the device is already registered, checks the packet number to avoid a reply attack and then sends a challenge.
 - The device sends a wrong a hash OTP;The network broker refuses the connection Fig.17.
- 3) Scenario 5: A malicious device connection:
 - When a malicious device sends a connect message and it is not registered the connection is refused and an error message is displayed Fig.18.

TABLE VIII. TOPIC 8

Topic Name	Device	Publish	Subscribe
humidity /Distribution	Distribution humidity sensor	X	-
	End device Manufactory firm	-	X
	End device Retailer	-	X
	End device customer	-	X

TABLE IX. TOPIC 9

Topic Name	Device	Publish	Subscribe
Temperature/warehouse Retailer	Retailer Temperature sensor	X	-
	End device Manufactory firm	-	X
	End device Retailer	-	X
	End device customer	-	X

TABLE X. TOPIC 10

Topic Name	Device	Publish	Subscribe
Humidity /warehouse Retailer	Retailer humidity sensor	X	-
	End device Manufactory firm	-	X
	End device Retailer	-	X
	End device customer	-	X

TABLE XI. TOPIC 11

Topic Name	Device	Publish	Subscribe
Customer feedback	End device customer	X	-
	End device Manufactory firm	-	X
	End device Retailer	-	X

TABLE XII. TOPIC 12

Topic Name	Device	Publish	Subscribe
Configuration/ raw Material	Supplier administrator Device	X	-
	Supplier Temperature sensor	-	X
	Supplier humidity sensor	-	X

TABLE XIII. TOPIC 13

Topic Name	Device	Publish	Subscribe
Configuration/ Manufactory	Manufactory administrator Device	X	-
	Manufactory Temperature sensor	-	X
	Manufactory humidity sensor	-	X

TABLE XIV. TOPIC 14

Topic Name	Device	Publish	Subscribe
Configuration/ Distribution	Distribution administrator Device	X	-
	Distribution Temperature sensor	-	X
	Distribution humidity sensor -	X	-
	GPS -	X	-

TABLE XV. TOPIC 15

Topic Name	Device	Publish	Subscribe
Configuration/ Retailer	Retailer administrator Device	X	-
	Retailer Temperature sensor	-	X
	Retailer humidity sensor	-	X

4) Scenario 6:Reply attack simulation:

- A registered device calls connectionMain function to connect to the broker network.
- A hacker forwards the same device’s message to deceive the broker network.
- The broker network refuses the connection

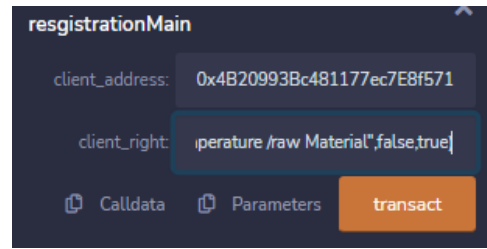


Fig. 11. Device registration transaction (Scenario 1).



Fig. 12. Device registration transaction output (Scenario 1).

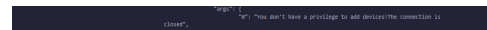


Fig. 13. An error message transaction output (scenario2).

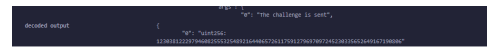


Fig. 14. Connection main transaction’s output (Scenario3).

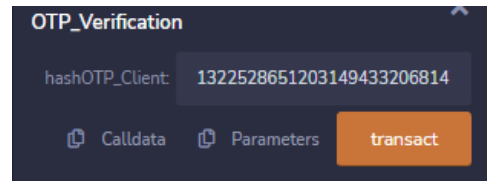


Fig. 15. OTP verification transaction (Senario 3).

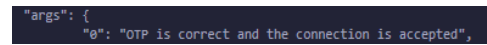


Fig. 16. Log message accepted connection (Senario 3).

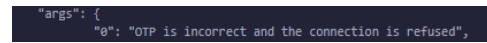


Fig. 17. Log message: Refused connection (Scenario 4).

and an error message is displayed Fig.19

3) Publishing phase:

1) Scenario 7:The nominal scenario:

- A registered Device (Supplier Temperature sensor for example)who has address (0x4B20993Bc481177ec7E8f571ceCaE8A9e22C02db) calls publish function in order to publish in “Temperature /raw Material”Fig. 20 .
- The smart contract verifies the Supplier Temperature sensor’s rights. Then it returns the addresses of all subscribers in the Topic Fig. 21.

2) Scenario 8: The topic doesn’t exist in the Devices’s list or it doesn’t have the right to publish on:

- A registered Device (End device Manufactory firm for example) publishes in “Temperature

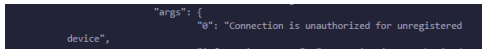


Fig. 18. Log Message: Unregistered client (Scenario 5).



Fig. 19. Log Message: Reply attack (Scenario 6).

/raw Material” topic (doesn’t have the right to publish on) .

- As depicted in Fig. 22, the network broker refuses the publishing in the topic, the connection is closed and an error message is displayed.

C. Discussion and Studies

The solution proposed in this paper responds to the problem raised, which consists in designing an architecture that meets the supply chains security requirements while taking into consideration the constrained environments properties. The attack simulation test scenarios have been designed to test the resistance of the new architecture to the attacks and security issues already mentioned in section III. Denial of service (DOS) and distributed denial-of-service (DDOS) attacks that jeopardize data availability have been addressed through the decentralized architecture of the blockchain and the use of a broker network instead of one. The use of the consortium blockchain perfectly meets the need of the supply chain, which stipulates the communication and sharing of data between several independent entities. Each entity designates an administrator who registers the entity’s devices as well as the topics on which they are allowed to publish or subscribe on (Scenario 1). Unlike the standard MQTT architecture where any device can connect, publish or subscribe in the broker, the new architecture allows connection only for already registered devices (Scenario 5). The devices publication is also allowed only for topics designated by administrators in the registration phase (Scenario 8). To avoid MAN in the middle attacks and preserve data confidentiality in a constrained environment, the authentication data exchange is carried out in several stages. Firstly, when registering devices, the administrator collects the keys necessary for calculating the OTP, which will be communicated to the device in out-of-band mode (Scenario 1). Then, during the connection phase, the broker network sends a challenge, another parameter used in the OTP calculation, to a device (Scenario 3). After calculating the OTP, the device sends the OTP hash for verification. Thus, even if a hacker is positioned between the device and the broker network, he does not have all the information needed to spoof the device’s identity. Reply attacks are avoided thanks to packet number verification (Scenario 6).

VIII. THE CONTRIBUTION OF THE ARTICLE

The goal of this paper is to enhance the MQTT security protocol in order to fit the supply chain requirements and without affecting the overall operation of this protocol. It proposed a holistic solution based on blockchain and smart contracts to automatize the authentication, connecting and

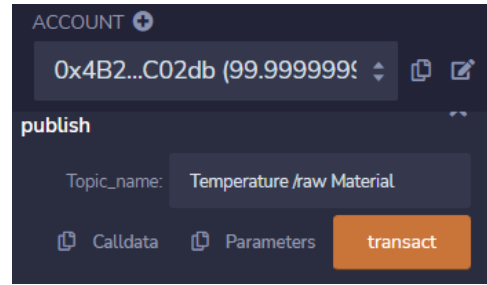


Fig. 20. Device publication transaction (Scenario7).



Fig. 21. List of subscribers (Scenario7).

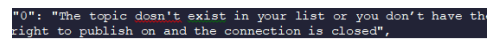


Fig. 22. Log message: Publishing is refused (Scenario 8).

publishing processes. It used a consortium blockchain since supply chain requires interaction between independent entities. Each entity appoints a trusted administrator who registers the devices in the blockchain as well as the topics that it has the right to publish or subscribe on. The TLS protocol can be used in this phase since the devices are not constrained. All the necessary keys used for OTP calculating are communicated in this phase. The paper proposed several security measures to protect the architecture from the most common attacks and to ensure data confidentiality, integrity and availability. Indeed, it used a broker network instead of one to avoid a denial of service. It used the OTP hash for connection to the broker network in order to protect the device from identity spoofing. It also automatically verified the packet number before any transaction to protect the architecture from reply attacks. The simulation tests attack showed that the new architecture is resilient to these attacks.

IX. CONCLUSION

In summary, this paper proposed a MQTT security improvement in order to fit supply chains requirements. It aims to address security challenges without affecting the overall operation or performances. The new architecture is based on blockchain and smart contracts in order to avoid a single point of failure, to ensure data immutability and to automatize the authentication mechanism as well as the publishing and the subscribing processes. The paper proposed several security measures to strengthen the MQTT architecture. Indeed, it requires the device registration before network integration. This task is performed by a trusted administrator of each supply chain entity. To lighten the protocol, the architecture has used a TLS only in this phase when the required keys for computing OTP and encrypting messages are communicated to device administrator. This later communicates this critical information to the device in out band mode. It is also based on OTP for authentication which protects the architecture from man-in-the-middle attacks. Packet number verification is another security measure used in this architecture in order to avoid a reply

attack. The attack simulation tests are shown the resistance of our architecture against malicious attacks. On the other hand, this paper is limited only to the implementation and the test of the smart contracts side, that is why in our future work, we will try to implement the end-to-end mechanism and perform the attack tests simulation on the whole architecture.

REFERENCES

- [1] Janvier-James, Assey Mbang. "A new introduction to supply chains and supply chain management: Definitions and theories perspective." *International Business Research* 5.1: 194-207 (2012).
- [2] Helo, Petri, and Yuqiuge Hao. "Blockchains in operations and supply chains: A model and reference implementation." *Computers & Industrial Engineering* 136: 242-251(2019).
- [3] Ageron, Blandine, Omar Bentahar, and Angappa Gunasekaran. "Digital supply chain: challenges and future directions." *Supply Chain Forum: An International Journal*. Vol. 21. No. 3. Taylor & Francis, 2020.
- [4] Wu, Lifang, et al. "Smart supply chain management: a review and implications for future research." *The International Journal of Logistics Management* (2016).
- [5] Buccafurri, Francesco, Vincenzo De Angelis, and Roberto Nardone. "Securing mqtt by blockchain-based otp authentication." *Sensors* 20.7 (2020).
- [6] Aknin Raouya, and Youssef Bentaleb. "Securing MQTT Architecture Using a Blockchain." *Advances in Information, Communication and Cybersecurity: Proceedings of ICI2C'21*. Springer International Publishing, (2022).
- [7] Al Enany, Marwa O., Hany M. Harb, and Gamal Attiya. "A New Back-off Algorithm with Priority Scheduling for MQTT Protocol and IoT Protocols." *International Journal of Advanced Computer Science and Applications* 12.11 (2021).
- [8] Zorkany, M., K. Fahmy, and Ahmed Yahya. "Performance evaluation of iot messaging protocol implementation for e-health systems." *International Journal of Advanced Computer Science and Applications* 10.11 (2019).
- [9] Elgendy, Mohamed Abdel Kader Mohamed, Mohamed Aborizka, and Ali Mohamed Nabil Allam. "A Blockchain-based Model for Securing IoT Transactions in a Healthcare Environment." *International Journal of Advanced Computer Science and Applications* 13.9 (2022).
- [10] Zou, Weiqin, et al. "Smart contract development: Challenges and opportunities." *IEEE Transactions on Software Engineering* 47.10: 2084-2106 (2019).
- [11] Wang, Zeli, et al. "Ethereum smart contract security research: survey and future research opportunities." *Frontiers of Computer Science* 15: 1-18(2021).
- [12] Hu, Teng, et al. "Transaction-based classification and detection approach for Ethereum smart contract." *Information Processing & Management* 58.2: 102462(2021).
- [13] Oliva, Gustavo A., Ahmed E. Hassan, and Zhen Ming Jiang. "An exploratory study of smart contracts in the Ethereum blockchain platform." *Empirical Software Engineering* 25: 1864-1904(2020).
- [14] Stadler, Hartmut. "Supply chain management: An overview." *Supply chain management and advanced planning: Concepts, models, software, and case studies* : 3-28 (2015).
- [15] Chen, Fu, et al. "A review on the study on MQTT security challenge." 2020 IEEE International Conference on Smart Cloud (SmartCloud). IEEE, (2020).
- [16] Bhawiyuga, Adhitya, Mahendra Data, and Andri Warda. "Architectural design of token based authentication of MQTT protocol in constrained IoT device." 2017 11th International Conference on Telecommunication Systems Services and Applications (TSSA). IEEE, (2017).
- [17] Laghari, ShaA, et al. "Cyberattacks and vociferous implications on SECS/GEM communications in industry 4.0 ecosystem." *International Journal of Advanced Computer Science and Applications* 12.7 (2021).
- [18] Andy, Syaiful, Budi Rahardjo, and Bagus Hanindhito. "Attack scenarios and security analysis of MQTT communication protocol in IoT system." 2017 4th International conference on electrical engineering, computer science and informatics (EECSI). IEEE, (2017).
- [19] Vaccari, Ivan, Maurizio Aiello, and Enrico Cambiaso. "SlowITe, a novel denial of service attack affecting MQTT." *Sensors* 20.10: 2932(2020).
- [20] Wu, Hanqing, et al. "Data management in supply chain using blockchain: Challenges and a case study." 28th International Conference on Computer Communication and Networks (ICCCN). IEEE, (2019).
- [21] Yeboah-Ofori, Abel, et al. "Cyber threat predictive analytics for improving cyber supply chain security." *IEEE Access* 9 : 94318-94337 (2021).
- [22] Turjo, Manoshi Das, et al. "Smart supply chain management using the blockchain and smart contract." *Scientific programming* 2021 :1-12 (2021).
- [23] Laxmi, Aishwarya Raj, and Ayaskanta Mishra. "Automation in supply chain management system using Internet of Things (IoT)." *International Journal of Engineering Technology* 7.2 :777-783(2018).

Hybrid Machine Learning-Based Approach for Anomaly Detection using Apache Spark

Hanane Chliah¹, Amal Battou², Maryem Ait el hadj³, Adil Laoufi⁴

RF-SIC Laboratory-Faculty of Science, Ibn Zohr University, Agadir, Morocco^{1, 2}

Laboratory for Sustainable Innovation and Applied Research (L.I.D.R.A), Universiapolis, Agadir, Morocco³

Equipe Systèmes Intelligents et Communicants (SIC), Ibn Zohr University, Agadir, Morocco⁴

Abstract—Over the past few decades, the volume of data has increased significantly in both scientific institutions and universities, with a large number of students enrolled and a high volume of related data. Furthermore, network traffic has increased with post-pandemic and the use of online learning. Therefore, processing network traffic data is a complex and challenging task that increases the possibility of intrusions and anomalies. Traditional security systems cannot deal with such high-speed and big data traffic. Real-time anomaly detection should be able to process data as quickly as possible to detect abnormal and malicious data. This paper proposes a hybrid approach consisting of supervised and unsupervised learning for anomaly detection based on the big data engine Apache Spark. Initially, the k-means algorithm was implemented in Sparks MLib for clustering network traffic, then for each cluster, K-nearest neighbors algorithm (KNN) was implemented for classification and anomaly detection. The proposed model was trained and validated against a real dataset from *Ibn Zohr* University. The results indicate that the proposed model outperformed other well-known algorithms in detecting anomalies based on the aforementioned dataset. The experimental results show that the proposed hybrid approach can reach up to 99.94 % accuracy using the k-fold cross-validation method in the complete dataset with all 48 features.

Keywords—Anomaly detection; big data; Apache Spark; k-means; KNN

I. INTRODUCTION

The growth of digital technology and the internet has resulted in an explosion of data creation and consumption in various fields, including science and education. Universities and research institutions are generating and collecting a vast amount of data, including research findings, academic papers, student records, and administrative data.

In addition, the COVID-19 pandemic has significantly impacted the way education is delivered, with many institutions transitioning to online learning to maintain social distancing and reduce the spread of the virus. This shift has led to an increase in network traffic, as students and faculty members access online resources, participate in virtual classes, and communicate through online platforms. Processing such network traffic data is a complex and challenging task that increases the possibility of intrusions and anomalies. Therefore, universities and research institutions need to ensure that their digital infrastructure can handle the increased data volume and network traffic, while maintaining data security and privacy, but these pillars are not sufficient without anomaly detection.

Several security strategies have been implemented to secure

networks; firewalls are an example of strategies used as a basic packet filter. Studies demonstrate that firewalls are not sufficient in providing secure environment [1][2]. Therefore, combining anomaly detection with firewalls can provide a safer network.

In machine learning (ML)-based anomaly detection, the anomaly is represented through a set of features that comprises the expected behavior of a system. Therefore, the ML model is expected to classify other events correctly if they present the same behavior during the training phase. Although, in a real-world environment, such as scientific institutions and universities, where the network traffic changes daily, either due to new attack discoveries or responding to new students' requests. Traditional security systems cannot deal with such high-speed and big data traffic. Real-time anomaly detection should be able to process data as quickly as possible to detect abnormal and malicious data.

The contribution of this paper is proposing a hybrid approach consisting of supervised and unsupervised learning for anomaly detection. Furthermore, the big data engine Apache Spark is used to provide continuous monitoring through real-time network traffic processing. The proposed approach is divided to three main components. (1) Data preprocessing to prepare data for feature extraction in machine learning. The dataset used in this analysis includes both numerical and symbolic representations, with 46 numerical attributes and 2 symbolic attributes. (2) Features selection based on K-means clustering in order to partitionate the dataset into K non-overlapping clusters based on their feature values. (3) Anomaly detection through clustering-based methods which are commonly used by analyzing the relationship between data instances and clusters. The existence of a large distance between an instance and the clusters can be used to identify an anomaly. In the suggested approach, the K-Nearest Neighbors (KNN) algorithm was applied with cross-validation techniques to detect network traffic anomalies.

The proposed model was trained and validated against a real dataset from *Ibn Zohr* University, and aimed to minimize false negatives and optimize true positives. The results indicate that the proposed model outperformed other well-known algorithms; in detecting anomalies based on the aforementioned dataset; compared to other well-known algorithms such as Random Forest [3], Support Vector Machine (SVM) [4], Naive Bayes [4] and Gradient Boosting [5]. The experimental results show that the proposed hybrid approach can reach up to 99.94 % accuracy using the k-fold cross-validation method in the

complete dataset with all 48 features.

The rest of this paper is organized as follows: Section II, present related work. Section III presents the proposed approach. Section IV reports and analyze experimental results. Section V presents general observations and considerations. Finally, the conclusion and expected future work are given in Section VI.

II. RELATED WORK

The issue of designing an efficient intrusion detection system has been discussed in many articles [6]. In order to achieve this goal, researchers have been working for many years on different aspects. Mainly, improving the accuracy of intrusion detection by minimizing the false detection rate. Another aspect concerns real-time intrusion detection, especially in a big data environment [7]. In this direction, this section reviews the prior work on applying machine learning techniques to support real-time processing.

Vimalkumar and Radhika [8] have designed a big data framework for intrusion detection using classification methods and Apache Spark as a platform for implementing intrusion detection in smart grids using big data analytics. However, the accuracy obtained is less than 80% and the DNN prediction time is higher compared to other models.

The authors in [9] have evaluated multiple classification algorithms using Apache Spark on the full UNSW-NB15 dataset. The performance achieved 97.49 % accuracy for Random Forest. Nevertheless, with the absence of any computation platform details, it is hard to validate the extremely high accuracies reported. Zhang et al. [10] proposed a real-time intrusion detection system for high-speed network environments using a distributed Random Forest detection model based on Spark. Their experimental results showed that the framework has a shorter detection time, achieves higher accuracy, and can realize real-time intrusion detection in a high-speed network environment. Gupta and Kulariya [11] proposed the use of Apache Spark for fast and efficient network intrusion detection. The authors compared the performance of five classifiers on the datasets KDD99 and NSL-KDD. The Random Forest method achieved the best accuracy on both datasets. In the same direction of using machine learning techniques to improve IDSs and to reduce the FPR and FNR, the authors of [12] developed a prototype IDS using the k-means algorithm implemented in Sparks MLlib.

In recent years, deep learning technologies are widely conducted in the field of intrusion detection [13][14]. For example, Mighan and Kahani [15] proposed a scheme that combines the advantages of a deep network and ML algorithms on Apache Spark. They used stacked auto-encoding network for feature extraction followed by several classification algorithms on The ISCX 2012 dataset. In the same direction, Chen et al. [16] proposed a method which uses the fusion convolutional neural network (FCNN) for feature extraction and stacked ensemble (SE) for classification.

Table I compares the related works in terms of data streams feature, ML algorithms and datasets that they have used, and the metrics used for evaluation. Some features in Table I such evaluation metrics are assigned * since there is no information available for that feature in related paper.

III. THE PROPOSED MODEL

This section presents the proposed hybrid approach consisting of supervised and unsupervised learning for anomaly detection. Let us recall that the aim of the proposed model is to provide continuous monitoring through real-time network traffic processing based on the big data engine Apache Spark. The general architecture can be described in Fig. 1.

The system presented in this article is designed as follows: for each new instance data captured in NetFlow format, Apache Spark Streaming is used as network traffic processing tool. The captured traffic from the input data is collected and streamed with Apache Kafka to Spark for processing. The captured packets from the input data as well as data transmission are mainly implemented by Kafka. This later is used because of its power to stream a huge amount of data in real time. Kafka mainly includes a producer and a consumer as shown in Fig. 2. The input application report messages through the producer API, and the output application subscribes to the message through consumer API.

To enhance detection efficiency, this proposed architecture makes use of the Apache Spark framework, as an open source powerful, scalable and fast distributed data processing engine in big data. Apache Spark provides a large number of libraries, such as MLlib and Spark Streaming. These tools allow Spark to support not only Streaming data processing but also machine learning algorithms. The current popular streaming frameworks mainly include Samza, Strom and Spark. Moreover, the use of Apache Spark framework is proved with the performance comparison of the literature [17].

1) *The proposed algorithm* : A joint algorithm is proposed for all layers to identify intruder flows. First, for each captured packet, the algorithm checks whether the incoming packet belongs to an already registered flow (lines 1-8 of Algorithm 1). A flow is defined as a five-tuples: IP source, IP destination, source port, and destination port (scr_{IP} , dst_{IP} , scr_{Port} , dst_{Port} , respectively). If the packet does not belong to a registered flow, then it is registered as a new distinct flow based on its scr_{IP} , dst_{IP} , scr_{Port} , dst_{Port} and a new sequence file is created for this flow by putting packet information into the first line (line 3 of Algorithm 1). On the other hand, if the packet belongs to a registered flow, then the packet information is simply sent to the particular sequence file corresponding to the registered flow (lines 6 and 7 of Algorithm 1). When the duration threshold deviates, the sequence file is sent to one of the preprocessing nodes to compute the flow parameters. The preprocessing nodes are equipped with a parameter calculation code to measure the feature values (lines 11 and 14 of Algorithm 1). In order to consider the big data environment, the code can be run in parallel by taking the sequence file as input in the distributed processing. Finally, the calculated feature values are sent to the decision server, which is equipped with several machine learning classifiers to decide whether the flow is an intrusion or normal flow based on its extracted features (line 17 of Algorithm 1). The complete system's flow is depicted in Fig. 3.

A. Data Preprocessing

The preprocessing step is critical for preparing data for feature extraction in machine learning. The real dataset used in

TABLE I. SUMMARY AND COMPARISON OF RELATED WORKS

Ref	Data streams	ML integration	ML algorithm used	Dataset description	Dataset used	Metrics comparison	Evaluation metrics
[8]		✓	DNN, SVM, RF DT and NB	✓	synchrophasor dataset.	✓	accuracy, recall, false rate, specificity
[9]		✓	SVM, RF DT and NB	✓	UNSW-NB15 dataset	✓	accuracy, recall, specificity
[10]	✓	✓	RF	✓	CICIDS2017open dataset	✓	precision, recall, F1-score
[11]		✓	LR, SVM, RF GBDT and NB	✓	DARPA's KDD'99, NSL-KDD	✓	accuracy, sensitivity, and specificity
[13]		✓	DNN	✓	DARPA's KDD'99, NSL-KDD		accuracy
[15]		✓	SVM and DT	✓	UNB ISCX 2012 dataset	✓ ✓	accuracy, recall, f-measure sensitivity, precision
[16]			CNN and stacked ensemble		NSL-KDD dataset	*	*
[17]		✓	RF, DT and NB	✓	BoT-IoT dataset		F-Measure (f1)
[18]	✓	✓	RF and DT	✓	CICIDS2017open dataset	✓	accuracy, precision, recall and F1-score
[19]		✓	CNN and LSTM	✓	NSL-KDD dataset	✓	accuracy and false alarm rate
[20]	✓	✓	Deep Learning	✓	MISP	✓	accuracy, precision recall and F1-score
[21]		✓	K-means	✓	Customize dataset		Anomaly rate
Our approach	✓	✓	K-means and KNN	✓	<i>Ibn Zohr</i> university Dataset	✓	accuracy, precision and recall

✓: Approach has this feature, Approach has not this feature, *: Information not available.

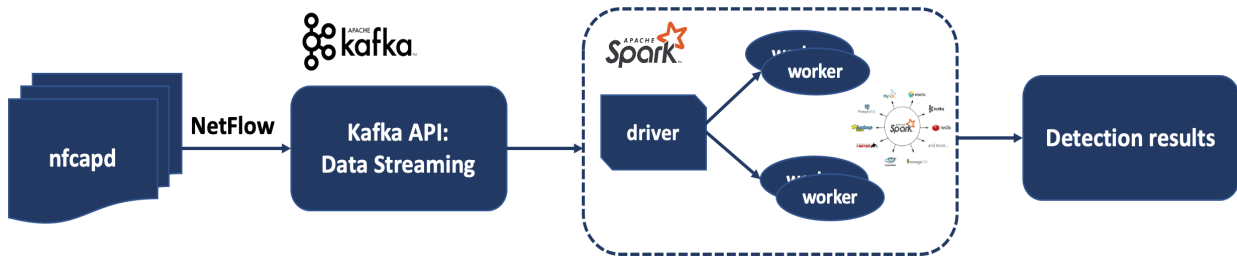


Fig. 1. The proposed approach architecture.

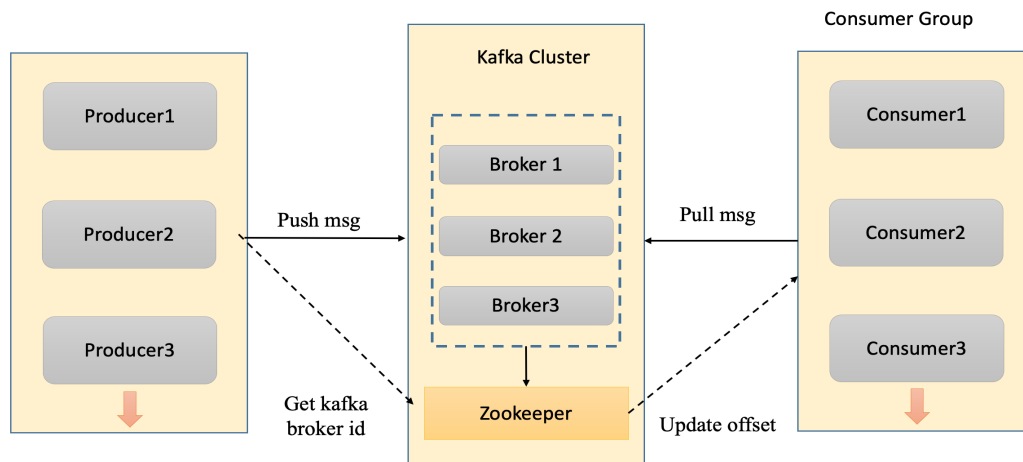


Fig. 2. Kafka architecture-Kafka Cluster.

this analysis includes both numerical and symbolic representations, with 46 numerical attributes and two symbolic attributes. To enable the use of symbolic attributes in the subsequent analysis, they are converted to numerical values, ensuring that all features are represented on the same scale. [21] emphasized that data normalization is an essential preprocessing step to

eliminate the dimensionality effect and improve the accuracy of the model. StandardScaler is a widely used normalization method that scales input features to have a mean of 0 and a standard deviation of 1, which helps to remove biases caused by different units or scales of measurement as given in Eq.1. For numerical attributes, each value is normalized over the

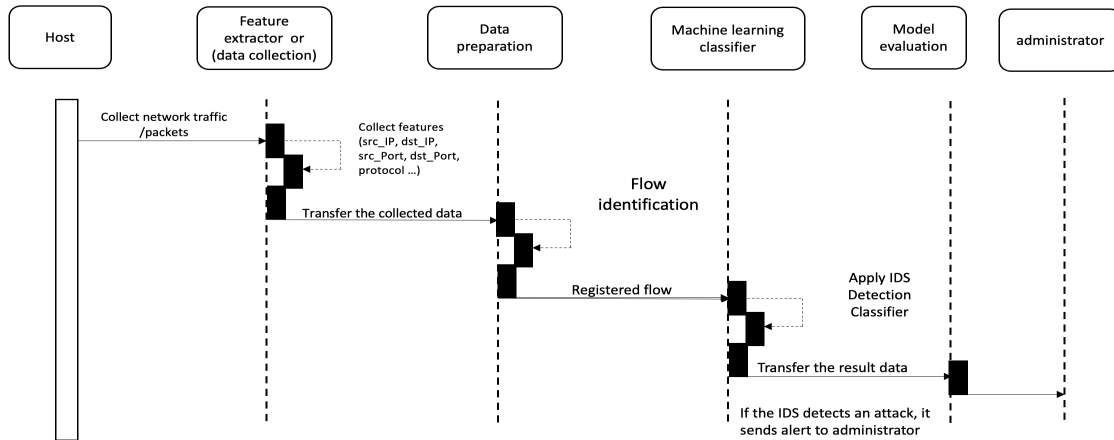


Fig. 3. The proposed algorithm sequence diagram.

Algorithm 1 Anomaly detection Algorithm

```

Input: continuous real-time network traffic/packets
Output: result, intrusions flows/normal flows.
1: for each incoming packet do
2:   if flow already register? = No then
3:      $flow_{list} += new_{flow}(pkt_{srcIP}, pkt_{dstIP}, pkt_{srcPort},$ 
4:        $pkt_{dstPort})$ 
5:     add packet parameters(new flow, new sequence file)
6:     return to next incoming packet
7:   else  $\triangleright$  flow already register? = Yes
8:     add packet parameters (registered flow, sequence file)
9:   end if
10:  if flow duration < time threshold then
11:    return to next incoming packet
12:  else  $\triangleright$  flow duration > time threshold
13:    send sequence file to reprocessing node
14:    for each sequence file do
15:      calculate flow parameter features
16:      send feature values to detection server
17:    end for
18:    result = ML classifier (parameters values)
19:    return next packet
20:  end if
end for
    
```

range [0,1] based on its deviation from the feature’s mean and standard deviation. By standardizing the data in this way, the effect of dimensionality can be reduced, leading to more accurate and reliable analysis of the data [22].

$$X' = \frac{X - \mu}{\sigma} \tag{1}$$

Where X is a feature vector containing the original values, μ is the mean of the feature vector X , σ is the standard deviation of the feature vector X and X' is the scaled feature vector, with mean 0 and variance 1.

B. Clustering

K-means clustering is a widely utilized unsupervised learning algorithm in data analysis and machine learning [22]. K-means partitions the dataset into K non-overlapping clusters based on their feature values, with objects that belong to the same cluster having similar feature values [21]. The K-means algorithm consists of four main steps: first, it initializes the number of clusters K and the initial centroids for each cluster.

Next, it iterates over all objects in the dataset and computes the distances between them and the centroids of each cluster. Based on this, it assigns the objects to the nearest cluster centroid. In the third step, it recalculates the centroids for each cluster based on the objects that are currently assigned to it. Finally, the algorithm repeats the second and third steps until the centroids of all clusters no longer change.

In order to determine the distance or similarity between two objects, it is necessary to use a distance function. The most commonly utilized distance metric is the Euclidean distance, which is defined as the square root of the sum of the squared differences between corresponding features in two input vectors, as illustrated in Equation 2. However, given that various features are often measured using different scales or metrics, it is crucial to normalize them before applying the distance function. An alternative to the Euclidean distance is the Mahalanobis distance function. This distance metric incorporates statistical correlations between different features by utilizing the inverse covariance matrix, as shown in Equation 3. However, the Mahalanobis distance can be computationally intensive, particularly for high-dimensional feature vectors. The Euclidean distance was utilized for the initial evaluation of the proposed anomaly detection method because the database was pre-normalized.

$$d(x, y) = \sqrt{\sum_{i=1}^m (x_i - y_i)^2} \tag{2}$$

$$d(x, y) = \sqrt{(x - y)^T S^{-1} (x - y)} \tag{3}$$

The K-means clustering algorithm is utilized for preprocessing training datasets that may consist of both normal and anomalous traffic. This is achieved without the need for prior labeling. The proposed approach is based on the assumption that normal and anomalous traffic can be distinguished by creating separate clusters in the feature space. The K-means clustering algorithm divides the training data into K clusters, without distinguishing between clusters that reflect time intervals of normal or anomalous traffic. For instance, an anomalous

cluster may be identified through a higher average number of packets. Clusters that are closely located to each other may arise due to either an unsuitable selection of K or the homogenous nature of the training data, where either there is no anomalous traffic or the anomalous traffic is similar to the normal traffic. A fundamental challenge of the K-means clustering method is to determine the appropriate number of K-clusters. Our study addresses this issue by concentrating on the evaluation criteria for optimizing clusters. Selecting the optimal value of k for K-means clustering is a crucial step in data analysis. This paper used the Silhouette approach to identify the optimal k value for *Ibn Zohr* dataset. The Silhouette method computes the silhouette coefficient for all instances across a range of k values. Equation 4 used to calculate the silhouette coefficient for each instance.

Silhouette coefficient:

$$\frac{b - a}{\max(a, b)} \quad (4)$$

C. Anomaly Detection and Attack Classification

Clustering-based methods are commonly used to detect anomalies by analyzing the relationship between data instances and clusters. The existence of a large distance between an instance and the clusters can be used to identify an anomaly. The proposed approach utilizes the K-Nearest Neighbors (KNN) algorithm with cross-validation techniques to detect network traffic anomalies.

KNN [23] algorithm is founded on the principle that objects or instances which are similar tend to be situated in close proximity. One commonly used distance metric in KNN is the Euclidean distance method. After computing the distance between the current data point and the query data point, the KNN algorithm sorts the distances and their corresponding indices and stores them in a collection. The classification of instances in the KNN algorithm involves taking the mode of the first K labels from the sorted collection. Although KNN is a non-parametric algorithm that is simple and easy to understand, it can become slow when processing large datasets and may struggle to make accurate predictions in high-dimensional datasets. Furthermore, the process of selecting an appropriate value for K in KNN can be a significant challenge. The KNN algorithm can be implemented in six steps

- Selecting the number of neighbors (K);
- Computing the Euclidean distance;
- Selecting the K nearest neighbors;
- Counting the number of data points in each category among the K nearest neighbors;
- Assigning the new data point to the category with the highest number of neighbors, and
- Creating the K -NN classification model for future predictions.

The selection of an appropriate value for the hyperparameter K is crucial in achieving high accuracy when using the KNN algorithm. In this study, cross-validation was used to assess the performance of different K values on the test dataset. Cross-validation methodology involves segmenting the

dataset into multiple subsets, and subsequently training the KNN model iteratively on one subset while testing it on the remaining subsets. This methodology permits the estimation of the KNN model's performance on new data, and facilitates the identification of the optimal K value that yields the best results. The use of cross-validation ensures that the KNN model neither overfits nor underfits the data, and guarantees optimal performance on previously unseen data. The dataset was partitioned into training and testing sets, with 70% of the data designated for training and 30% for testing. The training dataset was then passed through the KNN classifier for classification.

D. Decision Making

To evaluate the effectiveness and efficiency of the suggested method in detecting anomalies, a comparison was performed against well-known classification techniques such as Random Forest (RF), Support Vector Machines (SVM), Naive Bayes (NB), and Gradient Boosting (GB). The comparison enabled the determination of the relative performance of the proposed approach and the identification of its strengths and weaknesses. Additionally, experiments were conducted using various evaluation metrics to provide a comprehensive assessment of the model's performance. The aim was to demonstrate the superiority of the proposed method in accurately detecting anomalies in the *Ibn Zohr* university network. The results of the comparison are presented in the following section.

1) *Random forest* [3]: is an influential machine learning technique that combines decision trees with ensemble learning. The algorithm employs multiple decision trees that are trained on different subsets of the training data and features. Each tree independently provides a prediction, and the final output is the average prediction of all the trees. The algorithm works by partitioning the feature space into smaller subsets recursively, utilizing various metrics such as Gini impurity or entropy. The decision trees in Random Forest are created using a randomized feature selection process that reduces model variance and overfitting. The algorithm's benefits include its high accuracy, noise robustness, and capacity to handle high-dimensional data. However, the algorithm's computational cost can be significant, and its interpretability may be compromised due to its ensemble nature.

2) *Support vector machine* [4]: is a renowned machine learning algorithm that is widely used in both classification and regression tasks. The method works by creating a hyperplane that separates classes and maximizes the margin between them. This optimal hyperplane is identified by finding a subset of training data points called support vectors that lie nearest to the hyperplane. To handle non-linear decision boundaries, SVM transforms the input data into a higher-dimensional space using a kernel function, such as linear, polynomial, or Gaussian. The choice of kernel function depends on the nature of the problem.

3) *Naive bayes* [4] : is a well-known algorithm utilized in classification tasks. It is based on Bayes' theorem, which states that the probability of a hypothesis given the data is proportional to the product of the prior probability of the hypothesis and the likelihood of the data given the hypothesis. Naive Bayes is considered "naive" because it assumes that all features are independent of one another, simplifying the

computations but possibly leading to suboptimal performance in cases where the features are highly correlated.

4) *Gradient boosting* [5]: Gradient Boosting [5] is a machine learning algorithm commonly utilized in regression and classification tasks. The approach involves the combination of multiple weak learners, generally decision trees, to form a strong learner. Learners are added incrementally, with each new learner attempting to correct the errors of the preceding one. Gradient Boosting uses the gradient descent optimization algorithm to find optimal weak learner parameters. It achieves this by iteratively adjusting the parameters of the learners to minimize a loss function, such as mean squared error or log loss. The gradient of the loss function is calculated with respect to the output of the preceding learner, which is utilized to train the next learner. Gradient Boosting is highly customizable, with hyper parameters like the number of learners, learning rate, and maximum depth of trees, which can be tuned.

IV. EXPERIMENTAL RESULT

In this section, the dataset and the performance metrics used for the comparison are described in order to conduct the different experiments carried out. Then, the results that were obtained by using the proposed hybrid method in the big data ecosystem are presented and discussed. Apache Spark and Python language are used. All experiments are conducted using a machine with 1.5 GHz Intel Corei7 CPU@, 16 GB RAM, and with Ubuntu 20.04 trusty installed.

A. Dataset

In order to verify the effectiveness of the proposed approach a real dataset is used collected from digital data related to network traffic of *Ibn Zohr* university. The dataset was collected in the presidency of *Ibn Zohr* which consists of several faculties. The total size of the dataset is up to one million records. Data are sent mainly by network devices installed by *nfcapd* which is a *netflow* capture daemon of the *nfdump* tools. *nfcapd* reads *netflow* data from the network and stores it into files. The output file is automatically rotated and renamed every 5 minutes.

B. Performance Metrics

To evaluate the performance of an anomaly detection algorithm, four fundamental metrics can be used: true positive (*TP*), which represents the number of correctly identified attacks, true negative (*TN*), which represents the number of accurately identified normal connections, false positive (*FP*), which denotes the number of normal connections wrongly identified as attacks, and false negative (*FN*), which signifies the number of attack connections erroneously identified as normal [24]. Subsequently, the proposed approach is evaluated using the following metrics:

1) *Accuracy*: provides a measure of how well the classification model is able to accurately classify data instances, regardless of whether they are classified as positive or negative. It is calculated as the proportion of the total number of correctly classified records in a dataset over all the rows in that dataset.

$$Accuracy = \frac{TP + Tn}{TP + Tn + Fp + Fn} \quad (5)$$

2) *Precision*: measures how many of the positive predictions made by the model are actually correct. It is calculated as the ratio of true positive predictions to the total number of positive predictions made by the model.

$$Precision = \frac{TP}{TP + FP} \quad (6)$$

3) *Recall*: measures how many of the positive instances in the dataset were correctly identified by the model. It is calculated as the ratio of true positive predictions to the total number of actual positive instances in the dataset.

$$Recall = \frac{TP}{TP + FN} \quad (7)$$

C. Implementation and Experimental Results

1) *Selecting the optimal number of clusters*: Let us recall that data instances are grouped using K-means clustering technique. We have conducted a graph of Silhouette score vs *K* which represent an effective way to visualize and select the optimal number of clusters (*K*). Cluster cohesion and separation are measured using the Silhouette score. Cluster cohesion refers to how closely related the data points are within the same cluster, while cluster separation refers to how well-separated the clusters are from each other. A higher Silhouette score indicates a better clustering outcome (i.e. it implies that the data points are more closely related to their own cluster and less closely related to neighboring clusters). The ideal *K* number for finding abnormalities in *Ibn Zohr* data is thought to be the *K* value that corresponds to the highest Silhouette score. The cluster value *K* of 2 exhibits the highest Silhouette score, making it the best *K* value for anomaly identification, as shown in Fig. 4.

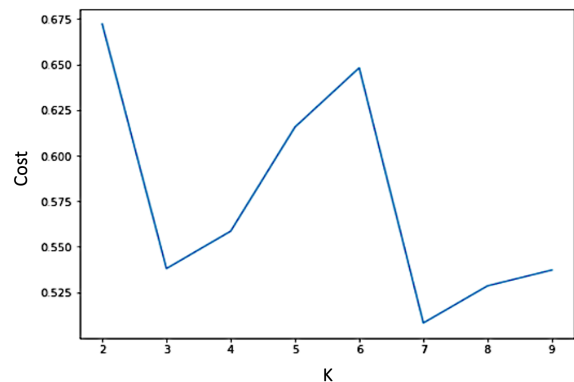


Fig. 4. Evolution of *K* value vs. silhouette score.

2) *Apply k-means clustering Using optimal cluster value of K*: To identify anomalies in each cluster, *KNN* with cross-validation is used. The ideal number of folds for *k*-fold cross-validation is determined by comparing the average performance metrics obtained for different values of *k*. The best *k* value is determined by the greatest average accuracy. The proposed approach first determines the optimal number of folds for cross-validation, then evaluates the performance of the *KNN* model on new, unseen data using cross-validation techniques such as *k*-fold cross-validation. The average distance

to the k nearest neighbors for each point is then calculated using Euclidean distance to determine the boundary between different classes and identify the most relevant. Finally, the plot method is applied to visualize the average distance to the k nearest neighbors for each point in the *Ibn Zohr* dataset.

Fig. 5 shows the results of the anomaly detection algorithm conducted on the *Ibn Zohr* dataset. It is shown that the algorithm was able to successfully separate the anomaly instances from the normal ones. Anomalies are presented by the points that are far from the group of points in the feature space.

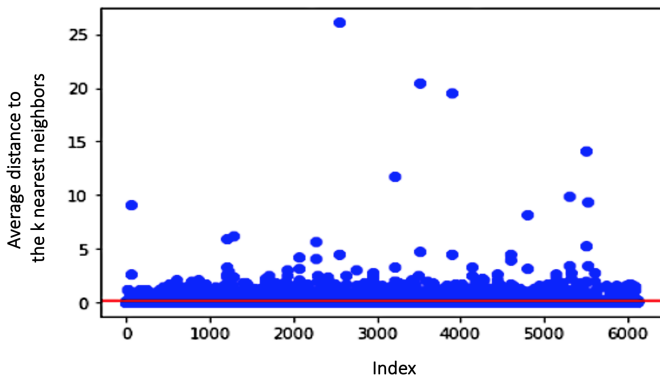


Fig. 5. Anomaly detection using KNN.

The proportion of anomalies identified by the model for various cluster values, as well as the non-linear connection between the cluster value and the number of observed anomalies, are depicted in Fig. 6. Anomalies, especially, tend to form their own clusters as the cluster value increases and may be incorrectly identified as normal data instances. Additionally, normal cases may form clusters, leading to some of them being wrongly classified as anomalies based on cluster threshold values.

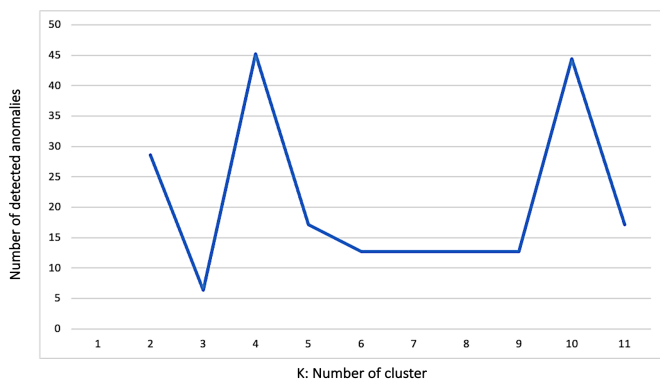


Fig. 6. The number of detected anomalies Vs. cluster number.

3) *Effectiveness analysis of the proposed anomaly detection model using optimal K:* In order to evaluate the effectiveness and efficiency of the suggested method in detecting anomalies, a comparison is performed against well-known classification techniques, such as Random Forest (RF), Support Vector Machines (SVM), Naive Bayes (NB), and Gradient Boosting (GB). The experimental data have been split into 70% as

training data and 30% as test data. The evaluation metrics include accuracy, precision, and recall.

The results of the comparison using *Ibn Zohr* data are presented in Table II. Our approach is found to outperform the other models with consistently higher accuracy score. In contrast, Naive Bayes is found to achieve a lower accuracy score compared to the other evaluated methods.

TABLE II. COMPARISON OF THE APPROACH WITH THE OTHERS

Method	Accuracy (%)	Precision (%)	Recall (%)
RF	99.88	99.66	99.46
SVM	99.81	99.33	99.30
NB	95.04	73.77	100
GB	99.89	99.73	99.5
Our approach	99.94	99.92	99.5

4) *Effectiveness analysis of the anomaly detection model for different cluster values:* The performance of clustering-based anomaly detection methods can be significantly influenced by the choice of the number of clusters, represented by K . Therefore, identifying the optimal cluster value is crucial for achieving optimal results. Additionally, the choice of cluster value has a significant impact on the classification of anomalies in *Ibn Zohr* dataset.

A comparison of various cluster values ranging from 2 to 11 was conducted, and the results are presented in Table III. Our model achieved the highest classification accuracy of 99.94% for $K = 2$, while GB had the highest accuracy value of 99.75% for both $K = 2$ and $K = 4$. RF had the best accuracy value of 99.61% for $K = 4$, while SVM achieved the best accuracy value of 99.49% $K = 3$. NB achieved the best accuracy value of 97.66% for both $K = 2$ and $K = 4$.

TABLE III. ACCURACY COMPARISON FOR DIFFERENT k VALUES

Cluster Value	RF	SVM	NB	GB	Our approach
2	99.49	99.38	97.66	99.75	99.94
3	99.57	99.49	93.69	99.66	99.80
4	99.61	99.36	97.66	99.72	99.75
5	99.44	99.35	96.48	99.48	99.71
6	99.35	99.17	90.54	99.30	99.70
7	99.26	99.23	94.29	99.38	99.67
8	99.12	97.94	83.38	99.21	99.53
9	99.15	98.98	87.04	99.30	99.75
10	99.36	99.33	82.68	99.51	99.78
11	99.25	98.90	78.24	99.35	99.70

Based on the above analysis, it can be concluded that the optimal cluster values are $K = 3$ and $K = 4$, as they yielded better performance across most classifiers. However, since the accuracy value never exceeded 99.75%, the effectiveness of the model in terms of the trade-off between accuracy and computation cost is demonstrated for the optimal cluster value of $K = 2$.

V. DISCUSSION

The proposed anomaly detection model is based on the K-means clustering algorithm [21]. To determine the most suitable number of clusters, the Silhouette method was employed, which yielded $K = 2$ as the optimal value. Following this direction, the K-nearest neighbors (KNN) algorithm is applied in each cluster to identify anomalies in the whole dataset.

To ensure the reliability of the proposed model, cross-validation was implemented, and the optimal number of folds was determined using the K-fold method [25][26]. By selecting the most suitable value of K through K-fold cross-validation, overfitting was reduced, and the accuracy of the model was improved. Furthermore, to obtain the best possible results, an investigation is conducted of the data behavior for the global cluster, varying the number of clusters from $K = 2$ to $K = 11$. The findings indicated that the combination of the K-means clustering algorithm, KNN anomaly detection technique, and cross-validation was a remarkably successful method for identifying anomalies in the *Ibn Zohr* dataset. However, increasing the cluster value could result in anomalies themselves forming their own clusters, which could lead to some anomalies being classified as normal data instances.

Although better accuracy was achieved for the proposed anomaly detection model for $K = 2, 3, 9,$ and 10 , with the best accuracy being 0.994 , a trade-off between accuracy and computation cost was observed. The evaluation of the K-means-based anomaly detection model showed that it was effective for the optimal cluster value of $K = 2$. However, the model's accuracy was found to be sensitive to the balance between the number of anomalies detected and the classification accuracy. Increasing the cluster value K could result in poor anomaly detection, as illustrated in Table III. Despite the use of the *Ibn Zohr* dataset to evaluate the effectiveness of the model, the methodology could be extended to other domains, such as IoT analytics and cybersecurity, and could prove useful for rule-based analysis based on specific datasets.

VI. CONCLUSION

The growth of digital technology and the Internet has led to a significant increase in data creation and consumption in universities and research institutions. Consequently, processing the resulting network traffic data has become a complex and challenging task, which increases the likelihood of intrusions and anomalies. Therefore, the primary objective of this paper is to address the issue of scalable network intrusion detection in a big data environment. The proposed hybrid model was validated using a real dataset from *Ibn Zohr* university. The evaluation results indicate that the suggested approach is efficient in detecting various anomalies. In addition, the effectiveness of the proposed model was verified by comparing its performance with other well-known models using different metrics.

The suggested framework's efficacy was evaluated using Apache Spark, a large data processing tool, and machine learning algorithms. The suggested model's performance was further examined using a hybrid machine learning method that blends k-means and KNN to identify anomalies.

Future work aims to integrate real-time anomaly identification to enable quick response and mitigation of any security breaches, thereby safeguarding all network traffic and protecting the privacy of all users on the *Ibn Zohr* university network.

ACKNOWLEDGMENT

The authors would like to express their gratitude to *Ibn Zohr* Presidency for providing their genuine dataset to assess the effectiveness of the proposed approach in this study.

REFERENCES

- [1] Connolly, Lena Y., and David S. Wall. "The rise of crypto-ransomware in a changing cybercrime landscape: Taxonomising countermeasures." *Computers & Security* 87 (2019): 101568.
- [2] Kristen, Erwin, et al. "Security assessment of agriculture iot (Aiot) applications." *Applied Sciences* 11.13 (2021): 5841.
- [3] Reis, Itamar, Dalya Baron, and Sahar Shahaf. "Probabilistic random forest: A machine learning algorithm for noisy data sets." *The Astronomical Journal* 157.1 (2018): 16.
- [4] Mandal, Jyotsna Kumar, and Debika Bhattacharya. *Emerging Technology in Modelling and Graphics*. Springer Singapore, 2020.
- [5] Sandhu, Amandeep Kaur, and Ranbir Singh Bath. "Software reuse analytics using integrated random forest and gradient boosting machine learning algorithm." *Software: Practice and Experience* 51.4 (2021): 735-747.
- [6] Aljanabi, Mohammad, Mohd Arfan Ismail, and Ahmed Hussein Ali. "Intrusion detection systems, issues, challenges, and needs." *International Journal of Computational Intelligence Systems* 14.1 (2021): 560-571.
- [7] Quincozes, Silvio E., et al. "A survey on intrusion detection and prevention systems in digital substations." *Computer Networks* 184 (2021): 107679.
- [8] Vimalkumar, K., and N. Radhika. "A big data framework for intrusion detection in smart grids using apache spark." 2017 International conference on advances in computing, communications and informatics (ICACCI). IEEE, 2017.
- [9] Belouch, Mustapha, Salah El Hadaj, and Mohamed Idhammad. "Performance evaluation of intrusion detection based on machine learning using Apache Spark." *Procedia Computer Science* 127 (2018): 1-6.
- [10] Zhang, Hao, et al. "Real-time distributed-random-forest-based network intrusion detection system using Apache spark." 2018 IEEE 37th international performance computing and communications conference (IPCCC). IEEE, 2018.
- [11] Gupta, Govind P., and Manish Kulariya. "A framework for fast and efficient cyber security network intrusion detection using apache spark." *Procedia Computer Science* 93 (2016): 824-831.
- [12] Azeroual, Otmame, and Anastasija Nikiforova. "Apache spark and mllib-based intrusion detection system or how the big data technologies can secure the data." *Information* 13.2 (2022): 58.
- [13] Rawat, Shisrut, et al. "Intrusion detection systems using classical machine learning techniques vs integrated unsupervised feature learning and deep neural network." *Internet Technology Letters* 5.1 (2022): e232.
- [14] Alferaidi, Ali, et al. "Distributed deep CNN-LSTM model for intrusion detection method in IoT-based vehicles." *Mathematical Problems in Engineering* 2022 (2022).
- [15] Mighan, Soosan Naderi, and Mohsen Kahani. "A novel scalable intrusion detection system based on deep learning." *International Journal of Information Security* 20 (2021): 387-403.
- [16] Chen, Chen, et al. "FCNN-SE: An Intrusion Detection Model Based on a Fusion CNN and Stacked Ensemble." *Applied Sciences* 12.17 (2022): 8601.
- [17] Abushwereb, Mohamed. *An accurate IoT intrusion detection framework using Apache Spark*. Diss. Princess Sumaya University for Technology (Jordan), 2020.
- [18] Seo, Wooseok, and Wooguil Pak. "Real-time network intrusion prevention system based on hybrid machine learning." *IEEE Access* 9 (2021): 46386-46397.
- [19] Alferaidi, Ali, et al. "Distributed deep CNN-LSTM model for intrusion detection method in IoT-based vehicles." *Mathematical Problems in Engineering* 2022 (2022).
- [20] Fotiadou, Konstantina, et al. "Incidents information sharing platform for distributed attack detection." *IEEE Open Journal of the Communications Society* 1 (2020): 593-605.
- [21] Sinaga, Kristina P., and Miin-Shen Yang. "Unsupervised K-means clustering algorithm." *IEEE access* 8 (2020): 80716-80727.
- [22] Goldstein, Markus, and Seiichi Uchida. "A comparative evaluation of unsupervised anomaly detection algorithms for multivariate data." *PloS one* 11.4 (2016): e0152173.

- [23] Imandoust, Sadegh Bafandeh, and Mohammad Bolandraftar. "Application of k-nearest neighbor (knn) approach for predicting economic events: Theoretical background." *International journal of engineering research and applications* 3.5 (2013): 605-610.
- [24] Sokolova, Marina, Nathalie Japkowicz, and Stan Szpakowicz. "Beyond accuracy, F-score and ROC: a family of discriminant measures for performance evaluation." *AI 2006: Advances in Artificial Intelligence: 19th Australian Joint Conference on Artificial Intelligence*, Hobart, Australia, December 4-8, 2006. *Proceedings 19*. Springer Berlin Heidelberg, 2006.
- [25] Pugazhenti, A., and Lakshmi Sutha Kumar. "Selection of optimal number of clusters and centroids for k-means and fuzzy c-means clustering: A review." *2020 5th International conference on computing, communication and security (ICCCS)*. IEEE, 2020.
- [26] De Bruin, Sytze, et al. "Dealing with clustered samples for assessing map accuracy by cross-validation." *Ecological Informatics* 69 (2022): 101665.

Egypt Monuments Dataset version 1: A Scalable Benchmark for Image Classification and Monument Recognition

Mennat Allah Hassan^{1,2,*}, Alaa Hamdy¹, Mona Nasr²

Faculty of Computer Science, Misr International University, Cairo, Egypt¹

Faculty of Computers and Artificial Intelligence, Helwan University, Cairo, Egypt²

Abstract—The success of machine learning (ML) as well as deep learning (DL) depends largely on data availability and quality. The system’s performance is frequently more affected by the amount and quality of its training data than by its architecture and training specifics. Consequently, demand exists for challenging datasets that both precisely measure performance and present unique challenges with real-world applications. The Egypt Monuments Dataset v1 (EGYPT-v1) is introduced as a new scalable benchmark for fine-image classification (IC) and object recognition (OR) in the domain of ancient Egyptian monuments. EGYPT-v1 dataset is by far the world’s first large specified such dataset to date, with over seven thousand images and 40 distinct instance labels. The dataset composes different categories of monuments such as pyramids, temples, mummies, statues, head statues, bust statues, heritage sites, palaces and shrines. Several advanced deep network architectures were tested to appraise the classification difficulty in the EGYPT-v1 dataset, namely ResNet50, Inception V3, and LeNet5 models. The models achieved accuracy rates as follows: 99.13%, 90.90%, and 92.64%, respectively. The dataset was predominantly created by manually collecting images from the popular global online video-sharing and social media platform, Youtube, as well as WATCHiT, Egypt’s top streaming entertainment service. Additionally, Wikimedia Commons, the largest crowdsourced media repository in the world, was used as a secondary source of images. The images that comprise the dataset can be accessed on the GitHub repository <https://github.com/mennatallahhassan/egypt-monuments-dataset>.

Keywords—Deep learning; landmark datasets; landmark recognition; monument datasets; monument recognition

I. INTRODUCTION

Within the realm of computer vision, IC and OR are key research topics that have been extensively investigated for years. The objective of IC is to [1], [2], [3], [4]. The objective of OR [5], [6] is the computer vision task of recognising a particular instance of an object, as opposed to its category. For example, it is interested in instance-level labels such as “Karnak Temple in Luxor” or “Khufu Pyramid in Giza” rather than simply “Karnak” or “Khufu” when labeling images.

When ML and DL techniques for image classification and instance-level recognition (ILR) tasks have progressed, methods have improved in their robustness and scalability, and they have started solving standard datasets.

Furthermore, despite the fact that increasingly largescale classification datasets, for instance, CIFAR-10 [7], ImageNet [8], in addition to OpenImages [9], have become standard

benchmarks, there still needs to be monument datasets for fine-grained instance recognition and classification. A monument refers to a structure that has been erected and can take many forms, including busts, crosses, statues, fountains, mausoleums, obelisks, pyramids, reliquaries, sarcophagi, steles, graves, or triumphal arches. In addition, smaller-scale forms such as medals and commemorative plaques can also be considered monuments [10]. Generally, the world is full of monuments; remarkably, Egypt contains a third of the world’s monuments; parts of these monuments are displayed in the most famous museums all over the world [11]. This paper presents the Egypt Monuments Dataset v1 (EGYPT-v1), a novel scalable dataset for IC and ILR. More than seven thousand images of forty-one different monuments and historical landmarks are available in EGYPT-v1, as seen in Fig. 1. Fig. 2 illustrates its geographic distribution across Egypt. The instance recognition task uses a training dataset of 5,833 labelled images and 1,945 labelled images as a test set, that includes ground truth information regarding the instance classification (IC) and ILR tasks. Although our primary objective for Egypt Monuments Dataset v1 is to recognize historical landmarks and monuments, the solutions developed to overcome the obstacles, can be easily adapted to address other instance-level recognition challenges, including artwork recognition.

The primary objective of Egypt Monuments Dataset v1 is to replicate real-world circumstances and, consequently, introduces several complex hindrances. There are thousands of images representing tens of classes in EGYPT-v1. The degree of intra-class variation is significantly elevated, with images of a single class exhibiting both indoor and outdoor views and images that possess a tangential association with a particular class, like museum paintings. The Egypt Monuments Dataset v1 is intended to be used as a novel benchmark for IC and ILR.

The dataset, training instance labels, classification and recognition ground truth data, and metric computation code are accessible to the public.

In summary, this paper presents the Egypt Monuments Dataset v1, a novel and challenging benchmark for fine-image classification and object recognition in the domain of ancient Egyptian monuments. The RELATED WORK section compares the existing monument/landmark image classification and recognition datasets with our novel proposed dataset. The dataset consists of many images and distinct instance labels



Fig. 1. An overview of the EGYPT-v1 dataset composed of over 7k images for 41 classes.

TABLE I. THE EGYPT-V1 DATASET IS BEING HIGHLIGHTED AMONG EXISTING MONUMENT AND HERITAGE SITES DATASETS. THE EGYPT MONUMENTS DATASET V1 IS THE FIRST PUBLIC DATASET REGARDING THE AVAILABILITY OF EGYPTIAN ANCIENT MONUMENT IMAGES AND HERITAGE SITES

Dataset	Year	# Monuments/Landmarks	# Images	Annotation Collection	Dataset Scale
Oxford [1]	2007	11	5,063	Manual	City
Paris [12]	2008	11	6,392	Manual	City
Holidays [13]	2008	500	-	Manual	Worldwide
European Cities 50k [14]	2010	20	50k	Manual	Continent
Geotagged StreetView [15]	2010	-	17k	StreetView	City
Rome 16k [16]	2010	69	16k	GeoTag + SfM	City
San Francisco [17]	2011	-	1.7M	StreetView	City
Landmarks-PointCloud [18]	2012	1k	205k	Flickr label + SfM	Worldwide
Singapore Landmark-40 [19]	2012	40	13,538	Internet sources + Manual	City
24/7 Tokyo [20]	2015	125	1k	Smartphone + Manual	City
Paris500k [21]	2015	13k	501k	Manual	City
Landmark URLs [3]	2016	586	-	Text query + Feature matching	Worldwide
Google Landmarks [22]	2017	30k	1M	GPS + semi-automatic	Worldwide
Revisited Oxford [4]	2018	11	1M	Manual + semi-automatic	Worldwide
Revisited Paris [4]	2018	11	1M	Manual + semi-automatic	Worldwide
Qutub Complex Monuments' Images [23]	2018	5	1,286	Google Images	City
Indian heritage monuments [24]	2020	143	7,150	Web Scraping	Country
Google Landmarks Dataset v2 [25]	2019	200k	5M	Crowdsourced + semi-automatic	Worldwide
Our Egypt Monuments Dataset v1	2022	41	7,778	Manual + semi-automatic	Country

representing various categories of monuments. All details have been explained in the DATASET OVERVIEW section. The performance of several advanced deep network architectures on the EGYPT-v1 dataset has been evaluated. The accuracy rates have been illustrated in the EXPERIMENT section.

This paper also discusses the dataset's creation, distribution, and potential applications in instance-level recognition challenges, including recognition. Overall, the EGYPT-v1 dataset aims to replicate real-world circumstances and provide a valuable tool for researchers and practitioners in computer vision.

II. RELATED WORK

Image recognition's challenges extend from simple image classification (e.g., "human face" or "building") through fine-grained tasks that distinguish between models, and styles (such as "Head Sculpture" and "Ancient Temple") to recognition on an instance level (as "Portrait Head of Queen Tiye at the Neues Museum, Berlin, German" and "The Great Temple

of Ramesses II, Aswan, southern Egypt"). Identifying ancient Egyptian monuments and historical landmarks is the primary focus of our novel dataset. Subsequently, datasets for image classification and recognition were scrutinized, with particular attention given to those most relevant to our research. Table I presents the existing datasets for monument/landmark image classification and recognition, along with our proposed novel dataset.

1) *City-scale datasets*: The datasets regarding Oxford [1], as well as Paris [12], consist of a huge number of landmark images found in both cities, collectively belonging to 11 categories. Additional datasets concentrate on photography from a specific city: Rome 16k [16]; Singapore Landmark-40 [19], including over 13,000 images from Singapore city. The dataset used in this study was created by collecting images from a range of different sources. Specifically, 40% of the images were obtained from Google Images, 40% were sourced from Flickr, and 5% came from Photobucket. The remaining 15% were acquired through manual means. This

portion comprised Geotagged Streetview Images [15], which consisted of approximately 17,000 photographs of Paris and San Francisco Landmarks [17], containing over 1.7 million images; Qutub Complex Monuments' Images containing 1,286 images for five famous monuments in Delhi, India; 24/7 Tokyo [20], including a thousand images under various lighting situations; and Paris500k [21], including 501,000 images.

2) *Country-scale dataset*: Indian heritage monuments dataset (IHMD) [24] contains 6,959 images of 413 classes. This dataset has been collected from image search engines using web scrappings such as Google Images, Bing Images, Wikimedia and Flickr.

3) *Continent-scale datasets*: The datasets that are more recent in origin contain images that have been sourced from a notably wider range of locales within the same continent than the older datasets. Within the European Cities (EC) 50k dataset, there are images of 20 landmarks that are distributed across 9 cities [14], including unannotated images from 5 additional cities that were used as distractors. Another version of this dataset has 1 million photos from 22 cities; however, all annotated photographs come from only one location [26].

4) *Worldwide-scale datasets*: The more expanded datasets have images of landmarks globally. The Landmarks-PointCloud dataset includes 205,000 images of 1,000 well-known landmarks [18]; The Landmark URLs dataset of approximately 192,000 images is classified into 586 landmarks. 168,882 images are utilized for fine-tuning in experiments, while the remaining 20,668 images are utilized to validate parameters [3]. The Revisited Oxford dataset, as well as Revisited Paris dataset are two other recent examples of global landmark datasets, each comprising eleven landmarks and roughly a million images [27]. The Google Landmarks Dataset, which is the dataset from which the data used in this study was drawn, originally consisted of 2.3 million photographs taken at 30,000 unique landmarks. However, this dataset is unstable due

to copyright limitations. It declines over time as photographs are destroyed by the users who upload them [22]. To our best knowledge, no such large unique, collected country-scale datasets with ground truth Egyptian ancient monuments and landmarks visibility information are publicly available yet.

III. DATASET OVERVIEW

A. Purposes

The purpose of the Egypt Monuments Dataset v1 is to simulate the following constraints of industrial monument/landmark recognition. It is *scalable* to encompass all ancient Egyptian monuments and landmarks worldwide, as they are not confined to Egypt alone. There are numerous discoveries and expeditions concerning ancient Egyptian monuments *Intra-class variability*. Images of monuments and historical landmarks are captured both indoors and outside, in a variety of lighting circumstances and from a variety of vantage points. In addition, there will be images that are connected to well-known ones. *Public availability*. The dataset aims to help the research community solve the scarcity of Egyptian ancient monument datasets that face researchers in this domain [28]. Our dataset was explicitly built to account for these difficulties.

B. Data Distribution

The Egypt Monuments Dataset v1 contains 41 diverse monuments and heritage sites from 6 out of the 28 governorates of Egypt; As indicated in Table II, this dataset of ancient Egyptian monuments and landmarks is truly unique and one-of-a-kind. By far, statues are the most frequent type, followed by temples, then pyramids. Approximately 37% of the monuments with over 2,000 images are located in Luxor, while about 27% are in Cairo.

C. Dataset Construction

The process of gathering data and constructing the ground truth is described in this section.

1) *Data sources*: WatchiT, a leading Egyptian streaming entertainment platform, and Youtube, a global online video-sharing platform, are the primary sources for the Egypt Monuments Dataset v1. Then there is Wikimedia Commons, the most extensive online collection of user-submitted images, videos, and other media. Millions of photos of famous landmarks, taken by an active community of photographers and partner organizations like libraries, archives, and museums, are available on Wikimedia Commons under Creative Commons and Public Domain licenses. The goal of Wiki Loves Monuments, a yearly competition, is to add more high-quality landmark images to the site, while classifying them based on a detailed taxonomy of cultural heritage sites around the world. Images were also sourced from Google Images in addition to the aforementioned Wikimedia Commons.

2) *Annotation*: Notably, ground-truth annotation is notoriously difficult. Given that it is difficult to predefine what monuments or heritage sites are and that they are only sometimes clearly apparent, identifying monuments is challenging. Furthermore, for certain heritage sites, such as The Giza Pyramids and The Bent Pyramid, images can be captured from a considerable distance.

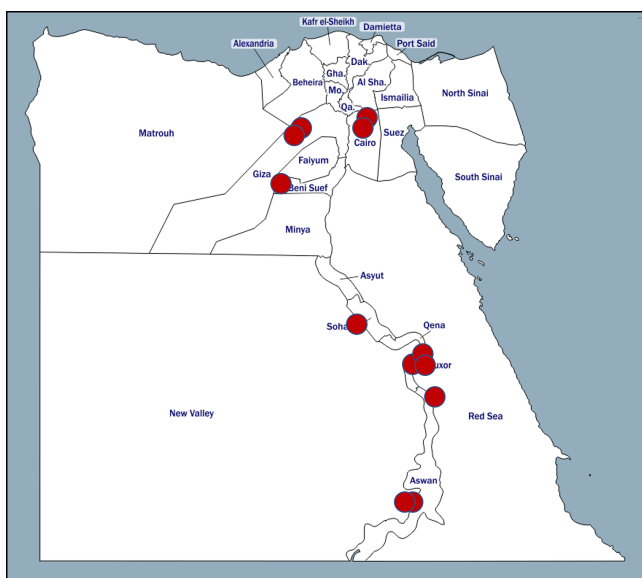


Fig. 2. Egypt map highlighting the distribution of ancient Egyptian monuments and heritage sites across Egypt regarding Egypt Monuments Dataset v1. (Starting from Cairo and reaching to Aswan).

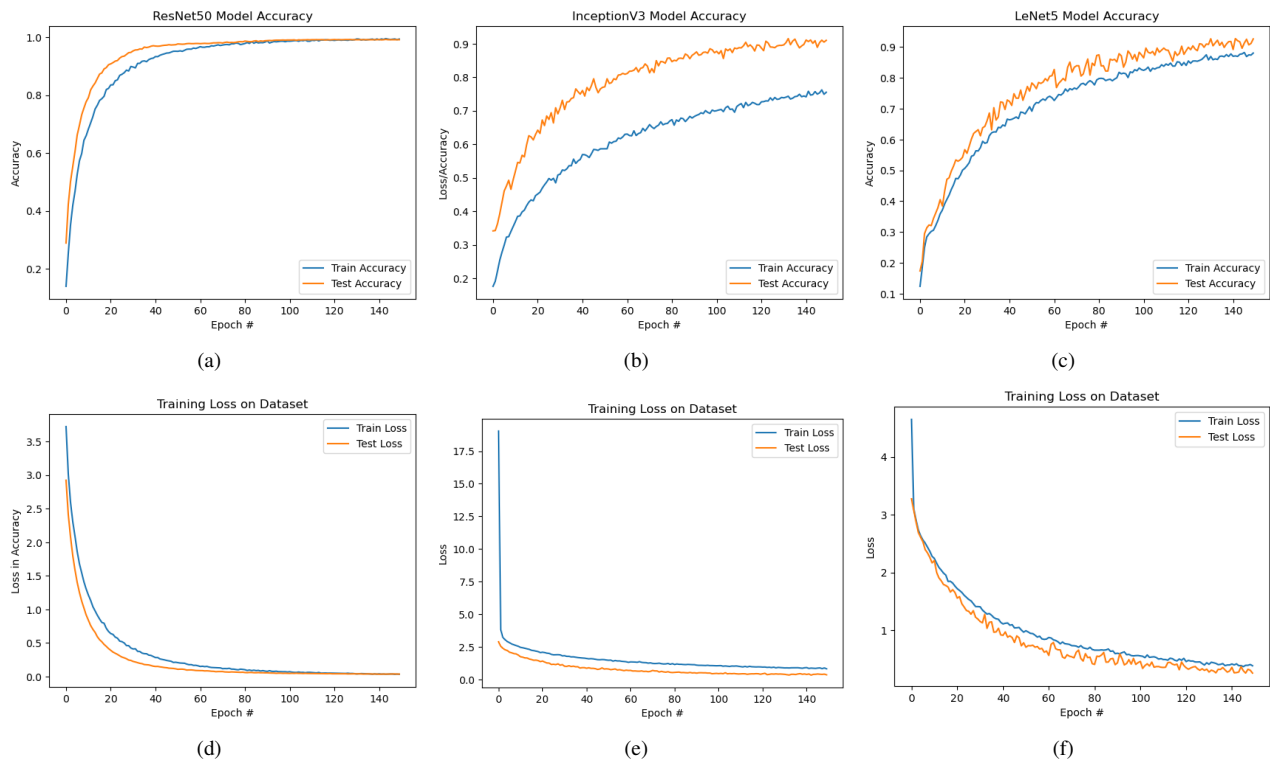


Fig. 3. (a) ResNet50 Model's Accuracy. (b) InceptionV3 Model's Accuracy. (c) LeNet5 Model's Accuracy. (d) Training Loss of ResNet50 Model. (e) Training Loss of InceptionV3 Model. (f) Training Loss of LeNet5 Model.

IV. EXPERIMENT

In this study, the utilization of the dataset is illustrated, and various baseline models that can be used as a reference for future research are introduced. Furthermore, an analysis of the outcomes obtained from the real-world challenge is provided. The findings discussed in this part are all relevant to the ground truth of version 1 of the dataset. Herein, the efficacy of recognizing the visual features of the EGYPT-v1 dataset using cutting-edge classification techniques is evaluated.

Experiments using various advanced deep network architectures have been conducted, such as ResNets [29], Inception V3 [30], and LeNet5 [31] models, to assess the level of classification difficulty in our EGYPT-v1 dataset. To train the models, data augmentation techniques have been employed.

Networks have been fine-tuned using pre-trained weights of ImageNet, optimized with Adam [32] and $1e-05$ for the learning rate. Training and testing have been conducted with images sized at 224×224 .

As shown in Table III, a performance comparison table was created to differentiate the three models. The models under examination are RESNET50, Inception V3, and LeNet 5, all of which are representative of deep learning techniques. The results of our analysis are presented in Table III in the document.

Upon close scrutiny of the graph presented in Fig. 3(a), it can be noted that the testing accuracy of ResNet50 surpasses the training accuracy after 90 epochs. In addition, a trend of improvement is observed in both the training and testing

accuracy curves as the number of epochs progresses. The ResNet50 model attains 99.69% for training accuracy and 99.13% for testing accuracy at the completion of the training process. The dissimilarity between these two accuracy values is negligible, which indicates that the model is not partial to training images, and can perform with comparable efficiency in recognizing unobserved images. In contrast, as depicted in Fig. 3(d), the loss function performance is nearly indistinguishable for both curves. It is worth noting that saturation is observed for both curves at around epoch 110. The ability of the loss function to produce consistent results suggests that the model is not prone to overfitting concerns and can distinguish unknown data as efficiently as it classifies recognized data.

In Fig. 3(b), upon completion of training, InceptionV3's accuracy on the training data was 93.31%, while its accuracy on testing data was 90.90%. The substantial gap in accuracy scores indicates overfitting, a situation in which a model excels on training data while it shows inferior performance on new, unseen data. Overfitting can be addressed by reducing the model's complexity, increasing the training data volume, or implementing regularisation techniques. In Fig. 3(e), if the loss graph is examined closely, it can be seen that initially, the testing loss was significantly lower than the training loss. The training and testing loss demonstrated a decreasing trend as the epochs' number increased. Based on Fig. 3(f), it appears that the model has not fully converged and may benefit from additional training epochs. To confirm this speculation, an additional experiment was conducted, as shown in Fig. 4. The InceptionV3 model was trained on the same dataset with 600 epochs and achieved an accuracy of 96% with a loss curve

TABLE II. INSIGHTS OF THE EGYPT-V1 DATASET, CONTAINING 41 DIVERSE CLASSES DISTRIBUTED IN 6 GOVERNORATES OF EGYPT

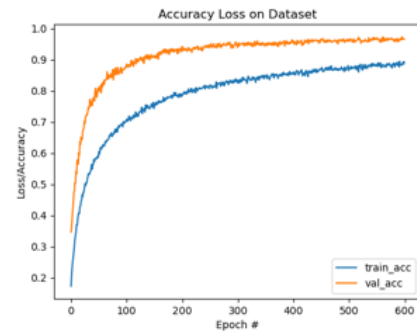
Monument/Heritage Site Name	Category	Located in	# Images
The Great Temple of Ramesses II	Temple	Aswan	1283
Hatshepsut Temple-Deir ElBahari Temple	Temple	Luxor	590
The Shunet El Zebib	Heritage Site	Sohag	558
Saqqara Pyramid-Pyramid of Djoser-Step Pyramid of Djoser	Pyramid	Giza	445
The Great Sphinx of Giza-Abou El Houf	Statue	Giza	358
Bent Pyramid of King Sneferu	Pyramid	Giza	305
The Small Temple of Abu Simbel-Temple of Nefertari-Temple of Hathor	Temple	Aswan	280
Menkaure Pyramid	Pyramid	Giza	267
Khafre Pyramid	Pyramid	Giza	247
Medinet Habu Temple-The Temple of Ramses III	Temple	Luxor	242
Meidum Pyramid of King Sneferu	Pyramid	Beni Suef	213
Head Statue of Akhenaten	Statue	Luxor	206
Karnak Temple	Temple	Luxor	205
Architect Senenmut with Princess Neferu-Ra	Statue	Cairo	193
Red Hatshepsut Shrines	Shrine	Luxor	166
Malkata Palace-Amenhotep III Palace	Palace	Luxor	152
Statue of King Zoser	Statue	Cairo	152
Mask of Tutankhamun	Mask	Cairo	147
King Amenhotemp Shrine	Shrine	Luxor	141
Abu Simbel Temples	Temple	Aswan	140
Sacred Lake	Lake	Luxor	139
Mastaba	Tomb	Giza	136
Khufu Pyramid	Pyramid	Giza	124
Statue of Queen Hatshepsut	Statue	Luxor	124
Amenhotep III Template	Temple	Luxor	119
Queen Hatshepsut Mummy	Mummy	Cairo	106
Court of King Thutmose I	Heritage Site	Luxor	105
Bust Statue of Akhenaten	Statue	Luxor	75
The Great Sun Court of Aton	Heritage Site	Luxor	72
Head Statue of King Hatshepsut	Statue	Cairo	68
Tutankhamun Coffin-Tutankhamun Sarcophagus	Sarcophagus	Cairo	66
Statue of Tutankhamun with Ankhese-namun	Statue	Luxor	58
Statue of Akhenaten	Statue	Cairo	55
Tomb of King Den	Tomb	Sohag	48
King Thutmose II Mummy	Mummy	Cairo	43
Giza Pyramids	Pyramid	Giza	41
Goddess Isis with her child	Statue	Cairo	41
King Thutmose II	Statue	Luxor	21
Another Statue of Akhenaten	Statue	Cairo	20
Statue of Princess Meketaton	Statue	Cairo	16
Temple of Edfu	Temple	Aswan	11
41	11	6	7,778

that plateaued at 60%. The model’s performance may benefit from more training epochs, as indicated by these results.

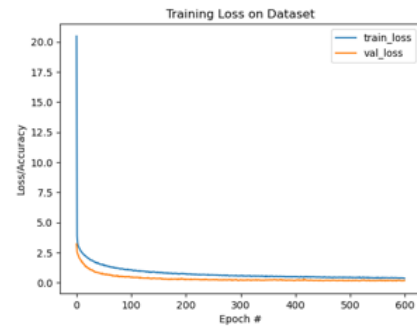
In Fig. 3(c), the graph demonstrates that the training and testing accuracy curves of the LeNet5 model display a consistent upward trend as the epochs’ number increases. Notably, the testing accuracy is continuously more unstable than the training accuracy curve. Upon completion of the training process, it attains 94.29% for training accuracy and 92.64% for testing accuracy. Besides Fig. 3(f), when analyzing the

TABLE III. EVALUATION METRICS

Model	Type	Measurement			
		Accuracy	Precision	Recall	F1-Score
ResNet50	Train	99.69%	99.73%	99.67%	99.70%
InceptionV3		93.31%	96.52%	90.53%	93.37%
LeNet5		94.29%	95.20%	93.17%	94.16%
ResNet50	Test	99.13%	99.28%	99.13%	99.20%
InceptionV3		90.90%	87.90%	94.21%	90.87%
LeNet5		92.64%	94.13%	91.34%	92.69%



(a)



(b)

Fig. 4. (a) Accuracy of inceptionV3 model for 600 epochs (b) Loss curve of inceptionV3 model for 600 epochs.

loss graph, it is evident that, in the beginning, the loss during testing was lower than the loss during training. However, as the epochs’ number progressed, both the training and testing loss curves showed a decreasing trend.

In terms of generalization, an experiment has been conducted with the ResNet50 model, as it is the highest accuracy among the three models. It has predicted new, unseen data of the same domain with an accuracy of 97.43%. Notably, the unseen data size is over 35 thousand images suggesting that the dataset is scalable.

V. CONCLUSION

The Egypt Monuments Dataset v1 is introduced as a new benchmark for classification and instance recognition on a country-wide scale. Unlike many current computer vision datasets, this dataset has the following features: 1) it was gathered by individuals who are not computer vision professionals for a specific goal, making it unbiased; 2) unlike previous datasets, it is a better representation of real-world challenges; 3) it presents a classification challenge with a long-tail distribution; and 4) it has practical applications in the fields of conservation and Egyptology.

In terms of domain coverage, the EGYPT-V1 dataset demonstrates scalability by covering most of the famous pharaonic monuments allocated in Egypt. In addition, the proposed approach performs well across different subcategories. The approach’s ability to perform well on a diverse range of data within the same domain suggests that the dataset is scalable.

Future plans involve undertaking object detection tasks. It is also planned to increase the size of the dataset where Egypt, in particular, is home to one-third of the world's monuments. This approach would promote the development of new methods for measuring errors. Finally, it is anticipated that this dataset will have value in examining how to train humans in recognizing intricate visual classes, and experimentation with human learning models is intended.

ACKNOWLEDGMENT

The authors extend their gratitude to Cercle, a live-stream media platform specializing in the film and broadcast of DJ sets and live performances within cultural heritage sites and landmarks through electronic music and video mediums. We would like to acknowledge Cercle for their collaboration in providing cultural heritage videos for our use. The authors also express their appreciation to WatchiT, a video-on-demand service for Arabic content online, with a unique collection of diverse entertainment options, including movies and TV shows, for their support of our work.

REFERENCES

- [1] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman, "Object retrieval with large vocabularies and fast spatial matching," *2007 IEEE conference on computer vision and pattern recognition*, pp. 1–8, 2007.
- [2] H. Jgou, F. Perronnin, M. Douze, J. Snchez, P. Prez, and C. Schmid, "Aggregating local image descriptors into compact codes," *IEEE transactions on pattern analysis and machine intelligence*, vol. 34, no. 9, pp. 1704–1716, 2012.
- [3] A. Gordo, J. Almazán, J. Revaud, and D. Larlus, "Deep image retrieval: Learning global representations for image search," *European conference on computer vision*, pp. 241–257, 2016.
- [4] F. Radenović, A. Iscen, G. Tolias, Y. Avrithis, and O. Chum, "Revisiting oxford and paris: Large-scale image retrieval benchmarking," *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 5706–5715, 2018.
- [5] Y. Kalantidis, L. G. Pueyo, M. Trevisiol, R. van Zwol, and Y. Avrithis, "Scalable triangulation-based logo recognition," *Proceedings of the 1st ACM international conference on multimedia retrieval*, pp. 1–7, 2011.
- [6] R. Del Chiaro, A. D. Bagdanov, and A. Del Bimbo, "Noisyart: A dataset for webly-supervised artwork recognition," *VISIGRAPP (4: VISAPP)*, pp. 467–475, 2019.
- [7] A. Krizhevsky and G. Hinton, "Convolutional deep belief networks on cifar-10," *Unpublished manuscript*, vol. 40, no. 7, pp. 1–9, 2010.
- [8] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein *et al.*, "Imagenet large scale visual recognition challenge," *International journal of computer vision*, vol. 115, pp. 211–252, 2015.
- [9] A. Kuznetsova, H. Rom, N. Alldrin, J. Uijlings, I. Krasin, J. Pont-Tuset, S. Kamali, S. Popov, M. Mallocci, A. Kolesnikov *et al.*, "The open images dataset v4," *International Journal of Computer Vision*, vol. 128, no. 7, pp. 1956–1981, 2020.
- [10] J. Turner, *The Dictionary of Art*. Macmillan, 1996.
- [11] A. Elnagar and A. Derbali, "The importance of tourism contributions in egyptian economy," *International Journal of Hospitality and Tourism Studies*, vol. 1, no. 1, pp. 45–52, 2020.
- [12] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman, "Lost in quantization: Improving particular object retrieval in large scale image databases," *2008 IEEE conference on computer vision and pattern recognition*, pp. 1–8, 2008.
- [13] H. Jegou, M. Douze, and C. Schmid, "Hamming embedding and weak geometry consistency for large scale image search-extended version," 2008.
- [14] Y. Avrithis, G. Tolias, and Y. Kalantidis, "Feature map hashing: Sub-linear indexing of appearance and global geometry," *Proceedings of the 18th ACM international conference on Multimedia*, pp. 231–240, 2010.
- [15] J. Knopp, J. Sivic, and T. Pajdla, "Avoiding confusing features in place recognition," *European Conference on Computer Vision*, pp. 748–761, 2010.
- [16] S. Agarwal, Y. Furukawa, N. Snavely, I. Simon, B. Curless, S. M. Seitz, and R. Szeliski, "Building rome in a day," *Communications of the ACM*, vol. 54, no. 10, pp. 105–112, 2011.
- [17] D. M. Chen, G. Baatz, K. Köser, S. S. Tsai, R. Vedantham, T. Pylvänäinen, K. Roimela, X. Chen, J. Bach, M. Pollefeys *et al.*, "City-scale landmark identification on mobile devices," *CVPR 2011*, pp. 737–744, 2011.
- [18] Y. Li, N. Snavely, D. P. Huttenlocher, and P. Fua, "Worldwide pose estimation using 3d point clouds," *Large-Scale Visual Geo-Localization*, pp. 147–163, 2012.
- [19] K.-H. Yap, Z. Li, D.-J. Zhang, and Z.-K. Ng, "Efficient mobile landmark recognition based on saliency-aware scalable vocabulary tree," *Proceedings of the 20th ACM international conference on Multimedia*, pp. 1001–1004, 2012.
- [20] A. Torii, R. Arandjelovic, J. Sivic, M. Okutomi, and T. Pajdla, "24/7 place recognition by view synthesis," *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1808–1817, 2015.
- [21] T. Weyand and B. Leibe, "Visual landmark recognition from internet photo collections: A large-scale evaluation," *Computer Vision and Image Understanding*, vol. 135, pp. 1–15, 2015.
- [22] H. Noh, A. Araujo, J. Sim, T. Weyand, and B. Han, "Large-scale image retrieval with attentive deep local features," *Proceedings of the IEEE international conference on computer vision*, pp. 3456–3465, 2017.
- [23] V. Sharma, "Qutub complex monuments' images dataset," Oct 2018, Accessed: Sept. 23, 2022. [Online]. Available: <https://www.kaggle.com/datasets/varunsharmaml/qutub-complex-monuments-images-dataset>
- [24] R. Gupta, P. Mukherjee, B. Lall, and V. Gupta, "Semantics preserving hierarchy based retrieval of indian heritage monuments," *Proceedings of the 2nd Workshop on Structuring and Understanding of Multimedia heritAge Contents*, pp. 5–13, 2020.
- [25] T. Weyand, A. Araujo, B. Cao, and J. Sim, "Google landmarks dataset v2-a large-scale benchmark for instance-level recognition and retrieval," *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 2575–2584, 2020.
- [26] Y. Avrithis, Y. Kalantidis, G. Tolias, and E. Spyrou, "Retrieving landmark and non-landmark images from community photo collections," *Proceedings of the 18th ACM international conference on Multimedia*, pp. 153–162, 2010.
- [27] F. Radenović, G. Tolias, and O. Chum, "Fine-tuning cnn image retrieval with no human annotation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 41, no. 7, pp. 1655–1668, 2018.
- [28] S. Hesham, R. Khaled, D. Yasser, S. Refaat, N. Shorim, and F. H. Ismail, "Monuments recognition using deep learning vs machine learning," *2021 IEEE 11th annual computing and communication workshop and conference (CCWC)*, pp. 0258–0263, 2021.
- [29] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.
- [30] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2818–2826, 2016.
- [31] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
- [32] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

Exploring the Joint Potential of Blockchain and AI for Securing Internet of Things

Md. Tauseef, Manjunath R Kounte, Abdul Haq Nalband, Mohammed Riyaz Ahmed
School of Electronics and Communication Engineering, REVA University, Bengaluru, India

Abstract—The emergence of the Internet of Things (IoT) has revolutionized the way we interact with the physical world. The rapid growth of IoT devices has led to a pressing need for robust security measures. Two promising approaches that can enhance IoT security are blockchain and artificial intelligence (AI). Blockchain can offer a decentralized and tamper-proof framework, ensuring the confidentiality and integrity of IoT data. AI can analyze large volumes of real-time data and detect anomalies in response to security threats in the IoT ecosystem. This paper explores the potential of these technologies and how they complement each other to provide a secured IoT system. Our main argument is that combining blockchain with AI can provide a robust solution for securing IoT networks and safeguarding the privacy of IoT users. This survey paper aims to provide a comprehensive understanding of the potential of these technologies for securing IoT networks and discuss the challenges and opportunities associated with their integration. It also provides a discussion on the current state of research on this topic and presents future research directions in this area.

Keywords—IoT; blockchain; AI; security; attacks; decentralization

I. INTRODUCTION

The emergence of Industry 4.0 in the 21st century marked a significant change in the industrial paradigm, leading to improvements in social, economic, and political conditions[1]. Industry 4.0 enables the use of cyber-physical systems such as the IoT, big data analytics, cloud manufacturing, and fog computing, supported by cutting-edge technologies such as blockchain, AI, and mobile networks [2]. The integration of IoT, blockchain, and AI enhances human-machine interaction and brings the physical and digital worlds closer than ever before [3]. These technologies are expected to offer numerous benefits and opportunities, including self-awareness, self-prediction, self-comparison, self-reconfiguration, and self-maintenance [4].

The IoT is the driving force behind Industry 4.0, enabling seamless inter-connectivity of various devices and objects to construct a network infrastructure that continuously regulates and manages sensing, processing, and communication processes without human intervention [5]. The daily introduction of new applications and services is one of the many advantages of the IoT system. According to Statista [6], in 2020, there were about 50 billion IoT devices worldwide. By the end of 2025, that number is expected to reach over 75 billion devices.. Fig. 1 describes the timely growth in the number of IoT devices. The IoT market is expanding at an almost exponential rate. It was valued at USD 743 billion in 2015, which increased to approximately USD 1710 billion by the end of 2019 [7]. The global IoT market is expected to be worth

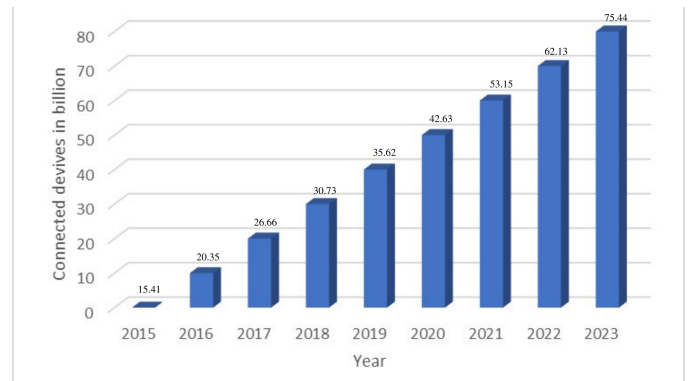


Fig. 1. Growth of IoT devices over the time 2015 to present.

186 billion dollars by the end of 2023, while the market for intelligent homes will be around 130 billion.

IoT has predominantly adopted the centralized architecture model for storing and processing sensor data. The central server serves as the network's manager, handling all requests from various nodes and overseeing task scheduling and distribution [8]. It saves costs by not requiring the installation of multiple workstations of hardware and software, as most processing tasks are managed by the centralized server. However, this architecture model has several challenges, including scaling issues due to the increasing number of IoT devices and various security and privacy concerns [9]. Table I summarizes the challenges posed by centralized IoT architecture. It is challenging to address the fundamental security concerns for such a large information ecosystem. Additionally, the centralized IoT model is susceptible to security breaches, single points of failure, and malicious assaults like DDoS and Sybil attacks[10][11].

Blockchain technology can help to address key security requirements in IoT, thanks to its "security by design" feature. The majority of IoT's architectural flaws can be fixed using blockchain's characteristics, including immutability, transparency, auditability, data encryption, and operational resilience. Blockchain is a decentralized network where all users have full control over peer-to-peer (P2P) monitoring of all network transactions[12]. Another technology that can strongly influence IoT is AI. Integration of AI into IoT devices can make them smart and intelligent [13]. AI can be used to train machines to understand novel material based on the training process they have already undergone. The goal of AI in IoT is to use IoT devices to gather relevant data and draw useful conclusions from that data. AIoT, or the merging of AI and the IoT, can be used to improve data analysis, human-machine

TABLE I. SUMMARY OF CHALLENGES IN CENTRALIZED MODEL OF IOT

Reference	IoT Challenge	Description
[9]	Security	Security is one of the primary issues (particularly DoS attacks) with centralized IoT design, because all data storage and processing operations are carried out by a single central server.
[10]	Single Point of Failure	Single point of failure is a problem since the server controls connections and carries out all processing tasks. If the server fails, the entire network of devices will stop working.
[14]	Access and Diversity	System should be accessible for all users with their dynamic needs. Yet, the centralized system requires that users access the data uniformly by adhering to the same protocols. Also, most of the centralized systems only support a single operating system, which restricts network diversity. IoT system includes heterogeneous and diverse devices, this will result in a serious issue that needs to be handled.
[15]	Inflexibility	A significant workload is generated by the centralised server's control over activities of communication and processing between all IoT network nodes. The centralized server schedules the workload to manage this workload and avoid peak-load problems. However, due to the constrained schedule and associated delay, this restricts user flexibility while performing their own duties.
[16]	Privacy	Sensitive data is among the real-time data types that IoT devices collect, such as habits, password, financial and personal information, etc. The centralised third-party server, which has complete control over this acquired data, keeps them all in one place while also violating their privacy. However, keeping it in one place could make it more vulnerable to intrusions.
[17]	Cost	The network's central server handles all of the communication and processing tasks, which place a heavy demand on the hardware and software needed to handle the workload. It also requires sizable storing storages that can hold data from diverse IoT devices.
[18]	Scalability	Among the main issues linked with the centralised architecture is scalability. Controlling and attempting to control all the network's nodes via a centralized server can scale successfully only in small networks. Using centralized system to big corporate organizations will be illogical. The centralised solution cannot scale and operate well since the number of Internet of Things devices is constantly growing.

interactions, and IoT operations [19]. The integration of decentralized AI has grown recently, enabling the execution and storage of an investigation or dynamic on verified, carefully annotated, and shared data on the blockchain in an automated, decentralized manner without middlemen[20][21]. Blockchain-based AI techniques can provide decentralized reasoning on how to promote safe and trust in the exchange of information and decision results over numerous independent operators who can contribute, organize, and vote on additional decisions [22]. This paper presents the joint potential of blockchain and AI for securing the industrial IoT system. The contributions made by this paper are as follows:

- Details of numerous security concerns in IoT ecosystem are described.
- Presented a concise literature review on security issues and their solutions.
- An architectural framework that integrates blockchain and AI for a secured IoT has been proposed.

The remainder of the study is organised as follows; in Section II, we go over the background of the IIoT, Blockchain, AI, and security concerns with IoT. We provide a full assessment of the literature on security-related issues, challenges, and solutions in Section III. In addition, this part illustrates how blockchain and AI combine to create a safe IoT environment. We provide the proposed architectural framework for combining blockchain and AI to improve IoT security in Section IV. We present high-level blockchain-driven AI and AI-driven blockchain as examples of how AI and Blockchain can be applied in IoT use cases. Section V provides the conclusion to the research findings.

II. BACKGROUND

The Industrial Internet of Things (IIoT) emerged from the intersection of manufacturing technology, industrial automation, and data sharing. Fig. 2 illustrates the timeline of the significant events that have contributed to IIoT's development. The evolution of Industry 4.0 has been fuelled by numerous game-changing advancements, including the availability of low-cost sensors, big data analytics, AI and machine learning (ML), the IoT, robotics, edge computing, standard communication protocols, and enhanced security technologies such

as encryption, authentication, and intrusion detection. As IIoT continues to progress, it has the potential to revolutionize various industries, such as manufacturing, logistics, healthcare, and energy. The impact of IIoT on these sectors is expected to be substantial, with efficiency, safety, and productivity improvements. For example, IIoT devices can help identify bottlenecks and optimize production processes in manufacturing, while in healthcare, connected devices can monitor patient health and aid in timely interventions. Overall, the IIoT represents a significant opportunity for businesses to transform their operations and unlock new value.

A. Evolution of Industry 4.0

The development of Industry 4.0 can be traced through several key milestones. It began with the digitization of production processes, which involved integrating digital technology like sensors and smart machines into industrial equipment and systems. This increased automation and data collection, laying the foundation for the next phase. The second phase involved the development of interconnected factories and supply networks. This was made possible through the adoption of IoT technologies, which allowed machines and gadgets to communicate with each other and share data in real time. Connected factories and supply chains became a reality, and this marked a significant advancement in Industry 4.0's evolution. The third phase saw the emergence of advanced analytics and artificial intelligence technologies that could analyse and make sense of the massive amounts of data produced by connected factories and supply networks. These technologies are employed to raise productivity, anticipate and avoid equipment breakdowns, and improve the production cycle. As machines and robots become more sophisticated, a collaboration between humans and machines has become a primary focus of Industry 4.0. This involves developing systems that allow people and machines to coexist in harmony, with machines handling routine jobs and people focusing on those that require creativity and problem-solving abilities. With the increased use of digital technology, cybersecurity has emerged as a critical issue for Industry 4.0. Companies have invested in technology and procedures to secure their systems from cyber-attacks, and this has become a hallmark of Industry 4.0's development.

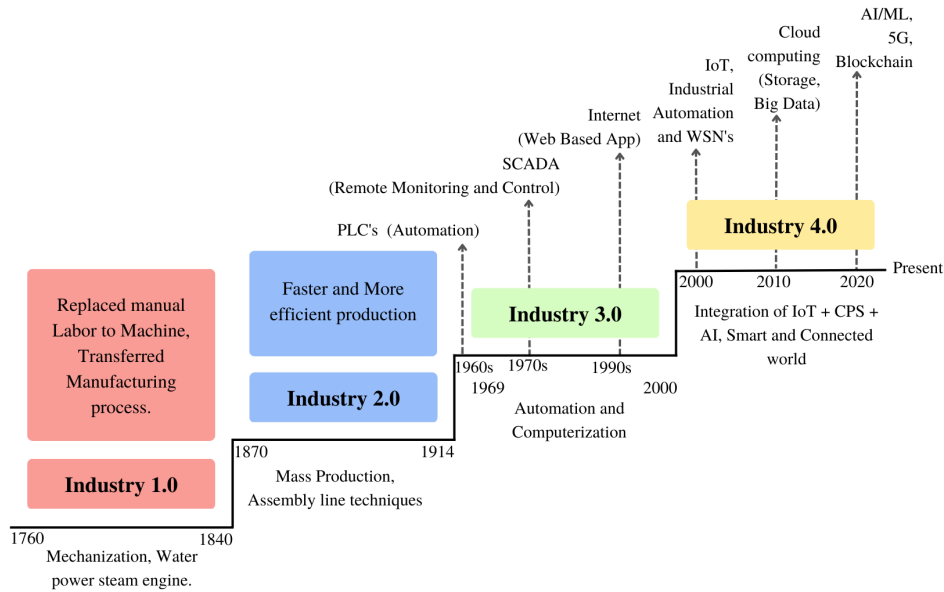


Fig. 2. Evolution of industrial revolution 4.0.

TABLE II. SECURITY ATTACKS IN IOT ECOSYSTEM

IoT Level	IoT Layer	IoT Protocol	IoT Security Attacks
Deployment	Application	CoAP, MQTT, AMQP, REST	DDoS Attack, Repudiation Attack, SQL Injection Attack, Cross-Site Scripting, Parameter Tampering, Slowloris Attack, HTTP Flood Attack
Data	Transport, Network	UDP, TCP, DCCP, RSVP, SCTP, CLNS, QUIC, DDP, IGMP, EIGRP, IPsec, ICMP, IPv6, IPv4, RIM, OSPF	Smurf Attack, SYN Flood, Mitnick Attack, Injection Attack, DoS Attack, Opt-ack Attack, IP Address Spoofing, Worm Hole Attack, Byzantine Attack, Resource Consumption Attack, Black Hole Attack
Device	Physical	ISDN, DSL, USB, IDA, CAN, Bluetooth, Ethernet	Access Control Attack, Disconnection of Physical Links, Physical damage

B. Security Concerns in IoT

The IoT devices gather and transmit vast amounts of data, including sensitive or private information such as medical records or video footage from home security systems. Ensuring the security of this data is critical to protect privacy and prevent identity theft. Unfortunately, IoT devices are frequently targeted by cybercriminals because they are often insecure and provide an easy entry point to a larger network. Successful cyber-attacks can result in data breaches, financial losses, and even physical harm. Additionally, many critical infrastructure systems, such as power grids and transportation systems, are controlled by a large number of IoT devices [23]. If these systems are compromised, the consequences can be severe, including widespread disruption, financial losses, and even loss of life. It is essential that IoT devices adhere to strict data privacy and security regulations in industries such as healthcare and banking to avoid legal repercussions and reputational harm. To ensure the security of IoT devices, it is necessary to implement security measures at three different levels: device, data, and deployment [24]. These levels correspond to the IoT's architectural layers, with specific protocols used in each tier of the IoT architecture to protect against related security attacks (Table II). Techniques such as encryption, authentication, and access control can be used to secure sensitive data from unauthorized access. IoT devices are often deployed in remote or harsh environments, making them vulnerable to physical attacks such as theft or manipulation. Physical security mea-

asures such as tamper-evident seals, firmware upgrades, and patches can help guard against these risks. Additionally, cyber attackers can compromise IoT devices and use malware such as DDoS, HTTP flood, SQL injection, and parameter tampering to control the devices and exploit their vulnerabilities. It is crucial to protect against these threats by regularly updating firmware and applying security patches.

C. Blockchain for IoT Security

Blockchain technology offers solutions to several security challenges facing IoT systems, such as data privacy, data integrity, and device authentication. By providing a secure, decentralized, and immutable system, it can defend against DDoS attacks, data manipulation, and unauthorized access. Its architecture combines hash algorithms with decentralized ledgers that use public and private keys, offering a potent alternative to the internet [25]. Blockchain ensures secure data storage and can prevent rogue IoT devices from entering the network. It can also reduce the costs of litigation caused by disagreements [26]. Transactions are protected by a consensus mechanism that ensures their integrity even in the presence of faults or hostile conditions, thus enabling a stable blockchain [27]. Moreover, blockchain can be combined with smart contracts to increase dependability and radio-frequency identification is safeguarded by attribute-based access control mechanisms [28]. In blockchain, each block contains its data, the previous block's hash, and the security hash code. It can offer secure

TABLE III. ENABLERS OF SECURED IOT SYSTEM

		Blockchain for IoT Security
Data privacy	PD	Data can be protected by using blockchain to store it in a decentralized, tamper-proof manner. Blockchain's public-key encryption ensures that only authorized parties can access the data.
Data integrity		Data integrity is guaranteed by the tamper-proof ledger in blockchain, which is immune to hacking and data modification. This guarantees the accuracy and dependability of the data that IoT devices collect.
Device authentication		Blockchain offers a safe and decentralized method for IoT device authentication. Each gadget can be given an own digital identity, making it simple to trace it down and confirm its legitimacy
DDoS		Blockchain establishes a distributed network of nodes capable of validating and verifying data sent by IoT devices. By limiting the amount of devices that may access the network, this can assist avoid DDoS attacks.
		AI for IoT Security
Anomaly detection		AI assists in identifying potential security concerns, such as malicious activity or unexpected network traffic, and alerting security teams to take appropriate action by detecting anomalies in the behaviour of IoT devices and networks.
Predictive maintenance		IoT device monitoring and forecasting of potential failures and maintenance needs. By doing this, security lapses brought on by unsecured or infected devices may be avoided.
Behavioral analytics		Investigate trends in the actions of IoT users and devices to look for potential security risks. An IoT device may have experienced a security breach if it suddenly starts transmitting significant amounts of data at odd hours.
Cyber threat intelligence		To address potential security concerns in IoT devices and networks, collect and evaluate data about known cyber threats and vulnerabilities. This can assist businesses in taking preventative measures to safeguard their networks and devices from threats.
Identity and access management		For IoT networks and devices, AI can be utilised to enhance identification and access control. In order to assist prevent unauthorised access to devices and networks, this may also include user authentication, authorisation, and access control.

data transfer over IoT device nodes, and its lightweight and sustainable algorithm employs decentralized techniques for authentication in distributed resource-constrained systems [29]. Certain companies might be capable to handle dependability issues with a double-chain design that uses data and transaction blockchain for storage, distribution, and data reliability, but there are risks related to privacy vulnerability [30]. Various studies have examined the use of blockchain in securing IoT, such as [31] and [32], which provided an overview of the security issues with IoT and how blockchain can be applied to address them. However, [33] critically analyzed the limits of using blockchain-based IoT platforms, offered a taxonomy of blockchain typologies, weighing their benefits and drawbacks in incorporating them into IoT. Several ideas have been proposed for using blockchain and AI technologies to address specific security issues of IoT, such as allowing smart contracts, IoT devices to update their firmware, or developing a system in which devices can acquire “money” by exchanging resources or data for goods or services [34]-[36]. Table III summarizes how blockchain and AI technologies individually help address security concerns of IoT.

D. AI for IoT Security

The integration of AI into the IoT devices can significantly enhance their performance and security. AI can simulate human learning processes such as decision-making, problem-solving, and object identification, making IoT devices intelligent and more efficient. By combining AI and IoT, businesses can generate useful data and obtain insights, making their operations more effective [37]. One of the main benefits of using AI in IoT is its ability to quickly identify potential security threats and holes. AI can examine network traffic and user activity trends, identify vulnerabilities, and prioritize them according to their severity [38]. With AI, businesses can automate the patching of IoT devices, ensuring that they are up-to-date with the latest security patches, reducing the chance of security breaches [39]. AI can also track user and IoT device behaviour, detecting any unusual activity that might point to a security breach, allowing organizations to respond quickly to threats. Secure access control, offloading, and virus detection can be provided by AI-based authentication employing machine learning to preserve data privacy [40]. By

TABLE IV. IOT SECURITY CONCERNS ADDRESSED BY BLOCKCHAIN AND AI

Security Issues Addressed by Blockchain	Security Issues Addressed by AI
Verification of Identity	Intrusion detection system
Self-healing and detection of firmware	Malware detection
Address space and privacy preservation	Anomaly detection
Secured communication and data integrity	Unauthorized IoT devices identification
Authorization and authentication	Distributed denial-of-service
Information sharing and access control	Jamming attack, Spoofing attack
Secure computation and storage	Authentication, Eavesdropping
Trust-management	False data injection, Impersonation

using the analytical capabilities of AI, businesses can uncover patterns and make better decisions about IoT data collection. AI can enhance the object identification technique utilized by smart cameras, enabling them to recognize dangerous objects like knives and guns instantly. However, AIoT frameworks are prone to data security and privacy issues. To address these issues, researchers are developing solutions such as blockchain, which offers greater protection than conventional security measures since it cannot be altered. Many AIoT applications can incorporate blockchain to boost security. The security concerns that blockchain and AI address when used independently with IoT are listed in Table IV. Overall, the combination of AI and IoT can significantly enhance businesses' operations and security. By using AI, businesses can identify potential threats quickly, automate patching, and make better decisions about data collection. Integrating blockchain into AIoT applications can enhance security, ensuring data privacy and reducing the risks of security breaches.

E. Convergence of Blockchain and AI for IoT Security

The combination of blockchain, IoT, and AI is becoming exceptionally important in the realm of digital transformation, particularly for IoT security. With this convergence, new business strategies are emerging that involve the creation of autonomous profit centers made up of autonomous agents, such as sensors, vehicles, machines, trucks, cameras, and other IoT devices. These autonomous agents will have digital twins through IoT, enabling them to independently send and receive money through blockchain technology and make decisions using AI and data analytics. As a result, we predict that

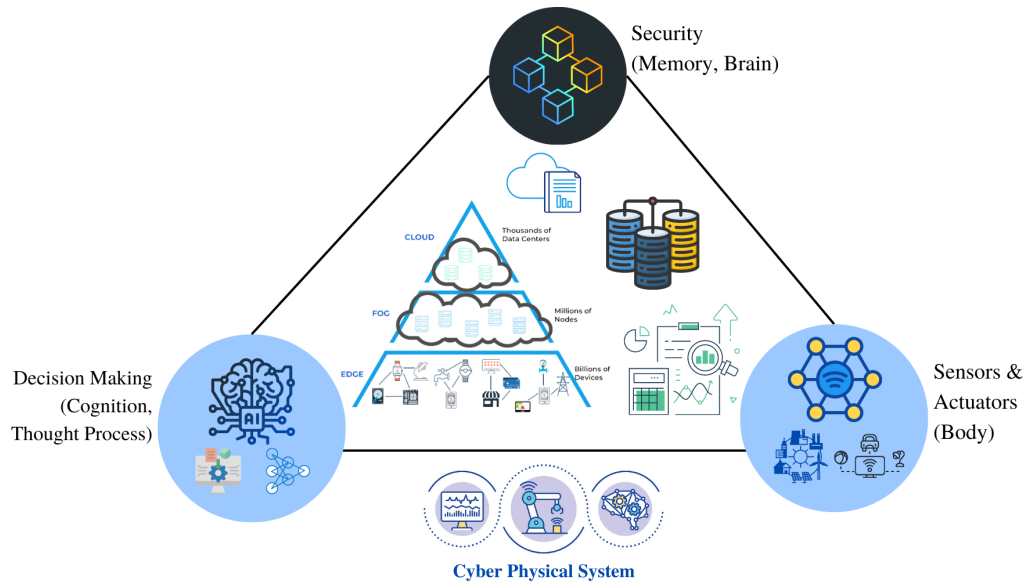


Fig. 3. Convergence of blockchain, AI towards secured IoT.

industrial firms will undergo a digital transition as these autonomous business models continue to evolve.

In recent years, IoT devices have collected an enormous amount of data centrally, causing security and space issues [41]. However, the integration of AI, blockchain and IoT offers a solution to this problem by creating a distributed database [42]. Decentralized AI, which is a growing concept, integrates these two technologies to allow for distributed, unmediated execution and storage of data on the blockchain without middlemen. Blockchain is expected to be a reliable platform for storing vast volumes of data that AI works with. With the blockchain's smart contract functionality [43], it is possible to keep track of member participation and transactions while accessing information. An autonomous system and machine can then choose exact and reliable option outcomes that are verified and acknowledged by all blockchain mining centers [44]. These decisions cannot be contested and can be supported by anyone with a stake in the outcome.

Blockchain, AI, and IoT have successfully combined, as seen in Fig. 3. The integration of these technologies can address several critical issues such as accuracy, latency, centralization, privacy and security issues in IoT. Decentralized reasoning provided by blockchain-based AI algorithms can benefit a large number of independent operators who may contribute, organize, and vote on new decisions [45]. Blockchain databases have hash values that are digitally signed to ensure secure and reliable processing of transactions. AI algorithms are used to solve issues related to accuracy, latency, security, and privacy, and to enhance big data analysis. The decentralization of blockchain networks eliminates the single point of failure in a cloud server, making it more efficient and reliable for data analysis. By integrating AI and blockchain technology, the IoT can be aided with better decision-making capabilities.

To further enhance the integration of blockchain, AI, and IoT, there are several areas that need to be addressed. For

instance, the integration of these technologies requires a strong and secure infrastructure, as well as effective governance and regulatory frameworks. There is also a need to address the ethical and social implications of using these technologies, such as issues related to data security, privacy and bias. Moreover, the adoption of blockchain, AI, and IoT will require a significant investment in research and development, as well as in the training of personnel to effectively manage and maintain these technologies. Nevertheless, the benefits of this convergence are substantial, including increased efficiency, improved security, and better decision-making processes. As these technologies continue to evolve, it is likely that we will see even more innovative ways to combine them. The convergence of blockchain, AI, and IoT has the potential to transform the way we live, work, and interact with one another. It is an exciting time to be at the forefront of this convergence and to witness the transformative power of technology.

The Big Data revolution has been a key driver of the AI revolution, as it enables businesses to divide vast databases into manageable, organized components. Additionally, the value of data has propelled the development of blockchain technology, as its distributed ledger offers a new, more efficient way to store data. This combination has the potential to significantly transform how information is handled, analyzed, and shared, ultimately making operations inside a company more efficient and effective. As AI continues to evolve, we may see the emergence of new AI systems, such as logic-based systems, bio-inspired systems, collaborative agents, cyber-physical intelligent systems, and ubiquitous AI or pervasive intelligence systems, which could fundamentally change our daily lives.

Machine Learning, a branch of AI, has gained significant popularity in recent years due to its ability to analyse huge data flows. It has found applications in various fields such as business, entertainment and research. Many countries and organizations, including Google, Amazon, Facebook, Microsoft,

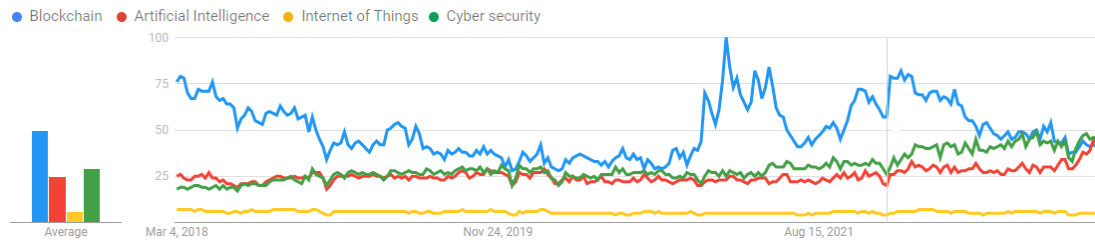


Fig. 4. Research trends in blockchain, AI, IoT and cyber security.

IBM, Apple, Intel, Alibaba, Baidu, and Uber, have invested heavily in developing cutting-edge ML goods, programs and platforms that can be used as cloud services to run ML models effectively, known as MLaaS. Cloud computing is a popular and cost-effective platform for running ML models used in many IoT and smart city services and applications, such as text classification, automated language processing, facial recognition, speech recognition, computer vision and speech synthesis. Additionally, several nations have launched ambitious national AI initiatives in less than a year [46], indicating the potential impact of ML and AI on various sectors.

The integration of Blockchain, AI, and IoT has been the subject of many experiments and hypotheses, but further research is necessary to develop a functional and reliable digital component that properly combines the three. Cloud computing is crucial in today's world and facilitates online connectivity. However, the vast amount of data handled by this system highlights the need for automated systems with quality of service (QoS) standards. The convergence of Blockchain, AI, and IoT is identified as a critical technology to meet this requirement. Consequently, the extensive use of AI and blockchain integration will bring about exciting advancements in Industry 4.0. Instead of only discussing software, algorithms, automation, robotics, and hardware, businesses are now exploring more sophisticated concepts such as producing and manufacturing items on-demand, dematerialization, and disintermediation. Industry 4.0 emphasizes distinct pillars including information transparency, support, and connection, and it represents the first major revolution that moves from a tech-centric state to a more sophisticated one.

III. LITERATURE REVIEW

A. Security Requirements in IoT

IoT devices need strong security measures to prevent unwanted access and data breaches since, like any linked system, they are susceptible to cyber-attacks. In the IoT ecosystem, some of the primary security requirements are; only devices or users that have been authenticated can access the network. This can be done via techniques including digital certificates, biometric identity, two-factor authentication, and password protection. To prevent unlawful interception, data transferred between IoT devices should be encrypted[47]. Data in transit and at rest should be protected using robust encryption techniques like Advanced Encryption Standard (AES) and RSA algorithm. Access control - It is important to prevent unauthorized access to network data. This group can include additional access control techniques as well as attribute- and

role-based access control. IoT devices should be built to constantly download and install firmware upgrades to address security holes and enhance functionality. In order to prevent malicious firmware updates, devices should provide a method for confirming the integrity of firmware upgrades. Physical security - In order to avoid unwanted access or tampering, IoT devices should be physically secure[48]. These can include steps like secure boot procedures, tamper-resistant designs, and physical security controls like locks and access cards. Privacy - IoT devices should respect user privacy by only gathering information that is required and handling it securely. Users should be clearly informed by devices about the types of data being gathered and how they are being used. Network security - IoT devices should be built to function safely on both public and private networks, among other network environments. This may involve taking precautions like installing firewalls, installing intrusion detection and prevention systems, or using secure communication protocols like VPNs. In order to stay abreast of the most recent security threats and vulnerabilities, a solid security strategy for the IoT ecosystem should be multi-layered, proactive, and constantly developing.

B. Avenues of Cyber Attacks in IoT

The connection between IoT devices offer several opportunities for cyber criminals to find flaws and launch assaults. Here are a few typical IoT ecosystem cyber-attack vectors. Poor authentication methods - IoT devices are susceptible to assaults like brute force attacks, dictionary attacks, and password cracking attacks if they employ weak or default passwords or have no authentication mechanisms at all. Unsecured communication protocols: IoT devices that employ these protocols are susceptible to replay, eavesdropping, and man-in-the-middle attacks. IoT devices that run out-of-date or unpatched software may contain vulnerabilities that attackers can take advantage of[49]. Buffer overflows, SQL injections, and cross-site scripting attacks are some examples of these flaws. IoT devices are susceptible to malware and ransomware infections, which might jeopardize the ecosystem's overall security. DDoS assaults, botnet attacks, and other malicious activities can all be carried out via malware on a system. Physical tampering: IoT devices that lack physical security are susceptible to physical tampering, which gives hackers access to the system without authorization, gives them the chance to steal data, or modifies the device's behaviour [50]. Social engineering - Social engineering attacks, such as phishing assaults or pretexting, which deceive users into disclosing sensitive information or installing malware, can be used to target IoT devices. Fig. 4 depicts research trend in the domains of Blockchain, AI, IoT and Cyber security

TABLE V. RELATED SURVEYS ON IOT SECURITY

Reference	Year	Description
[53]	2014	Three layers of IoT security are examined, as well as the appropriate fix.
[54]	2016	The proposed design is built on IoT middleware, and each layer's specifics are detailed. The authors also discussed the IoT middleware system's adaptability and security challenges.
[55]	2016	Presented the reference model and the edge-side security threads. The countermeasure to the potential solutions was also discussed in the study.
[56]	2017	Outlined the merging of the IoT with Cyber-Physical Systems. Detailed examination of the privacy and security issues. The integration of edge/fog computing with IoT is considered.
[57]	2017	The study surveyed participants about their privacy and security concerns with IoT applications and devices. The authors looked at the authentication process for the IoT system. IoT applications built on a four-layer architecture provide difficult security challenges that are thoroughly explained.
[58]	2017	The authors of this study looked into modern security concerns in IoT applications. The risks and vulnerabilities of the system are thoroughly investigated in terms of communications, architecture, and applications. The paper's conclusion offers a strategy for resolving several security problems.
[59]	2018	The paper offers a thorough analysis of IoT security layer-by-layer. Suitable countermeasures and a model of probable dangers are thoroughly examined.
[60]	2018	Look at the danger and security model for IoT applications. The article discussed a few IoT system problems, including access control, trust management, and authentication.
[61]	2018	IoT systems' many standardized architectures have been reviewed, and the current solution strategy for security and interoperability is described.
[62]	2019	Examined the hazard and security in IoT applications. The use of Blockchain, edge computing, fog computing, and alternative solution approaches was suggested.
[63]	2020	There is a newly emerging technology that can address IoT security concerns. During a thorough investigation, the authors discovered that artificial intelligence, blockchain technology, and machine learning are the current approaches being used to address the IoT security problem.
[64]	2021	Security, trust, and faultless communication issues affecting the integration of IoT with blockchain are examined. Described in detail the study process used to examine the problems with the integration of blockchain, AI, and IoT.
[65]	2022	The IoT, blockchain and AI-based authentication in cybersecurity are all combined in this paper to give researchers a full, high-quality study on authentication and session keys.

TABLE VI. SOLUTIONS TO IOT SECURITY ATTACKS/CHALLENGES

Reference	Technique Used	Security Attack/Challenge	Measure Taken	Description
[66]	Edge computing using permission based blockchain for Smart Grid Network	The use of viability and the allocation of funds are under attack	transparent agreements updated on the blockchain	Solves the problems of intelligent systems, information security, and viability security by combining square chain and limit registration approaches.
[67]	Distributed security model using blockchain, edge cloud, and software-defined networking	Security attacks at edge layer of IoT network	SDN based gateway to hinder the doubtful flows.	By computing suspicious network traffic flows and preventing suspicious flows, the SDN-based gateway's dynamic network traffic flow management aids in the detection of security assaults and lowers their frequency.
[68]	Privacy protection of location data mining	Protection of location data records	Differential privacy mechanism	Using a multi-level query tree's structure and a different privacy method, you can query and publish location data on databases.
[69]	Privacy protection technique for location data	Privacy protection for location data	Location sensitivity for location recommendation	It uses check-in frequencies and location trajectories to determine a threshold for categorizing the sensitivity level of the places.
[70]	Privacy protection is integrated to machine learning	Classification process for local differential privacy protection	Logistic regression is applied for noise addition and feature selection	To achieve classification utilizing noise addition and feature selection, local differential privacy protection is created.

C. Related Work

Some study has already been done by researchers on the topic of how AI and IoT may work together to improve computation and decision-making in IoT systems. The work employs an master attack in the IoT to enable AI-based smart city applications [51]. In Zou et al. [52], similar explanations of fog and edge computing in IoT may be found. The Blockchain network employs a number of consensus techniques to reach consensus among the nodes. The design and potential uses of the Blockchain are thoroughly explained in the authors' essay [71]. In their study [72], the authors examined the concept, relevant research on IoT security and a possible approach to a solution using blockchain. The authors proposed a secure architecture for internet of things applications built on a distributed Blockchain system [73].

The authors' paper [74] briefly discusses how Blockchain technology is used in the Internet of Things. In paper [75], the authors assess the several Blockchain options for IoT security challenges as well as their implementation issues. To analyse and calculate the massive data collection using a machine learning technique, an effective framework [76] is needed. The authors of paper [77] examine the security concerns raised by using machine learning to a smart grid application. The authors

of article [78] discuss intrusion detection in Internet of Things applications. Table V summarizes the related survey works in the area of IoT security.

IoT issues were divided into four groups in Kshetri et al. research's [79]: (i) Cost and Capacity Limitations (ii) Deficient Architecture (iii) Downtime and Unavailability of Services on Cloud Servers (iv) Susceptibility to Manipulation. Their research covered the significance of blockchain in enhancing general security in supply chain networks as well as potential Blockchain solutions to each IoT concern. IoT datasets, which are used by the academic and expert communities, are one of the issues Banerjee et al. [80] examined when researching IoT security solutions. A standard for communicating IoT data values among research and expert communities, other pertinent partners, and vendors is necessary given the potentially conscious character of IoT datasets. In the future, blockchain technology should be made accessible to secure the security of IoT applications. Li et al. [81] offered a thorough analysis of the security risks to blockchain technology and talked about analogous actual attacks by extending well-liked Blockchain systems. They examined the blockchain technology security augmentation options. Table VI summarizes the techniques and countermeasures to address security attacks in IoT.

TABLE VII. INNOVATIONS THROUGH THE CONVERGENCE OF BLOCKCHAIN, AI AND IOT

Reference	Enablers	Innovation Domain	Use case	Description
[82]	Big data, 5G, Cloud Computing	Software Engineering	Blockchain-as-a- service (BaaS)	Cloud-based infrastructure and management provided by a third party for businesses developing and running blockchain applications. BaaS performs the back-end management for a platform or app built on blockchain in a manner similar to that of a web host.
[83]	Digital Twin	Data analytics	Intelligent money transaction - own profit centers (Autonomous agent)	IoT-based digital twins that can send and receive money via blockchain technology and act independently as economic agents can all do this.
[83]	Deep Reinforcement Learning	Block producers, consensus algorithm, block size, and block interval	Blockchain enabled IoT System	Framework for performance tuning to increase throughput.
[84]	Decentralization, AI algorithms	Digitally signed hash values, Data validation	Secure banking system	Reduces difficulties with accuracy, latency, privacy and security, and centralization.
[85]	Open source software	Bitcoin Protocol	Trust Management	It increases confidence in the system when a miner completes Proof of Work (PoW) without the aid of a bank or other centralised authority. Blockchain technology makes consumers think that no one can be trusted and that no one can make a claim to be one as a result.
[86]	Edge computing	Control and manage computation workload distribution	Data integrity and ensuring its availability and user accountability	Giving dispersed IoT hardware a reliable way to distribute computation-workload control and management across a large number of nodes.

D. Blockchain for AI in Addressing IoT Security

The AI can complete all tasks without outside influence with the aid of the Blockchain. Hence, all analysis and decision-making may be done on a private and secure platform. A decentralized AI platform also prevents data manipulation. Although AI and machine learning are two distinct methodologies, they are somewhat interconnected. The AI program employs a method called machine learning to respond to the tasks that are put in front of it. Unsupervised learning and supervised learning are the two types of learning that can be employed when creating and training an AI. Blockchain can be used in the context of AI to protect data produced by IoT devices, which is frequently utilized to train AI models. Blockchain can aid in preventing data tampering, manipulation, and unauthorized access by building a secure, decentralized record of this data. This can safeguard users' privacy while enhancing the accuracy and dependability of AI models. Securing device communication is another potential use for blockchain-based AI in IoT security. Without a centralized authority or middleman, devices can establish safe communication channels by using blockchain-based smart contracts to verify one other's identities. Attacks like man-in-the-middle attacks, which are frequent in IoT environments, could be avoided as a result of this.

E. AI for Blockchain for Addressing IoT Security

AI can help blockchain technology overcome some of its drawbacks, such as consensus procedures. Nodes can validate transactions rapidly and effectively in Proof of Work (PoW) or Proof of Stake (PoS) by using AI. Also, as the mining process uses a lot of energy, AI can help blockchains use less energy. This is possible by implementing AI technologies that have proven successful in reducing energy consumption. Using federated learning, for instance, which can offer a decentralized learning system, can help alleviate the scalability problems of blockchain. Also, while blockchain offers greater security than current technologies, AI can add an extra layer of security. IoT device data is analysed by AI algorithms to spot patterns and anomalies that can point to security lapses or other risks. AI can assist in preventing and reducing security concerns by continuously monitoring the data produced by

IoT devices and using machine learning algorithms to spot anomalous activity. Digital signatures are used by blockchain-based systems to verify transactions. These signatures can be examined by AI systems to find patterns that might point to fraud or other sorts of harmful behaviour. This could aid in limiting illegal access to IoT networks or devices. IoT devices produce a lot of data that can be used to forecast when a gadget is most likely to break down. This might lessen security threats that might be brought on by a corrupted or broken device. Strong authentication procedures are needed for blockchain-based systems to make sure that only authorized users can access devices and networks. To identify individuals and prevent illegal access, AI algorithms can be employed to examine biometric data such as facial recognition or voice recognition. In summary, the blockchain and AI work together to enhance the security of the IoT ecosystem. As outlined in Table VII, the blockchain and AI work together to pave the road for innovations in the IoT space.

IV. PROPOSED ARCHITECTURAL FRAMEWORK

Security, privacy, and scalability concerns can be addressed by designing a new framework for a Blockchain-based AI-enhanced IoT system. Fig. 5 depicts the proposed layered architectural framework for secured IoT. The layers of proposed architecture are as follows:

1) *Perception layer*:: IoT devices that gather and send data to the system are part of the perception layer. The system's core elements are IoT devices. They are in charge of gathering (sensing), storing, acting upon, and sending information to the system. They gather information from a variety of sensors, including temperature, humidity, and motion sensors.

2) *Network layer*:: This layer includes communication network that links the IoT devices to the system. To protect against any unauthorized access or data tampering, the network needs to be trustworthy and secure. VPNs and cryptographic protocols like SSL/TLS can be used to increase network security.

3) *Security layer*:: This layer contains the security defenses against cyber-attacks, including access control, encryption, and authentication. It guarantees the system's availability, integrity,

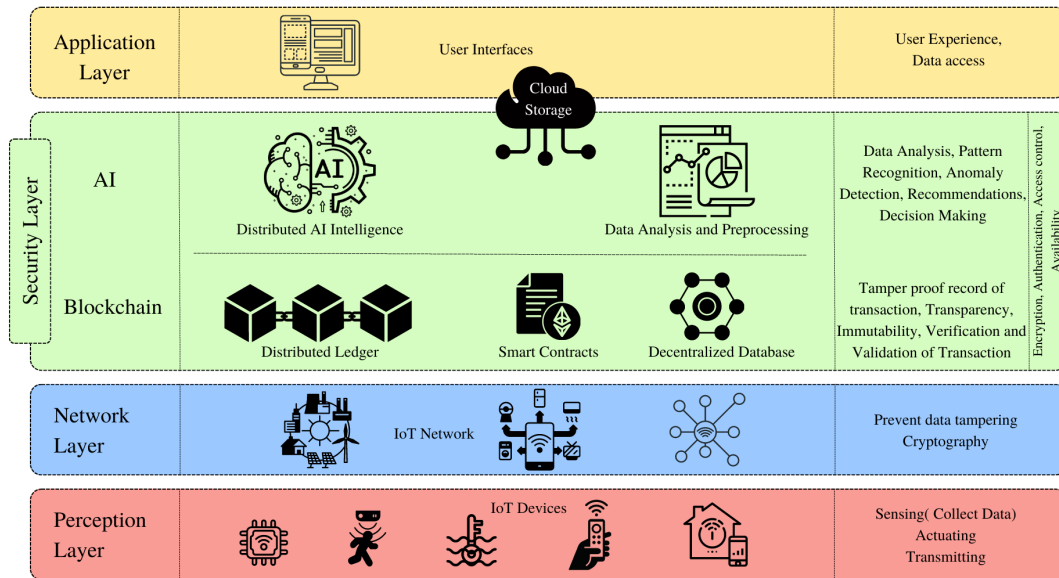


Fig. 5. Architectural Framework Integrating Blockchain and AI for Secured IoT.

and secrecy. The data gathered from the IoT devices is stored in the cloud and on the Blockchain. A distributed ledger in the Blockchain keeps a tamper-proof record of all transactions. It guarantees the system’s immutability, transparency, and security. Blockchain offers a secure, distributed database for storing all of the transactions. Large volumes of data are stored in the cloud storage, which also gives the system scalability. Smart contracts are self-executing agreements that uphold the system’s laws and regulations. They aid in the automation of the transactions’ validation and verification processes. The data gathered from IoT devices is stored and processed using cloud computing. It gives the system flexibility, scalability, and affordability. The distributed ledger of the blockchain offers a new way to store data in a more efficient manner. Blockchain and AI working together has implications for a variety of fields, including Security: AI and blockchain may combine to provide an additional line of defense against online dangers. Uncovering suspicious events is one of the most challenging issues facing upcoming businesses. Some, however, are already utilizing machine learning technologies that help in instantly spotting them. Speed – Combining the two technologies may speed up the delivery of information and data, improving the efficiency and speed of consumer interactions with enterprises. Service customization will increase over the coming years, and large companies’ or businesses’ recommendation systems will become widely used.

The artificial intelligence algorithms used to analyze the data gathered from IoT devices are included in this layer. AI assists in spotting trends, patterns, and abnormalities in the data. Additionally, it offers suggestions and forecasts based on the data analysis. AI offers insightful information that aids in decision-making. The data gathered from IoT devices is processed using data analytics. It assists with spotting trends, patterns, and abnormalities in the data.

4) *Application layer*:: User interface, data analytics, and other programs that enable the system’s functionality are all included in the application layer. The system’s data and analytics are accessed through the user interface, which is also used to interact with the system. IoT device data is processed using data analytics, which offers insightful information.

The proposed framework offers a safe, scalable, and effective solution to manage IoT devices, gather data, and use AI and data analytics to make wise decisions. While the use of AI and data analytics provide useful insights and aids in making decisions, the usage of Blockchain protects the security, transparency, and immutability of the system. An additional layer of defense against cyber-attacks is offered by the security layer.

V. CONCLUSION

Blockchain and AI coming together is a promising way to secure the IoT ecosystem. IoT devices are prone to cyber-attacks, and a decentralized architecture supported by blockchain and AI can offer a practical solution to improve the security of these devices. The purpose of this suggested framework is to investigate the potential benefits of this convergence for secure IoT. The framework can improve the data acquired by IoT devices in terms of security, transparency, and privacy. Additionally, it can lessen the need on centralized middlemen and offer a decentralized platform for data transfer and communication amongst IoT devices. The IoT ecosystem’s security and the development of the digital future may both benefit from the integration of blockchain and AI. Several theories and tests have been conducted in an effort to connect AI, IoT, and Blockchain; nevertheless, more research will be needed to build a digital strategy that might effectively combine the three for a reliable and useful digital component. Unlike any previous period in history, cloud computing is incredibly important today and enables online connections. The

massive amount of data handled by this computing system emphasizes the necessity for automated systems with (QoS) standards. Identification of essential technologies is necessary to meet this demand, and it currently looks that Blockchain, AI, and IoT are these convergent technologies.

ACKNOWLEDGMENT

Authors acknowledge the support from REVA University for the facilities provided to carry out the research.

REFERENCES

- [1] Schwab, Klaus. The fourth industrial revolution. Currency, 2017.
- [2] Akhtar, Muhammad Waseem, et al. "The shift to 6G communications: vision and requirements." *Human-centric Computing and Information Sciences* 10 (2020): 1-27.
- [3] Nguyen, Quoc Khanh, and Quang Vang Dang. "Blockchain Technology for the Advancement of the Future." 2018 4th international conference on green technology and sustainable development (GTSD). IEEE, 2018.
- [4] Li, Zhi, Ali Vatankeh Barenji, and George Q. Huang. "Toward a blockchain cloud manufacturing system as a peer to peer distributed network platform." *Robotics and computer-integrated manufacturing* 54 (2018): 133-144.
- [5] Zhu, Qingyi, et al. "Applications of distributed ledger technologies to the internet of things: A survey." *ACM computing surveys (CSUR)* 52.6 (2019): 1-34.
- [6] Gulati, Prerna, et al. "Approaches of blockchain with ai: Challenges & future direction." *Proceedings of the International Conference on Innovative Computing & Communications (ICICC)*. 2020.
- [7] Statista, I. H. S. "Internet of Things (IoT) connected devices installed base worldwide from 2015 to 2025 (in billions)." URL: <https://www.statista.com/statistics/471264/iot-number-of-connecteddevicesworldwide/>(Consulté 17/05/2020) (2018).
- [8] Hugoson, Mats-Åke. "Centralized versus decentralized information systems: A historical flashback." *History of Nordic Computing 2: Second IFIP WG 9.7 Conference, HiNC2, Turku, Finland, August 21-23, 2007, Revised Selected Papers 2*. Springer Berlin Heidelberg, 2009.
- [9] Atlam, Hany F., and Gary B. Wills. "IoT security, privacy, safety and ethics." *Digital twin technologies and smart cities* (2020): 123-149.
- [10] Atlam, Hany F., and Gary B. Wills. "Technical aspects of blockchain and IoT." *Advances in computers*. Vol. 115. Elsevier, 2019. 1-39.
- [11] Uddin, Md Ashraf, et al. "A survey on the adoption of blockchain in iot: Challenges and solutions." *Blockchain: Research and Applications* 2.2 (2021): 100006.
- [12] Da Xu, Li, and Wattana Viriyasitavat. "Application of blockchain in collaborative internet-of-things services." *IEEE Transactions on Computational Social Systems* 6.6 (2019): 1295-1305.
- [13] Xiong, Zuobin, et al. "Privacy threat and defense for federated learning with non-iid data in AIoT." *IEEE Transactions on Industrial Informatics* 18.2 (2021): 1310-1321.
- [14] Chen, Yinong. "IoT, cloud, big data and AI in interdisciplinary domains." *Simulation Modelling Practice and Theory* 102 (2020): 102070.
- [15] Nawaz, Anum, et al. "Edge AI and blockchain for privacy-critical and data-sensitive applications." 2019 Twelfth International Conference on Mobile Computing and Ubiquitous Network (ICMU). IEEE, 2019.
- [16] Bertino, Elisa, Ahish Kundu, and Zehra Sura. "Data transparency with blockchain and AI ethics." *Journal of Data and Information Quality (JDIQ)* 11.4 (2019): 1-8.
- [17] Zhang, Guozhen, et al. "Blockchain-based data sharing system for ai-powered network operations." *Journal of Communications and Information Networks* 3 (2018): 1-8.
- [18] Parker, Brian, and Christian Bach. "The synthesis of blockchain, artificial intelligence and internet of things." *European Journal of Engineering and Technology Research* 5.5 (2020): 588-593.
- [19] Lai, Ying-Hsun, et al. "Cognitive optimal-setting control of AIoT industrial applications with deep reinforcement learning." *IEEE Transactions on Industrial Informatics* 17.3 (2020): 2116-2123.
- [20] Nebula Ai (NBAI) Decentralized ai blockchain whitepaper, Nebula AI Team, Montreal, QC, Canada; 2018.
- [21] Dinh TN, Thai MT. Ai and blockchain: a disruptive integration. *Computer*. 2018;51(9):48-53.
- [22] McMahan, Brendan, et al. "Communication-efficient learning of deep networks from decentralized data." *Artificial intelligence and statistics*. PMLR, 2017.
- [23] Falco, Gregory, Carlos Caldera, and Howard Shrobe. "IIoT cybersecurity risk modeling for SCADA systems." *IEEE Internet of Things Journal* 5.6 (2018): 4486-4495.
- [24] Yu, Tianlong, et al. "Handling a trillion (unfixable) flaws on a billion devices: Rethinking network security for the internet-of-things." *Proceedings of the 14th ACM Workshop on Hot Topics in Networks*. 2015.
- [25] Nawari, Nawari O., and Shriram Ravindran. "Blockchain technology and BIM process: review and potential applications." *J. Inf. Technol. Constr.* 24.12 (2019): 209-238.
- [26] Vocke, Christian, Carmen Constantinescu, and Daniela Popescu. "Application potentials of artificial intelligence for the design of innovation processes." *Procedia CIRP* 84 (2019): 810-813.
- [27] Jesus, Emanuel Ferreira, et al. "A survey of how to use blockchain to secure internet of things and the stalker attack." *Security and communication networks* 2018 (2018).
- [28] Figueroa, Santiago, Javier Añorga, and Saioa Arrizabalaga. "An attribute-based access control model in RFID systems based on blockchain decentralized applications for healthcare environments." *Computers* 8.3 (2019): 57.
- [29] Dinh, Thang N., and My T. Thai. "AI and blockchain: A disruptive integration." *Computer* 51.9 (2018): 48-53.
- [30] Lv, Chen, et al. "Hybrid-learning-based classification and quantitative inference of driver braking intensity of an electrified vehicle." *IEEE Transactions on vehicular technology* 67.7 (2018): 5718-5729.
- [31] Panarello, Alfonso, et al. "Blockchain and iot integration: A systematic survey." *Sensors* 18.8 (2018): 2575.
- [32] Miraz, Mahdi H., and Maaruf Ali. "Blockchain enabled enhanced IoT ecosystem security." *Emerging Technologies in Computing: First International Conference, iCETiC 2018, London, UK, August 23-24, 2018, Proceedings 1*. Springer International Publishing, 2018.
- [33] Sharma, Pradip Kumar, Mu-Yen Chen, and Jong Hyuk Park. "A software defined fog node based distributed blockchain cloud architecture for IoT." *Ieee Access* 6 (2017): 115-124.
- [34] Zhang, Yu, and Jiangtao Wen. "The IoT electric business model: Using blockchain technology for the internet of things." *Peer-to-Peer Networking and Applications* 10 (2017): 983-994.
- [35] Novo, Oscar. "Blockchain meets IoT: An architecture for scalable access management in IoT." *IEEE internet of things journal* 5.2 (2018): 1184-1195.
- [36] Ozyilmaz, Kazim Rifat, and Arda Yurdakul. "Designing a Blockchain-based IoT with Ethereum, swarm, and LoRa: the software solution to create high availability with minimal security risks." *IEEE Consumer Electronics Magazine* 8.2 (2019): 28-34.
- [37] Wu, Chung Kit, et al. "The IDex case study on the safety measures of AIoT-based railway infrastructures." 2020 IEEE International Symposium on Product Compliance Engineering-Asia (ISPCE-CN). IEEE, 2020.
- [38] Xiong, Zuobin, et al. "Privacy threat and defense for federated learning with non-iid data in AIoT." *IEEE Transactions on Industrial Informatics* 18.2 (2021): 1310-1321.
- [39] Lai, Ying-Hsun, et al. "Cognitive optimal-setting control of AIoT industrial applications with deep reinforcement learning." *IEEE Transactions on Industrial Informatics* 17.3 (2020): 2116-2123.
- [40] Rathore, Shailendra, Pradip Kumar Sharma, and Jong Hyuk Park. "XSSClassifier: an efficient XSS attack detection approach based on machine learning classifier on SNSs." *Journal of Information Processing Systems* 13.4 (2017): 1014-1028.
- [41] Jeong, Young-Sik, and Jong Hyuk Park. "IoT and smart city technology: challenges, opportunities, and solutions." *Journal of Information Processing Systems* 15.2 (2019): 233-238.
- [42] Zheng, Zhibin, et al. "Blockchain challenges and opportunities: A survey." *International journal of web and grid services* 14.4 (2018): 352-375.

- [43] Wood, Gavin. "Ethereum: A secure decentralised generalised transaction ledger." Ethereum project yellow paper 151.2014 (2014): 1-32.
- [44] Li, Daming, et al. "Blockchain as a service models in the Internet of Things management: Systematic review." *Transactions on Emerging Telecommunications Technologies* 33.4 (2022): e4139.
- [45] McMahan, Brendan, et al. "Communication-efficient learning of deep networks from decentralized data." *Artificial intelligence and statistics*. PMLR, 2017.
- [46] Tanwar, Sudeep, et al. "Machine learning adoption in blockchain-based smart applications: The challenges, and a way forward." *IEEE Access* 8 (2019): 474-488.
- [47] Tyagi, Amit Kumar, Gillala Rekha, and N. Sreenath. "Beyond the hype: Internet of things concepts, security and privacy concerns." *Advances in Decision Sciences, Image Processing, Security and Computer Vision: International Conference on Emerging Trends in Engineering (ICETE)*, Vol. 1. Springer International Publishing, 2020.
- [48] Hossain, Md Mahmud, Maziar Fotouhi, and Ragib Hasan. "Towards an analysis of security issues, challenges, and open problems in the internet of things." 2015 IEEE world congress on services. IEEE, 2015.
- [49] Abdullah, T. A., et al. "A review of cyber security challenges attacks and solutions for Internet of Things based smart home." *Int. J. Comput. Sci. Netw. Secur* 19.9 (2019): 139.
- [50] Varga, Pal, et al. "Security threats and issues in automation IoT." 2017 IEEE 13th International Workshop on Factory Communication Systems (WFCS). IEEE, 2017.
- [51] Falco, Gregory, et al. "A master attack methodology for an AI-based automated attack planner for smart cities." *IEEE Access* 6 (2018): 48360-48373.
- [52] Zou, Zhuo, et al. "Edge and fog computing enabled AI for IoT-an overview." 2019 IEEE International Conference on Artificial Intelligence Circuits and Systems (AICAS). IEEE, 2019.
- [53] Jing, Qi, et al. "Security of the Internet of Things: perspectives and challenges." *Wireless Networks* 20 (2014): 2481-2501.
- [54] Ngu, Anne H., et al. "IoT middleware: A survey on issues and enabling technologies." *IEEE Internet of Things Journal* 4.1 (2016): 1-20.
- [55] Mosenia, Arsalan, and Niraj K. Jha. "A comprehensive study of security of internet-of-things." *IEEE Transactions on emerging topics in computing* 5.4 (2016): 586-602.
- [56] Lin, Jie, et al. "A survey on internet of things: Architecture, enabling technologies, security and privacy, and applications." *IEEE internet of things journal* 4.5 (2017): 1125-1142.
- [57] Yang, Yuchen, et al. "A survey on security and privacy issues in Internet-of-Things." *IEEE Internet of things Journal* 4.5 (2017): 1250-1258.
- [58] Alaba, Fadele Ayotunde, et al. "Internet of Things security: A survey." *Journal of Network and Computer Applications* 88 (2017): 10-28.
- [59] Grammatikis, Panagiotis I. Radoglou, Panagiotis G. Sarigiannidis, and Ioannis D. Moscholios. "Securing the Internet of Things: Challenges, threats and solutions." *Internet of Things* 5 (2019): 41-70.
- [60] Das, Ashok Kumar, Sherali Zeadally, and Debiao He. "Taxonomy and analysis of security protocols for Internet of Things." *Future Generation Computer Systems* 89 (2018): 110-125.
- [61] Di Martino, Beniamino, et al. "Internet of things reference architectures, security and interoperability: A survey." *Internet of Things* 1 (2018): 99-112.
- [62] Hassija, Vikas, et al. "A survey on IoT security: application areas, security threats, and solution architectures." *IEEE Access* 7 (2019): 82721-82743.
- [63] Mohanta, Bhabendu Kumar, et al. "Survey on IoT security: Challenges and solution using machine learning, artificial intelligence and blockchain technology." *Internet of Things* 11 (2020): 100227.
- [64] Guergov, Sasho, and Neyara Radwan. "Blockchain Convergence: Analysis of Issues Affecting IoT, AI and Blockchain." *International Journal of Computations, Information and Manufacturing (IJCIM)* 1.1 (2021).
- [65] Attkan, Ankit, and Virender Ranga. "Cyber-physical security for IoT networks: a comprehensive review on traditional, blockchain and artificial intelligence based key-security." *Complex & Intelligent Systems* 8.4 (2022): 3559-3591.
- [66] Gai, Keke, et al. "Permissioned blockchain and edge computing empowered privacy-preserving smart grid networks." *IEEE Internet of Things Journal* 6.5 (2019): 7992-8004.
- [67] Medhane, Darshan Vishwasrao, et al. "Blockchain-enabled distributed security framework for next-generation IoT: An edge cloud and software-defined network-integrated approach." *IEEE Internet of Things Journal* 7.7 (2020): 6143-6149.
- [68] Gu, Ke, Lihao Yang, and Bo Yin. "Location data record privacy protection based on differential privacy mechanism." *Information Technology and Control* 47.4 (2018): 639-654.
- [69] Yin, Chunyong, et al. "Location recommendation privacy protection method based on location sensitivity division." *EURASIP Journal on Wireless Communications and Networking* 2019.1 (2019): 1-13.
- [70] Yin, Chunyong, et al. "Local privacy protection classification based on human-centric computing." *Human-centric Computing and Information Sciences* 9 (2019): 1-14.
- [71] Mohanta, Bhabendu Kumar, et al. "Blockchain technology: A survey on applications and security privacy challenges." *Internet of Things* 8 (2019): 100107.
- [72] Banerjee, Mandrita, Junghee Lee, and Kim-Kwang Raymond Choo. "A blockchain future for internet of things security: a position paper." *Digital Communications and Networks* 4.3 (2018): 149-160.
- [73] Satapathy, Utkalika, et al. "A secure framework for communication in internet of things application using hyperledger based blockchain." 2019 10th international conference on computing, communication and networking technologies (ICCCNT). IEEE, 2019.
- [74] Fernández-Caramés, Tiago M., and Paula Fraga-Lamas. "A Review on the Use of Blockchain for the Internet of Things." *Ieee Access* 6 (2018): 32979-33001.
- [75] Khan, Minhaj Ahmad, and Khaled Salah. "IoT security: Review, blockchain solutions, and open challenges." *Future generation computer systems* 82 (2018): 395-411.
- [76] I. Kotenko , I. Saenko , A. Branitskiy , Framework for mobile internet of things security monitoring based on big data processing and machine learning, *IEEE Access* 6 (2018) 72714–72723 .
- [77] Hossain, Eklas, et al. "Application of big data and machine learning in smart grid, and associated security concerns: A review." *Ieee Access* 7 (2019): 13960-13988.
- [78] N. Chaabouni , M. Mosbah , A. Zemmari , C. Sauvignac , P. Faruki , Network intrusion detection for IoT security based on learning techniques, *IEEE Commun. Surv. Tutor.* 21 (3) (2019) 2671–2701 .
- [79] Rathore, Shailendra, Pradip Kumar Sharma, and Jong Hyuk Park. "XSSClassifier: an efficient XSS attack detection approach based on machine learning classifier on SNSs." *Journal of Information Processing Systems* 13.4 (2017): 1014-1028.
- [80] Banerjee, Mandrita, Junghee Lee, and Kim-Kwang Raymond Choo. "A blockchain future for internet of things security: a position paper." *Digital Communications and Networks* 4.3 (2018): 149-160.
- [81] Banerjee, M.; Lee, J.; Choo, K.-K.R. A blockchain future for internet of things security: A position paper. *Digit. Commun. Netw.* 2018, 4, 149–160.
- [82] Parker, Brian, and Christian Bach. "The synthesis of blockchain, artificial intelligence and internet of things." *European Journal of Engineering and Technology Research* 5.5 (2020): 588-593.
- [83] Sandner, Philipp, Jonas Gross, and Robert Richter. "Convergence of blockchain, IoT, and AI." *Frontiers in Blockchain* 3 (2020): 522600.
- [84] Singh, Sushil Kumar, Shailendra Rathore, and Jong Hyuk Park. "Block-iotintelligence: A blockchain-enabled intelligent IoT architecture with artificial intelligence." *Future Generation Computer Systems* 110 (2020): 721-743.
- [85] Uddin, Md Ashraf, et al. "A survey on the adoption of blockchain in iot: Challenges and solutions." *Blockchain: Research and Applications* 2.2 (2021): 100006.
- [86] Alrubei, Subhi M., Edward Ball, and Jonathan M. Rigelsford. "A secure blockchain platform for supporting AI-enabled IoT applications at the Edge layer." *IEEE Access* 10 (2022): 18583-18595.

Opportunities and Challenges in Human-Swarm Interaction: Systematic Review and Research Implications

Alexandru-Ionuț Șiean, Bogdănel-Constantin Grădinaru,
Ovidiu-Ionuț Gherman, Mirela Danubianu, Laurențiu-Dan Milici
Faculty of Electrical Engineering and Computer Science,
“Stefan cel Mare” University of Suceava, 720229 Suceava, Romania

Abstract—We conducted a Systematic Literature Review on scientific papers that examined the interaction between operators and drone swarms based on the use of a command and control center. We present the results of a meta-analysis of nine scientific papers published in the ACM DL and IEEE Xplore databases. Our findings show that research on human-drone swarm interaction shows a disproportionate interest in hand gestures compared to other input modalities for drone swarm control. Furthermore, all articles reviewed exclusively explored gestures and the size of the swarm used in the studies was limited, with a median of 3.0 and an average of 3.8 drones per study. We compiled an inventory of interaction modalities, recognition techniques, and application types from the scientific literature, which is presented in this paper. On the basis of our findings, we propose four areas for future research that can guide scientific investigations and practical developments in this field.

Keywords—Human swarm interactions; input modalities; swarm control

I. INTRODUCTION

Interacting with drone swarms is becoming fascinating for many people, but is still futuristic for many fields of scientific study [1]. However, the human-drone swarm interaction is made possible by the large amount of work done by researchers and practitioners in this field. The scientific community has investigated and proposed new technologies [2], artifacts [3], and interaction techniques [4] to make human-drone interaction possible and allow operators to accept drone interaction and rides in public spaces [5]. This work has revealed numerous benefits to society, such as finding survivors in emergency situations [6], [7], [8], [9] or escorting lone drones moving at night [10]. Other scientific papers have focused on the guidance of drones by people crossing roads [11] or providing personal assistance [12]. These proposals with drones used on a large scale have also come with many applications and fields of use. Despite the potential drone swarms they create, a quantitative analysis of how research has developed is lacking. We find drones in a multitude of fields, from professional environments [13], leisure [14], or semi- or fully autonomous modules [15]. However, we identify drones that are used singularly [16] or are part of a swarm [17]. This exemplification highlights the diversification of domains with drones, and we, therefore, propose a high-level perspective of current and future use cases for drone swarm interaction. See Fig. 1.

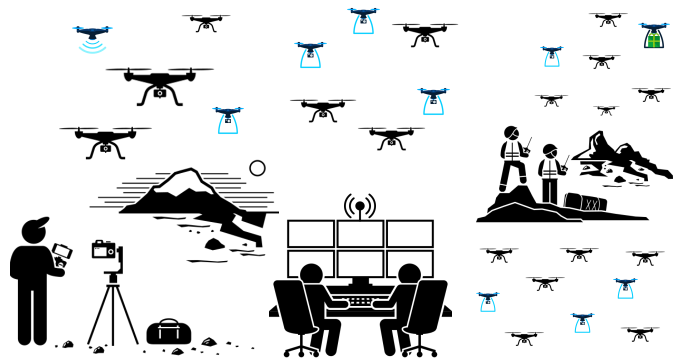


Fig. 1. Description of concept interaction between operators and drone swarms, based on the use of a command and control center.

First, this article describes the selection process, which represents a domain analysis of $N=59$ scientific articles on the interaction between operators and drone swarms based on the use of a command and control center. Second, our results have led to several future research directions. Finally, we conclude with fundamental observations and research opportunities for the field.

To the best of our knowledge, this is the first paper to analyze a scientific paper in this form; therefore, we propose the following.

- 1) We report results from the Systematic Literature Review conducted on the interaction between operators and drone swarms based on the use of a command and control center.
- 2) We draw implications for future research on the interaction between operators and drone swarms based on the use of a command and control center.
- 3) We propose four research directions for the scientific community to structure future scientific (A. *Input modalities and interaction technique*, B. *User studies and evaluations*, C. *Involvement of Multiple Drones in Simulations*, D. *Control and Command Devices and Prototypes*) investigations and practical developments in this area.

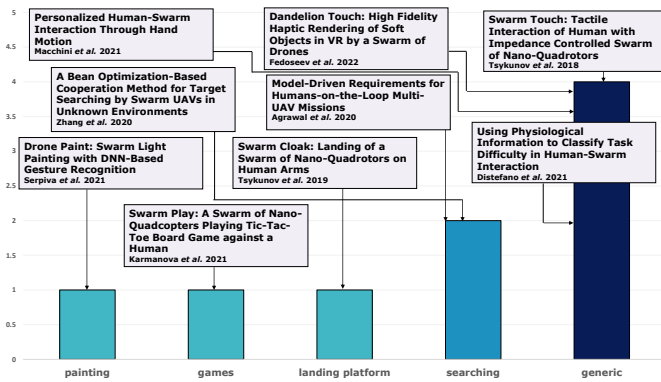


Fig. 2. Type of application interaction between operators and drone swarms, based on the use of a command and control center.

II. STUDY DESIGN AND SCOPING REVIEW

Previously published research in the field of human-drone swarm interaction has, for example, presented various activities in which drones are used for monitoring and assessing earthquake damage [18], investigating maritime spill [19], and delivering defibrillators [20]. All these applications involve human operators who will interact with drones through an interface, directly or indirectly, to capture images or plan drone routes. Previous research has described other forms of interaction [21], including voice and haptic interfaces [22], [23], but these have not been identified for interactions with drone swarms for emergency response.

Therefore, we present the method underlying the formation of our study during the systematic literature review of the literature to address the two research questions in the II-B section that are focused on our area of investigation for the interaction between operators and drone swarms, based on the use of a command and control center.

The substantial growth in drones in recent years¹ does not lead to a clear direction in which application domains these drone swarms are needed and what interactive modalities are required with them.

That is why we aim to map the systems and applications that have been described, studied, or envisioned in the scientific literature and that are based on the interaction between operators and drone swarms based on the use of a command and control center. There are a few scientific studies that have analyzed in detail, challenges, application domains, and detailed descriptions of previous works, all of which have focused on human-drone interaction [24], [25].

These surveys offer useful information to the community of researchers and practitioners. However, they do not provide a systematic exploration of previous work that looked at the interaction between operators and drone swarms based on the use of a command and control center in terms of the systems and applications that have been used. Here, we take a broader angle and propose a systematic literature review of the scientific literature that addresses the interaction between operators and drone swarms, based on the use of a command and control center, classifying the purposes of the applications,

the domains, how data were collected, and also the systems that were used to interact with drone swarms.

A. Chosen Methodology and Steps

We conducted a systematic literature review, using the methodology proposed by Siddaway *et al.* [26] and Liberati *et al.* [27], to perform a critical analysis of the current state-of-the-art in terms of the interaction between operators and drone swarms based on the use of a command and control center. Implementing this approach was based on the identification of references of interest for our study, and therefore we identified, using queries with a specific footnote query in the ACM Digital Library² and IEEE Xplore³ databases, where we identified and analyzed a body of N=59 scientific articles.

Our motivation for conducting this survey of systems and applications that address the interaction between operators and drone swarms is to provide an overview of the systems and applications, how they interact with them, and current and future use cases studied and envisioned in the scientific literature. In the following, we explain the survey method used, our research questions formulated, and how we selected and categorized the relevant articles for our study.

We have carried out data mapping that targets applications in the interaction between operators and drone swarms, based on the use of a command and control center according to the following steps:

B. First Stage: Identifying the Research Questions for Our Study

To explore the scope of applications targeting human-drone swarm interaction based on the use of a command and control center, we formed the following questions.

- RQ1:** “What are the categories of applications studied or considered in human-drone swarm interaction?”
- RQ2:** “What kind of interaction modalities has been described, studied, or designed for human-drone swarm interaction based on the use of a command and control center?”

C. Second Stage: Identifying Relevant Articles

We based our selection on academically evaluated research articles. We therefore selected the ACM Digital Library, which contains more than 3M records, and IEEE Xplore, these being the main digital libraries in Engineering and Computer Science. To identify works that could be integrated into our study we used to query:

```
"query": {  
  Abstract:  
  ((interaction* AND swarm*) AND  
  (drone* OR uav*) AND  
  (tool* OR platform* OR  
  application* OR system*))  
}  
"filter": {NOT VirtualContent: true}
```

¹<https://www.nytimes.com/2020/05/23/style/drones-coronavirus.html>

²<https://dl.acm.org/>

³<https://ieeexplore.ieee.org/Xplore/home.jsp>

Using asterisks (*) to avoid focusing the search on a single word. Note that we focused our search on drones, and UAVs are the most popular type of consumer drone, according to the analysis by Wojciechowska *et al.* [28]. The formed query was used for both databases.

D. Third Stage: Selecting Relevant Articles

Our analysis included reviewed articles and large-scale academic articles, such as extended abstracts (e.g., postings and demonstrations), but also articles from workshops. In our selection process, we did not include sites websites or other media, as we intend to focus on classifying the purposes of the applications, the domains, how data was collected but also the systems that were used to interact with drone swarms but also on image object recognition technologies encountered in the interaction between operators and drone swarms, based on the use of a command and control center. All the selections that we refer to will be called articles throughout this scientific article. For all articles identified using queries from ACM and IEEE Xplore, we used the following exclusion criteria.

- EC1:** *Mandatory article was written in English, abstract and the content article.* From our analysis, we have identified two articles that do not meet the criteria [29], [30]
- EC2:** *Considers interactions between humans and drone swarms.* Our analysis excluded 42.37% of the articles from the total aggregate gathered from the two queries. Among these, we identify articles such as [31], [32], [33].
- EC3:** *Does not show one or more types of interaction.* We have excluded all articles that show no interaction with drone swarms. Our analysis reduced cumulative work by 32.20%, excluding articles such as [34], [35], [36].

After the identification process, N=59, duplicate elimination, D=3, followed. As a result, the selection and analysis process of the articles remained, N=56, which were selected based on title and abstract. Subsequently, 56 articles were evaluated for eligibility criteria based on the full text. This process eliminated 47 articles that did not follow our research direction.

Therefore, this process resulted in nine articles, which were considered relevant to the formulation of the proposed aim, and based on these we completed the aims for interaction between operators and drone swarms, based on the use of a command and control center.

E. Fourth Stage: Charting the Data for Relevant Articles

Charting is a technique to synthesize and interpret quantitative data and involves sorting the material according to the issues and themes proposed by Hilary *et al.* [37]. The process started by establishing the scopes, methods of shape recognition, and other metrics to be presented in the scientific articles. Data extraction was carried out for each item that constitutes the body of scientific articles. Each time a new domain was identified, we recorded it in the table, so that later we could quantitatively report the results that we would present. Where applications did not specify a particular domain, we noted it

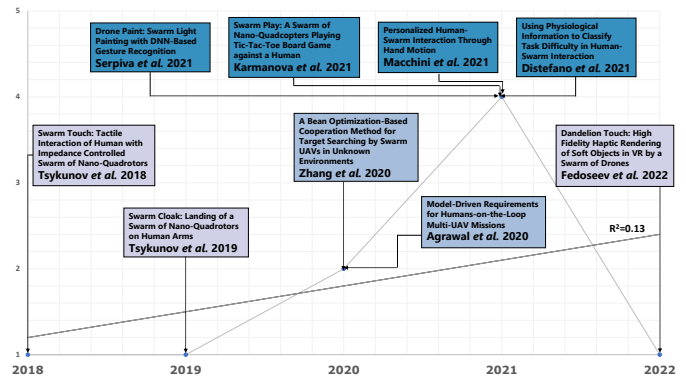


Fig. 3. Distribution by year of articles on the interaction between operators and drone swarms, based on the use of a command and control center.

as generic. In addition to application areas, algorithms, and other measurements, we also performed an analysis of the distribution of articles by year; see Fig. 3.

III. FINDINGS RESULTS

The analysis was conducted using nine scientific articles [4], [38], [39], [40], [41], [42], [43], [44], [45]. This section is dedicated to selected articles on the application areas of interaction between operators and drone swarms, based on the use of a command and control center.

In terms of ① *application types*, painting applications account for 11% of our body of work, in which we have the “Drone Paint” application by Serpiva *et al.* [4], in which he proposes a system capable of interacting with a swarm that controls the swarm’s trajectory through gestures. Another category of applications is dedicated to games, or entertainment, that occupies 11% of the total body of work, where “Swarm Play” has Karmanova *et al.* [38], which proposes the game Tic-Tac-Toe, using a swarm of drones, winning against a human user. Generic applications, which occupy the largest place in our body of work, are applications that aim to control robotic systems, such as drone swarms through navigation algorithms [40]. See Fig. 2.

The reviewed articles were published between 2018 and 2022, where the maximum ② *number of publications* was in 2021. A total of 33.33% articles were identified in the ACM Digital Library database and 66.66% articles were identified in IEEE Xplore. See Fig. 3.

The ③ *validation systems* were carried out by *user study* in the proportion of 22.22%, and we still have 66.67% of systems that were validated by *technical performance*, and 11.11% *demonstration*.

From the point of view of human-drone swarm interaction, we identify 4 scientific articles describing the system by which the swarm interacted with the drone. The drones used in the studies were simulated using the “ros framework” [4], [39], [43] and the “mocap framework” [38] and the drone type was “crazyflie 2.0.” This analysis aimed to present information that was selected using a research question *RQ1*.

A. Interaction Modalities and Recognition Technique

To address the second research question, *RQ2*, in II-B on technologies for object recognition in images we present the following information. In our case we will present the interaction modalities that have been identified to send different commands to drone swarms. From our body of work, which covers human-drone swarm interaction, we have identified only gestures, and several Machine Learning algorithms that we will present. This information was extracted from the four steps described.


Therefore, the largest set of gestures for human-drone swarm interaction was 8. The gestures are executed in front of a camera “Logitech HD Pro Webcam C920 of @30FPS”, which are recognized using the algorithm, Machine Learning, DNN (Deep Neuronal Network). The system aims to control a swarm of drones to paint different surfaces of buildings, and the commands it was “one,” “two,” “three,” “four,” “five,” “okay,” “rock,” and “thumbs up” were used to achieve this goal. The drones being simulated and their feedback were observed on the computer [4].

Dandelion Touch proposes a haptic display that is being used for a new type of haptic feedback for VR systems that does not require any wearable or portable interface. To achieve this approach, Fedoseev *et al.* [43], proposes a set of 4 gestures, “moving forward”, “moving backward”, “moving right”, “moving left” and uses an “Oculus Quest Vive Pro” headset.

Swarm Touch, proposes a system that targets human-drone swarm interaction, and receives vibrotactile feedback. The gestures used for swarm control are grouped into two categories, which are “extended state (increasing distance, constant distance, decreasing distance), contracted state (increasing distance, constant distance, decreasing distance)” [44].

In terms of algorithms used in the literature, we find SVM (System Vector Machine) and CNN (Convolutional Neural Network) [45], Robot Bean Optimization Algorithm [41], Body Machine Interface (BoMI) [40], Basic Algorithm [38], DNN (Deep Neural Network) [45]. The articles [39], [41], [43], [44], do not present the recognition algorithm.

B. Data Collection in Drone Swarm Applications

Before reporting on how the data was collected, we first focused on how the researchers validated the proposed systems. Thus, we submitted for analysis the  type of study of each scientific paper we analyzed.

Out of the amount of scientific articles we extracted, 88.9% of validated the system through an experimental study. For example, Macchini *et al.* [40], proposes an experimental study for the use of a Body Machine Interface (BoMI) to control a drone swarm. The main observation made in this study is given by the fact that users usually use their hands to control the drones, and therefore used a Leap Motion controller to track the movements of the hand movements. This implemented a Machine Learning algorithm to customize the BoMi interface, based on a rigorous calibration process. The study demonstrated that users received positive feedback on use compared to a remote control.

Another approach, by which the system is validated by experiment, we find in the work of Serpiva *et al.* [4], proposes a drone swarm control system by drawing the trajectory with hand gestures, and gesture recognition is based on the Deep Neural Network (DNN) algorithm. The experiment shows that the accuracy of gesture recognition is 99.7%, which allows the user to have a high accuracy in drawing the drone trajectory compared to the mouse drawing. The system gives the user the possibility to create their own art objects using drone swarms.

The only scientific work that uses a questionnaire is that of Agrawal *et al.* [42] in which they use a drone swarm for emergency scenarios such as search and rescue or fire surveillance. Using artificial intelligence, drone swarms operate autonomously, but nevertheless human intelligence and domain expertise are very important for mission deployment. The proposed system is a meta-model to describe the interactions between human operators and the autonomous swarm. The answers to the questions show that the modeling of interactions through artificial intelligence is supported in drone missions.

The proposed systems are also aimed at interacting with drone swarms, and at the same time being able to receive various information back from these drones. Therefore, our investigations in the literature show that data transmission is performed on PC systems [4], [38], through visual feedback [43], or even tactile feedback [39].

For data manipulation, which also aims at gestural interaction with drone swarms, we find algorithms such as Machine Learning, and Deep Neural Network [4] used to recognize hand gestures. Our analysis also identifies algorithms such as Basic [38], used for the Tic-Tac Toe game. Other algorithms identified in the literature are Body Machine Interface (BoMI) [40], which is used to interact with drone swarms using hand gestures captured by Leap Motion, Robot Bean Optimization algorithm [41], used to search for people using drone swarms in unknown environments, and last but not least, Support Vector Machine and Convolutional Neural Network [45], used to classify tasks that were analyzed using an EEG (electroencephalogram) headset.

IV. RESEARCH IMPLICATIONS

Our investigations show that research on the interaction between operators and drone swarms, based on the use of a command and control center has been limited, focusing on only a few types of applications, adopting entirely hand gestures, meanwhile, half of the works investigated in our SLR did not involve participants in the experimental studies presented.

Based on these findings, we formulate several research directions to encourage more work on the interaction between operators and drone swarms, based on the use of a command and control center about our research questions *RQ1* and *RQ2* from Section II-B.

We structure these research directions into the following:

- A. *Input Modalities and Interaction Techniques,*
- B. *User Studies and Evaluations,*
- C. *Involvement of Multiple Drones in Simulations,*
- D. *Control and Command Devices and Prototypes.*

For each category, we identify several opportunities for future work.

A. Input Modalities and Interaction Techniques

Our analysis revealed a limited number of interactions, all based on hand gestures, with a total of 15 gestures reported. For example, Serpiva *et al.* [4] proposes a dictionary of 8 gestures: “one,” “two,” “three,” “four,” “five,” “okay,” “rock,” and “thumbs up” for a painting app. Fedoseev *et al.* [43], proposes a set of 4 gestures: “moving forward,” “moving backward,” “moving right,” and moving left” for an application that achieves high-performance haptic rendering of objects in the virtual environment. Tsykunov *et al.* [44] proposes a set of 4 gestures: “extended state (increasing distance, constant distance, decreasing distance),” “contracted state (increasing distance, constant distance, decreasing distance),” for a human-kick interaction application in which vibrotactile feedback is received. Unfortunately, these interaction modes have not been evaluated by the user, and consequently, their implementation for drone control remains unknown. Thus, an immediate implication of our findings in relation to the formulated research questions is that more studies are needed to considerably increase the set of gestures for the interaction between operators and drone swarms but at the same time also new input modalities using e.g. wearable devices.

B. User Studies and Evaluations

The previous research direction referred to the importance of input modalities and interaction techniques for the interaction between operators and drone swarms. In the following, we highlight the need for user studies and evaluation of the results, since 55% of the articles presented user studies. For example, the largest number of users involved in a study on the interaction between operators and drone swarms was in the paper [38], in which an experiment was conducted on a game-like application. The second paper in which there were participants in the experiment is Macchini *et al.* [40] in which users’ hand movements are captured using Leap Motion. There were also participants in studies in articles [4], [43] in which seven participants were identified, and the smallest number of users in the studies was 6 [44]. It is therefore necessary to have more users in the studies to be proposed, while at the same time having more methods to evaluate the users. As we have mentioned, in all the studies in which there were participants, we only had experimental status. At the same time, we have to increase the average ages because for the mentioned articles we identify a mean of 8.6 (SD=2.8) and a median of 7.0 participants per study.

C. Involvement of Multiple Drones in Simulations

The use of multiple drones to compose the swarm can be beneficial for more complex missions. We, therefore, present the number of drones that composed the swarm, and then how many swarms were used in our selected articles that focus on the interaction between operators and drone swarms, based on the use of a command and control center. Thus, the largest number of drones was identified in Macchini’s paper *et al.* [40] in which the 20 drones were used for an application in which the drones were controlled by gestures captured by a Leap Motion controller. The articles that presented drone swarms consisting of only four flying devices were [38], [39], [44], and the article that identified the smallest drone formation was composed of three flying devices in Fedoseev *et al.* [43]

proposing “Dandelion-Touch” a new type of haptic feedback in VR systems where no portable or wearable interface is required. Therefore, we identify swarms of small drones that are all simulated. Thus, there is an urgent need to form much larger drone swarms, observe user behavior, and also understand how difficult it is to interact with large drone swarms. At the same time extending multiple drone groups, subjected the same user to interact with one or more drones, or even with different swarms he has under management. Using more drones that have a radar device integrated into the ambient intelligence environment [46].

D. Control and Command Devices and Prototypes

The use of devices to control and command drone swarms is necessary because our investigations have uncovered wearable devices such as webcams in the works [4], [38], but also devices such as leap motion [40]. Thus, we suggest that there is a great need for new devices to control and command drones. For example, in Tsykunov *et al.* [44] a smart glove is used to control the drone swarm. Our suggestion is that you can introduce another glove to control one drone in the swarm, and the other to control the whole swarm. Another presented work uses a headset for the interaction between human and drone swarm. An overview of the system functions and control commands for drone-based aerial photography and video, as well as a dictionary of gestures can be found in the scientific article [47]. Therefore, there is an urgent need for new devices to interact with drones but at the same time develop systems that combine known and tested devices.

V. CONCLUSION

This scientific paper presents a systematic literature review on the interaction between operators and drone swarms based on the use of a command and control center. We highlight that drone swarms are used in different types of applications, but we have not identified a defining one for the scientific community. In terms of drone interaction, only hand gestures are used and feedback is distributed to different wearable devices.

To understand the potential of drone swarms, the results show that there is still a need for contributions, and at the same time to test more interaction modalities, algorithms adapted for drone swarms, and more experimental results. Our investigations have shown that control systems are in their infancy and that few devices are used to form control systems, mainly focused on the visualization of drone information and to a very small extent on drone control.

The proposed implications serve the scientific community to encourage work in the field of drone interaction, which can be applied in different domains whereby we want to emphasize that more work is needed to understand the challenges of interacting with drone swarms, user interaction preferences, and performance in the use of these systems, for which we have proposed four research directions. We hope that our contributions will stimulate the development of new research directions for the interaction between operators and drone swarms based on the use of a command and control center.

ACKNOWLEDGMENT

This research was funded by the project “119722/Center for Knowledge Transfer to ICT-CENTRIC Enterprises, Subsidiary Contract 22081/05.10.2022, Autonomous Recognition and Support Platform - PARS”, contract no. 5/AXA 1/1.2.3/G/13.06.2018, SMIS code 2014+ 119722 (ID P-40-305)

The icons used in Fig. 1 were made by Jeremy (“Icon Pack: Drone Personal And Recreational Uses Usage And Applications”) from Flaticon (<https://www.flaticon.com>).

DECLARATIONS

The authors declare no conflict of interest. Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

REFERENCES

- [1] J. Cauchard, W. Gover, W. Chen, S. Cartwright, and E. Sharlin, “Drones in wonderland – disentangling collocated interaction using radical form,” *IEEE Robotics and Automation Letters*, pp. 1–1, 2021.
- [2] A.-I. Şiean, C. Pamparău, A. Sluÿters, R.-D. Vatavu, and J. Vanderdonck, “Flexible gesture input with radars: systematic literature review and taxonomy of radar sensing integration in ambient intelligence environments,” *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–15, 04 2023.
- [3] Y.-A. Chen, T.-Y. Wu, T. Chang, J. Y. Liu, Y.-C. Hsieh, L. Y. Hsu, M.-W. Hsu, P. Taelle, N.-H. Yu, and M. Y. Chen, “Arpilot: Designing and investigating ar shooting interfaces on mobile devices for drone videography,” in *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services*, ser. MobileHCI '18. New York, NY, USA: Association for Computing Machinery, 2018. [Online]. Available: <https://doi.org/10.1145/3229434.3229475>
- [4] V. Serpiva, E. Karmanova, A. Fedoseev, S. Perminov, and D. Tsetserukou, “Dronepaint: Swarm light painting with dnn-based gesture recognition,” in *ACM SIGGRAPH 2021 Emerging Technologies*, ser. SIGGRAPH '21. New York, NY, USA: Association for Computing Machinery, 2021. [Online]. Available: <https://doi.org/10.1145/3450550.3465349>
- [5] J. L. E. I. L. E, J. A. Landay, and J. R. Cauchard, “Drone and wo: Cultural influences on human-drone interaction techniques,” in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, ser. CHI '17. New York, NY, USA: Association for Computing Machinery, 2017, p. 6794–6799. [Online]. Available: <https://doi.org/10.1145/3025453.3025755>
- [6] O. Alon, S. Rabinovich, C. Fyodorov, and J. R. Cauchard, “Drones in firefighting: A user-centered design perspective,” in *Proceedings of the 23rd International Conference on Mobile Human-Computer Interaction*, ser. MobileHCI '21. New York, NY, USA: Association for Computing Machinery, 2021. [Online]. Available: <https://doi.org/10.1145/3447526.3472030>
- [7] W. Yamada, K. Yamada, H. Manabe, and D. Ikeda, “Isphere: Self-luminous spherical drone display,” in *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*, ser. UIST '17. New York, NY, USA: Association for Computing Machinery, 2017, p. 635–643. [Online]. Available: <https://doi.org/10.1145/3126594.3126631>
- [8] J. Cacace, A. Finzi, V. Lippiello, M. Furci, N. Mimmo, and L. Marconi, “A control architecture for multiple drones operated via multimodal interaction in search & rescue mission,” in *2016 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)*. IEEE Press, 2016, p. 233–239. [Online]. Available: <https://doi.org/10.1109/SSRR.2016.7784304>
- [9] M. Nafiz Hasan Khan and C. Neustaedter, “An exploratory study of the use of drones for assisting firefighters during emergency situations,” in *Proceedings of the Conference on Human Factors in Computing Systems, Glasgow, UK, 2019*, pp. 4–9.
- [10] B. Kim, H. Y. Kim, and J. Kim, “Getting home safely with drone,” in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*, ser. UbiComp '16. New York, NY, USA: Association for Computing Machinery, 2016, p. 117–120. [Online]. Available: <https://doi.org/10.1145/2968219.2971426>
- [11] A. Colley, L. Virtanen, P. Knierim, and J. Häkkinä, “Investigating drone motion as pedestrian guidance,” in *Proceedings of the 16th International Conference on Mobile and Ubiquitous Multimedia*, ser. MUM '17. New York, NY, USA: Association for Computing Machinery, 2017, p. 143–150. [Online]. Available: <https://doi.org/10.1145/3152832.3152837>
- [12] A. Yeh, P. Ratsamee, K. Kiyokawa, Y. Uranishi, T. Mashita, H. Takemura, M. Fjeld, and M. Obaid, “Exploring proxemics for human-drone interaction,” in *Proceedings of the 5th International Conference on Human Agent Interaction*, ser. HAI '17. New York, NY, USA: Association for Computing Machinery, 2017, p. 81–88. [Online]. Available: <https://doi.org/10.1145/3125739.3125773>
- [13] M. N. H. Khan and C. Neustaedter, “An exploratory study of the use of drones for assisting firefighters during emergency situations,” in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, ser. CHI '19. New York, NY, USA: Association for Computing Machinery, 2019, p. 1–14. [Online]. Available: <https://doi.org/10.1145/3290605.3300502>
- [14] C.-F. Chen, K.-P. Liu, and N.-H. Yu, “Exploring interaction modalities for a selfie drone,” in *SIGGRAPH Asia 2015 Posters*, ser. SA '15. New York, NY, USA: Association for Computing Machinery, 2015. [Online]. Available: <https://doi.org/10.1145/2820926.2820965>
- [15] B. Jones, K. Dillman, R. Tang, A. Tang, E. Sharlin, L. Oehlberg, C. Neustaedter, and S. Bateman, “Elevating communication, collaboration, and shared experiences in mobile video through drones,” in *Proceedings of the 2016 ACM Conference on Designing Interactive Systems*, ser. DIS '16. New York, NY, USA: Association for Computing Machinery, 2016, p. 1123–1135. [Online]. Available: <https://doi.org/10.1145/2901790.2901847>
- [16] J. Auda, M. Weigel, J. R. Cauchard, and S. Schneeggass, “Understanding drone landing on the human body,” in *Proceedings of the 23rd International Conference on Mobile Human-Computer Interaction*, ser. MobileHCI '21. New York, NY, USA: Association for Computing Machinery, 2021. [Online]. Available: <https://doi.org/10.1145/3447526.3472031>
- [17] C. Rubens, S. Braley, J. Torpegaard, N. Lind, R. Vertegaal, and T. Merritt, “Flying lego bricks: Observations of children constructing and playing with programmable matter,” in *Proceedings of the Fourteenth International Conference on Tangible, Embedded, and Embodied Interaction*, ser. TEI '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 193–205. [Online]. Available: <https://doi.org/10.1145/3374920.3374948>
- [18] Z. Xu, J. Yang, C. Peng, Y. Wu, X. Jiang, R. Li, Y. Zheng, Y. Gao, S. Liu, and B. Tian, “Development of an uas for post-earthquake disaster surveying and its application in ms7.0 lushan earthquake, sichuan, china,” *Computers and Geosciences*, vol. 68, no. Complete, pp. 22–30, 2014.
- [19] G. Dooly, E. Omerdic, J. Coleman, L. Miller, A. Kaknjo, J. Hayes, J. Braga, F. Ferreira, H. Conlon, H. Barry, J. Marcos-Olaya, T. Tuohy, J. Sousa, and D. Toal, “Unmanned vehicles for maritime spill response case study: Exercise cathach,” *Marine Pollution Bulletin*, vol. 110, no. 1, pp. 528–538, 2016. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0025326X16301242>
- [20] M. Fleck, “Usability of lightweight defibrillators for uav delivery,” in *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, ser. CHI EA '16. New York, NY, USA: Association for Computing Machinery, 2016, p. 3056–3061. [Online]. Available: <https://doi.org/10.1145/2851581.2892288>
- [21] D. Tezza and M. Andujar, “The state-of-the-art of human-drone interaction: A survey,” *IEEE Access*, vol. 7, pp. 1–1, 01 2019.
- [22] M. Funk, “Human-drone interaction: Let’s get ready for flying user interfaces!” *Interactions*, vol. 25, no. 3, p. 78–81, apr 2018. [Online]. Available: <https://doi.org/10.1145/3194317>
- [23] J. Cauchard, J. E. K. Zhai, and J. Landay, “Drone and me: an exploration into natural human-drone interaction,” in *Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing*. Association for Computing Machinery, Inc, 2015, pp. 361–365.

- [24] S. Mirri, C. Prandi, and P. Salomoni, "Human-drone interaction: State of the art, open issues and challenges," in *Proceedings of the ACM SIGCOMM 2019 Workshop on Mobile AirGround Edge Computing, Systems, Networks, and Applications*, ser. MAGESys'19. New York, NY, USA: Association for Computing Machinery, 2019, p. 43–48. [Online]. Available: <https://doi.org/10.1145/3341568.3342111>
- [25] V. Herdel, L. J. Yamin, and J. R. Cauchard, "Above and beyond: A scoping review of domains and applications for human-drone interaction," in *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, ser. CHI '22. New York, NY, USA: Association for Computing Machinery, 2022. [Online]. Available: <https://doi.org/10.1145/3491102.3501881>
- [26] A. Siddaway, A. Wood, and L. Hedges, "How to do a systematic review: A best practice guide for conducting and reporting narrative reviews, meta-analyses, and meta-syntheses," *Annual Review of Psychology*, vol. 70, 01 2019.
- [27] A. Liberati, D. Altman, J. Tetzlaff, C. Mulrow, P. Gøtzsche, J. Ioannidis, M. Clarke, P. Devereaux, J. Kleijnen, and D. Moher, "The prisma statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration," *Journal of clinical epidemiology*, vol. 62, pp. e1–34, 08 2009.
- [28] A. Wojciechowska, J. Frey, E. Mandelblum, Y. Amichai-Hamburger, and J. R. Cauchard, "Designing drones: Factors and characteristics influencing the perception of flying robots," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 3, no. 3, sep 2019. [Online]. Available: <https://doi.org/10.1145/3351269>
- [29] Z. Li, J. Xu, and Q. Guo, "Information fusion algorithm based on intelligent algorithm for multiple uavs information interaction deception," in *2018 5th IEEE International Conference on Cloud Computing and Intelligence Systems (CCIS)*, 2018, pp. 38–42.
- [30] S. A. CELTEK, A. DURDU, and E. KURNAZ, "Design and simulation of the hierarchical tree topology based wireless drone networks," in *2018 International Conference on Artificial Intelligence and Data Processing (IDAP)*, 2018, pp. 1–5.
- [31] J. J. Corner and G. B. Lamont, "Parallel simulation of uav swarm scenarios," in *Proceedings of the 36th Conference on Winter Simulation*, ser. WSC '04. Winter Simulation Conference, 2004, p. 355–363.
- [32] F. Fabra, J. Wubben, C. T. Calafate, J. C. Cano, and P. Manzoni, "Efficient and coordinated vertical takeoff of uav swarms," in *2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring)*, 2020, pp. 1–5.
- [33] H. N. Saha, N. K. Das, S. K. Pal, S. Basu, S. Auddy, R. Dey, A. Nandy, D. Pal, N. Roy, D. Mitra, S. Biswas, and T. Maity, "A cloud based autonomous multipurpose system with self-communicating bots and swarm of drones," in *2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC)*, 2018, pp. 649–653.
- [34] H. A. Shehata and M. El-Helw, "Modeling collaborative ai for dynamic systems of blockchain-ed autonomous agents," in *2021 3rd Novel Intelligent and Leading Emerging Sciences Conference (NILES)*, 2021, pp. 421–426.
- [35] C.-Q. Dai, X. Li, and Q. Chen, "Intelligent coordinated task scheduling in space-air-ground integrated network," in *2019 11th International Conference on Wireless Communications and Signal Processing (WCSP)*, 2019, pp. 1–6.
- [36] C. N. Mavridis, A. Tirumalai, and J. S. Baras, "Learning swarm interaction dynamics from density evolution," *IEEE Transactions on Control of Network Systems*, pp. 1–12, 2022.
- [37] H. Arksey and L. O'Malley, "Scoping studies: towards a methodological framework," *International Journal of Social Research Methodology*, vol. 8, no. 1, pp. 19–32, 2005. [Online]. Available: <https://doi.org/10.1080/1364557032000119616>
- [38] E. Karmanova, V. Serpiva, S. Perminov, R. Ibrahimov, A. Fedoseev, and D. Tsetsrukou, "Swarmplay: A swarm of nano-quadcopters playing tic-tac-toe board game against a human," in *ACM SIGGRAPH 2021 Emerging Technologies*, ser. SIGGRAPH '21. New York, NY, USA: Association for Computing Machinery, 2021. [Online]. Available: <https://doi.org/10.1145/3450550.3465346>
- [39] E. Tsykunov, R. Agishev, R. Ibrahimov, L. Labazanova, T. Moriyama, H. Kajimoto, and D. Tsetsrukou, "Swarmcloak: Landing of a swarm of nano-quadrotors on human arms," in *SIGGRAPH Asia 2019 Emerging Technologies*, ser. SA '19. New York, NY, USA: Association for Computing Machinery, 2019, p. 46–47. [Online]. Available: <https://doi.org/10.1145/3355049.3360542>
- [40] M. Macchini, L. De Matteis, F. Schiano, and D. Floreano, "Personalized human-swarm interaction through hand motion," *IEEE Robotics and Automation Letters*, vol. 6, no. 4, pp. 8341–8348, 2021.
- [41] X. Zhang and M. Ali, "A bean optimization-based cooperation method for target searching by swarm uavs in unknown environments," *IEEE Access*, vol. 8, pp. 43 850–43 862, 2020.
- [42] A. Agrawal, J. Cleland-Huang, and J.-P. Steghöfer, "Model-driven requirements for humans-on-the-loop multi-uav missions," in *2020 IEEE Tenth International Model-Driven Requirements Engineering (MoDRE)*, 2020, pp. 1–10.
- [43] A. Fedoseev, A. Baza, A. Gupta, E. Dorzhieva, R. N. Gujarathi, and D. Tsetsrukou, "Dandeliontouch: High fidelity haptic rendering of soft objects in vr by a swarm of drones," in *2022 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2022, pp. 1078–1083.
- [44] E. Tsykunov, L. Labazanova, A. Tleugazy, and D. Tsetsrukou, "Swarmtouch: Tactile interaction of human with impedance controlled swarm of nano-quadrotors," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2018, pp. 4204–4209.
- [45] J. P. Distefano, H. Manjunatha, S. Chowdhury, K. Dantu, D. Doermann, and E. T. Esfahani, "Using physiological information to classify task difficulty in human-swarm interaction," in *2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2021, pp. 1198–1203.
- [46] A.-I. Şiean, C. Pamparau, and R.-D. Vatavu, "Scenario-based exploration of integrating radar sensing into everyday objects for free-hand television control," *ACM International Conference on Interactive Media Experiences*, 2022.
- [47] A.-I. Şiean, R.-D. Vatavu, and J. Vanderdonckt, "Taking that perfect aerial photo: A synopsis of interactions for drone-based aerial photography and video," in *ACM International Conference on Interactive Media Experiences*, ser. IMX '21. New York, NY, USA: Association for Computing Machinery, 2021, p. 275–279. [Online]. Available: <https://doi.org/10.1145/3452918.3465484>

IM2P-Medical: Towards Individual Management Privacy Preferences for the Medical Web Apps

Nguyen Ngoc Phien¹, Nguyen Thi Hoang Phuong², Khiem G. Huynh³

Khanh H. Vo⁴, Phuc T. Nguyen⁵, Khoa D. Tran⁶, Bao Q. Tran⁷

Loc C. P. Van⁸, Duy T. Q. Nguyen⁹, Hieu M. Doan¹⁰, Bang K. Le¹¹

Trong D. P. Nguyen¹², Ngan T. K. Nguyen¹³, Huong H. Luong¹⁴, Duong Hon Minh¹⁵

Center for Applied Information Technology, Ton Duc Thang University, Ho Chi Minh City, Viet Nam¹

Faculty of Information Technology, Ton Duc Thang University, Ho Chi Minh City, Viet Nam¹

FPT University, Can Tho City, Viet Nam^{3,4,5,6,7,8,9,10,11,12,14}

FPT Polytechnic, Can Tho City, Viet Nam¹³

Faculty of Information Technology, Pham Van Dong University, Quang Ngai Province, Viet Nam²

Faculty of Pharmacy, Nguyen Tat Thanh University, Ho Chi Minh City, Viet Nam¹⁵

Abstract—With the advancement of technology, people are now able to monitor their health more efficiently. Mobile phones and smartwatches are equipped with sensors that can measure real-time changes in blood pressure, SPO2, and other attributes and public them to service providers via web applications (called web apps) for health improvement suggestions. Moreover, users can share the collected health data with other people, such as doctors, relatives, or friends. However, using technology in healthcare has raised the issue of privacy. Some health web apps, by default, intrusively gather and share data. Additionally, smartwatches may monitor people's health status 24/7. Therefore, users want to control how their health is processed (e.g., collected and shared). This can be cumbersome as they would have to configure each device manually. To address this problem, we have developed a privacy-preference prediction mechanism in the web apps called IM2P-Medical: towards Individual Management Privacy Preferences for the Medical web apps. To capture individual privacy preferences in the web apps, our model learns users' privacy behavior based on their responses in different medical scenarios. In practice, we exploited several machine learning algorithms: SVM, Gradient Boosting Classifier, Ada Boost Classifier, and Gradient Boosting Regressor. To prove the effectiveness of the proposed model, we set up several scenarios to measure the accuracy as well as the satisfaction level in the two participant groups (i.e., expert and normal users). One key point in this research's selection of participants is its focus on those living in developing countries, where privacy violation issues are not a common topic. The main contribution of our model is that it allows users to preserve their privacy without configuring privacy settings themselves.

Keywords—Letter-of-credit; cash-on-delivery; blockchain; smart contract; NFT; ethereum; fantom; polygon; binance smart chain

I. INTRODUCTION

Monitoring health with technological devices and web apps has gained great popularity in this day and age. In fact, the market value of health monitoring devices in 2019 was \$25,78.56 million and is predicted to soar to \$44,861.56 million in 2027¹. The devices can track multiple values like blood pressure, heart rate, and sleep quality and send them to a web app for visualization. Based on the collected data

(called Evidence-based disease management [1], [2]) from these sensors, several approaches could detect the corresponding diseases. Furthermore, these devices can monitor users' metrics everywhere due to their portability. Thanks to these gadgets, hospitals can better support their patients, and people can take care of their health more efficiently. However, due to the vast amount of data that can be collected, the use of health-monitoring devices and web apps faces doubt from those value data privacy [3]. Thus, there is a demand for a solution that manages how personal data is collected by health devices and web apps.

Conventionally, users can manually adjust their privacy preferences via web apps' or devices' settings. However, this can be cumbersome and may not be effective. On the other hand, an automated solution that can suggest security settings for a user based on his/her personality or privacy preference can bring better results [4]. In this paper, we introduce our privacy preference prediction solution. Our system learns a user's security perspective and makes suitable suggestions for changing privacy settings.

Due to some economic barriers in developing countries, their citizens lack healthcare services and institutions. Thus, the privacy issues of the medical data are ignored. Currently, there are no medical data protection standards like the developed countries (e.g., European countries - GDPR²) to protect the user privacy issues. To address this drawback, our model focuses on individual privacy preferences w.r.t medical data - especially, in developing countries. Our dataset explores the feedback from the developing countries' citizens, e.g., Asia, Africa, and Latin America.

Moreover, most current health systems are focused on protecting users' medical data [5], [6]. For example, Son et al. [7] emphasizes the importance of user privacy preferences that are placed alongside the system's privacy policy. This means that the requestor must satisfy both privacy policy and privacy preferences in order to be able to access the patient's medical data. Besides, Hoang et al. [8] provide a mechanism to handle

¹<https://www.alliedmarketresearch.com/patient-monitoring-devices-market>

²The GDPR document is available at https://edps.europa.eu/data-protection/data-protection/glossary/d_en#data_minimization

conflicts between the privacy policy and privacy preferences where it depends on prioritizing patient treatment or reducing the risk of personal information leakage. In addition to the above studies, the systems that build smart contract models for medical facilities using Blockchain technology also take care of users' privacy preferences, for example, Nghia et al. [9], [10], [11]. In these studies, the patient role was given full discretion in sharing their data with stakeholders. Moreover, the priority of treatment is also exploited in the approach of [12], [13], where patients allow access to their personal data in case of an emergency. Also, in the medical-related system, several studies deployed the individual privacy preference based on blockchain technology (e.g., blood donate [14], [15] or IoT medical sensors [16] via the transmission messages protocol [17]).

To make this suggestion function work, we introduce IM2P-Medical: Towards individual management privacy preferences w.r.t medical data based on a Machine Learning system, i.e., built based on Semi-Supervised Learning. The reason for choosing the Semi-Supervised Learning method is to be able to reduce the amount of data required while preserving the prediction accuracy. The data for this system was gathered via a questionnaire. This questionnaire aims to learn respondents' perspectives or attitudes toward privacy. We distribute the questions to two types of people. The former type was people who had a background in IT and privacy, such as IT students (i.e., expert participants), while the latter group included various types of people called average users (i.e., normal participants) - see Sect. IV-B. who responded to the questionnaire on the Internet.

The key contributions of this paper are three-fold, including i) designing the individual privacy preferences architecture w.r.t medical data (i.e., IM2P-Medical - Sect. II); ii) balancing the user burden and accuracy based on the semi-supervised learning approach (i.e., Implementation - Sect. III); and iii) proving the effectiveness of the IM2P-Medical based on the two participant dataset.

The remainder of this paper is organized as follows. Sections II and III introduce the IM2P-Medical architecture and implementation (i.e., training strategy, algorithm, and questionnaire). Section IV presents experimental results, whereas related work are discussed in Section V. Finally, Section VI concludes the paper.

II. IM2P-MEDICAL ARCHITECTURE

Fig. 1 shows the interaction process among the parties in the proposed model, including i) Users; ii) Personal data; iii) People; iv) Service Providers, and v) IM2P-Medical. The main roles and responsibilities of these parties in the system are presented as follows:

User (also known as the data owner): they have the right to make a decision whether to share their personal data by responding with consent (Yes) or disagree (No). The data in their possession includes normal data (i.e., easy to share) and personal data (i.e., medical data).

Personal data: the data that needs to be protected because they are highly identifiable. As a result, a malicious user can

obtain other types of user data based on the exploitation of this data pool. This study focused on grouping personal data in the medical environment which can be exploited by sensors or smartphones (e.g., heart rate, SPO2, calories burned).

People: Can act as a user (in some specific cases). **People** represent other users (in the same or not the same system) who have a relationship (e.g., relative, friend,) with the owner of the data. This party can have more than one relationship with the data owner.

Service providers: the party provides the necessary medical services to users (e.g., health monitoring, disease diagnosis, online doctor)³. This target group provides a specific type of health care service; in return, the user must provide the data requested by the service provider. In a traditional environment, users have to provide virtually any type of data to service providers, ignoring privacy risks [18]. Previous studies have shown that applications collect more data than what they need for the supported services [19]. In this paper, any data manipulation request must be accompanied by a corresponding purpose to eliminate this drawback.

IM2P-Medical: this party automatically identifies privacy preferences for each individual. Specifically, this model identifies users' privacy behaviors based on their responses to service providers' access requests. Besides, Fig. 1 also depicts five main components of IM2P-Medical including: data types; relationship(s); context; access request(s); and purpose(s). The relationship between IM2P-Medical and the remaining parties is presented as follows:

- **Data types:** the types of data (e.g., location, heart rate, etc). In fact, the identification of personal data also depends on the user's sense of privacy. Each individual will make a different decision depending on many factors. It is not possible to define all possible possibilities in this study. So we're targeting the kind of medical data that are being exploited by the user's sensors, wearable devices, and smartphones. To ensure that our survey achieves its stated objectives (i.e., risks of personal information disclosure), we also emphasized in our survey that these medical data can easily be exploited through their device without any notification to the user.
- **Relationship(s):** This group of attributes also greatly affects the issue of sharing personal data. A user can easily share their walking record (e.g., steps, distance traveled, start and end locations) for a day with his/her friends or personal training etc. Users will share or not share their data depending on each specific relationship.
- **Context:** depending on the specific context, users can (not) share their data regardless of the same data type and relationship. For example, in healthcare scenarios (hospitals, clinics), heart rate data can be shared with **People** as healthcare workers (e.g., doctors, nurses); however, same data types and relationships but different contexts (e.g., sports participation) - users can opt-out of sharing. To be able to capture each user's data

³The service provider can reserve several services, but we target the medical environment

sharing behavior, we exploit sub-attributes (i.e., data types; relationship(s); access request(s); purpose(s)) on the context-specific (see Sect. III-C for more details).

- **Access request(s):** This component is closely associated with the service provider. In contrast to **People**, where users voluntarily share their data with a specific purpose (i.e., decided by themselves), the data retrieval process for service providers is the opposite. Specifically, the structure must include the party of the request for access (e.g., medical or fitness apps) and the corresponding purpose (discussed in the next section). Users judge between the benefits and risks of privacy to make a decision.
- **Purpose(s):** One of the important pieces of information to decide whether users share their data or not is the purpose of access. In particular, a series of analyses have shown that requests for supporting the application's service will be accepted more than advertising purposes. To clarify this, we also emphasize the importance of access intent in our survey scenarios (see Sect. III-C for more details).

III. IMPLEMENTATION

A. Self-training

Self-training or "Self-learning" is the most basic of pseudo-labeling approaches [20]. They consist of a single supervised classifier that is iteratively trained on both labeled data and unlabeled data that has been pseudo-labeled in previous iterations of the algorithm. At the initial procedure, a supervised classifier is trained on only the labeled data. The outcome of the classifier is used to obtain predictions for the unlabeled data. Then, the most confident of these predictions is added to the labeled data set, and the supervised classifier is re-trained on both the original labeled data and the newly obtained pseudo-labeled data. This procedure is typically iterated until no more unlabeled data remain.

Several applications and variations of self-training have been put forward. For instance, Rosenberg et al. [21] applied self-training to object detection problems, and showed improved performance over a state-of-the-art (at that time) object detection model. Dopido et al. [22] developed a self-training approach for hyperspectral image classification. They used domain knowledge to select a set of candidate unlabeled samples, and pseudo-labeled the most informative of these samples with the predictions made by the trained classifier.

B. Algorithm

Algorithm 1 applies self-training model to label the *apps* in *UApp* dataset. Specifically, it applies the SVM algorithm (several supervised methods) to pseudo-label the apps in the unlabeled data set (*UApp*) (see line 4). For the other supervised learning algorithm, we do the same idea.

C. Questionnaire

To make accurate suggestions, effectively learning users' behaviours is of paramount importance. We have meticulously

Algorithm 1 selfTraining(*LApp*, *UApp*, *supAlg*)

```
1: input: training apps LApp, target apps UApp, list of supervised algorithms supAlg.
2: output: label for UApp.
3: for each  $app_i \in UApp$  do
4:    $label_{app_i} = SVM(app_i)$ ;
5:    $UApp - \{app_i\}$ ;
6:    $LApp \cup \{(app_i, label_{app_i})\}$ ;
7: end for
```

designed a questionnaire for the learning purpose. Our questionnaire focuses on observing how users will adjust their privacy preferences in multiple contexts. To be specific, from our questions, we expect to answer three queries:

- 1) Given a specific context, how will participants share different types (e.g., heart rate, SPO2, burned calories) of data with other people (e.g., friends, relatives, doctors)?
- 2) Given a specific context, how will participants share the data for a certain purpose (e.g., analysis, education, ads)?
- 3) Given a specific context, how will participants share data with service providers (e.g., medical apps, fitness apps)?

For each query, we develop an appropriate type of question. In each question, there are a number of parameters whose values can be activities, individuals, permissions, etc. Participants have three options, they may completely agree, completely disagree or partly agree with the statement.

In the following parts of this section, we will explain the given questions and introduce the list of parameters.

1) *Sharing data with others in a specific context:* In this type of question, we attempt to understand how people share personal data with others when they are doing certain activities. We classify possible activities into two groups: indoor and outdoor. An example of an outdoor activity question is: "You are playing sports. Do you want to share your location with your doctors?". In this example, we want to know if the user is willing to share their location with a doctor while playing sports.

The general structure of the questions is: "You are *an activity*. Do you want to share your *information* with a *person*". Three parameters are required: the activity, the information to share and the person to share with. We have prepared a set of possible parameters' values as given below (Table I to III).

TABLE I. THE SAMPLES OF ACTIVITIES

ID	Activity name
1	playing sport
2	relaxing
3	doing daily activities
4	at home
5	at work
6	having treatment at home
7	at the hospital
8	under an emergency

Questionnaire participants have three options to choose: Yes (without restriction), No or Yes (with restriction). If they

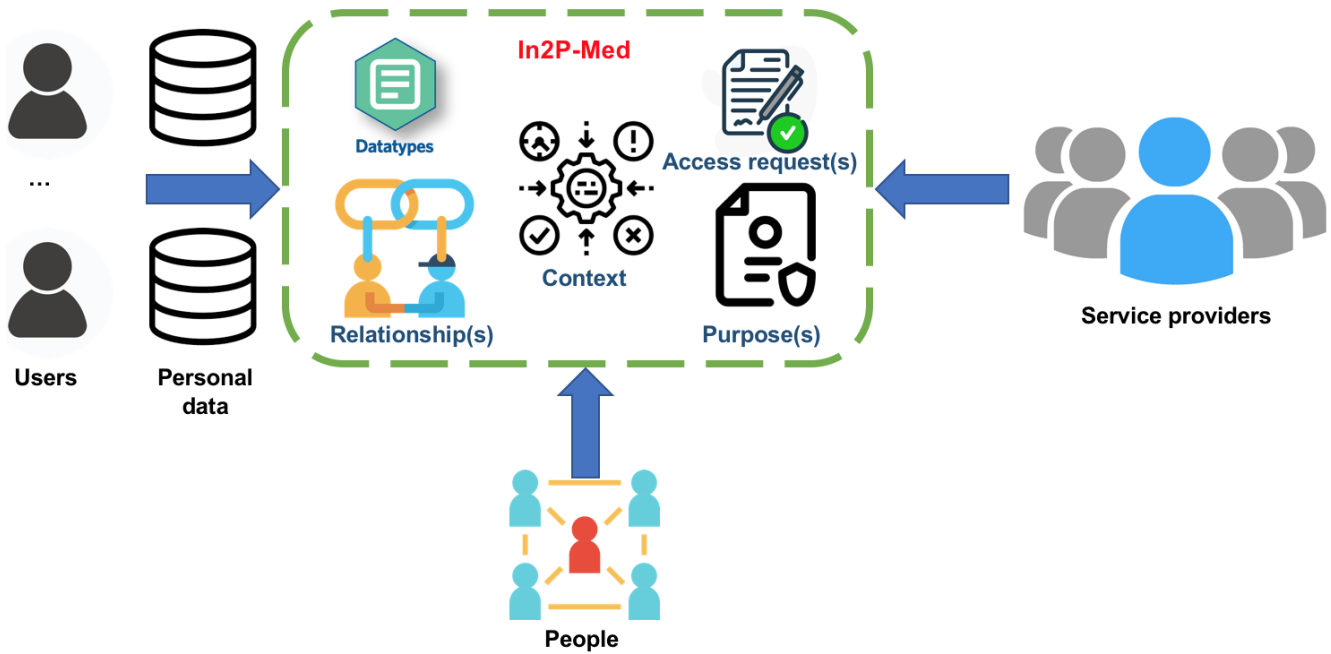


Fig. 1. IM2P-Medical architecture.

TABLE II. THE SAMPLE OF COLLECTIBLE INFORMATION

ID	Data item	ID	Data item	ID	Data item
1	name	7	steps	13	sugar level
2	phone number	8	heart rate	14	fat level
3	email	9	weight	15	travel distance
4	location	10	height	16	sleep
5	age	11	SPO2	17	health goal
6	gender	12	calories		

choose the first option, they are unconditionally willing to share data with an individual. However, if they choose to share with restrictions, they may want to implement some controls. For instance, they may share only once every 1 hour. Choosing “no” means that sharing under the given context is out of the question.

2) *Granting permissions in a specific context*: This type of question is intended to learn if users allows applications to access devices’ components such as Bluetooth or Wifi.

The general structure of the question is: “You are *an activity*. Do you allow your device to access your *a component*?”. There are two parameters required: the activity like in the first type of question and the device’s components being granted access. We have prepared a set of possible parameters’ values as given below (Table IV). Besides, we use the same set of answers in this type of question.

3) *Sharing data with applications in a specific context*: In this type of question, participants are surveyed if they allow applications in a specific category to collect data for a particular purpose. As an example, we may ask if they permit vendors making Health apps to collect usage information for marketing purposes.

The structure of the questions is: “Do you want your *data* to be collected by *an app category* for the *purpose*

purposes ? (service provider x information collected)”. Three parameters are required: the collected data, the app’s category and the collecting purpose. We have prepared a set of possible parameters’ values as given below (Table V).

We also use the same set of answers in this type of question (Table VI).

IV. EXPERIMENTAL RESULTS

A. Experiment Setting

In our tests, each participant had to take part in two phases i) they make their choice about whether to share medical data in each context different in the training period ii) they participate in the evaluation of the prediction results from our algorithms and give the satisfaction level of the corresponding algorithms. To achieve this goal, we have developed a web application to that requires interaction with participants through the two phases mentioned above. Specifically, participants label questions that share data during the learning phase (i.e. data set training data), then give their feedback on the labels generated by the models predict in the testing phase (i.e. test data set), and finally rate their satisfaction level on our predictive models.

More precisely, in the first phase, participants were asked to label each sentence (i.e. Yes (Y), No (N), or Maybe (M)) about sharing data in each term-specific scene as described in III-C. During the training phase, participants have to give answers to all 20 questions over a while. The minimum time is 10 minutes (an average of 30 seconds for an answer). After the labeling, the collected training dataset is built using algorithms learning is covered in Section III-B, specifically the -based approach to supervised learning and semi-supervised learning. To evaluate learning strategies, during the beta phase, the web app displays 20 new questions for those who participate.

TABLE III. THE ROLES OF REQUESTER

ID	Relationship	ID	Relationship	ID	Relationship	ID	Relationship
1	Friends	2	Family members	3	Doctors	4	Nurses

TABLE IV. PERMISSIONS

ID	Permission	ID	Permission	ID	Permission
1	Bluetooth	4	File Storage	7	Wifi
2	Camera	5	Microphone	8	Identity
3	Location	6	Sensors	9	Contacts

We randomly select prediction strategies instead of trying to define strategies according to the expected degree of accuracy. The main purpose is to remove all user prejudices about the algorithms in the back, which will be better than the previous algorithms. Specifically, four predictive models (five assessment questions/per model) are applied in this paper, including SVM, Gradient Boosting Classifier, Ada Boost Classifier, and Gradient Boosting Regressor.

For each new question in the experimental phase, the participants gave feedback on the labels, i.e. agree (Y) or disagree (N) - and in the case of disagreement, they must provide the correct label. For example, our forecast label is "Y", but their expected result is "M". They will give feedback on the label as no agree (N) and reselect the outcome they expected. In addition, in the 20 questions at a stage of the test phase, we reused four questions that appeared in the test phase corresponding to four predictive models. Specifically, each prediction model will have four sentences of new questions, and 1 question is randomly selected out of 20 questions during the experimental stage. The main purpose of this work is to divide into two groups of people based on their choices for that question for both periods, specifically, the selection group, the same selection, and the different selection group. The details of this comparison will be presented in section IV D 2. Finally, we collected the satisfaction level of the participants with the project guesses generated by each model. Participants can answer Yes (100%), No (0%), or Maybe (50%). The average time for answering each question at the test stage was 30 seconds. So each person participating in the survey must spend at least 20 minutes completing both learning and testing. To remove unsatisfactory answers, we have set the timer to track participants' time answering questions. If they spent less than the desired time (i.e. less than 30 seconds for a question), participants could not move on to the next question.

B. Participants

The primary purpose of this paper is to build an automatic medical data-sharing model that meets the privacy requirements of the users. We also want to explore the issue of sharing private data in developing countries where privacy is not widely aware of, especially in terms of sensitive data like medical. To achieve the above purposes, we conducted surveys. Our models are in countries in Asia, Africa, and Latin America. Besides that, there is a difference between the survey respondents on security and privacy, we categorized the differences between these two groups of users. Specifically, in the expert user group, we collected feedback from students as well as teachers who are studying and working at FPT University

(Vietnam) majoring in Information Security at two campuses in Ho Chi Minh City. Ho Chi Minh City and Can Tho. For the normal user group, we used the tool Microworkers⁴ to collect user feedback participants from developing countries.

1) *Expert users*: : For expert users (students and teachers of information security), we sent an email to the students who participated in the survey for four weeks (September 2021). There were a total of 20 qualified participants out of 32. The majority of participants were disqualified for not answering enough required questions. The average age of participants is 21.5, with the oldest and youngest being 29 and 18, respectively. Besides, about 15% of the participants were female (3/20).

2) *Normal users*: : The main purpose of this user group is to satisfy the requirement of diversity in terms of age, education, gender, and culture. We choose developing countries in two regional groups, including Latin America and Asia-Africa. We got 209 valid responses out of 296 participants. Each participant was paid \$3. The number of participants belonging to the above two regional groups is 85 and 124, respectively. The mean age is 31.06 (minimum is 18 and oldest is 70) and 28.29 (the smallest age is 18 and the largest is 59). Out of a total of 85 responses, 42 were female (49.41%). The number for Asia-Africa is 25.6% (32 out of 125 participants).

C. Confusion Matrix

We used conventional measures to evaluate the accuracy of the proposed learning methods. Specifically, we exploited the 3X3 confusion matrix corresponding to the three labels (Y, N and M) (Table VII), where the columns represent the predicted labels (generated from approaches) and possible rows of values actual (participant opinion) and cells represent Error (E) or True Positive (TP). From the confusion matrix, we determined the evaluable metrics given in Table VIII.

D. Evaluation

In the evaluation, we performed a series of measurements to find the most appropriate algorithm in detecting the sharing behavior of personal medical data, given a specific context. Specifically, in the first test, we compared the accuracy obtained by different learning approaches (specifically between supervised learning and semi-supervised learning). We first compared the semi-supervised soft clustering method and the hard clustering techniques. This comparison aims to evaluate whether a semi-supervised system has good accuracy even with a reduced training set.

In the second test, we compared the accuracy of the proposed prediction models, namely SVM, Gradient Boosting Classifier, Ada Boost Classifier, and Gradient Boosting Regressor. As the results displayed in Section IV D 3, the semi-supervised-based approach gave better results than the supervised approach. Therefore, we apply the semi-supervised model to all four proposed algorithms.

⁴<https://www.microworkers.com/>

TABLE V. APP CATEGORIES

ID	Category	ID	Category	ID	Category
1	Communication	10	Libraries & Demo	19	Productivity
2	Dating	11	Lifestyle	20	Shopping
3	Education	12	Maps & Navigation	21	Social
4	Entertainment	13	Medical	22	Sports
5	Events	14	Music & Audio	23	Tools
6	Finance	15	News & Magazines	24	Travel & Local
7	Food & Drink	16	Parenting	25	Video Players & Editors
8	Health & Fitness	17	Personalization	26	Wear OS
9	House & Home	18	Photography	27	Weather

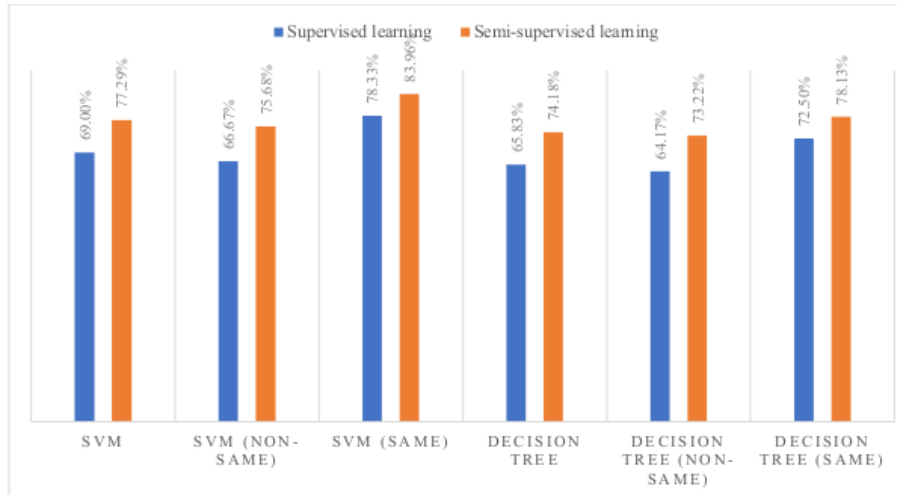


Fig. 2. The accuracy of SVM and Decision tree in supervised and semisupervised learning approaches.

TABLE VI. COLLECTING PURPOSES

ID	Purpose	ID	Purpose
1	Education	5	Scientific Research
2	Government	6	Treatment
3	Marketing/Advertising	7	Analytics
4	Product Development	8	Apps' functional

TABLE VII. CONFUSION MATRIX

	Predicted value: Y	Predicted value: N	Predicted value: M
Actual value: Y	TP_Y	$E_{Y,N}$	$E_{Y,M}$
Actual value: N	$E_{N,Y}$	TP_N	$E_{N,M}$
Actual value: M	$E_{M,Y}$	$E_{M,N}$	TP_M

1) *Supervised learning and semi-supervised learning comparison:* In this section, in addition to demonstrating which approaches (in particular supervised learning and semi-supervised learning) provide better prediction results with small data sets, we also wanted to test the difference between

TABLE VIII. METRICS DEFINITION

Accuracy	$(TP_Y + TP_N + TP_M) / \#samples$
Pre_Y	$TP_Y / (TP_Y + E_{N,Y} + E_{M,Y})$
Pre_N	$TP_N / (TP_N + E_{Y,N} + E_{M,N})$
Pre_M	$TP_M / (TP_M + E_{Y,M} + E_{N,M})$
Re_Y	$TP_Y / (TP_Y + E_{Y,N} + E_{Y,M})$
Re_N	$TP_N / (TP_N + E_{N,Y} + E_{N,M})$
Re_M	$TP_M / (TP_M + E_{M,Y} + E_{M,N})$
$F1_X$	$2 * (Pre_X * Re_X) / (Pre_X + Re_X)$, where $X \in \{Y, N, M\}$

homogeneous and heterogeneous user groups in terms of data sharing decisions.

To achieve the above goals, we first compare the accuracy between supervised learning and semi-supervised learning approaches by building a training set that is a subset of the original training set (with 10, instead of 20 questions). Specifically, in the new dataset, the number of questions in the shuttered train is ten and in the test set is 30 (including ten questions transferred from the train set). The main purpose for this reallocation of questions is that we wanted to aim for an approach that can balance the effort spent by the user to build the training set and the accuracy of the applied algorithm. A good approach that can satisfy the above criteria is the one that only has a small number of questions in the training set and ensures an acceptable accuracy. To meet the above requirements, we used a semi-supervised learning approach for the SVM algorithm, decision tree, and supervised learning for both SVM and Decision tree. We compared the accuracy of the SVM algorithm for both approaches as well as two different algorithms (SVM and decision tree) to get the most general view when choosing strategies to build predictive models.

Fig. 2 shows the accuracy between SVM and Decision tree on both approaches: semi- and supervised learning. It shows that the semi-supervised-based approach is always better than supervised learning for both SVM and Decision tree algorithms as well as groups of participants (same and non-same). SVM algorithm has higher accuracy than the Decision tree for all cases. Besides, the same answer group has the highest accuracy in each algorithm. This proves that the approach

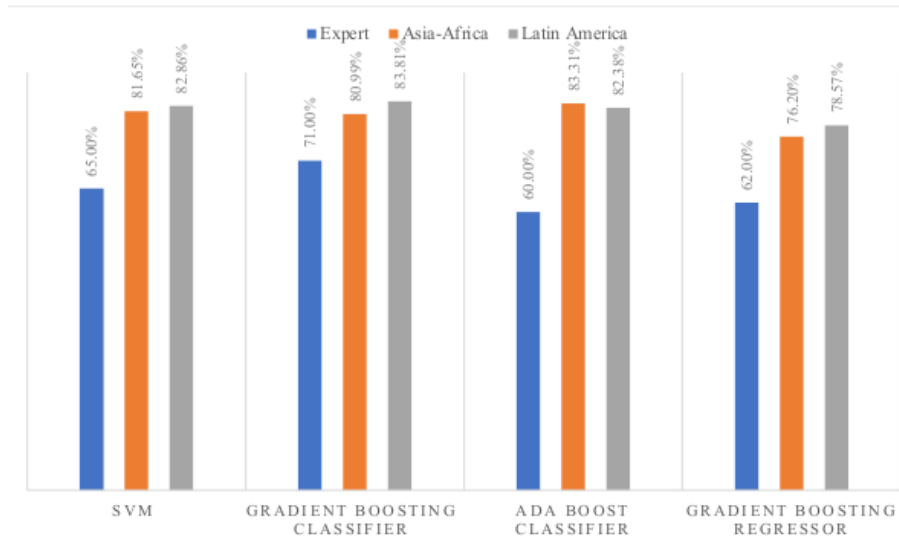


Fig. 3. Accuracy level of four classification models.

based on semi-supervised learning gives high accuracy even when trained on a small data set. In the following section, we apply a semi-supervised learning approach to delve into the analysis of accuracy, F1 score as well as the satisfaction of survey participants.

2) *Answer prediction model accuracy*: Fig. 3 depicts the accuracy of four predictive models for all three datasets (expert, Asia-Africa, and Latin America). The group with the lowest accuracy was experts from 60-71%, while the highest accuracy group was the group of participants from Asia-Africa and Latin America from 76.20% to 83.81%. This proves that the behavior of the user group is often easier to capture than the expert user group. Indeed, based on a manual analysis of users' comments on the reasons for their choice, we found that the difference between the two groups of users lies in the context of sharing personal data in the medical environment. In the case of ordinary users, they only care about the object to be shared or the type of data requested. Meanwhile, the experts evaluated all three groups of data, the shared audience, and especially their context. In particular, they are very careful when sharing high-risk data (motion data, location) with any individual (including relatives and friends). On the other hand, user groups are often more comfortable sharing personal data. They are willing to trade personal information to choose the services or utilities that the services or applications bring. Moreover, they trust the personal data protection mechanism of the service or application as well as the group of people who share it closely (relatives, friends).

3) *Satisfaction level*: This section evaluates the satisfaction level of four prediction algorithms for all three datasets (see Fig. 4). Specifically, all approaches have high satisfaction for all three datasets (from 90% to 98.81%). The algorithm with the highest satisfaction is the Gradient Boosting classifier (95% - 98.81%). Meanwhile, the algorithm with the lowest satisfaction level is SVM (95% - 98.81%). This experiment proves that the semi-supervised learning-based approach brings about a high level of satisfaction for all three groups of participants.

4) *F1 score*: Finally, we measured the score F_1 for each label (Y, N, M) in all three datasets (Expert, Asia-Africa, America Latin). The F_1 score considers both the precision and the recall aspect (see Table VII). There is a difference in approach four compared to the other approaches: this algorithm does not correctly predict any Yes or No answer options, but only predicts all possible answer options. This can be seen as a minus when applied to these models to identify user behavior in complex contexts.

Table IX shows the comparison of the four prediction models for the test dataset.

V. RELATED WORK

Privacy problems have always been captivating researchers. To discover potential privacy infringement from browsing the Internet with a mobile phone, Collin Mulliner [23] tracked all HTTP headers sent to web services providers. From this activity, he could estimate the amount of covertly leaked personal information. Threats that come from unsecured applications have also been meticulously summarized by Jain et al. in their paper [24].

There have been multiple papers introducing various approaches to adjust privacy preferences dynamically. These proposed approaches do not only apply to applications but also to a wide range of other cases.

By introducing a Context-aware Privacy Policy Language (CPPL), Behrooz et al. [25] aimed to minimize the number of privacy policies that need analysis. In their work, the language filters policies that are relevant to the current scenario using context. The expectation of this research is to enhance the user experience of mobile users in general.

To cope with the issue of ever-changing contexts, Alom et al. [26] proposed a context-based privacy management system that utilizes machine learning algorithms. In their system, privacy preferences for a new context are automatically determined based on existing ones. To be specific, the authors

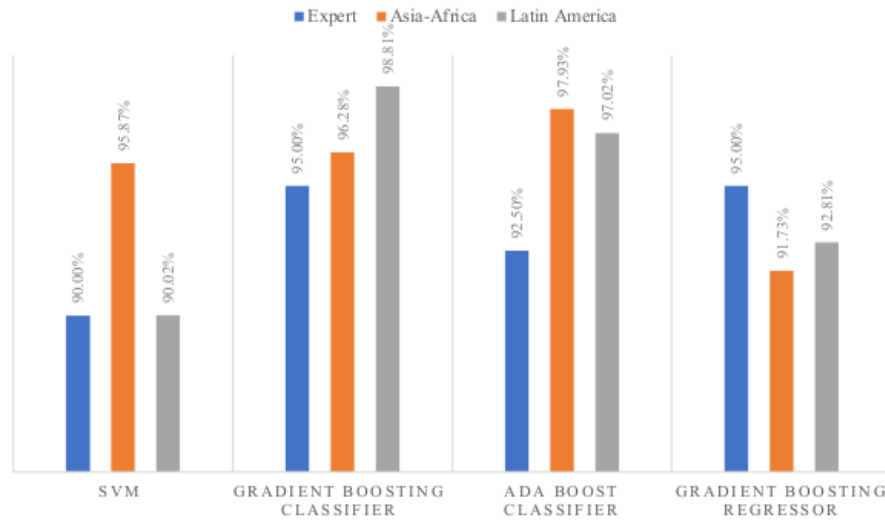


Fig. 4. Satisfaction level of four classification models.

TABLE IX. COMPARISON OF THE FOUR PREDICTION MODELS FOR THE TEST DATASET

		Approach 1			Approach 2			Approach 3			Approach 4		
		Y (%)	N (%)	M (%)	Y (%)	N (%)	M (%)	Y (%)	N (%)	M (%)	Y (%)	N (%)	M (%)
Expert-based participants (N=20)	Precision	75.86%	55.10%	72.73%	88.46%	63.64%	66.67%	90.91%	41.18%	70.37%	NaN	62.99%	NaN
	Recall	62.86%	77.14%	53.33%	67.65%	75.68%	68.97%	60.61%	72.41%	50.00%	0.00%	100.00%	0.00%
	F1	68.75%	64.29%	61.54%	76.67%	69.14%	67.80%	72.73%	52.50%	58.46%	NaN	76.54%	NaN
Crowd-based in Latin American participants (N=85)	Precision	85.00%	75.63%	87.65%	90.75%	73.17%	80.00%	86.29%	78.00%	75.00%	NaN	78.57%	NaN
	Recall	92.12%	83.33%	65.14%	92.37%	83.33%	62.92%	95.54%	75.73%	58.06%	0.00%	100.00%	0.00%
	F1	88.42%	79.30%	74.74%	91.56%	77.92%	70.44%	90.68%	76.85%	65.46%	NaN	88.00%	NaN
Crowd-based in Asia -Africa participants (N=124)	Precision	89.07%	57.48%	84.47%	87.79%	62.81%	87.63%	90.15%	61.29%	83.53%	NaN	76.20%	NaN
	Recall	93.30%	71.57%	60.00%	93.50%	78.63%	56.29%	96.75%	72.38%	54.20%	0.00%	100.00%	0.00%
	F1	91.13%	63.76%	70.16%	90.56%	69.83%	68.55%	93.33%	66.38%	65.74%	NaN	86.49%	NaN

build a classifier that can detect which users’ preferences may be changed as well as the extent of that change. Therefore, given a new scenario, the classifier can predict users’ choices. In their work, Lin et al. [27] also use machine learning to determine the most appropriate privacy preferences for the users. This is done after the system has created a collection of candidate configurations for a particular user.

Bahirat et al. [28] proposed a data-driven approach to designing privacy-setting profiles for IoT devices. Using scenario-based input, the system generates a collection of default privacy settings for the devices, and it is the responsibility of the users to pick one manually.

Knijnenburg et al. [29] introduced a system that supports privacy decisions by modeling privacy concerns. This approach is also known as user-tailor privacy, in which users are provided with private information and non-invasive controls. However, given the variance in people’s perspectives on privacy, creating a general privacy model can be complicated.

There are also methods of adjusting privacy preferences specifically for smartphones. For instance, by taking contextual information into account, Yuan et al. [30] developed a machine learning-based privacy model for sharing photos. In their work, Sanchez et al. [31] developed a privacy preference recommendation system for personalized fitness apps. In their approach, the first profile users’ traits and data permission preferences with machine learning clustering algorithms before designing

privacy setting recommendation strategies. In their attempts to preserve users’ privacy when using health applications, V. Koufi et al. [32] proposed an access control framework used in PHRManager. PHRManager is an Android app that gives authorized users access to Personal Health Records.

VI. CONCLUSION

This paper has introduced IM2P-Medical, a solution on how to learn users’ privacy preferences and suggest appropriate settings for medical data (i.e., health-monitoring devices and apps scenarios). Specifically, semi-supervised learning can help understand people’s perspectives while requiring fewer data to be explained in great detail. Moreover, a collection of questions for understanding users’ thinking on privacy was also shown. The questions were then distributed to two types of participants (normal and expert).

At the end of the project, the satisfaction of users was gathered. Additionally, four models on how to minimize users’ burdens were explained in the evaluation section. In this section, we also compare semi-supervised learning to prove the effectiveness of our model. The result indicates that semi-supervised learning can potentially conserve users’ privacy.

The paper is the first attempt toward a user-centric model for healthcare systems, so it is extremely urgent to identify future development directions. Specifically, we plan to analyze user behavior to build a set of privacy settings recommen-

dations for new users to apply to the medical system. A blockchain-based solution is a potential option to validate service providers' claims about how much data mining is required. On the other hand, an extensive and in-depth study (e.g., increasing the number of participants, compared with users in developing countries) will also be launched soon.

ACKNOWLEDGEMENT

This work would not have been possible without the Mr. Le Thanh Tuan support in implementation and evaluation process. We also express our sincere gratitude to the students and crowd-workers who joined our survey.

REFERENCES

- [1] H. H. Luong *et al.*, "Feature selection using correlation matrix on metagenomic data with pearson enhancing inflammatory bowel disease prediction," in *International Conference on Artificial Intelligence for Smart Community*. Springer, 2022, pp. 1073–1084.
- [2] H. T. Nguyen *et al.*, "Enhancing inflammatory bowel disease diagnosis performance using chi-squared algorithm on metagenomic data," in *Intelligent Systems and Networks*. Springer, 2022, pp. 669–678.
- [3] S. Lim, T. H. Oh, Y. B. Choi, and T. Lakshman, "Security issues on wireless body area network for remote healthcare monitoring," in *International Conference on Sensor Networks, Ubiquitous, and Trustworthy Computing*. IEEE, 2010, pp. 327–332.
- [4] H. X. Son *et al.*, "Priapp-install: Learning user privacy preferences on mobile apps' installation," in *Information Security Practice and Experience: 17th International Conference*. Springer, 2022, pp. 306–323.
- [5] H. X. Son and E. Chen, "Towards a fine-grained access control mechanism for privacy protection and policy conflict resolution," *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 2, 2019.
- [6] H. X. Son *et al.*, "Toward a privacy protection based on access control model in hybrid cloud for healthcare systems," in *12th International Conference on Computational Intelligence in Security for Information Systems (CISIS 2019)*. Springer, 2019, pp. 77–86.
- [7] H. X. Son and N. M. Hoang, "A novel attribute-based access control system for fine-grained privacy protection," in *Proceedings of the 3rd International Conference on Cryptography, Security and Privacy*, 2019, pp. 76–80.
- [8] N. M. Hoang and H. X. Son, "A dynamic solution for fine-grained policy conflict resolution," in *Proceedings of the 3rd International Conference on Cryptography, Security and Privacy*, 2019, pp. 116–120.
- [9] N. Duong-Trung, H. X. Son, H. T. Le, and T. T. Phan, "Smart care: Integrating blockchain technology into the design of patient-centered healthcare systems," in *Proceedings of the 4th International Conference on Cryptography, Security and Privacy*, ser. ICCSP 2020, 2020, p. 105–109.
- [10] —, "On components of a patient-centered healthcare system using smart contract," in *Proceedings of International Conference on Cryptography, Security and Privacy*, 2020, p. 31–35.
- [11] N. Duong-Trung, N. Quynh, T. Tang, and X. Ha, "Interpretation of machine learning models for medical diagnosis," *Advances in Science, Technology and Engineering Systems Journal*, vol. 5, no. 5, pp. 469–477, 2020.
- [12] H. X. Son, T. H. Le, N. T. T. Quynh, H. N. D. Huy, N. Duong-Trung, and H. H. Luong, "Toward a blockchain-based technology in dealing with emergencies in patient-centered healthcare systems," in *International Conference on Mobile, Secure, and Programmable Networking*. Springer, 2020, pp. 44–56.
- [13] H. T. Le *et al.*, "Patient-chain: Patient-centered healthcare system a blockchain-based technology in dealing with emergencies," in *International Conference on Parallel and Distributed Computing: Applications and Technologies*. Springer, 2022, pp. 576–583.
- [14] N. T. T. Quynh, H. X. Son, T. H. Le, H. N. D. Huy, K. H. Vo, H. H. Luong, K. N. H. Tuan, T. D. Anh, N. Duong-Trung *et al.*, "Toward a design of blood donation management by blockchain technologies," in *International Conference on Computational Science and Its Applications*. Springer, 2021, pp. 78–90.
- [15] H. T. Le, T. T. L. Nguyen, T. A. Nguyen, X. S. Ha, and N. Duong-Trung, "Bloodchain: A blood donation network managed by blockchain technologies," *Network*, vol. 2, no. 1, pp. 21–35, 2022.
- [16] L. N. T. Thanh, N. N. Phien, H. K. Vo, H. H. Luong, T. D. Anh, K. N. H. Tuan, H. X. Son *et al.*, "Ioht-mba: an internet of healthcare things (ioht) platform based on microservice and brokerless architecture," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 7, 2021.
- [17] L. T. T. Nguyen *et al.*, "Bmdd: a novel approach for iot platform (broker-less and microservice architecture, decentralized identity, and dynamic transmission messages)," *PeerJ Computer Science*, vol. 8, p. e950, 2022.
- [18] H. X. Son *et al.*, "A risk estimation mechanism for android apps based on hybrid analysis," *Data Science and Engineering*, vol. 7, no. 3, pp. 242–252, 2022.
- [19] —, "A risk assessment mechanism for android apps," in *International Conference on Smart Internet of Things*. IEEE, 2021, pp. 237–244.
- [20] I. Triguero *et al.*, "Self-labeled techniques for semi-supervised learning: taxonomy, software and empirical study," *Knowledge and Information Systems*, vol. 42, no. 2, pp. 245–284, 2015.
- [21] C. Rosenberg *et al.*, "Semi-supervised self-training of object detection models," 2005.
- [22] I. Dópido, J. Li, P. R. Marpu, A. Plaza, J. M. B. Dias, and J. A. Benediktsson, "Semisupervised self-learning for hyperspectral image classification," *IEEE transactions on geoscience and remote sensing*, vol. 51, no. 7, pp. 4032–4044, 2013.
- [23] C. Mulliner, "Privacy leaks in mobile phone internet access," in *2010 14th International Conference on Intelligence in Next Generation Networks*. IEEE, 2010, pp. 1–6.
- [24] A. K. Jain and othes, "Addressing security and privacy risks in mobile applications," *IT Professional*, vol. 14, no. 5, pp. 28–33, 2012.
- [25] A. Behrooz and A. Devlic, "A context-aware privacy policy language for controlling access to context information of mobile users," in *International Conference on Security and Privacy in Mobile Information and Communication Systems*. Springer, 2011, pp. 25–39.
- [26] M. Z. Alom *et al.*, "Helping users managing context-based privacy preferences," in *2019 IEEE International Conference on Services Computing (SCC)*. IEEE, 2019, pp. 100–107.
- [27] J. Lin, B. Liu, N. Sadeh, and J. I. Hong, "Modeling users' mobile app privacy preferences: Restoring usability in a sea of permission settings," in *10th Symposium On Usable Privacy and Security (SOUPS) 2014*, 2014, pp. 199–212.
- [28] P. Bahirat, Y. He, A. Menon, and B. Knijnenburg, "A data-driven approach to developing iot privacy-setting interfaces," in *23rd International Conference on Intelligent User Interfaces*, 2018, pp. 165–176.
- [29] B. P. Knijnenburg, "Privacy? i can't even! making a case for user-tailored privacy," *IEEE Security & Privacy*, vol. 15, no. 4, pp. 62–67, 2017.
- [30] L. Yuan *et al.*, "Context-dependent privacy-aware photo sharing based on machine learning," in *IFIP International Conference on ICT Systems Security and Privacy Protection*. Springer, 2017, pp. 93–107.
- [31] O. R. Sanchez, I. Torre, Y. He, and B. P. Knijnenburg, "A recommendation approach for user privacy preferences in the fitness domain," *User Modeling and User-Adapted Interaction*, pp. 1–53, 2019.
- [32] V. Koufi *et al.*, "Privacy-preserving mobile access to personal health records through google's android," in *2014 MOBIHEALTH*. IEEE, 2014, pp. 347–347.

Anomaly Discover: A New Community-based Approach for Detecting Anomalies in Social Networks

Hedia Zardi, Hajar Alrajhi

Department of Computer Science-College of Computer, Qassim University, Buraydah 51452, Saudi Arabia

Abstract—In this paper, a new method called Anomaly Discover is provided for detecting anomalies in communities with mixed attributes (binary, numerical and categorical attributes). Our strategy tries to identify unusual users in Online Social Networks (OSN) communities and score them according to how far they deviate from typical users. Our ranking is based on both users' attributes and network structure. Moreover, for effective anomaly detection, the context-selection process is performed for choosing relevant attributes that demonstrate a strong contrast between normal and abnormal users. So the anomaly score is defined as the degree of divergence in the network structure as well as a context-specific subset of attributes. To assess the efficacy of our model, we used real and artificial networks. We then compared the outcomes to those of two state-of-art models. The outcomes show that our model performs well since it outperforms other models and can pick up anomalies that competing models miss.

Keywords—Anomaly detection; community anomaly; anomaly ranking; social networks; relevant attributes

I. INTRODUCTION

In data analysis, anomaly detection (also called outlier detection) is identifying items that raise suspicions by differing from the majority of the data. The rising popularity of social networks has attracted some malicious users who have been abusing them. Because of this, it becomes essential on social networks to identify anomalous users. It aims to find users whose activities deviate significantly from regular users, which affects widely different areas such as the detection of social email senders, and detection of fake accounts.

For analyzing relationships in networks, graph-based anomaly detection (GBAD) approaches were used to detect abnormal patterns. In fact, social networks often have data objects linked to each other, so they can be represented as graph networks. The nodes in graph $G(N, E)$ refer to individuals, whereas the edges refer to relationships. The nodes and connecting edges make up a simple graph. On the other hand, nodes and/or edges with related features such as work status, individual ages, type of interaction, and duration, make up the attributed graph(also known as the labeled graph). The attributed networks combine a topological structure with a rich set of features.

The anomaly detection methods in plain graphs analyze the interactions between nodes and employ the network structure to extract graph-centric features and quantify the nodes' closeness. However, the anomaly detection methods in attributed graphs employ the graph's structure and the coherence of

attributes as auxiliary information to find patterns and identify anomalies. Unfortunately, few research has considered attributed graphs. In addition, most of these methods focus only on numerical node attributes; but, real-world networks' node attributes are made up of many attribute types. Therefore, the approaches that consider the network structure and users' attributes, depend on the inclusion of all of a given network's attributes. This makes them inadequate for application to today's networks, which feature ever-growing numbers of attributes. Furthermore, the existence of irrelevant attributes in these networks is inevitable and obstructs anomaly detection. Thus, the context selection process of selecting relevant attributes that show a high level of contrast between normal and anomalous users is crucial for effective anomaly detection.

The community-based approaches are well-known graph-based strategies suggested to address the challenge of anomaly detection [1], [2], [3], [4], [5], [6], [7], [8], [9], [10], [11]. These approaches are based on detecting groups of nodes that are densely connected to each other in the graph and to different communities. Actually, anomalous nodes under this setting can be defined as identifying "bridge" nodes that do not have a direct relationship to a specific community [4].

The aim of this work is to present a new approach to the detection of anomalous users in social networks. Our approach is a community-based approach in attributed graphs which considers both the nodes' attributes and the network structure. A node behavior, which would be considered normal across the entire network, may appear as anomalies within the community context. For this reason, our approach is based on the selected context of relevant attributes. In our approach's first step, we start with the detection of the different communities in the network. In the second step, the nodes will be ranked based on the network structure and attributes similarity. The ranking scores given to the nodes represent the anomaly degree for each node, where a high score indicates an anomalous node, and a low score indicates a normal node.

This paper is organized as follows. In Section II, a summary of the state-of-the-art that reviews the existing works in the field of anomaly detection in a social network is provided. Section III is dedicated to the details of our proposed method for anomaly detection. In Sections IV and V, the performance of our approach is examined and the results obtained by our experiments are presented. A summary and our recommendations for future works are provided in Section VI.

II. RELATED WORKS

Due to the great interest in the discovery of anomalies in recent times, a variety of algorithms have resulted. We will discuss some of the suggested methods in this section. These methods can be divided into two groups according to the basic idea of the approach, structure-based approaches and community-based approaches for unlabeled (plain) and labeled (attributed) graphs.

A. Structure-based Approaches

The concept behind structure-based approaches [12], [13], [14] is that to inspect both normal and anomalous nodes' characteristics, the structural properties are checked. To find abnormal nodes, specific graph matrices are calculated. Akoglu et al. [12] proposed a feature-based model called OddBall. This method finds patterns that the majority of the graph's egonets follow in relation to the egonet-based features it extracts. The 1-step neighborhood surrounding a node, which includes the node, is referred to as a node's egonet. The main eigenvalue of the weighted adjacency matrix of the egonet, the weight of the egonet, the number of each node's neighbours, and the number of egonet's edges are all examples of egonet features. These features are then analyzed for each egonet, and based on how much they deviate from a particular pattern, an outliers score is given. In [13], the authors use the same previous approach, but with a different metric called a brokerage, which is defined as how many times a node bridges a connection between two other nodes when there is no direct link between them. In [14], the neural network model is used to construct graph-centric attributes to identify the nodes as abnormal or normal. Degree centrality and closeness centrality, in addition to betweenness centrality, were used singly and in combination to improve the model's accuracy. Anomalous nodes have a greater degree of proximity centrality, and betweenness centrality.

B. Community-based Approaches

Community approaches are based on the concept of finding groups or communities of densely connected nodes in a network and then detecting the abnormal nodes within these communities. Anomalous nodes are described as nodes that do not show the same characteristics compared to other nodes belonging to the same community or that cannot be assigned to any community [3]. So, in the first stage, Community-based approaches start with the community detection process that groups the nodes into groups that contain dense relationships inside those groups and a few connections between other groups (see Fig. 1).

The second stage consists of detecting community anomalies by finding the nodes that do not deserve to be in this community. An anomaly is a node that has different properties compared to the members of its community and that is not very connected to them. Some approaches identify the nodes as either anomalous users or normal users (see Fig. 2). Some other approaches give a score to each user that determines the degree of its abnormality.

Community-based approaches are more effective at spotting anomalies in attributed graphs. In these approaches, the context of the node is specified by the community because it should share qualities with other nodes within its community.

A node is considered abnormal when it deviates from these typical characteristics. Community-based approaches can be divided into approaches for plain graphs and approaches for attributed graphs:

- Community-based approaches in plain graphs [4], [5] only depend on the structural information of the network. They look at how nodes are related and use the network structure to extract useful information.
- Detecting community anomalies in attributed graphs requires considering both the nodes' attributes and the network structure [9], [10], [11], [1]. Some approaches use the entire attributes space, which is a disadvantage because they are subjected to the curse of dimensionality, which comes with a slew of issues, including longer runtime. Some other approaches select a set of relevant attributes (called context) to filter the full attributes space and select only relevant ones that show a high level of contrast between normal and anomalous users.

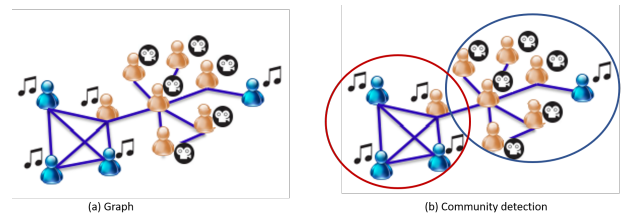


Fig. 1. Community detection in the graph.

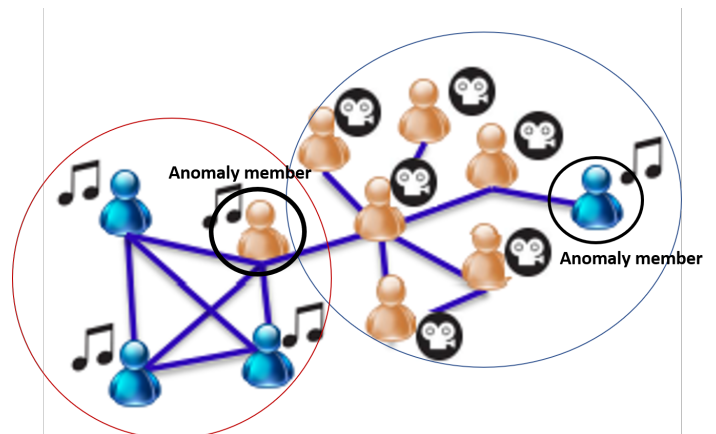


Fig. 2. Community-based anomaly detection in the graph.

III. "ANOMALY DISCOVER" APPROACH

A. Basic Idea

In this work, we present the "Anomaly Discover" approach which is a community-based method for detecting anomalous nodes in social networks. These community anomalies are the nodes that are entrenched within certain graph communities and have deviating attribute values, making them undetectable by global techniques that look for deviation throughout the whole graph node range. As a result, we concentrate on

anomalous nodes that differ from their peers in terms of the structure of the graph and the attributes of the node. While some nodes are strongly connected to their communities, the attribute values can vary significantly amongst nodes, for that, we examine these two factors. By introducing context selection, we concentrate on a subset of important attributes. Context selection reduces the issue of unrelated attributes spreading the full-attributes space and concealing anomalous nodes. Context selection is another technique used to reduce the algorithm's time complexity. There are two phases in our methodology. Firstly, the network is split into communities. A community is a collection of users (that is nodes) who are closely connected and who have similar attributes. So, nodes that lack these features should be regarded as abnormal. So this phase aids in the detection of aberrant nodes that render the community definition invalid. We adopt a modularity-based strategy to split the network since it is fast and can handle enormous networks. In the second phase, for each node, the anomaly score is calculated. By considering both attribute and structure information, this score is utilized to rank nodes depending on the degree of their abnormality. For that, the abnormal nodes are given higher rankings than the usual nodes, which are given lower rankings. Two components make up our score: (i) a structure-based score that considers the graph structure and (ii) an attribute-based score that considers attribute data.

B. Approach Description

In our model, the first phase involves obtaining the communities, and the second phase involves associating the nodes with the anomaly scores. We outline each phase of our method in this section.

1) *First phase: Community detection:* We divide the graph into communities in the first phase to identify the most linked nodes. In our model, we apply the Louvain Algorithm [15], which is a well-known modularity-based strategy for detecting communities in graphs. We employ the Louvain algorithm because of its speed and efficacy. This approach initially allocates each node to a distinct community in order to maximize modularity gain, and then iteratively moves each node to its neighbor. It comes to a halt when further modularity gains are no longer attainable. Gains in modularity are calculated as follows:

$$\delta Q = \left[\frac{\sum_{in} + 2k_{i,in}}{2m} - \left(\frac{\sum_{tot} + k_i^2}{2m} \right) \right] - \left[\frac{\sum_{in}}{2m} - \left(\frac{\sum_{tot}}{2m} \right)^2 - \left(\frac{k_i}{2m} \right)^2 \right] \quad (1)$$

where \sum_{in} represents the sum weights of the edges inside C , \sum_{tot} represents the sum of weights of the edges incident to nodes in C , k_i represents the sum of weights of the edges incident to node i , $k_{i,in}$ represents the sum of weights of the edges from i to other nodes in C and the sum of the weights of all the network's edges are represented by m . Then, when this value is determined for all communities i is linked to, the community that resulted in the greatest modularity increase is assigned to i . In the absence of an increase, i stays inside its original community. All nodes are subjected to this process regularly and sequentially until no more increases in modularity are possible. Once this local modularity maximum is attained, the first stage is finished.

In the second step of the algorithm, each node is combined into a single community, and a new network made up of nodes from the first phase's communities is created. Next, the first phase is applied once more to the new network.

2) *Second phase: Anomaly scoring:* In this step, we characterize the anomalous nodes within the communities using the attribute and structural data. As a result, the node's anomaly score must take into account its deviation within the necessary attributes as well as its community relationships. This score introduces new issues, as it requires combining data from both of the components described in this section: the variance in relevant attribute values of and community connections. The attribute-based score and the structure-based score, that are explained in the following subsections, make up our anomalous score.

Structure-based score: In structure-based scoring, we use a similarity measure to examine the node's relevance. We also look at how connected a node is to the members of its community. Within a community anomaly nodes are consequently less comparable, connected, and influential. In this work, the Jaccard similarity measure and node connectivity are used to calculate the structure-based score, as follows:

$$StrAnomaly(v) = 1 - Jacc(v) * con(v) \quad (2)$$

Where $Jacc(v)$ is the Jaccard index of node v and $con(v)$ is the connectivity of node v . A metric known as node connectivity is used to evaluate how strongly connected a node is to its surrounding neighbourhood. A node produces a high value when its connection to its community is higher than the average connection for the community. The following is a description of the node connectivity:

$$con(v) = \frac{nb(v)}{\frac{\sum_{j=1}^C nb(v_j)}{|C|}} \quad (3)$$

where nb represents neighbors of the node within the community, and $|C|$ represents the nodes' number within the community. The node with high community connectivity has a high score $con(v)$ and vice versa. Anomaly nodes have a low connectivity score compared to normal nodes, which are high in connectivity.

The Jaccard Similarity measure is used to measure the similarity of the node to its community in a graph and hence similar and dissimilar nodes can be detected. This similarity index is used to identify anomalies nodes in the community. The Jaccard index for each node in a community is the sum of the edges the node has with other nodes in the similar community and all other nodes in the graph. In other words, the intersection of the number of node neighbors in the graph over the union of the number of nodes in the similar community. The following is how the Jaccard index is calculated:

$$Jacc(v) = \frac{|ng(v) \cap cv(v)|}{|ng(v) \cup cv(v)|} \quad (4)$$

Where $ng(v)$ represents the neighbors of the node v in the graph, and $cn(v)$ represents all nodes in the same community with node v . The value of the Jaccard index should range in: $0Jacc(v)1$. A Jaccard index of 1 means the node is fully

connected to other nodes in the similar community and vice versa.

Attribute-based score: The objective of the attribute-based score is to quantify the difference in a node's attributes from those of other nodes within the similar community considering subset of related attributes that are chosen in accordance with the context. Considering that many real-world networks are heterogeneous, this attribute similarity function facilitates the combination of attributes of various types. We employ Yi Yang's original Unsupervised Discriminative Feature Selection (UDFS) algorithm to identify relevant attributes [16]. Its goal is to select the data representation features with the highest discriminating. The algorithm optimises the features and produces a ranking and weighting of the features. Additionally, to filter the entire space of attributes, they are ranked according to their UDFS scores as follows:

$$\min_{W^T W=1} Tr(W^T M W) + \gamma \|W\|_{2,1} \quad (5)$$

where ω^i is denoted by the i -th row of W , i.e., $W = [\omega^1, \dots, \omega^d]^T$, the objective function represented in eq. 5 can also be expressed as:

$$\min_{W^T W=1} Tr(W^T M W) + \gamma \sum_{i=1}^d \|\omega^i\|_2 \quad (6)$$

As a result, a new representation of χ_i employing only a select few features is provided for a datum $\chi_i, \chi_i = W^T x_i$. As an alternative, we can rank each feature $f_i |_{(i=1)}^d$ based on $\|\omega^i\|_2$ in descending order and choose the features with the highest rankings. As a consequence, we'll select a subset of the most crucial attributes with size N , where N is a parameter that our algorithm takes into account as input. Given the context (relevant attributes), we can estimate the attribute-based score, which can be defined as the average of all the scores for the relevant attributes. This attribute-based score is defined as:

$$AttrAnomaly(v) = \frac{\sum_{a_r \in A} S(v, a_r)}{n}, \forall a_r \in A \quad (7)$$

where $S(v, a_r)$ represents the attribute score of node v for the attribute $a_r, \forall a_r \in A$, and in the set A of relevant attributes the number of attributes is represented by n , which is define as:

$$S(v, a_r) = \frac{\sum_{j=1}^C d(v, v_j)}{|C|}, \forall v_j \in C \quad (8)$$

where v_j stands for the other nodes within the community, the nodes' number in the community is denoted by $|C|$, and the distance between the nodes v and v_j is represented by $d(v, v_j)$, which is assigned either to zero or one. When $d(v, v_j)$ is zero, this means that the distance between the these nodes is equal to or less than the mean distance (Md), otherwise, it is set at one.

$$d(v_j, v_j) = \begin{cases} 0 & \text{if } |a_r(v_i) - a_r(v_j)| \leq Md(C_{(v)}, a_r) \\ 1 & \text{Otherwise.} \end{cases}$$

where $a_r(v_i)$ represents the the attribute a_r 's value in the node v_i attribute vector, and the mean distance of attribute a_r of node v is denoted by $Md(C_{(v)}, a_r)$ within the community which is illustrated as follows:

$$Md(C_v, a_r) = \frac{\sum_{i,j=1}^C (a_r(v_i) - a_r(v_j))}{p} \quad (9)$$

where the distance between the node v_i and the node v_j for the attribute a_r is represented by $(a_r(v_i) - a_r(v_j))$, and in each community the number of node pairs is represented by p . This distance for binary attributes is determined by the *simple matching coefficient* between the nodes v_i and v_j for the a_r attribute:

$$(a_r(v_i) - a_r(v_j)) = \frac{\sum_{k=1}^d (a_{rk}(v_i) \wedge a_{rk}(v_j)) \vee (\neg a_{rk}(v_i) \wedge \neg a_{rk}(v_j))}{d}$$

The *Jaccard similarity index* between the "1-of-N" binary encodings of $(a_r(v_i)$ and $a_r(v_j))$ gives the distance for categorical attributes; in other words:

$$(a_r(v_i) - a_r(v_j)) = \frac{\sum_{k=1}^d a_{rk}(v_i) \wedge a_{rk}(v_j)}{\sum_{k=1}^d a_{rk}(v_i) \vee a_{rk}(v_j)}$$

The *Euclidean distance* between $(a_r(v_i)$ and $a_r(v_j))$ is what determines this distance for numeric characteristics; that is,

$$(a_r(v_i) - a_r(v_j)) = \frac{1}{1 + \sqrt{\sum_{k=1}^d (a_{rk}(v_i) \wedge a_{rk}(v_j))^2}}$$

where d is the attribute's a_r dimensions, $(a_{rk}(v_i))$ is the value of the attribute's a_r k -th coordinate for node v_i , and \neg , \wedge and \vee are the logical operators for NOT, AND, and OR, accordingly.

Anomaly score:: To rank nodes in every community and recognize anomalous nodes, we next use both the structure-based and attribute-based scores to get an aggregate anomaly score (those with a higher anomaly score). Calculate each node v anomaly score as below:

$$AnomalyScore(v) = StrAnomaly(v) + (1-\alpha)AttAnomaly(v) \quad (10)$$

where the weight used to regulate the relative importance of attribute-based anomaly and the structure-based anomaly is represented by α .

C. Algorithm

First, we use the Louvain algorithm to partition the graph into communities (line 1). Then, we determine each node's Jccard similarity (line 3). Before assigning the structure-based score, we iterate over each node to determine its connectedness(line 6). Afterward, we determine each attribute's UDFS score (line 8), and we choose the N attributes corresponds to lowest score of UDFS to be the relevant attributes (line 9). Then, for each community, we iterate, comparing the distance

between the node and its community nodes with the community mean distance using the relevant attributes (lines 10-16). Finally, the anomalous score of the nodes from G is returned (line 20) by combining the structure-based score and attributes-based score (lines 17-19). The name (AnomalyDiscover), an integration of the words community and anomaly, refers to our algorithm for community-based anomaly detection.

Algorithm 1 Anomaly Discover

Require: $G : (V, E)$, A : Attributes, N : number of select attributes

Ensure: Ranking of all $v \in V$

```
1:  $C \leftarrow$  Louvain Method( $G$ )
2: Initialize empty vectors  $StrAnomaly$ ,  $AttrAnomaly$ ,
    $AnomalyScore$  for all  $v \in V$ 
3:  $Jacc \leftarrow$  Jaccard similarity index for all  $v \in G$  (Eq. 4)
4: For each  $v \in G$  do
5:   Compute connectivity( $v$ ) (Eq. 3)
6:   Compute  $StrAnomaly(v)$  (Eq. 2)
7: End For
8: For all  $a_r \in A$  calculate UDFS score (Eq. 6)
9:  $A' \leftarrow$  subset of  $N$  attributes with the lowest UDFS score
10: For each community  $C_k$  in  $C$  do
11:    $Md \leftarrow$  mean values of attributes from  $A'$  in  $C_k$  (Eq.
   9)
12:   For each  $v_i$  in  $C_k$  do
13:      $s_{v_i} \leftarrow$  a dictionary containing for each attribute
        $a_r$  of  $v_i$  its anomaly score
14:     Compute  $AttrAnomaly(v)$  (Eq. 7)
15:   End For
16: End For
17: For each  $v \in G$  do
18:   Calculate the anomaly score using the Eq. 10
19: End For
20: Return  $AnomalyScore$ 
```

D. Complexity Analysis

First, the graph is partitioned into communities by applying the Louvain algorithm, which has a running-time cost $O(v \log v)$ for the number of graph nodes. After that, the Jaccard similarity measure is calculated, and that costs $O(v + e)$, where the number of edges in the graph is represented by e . Next, the $StrAnomaly$ score is calculated with the nodes' number linear cost. The context is then defined using the UDFS score, which has a cost of $O(mv^2)$, where m represents the total attributes' number. Consequently, the computational complexity of the Anomaly Discover model is $O(\max(mv^2, v + e))$. When all of the graph nodes are allocated to one community, which happens when a quadric analysis is carried out for each community, this is the worst-case scenario. As a result, the algorithm performs better on a real network with a large number of communities.

IV. EXPERIMENTAL EVALUATION OF PERFORMANCES

To study the performance of the "Anomaly Discover" method, we compared it with two well-known algorithms which are CODA and ConSub that we briefly introduce in this section:

- CODA [17] is one of the most popular models used for anomaly detection in social network communities. In this model, community detection and anomalous node identification are done in a single step. It utilizes the entire attribute set of nodes.
- ConSub's [7] concept is a statistically-based selection of a subset of all attributes of the nodes. This subset demonstrates dependencies inside the graph structure. To find abnormal nodes in the communities, a subset of attributes is chosen and used with the DistOut distance-based outlier model.

A. Evaluation Measures

In order to evaluate the performance of the community-anomaly detection model and establish its validity on synthetic and real datasets, we compare the acquired nodes' ranks of the model with the ground truth. The Area Under the Curve (AUC) is one of the most important performance metrics for anomaly ranking and classification models. AUC measures how well the model can differentiate between two classes; a greater value of AUC means a more effective model. In a machine learning classification task, comparing the actual classes to the predicted classes of the model. Hence, the results can be categorized into four groups: true positives, false positives, true negatives, and false negatives. True positives are actual anomalies that the model predicted correctly as anomalies, while true negatives are actual normals that the model predicted correctly as normal. On the other hand, false positives are actual normals that the model predicted as an anomaly, while, false negatives are actual anomalies that the model predicted as normal. Specificity is the proportion of correctly identified negatives, whereas sensitivity is the proportion of correctly detected positives. We provide several thresholds in the classification model and to create the ROC curve, sensitivity (also referred to as the true positive rate) is plotted against the false positive rate which is calculated as $(1 - \text{Specificity})$.

So, the optimum model is the one that reliably detects all positives and all negatives at a specified threshold value while still obtaining the highest levels of specificity and sensitivity. The top-left portion of the ROC plot contains the greatest value. The area under the curve (top-left corner) consequently represents the ROC curve's ability to reach the highest level of specificity and sensitivity.

The model's runtime is the second metric considered in this evaluation since it's crucial to see if the model can accurately identify the community abnormality in a timely manner. If a shorter runtime is possible, it is better if the outcomes are high-caliber.

B. Real Benchmark Dataset

The Book network and Disney network serve as our testbeds in this part, and the performance of our model is compared to that of the CODA and ConSub models. The only variable in our suggested model is how many attributes to include. The number of characteristics was set to half-number and 10 attributes maximum; adding more characteristics increased the run-time without notably enhancing the quality.

TABLE I. MODELS PARAMETER SETTING

CODA	Number of communities= 8	Anomaly percentage= 0.05	Link importance= 0.01
ConSub	Size of interval= 10	The number of Monte Carlo iterations= 150	The significance level=0.05

Table I provides information about the other models' parameter settings.

We compare our findings with those of ConSub and CODA in order to assess our methodology using the following real networks:

- **Disney Network:** the Amazon co-purchase network was divided up into a Disney network that exclusively considered Disney DVDs. In the graph, each product is characterized by 30 properties, including review ratings, product prices, and other information. The network has 124 nodes and 334 edges. The network, although being a tiny dataset, is used to test the majority of anomaly detection models because of its intricate graph and attribute structure. The ground truth of whether an object is an abnormality or normal is not available for this real-world dataset. As part of a user experiment to determine the dataset's ground truth, high school students personally classified each object as normal or an abnormality.[18] presents a thorough explanation of the dataset and the user experiment.
- **Book Network:** this network which was based on Amazon Co-Purchase Network, includes books that users have tagged as "amazon fails" [19]. On Amazon, customers could use tags to describe items, and different tags like the "amazon fail" tag was used to indicate dubious products. The network has 1468 nodes, 3695 edges, and 28 attributes that are used to describe each object. The basis for this dataset was established by classifying a book as an anomaly when at least 20 users had labeled it as an "amazon fail".
- **Enron Network:** we employ email transmission as edges between email addresses on the Enron communication network. Spam dataset outliers were defined as addresses that have sent spam. This network contains 13 533 nodes, 176 967 edges. There are 20 attributes present in each node that provide aggregate information about the average number of recipients, the average content length, or the time interval between two mails [19].

C. Synthesis Benchmark Dataset

Evaluating anomaly detection methods is not a straightforward process due to the lack of suitable datasets containing anomalies and the lack of ground truth that defines which data points are actual abnormalities. As a result, performance evaluation is typically the purpose of synthetic datasets. These datasets are utilized to compare a model's performance on synthetic versus real data. Based on [7], synthetic datasets of various attribute counts and sizes are created. To replicate the characteristics of real networks, the graph is created by following a power-law distribution. Relevant attributes obtained values from a Gaussian distribution, while irrelevant attributes

obtained values from a uniform distribution. To ensure there were no abnormal values in the relevant attributes, the tails of each Gaussian distribution were truncated using a hyper ellipsoid (see Fig. 3). Anomaly nodes' characteristic values were modified to be random numbers beyond the boundaries of their communities' hyper ellipsoids. The anomalous nodes number within the communities is determined by the anomaly ratio, which is 10%. Only when at least two pertinent attribute combinations are taken into account can the anomalies be found.

The graphml file and the true file are the two files that make up any synthetic dataset. The graph and each node's properties can be found in the graphml file. The true file includes the actual nodes, with a ground truth value of 0 for normal nodes and 1 for anomalous nodes. We use synthetic datasets that contain 1000 number of nodes and various characteristics 2, 10, 20, 40, 60, and 80 to assess the performance of the model's as the number of attributes increases. To evaluate how well the model performs when the network size is increased, we use synthetic datasets with varying numbers of nodes and ten attributes. We configured our model parameter to be the half-number of attributes to test the impact of increase in attributes a while utilising the same configuration in the real network trials for the other models.

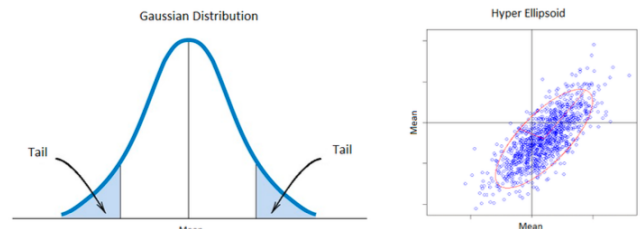


Fig. 3. Gaussian distribution with cutting tails.

V. RESULTS FOR REAL AND SYNTHESIS DATASETS

A. Results for Real Datasets

1) *Disney network:* Fig. 4 and 5 illustrate our model's results applied on the Disney network in contrast to CODA and ConSub. The findings show that the "Anomaly Discover" model produces good-quality outcomes with an AUC of .83 (see Fig. 4). In contrast, the other models produce results of poorer quality, with an AUC of.82 for ConSub and.50 for CODA, respectively. The ConSub model performs better than the CODA model, which produces the lowest-quality outputs. In Fig. 5, the runtime evaluation is displayed. The CODA model comes in second with a runtime of 6.05 seconds, just 0.04 seconds behind the "Anomaly Discover" model.

At 8.93 seconds, the ConSub model is the slowest model. The “Anomaly Discover” model yields the best outcomes for identifying community abnormalities, according to this experiment on the Disney Network, out of the three methods. Among the three, the “Anomaly Discover” model is the fastest. While the ConSub model delivers high-quality findings but operates slowly, the CODA model is recognised as the least effective model for identifying community-anomalies in a real-world network.

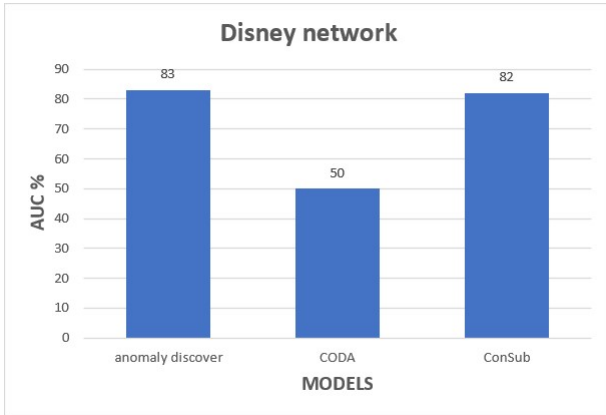


Fig. 4. AUC and ROC curve for the disney network.

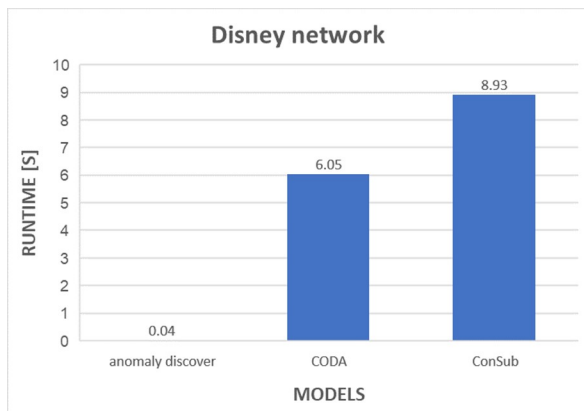


Fig. 5. Evaluation of runtime for disney network.

2) *Book Network*: By computing the AUC, Fig. 6 shows how well the “Anomaly Discover”, CODA, and ConSub models perform. In comparison to the other models, the “Anomaly Discover” model produces least findings (see Fig. 6), whereas ConSub produces the highest-quality results (AUC = 0.60). The “Anomaly Discover” model outperforms the CODA models in the Book network’s runtime evaluation, which shows that it executes in 12.48 seconds and yields the best results. The CODA model is the slowest at 36 seconds (see Fig. 7).

3) *Enron network*: The outcomes of our model on the Enron network in comparison to CODA and ConSub are shown in Fig. 8 and 9. The figures demonstrate that our model produces good-quality outcomes with an AUC of .78 (see Fig. 8), but it was the slowest (see Fig. 9). In contrast, the CODA produces a result that has lower quality with an AUC .46 but it was the fastest. ConSub produces a result inferior to the “Anomaly Discover” model, but it was faster than our model.

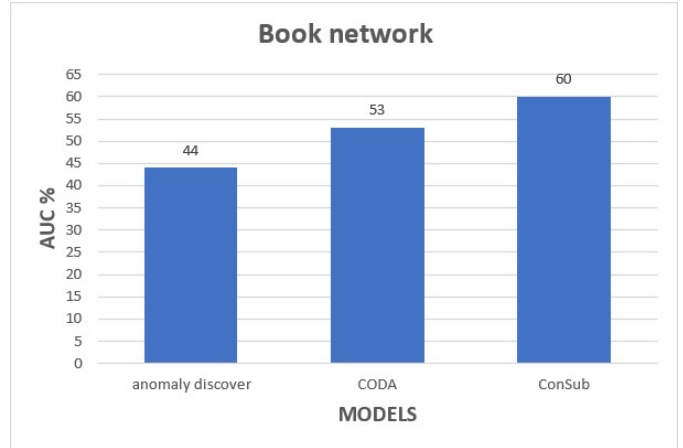


Fig. 6. AUC and ROC curve for the book network.

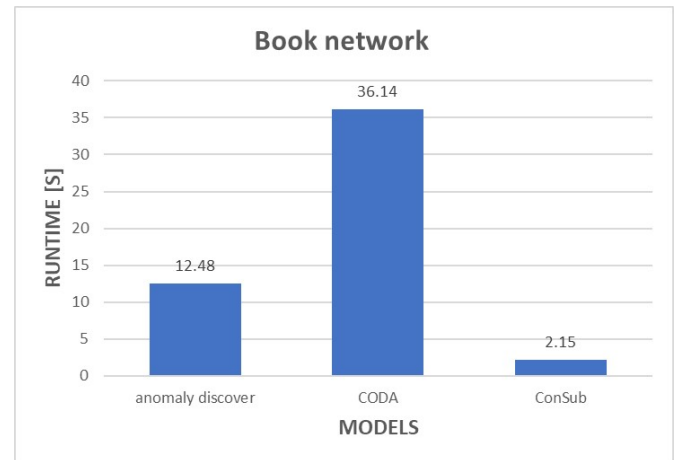


Fig. 7. Evaluation of runtime for book network.

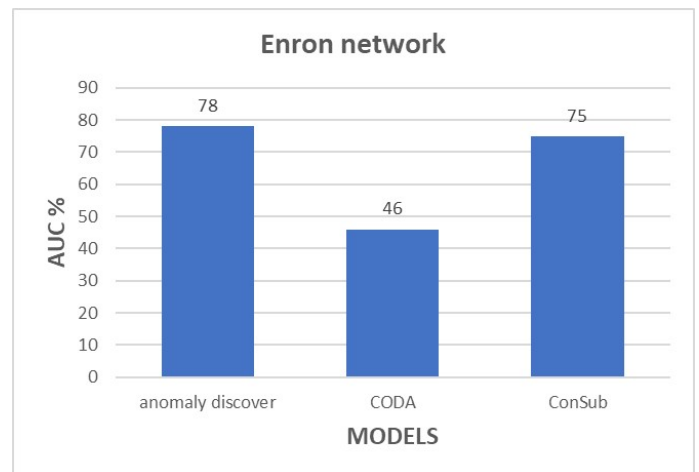


Fig. 8. AUC and ROC curve for the enron network.

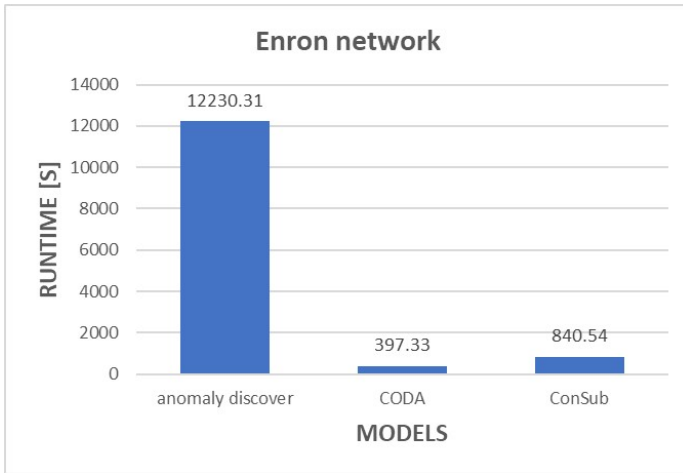


Fig. 9. Evaluation of runtime for enron network.

B. Results for Synthesis Datasets

1) *First experiment:* We evaluate our model performance as we add more attributes using the AUC curve with datasets of 1000 numbers of nodes. The variance in AUC of the tested models as the number of features rises is depicted in Fig. 10. In comparison to ConSub and CODA, our model results in the best AUC, which ranges from 0.88 to 0.98. Fig. 11 illustrates the runtime evaluation by showing the runtimes with increasing numbers of attributes. In spite of the fact that CODA runtime often grows as the number of characteristics does as well, “Anomaly Discover” and ConSub both provide the best scalability in this regard. It’s important to keep in mind that matrix operations are costly, and with CODA, they are performed for each attribute, increasing the runtime.

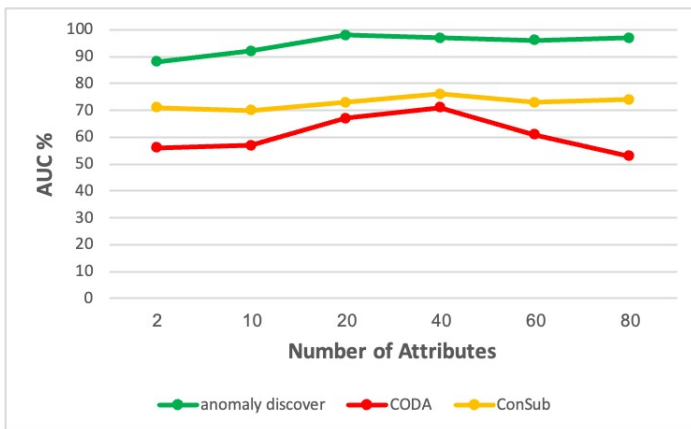


Fig. 10. Variations in AUC using different number of attributes for each tested models.

2) *Second experiment:* Networks with 500, 1000, 2000, 3500, 6000, and 10000 nodes are utilised to evaluate the model with a larger network. In the smallest network, the AUC of the “Anomaly Discover” model is 0.93, and in the largest network, it is 0.90. In fact, when compared to the other models, “Anomaly Discover” has the greatest AUC (see Fig. 12). Fig. 13 displays the evaluation of the runtime as the network size increases. ConSub has overall faster runtimes than the other

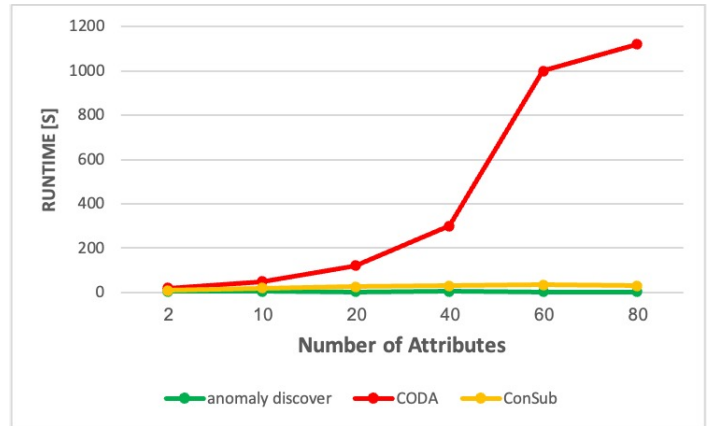


Fig. 11. Variations in runtime using different number of attributes for each tested models.

models, although the “Anomaly Discover” outperforms CODA in networks with 500, 1000, 2000, and 3500 nodes, while CODA outperforms “Anomaly Discover” in networks with 6000 and 10000 nodes.

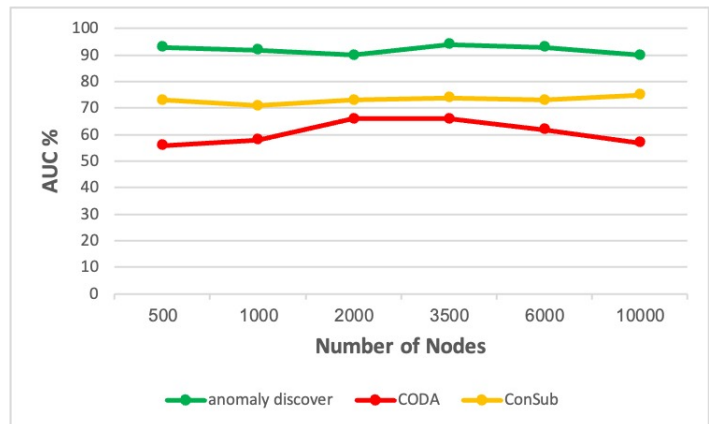


Fig. 12. Variations in AUC using the different number of nodes for each tested models.

C. Discussion

The outcomes of these tests show how well “Anomaly Discover” works to identify community anomalies in both real and synthetic networks. In fact, we’ve achieved things that are extremely intriguing, which earlier approaches like ConSub [7] couldn’t. As the model defines the pertinent network properties rather than taking into account the entire attribute space, it increases the number of attributes while still producing high-quality results and scalability. As a result, the model is appropriate for applications used today, when the number of attributes is increasing. Since the ConSub model also describes the network context, whereas CODA [17] simply considers the network attributes, it performs better than CODA in terms of performance.

VI. CONCLUSION

In this study, we focused on finding anomalous users in online networks. In particular, we are seeking to identify

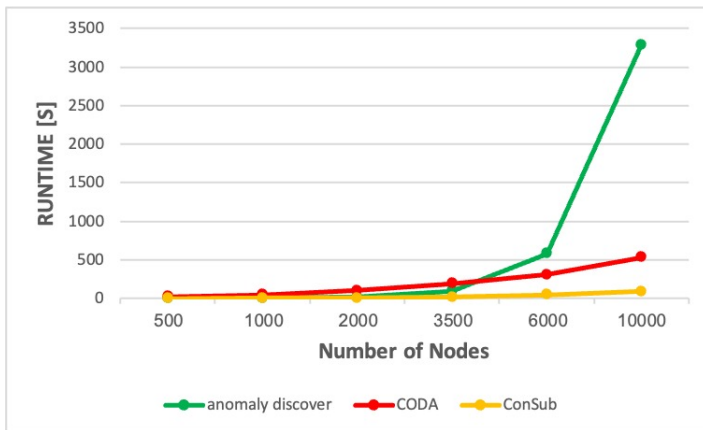


Fig. 13. Variations in runtime using different number of nodes for each tested models.

anomalies that diverge from their communities in comparison to normal users who frequently share numerous attributes with their members of the community. In fact, by using both structure information and attributes information, we were able to create an anomalous ranking score to efficiently find complicated anomalies that differed in either their characteristics values or their structure, or both. To highlight any deviation in these values, the context is selected, which is a subset of the relevant attributes. Given that many real-world networks are heterogeneous, this approach enables the combining of the attributes of mixed types.

We next go over different ways that an extension of our suggested approach could be implemented. Other features of online social networks, such as user communication through comments or message exchange, could be evaluated to find anomalies, though. While they might have common attribute values and structural characteristics with their community, these data might signify an unexpected communication pattern, making them a useful indicator of anomalous nodes.

REFERENCES

- [1] S. A. Moosavi, M. Jalali, N. Misaghian, S. Shamshirband, and M. H. Anisi, "Community detection in social networks using user frequent pattern mining," *Knowl. Inf. Syst.*, vol. 51, no. 1, p. 159–186, apr 2017. [Online]. Available: <https://doi.org/10.1007/s10115-016-0970-8>
- [2] M. Bouguessa, *A Model-Based Approach for Mining Anomalous Nodes in Networks*, 01 2020, pp. 213–237.
- [3] L. Akoglu, H. Tong, and D. Koutra, "Graph-based anomaly detection and description: A survey," *CoRR*, vol. abs/1404.4679, 2014. [Online]. Available: <http://arxiv.org/abs/1404.4679>
- [4] H. N. Win and K. T. Lynn, "Community detection in facebook with outlier recognition," in *2017 18th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD)*, 2017, pp. 155–159.

- [5] M. Bouguessa, *A Model-Based Approach for Mining Anomalous Nodes in Networks*, 01 2020, pp. 213–237.
- [6] E. Müller, P. I. Sánchez, Y. Mülle, and K. Böhm, "Ranking outlier nodes in subspaces of attributed graphs," in *2013 IEEE 29th International Conference on Data Engineering Workshops (ICDEW)*, 2013, pp. 216–222.
- [7] P. I. Sánchez, E. Müller, F. Laforet, F. Keller, and K. Böhm, "Statistical selection of congruent subspaces for mining attributed graphs," in *2013 IEEE 13th International Conference on Data Mining*, 2013, pp. 647–656.
- [8] P. I. Sánchez, E. Müller, O. Irmeler, and K. Böhm, "Local context selection for outlier ranking in graphs with multiple numeric node attributes," in *Proceedings of the 26th International Conference on Scientific and Statistical Database Management*, ser. SSDBM '14. New York, NY, USA: Association for Computing Machinery, 2014. [Online]. Available: <https://doi.org/10.1145/2618243.2618266>
- [9] S. Mekouar, N. Zrira, and E. H. Bouyakhf, "Community outlier detection in social networks based on graph matching," *International Journal of Autonomous and Adaptive Communications Systems*, vol. 11, p. 209, 01 2018.
- [10] G. V. Daniel and M. Venkatesan, "Robust graph based deep anomaly detection on attributed networks," in *2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, 2021, pp. 1029–1033.
- [11] Y. Wang and Y. Li, "Outlier detection based on weighted neighbourhood information network for mixed-valued datasets," *Information Sciences*, vol. 564, pp. 396–415, 2021. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0020025521001870>
- [12] L. Akoglu, M. McGlohon, and C. Faloutsos, "oddball: Spotting anomalies in weighted graphs," in *Advances in Knowledge Discovery and Data Mining*, M. J. Zaki, J. X. Yu, B. Ravindran, and V. Pudi, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 410–421.
- [13] R. Kaur and S. Singh, "A comparative analysis of structural graph metrics to identify anomalies in online social networks," *Computers & Electrical Engineering*, vol. 57, pp. 294–310, 2017. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0045790616307959>
- [14] A. Chaudhary, H. Mittal, and A. Arora, "Anomaly detection using graph neural networks," in *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)*, 2019, pp. 346–350.
- [15] V. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *Journal of Statistical Mechanics Theory and Experiment*, vol. 2008, 04 2008.
- [16] Y. Yang, H. T. Shen, Z. Ma, Z. Huang, and X. Zhou, "L2,1-norm regularized discriminative feature selection for unsupervised learning," in *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence - Volume Volume Two*, ser. IJCAI'11. AAAI Press, 2011, p. 1589–1594.
- [17] J. Gao, F. Liang, W. Fan, C. Wang, Y. Sun, and J. Han, "On community outliers and their efficient detection in information networks," in *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '10. New York, NY, USA: Association for Computing Machinery, 2010, p. 813–822. [Online]. Available: <https://doi.org/10.1145/1835804.1835907>
- [18] J. Z. Huang, "Extensions to the k-means algorithm for clustering large data sets with categorical values," *Data Mining and Knowledge Discovery*, vol. 2, pp. 283–304, 1998.
- [19] P. Iglesias Sánchez, "Context selection on attributed graphs for outlier and community detection," Ph.D. dissertation, 2015.

A Particle Swarm Optimization with Imbalance Initialization and Task Rescheduling for Task Offloading in Device-Edge-Cloud Computing

Hui Fu, Guangyuan Li, Fang Han*, Bo Wang

Faculty of Engineering, Huanghe Science and Technology College, Zhengzhou, China 450006

Abstract—Smart devices, e.g., smart-phones, internet-of-thing device, has been prevalent in our life. How to take full advantage of the limited resources to satisfy as many as requirements of users is still a challenge. Thus, in this paper, we focus on the task offloading problem to address the challenge by device-edge-cloud computing, by PSO improved with the imbalance initialization and the task scheduling. The imbalance initialization is to increase the probability that a task is assigned to a computing node such that the node provides a longer slack time. The task scheduling is to reassign tasks with deadline violations into other nodes, to improve the number of accepted tasks for each offloading solution. Extensive experiment results show that our proposed algorithm has better performance than other ten classical and up-to-data algorithms in both the maximization of the accepted task number, the resource utilization, as well as the processing rate.

Keywords—Cloud computing; edge computing; task offloading; task scheduling; particle swarm optimization

I. INTRODUCTION

In our life, smart devices, e.g., smart-phones, Internet-of-Thing (IoT) devices, are becoming popular, and they are appearing everywhere. As the life quality improves, people request a wide variety of services by their devices. But smart devices generally have limited resource capacity due to their small spaces. Thus, smart devices cannot satisfy all requirements of their users. Edge computing and cloud computing are two of commonly used ways to address this issue. Edge computing places a few servers close to devices, aiming at providing low latency services [1]. Cloud computing extends the capacity of devices by its abundant computing resources, but it usually has poor network performance [2].

Device-edge-cloud computing (DECC) combines benefits of edge and cloud computing, which provides services by not only local device resources but also low latency edge servers and abundant cloud servers [3]. How to make these heterogeneous resources cooperate well is a kind of very challenging work for high service quality and resource efficiency in DECC. Therefore, several works focused on the task offloading or scheduling to address this challenge. The task offloading problem is to decide which and how many resources for each task's processing [3]. The offloading problem has been proofed as NP-hard, because tasks and resources generally are discrete, and the task processing models and the optimization objectives are usually non-convex.

To solve the task offloading problem, existing works mainly exploited two kinds of approaches, heuristics and meta-heuristics, based on their desires. Heuristics use local search strategies to provide local best solutions with a very few overheads. For example, Wang et al. [4] proposed two heuristic algorithms for the offloading problem. The first one is iteratively assigning the task to an edge server (ES), which provides the minimal response time. The second one, load balance, iteratively assigns the task to the ES such that the ES can satisfy the most tasks' requirements. Only when there is no available edge resources, these two algorithms assign tasks to the cloud for their processing. Sang et al. [5] and Chen et al. [6] presented multi-stage heuristic algorithms for task offloading in DECC, which both considers to use abundant cloud resources at first for task processing, to improve the overall accept ratio. Heuristics generally provide solutions with limited performance, because they only exploit local search strategies.

Meta-heuristics exploit global search strategies, inspired by some natural or social laws. Meta-heuristics can provide better solutions than heuristics many times, benefiting from their global search abilities, with a few costs. Such as, Wang et al. [7] and Song et al. [8] applied genetic algorithm (GA) to improve the deadline violation and the delay, respectively. Alqarni et al. [9] and Wang et al. [10] made use of Particle Swarm Optimization (PSO) for the delay minimization and the user satisfaction maximization.

No heuristic or meta-heuristic has the best performance as there is no such thing as a free lunch. Thus, a promising way to achieve better performance is combining two or more heuristics and meta-heuristics for hybrid heuristics. Mahenge et al. [11] proposed a hybrid swarm intelligence offloading algorithm, by using the position update strategies of both PSO and Grey Wolf Optimizer to optimize the energy consumption of devices. To combine benefits of both swarm intelligence and evolutionary algorithm, Wang et al. [13] used GA at first, and then applied PSO with the population provided by GA. Hafsi et al. [12] sequentially employed GA and PSO on each iteration of the population evolution. Nwogbaga et al. [14] performed a mutation operator on each individual at the end of each iteration for PSO, to avoid trapping to local best solution. These works only sequentially performed two meta-heuristics without considering their integration, which provides a low combination efficiency.

Therefore, in this paper, we aim at designing a hybrid heuristic algorithm for providing a good task offloading solution in DECC, by improved PSO. Specifically, we rep-

*Corresponding Authors.

resent task assignment solutions as integer coded positions of particles, and use PSO for the position updates. And by the earliest deadline first (EDF) heuristic algorithm for the task scheduling on each computing node, we can achieve a task offloading from each particle position. To improve the performance of PSO, we propose to use an imbalance idea for the population initialization, where the computing node for a task's processing has a possibility positively associated with its capacity. This can take full use of the easily trapping into local best position, by increasing the probability that the initialized particles are near the global best position. To improve the offloading solution, we reschedule tasks with deadline violations to other computing nodes for each assignment solution, i.e., each particle position. We conduct extensive experiments to evaluate the performance of our proposed algorithm, and the results confirm the superior performance in optimizing the number of tasks with requirement satisfactions and the resource efficiency, compared with ten of classical and up-to-data heuristics and meta-heuristics.

In the following, we state the task offloading problem of DECC in Section II. The improved PSO is illustrated in Section III. Experimental evaluations are presented in Section IV, and we conclude our work in Section V.

II. PROBLEM STATEMENT

In a considered DECC system, there are D devices, E ESs, and C CSs, which can be seen as $D + E + C$ computing nodes (respectively represented as $n_i, 1 \leq i \leq D + E + C$). For node n_i , there are g_i computing capacity. For data transferring, the network transmission speed between n_i and $n_{i'}$ is $b_{i,i'}$. If there is no network connection between n_i and $n_{i'}$, $b_{i,i'} = 0$. And within one node, there is no latency for data transfer, i.e., $b_{i,i} = 1$.

In the DECC system, there are T tasks ($t_j, 1 \leq k \leq T$) launched by these D devices for processing. We use binary constants $l_{i,j}$ to indicate the relationships between tasks and devices, where $l_{i,j} = 1$ means t_j is launched by n_i , and $l_{i,j} = 0$, otherwise. For every task, say t_i , the nodes that can accept its request for processing include the device launching it, ESs having network connections with this device, and CSs. Therefore, we can use $l_{i,j}, 1 \leq i \leq D + E + C, 1 \leq k \leq T$, to indicate whether n_i can be used for processing t_j , where $l_{i,j} = \sum_{i'=1}^D (l_{i',j} \cdot (b_{i',i} \neq 0))$, for $D + 1 \leq i \leq D + E + C, 1 \leq j \leq T$.

For task t_j , the amount of required computing resources, i.e., its computing size, is c_j . The input data amount of t_j is a_j , and the deadline is d_j which means t_j must be finished before d_j . Without loss of generality, we assume $d_1 \leq d_2 \leq \dots \leq d_T$. Then, if task t_j is offloaded into ES or CS n_i , the data transfer latency is

$$\tau_{i,j}^D = \frac{a_j}{\sum_{i'=1}^D (l_{i',j} \cdot b_{i',i})}. \quad (1)$$

where $\sum_{i'=1}^D (l_{i',j} \cdot b_{i',i})$ is the transmission speed between the device launching t_j and the ES or CS n_i . The computing latency of t_j in node n_i is

$$\tau_{i,j}^C = \frac{c_j}{g_i}. \quad (2)$$

In this paper, we ignore the transfer latency of the output data for each task, as the result generally has much less amount than the input [15], [16].

For multiple tasks assigned to one computing node, the data transfers and the computing are processed sequentially. It has been proofed that EDF yields an optimal schedule for minimizing the number of task deadline violations in a computing node [17], which is the major objective in the paper. Therefore, we can deduce the finish time of each task on every node with EDF processing order. Then, the earliest complete time (ft_j^D) of the data transfer for a task can be deduced by Eq. (3). Where $x_{i,j}$ is the binary variable to indicate where t_j is assigned to n_i for its processing (1 is yes, and 0 is no).

$$ft_j^D = \sum_{i=1}^{D+E+C} (x_{i,j} \cdot \sum_{j' \leq i} (x_{i,j'} \cdot \tau_{i,j'}^D)), 1 \leq j \leq T. \quad (3)$$

For a task's computing on a node, it can be started only when its data transfer finishes and the node is available, where the node is available only when the its previous task finishes its computing. Thus, the finish time of a task's computing, i.e., its finish time, can be calculated iteratively by Eq. (4).

$$ft_j = \max\{ft_j^D, \sum_{i=1}^{D+E+C} (x_{i,j} \cdot \max_{j' < j} \{x_{i,j'} \cdot ft_{j'}\})\} + \sum_{i=1}^{D+E+C} (x_{i,j} \cdot \tau_{i,j}^C), 1 \leq j \leq T. \quad (4)$$

Then, based on above formulations, the task offloading problem in the DECC system can be expressed as

$$\text{Maximizing } N = \sum_{i=1}^{D+E+C} \sum_{j=1}^T x_{i,j}, \quad (5)$$

subject to

$$ft_j \leq d_i, 1 \leq j \leq T, \quad (6)$$

$$x_{i,j} \in \{0, 1\}, 1 \leq i \leq D + E + C, 1 \leq j \leq T. \quad (7)$$

Where the objective (5) is maximizing the number of tasks whose deadlines are met, and the constraints mainly include the deadline requirements (Eq. 6) and the atomicity of every task (Eq. 7). In this work, we consider the hard deadline, where if the deadline of a task is satisfied, the task is accepted for its processing, and it is rejected, otherwise. Due to the discrete decision variables ($x_{i,j}$) and the non-convex constraints (see Eq. 4 and 6), the task offloading problem generally is hard to solve. Existing tools, e.g., lpsolve [18] and MathWorks [19], can provide exact solutions, but has exponential complexity at worst. Therefore, in the following section, we present a hybrid heuristic offloading algorithm to solve the problem with a polynomial time.

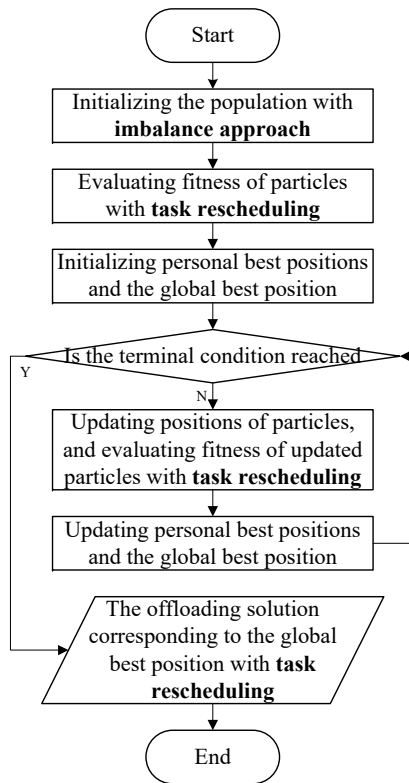


Fig. 1. The flow chart of PSOINRS.

III. THE PSO WITH IMBALANCE INITIALIZATION AND TASK RESCHEDULING

PSO has been applied to solve optimization problems in various fields, due to its easy implementation and good performance. But PSO has some drawbacks which prevent its more wide application. Therefore, in this section, we integrate two improvement strategies on PSO to achieve a better offloading solution for DECC, and proposed a hybrid heuristic offloading algorithm, PSOINRS (PSO with imbalance Initialization and task Rescheduling). The algorithm flow chart is shown in Fig. 1.

In PSOINRS, a position of particles represents a task assignment solution. The dimension number of particles is identical to the task number, and the value in a dimension indicates the node that the corresponding task is assigned to.

At first, for the population initialization of PSO, we propose an imbalance approach to improve the particle density in the possible best positions. Intuitively, a task has more possibility to be processed by the node providing longer slack time (the closeness to the deadline). Therefore, PSOINRS initializes the position of a particle by setting the possibility of a node positively associated with the deadline tightness that the node provides, where the probability that t_j is assigned to n_i is

$$p_{i,j} = \frac{s_{i,j}}{\sum_{j'=1}^T s_{i',j}}, \quad (8)$$

where $s_{i,j} = d_j - ft_j$ if t_j is assigned to n_i and $ft_j \leq d_j$, and $s_{i,j} = 0$, otherwise.

After the population initialization, PSOINRS evaluates the fitness of each particle according to its position. Given a position, we can easily achieve a task assignment solution. By EDF ordering scheme, we achieve a offloading solution from the task assignment solution. Then, we proposed to use the task rescheduling in PSOINRS to improve the solution quality for each position. The task rescheduling is scheduling tasks with deadline violations to other nodes by EDF to improve the optimization objective (5).

Except the imbalance initialization and the task rescheduling for mapping a position into a task offloading solution, PSOINRS updates the positions of particles identical to PSO, as shown in Fig. 1.

IV. PERFORMANCE EVALUATION

For evaluating the performance of our proposed algorithm presented in the previous section, we simulate a DECC environment based on related works and the reality. In a simulated DECC, there are 10 devices, 5 ESs, and 10 CS types. For each device, there are [1.8, 2.5]GHz computing capacity. Each device is randomly connected with an ES. Each ES or CS type has [1.8, 3.0]GHz computing capacity. The data transfer rate of a device to an ES or CS is set as [80, 120]Mbps, and [10, 20]Mbps, respectively. There are 1000 tasks randomly launched by these devices. Each task has [0.5, 1.2] GHz computing size, and [1.5, 6] MB input data. The deadline of a task is set as [1, 5]s.

We compare our proposed algorithm with First Fit (FF), First Fit Decreasing (FFD), EDF, Short Job First (SJF), random (RAND), GA, GAR [20], PSO, PSOM [14], and GAPSO [13]. RAND is randomly initializing a population, and provide the best individual. We use the following performance metrics for the evaluation of each algorithm, the number of tasks with deadline satisfactions, the overall computing resource utilization, the computing rate, and the data processing rate. The rate of computing and data processing is the computing size and the input data amount of tasks with deadline satisfactions.

Fig. 2 shows the number of tasks with deadline satisfactions when applying different offloading algorithms, which is one of the most commonly used metrics for evaluating the user satisfaction or the quality of service (QoS). From this figure, we can see that PSOINRS has 8.73%–36.8% better performance than others. This verifies the performance superiority of our proposed algorithm. The benefits of PSOINRS are mainly the imbalance initialization and the task rescheduling, which will be both evaluated and illustrated in the followings.

Fig. 3 gives the resource utilizations achieved by these offloading algorithms, which is one of the most frequently used metrics for the quantification of the resource efficiency. As shown in the figure, heuristics (FF, FFD, EDF, and SJF) has better utilizations than meta-heuristics (GA, GAR, PSO, PSOM, GAPSO, and PSOINRS). This is mainly because heuristics prioritise processing tasks locally or in low-latency edge resources, while meta-heuristics pursues the global optimization objective for maximizing the number of accepted tasks, because the task processing with a low data transfer latency has a high computing resource utilization, which is mainly decided by the ratio between the data transfer latency and the computing delay.

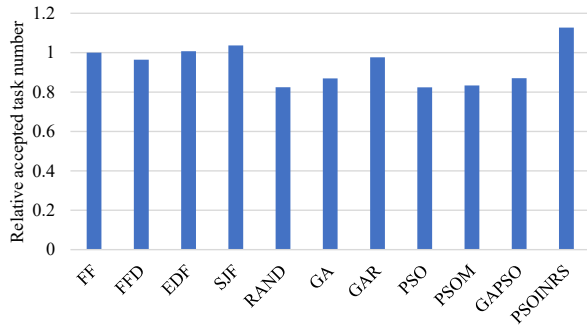


Fig. 2. The accepted task numbers achieved by various methods.

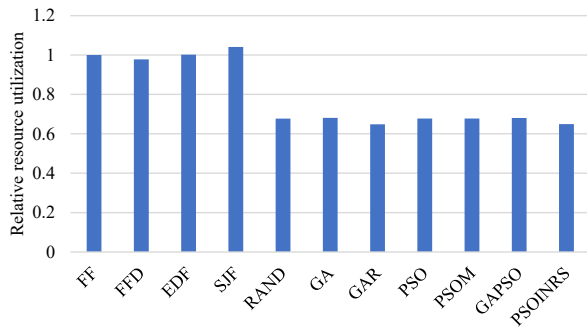


Fig. 3. The resource utilizations achieved by various methods.

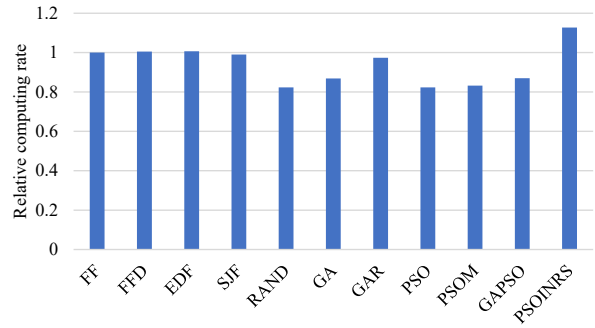


Fig. 4. The computing rates achieved by various methods.

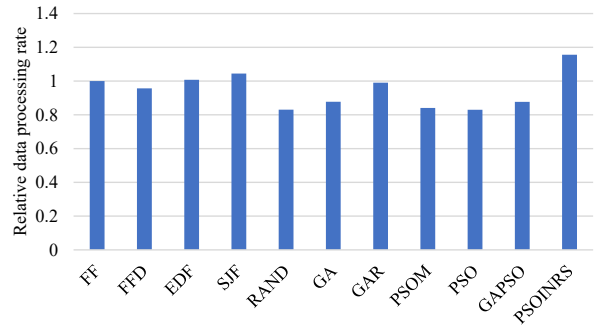


Fig. 5. The data processing rates achieved by various methods.

Fig. 4 and 5 presents the rates of the computing and the data processing when applying these offloading methods. From these figures, we can see that our proposed algorithm has the fast processing rate. Our algorithm has 12.0%–36.9% faster computing rate, and 10.7%–39.6% faster data processing rate than others.

Next, we evaluate the performance of our improvement strategies, the imbalance initialization and the task rescheduling, and the experimental results are shown in Fig. 6 and 7, where PSOIN and PSORS are respectively PSO with the imbalance initialization and the task rescheduling. From these figures, we can see that the imbalance initialization can improve GA by 1.73% and PSORS 12.1%, which verifies the validity of the imbalance initialization. In Fig 7, we can see that offloading algorithms with the task rescheduling can achieve 18.5%–189% better performance than that without it in processing rate. Thus, the task rescheduling strategy is deserved to be integrated into offloading algorithms.

From Fig. 6 and 7, we also can see that the imbalance initialization can decrease the performance of PSO, while the task rescheduling can make up the degradation and improve both the accepted task number, the utilization and the processing rate. This inspires us that the combination of multiple improvement strategies may produce good solutions, even though a signal improvement strategy degrades the overall performance.

V. CONCLUSION

In this paper, we focus on the task offloading problem in DECC systems. We first formulate the problem as a discrete non-convex optimization model, which is hard to be solved. Then, we propose a PSO-based algorithm to solve the task offloading problem, where we use the imbalance initialization and the task rescheduling to improve the performance of PSO on the solving offloading problem in DECC. Extensive experiments are conducted and results verify the superior performance of our proposed algorithm.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their valuable comments and suggestions. The research was supported by the key scientific and technological projects of Henan Province (Grant No. 232102211084, 232102210023, 232102210125), the Key Scientific Research Projects of Henan Higher School (Grant No. 22A520033), Zhengzhou Basic Research and Applied Research Project (ZZSZX202107) and China Logistics Society (2022CSLKT3-334).

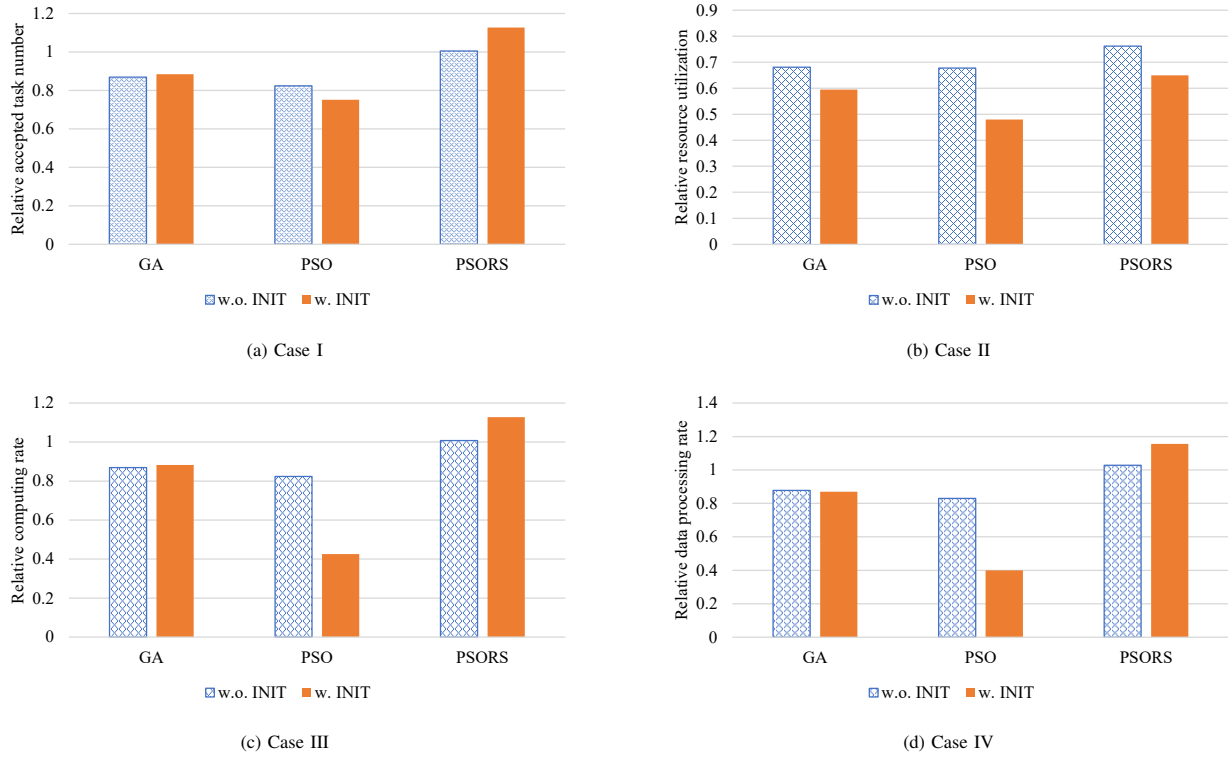


Fig. 6. The improvement of the imbalance initialization.

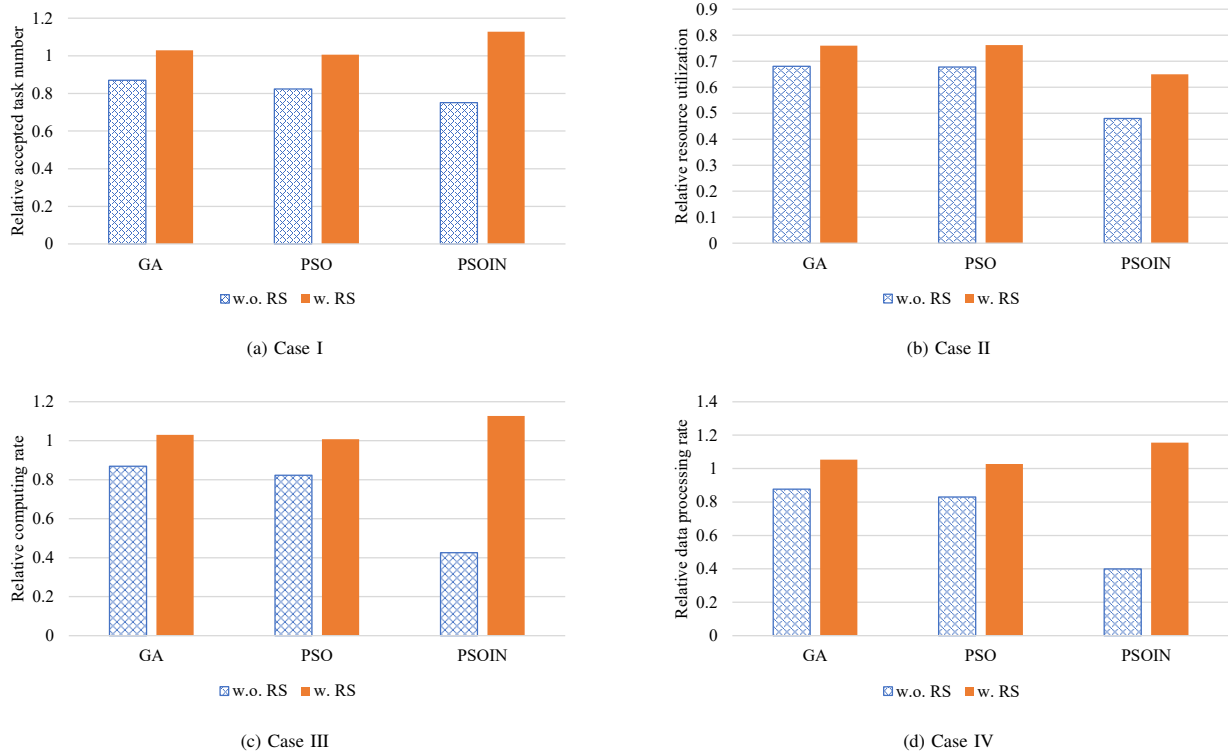


Fig. 7. The improvement of the task rescheduling.

REFERENCES

- [1] K. Cao, Y. Liu, G. Meng and, Q. Sun. An Overview on Edge Computing Research. *IEEE Access*, 2020, 8: 85714-85728. Doi: 10.1109/ACCESS.2020.2991734.
- [2] N. Fernando, S. W. Loke, and W. Rahayu. Mobile cloud computing: A survey. *Future Generation Computer Systems*, 2013, 29(1): 84-106. Doi: 10.1016/j.future.2012.05.023.
- [3] B. Wang, C. Wang, W. Huang, Y. Song, and X. Qin. A Survey and Taxonomy on Task Offloading for Edge-Cloud Computing. *IEEE Access*, 2020, 8: 186080-186101. Doi: 10.1109/ACCESS.2020.3029649.
- [4] C. Wang, R. Guo, H. Yu, Y. Hu, C. Liu, and C. Deng. Task offloading in cloud-edge collaboration-based cyber physical machine tool. *Robotics and Computer-Integrated Manufacturing*, 2023, 79: 102439. Doi: 10.1016/j.rcim.2022.102439.
- [5] Y. Sang, J. Cheng, B. Wang, and M. Chen. A three-stage heuristic task scheduling for optimizing the service level agreement satisfaction in device-edge-cloud cooperative computing. *PeerJ Computer Science*, 2022, 8:e851. Doi: 10.7717/peerj-cs.851.
- [6] X. Chen, T. Gao, H. Gao, B. Liu, M. Chen, and B. Wang. A multi-stage heuristic method for service caching and task offloading to improve the cooperation between edge and cloud computing. *PeerJ Computer Science*, 2022, 8:e1012. Doi: 10.7717/peerj-cs.1012.
- [7] B. Wang, B. Lv, and Y. Song. A Hybrid Genetic Algorithm with Integer Coding for Task Offloading in Edge-Cloud Cooperative Computing. *IAENG International Journal of Computer Science*, 2022, 49(2): 503-510.
- [8] S. Song, S. Ma, J. Zhao, F. Yang, and L. Zhai. Cost-efficient multi-service task offloading scheduling for mobile edge computing. *Applied Intelligence*, 2022, 52(4): 4028-4040. Doi: 10.1007/s10489-021-02549-2.
- [9] M. A. Alqarni, M. H. Mousa, and M. K. Hussein. Task offloading using GPU-based particle swarm optimization for high-performance vehicular edge computing. *Journal of King Saud University - Computer and Information Sciences*, 2022, 34(10-Part B): 10356-10364. Doi: 10.1016/j.jksuci.2022.10.026.
- [10] B. Wang, J. Cheng, J. Cao, C. Wang, and W. Huang. Integer particle swarm optimization based task scheduling for device-edge-cloud cooperative computing to improve SLA satisfaction. *PeerJ Computer Science*, 2022, 8: e893. Doi: 10.7717/peerj-cs.893.
- [11] M. P. John Mahenge, C. Li, and C. A. Sanga. Energy-efficient task offloading strategy in mobile edge computing for resource-intensive mobile applications. *Digital Communications and Networks*, 2022, 8(6): 1048-1058. Doi: 10.1016/j.dcan.2022.04.001.
- [12] H. Hafsi, H. Gharsellaoui, and S. Bouamama. Genetically-modified Multi-objective Particle Swarm Optimization approach for high-performance computing workflow scheduling. *Applied Soft Computing*, 2022, 122: 108791. Doi: 10.1016/j.asoc.2022.108791.
- [13] B. Wang, P. Wu, and M. Arefzaeh. A new method for task scheduling in fog-based medical healthcare systems using a hybrid nature-inspired algorithm. *Concurrency and Computation: Practice and Experience*, 2022, 34(22): e7155. Doi: 10.1002/cpe.7155.
- [14] N. E. Nwogbaga, R. Latip, L. S. Affendey, and A. R. Abdul Rahiman. Attribute reduction based scheduling algorithm with enhanced hybrid genetic algorithm and particle swarm optimization for optimal device selection. *Journal of Cloud Computing*, 2022, 11: 15. Doi: 10.1186/s13677-022-00288-4.
- [15] H. Zhou, Z. Zhang, D. Li and Z. Su. Joint Optimization of Computing Offloading and Service Caching in Edge Computing-based Smart Grid. *IEEE Transactions on Cloud Computing*, 2022, In Press. Doi: 10.1109/TCC.2022.3163750.
- [16] H. Hu, and P. Wang. Computation Offloading Game for Multi-Channel Wireless Sensor Networks. *Sensors*, 2022, 22(22):8718. Doi:10.3390/s22228718
- [17] M. L. Pinedo. *Scheduling Theory, Algorithms, and Systems*, Fifth Edition. Springer International Publishing AG Switzerland, 2016.
- [18] Ipsolve: Mixed Integer Linear Programming (MILP) Solver. 2021. Available online: <https://sourceforge.net/projects/ipsolve/> (accessed on 31 March 2023).
- [19] MathWorks, Inc. *Optimization Toolbox: Solve Linear, Quadratic, Conic, Integer, and Nonlinear Optimization Problems*. 2022. Available online: <https://ww2.mathworks.cn/en/products/optimization.html> (accessed on 31 March 2023).
- [20] A. A. Hussain and F. Al-Turjman. Hybrid Genetic Algorithm for IOMT-Cloud Task Scheduling. *Wireless Communications and Mobile Computing*, 2022, 2022: 6604286. Doi: 10.1155/2022/6604286

Prediction of Air Quality and Pollution using Statistical Methods and Machine Learning Techniques

Mr. V. Devasekhar¹, Dr. P. Natarajan²

Research Scholar, SCOPE, VITU, Vellore, Associate Professor, Department of CSE-GNITC, Hyderabad, T.S, India¹
Associate Professor, Department of SCOPE, VITU Vellore, Tamil Nadu (State), India²

Abstract—Air pollution is a major environmental issue and machine learning techniques play an important role in analyzing and forecasting these data sets. Air quality is an outcome of the complex interaction of several factors involving the chemical reactions, meteorological parameters, and emissions from natural and anthropogenic sources. In this paper, we propose an efficient combined technique that takes the benefits of statistical techniques and machine learning techniques to predict/forecast the Air Quality and Pollution in particular regions. This work also indicates that prediction performance varies over different regions/cities in India. We used time series analysis, regression and Ada-boosting to anticipate PM 2.5 concentration levels in several locations throughout Hyderabad on an annual basis, depending on numerous atmospheric and surface parameters like wind speed, air temperature, pressure, and so on. Dataset for this investigation is taken from Kaggle and experimented with proposed method and comparison results of our experiments are then plotted.

Keywords—Air quality; forecasting; machine learning; statistical techniques

I. INTRODUCTION

Today, all cities have attracted significant attention in the context of urban development approaches [2,3]. The Internet and developments in broadband networking are seen as enablers of e-services and are becoming increasingly important for urban development and addressing the massive air pollution problem. Citizens and governments around the world have witnessed and expressed growing concern about the impact of air pollution [4] on human health, as well as advocated sustainable development to address air pollution challenges. The result of modern manufacturing is a mixture of liquid droplets, solid particles, and gas molecules that is dispersed throughout the atmosphere. The high concentration of particulate matter of size $PM_{2.5}$ has a major negative impact on human health.

Recent studies have focused on rigorous statistical learning algorithms for evaluating air quality and predicting pollution levels. Neural networks have been employed by Raimondo et al. [5], Garcia et al. [6], and Park et al. [7] to build models for forecasting the occurrence of individual pollutants, such as particles less than 10 microns (PM_{10}). To train their models, Raimondo et al. [5] employed a support vector machine (SVM) and an artificial neural network (ANN). Their best ANN model had a specificity of nearly 79 percent and a false-positive rate of only 0.82 percent, while their best SVM model had a specificity of 80 percent and a false-positive rate of only 0.13 percent. For AQI category prediction, Yu et al. [8] suggested

RAQ, a random forest technique. After that, Yi et al. [9] used deep neural networks to predict AQI categories. For forecasting AQI levels, Veljanovska and Dimoski [10] used several settings to surpass k-nearest neighbour (k-NN), decision tree, and SVM. Their ANN model outperformed all other algorithms examined, with an accuracy of 92.3 percent.

The deep learning architecture is suitable for solving air pollution prediction problems and nonlinear problems, and to learn long-term dependencies from time-series data[23]. In [24], authors presented deep learning solution to predict the hourly forecast of $PM_{2.5}$ concentration in Beijing, China, based on CNN-LSTM, and other hybrid deep learning techniques for air pollution analysis is presented in [25].

A. Air Quality Monitoring

The CPCB recommends a combination of physical, wet-chemical, and continuous online measuring procedures for each parameter. Analyzers for measuring PM_{10} , $PM_{2.5}$, SO_2 , CO , NO_2 , O_3 , NH_3 , and Benzene are provided in air quality monitoring systems [10,11,13]. Using filter-based air samplers, the metallic parameters Pb, Ni, and As are assessed offline.

The ambient air quality monitoring station (AQMS) having the following systems:

- PM_{10} & $PM_{2.5}$: It Measures particle mass concentrations ranging from 0 to 5 mg/m^3 with a lowest detection limit of 1 g/m^3 . It works on the principle of Beta Ray Attenuation. A PM_{10} intake and a $PM_{2.5}$ inlet are included in the equipment.
- NO_x and NH_3 : Based on the chemiluminescence technique, with a detection range of 0 to 2000 g/m^3 and a minimum detection limit of 0.5 g/m^3 .
- SO_2 Analyzer: Operates on the UV Fluorescence technique, with a detection range of 0 to 2000 g/m^3 and a minimum detection limit of 0.5 g/m^3 .
- CO Analyzer: Uses the Non-Dispersive Infrared Spectrometry (NDIR) method to measure CO levels ranging from 0 to 100 mg/m^3 with a detection limit of 0.03 g/m^3 .
- O_3 Analyzer: Works on the UV Photometry principle, with a range of 0 to 2500 g/m^3 and a minimum detection limit of 0.5 g/m^3 .

- BTEX (Benzene, Toluene, Ethylbenzene, Xylene): GC/PID for automatic monitoring of BTEX in air with a minimum detection threshold of 10 ppt in ambient air.
- Multigas Calibrator: used to manually, remotely, or automatically calibrate gas analyzers for quality assurance. Up to 20 points of multi-calibration Ultrasonic Wind Sensor, Barometric Pressure, Temperature, Relative Humidity, Rainfall, Solar Radiation, and other features of an automatic weather station (AWS).
- Except for the AWS, all of these instruments are kept in a room or walk-through shelter with an appropriate sampling system for gaseous and particulate matter measurements.

B. Our Contribution

In this research,

- 1) We propose an effective combination method to forecast and anticipate air quality and pollution in specific areas by combining the advantages of statistical and machine learning methods.
- 2) Additionally, this research suggests that prediction accuracy differs between Indian cities and regions.
- 3) We predicted annual PM 2.5 concentration levels in different places throughout Hyderabad using time series analysis, regression, and Ada boosting, depending on a variety of meteorological and surface characteristics like wind speed, air temperature, pressure, and so on.
- 4) The investigation's data set was obtained via Kaggle, and the suggested strategy was tested.

C. Organization of the Paper

The remaining paper is structured as follows- Literature review in this domain is presented in Section II. Section III provides preliminaries related to machine learning and other statistical techniques. Section IV presents our proposed hybrid method. The experiment results with comparative results are mentioned in Section V. Conclusions are discussed in Section VI.

II. RELATED WORK

Gopalakrishnan (2021) [26] used Google Street View data and machine learning to predict air quality in various locations throughout Oakland, California. The author created a web application that can predict pollution levels in any city and neighbourhood. Sanjeev (2021) [27] examined a set of data that would include pollutant concentrations as well as meteorological factors. Castelli et al. (2020) [28] used the Support Vector Regression (SVR) ML algorithm to forecast air quality in California in terms of pollutants and particulate levels. The researchers claimed to have created a novel method for modeling hourly atmospheric pollution. In [29] Doreswamy et al. (2020) investigated ML predictive models for PM concentration forecasting in the air. The authors examined six years of Taiwanese air quality monitoring data and applied existing models. They claimed that predicted and actual values were extremely close. Based on 11 years of data,

Liang et al. (2020) [30] investigated the performance of six ML classifiers in predicting Taiwan's AQL. Madan et al. (2020) [31] compared the performance of ML algorithms and twenty different literary works over pollutants studied. The authors discovered that many works used meteorological data such as humidity, wind speed, and temperature to more correctly estimate pollutant concentrations. They discovered that the Neural Network (NN) and boosting models outperformed the other leading machine learning (ML) algorithms. Monisri et al. (2020)[32] gathered air pollution data from a variety of sources in order to create a mixed model for predicting air quality. The proposed model, according to the authors, aims to assist people in small towns in analysing and forecasting air quality. Based on ML classifiers, Nahar et al. (2020) [33] created a model to predict AQI. Their proposed model accurately detected the most contaminated areas. Patil et al. (2020) presented some research papers on various machine learning techniques for AQI modeling and forecasting. Multi-agent systems[11,12] have been proposed as a helpful apparatus for huge scope frameworks like Important traffic and air quality control). The significant objective of such a framework is to help street administrators with traffic the board tasks while likewise further developing air quality on the course. Many examinations have proposed the use of MAS innovation in traffic signal and the executives frameworks, for example, (Namoun et al. 2013), which proposes an incorporated technique for demonstrating transportation framework and streamlining transportation in metropolitan regions to diminish fossil fuel byproducts.

In the hybrid classification PM2.5 fixation determining models, highlight determination is seldom applied in several techniques. Nonetheless, assuming that a PM2.5 fixation determining model's feedback incorporates an enormous number of elements (PM 2.5 etc.), it could be hard to prepare the model and increment the preparation time. This affects the PM2.5 fixation anticipating model's power [14]. At the same time, muddled info information might bring about overfitting of the model and a decrease in model precision [15]. The guideline parts investigation (PCA), stage space transformation (PSR), and angle helped relapse tree are at present famous component choice methodologies (GBRT).

Notwithstanding, on the grounds that these techniques assume a direct framework, they might be insufficient for air contamination focus arrangements, bringing about issues, for example, inability to accomplish worldwide ideal decrease. The fluffy hypothesis based unpleasant sets characteristic decrease (RSAR) strategy offers the benefits of unambiguous stop measures and no boundaries [16]. Through the reliance between particular ascribes, RSAR can decide the objective property's fundamental trait set. The RSAR calculation [17] is a well known review point. Information mining and examination as often as possible utilize grouping methods [33]. K-implies grouping (KC) [18], probabilistic c-means (PCM) [19], fix clustering [20], and other bunching approaches exist. The KC algorithm, when compared to others, gives an advantage based on easy procedure, quick computation speed, and great clustering results; as a result, it is currently the most extensively used clustering algorithm[21]. Combining the RSAR and KC algorithms allows the RSAR method to generate suitable, which is a promising exploration course.

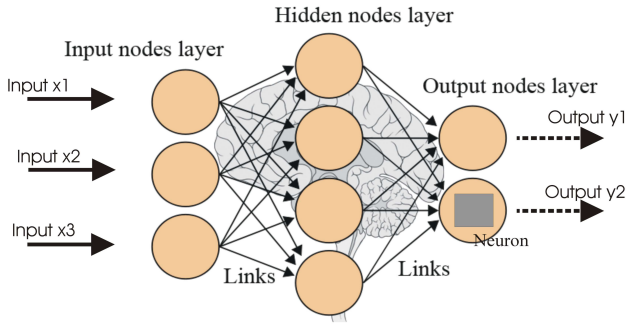


Fig. 1. Artificial neural network structure.

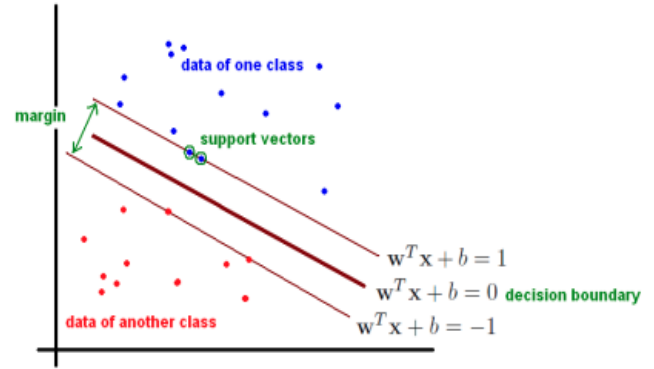


Fig. 2. SVM pictorial representation.

III. MACHINE LEARNING TECHNIQUES

Atmospheric particulate matter (PM) is one of the pollutant that may have a significant impact on human health. Data collected from various regions of country can be analysed using various models : a multiple linear regression model, a neural network models and other machine learning models. In this section, we present some significant statistical preliminaries, soft computing based methods and mathematical notations [2,4,7].

A. Artificial Neural Network

It is a collection of classifiers whose standard project for future and utility are similar to the algorithmic model of human cerebrum structure [1]. The precise organisation of neuronal organisation varies for order enquiry. To begin, the topological structure and number of organisation hubs present in the core layer are resolved throughout the preparation. For example, n-dimensional planes and hyperplanes have no eccentricity, unlike SVM. In any event, preparing informational collection measures takes time and results in less precise and proficient results. Fig. 1 provides ANN framework and its input parameters.

- For singleton data sequence, $x_0, x_1, x_2, x_3 \dots x_{(n)}$ depicts several data items to the computational network. Each input is being multiplied with a corresponding weights. These weights are depicted as $w_0, w_1, w_2, w_3 \dots w_{(n)}$. These weight represents the major strength of any specific node.
- Here b is considered as a bias value. A certain bias value permits to move the enactment increasing or decreasing work.
- In most straightforward cases, these items are added, taken care of a transfer function (activation function) to produce an outcome, and this outcome is sent as yield.

$$x_1.w_1 + x_2.w_2 + x_3.w_3 \dots x_n.w_n = \sum x_i.w_i$$

- Presently, enactment (activation) function is applied, i.e. $\phi(\sum x_i.w_i)$

B. Support Vector Regression

Its well known that the categorization method comes in either supervised method or unsupervised method. Therefore, in the field of ML, support vector network architecture comes in the category of supervised machine learning standards. An SVM [2, 3] is the description of features / attribute states in the plane, alongside the non-direct hyperplanes for detachment task in order. A few boundaries like gaussian parts, standard deviation and fluctuation of information, bit capacities are some critical boundaries which influence the exhibition of SVM.

Fig. 2 provides the SVM framework and its representation.

1) Linearly-separable data aided Binary classification:

- **Goal:** we need to discover the hyperplane (for example choice limit) directly isolating to listed class. The limit of listed the condition: $w^T x + b = 0$.
- Any of the classes that falls this decision boundary must be label to 1. i.e., x_i such that $w^T x + b > 0$ will be having respective $y_i = 1$.
- Likewise, anything underneath the choice limit ought to have labeling - 1. i.e., x_i s.t. $w^T x + b < 0$ will possesses respective $y_i = -1$.

C. Learning

This section lists out various types of learning and tries to find out the answer of the question that, why deep learning outperforms over other traditional algorithms? The discussion is given below.

Types of learning:-

Various learning methods are mentioned below:-

1) *Active learning:* It picks a subset of an unstructured and basic event for reason for marking. The dynamic learner acquires bigger precision utilizing decreased measure of events.

2) *Kernel-based learning:* It is demonstrated to be a predominant approach to upgrade the computational potential proficiently. It is profitable with respect to that, both direct just as non-straight vector bit utilitarian strategies are available

to manage the non-linearity of information in N-dimensional element space.

3) *Transfer learning*: It is primarily advantageous as it can productively apply information, which has been adapted already so as to discover answer for new issues in quick, optimal and successful way.

4) *Distributed learning*: This sort of learning restrains the bunch arrangement, in which one processes thread is designated to each group in plan to perform multi-stringing in parallel and distributed fashion.

5) *Deep learning*: This learning considers more muddled, compartmented measurable examples of data sources and figures out how to be robust for new areas when contrasted with customary learning frameworks.

6) *Supervised learning*: It deals with learning a function from available training set data. A supervised learning algorithm uses the accessible training information and makes an induced capacity, which can be used further for the purpose of mapping the new ones.

7) *Unsupervised learning*: It manages unlabeled information without taking any predefined dataset for its preparation. This learning can be considered as an intense apparatus for look for patterns and trends and analysing available data. It is commonly employed for clustering similar input into logical groups.

8) *Classification*: is learning an specific function that maps (orders) an information thing into one of a few predefined categories [24]. Instances of grouping techniques utilized as a feature of information revelation applications remember the arranging of patterns for money related business sectors and the computerized recognizable proof of objects of revenue in enormous image oriented data sets [25]. The bank may exploit the characterization districts to consequently choose whether future credit candidates will be given an advance or not.

9) *Regression*: is learning a limit that maps a data thing to a real-valued esteemed forecast variable. Relapse applications are many, for instance, predicting the proportion of biomass present in a backwoods given distantly detected microwave estimations, assessing the likelihood that a patient will endure given the after effects of a bunch of analytic tests, anticipating shopper interest for another item as a component of promoting use, and foreseeing time arrangement where the information factors can be time-slacked adaptations of the prediction variable.

10) *Clustering*: is a typical graphic assignment where one looks to distinguish a limited arrangement of classifications or bunches to portray the information. The classes can be commonly exclusive and thorough or comprise of a more extravagant portrayal, for example, various leveled or covering classifications. Instances of bunching applications in an information disclosure setting incorporate finding homogeneous sub populaces for various customers in advertising data sets and distinguishing subcategories of spectra from infra-red sky estimations.

D. Information and Decision System

An instance of *information system* is represented as pair (U, A) where,

U : denotes - a kind of non-empty set.

A : denotes - a non-empty type finite set of features.

However, A *decision table* is a system possessing the form $S = (U, C \cup \{d\})$ where -

C : denotes set of conditional variables.

d : denotes the decision variable.

E. Uncertainty Approximation

The approximation extent of an uncertain variety of concept in a knowledge space is performed as follows - Let $S = (U, R)$ be a certain approximation space, X be a type of concept in that space, then, the *lower approximation*(inf) is-

$$\underline{R}X = \{x \in U \mid [x] \subseteq X\}$$

upper approximation(sup) is-

$$\overline{R}X = \{x \in U \mid [x] \cap X \neq \emptyset\}$$

where, $[x]$ can be considered as equivalence class, possessing an element e .

IV. PROPOSED METHOD

We propose an efficient method that uses statistical method and machine learning techniques for Air Quality Predication. In this section, we present our proposed method in phased manner as follows:

A. Data Preprocessing

Various preprocessing procedures, in overall, come before the learning phase. In India, O₃, PM_{2.5}, PM₁₀, CO, SO₂, and NO₂ pollutants are monitored with respect to its concentration. Our method used the air quality data [22] between Nov 2019 to April 2020, that are collected from several monitoring stations across major metropolitan cities such as Delhi, Mumbai, Chennai, Bangalore and Hyderabad of India and reported via the Govt website [22]. At any given time, an inaccurate parameter will not influence the global data group.

B. Feature Selection

To get an efficient representation of data, the auto encoder procedure must capture very important features using several measures which are part of the method.

Given sequence of data points $x(1), x(2), x(3), \dots, x(N)$, where $x^i \in R^D$, an auto encoder first encodes the info vector x to a more elevated level secret portrayal y in view of condition (1), and after ward it unravels the portrayal y back to a remaking z , determined as in condition (2):

$$y = f(W_1x + b) \quad (1)$$

$$z = g(W_2y + c) \quad (2)$$

where W_1 and W_2 are weights chosen for the procedure and b and c are the vectors of corresponding function. We utilized the strategic sigmoid capacities for $f(x)$ and $g(x)$ in this method.

$$\text{Sigmoid function} = 1/(1 + \exp(-x))$$

The boundaries of this neural organization are enhanced to limit the normal remaking error:

$$j(\theta) = \frac{1}{N} \sum_{i=1}^N M(x^{(i)}, z^{(i)}) \quad (3)$$

In this a loss function is M . We involved the customary squared blunder in our method.

C. Statistical Modeling

The underlying dividing step can be done at least numerous times to wipe out the non-stationarity of the mean capacity, and the accompanying quantifiable approach for showing is used as the info data exhibits proof of non-stationarity in the sensation of mean.

BEGIN PROCEDURE {

Consider, process PROC(α, β, γ) where; α, β, γ are +ve integers.

α : no. of lags.

β : difference degree

γ : order of the moving average model

for an input time series pollution data D_t , here index of integer is t and D_t are real numbers.

PROC(α', γ) is given by -

$$D_t - a_1 D_{t-1} - \dots - a_{\alpha'} D_{t-\alpha'} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_\gamma \varepsilon_{t-\gamma} \quad (4)$$

above equation can be equivalently represented as -

$$\left(1 - \sum_{i=1}^{\alpha'} a_i M^i\right) D_t = \left(1 + \sum_{i=1}^{\gamma} \theta_i M^i\right) \varepsilon_t \quad (5)$$

here, M is lag operator

a_i - autoregressive parameters

θ_i - average part parameters

ε_t are represented as error terms

$\nabla \varepsilon_t$ are considered to be iid (independent, identically distributed) variables and these are normal distribution samples with zero mean.

Consider the polynomial $\left(1 - \sum_{i=1}^{\alpha'} a_i M^i\right)$ has a unit root of multiplicity β . The mentioned polynomial also can be rewritten as -

$$\left(1 - \sum_{i=1}^{\alpha'} a_i L^i\right) = \left(1 - \sum_{i=1}^{\alpha' - \beta} \varphi_i M^i\right) (1 - M)^\beta \quad (6)$$

with polynomial factorization property,

$\alpha = \alpha' - \beta$ and given as -

$$\left(1 - \sum_{i=1}^{\alpha} \varphi_i M^i\right) (1 - M)^\beta D_t = \left(1 + \sum_{i=1}^{\gamma} \theta_i M^i\right) \varepsilon_t \quad (7)$$

this can be further generalized as -

$$\left(1 - \sum_{i=1}^{\alpha} \varphi_i M^i\right) (1 - M)^\beta D_t = \delta + \left(1 + \sum_{i=1}^{\gamma} \theta_i M^i\right) \varepsilon_t \quad (8)$$

where, drift $\rightarrow \frac{\delta}{1 - \sum \varphi_i}$

} END PROCEDURE

D. Model Evaluation

To build a model to predict concentrations, we used two different machine learning algorithms including a simple linear regression model and non linear regression. The NO2 dataset was split into test/train data and a cross-validation approach was applied to the training dataset. We use Linear Regression and A da-boosting for improving efficiency of the results.

When the quantity of PM 2.5 and PM 10 pollutant particles in the atmosphere is indeed very high, it has a negative impact on our health and can cause life-risk issues in a less period of time. Particulate matter has been shown to have an effects on peoples health, often at the genetic level, according to studies. So we are emphasising our work to forecast the concentration of PM 2.5 levels in the atmosphere. "Hyderabad Weather with Air Quality index and Covid" dataset is used. The dataset [22] is downloaded into .csv format. Brief about the datasets are as follows: It consists of a total of 5 months of data between October 2019 to April 2020 as described below:

- Date: dd/mm/yyyy
- Humidity
- Wind Speed
- Dew Point
- Temperature
- Pressure
- Festival (Rating out of 5)
- Lockdown (0 for No, 1 for Yes)
- Covid-19 Cases in Hyderabad
- Air quality (PM2.5)

Fig. 3 shows the first 10 rows along with the column values of the dataset when reading the csv file into the colab file.

```
df = pd.read_csv('/content/Hyderabad-AirQ -2019-20.csv')
df.head(10)
```

	Date	Humidity	Wind Speed	Dew Point	Temperature	Pressure	Festival	Lockdown	Covid-Case	PM2.5
0	01-10-2019	83.0	4.5	76.0	81.9	28.1	0	0	0	84
1	02-10-2019	81.6	4.6	77.4	83.6	28.1	0	0	0	83
2	03-10-2019	82.0	3.7	75.3	81.7	28.1	0	0	0	81
3	04-10-2019	85.4	2.7	73.9	78.6	28.1	0	0	0	94
4	05-10-2019	87.4	3.5	75.3	79.4	28.1	0	0	0	112
5	06-10-2019	86.3	2.5	73.3	77.9	28.1	4	0	0	97
6	07-10-2019	82.9	2.4	76.0	81.6	28.1	0	0	0	126
7	08-10-2019	86.2	3.1	76.3	80.6	28.0	3	0	0	125
8	09-10-2019	88.7	1.7	74.5	78.1	28.0	0	0	0	132
9	10-10-2019	85.6	3.4	72.9	77.3	28.1	0	0	0	138

Fig. 3. First 10 rows along with column of the dataset.

We pre-processed it, check if there is any null value inside it and removed it. After the data has been cleaned and pre-processed, it is submitted to later experiment, which includes time series analysis and determining the total impact

of each characteristic on the PM 2.5 value. Date is also an impacting factor which shows the PM 2.5 values increasing or decreasing with time. So we grouped them and find the values of each attribute. Fig. 4 shows the grouped values of each attribute. With the help of these values we are plotting the variation of PM 2.5 w.r.t date. Fig. 5 shows graph plot of PM 2.5 vs Date.

```
[ ] df= df.groupby("Date").mean()
df.head(10)
```

Date	Humidity	Wind Speed	Dew Point	Temperature	Pressure	Festival	Lockdown	Covid-Case	PM2.5
01-03-2020	62.5	5.3	67.3	82.7	28.1	0	0	1	113
01-04-2020	62.0	3.9	72.0	92.6	28.1	0	1	96	120
01-10-2019	83.0	4.5	76.0	81.9	28.1	0	0	0	84
01-11-2019	82.7	4.8	75.3	81.3	28.1	1	0	0	57
01-12-2019	80.7	5.5	71.7	79.2	28.2	0	0	0	147
02-03-2020	58.9	4.6	67.3	84.0	28.1	0	0	1	114
02-04-2020	55.0	4.7	72.0	96.8	28.1	0	1	96	130
02-10-2019	81.6	4.6	77.4	83.8	28.1	0	0	0	83
02-11-2019	79.8	4.3	75.3	82.4	28.1	2	0	0	61
02-12-2019	86.3	5.8	71.1	75.3	28.2	0	0	0	144

Fig. 4. Grouped values of the dataset.

The format of date is $xx - zz - yy$ where xx is date, zz is month and yy is year. The value of PM 2.5 is increasing yearly. In December its value is higher as compared to other months. December is a winter month, so we say that in the winter months PM 2.5 value is high.

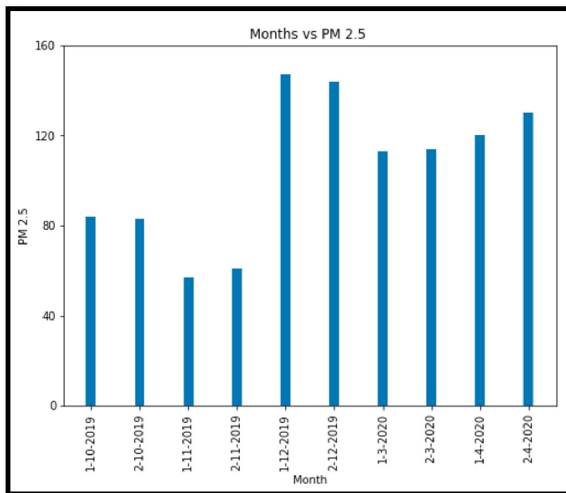


Fig. 5. PM 2.5 vs Month.

The effect of wind on PM 2.5 is also investigated. The PM 2.5 value was found to be lower when the wind speed was high, and vice versa. Fig. 6 shows a scatter plot of wind speed vs PM 2.5 value, which confirms that we have higher wind speed concentrations. It's also worth noting that when the wind speed is between 6–10m/s, PM 2.5 levels are essentially non-existent.

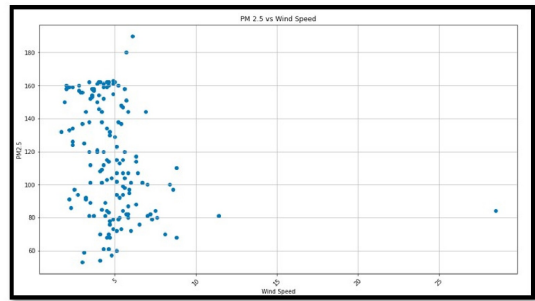


Fig. 6. Effects of PM 2.5 due to wind speed.

The impacts of humidity on PM 2.5 is investigated. The value of PM 2.5 value was found to be lower when the humidity value ranges between 30 gm^3 to 60 gm^3 . Its value increases to $120 \mu\text{m}$ when humidity value ranges between 60 gm^3 to 80 gm^3 . Fig. 7 shows a scatter plot of humidity vs PM 2.5 value.

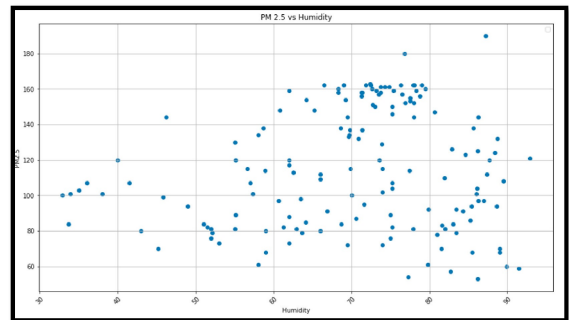


Fig. 7. Effect of humidity on PM 2.5 values.

The effect of dew drops on PM 2.5 was found to be lower when dew drops value ranges between $60^\circ\text{F} - 65^\circ\text{F}$ and PM 2.5 value increases when dew points value ranges between 65° to 75° . Fig. 8 shows a scatter plot between PM 2.5 vs Dew Points

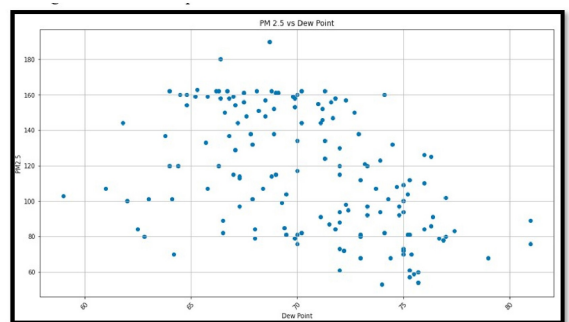


Fig. 8. Effect of dew points on PM 2.5.

Similarly the impacts of temperature on PM 2.5 is investigated. Its value was found lower when the value of temperature

increases whereas the value of PM 2.5 increases when the temperature is lower. It shows that at winter season PM 2.5 value increase into the atmosphere. Fig. 9 shows a scatter plot between PM 2.5 vs Temperature.

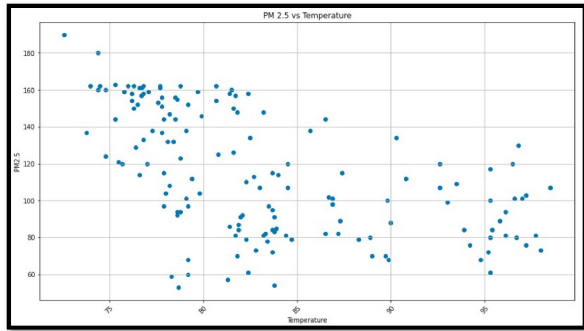


Fig. 9. PM 2.5 vs temperature.

Also the effects of pressure is investigated on PM 2.5. Fig. 10 shows a scatter plot between PM 2.5 and Pressure.

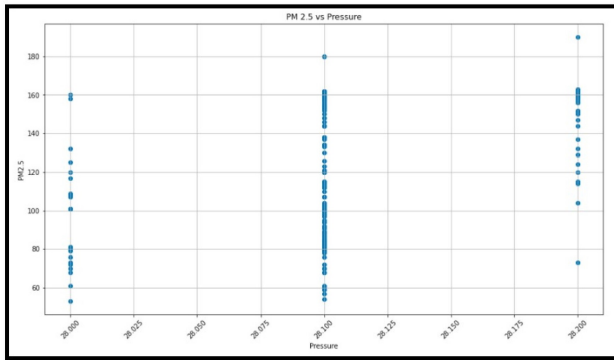


Fig. 10. PM 2.5 vs pressure.

V. EXPERIMENT RESULTS AND COMPARATIVE ANALYSIS

We present experiment results of our proposed methods and efficiency analysis among several methods.

A. Metrics

Firstly, we brief the metrics used for measuring the results are discussed below:

1) *Mean Absolute Error (MAE)*:: It refers to the size of the difference between the predicted and true value of an observation. The average of absolute errors for the entire group is used to calculate the size of mistakes for a group of forecasts and observations.

$$MAE = \frac{1}{n} \sum_{i=1}^n |\tilde{Y}_i - Y_i|$$

In this N is the total number of data items, y_i is i -th measurement, and \tilde{Y}_i is its respective prediction.

2) *Root Mean Square Error (RMSE)*:: One of the most often used approaches for assessing the validity of estimates is root mean square error, also known as root mean square deviation. It uses Euclidean distance to demonstrate how far predictions differ from observed true values.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n |Y_i - \tilde{Y}_i|^2}{n}}$$

In this N is the total number of data items, y_i is i -th measurement, and \tilde{Y}_i is its respective prediction.

3) *R2_score*:: The coefficient of determination, often known as the R2 score, is used to assess the performance of a linear regression model.

$$R2_{score} = 1 - \frac{\sum_{i=1}^n (Y_i - \tilde{Y}_i)^2}{\sum_{i=1}^n (Y_i - \mu)^2} \quad (9)$$

In this N is the total number of data items, y_i is i -th measurement, and \tilde{Y}_i is its respective prediction and μ is the mean of actual values.

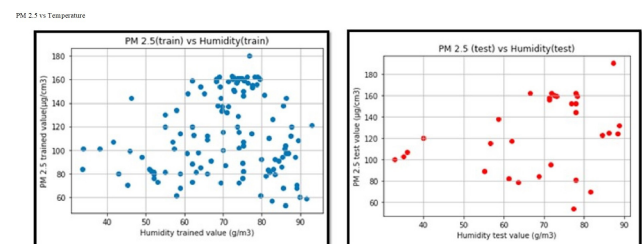
B. Methods: Experiment Results

We present experiment results of our proposed methods by adopting linear regression, Ada-boosting and XG-boosting on standard data sets available on Kaggle [22].

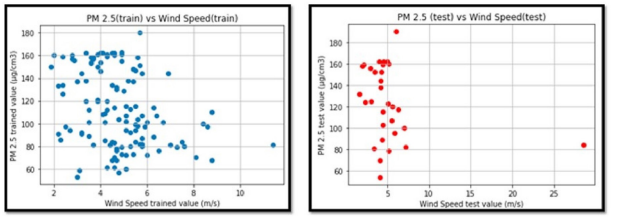
1) *Linear regression*: We get the values of R2 is very less or nearly equal to 0.5. Absolute mean error is 27.33 when temperature vs PM 2.5 Linear Regression calculated. Root mean square error is 30.31 when temperature vs PM 2.5 Linear Regression is calculated. Table I shows the comparative result of mean absolute error, mean square error and r2 score. Fig. 11(a), (b) show the comparative analysis graph. As absolute mean error and root mean square error are the errors so when their value is small, then the model is good. And $r2_score$ calculates the accuracy of the model so when their value is high, model is good.

TABLE I. COMPARATIVE RESULT OF EVALUATION METRICS.

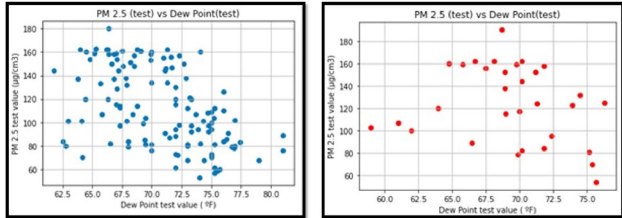
Factors/Error	MAE	RMSE	R2_score
Humidity	29.51	35.99	-0.07
Wind Speed	35.39	41.93	-0.54
Dew Point	31.58	34.58	-0.06
Temperature	27.33	30.31	0.2
Pressure	29.28	33.86	0.0



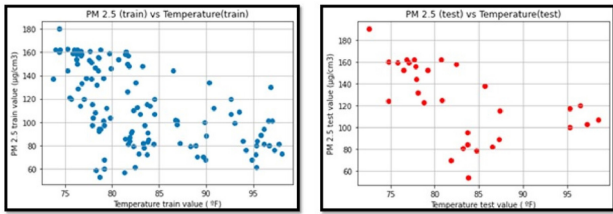
(a) Training (left) and testing (right) results for PM2.5 vs. Humidity



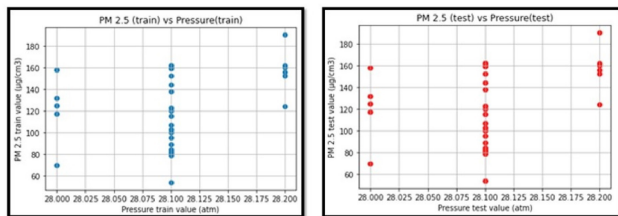
(b) Training (left) and testing (right) results for PM2.5 vs. Wind speed



(c) Training (left) and testing (right) results for PM2.5 vs. Dew-point



(d) Training (left) and testing (right) results for PM2.5 vs. Temperature



(e) Training (left) and testing (right) results for PM2.5 vs. Dew Pressure

Fig. 11. Training and testing results.

Fig. 11(a) represents training and testing results between PM2.5 and Humidity. Fig. 11(b) represents training and testing results between PM2.5 and Wind speed. Fig. 11(c) represents training and testing results between PM2.5 and Dew-point. Fig. 11(d) represents training and testing results between PM2.5 and Temperature. Fig. 11(e) represents training and testing results between PM2.5 and Pressure.

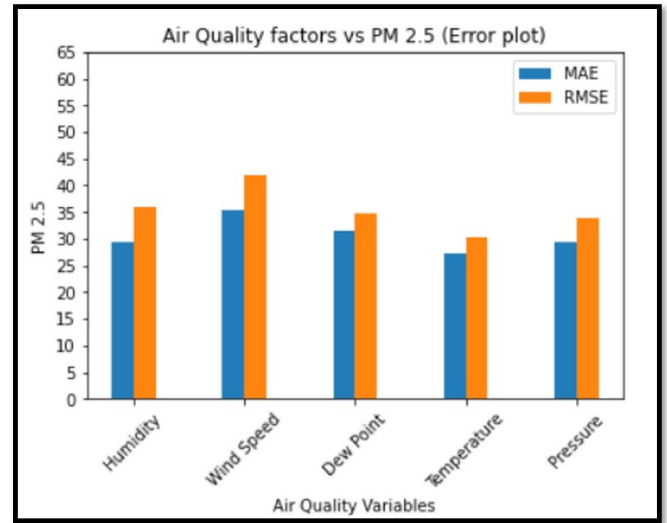


Fig. 12. Comparative analysis of mean square errors and root mean square error.

Fig. 12 shows the comparative analysis of mean square errors and root mean square error.

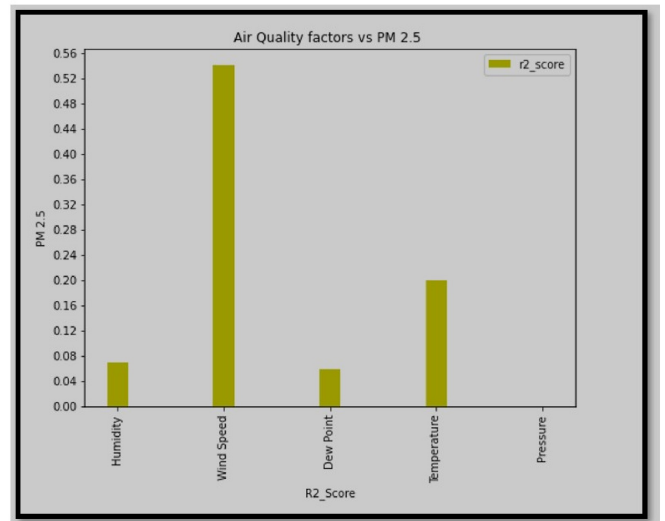


Fig. 13. Comparative analysis of r_2_score on atmospheric factors.

2) *Ada-Boosting*: We get the values of R2 is very less or nearly equal to 0.38. Absolute mean error is 22.39 when temperature vs PM 2.5 Ada-Boosting Regression calculated. Root mean square error is 26.42 when temperature vs PM 2.5 Ada-Boosting Regression is calculated.

Table II shows the comparative result of mean absolute error, mean square error and r_2 score. Fig. 13 and 14 shows the comparative analysis graph.

TABLE II. COMPARATIVE RESULT OF EVALUATION METRICS

Factors/Error	MAE	RMSE	$R2_score$
Ada- Humidity	27.68	35.55	-0.10
Ada-Wind Speed	29.33	36.41	-0.16
Ada-Dew Point	27.35	30.76	-0.17
Ada-Temperature	22.39	26.42	0.38
Ada-Pressure	28.99	33.44	-0.02

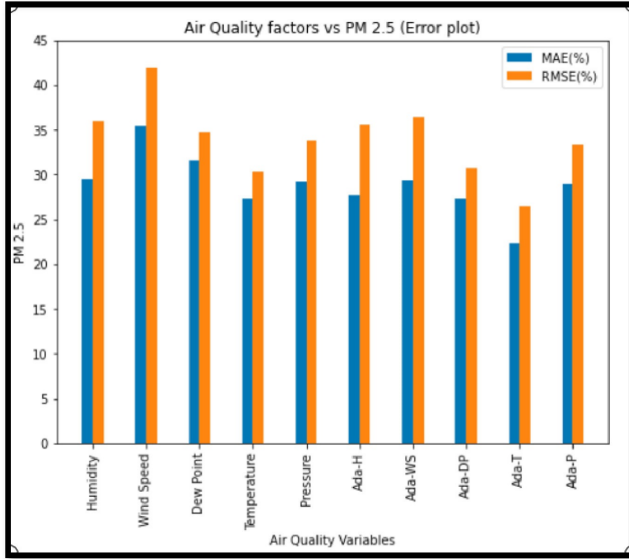


Fig. 14. Comparative analysis of mean square errors and root mean square error using Ada-Boosting.

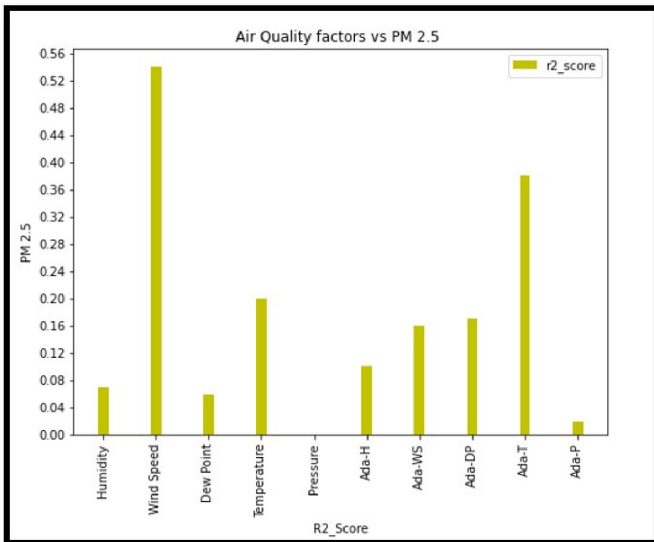


Fig. 15. Comparative analysis of $r2_score$ using Ada-Boosting.

3) *XG-Boosting regression*: We get the values of $r2_score$ is very less or nearly equal to 0.23 when PM 2.5 vs Humidity is calculated. Mean Absolute error is 28.09 when humidity

vs PM 2.5 XG-Boosting Regression calculated. Root mean square error is 35.07 when dew points vs PM 2.5 XG-Boosting Regression is calculated. Table III shows the comparative result of mean absolute error, mean square error and $r2$ score. Fig. 15 and 16 show the comparative analysis graph.

TABLE III. COMPARATIVE RESULT OF EVALUATION METRICS

Factors/Error	MAE	RMSE	$R2_score$
XGB- Humidity	28.09	37.62	-0.23
XGB-Wind Speed	32.31	42.10	-0.5
XGB-Dew Point	28.67	35.07	-0.07
XGB-Temperature	29.65	38.92	-0.32
XGB-Pressure	28.30	32.87	-0.05

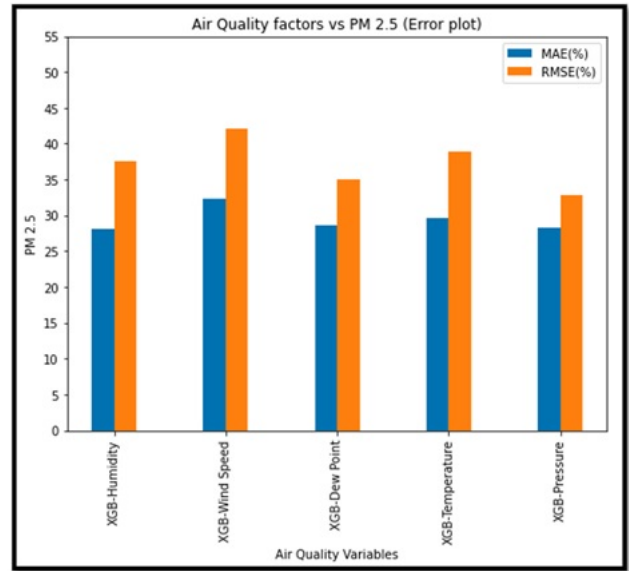


Fig. 16. Comparative analysis of mean square errors and root mean square error using XG-Boosting.

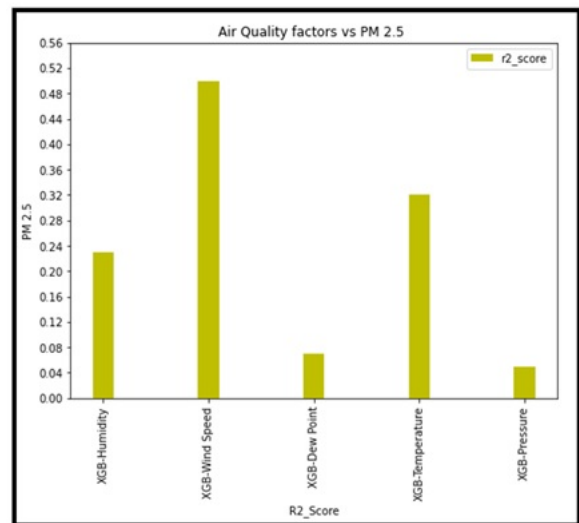


Fig. 17. Comparative analysis of $r2_score$ using XG-Boosting.

4) *Efficiency analysis of models:* Final comparative result of all the models are given in a Table IV and the comparative graph is shown in Fig. 17.

TABLE IV. COMPARATIVE RESULT OF EVALUATION METRICS

Factors/Error	MAE	RMSE	R2_score
LR-Humidity	29.51	35.99	-0.07
LR-Wind Speed	35.39	41.93	-0.54
Dew Point	31.58	34.58	-0.06
L-Temperature	27.33	30.31	0.2
Pressure	29.28	33.86	0.0
Ada- Humidity	27.68	35.55	-0.10
Ada-Wind Speed	29.33	36.41	-0.16
Ada-Dew Point	27.35	30.76	-0.17
Ada-Temperature	22.39	26.42	0.38
Ada-Pressure	28.99	33.44	-0.02
XGB- Humidity	28.09	37.62	-0.23
XGB-Wind Speed	32.31	42.10	-0.5
XGB-Dew Point	28.67	35.07	-0.07
XGB-Temperature	29.65	38.92	-0.32
XGB-Pressure	28.30	32.87	-0.05

VI. CONCLUSION

Worldwide, air pollution is responsible for around 1.3 million deaths annually according to the World Health Organization (WHO) [11]. The depletion of air quality is just one of harmful effects due to pollutants released into the air. In this paper, we propose an efficient combined technique that takes the benefits of multi-agent systems, statistical techniques and machine learning techniques for forecasting Air Quality utilizing supervised machine learning procedures. We used Linear Regression and Ada-boosting for improving efficiency of the results. As part of our future work, we will improve our results using various deep learning techniques. Further, we will explore whether adopting lag meteorological variables and tuning the hyper parameters to improve the accuracy of the model.

REFERENCES

[1] Yi X, Zhang J, Wang Z, Li T, Zheng Y. Deep distributed fusion network for air quality prediction. the 24th ACM SIGKDD International Conference: ACM; 2018. <https://doi.org/10.1145/3219819.3219822>.

[2] Lin Y, Mago N, Gao Y, Li Y, Chiang YY, Shahabi C, et al. Exploiting spatiotemporal patterns for accurate air quality forecasting using deep learning. In: The 26th ACM SIGSPATIAL International Conference; 2018. p. 359–68. <https://doi.org/10.1145/3274895.3274907>.

[3] Liao, Q., Zhu, M., Wu, L. et al. Deep Learning for Air Quality Forecasts: a Review. *Curr Pollution Rep* 6, 399–409 (2020). <https://doi.org/10.1007/s40726-020-00159-z>.

[4] Athira V, Geetha P, Vinayakumar R, Soman KP. DeepAirNet: applying recurrent networks for air quality prediction. *Procedia Comput Sci*. 2018;132:1394–403. <https://doi.org/10.1016/j.procs.2018.05.068>.

[5] Raimondo, G.; Montuori, A.; Moniaci, W.; Pasero, E.; Almkvist, E. A Machine Learning Tool to Forecast PM10 Level. In Proceedings of the Fifth Conference on Artificial Intelligence Applications to Environmental Science, San Antonio, TX, USA, 14–18 January 2007; pp. 1–9.

[6] Garcia, J.M.; Teodoro, F.; Cerdeira, R.; Coelho, R.M.; Kumar, P.; Carvalho, M.G. Developing a Methodology to Predict PM10 Concentrations in Urban Areas Using Generalized Linear Models. *Environ. Technol.* 2016, 37, 2316–2325.

[7] Park, S.; Kim, M.; Kim, M.; Namgung, H.-G.; Kim, K.-T.; Cho, K.H.; H, K.; Kwon, S.-B. Predicting PM10 Concentration in Seoul Metropolitan Subway Stations Using Artificial Neural Network (ANN). *J. Hazard. Mater.* 2018, 341, 75–82.

[8] Yu, R.; Yang, Y.; Yang, L.; Han, G.; Move, O.A. RAQ A Random Forest Approach for Predicting Air Quality in Urban Sensing Systems. *Sensors* 2016.

[9] Yi, X.; Zhang, J.; Wang, Z.; Li, T.; Zheng, Y. Deep Distributed Fusion Network for Air Quality Prediction. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, London, UK, 19–23 August 2018; pp. 965–973.

[10] Veljanovska, K.; Dimoski, A. Air Quality Index Prediction Using Simple Machine Learning Algorithms. *Int. J. Emerg. Trends Technol. Comput. Sci.* 2018, 7, 25–30.

[11] Ivo T, Miguel A, Rosaldo R, Eugénio O (2011) Using TraSMAP for developing multi-agent intelligent traffic management solutions. In: Demazeau Y, Pěchouček M, Corchado J, Pérez J (eds) PAAMS 2011: 9th international conference on practical applications of agents and multiagent systems.

[12] Salamanca, Spain, April 2011. *Advances in intelligent and soft computing (advances on practical applications of agents and multiagent systems)*, vol 88, Springer, Heidelberg.

[13] Jin X, Jie L (2012) A study of multi-agent based model for urban intelligent transport systems. *Int J Adv Comput Technol* 4:126–134. doi:10.4156/ijact.vol4.issue6.15.

[14] L. Da, J. Wang, W. Hui Short-term wind speed forecasting based on spectral clustering and optimised echo state networks *Renew Energ*, 78 (2015), pp. 599-608.

[15] L. Xu, Y. Yu, J. Yu, J. Chen, Z. Niu, L. Yin, et al. Spatial distribution and sources identification of elements in PM2.5 among the coastal city group in the Western Taiwan Strait region, *China Sci Total Environ*, 442 (1) (2013), pp. 77-85.

[16] C. Li, Z. Zhu Research and application of a novel hybrid air quality early-warning system: a case study in China *Sci Total Environ*, 626 (2018), pp. 1421-1438.

[17] C. Wang, M. Shao, Q. He, Y. Qian, Y. Qi Feature subset selection based on fuzzy neighborhood rough sets *Knowl-Based Syst*, 111 (2016), pp. 173-179.

[18] S. Wang, Q. Li, H. Yuan, D. Li, J. Geng, C. Zhao, et al. δ -Open set clustering a new topological clustering method *WIREs Data Mining Knowl Discov*, 8 (6) (2018).

[19] S. Yu, S. Chu, C. Wang, Y. Chan, T. Chang Two improved k-means algorithms *Appl Soft Comput*, 68 (2018), pp. 747-755.

[20] Q. Zhang, L.T. Yang, Z. Chen, P. Li High-order possibilistic c-means algorithms based on tensor decompositions for big data in *IoT Inf Fusion*, 39 (2018), pp. 72-80

[21] Majumdar J, Udandakar S, Bai BM. Implementation of cure clustering algorithm for video summarization and healthcare applications in big data. In: *Emerging Research in Computing, Information, Communication and Applications*. Singapore: Springer; 2019. p. 553–64.

[22] <https://www.kaggle.com/rohanrao/air-quality-data-in-india>

[23] Ghufuran Isam Drewil, Riyadh Jabbar Al-Bahadili, Air pollution prediction using LSTM deep learning and metaheuristics algorithms, *Measurement: Sensors Volume 24*, December 2022, 100546.

[24] Abdellatif Bekkar, Badr Hssina, Samira Douzi & Khadija Douzi, Air-pollution prediction in smart city, deep learning approach, *Journal of Big Data volume 8*, Article number: 161 (2021)

[25] Qiuju Xie, Ji-Qin Ni, Enlin Li, Jun Bao, Ping Zheng, Sequential air pollution emission estimation using a hybrid deep learning model and health-related ventilation control in a pig building, *Journal of Cleaner Production*, Volume 371, 15 October 2022, 133714

[26] Gopalakrishnan V (2021) Hyperlocal air quality prediction using machine learning. *Towards data science*. <https://towardsdatascience.com/hyperlocal-air-quality-prediction-using-machine-learning-ed3a661b9a71>.

[27] Sanjeev D (2021) Implementation of machine learning algorithms for analysis and prediction of air quality. *Int. J. Eng. Res. Technol.* 10(3):533–538

[28] Castelli M, Clemente FM, Popovič A, Silva S, Vanneschi L (2020) A machine learning approach to predict air quality in California. *Complexity* 2020(8049504):1–23. <https://doi.org/10.1155/2020/8049504>

[29] Doreswamy HKS, Yogesh KM, Gad I (2020) Forecasting Air pollution particulate matter (PM2.5) using machine learning regression models. *Procedia Comput Sci* 171:2057–2066. <https://doi.org/10.1016/j.procs.2020.04.221>

- [30] Liang Y, Maimury Y, Chen AH, Josue RCJ (2020) Machine learning-based prediction of air quality. *Appl Sci* 10(9151):1–17.
- [31] Madan T, Sagar S, Virmani D (2020) Air quality prediction using machine learning algorithms—a review. In: 2nd international conference on advances in computing, communication control and networking (ICACCCN) pp 140–145.
- [32] Monisri PR, Vikas RK, Rohit NK, Varma MC, Chaithanya BN (2020) Prediction and analysis of air quality using machine learning. *Int J Adv Sci Technol* 29(5):6934–6943.
- [33] Nahar K, Ottom MA, Alshibli F, Shquier MA (2020) Air quality index using machine learning—a jordan case study. *COMPUSOFT, Int J Adv Comput Technol* 9(9):3831–3840.

Fraud Mitigation in Attendance Monitoring Systems using Dynamic QR Code, Geofencing and IMEI Technologies

Augustine Nwabuwe, Baljinder Sanghera, Temitope Alade, Funminiyi Olajide
Department of Computer Science, Nottingham Trent University, Nottingham, NG11 8NS UK

Abstract—Attendance monitoring is a vital activity in several organizations. Due to its importance, many attendance monitoring systems have been developed to automate this process. Despite several advancements in automated attendance management solutions, attendance fraud remains an issue as some end users can manipulate known vulnerabilities, such as proxy attendance, buddy-punching, early departure, and so on. In this paper, a fraud-resistant attendance management solution is developed by harnessing technologies such as geofencing, dynamic QR code and IMEI Checking. The proposed solution is comprised of a single-page web application where QR code can be enabled for attendance registration, and a mobile application, where end-users can scan generated QR code to register their attendance. Attendance cheating via QR code sharing is prevented by encoding the polygonal coordinates of the event venue in the QR code to determine if the user is within the venue. The proposed system solves the problem of proxy attendance by registering and verifying the end user's device IMEI number. Results obtained from testing indicate that attempts at committing a variety of attendance frauds are effectively mitigated.

Keywords—Attendance management systems; fraud prevention; dynamic QR code; geofencing; IMEI verification; software algorithms; mobile application

I. INTRODUCTION

Mobile technologies have changed the way attendance is monitored. Whether it is in schools, colleges, universities or organisations, attendance monitoring systems play a vital role in effective student or staff management. Certain metrics and statistics from attendance data can reveal vital information that can help management identify underlying problems that otherwise may not have been discovered [1]. Attendance monitoring has become ever more important in higher education institutes due to evidence of a positive correlation between attendance, engagement, and academic achievement [2]. Moreover, the UK visas and immigration (UKVI) require UK universities to ensure that all international student visa holders are engaging satisfactorily with their courses. Poor attendance or failure to engage with programmes could lead to being withdrawn from the university and the student visa being curtailed [3]. Despite the importance of attendance, high levels of student non-attendance continues to be a problem in schools and universities [4]. It is also worth noting that with a large student population, effective monitoring of attendance is a problematic task as users constantly look for ways to manipulate attendance monitoring systems, leading to inaccurate and false records. Current electronic techniques for attendance monitoring include the use of quick response (QR) code scan [5], radio frequency identification (RFID) [6], near field communication

(NFC) [7], biometrics technology [8], and global positioning systems (GPS) [9]. Each method has its own pros and cons such as cost, power, ease of deployment and operation, communication, privacy and accuracy, and attendance cheating. For instance, QR code attendance management systems which involve users using their smartphone to scan a QR code are vulnerable to fraud as the generated code can be easily shared amongst colleagues to commit attendance fraud. RFID-based attendance systems where attendance is completed by placing an RFID card on an RFID reader is a common approach. However, a crucial challenge with RFIDs is that students are unable to mark their attendance on days when they have lost or forgotten their ID cards, as RFID tags are embedded in the user's ID card. RFIDs are relatively expensive, and susceptible to buddy-punching, as users can fraudulently help their colleagues mark their attendance [6]. Biometric-based solutions such as face recognition, fingerprint recognition, and voice recognition are convenient as there are no lost cards and replacement problems given that users are essentially their own authentication. Face recognition technology uses image processing techniques to extract features and match a human face from a digital image or video against a database of faces. However, they are prone to the issue of false rejects, mismatch errors, and may not work if a user has an injury or scar [8]. There are also ways to invalidate face recognition systems, such as using a prepared photo or recorded video [10]. Biometric-based solutions are relatively costly and may also expose students' privacy. Furthermore, none of the above techniques effectively capture the problem of early exits and lateness. Tracking and maintaining accurate attendance records remains a challenge. Consequently, it is vital to understand the benefits and challenges of various approaches in order to effectively develop and deploy an attendance monitoring system that is highly resistant to attendance fraud and manipulation. This paper proposes to combine multiple fraud-prevention techniques including geofencing, dynamic QR code and international machine equipment identity (IMEI) checking to increase the reliability of attendance data and significantly reduce the possibility of committing attendance fraud. The main contributions are as follows:

- 1) A mobile application is built to allow users to be authenticated and carry out QR code scanning.
- 2) A single-page web application is developed to generate QR code, manage end user devices and user access, monitor and track attendance activities.
- 3) Geofencing around an area of interest, such as lecture venue or company location is implemented to prevent

end users from registering their attendance outside a specified virtual geographic boundary.

- 4) IMEI checking feature is implemented to ensure that end-users can only register their attendance from their own personal mobile device.
- 5) A new algorithm is created as a final layer of fraud prevention to ensure that generated QR code changes dynamically at set intervals.

The rest of the paper is organised as follows: Relevant related work is presented and analysed in Section II. Details of the system architecture and the core functionality of the proposed attendance monitoring system are discussed in Section III. The system performance under different scenarios of attendance fraud is examined in Section IV, and conclusions are presented in Section V.

II. BACKGROUND AND RELATED WORK

A. Radio Frequency Identification (RFID) based Solutions

Various approaches have been widely explored in literature to automate attendance monitoring in educational settings and organisations. Among these, radio frequency identification (RFID) based attendance monitoring systems [11] – [13], which utilize RFID Tags, also known as a transponder, attached to the identity (ID) card of the individual to be tracked is a popular approach. Students complete their attendance monitoring by placing their RFID card on a RFID reader. In [11], a RFID-based attendance monitoring system is proposed where each student record, and lecture schedules is electronically linked to the RFID tag of the student ID card. The RFID readers are then connected to a server for the information to be stored and processed in a database. Authors in [12] proposed and developed a RFID-based attendance management system with additional functionalities such SMS and email that notify stakeholders including parents and managers when specific metrics are triggered. More recently, RFID-based attendance monitoring systems that combine Internet of things (IoT) technology have also been proposed [13], where IoT devices are used to log, track and fetch attendance data on the cloud and made available for the user anytime and anywhere. Although RFID-based solutions can greatly improve attendance monitoring, the drawbacks are also obvious including but not limited to queuing problems during peak periods, recognition distance, lost or forgotten student ID cards. The cost of deploying RFID readers is relatively high. More critically, RFID-based solutions are susceptible to fraud via buddy-punching, as users could help their colleagues clock in and out of events [14]. As attendance is recorded and checked once, it is not possible to respond to the problem of early departure from scheduled events.

B. Near-Field Communication (NFC) based Solutions

The near-field communication (NFC) empowered attendance system similar in concept to RFID-based solution is proposed in [15], where a strategically located reading device exists to capture access attempts of users with embedded NFC devices, like phones and ID cards. Face identification was incorporated into the NFC-based attendance solution proposed in [16] to create a much more robust system. To enhance the security of NFC-based attendance system, the one-time

passwords (OTP) technology was integrated in [17]. NFC-based solutions are a more cost-effective alternative to RFIDs and are better optimised for power consumption. However, they are also susceptible to the problem of fake attendance or buddy-punching by students and employees.

C. Biometric based Solutions

To overcome the prolonged process of attendance marking associated with ID-based attendance management systems, biometric-based attendance methods are preferred. Biometric-based attendance management systems use users' distinct biological or physiological characteristics such as face, fingerprint, and iris to verify their requests [18]. A typical biometric system will include a reader, software for converting the scanned biometric data into a digital format and a database to store biometric data for future comparison. In [19], images of the user's fingertips are captured, and characteristics such as whorls, arches and loops are recorded. A major challenge with fingerprint-based attendance system is that it cannot recognise a wet or dirty finger [20]. An Iris recognition-based attendance system that utilizes biometric entropy was proposed in [21], where specific eye highlights peculiar to each user were extracted and transformed into a 512-digit Iris Code number. This code was stored as a unique identifier in a database and used to identify users. Face recognition technology is one of the most widely used in biometric-based attendance management systems due to its advantages of greater security, improved accuracy, and capability to easily integrate with other systems [22]. The underlying technologies in face recognition system are based on artificial intelligence and machine learning [23]. A face recognition attendance management system that utilizes the local binary pattern (LBP) algorithm and techniques like bilateral filtration and histogram equalization was proposed in [24]. The technique helped to address some of the issues associated with face recognition accuracy, like varying lighting conditions, image background, and noise in face images. This improved the recognition efficiency to as much as 95%. Biometric-based solutions are convenient; however, they are prone to the issue of false rejects (mismatch errors) and may not work if a user gets injured or scarred [9]. There are also ways to invalidate face recognition system such as using a prepared photo or recorded video [10]. Biometric-based solutions are relatively costly and may expose students' privacy [25].

D. QR Code based and Geolocation based Solutions

In recent years, QR code technology has significantly improved the efficiency and cost of deploying attendance monitoring systems. With the increasing popularity of mobile devices, attendance checking-related applications are deployed on mobile devices, and users can complete attendance checking by scanning a QR code. In [26], a QR code-based attendance management system was proposed which allows users to log on to a web-based application with their mobile phones and complete their attendance by scanning a QR Code generated by the class tutor using the QR code scanner on the web application. Mobile support for QR code makes it cost-effective as no special reading devices or ID cards are required but the problem of an imposter signing in remains. A user can bring the mobile phones of others into the classroom

to complete the attendance checking for them. Furthermore, given that QR code is unencrypted [27], the generated code can be easily shared amongst colleagues to commit attendance fraud. Aiming at solving this problem, location aware QR code attendance management systems have been proposed in [28], where users are required to scan a static QR code generated by the event organizer with a mobile application. Event details and the user’s location are stored in a remote database. Users are not allowed to register their attendance outside a specified virtual geographic boundary, known as a geofence. In [29], a mobile presence control information system that utilize the real-time location capabilities of mobile devices was used to demonstrate the effectiveness and reliability of mobile-based geolocation service. A geolocation-based attendance management system which uses geofencing to create a virtual box representing the classroom was proposed in [30] where the user’s exact location is then retrieved using GPS coordinates from their mobile devices. Other location-based attendance monitoring systems that use static QR Code and geolocation were proposed by [31] and [32]. The shortcoming of location-based attendance checking systems is proxy attendance where users can easily log in with their colleagues’ credentials on their devices and register their attendance.

E. International Mobile Equipment Identity (IMEI) based Solutions

The problem of proxy attendance in geolocation-based attendance management system can be addressed via the IMEI number. IMEI is a unique identification number allocated to each mobile device that serves as a base identification for every mobile phone [33]. It is a fifteen-digit number comprised of a type allocation number (TAC), a serial number and a check digit. It enables each device to be uniquely distinguishable from other mobile devices, as no two mobile phones have the same IMEI numbers. In [33], a solution to monitor the attendance of traffic officers was proposed that utilizes geolocation and IMEI number of the user’s mobile phone to monitor their attendance and effectively reduce cases of attendance cheating. A web-based student presence system that utilizes a combination of QR code and IMEI numbers was proposed in [34]. The IMEI number uniquely identified each student’s mobile device, and every attendance registration attempt was validated against it. Attendance fraud was reduced by ensuring that one student could use only one mobile device.

Therefore, this paper presents the design of a robust fraud-resistant attendance monitoring solution that integrates dynamic QR code, geolocation, geofencing, and IMEI checks to reduce attendance fraud. It is worth noting that no previous studies have explored the combination of these technologies and their potential to effectively mitigate fraud in attendance monitoring systems as illustrated in Table I.

III. PROPOSED ATTENDANCE MONITORING SYSTEM WITH FRAUD MITIGATION

A. System Architecture and Techniques

In this investigation, a robust attendance monitoring system that can effectively guard against fraud is achieved by combining dynamic QR code, geofencing, and IMEI technologies. The system primarily consists of two main parts as shown

TABLE I. COMPARISON OF CURRENT ATTENDANCE MONITORING SOLUTIONS WITH PROPOSED SOLUTION

Features / Solutions	[Proposed]	[28]	[29]	[30]	[31]	[32]
Cross Platform	✓	✓	×	×	×	×
QR Code	✓	✓	×	×	✓	✓
Geolocation	✓	✓	✓	✓	✓	✓
Early Exit Detection	✓	×	×	×	✓	×
IMEI	✓	×	✓	×	×	×
Dynamic QR Code	✓	×	×	×	×	×
Geofencing	✓	×	×	✓	×	×
Manual Fallback option	✓	×	×	×	×	✓

in Fig. 1(a): a cross platform mobile application (app) for registering and collecting user information; and a single-page application (SPA) for administration and data analysis. The mobile app is developed to register the unique IMEI

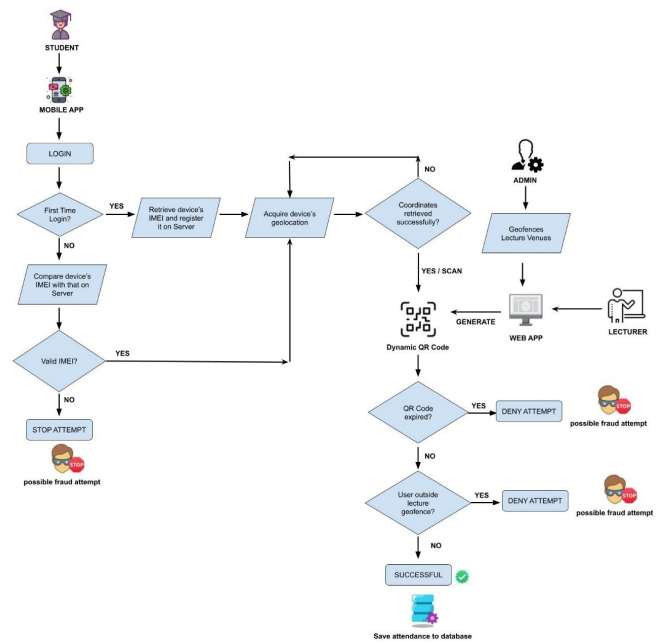


Fig. 1(a). System architecture of the proposed attendance monitoring system with fraud mitigation.

number that precisely identifies the user’s mobile device (UE), detect the entry and exit of a UE by collecting geographical location (geo-location) of the UE every minute, scan a dynamic QR code to record attendance, and upload these data for processing. The SPA is developed to undertake administrative functions such as dynamic QR code generation, user and role management, event management, attendance management and analysis. The system is developed using the MERN (MySQL, Express, React, Node) development architecture as shown in Fig. 1(b), consisting of MySql for storing data, ExpressJS for URL routing and HTTP (hypertext transfer protocol) requests and responses handling, ReactJS for building the system’s user interfaces, and NodeJS as the runtime environment [35].

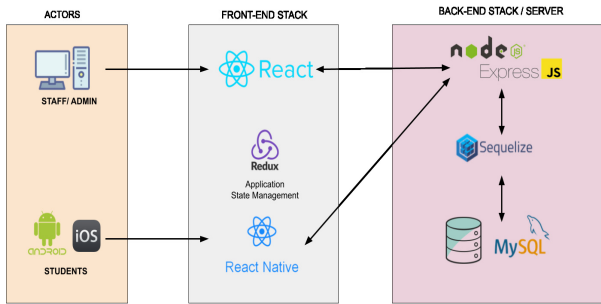


Fig. 1(b). Developmental framework of the proposed attendance monitoring system.

B. System Implementation

The techniques used for fraud mitigation including authentication and authorization, dynamic QR code generation, geofencing and IMEI checking are presented in the following sub-section.

1) *Authentication and authorization:* One of the main components of the proposed system is a robust authentication and authorization mechanism. To enable secure authentication, JavaScript object notation (JSON) web tokens (JWTs) are used, where for a successful login attempt by a user that provides correct credentials, a JWT access token and another JWT refresh tokens is generated and returned to the user agent [36]. To prevent security issues with JWT access tokens, they are stored in the application state managed by the Redux middleware, while the refresh tokens are stored in HTTP-ONLY cookies. This call happens automatically without affecting the user experience. The access tokens are given a short expiration time of 1 hour, while the refresh tokens are given a longer expiration time of 48 hours. The refresh tokens are automatically retrieved and used to generate a new access token whenever the access tokens expire. Fig. 2 presents the JWT authentication design process. The system uses a permission-based architecture [34] to maintain authorization which can be attached to defined roles and assigned to users. This design pattern enables flexible, dynamic, and robust authorization management. Permissions are checked at both the client and server sides to ensure that only authorized users can perform restricted tasks. Unauthorized user attempts at restricted resources are denied.

2) *Geofencing enforcement:* Geofencing is one of the main techniques for fraud mitigation in the proposed system where attendance registration is available only within a defined geofenced area. Users outside of the defined geo-fenced area will not be allowed to register their attendance for a scheduled event. Geofencing is implemented using Google Maps API [37] where each point on the geofence corresponds to a latitude and longitude coordinate as illustrated in Fig. 3. The entire geofence is thus an array of different latitudes and longitudes, forming a polygonal virtual boundary. For a given event, an authorized user must assign a venue in addition to other event information such as delivery method, date and time when creating a timetabled event. The system automatically records latitudes and longitudes to detect when the user enters and

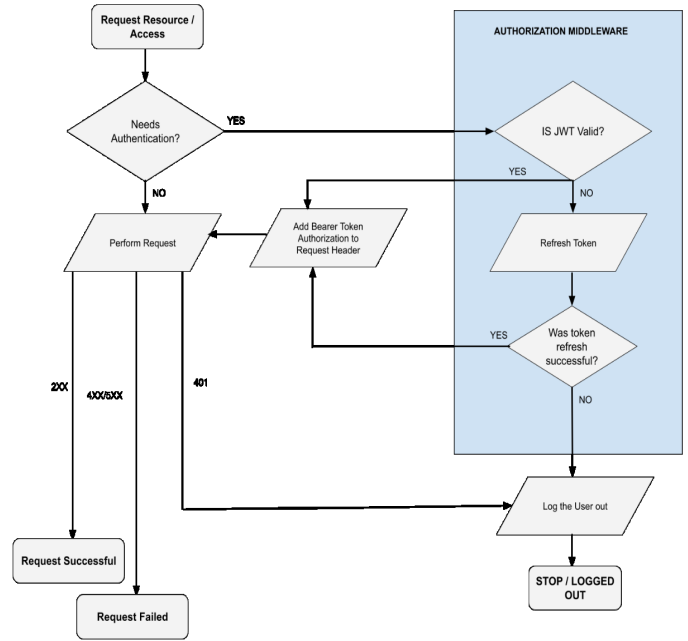


Fig. 2. JWT authentication workflow.

exists the defined geo-fenced area. The system enables the authorized administrative user to disable geofencing for events conducted online. As such, events marked as online will skip the geofence enforcement during attendance registration.

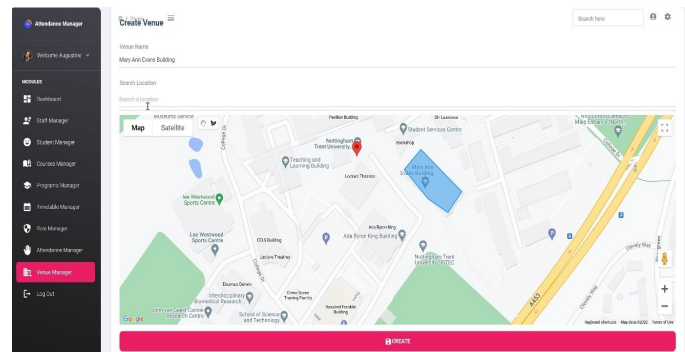


Fig. 3. Geofencing an event venue.

3) *Dynamic QR Code Generation and Scanning:* A critical fraud mitigation component of the proposed attendance monitoring system is dynamic QR code generation. For a given scheduled event, an authorized user such as a tutor can generate and display a two-dimensional QR code for attendance registration. The QR code generated encodes event information such as event title, date, start and end times, and more critically, the geofence polygonal coordinates of the event venue. Given that QR code are not encrypted, the contents of the generated QR code are encoded using base64 encoding format [38]. This makes the encoded content to appear encrypted and serves to prevent fraudulent users who may attempt to scan the QR code using a different QR code scanner. Fig. 4(a) illustrates the encoded data when scanned with a different scanner. A closer examination of the decoded

data shows information separated using slashes, as shown in Fig. 4(b). The individual components of this data are retrieved using JavaScript's split method [36]. The QR code is displayed on a screen where authenticated users can scan it using their UE. The entire process of scanning, decoding and retrieving the QR code encoded data occurs within milliseconds. One of the valuable pieces of information in the decoded data is the geofence polygonal coordinates of the event venue. With this data and the actual geocoordinate of the user, it is possible to determine if the user is within the venue's geofence. The ray casting algorithm [39] is modified and used to solve the problem of identifying whether a point is inside or outside of a polygon, a common geospatial problem in geofencing. Algorithm 1 shows the implementation of this algorithm.

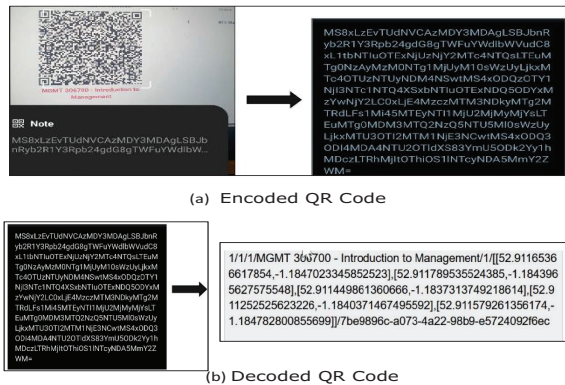


Fig. 4 QR code data.

The QR code generated automatically refreshes every 20s to solve the problem of proxy attendance via QR code sharing. In the unlikely event that geofencing functions are circumvented, the time sensitive QR serves to invalidate any shared QR code. A manual override functionality is implemented to allow an authorized user to manually register the attendance of users who may experience difficulties with the system or have misplaced or forgotten their UE. Once the UE's geo-coordinate is determined to fall within the lecture venue, their information is sent to the server to run final verification checks to determine if the attendance registration request should be approved. The following checks are done before final registration is approved:

- Check if the user is allowed to take the course
- Check if attendance registration is still enabled for the lecture
- Check if the attendance QR code ID has expired
- Check if the student has already marked attendance for the course.

If the above conditions are met, the attendance request is granted, and a successful notification is sent to the mobile app.

4) *IMEI registration and verification:* The proposed system implements IMEI registration and verification technique to solves the issue of proxy attendance which may occur when users log on to the system with their absent colleague's credentials to register attendance for them. When a user lunches

Algorithm 1 : Ray Casting Algorithm for Determining if a Point is Within a Polygon

```

1: Input: p is a simple polygon
2:  $G_i$  is the position of interest
3: Output: true if p contains  $G_i$ , otherwise false
4:  $count = 0$ 
5:  $s$  is an infinite ray in the  $+y$  direction, originating at  $G_i$ 
6: for each edge  $e$  in  $p$  do
7:   if  $G_i$  is within buf of  $e$  then
8:      $e_{x,buf} = e_x - 2 * buf$ 
9:   else
10:     $e_{buf} = e$ 
11:   end if
12:   if  $G_i \prec buf$  of  $e$  or  $e_{buf}$  then
13:     return false
14:   end if
15: end for

```

the mobile app for the first time, they are prompted to confirm whether to register the current UE as the primary device as illustrated in Fig. 5. If confirmed, the IMEI of the UE is requested and stored. This technique prevents fake attendance by ensuring that users can only register their attendance with their verified UE. For subsequent mobile app lunches, the IMEI of the UE is checked and compared against the stored value.

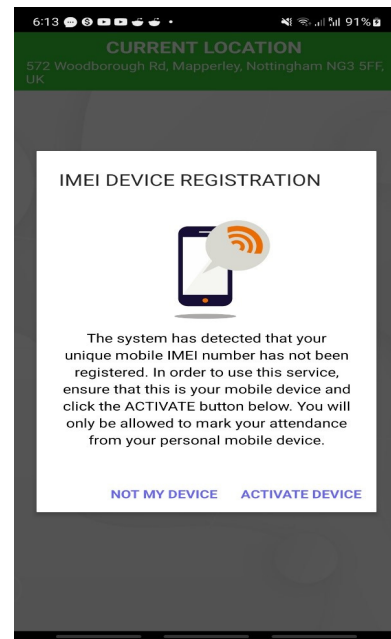


Fig. 5 IMEI registration window.

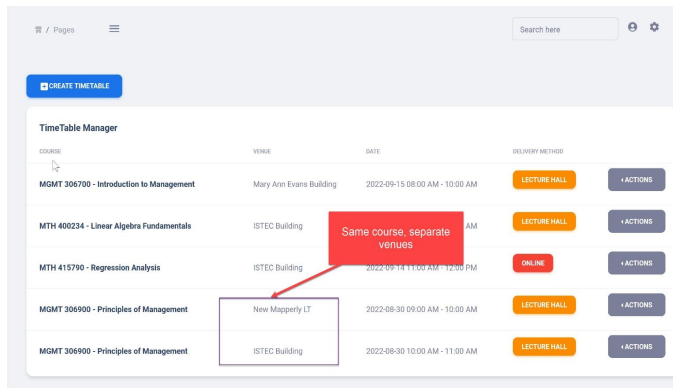
IV. EXPERIMENTAL RESULTS AND DISCUSSION

In this section, the performance of the proposed attendance monitoring system is investigated under three scenarios of attendance restriction. First, the effectiveness of geofencing restriction is tested as an end-user attempts to registering their attendance outside a specified geographic boundary of an event venue. Second, the dynamic QR code restriction is tested in a scenario where a generated QR code is sent to colleagues

in order to commit attendance fraud. Third, the capability of the attendance monitoring system to prevent registering a colleague's attendance from a their own personal UE is tested.

A. Experimental Setup

The web application is deployed on a server equipped with 3 cores and 4GB dedicated memory. The mobile app is deployed on mobile devices equipped with operating systems of Android 10 or later. Event venues named 'New Mapperly LT', 'Mary Ann Evans Building', and 'ISTEC Building' within a university environment, courses of study, event dates and times, are created by an authorized administrative user on the web application and used for testing as illustrated in Fig. 6. Two end-users are asked to register their UEs in order to register their attendance. Geofencing was implemented using the Javascript Google Maps API and location services was enabled on the UEs.



COURSE	VENUE	DATE	DELIVERY METHOD	ACTIONS
MGMT 306700 - Introduction to Management	Mary Ann Evans Building	2022-09-15 08:00 AM - 10:00 AM	LECTURE HALL	+ ACTIONS
MTH 400234 - Linear Algebra Fundamentals	ISTEC Building	2022-09-14 11:00 AM - 12:00 PM	LECTURE HALL	+ ACTIONS
MTH 415790 - Regression Analysis	ISTEC Building	2022-09-14 11:00 AM - 12:00 PM	ONLINE	+ ACTIONS
MGMT 306900 - Principles of Management	New Mapperly LT	2022-08-30 09:00 AM - 10:00 AM	LECTURE HALL	+ ACTIONS
MGMT 306900 - Principles of Management	ISTEC Building	2022-08-30 10:00 AM - 11:00 AM	LECTURE HALL	+ ACTIONS

Fig. 6 Timetable view for courses and event venues.

B. Scenario A: Geofencing Restriction

In order to test the effectiveness of the proposed attendance monitoring system to prevent proxy attendance, two event venues were created and mapped to different geolocations as shown in Fig. 6 to test geofencing restrictions. An end-user located at the New Mapperly LT venue attempts to register their attendance for both events holding at the New Mapperly LT and the ISTEC building venues. It is observed that attendance was successfully registered for the event held at the New Mapperly LT, however, the proxy attendance attempt for event held at the ISTEC building was prevented as shown in Fig. 7.

C. Scenario B: Dynamic QR Code Restriction

In order to assess the effectiveness of the dynamic QR code fraud mitigation mechanism, an end-user captured a generated QR code with their UE and sent it to another user within the same event venue. An attempt was subsequently made to register attendance registration using the shared QR code. It can be seen from the error message shown in Fig. 8 that the system successfully prevents this type of attendance fraud as the QR code generated refreshes every 20s.

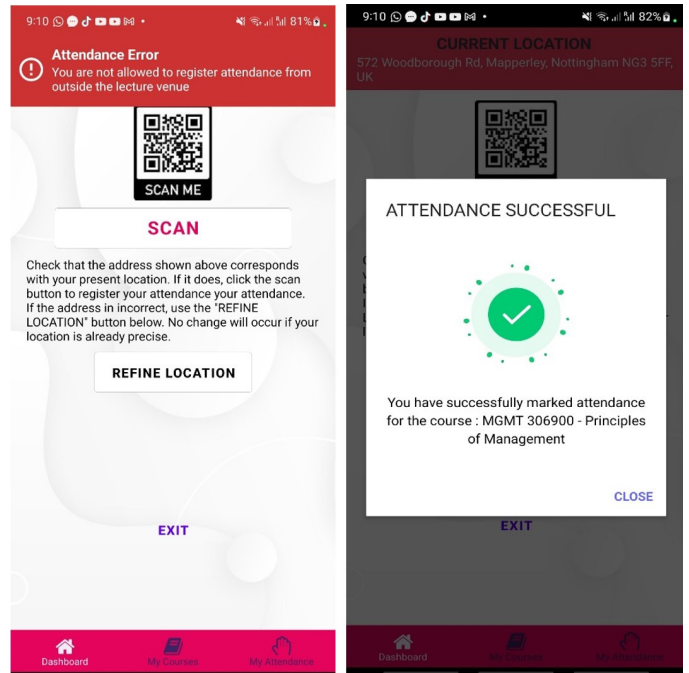


Fig. 7 Effect of geofencing restriction technique.

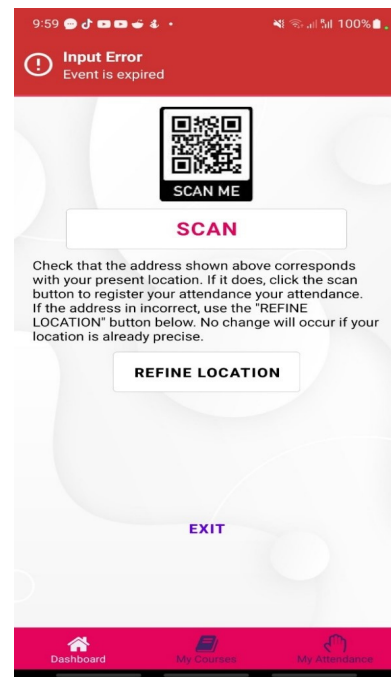


Fig. 8 Effect of dynamic QR code restriction technique.

D. Scenario C: IMEI Restriction

In order to demonstrate IMEI checks restriction, an end-user already registered on the application with their primary device attempts to register their attendance from another user's UE. It can be seen from Fig. 9 that system has effectively mitigated this fraudulent attempt as users are only allowed to register their attendance from their own personal UE.

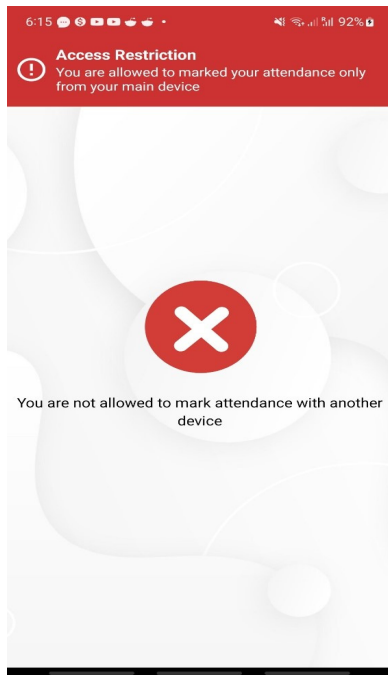


Fig. 9 Effect of IMEI restriction technique.

V. CONCLUSION

This study investigates attendance monitoring systems and proposes a fraud-resistant attendance monitoring system that utilizes a combination of geofencing, dynamic QR code and IMEI checks as preventive techniques to mitigate attendance fraud. A mobile app is developed for registering and collecting user information; and a single-page web app is built for administration and data analysis. Experimental results reveal that the solution effectively mitigates attendance fraud. End-users are effectively and securely authenticated and authorized to use the mobile and web app, and unauthorized user attempts at restricted resources are denied. Geofencing enforcement ensures that attendance registration is only available within a defined geofenced area and solves the proxy attendance problem. The dynamic QR code can only be used within a defined geofenced area, and is used only once within a set time interval to solve the problem of QR code sharing fraud. IMEI restrictions effectively restrict attendance registration to user's own personal UE and effectively eliminates the risk of buddy-punching fraud. This study is of significant importance to institutions and organizations where strict attendance compliance and reliable attendance data are of utmost importance. The efficiency of the application can be fundamentally improved by implementing an in-memory database like Memcached or Redis to cache blacklisted access and refresh tokens. The proposed system can be further enhanced by incorporating machine learning techniques to offer granular visibility into attendance patterns, highlighting those who frequently attempt to commit attendance cheating, as well as those who arrive late and leave scheduled events early. Behaviors like these in attendance data can be duly noted and rectified by the management, upholding engagement, academic achievement and optimal productivity levels in the organization.

REFERENCES

- [1] C. Kearney and J. Childs, *Improving school attendance data and defining problematic and chronic school absenteeism: the next stage for educational policies and health-based practices*, *Preventing School Failure: Alternative Education for Children and Youth*, 2022, DOI: 10.1080/1045988X.2022.2124222.
- [2] K. Alice, S. Sharry, A. Arman, P. Celia, and P. Lillian, *Understanding the impact of attendance and participation on academic achievement*, *Scholarship of Teaching and Learning in Psychology*, 6(4), 272–284, 2020, <https://doi.org/10.1037/stl0000151>.
- [3] M. Ferguson, F. Phiri, *Limitations of attendance monitoring as a singular tool for motivating students' academic engagement: The case study of one overseas student*, *International Journal of Teaching and Education*, Vol. IV(1), pp. 16-25. , 2016, DOI: 10.52950/TE.2016.4.1.002.
- [4] D. Sloan, H. Manns, A. Mellor and M. Jeffries, *Factors influencing student non-attendance at formal teaching sessions*, *Studies in Higher Education*, 45:11, 2203-2216, 2020, DOI: 10.1080/03075079.2019.1599849.
- [5] A. Nuhi, A. Memeti, F. Imeri and B. Cico, *Smart Attendance System using QR Code*, 9th Mediterranean Conference on Embedded Computing (MECO), Budva, Montenegro, 2020, pp. 1-4, doi: 10.1109/MECO49872.2020.9134225.
- [6] D. Mijić, O. Bjelica, J. Durutović and M. Ljubojević, *An Improved Version of Student Attendance Management System Based on RFID*, 18th International Symposium INFOTEH-JAHORINA (INFOTEH), East Sarajevo, Bosnia and Herzegovina, 2019, pp. 1-5, doi: 10.1109/INFOTEH.2019.8717750.
- [7] C. Keau, C.K. On, M. Hijazi, and M. Singh, *Smart-Hadir – Mobile Based Attendance Management System*, *International Journal of Interactive Mobile Technologies (IJIM)*, 15(14), pp. 4–16, 2021, <https://doi.org/10.3991/ijim.v15i14.22677>.
- [8] M. Andrejevic and N. Selwyn, *Facial recognition technology in schools: critical questions and concerns*, *Learning, Media and Technology*, 45:2, 115-128, DOI: 10.1080/17439884.2020.1686014
- [9] Z. Gao et al., *A Student Attendance Management Method Based on Crowdsensing in Classroom Environment*, in *IEEE Access*, vol. 9, pp. 31481-31492, 2021, doi: 10.1109/ACCESS.2021.3060256.
- [10] J. Galbally, S. Marcel and J. Fierrez, *Image Quality Assessment for Fake Biometric Detection: Application to Iris, Fingerprint, and Face Recognition*, in *IEEE Transactions on Image Processing*, vol. 23, no. 2, pp. 710-724, Feb. 2014, doi: 10.1109/TIP.2013.2292332.
- [11] H. U. Zaman, J. S. Hossain, T. T. Anika and D. Choudhury, *RFID based attendance system*, 2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Delhi, India, 2017, pp. 1-5, doi: 10.1109/ICCCNT.2017.8204180.
- [12] Q. Miao, F. Xiao, H. Huang, L. Sun and R. Wang, *Smart attendance system based on frequency distribution algorithm with passive RFID tags*, in *Tsinghua Science and Technology*, vol. 25, no. 2, pp. 217-226, April 2020, doi: 10.26599/TST.2018.9010141.
- [13] T. Sharma and S. L. Aarthy, *An automatic attendance monitoring system using RFID and IOT using Cloud*, 2016 Online International Conference on Green Engineering and Technologies (IC-GET), Coimbatore, India, 2016, pp. 1-4, doi: 10.1109/GET.2016.7916851.
- [14] Q. Y. Tan, P. S. Joseph Ng and K. Y. Phan, *JomRFID Attendance Management System*, 2021 Innovations in Power and Advanced Computing Technologies (i-PACT), Kuala Lumpur, Malaysia, 2021, pp. 1-6, doi: 10.1109/i-PACT52855.2021.9696816.
- [15] M. Mohandes, *Class attendance management system using NFC mobile devices*, *Intelligent Automation and Soft Computing*, 23.2 (2017): 251-259.
- [16] S. U. Masruroh, A. Fiade and I. R. Julia, *NFC Based Mobile Attendance System with Facial Authorization on Raspberry Pi and Cloud Server*, 2018 6th International Conference on Cyber and IT Service Management (CITSM), Parapat, Indonesia, 2018, pp. 1-6, doi: 10.1109/CITSM.2018.8674293.
- [17] J. Jacob, K. Jha, P. Kotak and S. Puthran, *Mobile attendance using Near Field Communication and One-Time Password*, 2015 International Conference on Green Computing and Internet of Things (ICG-CIoT), Greater Noida, India, 2015, pp. 1298-1303, doi: 10.1109/ICG-CIoT.2015.7380666.

- [18] O. Nodirbek, M. Faxriddin, A. Gulzira, and U. Ra'no, *Biometrics authentication: A study*, *ACADEMICIA: An International Multi-disciplinary Research Journal*, vol. 10, (5), 2020. DOI: 10.5958/2249-7137.2020.00374.2
- [19] D. Feng, P. Wang and L. Zu, *Design of Attendance Checking Management System for College Classroom Students Based on Fingerprint Recognition*, *2020 Chinese Control And Decision Conference (CCDC), Hefei, China*, 2020, pp. 555-559, doi: 10.1109/CCDC49329.2020.9164638.
- [20] M. Chandra, F. Feisal, M. Gunawan, F. Gaol and T. Oktavia, *Application of "face recognition" technology for attendance management system*, *Journal of Advances in Information Technology*, vol. 12, (3), pp. 260-266, 2021. DOI: 10.12720/jait.12.3.260-266
- [21] A. K. M Zamin et al, *Design and Implementation of an IRIS Recognition Attendance Management System*, *International Journal of Computer Science Issues*, vol. 15, (4), pp. 64-67, 2018. . DOI: 10.5281/zenodo.1346059.
- [22] S. M. Anzar, N. P. Subheesh, A. Panthakkan, S. Malayil and H. A. Ahmad, *Random Interval Attendance Management System (RIAMS): A Novel Multimodal Approach for Post-COVID Virtual Learning*, in *IEEE Access*, vol. 9, pp. 91001-91016, 2021, doi: 10.1109/ACCESS.2021.3092260.
- [23] A. Chowanda, J. Moniaga, J. C. Bahagiono and J. Sentosa Chandra, *Machine Learning Face Recognition Model for Employee Tracking and Attendance System*, *2022 International Conference on Information Management and Technology (ICIMTech), Semarang, Indonesia*, 2022, pp. 297-301, doi: 10.1109/ICIMTech55957.2022.9915078.
- [24] S. M. Bah and F. Ming, *An improved face recognition algorithm and its application in attendance management system*, *Array*, vol. 5, pp. 100014, 2020.
- [25] T. -C. Li, H. -W. Wu and T. -S. Wu, *The Study of Biometrics Technology Applied in Attendance Management System*, *2012 Third International Conference on Digital Manufacturing and Automation, Guilin, China*, 2012, pp. 943-947, doi: 10.1109/ICDMA.2012.223.
- [26] A. Nuhi, A. Memeti, F. Imeri and B. Cico, *Smart Attendance System using QR Code*, *2020 9th Mediterranean Conference on Embedded Computing (MECO), Budva, Montenegro*, 2020, pp. 1-4, doi: 10.1109/MECO49872.2020.9134225.
- [27] R. Focardi, F. L. Luccio and H. A. M. Wahsheh, *Usable cryptographic QR codes*, *2018 IEEE International Conference on Industrial Technology (ICIT), Lyon, France*, 2018, pp. 1664-1669, doi: 10.1109/ICIT.2018.8352431.
- [28] G. W. Wiriasto, R. W. S. Aji and D. F. Budiman, *Design and development of attendance system application using android-based flutter*, in *2020 Third International Conference on Vocational Education and Electrical Engineering (ICVEE)*, Surabaya, Indonesia, 2020, pp. 1-6, doi: 10.1109/ICVEE50212.2020.9243190.
- [29] S. Ríos-Aguilar and F. Lloréns-Montes, *A Mobile Business Information System for the Control of Local and Remote Workforce through Reactive and Behavior-based Monitoring*, *Expert Systems with Applications*, vol. 42, (7), pp. 3462-3469, 2015, ISSN 0957-4174, doi: https://doi.org/10.1016/j.eswa.2014.12.030.
- [30] A. Morankar, R. Baviskar, R. Vishwakarma, S. Patil, and N. Ujgare, *Geolocation Based College Attendance System*, *International Research Journal of Modernization in Engineering Technology and Science*, Vol 03, Issue 04, April-2021, e-ISSN: 2582-5208.
- [31] Z. Ayop, C. Lin, S. Anawar, E. Hamid, and M. Azhar *Location-aware event attendance system using QR code and GPS technology*, *International Journal of Advanced Computer Science and Applications*, vol. 9(9), pp. 466-473, 2018, doi: https://10.14569/IJACSA.2018.090959.
- [32] H. Elbehery, *Enhancement of QR code Student's Attendance Management System using GPS*, *IOSR Journal of Computer Engineering (IOSR-JCE)*, vol. 21(4), pp. 18-30, 2019.
- [33] A. B. Nasution, *Traffic Officers Attendance System Design Using GPS and IMEI Smartphone*, *Infokum*, vol. 10, (1), pp. 206-214, 2021.
- [34] N. Hermanto and W. M. Baihaqi, *Implementation of QR code and imei on android and web-based student presence systems*, in *2018 3rd International Conference on Information Technology, Information System and Electrical Engineering (ICITISEE)*, Yogyakarta, Indonesia, 2018, pp. 276-280, doi: 10.1109/ICITISEE.2018.8721009.
- [35] N. Kozma and D. Krstić, *Design of Information System for Bookstore support Student paper*, *2022 21st International Symposium INFOTEH-JAHORINA (INFOTEH)*, East Sarajevo, Bosnia and Herzegovina, 2022, pp. 1-6, doi: 10.1109/INFOTEH53737.2022.9751271.
- [36] I. Darmawan, A. P. A. Karim, A. Rahmatulloh, R. Gunawan and D. Pramesti, *JSON Web Token Penetration Testing on Cookie Storage with CSRF Techniques*, *2021 International Conference Advancement in Data Science, E-learning and Information Systems (ICADEIS)*, Bali, Indonesia, 2021, pp. 1-5, doi: 10.1109/ICADEISS2521.2021.9701965.
- [37] R. Shinde, A. Nilose and P. Chandankhede, *Design and Development of Geofencing Based Attendance System for Mobile Application*, *2022 10th International Conference on Emerging Trends in Engineering and Technology - Signal and Information Processing (ICETET-SIP-22)*, Nagpur, India, 2022, pp. 1-6, doi: 10.1109/ICETET-SIP-2254415.2022.9791781.
- [38] A. Azizi, Y. Yusof, and F. Ahmad, *Expanding the data capacity of QR codes using multiple compression algorithms and base64 encode/decode*, *Journal of Telecommunication, Electronic and Computer Engineering (JTEC)* 9.2-2 (2017): 41-47.
- [39] Y. Ye, F. Guangrui and O. Shiqi, *An Algorithm for Judging Points Inside or Outside a Polygon*, *2013 Seventh International Conference on Image and Graphics, Qingdao, China*, 2013, pp. 690-693, doi: 10.1109/ICIG.2013.140.

Reverse Supply Chain Management Through a Quantity Flexibility Contract: A Case of Stochastic Remanufacturing Capacity

Changhao Zhang

Henan University of Animal Husbandry and Economy,
Zhengzhou, Henan, 450000, China

Abstract—This article investigates a two-echelon reverse supply chain (RSC) where a third-party logistics provider charges customers to return outdated products. A green manufacturer refurbishes qualified returned products through the remanufacturing process. Remanufacturing capacity is considered a stochastic variable. Under the volatility of remanufacturing capacity, some likely examined, and qualified products could not be remanufactured. If a collected product cannot be processed, it should be salvaged at a lower value and be perceived as a lost profit. In such scenarios, increasing the quantity of returned outdated products is suitable if there is a strong possibility of enough capacity in the remanufacturing process. This paper develops a stochastic model to identify the optimal order quantity under diverse contracts, including wholesale price, centralized, and quantity flexibility contracts. Under the quantity flexibility contract, the green supplier might cancel its preliminary order in a restricted quantity. Additionally, third-party logistics supplier offers a restricted quantity above the initial order to minimize understocking during peak seasons. Our numerical experiments demonstrate that the suggested quantity flexibility can coordinate the examined RSC under the volatility of remanufacturing capacity. Contrary to wholesale and centralized contracts, quantity flexibility is a more practical alternative from the perspective of participants' profitability.

Keywords—Reverse supply chain; channel coordination; uncertain remanufacturing capacity; quantity flexibility contract

I. INTRODUCTION

A reverse supply chain (RSC) comprises a series of actions that collect obsolete products from customers and return them to the original manufacturer or recycler for reprocessing or proper disposal [1]–[3]. The application of the RSC has financial, social, and environmental benefits. One of the fundamental obligations of a company to recycle its commodities is setting up an RSC; Xerox, Nike, Adidas, Sony, and Siemens are successful firms that built up their RSC [4]. Moreover, Consumers' sensitivity to environmental problems has driven corporations to highlight their recycling potential [5].

Typically, young customers in South Asia are aware of the eco-friendliness of a product's components; in some regions of the Middle East, this figure approaches 80 percent. Due to the increased environmental awareness among consumers and government intervention, many businesses have begun creating green goods [6]. Numerous goods are suitable for

remanufacturing or recycling. Electronic items, for instance, are among the most ideal for recycling due to their short life cycle, modular design, and the kind of raw material utilized [7]. For instance, 76 percent of a camera's components may be used multi times [8]. As a result, companies now put their efforts into green operations, specifically RSC, fully take advantage of their benefits.

Since 2020, the COVID-19 scenario has wreaked havoc on SCs; it produces extreme supply and demand oscillations, disrupting the corporate system. Upstream participants cannot foresee demand, and downstream participants cannot fulfill their responsibilities [9], [10]. A measure we can take for this problem is adopting coordination mechanisms [11]. SC members can set up coordinated contracts to make an SC that works well together. For example, a study by Bakhshi and Heydari (2021) demonstrated the value of investigating an option contract for coordinating the interaction between an e-retailer and a 3PL; they also compared the investigated contract with a penalty-based contract; the results show that the option contract incredibly increases the SC's total profit in comparison with other contracts. Some examples of these kinds of contracts are quantity flexibility contracts. Expressly, a great deal of research work has confirmed their applications in inventory management problems, and their findings have revealed that they can handle uncertainty well [12], [13]. Quantity flexibility contract has many capabilities and has been validated in various applications. For instance, Kord and Samouei (2023) studied a quantity flexibility contract for managing a humanitarian supply chain to buy a spot market under demand uncertainty. The results indicated the powerful performance of NPGA in terms of most evaluation indicators. The results indicated the powerful performance of NPGA in terms of most evaluation indicators.

Based on a quantity flexibility contract, a buyer can adjust its initial purchase up or down within a specified volume range. Thus, the buyer must purchase a minimum quantity, while the supplier must provide additional amounts if necessary [5], [14]. The industry can improve numerous real-world instances by implementing quantity flexibility contracts. Sun Microsystems purchases its workstations through QF contracts. Nippon Otis, a maker of elevator equipment, utilizes a quantity flexibility contract with the Tsuchiya plant, a manufacturer of components and switches. Toyota Motor Corporation, IBM, and Hewlett-Packard have also utilized quantity flexibility contracts [15].

This study aims to determine an effective strategy for stabilizing 3PL's position in an RSC. Meanwhile, this work highlights the necessity of optimal ordering decisions. In this respect, we try to answer the following questions:

QS1: What is the GS's optimal order quantity within concluded contracts?

QS2: Can the suggested quantity flexibility contract enhance the parties' profitability?

QS3: Which contract is preferred?

We evaluate a two-echelon RSC involving a green supplier who can refurbish eligible outdated products and a third-party logistics provider that collects such products from customers despite the unpredictability of remanufacturing capacity. We investigate the impact of a quantity flexibility contract in which the third-party logistics provider is the Stackelberg game leader. Under the quantity flexibility contract, the green supplier is permitted to reserve items without incurring a reservation charge and must buy a minimum amount. In contrast, the third-party logistics provider will be needed to collect additional products if required.

To the best of our knowledge, prior research has focused primarily on the significance of quantity flexibility contracts in forwarding supply chains with uncertain demand; in contrast, the current study investigates the impact of the quantity flexibility contract on an RSC with uncertain remanufacturing capacity. In other words, contrary to earlier studies, uncertainty has migrated from demand to supply. We based our model on this gap in the research.

The remainder of the article is categorized as follows: Our model setting and assumptions are illustrated in detail in Section II. Section III derives the optimum decisions in a variety of models. In Section IV, numerical analysis is used to validate the models. Section V reveals managerial insights and conclusions.

II. MODEL SETTING

A. RSC Structure

This study examines a two-echelon RSC, which includes a green supplier (GS) and a third-party logistics provider (3PL). GS recycles outdated products qualified enough to return to the market. However, the remanufacturing capacity on the GS side is susceptible to uncertainty. On the other hand, 3PL collects outdated products from consumers and returns the acceptable ones to GS after a final assessment. GS makes profits by selling remanufactured products directly through the market. Indeed, GS purchases outdated products from 3PL at the cost of c_{st} /unit and, after the recycling process, sells recycled products at the price of w /unit into the market.

In the current RSC, GS sets an order amount for 3PL and determines the order quantity. However, 3PL may not provide the whole order quantity because of the ineligible collected product. Accordingly, all qualified orders are shipped directly to GS. To sum up, GS may get fewer amounts of the placed ordered products. The parameters used in this study are summarized in Table I.

TABLE I. LISTED PARAMETERS AND DECISION VARIABLES

Parameters	Description
m	A continuous random variable with distribution function $g(\cdot)$ and cumulative function $G(\cdot)$, standing for remanufacturing capacity.
c_{st}	The unit fee paid by GS to 3PL to collect outdated products.
c_i	Inspection fee per unit of collected products at the 3PL site.
α	The clearance percentage of products inspected by 3PL.
c_m	Inventory cost of amassed products collected by 3PL.
c_r	The unit cost of refurbishing at the GS's site
c_p	Cost per unit of preparation for refurbishment by the GS.
c_s	Unit shipping cost for 3PL.
c_{tc}	Reward offered by the 3PL for the return of each outdated product.
w	The wholesale price of refurbished products.
s	The value of salvage per unit for the 3PL.
d	Downward adjustment parameter in the quantity flexibility contract.
u	Upward adjustment parameter in the quantity flexibility contract.
Decision variables	
q_w	GS's order amount for 3PL within the wholesale price contract.
q_c	GS's order amount for 3PL within the centralized contract.
q_f	GS's order amount for 3PL within the quantity flexibility contract.

B. Assumptions

The primary purpose of this study is to highlight the significance of optimum ordering/pricing choices and the management of the possible costs of overstocking or understocking in the analyzed RSC under uncertainty. Hence, the generated models include the following modeling assumptions:

Assumption 1. GS’s remanufacturing capacity is contaminated by uncertainty.

Based on Assumption 1, x is a continuous random variable with a distribution function $g(x)$ and cumulative function $G(x)$ [4].

Assumption 2. 3PL rewards customers with a per-unit fee for the return of outdated products.

Assumption 3. 3PL delivers a definite proportion of qualified collected products to GS after the final assessment.

According to the third assumption, all of the obsolete products collected by 3PL are ineligible for recycling. On the 3PL side, it is thus essential to analyze gathered products and detect discarded ones [16].

Assumption 4. Customers can buy all recycled products at a pre-set price w .

III. MATHEMATICAL MODELING

Suppliers need an ordering system to control the production/inventory process; thus, they pressure downstream partners to place their orders before the selling season. Nonetheless, uncertainties raise overstock/understock risks and discourage partners from setting early orders. Ordering in SCs may be resolved by concluding a suitable contract between two parties. Offering a reservation policy is a measure we can take to encourage GS to purchase more/earlier. To solve the problem, this research compares a regular wholesale pricing contract with a quantity flexibility contract [5] (Appendix).

A. Wholesale Price Contract

Under the wholesale price contract, each participant decides independently of the other participants’ interests. GS determines the order quantity placed for 3PL to optimize its profit, and 3PL inspects the collected products. Then, GS refurbishes the received products following the acknowledged remanufacturing capability [17]. Based on the order quantity and realized remanufacturing capacity, the profit function of GS is formulated as follows:

$$\Pi_{GS}(q_w) = \begin{cases} (w - c_r)m - (c_{st} + c_p)\alpha q_w + s(\alpha q_w - m) & \alpha q_w > m \\ (w - c_r - c_{st} - c_p)\alpha q_w & \alpha q_w < m \end{cases} \quad (1)$$

According to Eq. (1), only m units of products are remanufactured and sold if there is inadequate remanufacturing capacity; the remaining products are salvaged at a value of s ($(\alpha q_w - m) s$). Consequently, the remanufacturer’s profit function comprises income from selling m units on the market and salvage value. Moreover, A unit cost will be paid to 3PL

for each unit received, which under any condition would be αq_w . Now, the expected profit function of GS is:

$$E(\Pi_{GS}(q_w)) = (w - c_r) \left(\int_0^{\alpha q_w} m g(m) dm + \int_{\alpha q_w}^{\infty} \alpha q_w g(m) dm \right) - (c_{st} + c_p)\alpha q_w + \int_0^{\alpha q_w} s(\alpha q_w - m) g(m) dm \quad (2)$$

Proposition 1. $E(\Pi_{GS}(q_w))$ is concave in q_w and q_w^* That maximizes GS’s profit function will be calculated as follows:

$$q_w^* = \frac{G^{-1}\left(\frac{w - c_r - c_{st} - c_p}{w - c_r - s}\right)}{\alpha} \quad (3)$$

Contingent on Eq. (3), the cost of c_{st} is a criterion to determine the trade-off between overstock and understock costs. In this respect, the overstock cost is $c_{st} + c_m - w$ and the understock cost will be $w - c_r - c_{st} - c_p$. Now let us show the 3PL profit function, which is obtained as follows:

$$\Pi_T = \left(c_{st} - c_{tc} - \frac{c_i}{\alpha} - c_m - c_s \right) \alpha q_w \quad (4)$$

The profit function of 3PL reflects the revenue generated by selling outdated products to GS and the associated expenses, such as the cost of returning products, inspection, holding, and shipment. Now, we go further and investigate the centralized contract in the subsequent sections.

B. Centralized Contract

Under the centralized decision-making system, we attempt to maximize the RSC’s overall profit. Under the centralized scenario, the predicted profit function of RSC may be computed as follows:

$$\Pi_{RSC}(q_c) = \begin{cases} (w - c_r)m - c_p \alpha q_c + s(\alpha q_c - m) + \left(-c_{tc} - \frac{c_i}{\alpha} - c_m - c_s\right) \alpha q_c & \alpha q_c > m \\ (w - c_r - c_p)\alpha q_c + \left(-c_{tc} - \frac{c_i}{\alpha} - c_m - c_s\right) \alpha q_c & \alpha q_c < m \end{cases} \quad (5)$$

In Eq. (5), 3PL reviews all returned outdated items and accepts quantities that can be refurbished; following storage, the qualified outdated products are dispatched to GS. Then, GS prepares the items for refurbishing. Consequently, when there is insufficient remanufacturing capacity, as indicated by the expression $\alpha q_c > m$, only m items may be introduced into the remanufacturing process. Accordingly, the total expected profit function of RSC is determined by the following:

$$E(\Pi_{RSC}(q_c)) = (w - c_r) \left(\int_0^{\alpha q_c} m g(m) dm + \int_{\alpha q_c}^{\infty} \alpha q_c g(m) dm \right) - c_p \alpha q_c + \int_0^{\alpha q_c} s(\alpha q_c - m) g(m) dm + \left(-c_{tc} - \frac{c_i}{\alpha} - c_m - c_s\right) \alpha q_c \quad (6)$$

Proposition 2. $E(\Pi_{RSC}(q_c))$ is concave in q_c and q_c^* that maximizes RSC’s profit function will be calculated as follows:

$$q_c^* = \frac{G^{-1}\left(\frac{w - c_r - c_p - c_{tc} - \frac{c_i}{\alpha} - c_m - c_s}{w - c_r - s}\right)}{\alpha} \quad (7)$$

It should be mentioned that the primary insight gained from Proposition 2 is that $G(q_c)$ belongs to a newsvendor concern, i.e., the trade-off acquired between overstock and understock

costs is conducted depending on $(c_{tc} + \frac{c_i}{\alpha} + c_m + c_s)$ in the situation of volatile remanufacturing capacity.

C. Quantity Flexibility Contract

This section examines the scenario in which GS and 3PL implement the quantity flexibility contract. The 3PL is the leader, while the GS is the follower. The following is the order of events under this contract: (1) An offer is made for a contract with parameters (w, q_f, d, u) . w is the wholesale price following the realization of remanufacturing capacity. $0 \leq d \leq 1$ is a parameter that determines the acceptable range and gauges the flexibility. $u \geq 0$ is the upward adjustment parameter. (2) Given the contract, GS determines the reservation quantity q_f . Therefore, the permissible range is set as $[daq_f, (1+u)\alpha q_f]$. Note that GS is not required to pay the reservation fee. (3) After observing the reserved quantity, 3PL gathers outdated products at least equal to the lower limit of the permissible range, i.e., daq_f . (4) The M capacity for remanufacturing is realized. (5) 3PL delivers products to GS in accordance with M's realized capacity and after the final examination. According to the order quantity q_f and realized remanufacturing capacity, the profit function of GS will be:

$$\Pi_{GS}(q_f) = \begin{cases} wm + (-c_r - c_p)m - c_{st}daq_f + s(daq_f - m) & 0 < m < daq_f \\ wm + (-c_r - c_{st} - c_p)m & daq_f < m < (1+u)\alpha q_f \\ (w - c_r - c_{st} - c_p)(1+u)\alpha q_f & m > (1+u)\alpha q_f \end{cases} \quad (8)$$

In the first condition (i.e., $0 < m < daq_f$) of Eq. (8), the first term represents the total revenue from selling m units, the second and third terms stand for operation costs related to purchasing and refurbishing obsolete products, and the third term means the total revenue from salvaging unsold products. In the second condition ($daq_f < m < (1+u)\alpha q_f$), the first term represents the entire income from selling m units, and the second reveals the operation costs. Finally, based on the third condition ($m > (1+u)\alpha q_f$) GS sells $(1+u)\alpha q_f$ units in the market, which is the highest band of purchased products according to the concluded contract. In this regard, the expected profit function of GS will be determined as follows:

$$\begin{aligned} E(\Pi_{GS}(q_f)) &= (w - c_r - c_p) \left(\int_0^{(1+u)\alpha q_f} mg(m)dm \right. \\ &\quad + \int_{(1+u)\alpha q_f}^{\infty} (1+u)\alpha q_f g(m)dm \Big) \\ &\quad - c_{st} \left(\int_0^{daq_f} daq_f g(m)dm \right. \\ &\quad + \int_{daq_f}^{(1+u)\alpha q_f} mg(m)dm \Big) \\ &\quad + \int_{(1+u)\alpha q_f}^{\infty} (1+u)\alpha q_f g(m)dm \\ &\quad + \int_0^{daq_f} s(daq_f - m)g(m)dm \end{aligned} \quad (9)$$

The following Proposition will obtain the optimum amount of GS's order under the quantity flexibility contract.

Proposition 3. $E(\Pi_{GS}(q_f))$ is concave in q_f and q_f^* that maximizes GS's profit function will be calculated by solving the following equation:

$$\begin{aligned} (w - c_r - c_p - c_{st})(1+u)\alpha^2 q_f G((1+u)\alpha q_f) \\ - c_{st}(d\alpha)^2 q_f \left(1 - G((1+u)\alpha q_f) \right) \\ + s(d\alpha)^2 q_f G(daq_f) = 0 \end{aligned} \quad (10)$$

d is a crucial aspect of the quantity flexibility contract. Therefore, as d increases, GS decreases its order quantity. In fact, by increasing d , GS becomes reluctant to place an order with 3PL. Now let us show the 3PL profit function, which is established as follows:

$$\Pi_T = \begin{cases} (c_{st} - c_s)daq_f - (c_{tc} + \frac{c_i}{\alpha} + c_m)(1+u)\alpha q_f + s((1+u)\alpha q_f - daq_f) & 0 < m < daq_f \\ (c_{st} - c_s)m - (c_{tc} + \frac{c_i}{\alpha} + c_m)(1+u)\alpha q_f + s((1+u)\alpha q_f - m) & daq_f < m < (1+u)\alpha q_f \\ (c_{st} - c_{tc} - \frac{c_i}{\alpha} - c_m - c_s)(1+u)\alpha q_f & m > (1+u)\alpha q_f \end{cases} \quad (11)$$

As shown in Eq (11), when GS's remanufacturing capacity m is lower than daq_f , it purchases its minimum promised volume. In such circumstances, GS must salvage unsold inventory at the end of the selling season.

IV. NUMERICAL ANALYSIS

We explore various numerical experiments to show the recommended model's effectiveness. The parameters for two numerical examples are presented in Table II. The datasets utilized in the examples meet all the examined assumptions and models' criteria. The employed data are compatible with datasets applied in the prior research, mainly produced based on genuine instances. In addition, the selected values are relatively broad to be utilized for numerous items in the remanufacturing and refurbishment sector, although by scaling and specific adjustments [18] Meanwhile, equations acquired for the optimal order values are all closed-form relations that there is no need for software to solve. This study assumes that remanufacturing capacity follows a normal probability distribution $N(\mu, \sigma)$ in the first case and a uniform probability distribution $U(T1, T2)$ in the second. Table III displays the numerical findings, which include decision variables and profit functions.

In the first case, the tremendous potential of the channel is attained when all partners are handled through an interconnected solution, i.e., a centralized contract. Nevertheless, each partner seeks to advance its interests via decentralized contracts; hence, Table III indicates that the quantity flexibility contract enables all RSC participants to function nearly as a single unit. This incentive-based method has enhanced the number of orders and produced much more flexibility than the decentralized contract — the quantity flexibility contract functions as a risk-sharing strategy. Specifically, GS and 3PL share the risk of unpredictable remanufacturing capacity. By utilizing a reserving approach, GS can order obsolete items from GS with greater flexibility, increasing order volume and decreasing the likelihood of experiencing capacity instability. Results indicate that GS and 3PL will gain profit about equally if the deal is accepted.

Table IV illustrates the effect of changing d on profit functions and the GS's order quantity. Based on Table IV, as expected, by decreasing d , GS's profit under the quantity flexibility contract will increase. On the other hand, reducing d decreases the profit of 3PL under quantity flexibility.

TABLE II. AMOUNTS OF PARAMETERS

Parameters	First example	Second example
c_{st}	50	80
c_i	8	16
α	0.88	0.92
c_m	4	6
c_r	15	20
c_p	5	7
c_s	3	6
c_{tc}	20	30
w	80	110
s	6	8
d	0.7	0.8
u	0.9	1.5
m	$\sim N(\mu = 135, \sigma = 40)$	$\sim U(200,400)$

TABLE III. OPTIMAL EQUILIBRIUMS WITHIN DIFFERENT CONTRACTS

	Contracts	q^*	Π_{GS}^*	Π_T^*	Π_{RSC}^*
First Example	Wholesale price	140.11	1309.11	1715.58	3024.69
	Centralized	170.44	1400.87	2370.82	3771.69
	Quantity flexibility	216.88	1755.93	3567.97	5323.9
Second Example	Wholesale price	161.23	1507.66	3453.15	4960.81
	Centralized	194.41	1666.77	4163.79	5830.56
	Quantity flexibility	244.86	1903.21	5123.33	7026.54

TABLE IV. EQUILIBRIUMS SENSITIVITY WHEN d VARIES

d	Contracts	q^*	Π_{GS}^*	Π_T^*	Π_{RSC}^*
$d = 0.6$	Wholesale price	140.11	1309.11	1715.58	3024.69
	Centralized	170.44	1400.87	2370.82	3771.69
	Quantity flexibility	230.11	1833.60	3213.42	5047.02
$d = 0.55$	Wholesale price	140.11	1309.11	1715.58	3024.69
	Centralized	170.44	1400.87	2370.82	3771.69
	Quantity flexibility	236.77	1864.22	3013.37	4877.59
$d = 0.5$	Wholesale price	140.11	1309.11	1715.58	3024.69
	Centralized	170.44	1400.87	2370.82	3771.69
	Quantity flexibility	241.88	1891.92	2892.99	4784.91

V. CONCLUSION AND FUTURE STUDIES

In this study, a reverse supply chain is built to address the problems of a stochastic remanufacturing capacity. Initially, a decentralized contract is developed in which the 3PL gives consumers a cost charge for returning outdated items. In the subsequent phase, we model a centralized contract to obtain a globally optimum solution. However, the unpredictability of the upstream remanufacturing capacity is complicated from a decentralized to a centralized contract. Developing a quantity flexibility contract that allows the green supplier to cancel its preliminary order in a restricted quantity. Additionally, 3PL offers a restricted quantity above the initial order to prevent understocking during peak seasons. Numerical examples and sensitivity analysis indicate that the suggested quantity flexibility contract not only enhances the economic

performance of each SC participant but also produces a Pareto-improving condition in which both SC participants earn higher profitability. We further demonstrate that when the quantity flexibility contract's downward adjustment parameter is low, the quantity flexibility contract is profitable for both 3PL and supplier. Thus, both prefer to participate in the quantity flexibility contract channel.

We were excused from investigating several limitations: First, it would be interesting to consider both the remanufacturing capacity and demand uncertainty and compare the results with this study. Another limitation is that we were also excused from comparing the proposed reservation-based contract (quantity flexibility) with another, such as an option contract or even a penalty-based contract.

Future research opportunities in reverse supply chain coordination might include analyzing the simultaneous effects of numerous sources of uncertainty. Another expansion may concentrate on the competitive context and use game theory to address the issue in competitive settings. In addition, how outdated items are gathered is an essential aspect of reverse logistics. Hence, this perspective might be the subject of intriguing research. Customers will be more inclined to return their used items for a discount if they can easily store out-of-date merchandise. In this study, the desire of consumers to return their outdated items is assumed to be a deterministic linear function of the provided incentive amount; however, stochastic and nonlinear processes can also be considered. Lastly, exploring the potential for outsourcing capacity and contemplating a backup manufacturer for the existing model might be examined as part of future relevant and intriguing work.

REFERENCES

[1] N. K. Dev, R. Shankar, and F. H. Qaiser, "Industry 4.0 and circular economy: Operational excellence for sustainable reverse supply chain performance," *Resour Conserv Recycl*, vol. 153, p. 104583, 2020.

[2] C. Garrido-Hidalgo, T. Olivares, F. J. Ramirez, and L. Roda-Sanchez, "An end-to-end internet of things solution for reverse supply chain management in industry 4.0," *Comput Ind*, vol. 112, p. 103127, 2019.

[3] D. Wu, J. Chen, P. Li, and R. Zhang, "Contract coordination of dual channel reverse supply chain considering service level," *J Clean Prod*, vol. 260, p. 121071, 2020.

[4] J. Heydari and M. Ghasemi, "A revenue sharing contract for reverse supply chain coordination under stochastic quality of returned products and uncertain remanufacturing capacity," *J Clean Prod*, vol. 197, pp. 607–615, 2018.

[5] J. Li, X. Luo, Q. Wang, and W. Zhou, "Supply chain coordination through capacity reservation contract and quantity flexibility contract," *Omega (Westport)*, vol. 99, p. 102195, 2021.

[6] M. Dehghan-Bonari, A. Bakhshi, A. Aghsami, and F. Jolai, "Green supply chain management through call option contract and revenue-sharing contract to cope with demand uncertainty," *Cleaner Logistics and Supply Chain*, vol. 2, p. 100010, 2021.

[7] A. Mann, P. Saxena, M. Almani, O. Okorie, and K. Saloni, "Environmental Impact Assessment of Different Strategies for the Remanufacturing of User Electronics," *Energies (Basel)*, vol. 15, no. 7, p. 2376, 2022.

[8] J. Heydari, K. Govindan, and R. Sadeghi, "Reverse supply chain coordination under stochastic remanufacturing capacity," *Int J Prod Econ*, vol. 202, pp. 1–11, 2018.

[9] K. Xie, S. Zhu, and P. Gui, "A Game-Theoretic Approach for CSR Emergency Medical Supply Chain during COVID-19 Crisis," *Sustainability*, vol. 14, no. 3, p. 1315, 2022.

[10] H. A. Mahdiraji, A. A. Kamardi, M. Beheshti, S. H. R. Hajiagha, and L. Rocha-Lona, "Analysing supply chain coordination mechanisms dealing with repurposing challenges during Covid-19 pandemic in an emerging economy: a multi-layer decision making approach," *Operations Management Research*, pp. 1–20, 2022.

[11] S.-M. Hosseini-Motlagh, M. Nematollahi, and S. Ebrahimi, "Tri-party reverse supply chain coordination with competitive product acquisition process," *Journal of the Operational Research Society*, vol. 73, no. 2, pp. 382–393, 2022.

[12] Z. Yuan, Y. Gong, and M. Chen, "Quantity-Flexibility Contract Models for the Supply Chain with Green-Sensitive Demand in the Automotive Manufacturing Industry," in *IFIP International Conference on Advances in Production Management Systems*, 2021, pp. 441–449.

[13] K. Choi and S. Moon, "A Systematic Literature Review on Quantity Flexibility Contracts in a Supply Chain System," *Industrial Engineering & Management Systems*, vol. 21, no. 2, pp. 220–227, 2022.

[14] F. Nikkhoo, A. Bozorgi-Amiri, and J. Heydari, "Coordination of relief items procurement in humanitarian logistic based on quantity flexibility contract," *International Journal of Disaster Risk Reduction*, vol. 31, pp. 331–340, 2018.

[15] J. Heydari, K. Govindan, H. R. E. Nasab, and A. A. Taleizadeh, "Coordination by quantity flexibility contract in a two-echelon supply chain system: Effect of outsourcing decisions," *Int J Prod Econ*, vol. 225, p. 107586, 2020.

[16] Dehghan-Bonari, M., Bakhshi, A., Aghsami, A., & Jolai, F. (2021). Green supply chain management through call option contract and revenue-sharing contract to cope with demand uncertainty. *Cleaner Logistics and Supply Chain*, 2, 100010.

[17] S. Sarkar and S. Bhala, "Coordinating a closed loop supply chain with fairness concern by a constant wholesale price contract," *Eur J Oper Res*, vol. 295, no. 1, pp. 140–156, 2021.

[18] L. Feng, K. Govindan, and C. Li, "Strategic planning: Design and coordination for dual-recycling channel reverse supply chain considering consumer behavior," *Eur J Oper Res*, vol. 260, no. 2, pp. 601–612, 2017.

[19] Heydari, J., & Bakhshi, A. (2022). Contracts between an e-retailer and a third party logistics provider to expand home delivery capacity. *Computers & Industrial Engineering*, 163, 107763. <https://doi.org/10.1016/j.cie.2021.107763>.

[20] Kord, H., & Samouei, P. (2023). Coordination of humanitarian logistic based on the quantity flexibility contract and buying in the spot market under demand uncertainty using NSGA-II and NPGA algorithms. *Expert Systems with Applications*, 214, 119187.

APPENDIX

Proof of Proposition 1. To prove concavity, from Eq. (1), we must determine $\frac{\partial E(\Pi_{GS}(q_w))^2}{\partial q_w^2}$. We have: $\frac{\partial E(\Pi_{GS}(q_w))^2}{\partial q_w^2} = -\alpha^2(w - c_r - c_{st})g(\alpha q_w) < 0$; therefore, GS's profit is concave in q_w . Besides, q_w^* will be obtained by first-order optimality condition, i.e., $\frac{\partial E(\Pi_{GS}(q_w))}{\partial q_w} = (w - c_r) \int_{\alpha q_w}^{\infty} g(m)dm + s \int_0^{\alpha q_w} g(m)dm - (c_{st} + c_p) = 0$. ■

Proof of Proposition 2. To show concavity, from Eq. (6), we calculate $\frac{\partial E(\Pi_{RSC}(q_c))^2}{\partial q_c^2}$. Since $-\alpha^2(w - c_r - s)g(\alpha q_c) < 0$ therefore, the RSC's profit function is strictly concave in q_c . q_c^* to be calculated through the first-order optimality state, which is $\frac{\partial E(\Pi_{RSC}(q_c))}{\partial q_c} = (w - c_r) \int_{\alpha q_c}^{\infty} g(m)dm + s \int_0^{\alpha q_c} g(m)dm + (c_r - c_p - c_{tc} - \frac{c_i}{\alpha} - c_m - c_s) = 0$. ■

Proof of Proposition 3. Since the second-order derivative of Eq. (9) in q_f is negative ($\frac{\partial E(\Pi_{GS}(q_f))^2}{\partial q_f^2} = -\alpha^2(-c_r - c_p - c_{st})g(\alpha q_w) < 0$), the $E(\Pi_{GS}(q_f))$ is concave in q_f . Now the equilibrium will be obtained by solving following equation: $(w - c_r - c_p - c_{st})(1 + u)\alpha^2 q_f G((1 + u)\alpha q_f) - c_{st}(d\alpha)^2 q_f (1 - G((1 + u)\alpha q_f)) + s(d\alpha)^2 q_f G(d\alpha q_f) = 0$

Simulation Method of Port Petrochemical Industry Throughput Development under the Background of Integration of Port, Industry and City

Tingting Zhou, Chen Guo*

College of Economics and Management, Yingkou Institute of Technology, Yingkou, 115014, China

Abstract—In order to accurately predict the changes in the throughput of port petrochemical products and facilitate the formulation of relative decisions, this paper analyzes the factors affecting the throughput of port petrochemical products in a city through the GRA method. After sorting and selection, PCA method is used for pretreatment. In the SVM algorithm, ICSO is used to obtain the best parameters and improve the prediction accuracy and efficiency. In view of the variability of future development, three development scenarios are set up to prepare for the throughput forecast of petrochemical products in a city's port. The results show that the optimization speed of ICSO algorithm is very fast. When the training iteration is 20, the best fitness value is obtained, which is 0.0572. The training effect of ICSO-SVM algorithm is good, the gap between it and the original data is small, and the overall trend is close to the original data. In the test prediction, ICSO-SVM algorithm has the best prediction effect, and its MAE, RMSE and MAPE are the smallest. The minimum MAE is 762.2, 477.0 smaller than CSO-SVM algorithm, and the latter's MAE is 1239.2. The minimum MAPE of the proposed algorithm is 1.05%, while that of CSO-SVM algorithm is 1.71%. In general, the prediction error of ICSO-SVM algorithm is smaller. After the prediction of different development scenarios, the throughput of petrochemical products in a port of a city shows an increasing trend in the next five years. This method can be applied to the development forecast of port petrochemical products and provide reference for decision-making.

Keywords—Support vector machine algorithm; port throughput; chicken swarm optimization algorithm; grey correlation analysis; petrochemical products

I. INTRODUCTION

With the continuous development of the economy, the heavy chemical industries on both sides of the Yangtze River are facing deep adjustment. Due to the low economic efficiency of the chemical siege movement, the waste of land resources and the pollution of the ecological environment have resulted in the lack of impetus for the integration and development of port clusters, petrochemical industry clusters and urban clusters, that is, it is difficult for the three to coordinate [1-3]. In order to promote the coordinated development of port, industry and city, it is necessary to analyze the development of each integrated part in depth. Some scholars choose Support Vector Machine (SVM) to predict the port throughput when studying it. In order to explore the impact of market indicators, they use this indicator as the input of the prediction model. Through experimental

tests, the model has achieved good prediction results and can predict port throughput [4]. To predict the power in the Kanto region, some scholars choose SVM algorithm to build relevant multi-network configuration predictors. After testing, the accuracy of prediction has been improved, and the test effect is good [5]. Therefore, when studying the throughput of port petrochemical products, SVM algorithm is selected as the prediction algorithm. In the analysis and processing of the factors affecting the throughput of port petrochemical products, the grey correlation analysis (GRA) and principal component analysis (PCA) are used for correlation analysis and processing to predict the throughput of port petrochemical products more accurately. By predicting the throughput of petrochemical products in ports, it is beneficial for ports to grasp the direction of business development and actively promote the rational layout of port logistics networks. The study was divided into four parts. The second part is a literature review, which introduces the research of domestic and foreign scholars on port development and port throughput, the good performance of SVM algorithm in forecasting and the relevant application research of PCA and GRA method in data analysis, among which these three methods are also suitable for port throughput forecasting research. In the third part, PCA method and GRA method are proposed to deal with the factors affecting the throughput of petrochemical products at port. ICSO-SVM algorithm is used to predict the throughput of petrochemical products at port. The fourth part carries on the empirical analysis of port petrochemical product throughput prediction, and the results show the superiority of ICSO-SVM algorithm.

II. RELATED WORK

To promote the development of the integration of port, industry and city, the transformation of petrochemical industry is necessary. Relevant parties need to adjust their industrial layout, strengthen the internal power of coordinated development, and protect the ecological environment while coordinating development. During the transformation of the petrochemical industry, some scholars simulated the development of the transportation volume of the port petrochemical industry to understand its development trend and make better decisions. In the face of the transportation problem of petrochemical products, An H and others adopted the relevant comprehensive short-term scheduling model for scheduling. They used heuristic algorithms. The data set test confirmed the superiority and low cost of this method [6]. Zhang selected the port and shipping industry of Ningbo as the

research object to study its development strategy. Taking the spatial scale as the starting point, they put forward their relevant strategic objectives, analyzed the relevant paths, and finally gave the relevant strategic integration model [7]. Wang and others analyzed the national economy and the role of port industry in it. They focus on their input and output, as well as the relevant time evolution process. They summarized relevant policy applications [8]. Ngoc et al. studied container port throughput management and applied control theory and chaos analysis. They optimize the port operation through the theory of sliding membrane control. The test verifies the effectiveness of this method [9].

Zhang et al., faced with the problem of carbon price prediction, chose the least squares SVM to predict it. When processing raw data, they choose the empirical model. The results show that the method is effective and feasible [10]. Praveena et al. detected epileptic seizures and processed the collected data through PCA dimensionality reduction by intelligent means. They use SVM to classify them. After verification, the application effect of this method is good [11]. Song et al. carried out earthquake early warning P-wave prediction on the basis of SVM. After model training, the test error of this method is small [12]. In order to predict PM2.5 of air pollution, Lai X et al. carried out relevant prediction by improving SVM algorithm on the basis of feature selection. The results confirmed the availability of this method [13]. Huang J et al. faced the problem of analyzing the factors affecting the calcination temperature of a vertical furnace, and based on orthogonal design, chose the GRA method to quantify the significant factors involved. According to the analysis results, the influence of volatile matter content is significant and is a key factor [14]. Chen C et al. used the PCA method to reduce the dimensionality of the relevant influencing factors in order to predict the phosphorus content at the endpoint of the Condi electric furnace, and input the processed results into the extreme randomization tree. The results show that the prediction effect is good [15]. Luo S et al. chose to improve the SVM algorithm when facing the problem of predicting the thermal state inside the blast furnace skull. After testing on the dataset, this method has high prediction accuracy [16].

To sum up, in the study of port throughput, there are relatively few researches on its cargo, which mainly focus on the port itself and analyze the development of the port. The researches on cargo throughput are not deep enough. In order to dig into the key factors affecting the development and change of the throughput of petrochemical products in port and obtain more accurate throughput prediction results, this study takes the throughput of petrochemical products in port as the research object and studies its development trend. Through GRA method and PCA method, the key influencing factors are dug out. Considering the good effect of SVM algorithm in prediction, the improved SVM algorithm is used to carry out relevant research, which makes the prediction technology be innovatively improved and the prediction accuracy and generalization ability of the prediction model be enhanced.

III. SIMULATION OF PORT PETROCHEMICAL INDUSTRY TRAFFIC VOLUME DEVELOPMENT BASED ON ICSO-SVM ALGORITHM UNDER THE BACKGROUND OF PORT INDUSTRY AND CITY INTEGRATION

A. Treatment of Factors affecting Port Petrochemical Product Throughput based on GRA and PCA Methods

In the continuous economic development, the integration of port and industry is an important way to develop regional economic, which is conducive to the transformation of petrochemical industry and accelerating the integration process. To understand the development trend of the petrochemical industry in the development of transportation volume, the article selects the port petrochemical product throughput as the research object and analyzes it. The change trend of the throughput of petrochemical products in a city's port in recent years is shown in Fig. 1.

In Fig. 1, the throughput of port petrochemical products in 2020 was lower than that in 2019. In 2021, the throughput of petrochemical products in the port returned to normal and increased to a certain extent, surpassing 2019. Although there were gaps in the throughput of petrochemical products in ports in different years, there was an overall growth trend. To study the causes of the changes in the throughput of port petrochemical products, this paper analyzes the influencing factors. Through consulting information and data acquisition, it is found that the throughput of port petrochemical products is affected by many factors. Based on these factors, the paper constructs a pre-selection index system of relevant influencing factors, as shown in Fig. 2.

In Fig. 2, the indicator system includes four primary indicators, namely economic and trade level (EATL), regional development vitality (RDV), population and employment level (PAEL), port infrastructure conditions (PIC), and 14 secondary indicators. In order to understand the correlation between these indicators and the throughput of port petrochemical products, the article selects the GRA method to analyze them. The greater the correlation, the stronger the corresponding correlation. In the GRA method, first set the original sequence $X'_i = [x'_i(1), x'_i(2), \dots, x'_i(n)]$, $i = 0, 1, 2, \dots, l$. The length of the number sequence is set to n , and there are $l+1$ index number sequences collected^[17-19]. The difference sequence is solved for the initial value of each sequence, and its calculation formula is shown in Eq. (1).

$$\Delta_i(k) = |x'_0(k) - x'_i(k)|, \Delta_i = (\Delta_i(1), \Delta_i(2), \dots, \Delta_i(n)) \quad (1)$$

In formula (1), Δ_i represents the difference sequence and k represents the sequence number. Solve the maximum difference and minimum difference between the two poles, as shown in Formula (2).

$$M = \max_l \max_k \Delta_i(k), M = \min_l \min_k \Delta_i(k) \quad (2)$$

In formula (2), M represents the maximum difference between the two poles, and m represents the minimum difference between the two poles. Solve the correlation coefficient $\gamma_{0i}(k)$ and calculate the correlation degree. The

calculation formula is shown in formula (3).

$$\gamma_{oi} = \frac{1}{n} \sum_{k=1}^n \gamma_{oi}(k), i = 1, 2, \dots, l \quad (3)$$

In formula (3), γ_{oi} represents the degree of correlation. The software used for correlation calculation is MATLAB. Sort the relevant results according to the order from the largest to the smallest. According to the ranking of the correlation degree obtained, 13 indicators are selected as the prediction indicators required by the article. These indicators are in the top 13 in the ranking of correlation degree and have high correlation with the throughput of port petrochemical products. Among them, the top three indicators are GDP, total import

and export of petrochemical products, and coastal berths. In particular, the correlation value corresponding to GDP is the largest. In order to better reflect the impact factors of port throughput, 13 indicators are included in the prediction model. However, the dimension of data is too large to have a certain impact on the prediction effect of SVM. Therefore, PCA method is adopted to reduce data redundancy and reduce the complexity of the problem [20-22]. It is a linear dimensionality reduction algorithm with practical significance, which can perform orthogonal transformation on high-dimensional data. The transformed data maintains the basic information of the original data. After the change of linearly related variables, the new variable formed has linear independence, which is also called principal component.

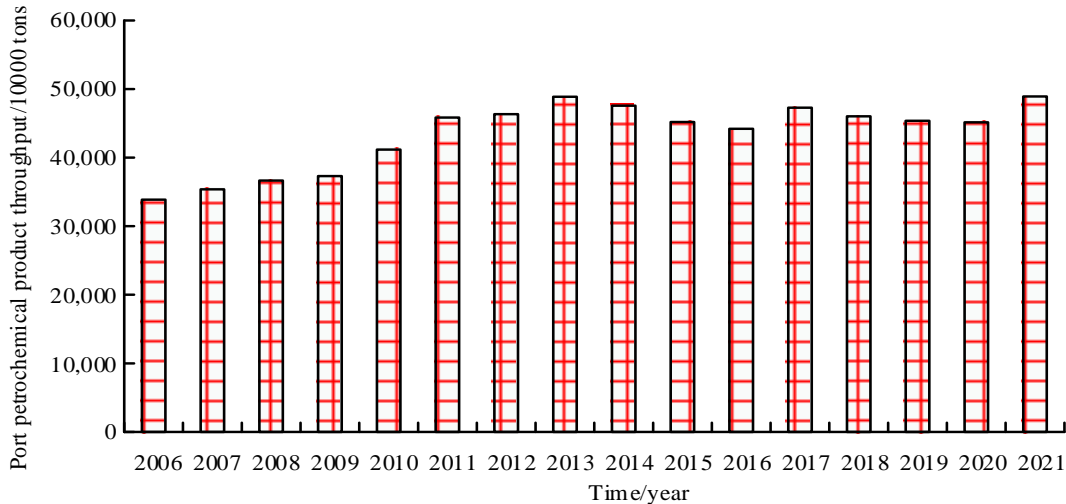


Fig. 1. Change trend of petrochemical product throughput in a port.

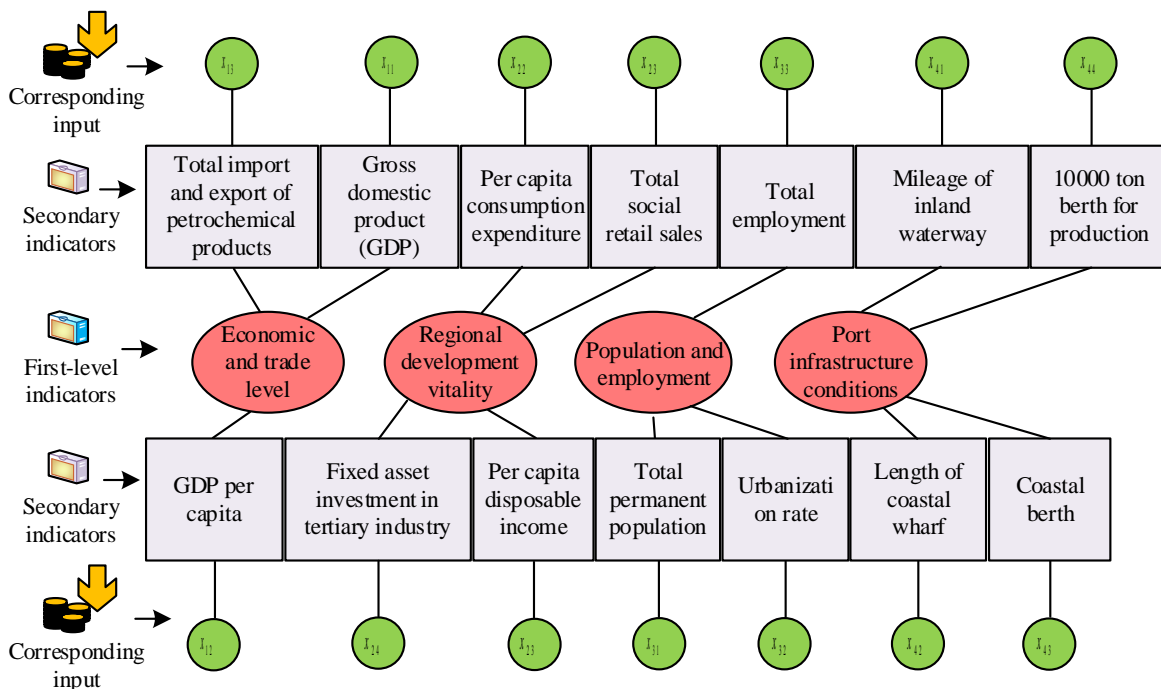


Fig. 2. Pre-selection index system of influencing factors.

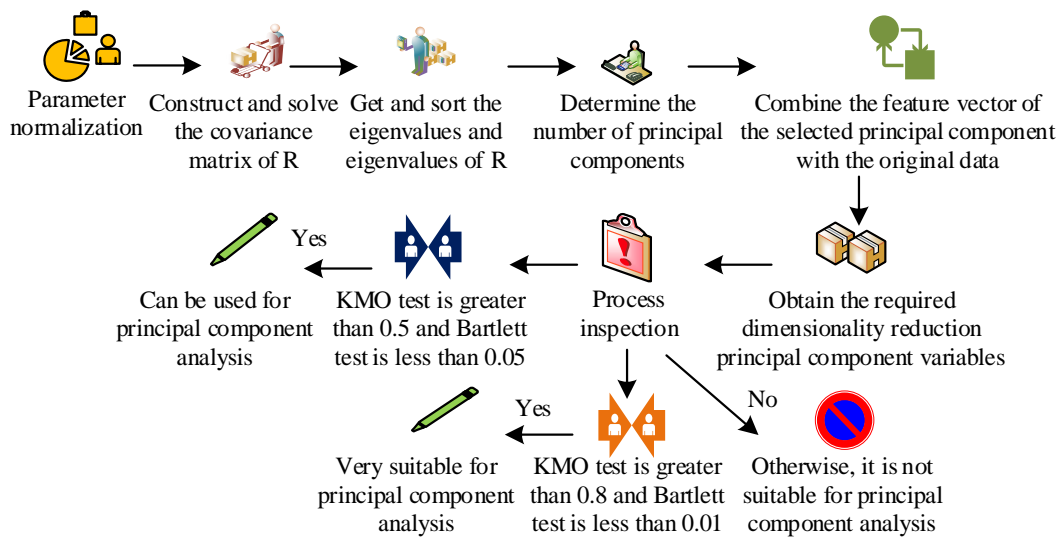


Fig. 3. Relevant process of PCA method.

PAC method first normalizes all parameters to form a normalized matrix. Secondly, it obtains the eigenvalues and eigenvectors of R by constructing and solving the covariance matrix of R, and ranks them according to the variance of principal components. The number of principal components is determined by their contribution degree and cumulative variance contribution rate, and the eigenvalues are taken as the judgment criteria, which need to exceed 1, and the contribution rate need to be no less than 85%. The feature vector of the selected principal component is combined with the original data to obtain the required dimension-reduced principal component variable. Then, there are two test methods in dimension reduction, namely, Kaiser-Meyer-Olkin (KMO) test and Bartlett test. If the value of the former test method is greater than 0.5 and the value of the latter test method is less than 0.05, the principal component analysis can be performed. If the value of the former test method is greater than 0.8, and the value of the latter test method is less than 0.01, it is particularly suitable for principal component analysis. The relevant process of PCA method is shown in Fig. 3.

After dimension reduction by PCA method, the principal components are obtained, and the correlation between the factors is eliminated, which maximizes the original main information. This method is simple, easy to implement, without parameter constraints, and has good objectivity. It can not only simplify the subsequent data processing, but also help the SVM model have better robustness. The PCA method is used to reduce the dimension of the selected 13 indicators. To better save the relevant information of the original data, the PCA method is used to process the secondary indicators under a primary indicator, and then the relevant secondary indicators of the remaining primary indicators are processed according to this method. The software used for PCA analysis is SPSS software.

B. Application of ICSO-SVM Algorithm in Throughput Prediction of Port Petrochemical Products

According to the relevant data characteristics of port petrochemical product throughput, forecasting and analysis by

SVM algorithm can effectively overcome the problems of other models. Problems include insufficient sample size and poor model stability. This method is good at processing a kind of nonlinear data, which is characterized by small samples and high dimensions. The SVM algorithm is to find an optimal hyperplane. On both sides of the plane, the samples are farthest away from it, and the classification effect and robustness of the algorithm are also better [23-24]. In the distance between sample point and hyperplane, the relevant calculation formula is shown in formula (4).

$$d = \frac{\|w^T x + b\|}{\|w\|} \quad (4)$$

In formula (4), x_i represents sample point and d represents sample interval. The normal vector is expressed as w , b means parameter. To further describe the sample interval, some training samples closest to the hyperplane on the hyperplane are found, and a formula that meets the conditions is given as formula (5).

$$\begin{cases} w^T x_i + b = 1, y_i = 1 \\ w^T x_i + b = -1, y_i = -1 \end{cases} \quad (5)$$

In formula (5), y_i represents the sample point. According to Formula (4) and Formula (5), the sum of the distance between the two support vectors and the hyperplane is $\frac{2}{\|w\|^2}$.

When this value is maximum, the corresponding sample discrimination is maximum. Analyze the nature of the sample and the degree to which it can be separated, and use the kernel function SVM as its learning method. Support vector machine regression (SVR) is widely used in classifiers. When constructing the regression function $f(x)$, set the loss boundary as ϵ . When the distance between $f(x)$ and sample is less than ϵ , an isolation band with the width of 2ϵ is formed. When the sample is in the isolation zone, the

prediction is correct. After the kernel function is applied, the relevant objective function is shown in equation (6).

$$\min \frac{\|w\|^2}{2} + C \sum_{i=1}^N \ell_\varepsilon(f(x_i) - y_i) \quad (6)$$

In formula (6), the penalty factor is expressed as C , and the ε insensitive loss function is expressed as ℓ_ε . The relevant expression is shown in Eq. (7).

$$\ell_\varepsilon = \begin{cases} 0, & \text{if } |z| \leq \varepsilon \\ |z| - \varepsilon, & \text{else} \end{cases} \quad (7)$$

In formula (7), z represents the parameter. By adding relaxation variables δ_i and $\hat{\delta}_i$, the SVR problem is transformed. Combining Lagrange multipliers α_i and $\hat{\alpha}_i$, the relevant dual problem can be obtained, and then the optimal solution of SVR problem is shown in Eq. (8).

$$f(x) = \sum_{i=1}^N (\alpha_i - \hat{\alpha}_i) K(x_i, x) + b \quad (8)$$

In formula (8), K represents the parameter. In SVM, C and kernel function width σ have a greater impact on its prediction effect. In view of this situation, radial basis function (RBF) is selected as the kernel function in this paper because of its strong generalization ability. In addition, to obtain the best parameters, the paper selects CSO algorithm for parameter optimization. Because of its poor processing effect on complex optimization problems, this paper optimizes it through two improvement strategies and obtains the improved CSO (Improved Chicken Swarm Optimization, ICSO) algorithm. Using chaos theory, the diversity of the algorithm is strengthened, and adaptive learning strategy is introduced. In the first strategy, due to the randomness of the initial population generation of CSO algorithm, it is easy to have a local optimal solution. Adding chaotic variables can avoid this situation. This variable can be generated by a specific chaotic map, and its initial population distribution range is much larger than the random variable, and can be combined with intelligent optimization methods. Among them, tent mapping is a typical method of generating chaotic sequence similar to tent, and the formula involved is shown in formula (9).

$$x''_{i+1} = \begin{cases} x''_i/a, & x''_i < a \\ (1-x''_i)/(1-a), & x''_i \geq a \end{cases} \quad (9)$$

In formula (9), x''_i and x_0 represent chaotic variables and initial variables respectively. The mapping coefficients are expressed as $a \in (0, 1)$, and a group of chaotic sequences can be obtained by x_0 iterating n times. Compared with random sequence, the data distribution characteristics of the chaotic sequence of Ten-map are relatively stable, so the corresponding x''_i is selected to replace the random number of CSO algorithm. In the adaptive learning strategy, the chicken's movement mode is improved, and the self-learning

coefficient is added, so that the direction of its movement has a certain probability towards the cock, and the relevant calculation formula of the relevant improved strategy is shown in Formula (10) and Formula (11).

$$x''_{ij}{}^{t+1} = w_i * x''_{ij}{}^t + F * (x''_{mj}{}^t - x''_{ij}{}^t) + R * (x''_{rj}{}^t - x''_{ij}{}^t) \quad (10)$$

In formula (10), w_i represents the self-learning coefficient of chicks, and x_m represents the position of hens. At the time of t , the position of the individual is expressed as x''_{ij} , the fitness of the i th hen is expressed as f_i , and the chick random following parameter is expressed as F . r represents the cock's coefficient, and R means the coefficient that follows the cock's movement, which will replace the chaotic variable.

$$w_i = \begin{cases} w_{\min} + (f_i - f_{\min}) * (w_{\max} - w_{\min}) / (f_{\text{avg}} - f_{\min}), & \text{if } f_i \leq f_{\text{avg}} \\ w_{\max}, & \text{else} \end{cases} \quad (11)$$

In formula (11), w_{\min} and w_{\max} represent the minimum and maximum weight. The maximum fitness, minimum fitness and average fitness of chicken flocks are expressed as f_{\max} , f_{\min} and f_{avg} respectively. Combined with the improved CSO algorithm, the problem to be solved is the optimal problem on the bounded two-dimensional plane. Its goal is to find the f_{\min} best individual. The individual f of the chicken flock corresponds to the mean square error of prediction, that is, to find the minimum prediction error, and then the optimal parameters of SVM can be obtained. The throughput of port petrochemical products is predicted by ICSO-SVM algorithm. The data of the first 11 years in Fig. 1 are classified as training sets, and the rest are test sets. First, data preprocessing and parameter setting are carried out. The basic model is the ε -SVR model. The system default value is 0.1. Set the ICSO algorithm parameters, as shown in Fig. 4.

In Fig. 4, according to the Ten-map chaotic algorithm, all chickens form initial positions. In the iterative process, individuals search for optimization in the search space according to their own way of movement. The individual identity is updated every 10 generations. The update is based on the current f . Keep iterating until the prediction error is less than 10-4 or the maximum iteration is reached, then the operation will be stopped. For data preprocessing, Z-core standardization method is selected, as Formula (12).

$$H' = \frac{H - \bar{H}}{\sigma} \quad (12)$$

In formula (12), H' , H and \bar{H} are mean normalized, original and original average data respectively. Mean Absolute Error (MAE) is used to evaluate the prediction effect, as Formula (13).

$$MAE = \frac{1}{T} \sum_{t=1}^T |s_t - n_t| \quad (13)$$

In formula (13), T represents the number of years, and

the real original value and predicted value of the t year are s_t and n_t respectively. The evaluation index is Mean Absolute Percentage Error (MAPE), as shown in formula (14).

$$MAPE = \frac{1}{T} \sum_{t=1}^T \left| \frac{s_t - n_t}{s_t} \right| * 100\% \quad (14)$$

The root mean square error (RMSE) as the evaluation index is shown in formula (15).

$$RMSE = \sqrt{\frac{1}{T} \sum_{t=1}^T \left| \frac{s_t - n_t}{s_t} \right|^2} \quad t \in 1, 2, \dots, T \quad (15)$$

After the prediction results are obtained, they are compared with other prediction methods to verify the reliability of the methods used in the article. Overall, the model building process is shown in Fig. 5.

In Fig. 5, several comparison models are applied to test and verify the performance of the methods used in the article. During the comparative experiment, the same data sets were used. These comparison methods are shown in Fig. 6.

After testing the method, it is applied to the throughput prediction of petrochemical products in a port of a city. Considering the uncertainty of future development, this paper uses scenario prediction method to predict the throughput of petrochemical products in a city's future port. First, according to the policy documents of a city and its development rules, different scenarios are set to reflect different development situations. The middle-aged change rate of different scenarios is set to obtain the corresponding future value. These values are entered into the model. Through the trained model, the throughput under different scenarios is predicted and the relevant outputs are obtained.

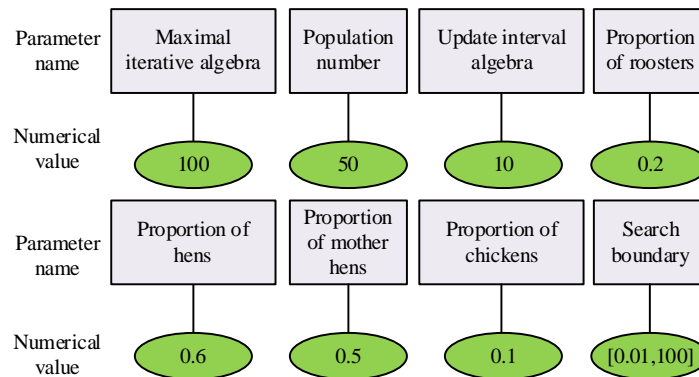


Fig. 4. ICSO algorithm parameters.

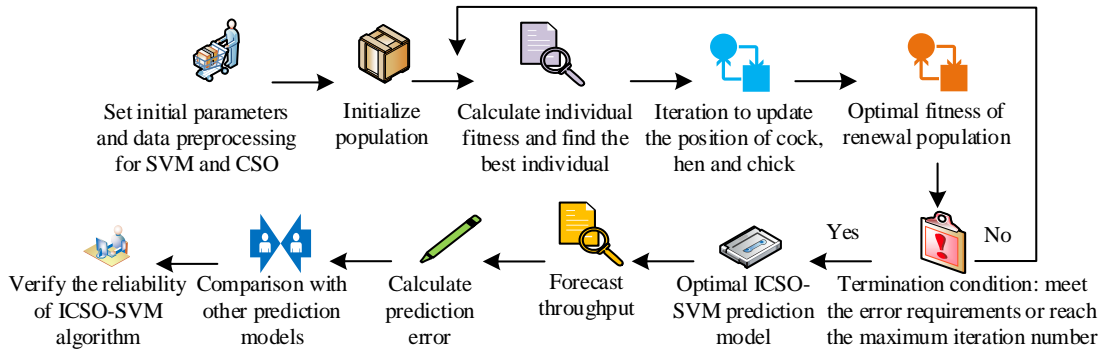


Fig. 5. Model building process.

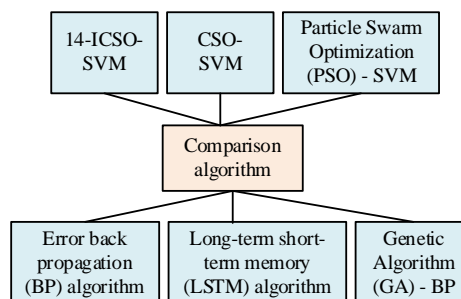


Fig. 6. Relevant comparison methods.

IV. PREDICTION AND ANALYSIS OF PORT PETROCHEMICAL PRODUCT THROUGHPUT

This paper uses the GRA method to deal with the impact indicators of the throughput of petrochemical products in a port of a city, so as to analyze the correlation between the various impact indicators. After processing by the GRA method, the correlation value of each indicator in the impact indicator system is shown in Fig. 7.

Fig. 7 shows that the grey correlation degree values of the impact indicators on the throughput of petrochemical products in different ports are different. Among these indicators, the grey correlation degree value of x'_{11} indicator is the largest, 0.882. The second is x'_{13} , whose grey correlation value is 0.011 less than x'_{11} . The grey correlation value of x'_{13} is 0.871. The grey correlation degree values of indicators x'_{22} , x'_{12} and x'_{24} are 0.854, 0.849 and 0.817 respectively. The

grey correlation degree value of indicator x'_{12} is 0.032 higher than x'_{24} . The index with the lowest grey correlation value of the impact index is x'_{32} , and its grey correlation value is 0.571, which is less than 0.6, that is, the urbanization rate has a weak correlation with the throughput of petrochemical products in a city's port. In addition to the urbanization rate, the other 13 impact indicators have a strong or general correlation with the throughput of petrochemical products in the port of the city. Among them, GDP, total import and export of petrochemical products, coastal berths and the throughput of petrochemical products in the port of a city have a strong correlation. GDP is the most critical impact indicator. Therefore, 13 impact indicators other than urbanization rate are selected as the input data of the prediction model. The training data selects the data from 2006 to 2016. After preprocessing, the ICSO-SVM algorithm is trained through the processed data, and the relevant results are shown in Fig. 8.

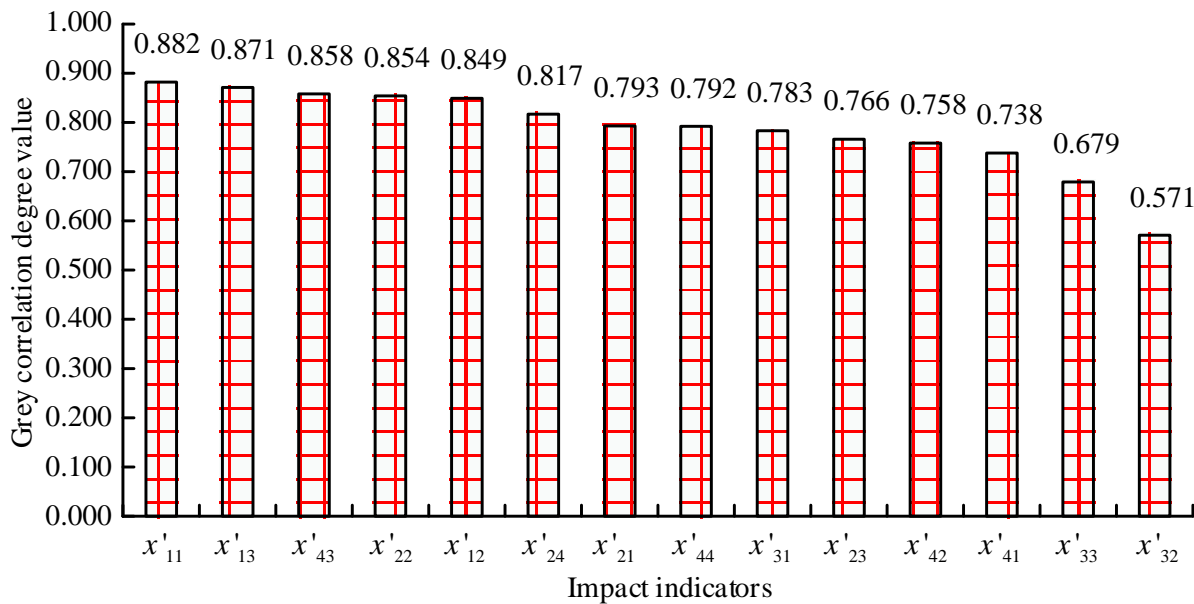


Fig. 7. Grey correlation degree value of impact index.

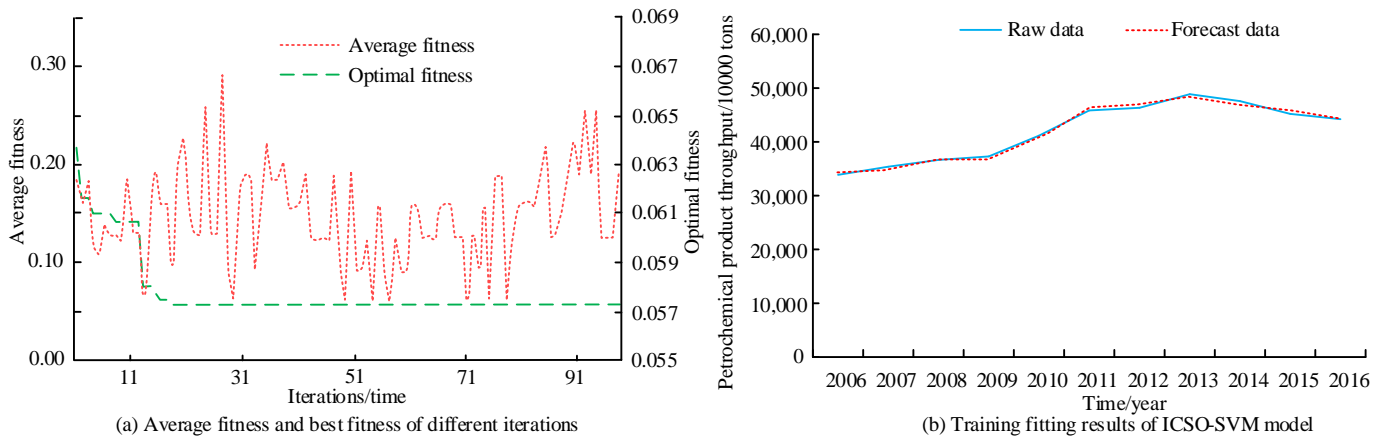


Fig. 8. Algorithm iterative optimization and training fitting results.

Fig. 8(a) shows the iterative optimization process of ICSO algorithm. Fig. 8(b) shows the training results of ICSO-SVM algorithm. In Fig. 8(a), the average fitness value of the ICSO algorithm is different under different iterations. The optimal fitness decreased slowly and then stabilized. When the iteration number is 12, the average fitness value of the ICSO algorithm is 0.13, which is 0.05 less than the average fitness value of the iteration number of 10. When the iteration number is 28, the average fitness value of ICSO algorithm is the largest, and its average fitness value is 0.24. For the best fitness of ICSO algorithm, when the iteration number is 10, the best fitness value is 0.0607, which is 0.0029 less than that when the iteration number is 1, and the best fitness value of the latter is 0.0636. When the iteration number is 20, the best fitness value of the ICSO algorithm is 0.0572, and the ICSO algorithm begins to converge. Therefore, the optimization speed of the proposed algorithm is fast and the error is low. In Fig. 8(b), according to the line chart of the predicted value of different years and the original data, the two fit well and the fitting result is good. In 2007, the original data was 353.71 million tons, and the forecast data was 348.09 million tons, which was 5.62 million tons less than the former. In 2008, the forecast data was 366.65 million tons, 180000 tons more than the original data. Therefore, the gap between the two is small. The prediction results of different algorithms in the test set are shown in Fig. 9.

In Fig. 9, according to the broken line of the prediction results of different algorithms, the broken line of the prediction results of the ICSO-SVM algorithm is closest to the broken line of the original data. The distance between other polylines and the polyline where the original data is located is different. In 2017, the forecast result of ICSO-SVM algorithm was 478.53 million tons, and the original data was 472.82 million tons, the former was 5.71 million tons more than the latter. In 2019, the prediction results of ICSO-SVM algorithm, CSO-SVM algorithm and PSO-SVM algorithm were 449.73 million tons, 441.03 million tons and 412.6 million tons respectively, while the original data was 453.8 million tons. The gap between the predicted results of ICSO-SVM algorithm and the original data is the smallest. In 2021, the predicted results of SVM algorithm and ICSO-SVM algorithm are 378.56 million tons and 482.05 million tons respectively, which are 110.54 million tons and 7.05 million tons less than the original data. Quantify the prediction error of the algorithm and the relevant results are shown in Fig. 10.

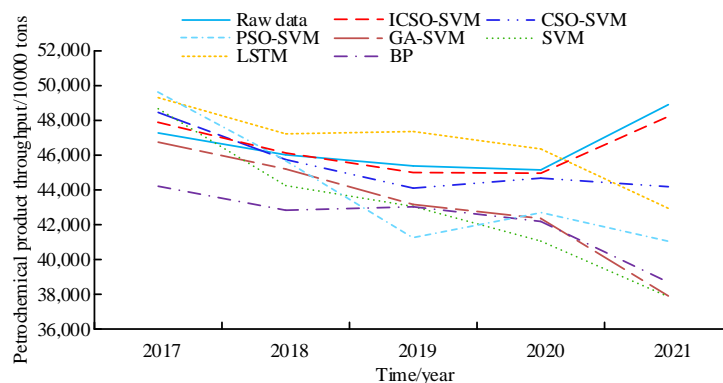


Fig. 9. Prediction results of different algorithms.

In Fig. 10(a), different algorithms correspond to different MAE and RMSE. In terms of MAE, the MAE of ICSO-SVM algorithm is 762.2, 477.0 smaller than CSO-SVM algorithm, and the MAE of the latter is 1239.2. The MAE of GA-SVM algorithm is 1905.0, and that of SVM algorithm is 2116.4. The maximum MAE of BP algorithm is 2473.0, which is 1710.8 larger than that of ICSO-SVM algorithm, and the latter has the minimum MAE. In terms of RMSE, the minimum RMSE of ICSO-SVM algorithm is 814.7, while the maximum RMSE of BP algorithm is 2792.4. In Fig. 10(b), the minimum MAPE of ICSO-SVM algorithm is 1.05%, which is 0.66% smaller than CSO-SVM algorithm. Therefore, ICSO-SVM algorithm has the best prediction effect. The algorithm without influencing factor analysis is 14-ICSO-SVM algorithm, and the prediction results under different pretreatment methods are shown in Fig. 11.

In Fig. 11(a), the prediction results obtained by the two pretreatment methods differ greatly. Compared with 14-ICSO-SVM algorithm, ICSO-SVM algorithm has smaller MAE and RMSE values. The MAE value of ICSO-SVM algorithm is 762.2, 1016.4 less than that of 14-ICSO-SVM algorithm, and the MAE value of the latter is 1778.6. The RMSE values of ICSO-SVM algorithm and 14-ICSO-SVM algorithm are 814.7 and 1901.7 respectively. In Fig. 11(b), the MAPE values of ICSO-SVM algorithm and 14-ICSO-SVM algorithm are 1.05% and 2.38% respectively, the latter is 1.33% higher than the former. The results further show that the GRA and PCA analysis of the data before the model prediction is conducive to reducing the prediction error. Input the influence factors after pretreatment into the trained model, and predict the throughput of petrochemical products in a city's port from 2023 to 2027 according to different development scenarios. Fig. 12 gives the results.

In Fig. 12, the predicted value under the high-speed development scenario is the largest. Under the same development scenario, the throughput of petrochemical products increased year by year. Under the low-speed development scenario, the forecast result in 2023 is 483.25 million tons, 27.14 million tons less than that in 2025. The forecast result in 2027 is 53.152 million tons. In 2026, the forecast results under the baseline development scenario and the high-speed development scenario are 528.75 million tons and 533.15 million tons respectively, 12.53 million tons and 59.33 million tons less than the corresponding development scenario in 2027.

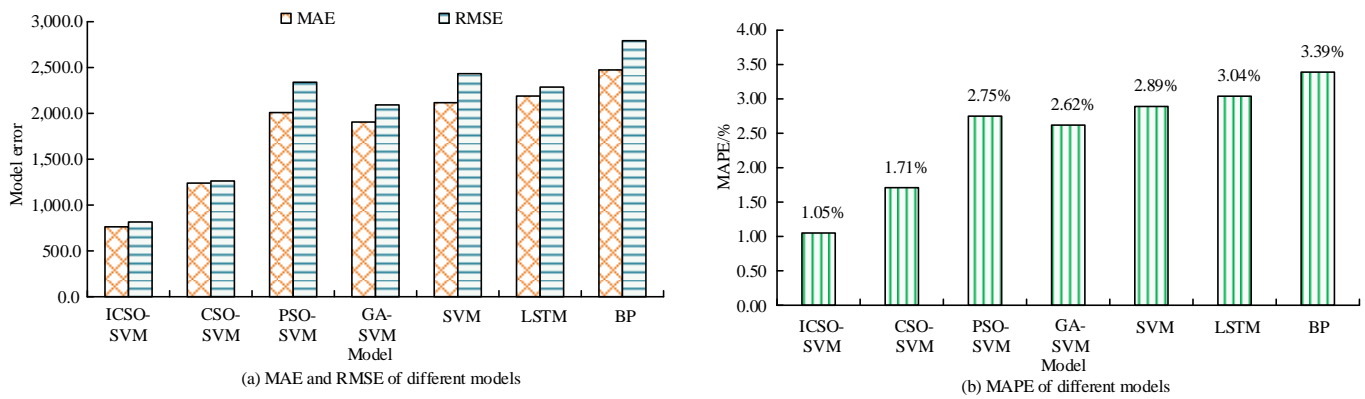


Fig. 10. Prediction error of different algorithms.

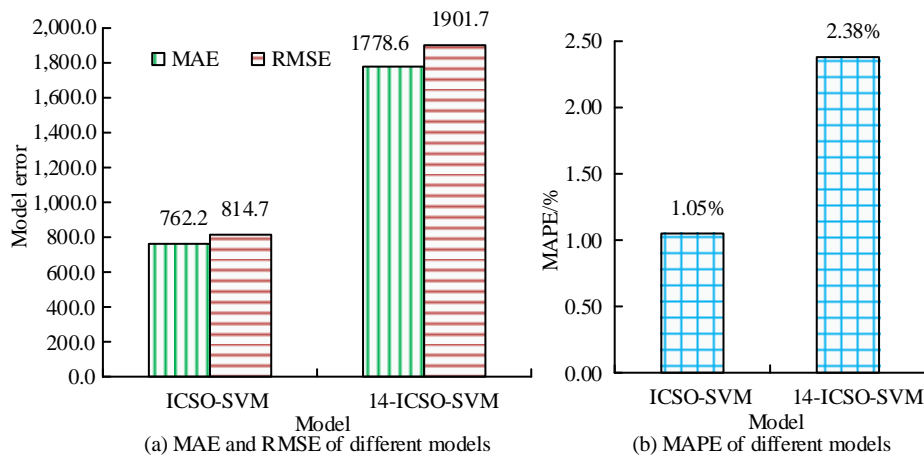


Fig. 11. Prediction results under different pretreatment methods.

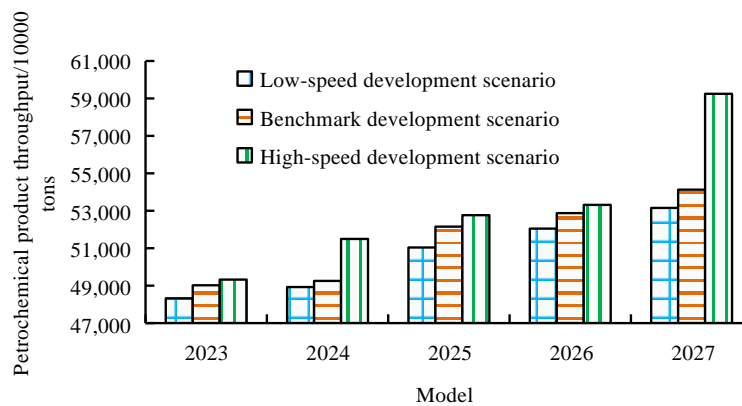


Fig. 12. Prediction results under different scenarios.

In the throughput prediction of petrochemical products at a port in a certain city, the influencing factors of throughput development will affect the prediction results. By screening out key influencing factors, the prediction results are superior to those without considering the influencing factors. Some scholars take into account the influencing factors of FDT when predicting the finish rolling discharge temperature (FDT). From the results, it can be seen that considering the influencing factors, the prediction accuracy of FDT is higher [25]. It can be seen that mining key influencing factors is beneficial for improving prediction accuracy. The model

parameters will affect the prediction accuracy of the SVM algorithm, and the more suitable the parameters are, the higher the prediction accuracy of the algorithm. Some scholars choose SVM algorithm to optimize algorithm parameters through simulated annealing algorithm when predicting student grades, in order to obtain the best parameters. After verification, the optimized SVM algorithm has better prediction performance [26]. It can be seen that optimizing parameters can improve the prediction accuracy of the algorithm. When predicting the throughput of petrochemical products, a more comprehensive and scientific consideration

of key influencing factors and optimization of SVM algorithm parameters can yield more accurate prediction results, making the obtained results more valuable for reference and providing an effective universal prediction tool for decision-makers.

V. CONCLUSION

Under the background of the integration of port and industry, to understand the development of the port petrochemical industry, this paper takes the throughput of port petrochemical products as the research object, analyzes its impact factors through the GRA method, and studies its correlation with different impact indicators. The index is sorted and selected based on the grey correlation degree, and the selected influence index is pretreated by PCA method as the input of the prediction model. The research improves the SVM algorithm through the ICSO algorithm, so as to train and test the preprocessed data. At the end of the study, the throughput of petrochemical products in the port of a city in the next five years is predicted. The results show that the optimal fitness value of ICSO algorithm is 0.0572, and the corresponding iteration number is 20. The optimization speed of the algorithm is fast and the error is low. In the training results of ICSO-SVM algorithm, the predicted value is close to the line corresponding to the original data, and the difference between the two data is small. Among the test results of different algorithms, the gap between the predicted results of ICSO-SVM algorithm and the original data is the smallest. In terms of prediction error, ICSO-SVM algorithm has the smallest MAE, RMSE and MAPE. The minimum MAE is 762.2, 1710.8 less than BP algorithm. Its minimum MAPE is 1.06%, which is 0.65% smaller than CSO-SVM algorithm. Among the different preprocessing methods, the prediction result of ICSO-SVM algorithm is better than that of 14-ICSO-SVM algorithm. The former has the lowest MAE, RMSE and MAPE. The MAPE values of ICSO-SVM algorithm and 14-ICSO-SVM algorithm are 1.05% and 2.38% respectively, and the latter is 1.33% higher than the former. Through the prediction of the throughput of petrochemical products in a city's port, its throughput is increasing year by year. Under the high-speed development scenario, the forecast result in 2027 is 592.48 million tons. Thus, the prediction effect of the method used in the article is good. In the future, more scenarios can be set, and the diversity of influencing factors can be considered to make the development scenario prediction more reasonable.

ACKNOWLEDGMENT

The research is supported by: Natural Science Foundation of Liaoning Province "Science and Technology Innovation Ecosystem Construction and Governance Mechanism Research" (2020-YKLH-21); The Key projects of basic scientific research projects in colleges and universities in the Department of Education of Liaoning Province in 2021 "Exploring the path of integrated development of manufacturing and service industries in Liaoning" (LJKR0585); Yingkou Soft Science Plan Project "Research on the Countermeasures of transformation and upgrading of Petrochemical Industry Promoted by 'Yingkou Port-industry-city' Integrated development" in 2022 (2022JH2/0100001).

REFERENCES

- [1] Wang Z, Li J, Mu X, et al. A WRF-CMAQ modeling of atmospheric PAH cycling and health risks in the heavy petrochemical industrialized Lanzhou valley, Northwest China. *Journal of Cleaner Production*, 2021, 291(5):125989.1-125989.9.
- [2] Zhang T C, Zheng B Y, Li K, et al. Using CAESAR II software to do stress analysis on centrifugal pump pipeline in Tianjin Port-North China Petrochemical crude oil pipeline project. *Xiandai Huagong/modern Chemical Industry*, 2017, 37(9):211-212+214.
- [3] Maritime, Singapore P. Port Initiatives in Singapore as A Driver of Technology and Growth. *Marine Money International*, 2018, 34(4):16-19.
- [4] Caliskan A, Karaz B. Can market indicators forecast the port throughput?. *International Journal of Data Mining Modelling and Management*, 2019, 11(1):45-63.
- [5] Takamatsu T, Ohtake H, Oozeki T, et al. Regional Solar Irradiance Forecast for Kanto Region by Support Vector Regression Using Forecast of Meso-Ensemble Prediction System. *Energies*, 2021, 14(11):1-18.
- [6] An H, Choi S S, Lee J H. Integrated scheduling of vessel dispatching and port operations in the closed-loop shipping system for transporting petrochemicals. *Computers & Chemical Engineering*, 2019, 126(JUL.12):485-498.
- [7] Zhang N. Research on the Development Strategy of Ningbo Transportation Port & Shipping Industry. *Journal of Transportation Technologies*, 2019, 09(4):474-488.
- [8] Wang Y, Wang N. The role of the port industry in China's national economy: An input-output analysis. *Transport Policy*, 2019, 78(JUN.):1-7.
- [9] Ngoc C T, Xu X, Kim H S, et al. Container port throughput analysis and active management using control theory. *Proceedings of the Institution of Mechanical Engineers, Part M: Journal of Engineering for the Maritime Environment*, 2022, 236(1):185-195.
- [10] Zhang W, Wu Z, Bunn D W. Optimal hybrid framework for carbon price forecasting using time series analysis and least squares support vector machine. *Journal of Forecasting*, 2022, 41(3):615-632.
- [11] Praveena H D, Subhas C, Naidu K R. Classification and discrimination of focal and non-focal EEG signals using hybrid features and support vector machine. *International Journal of Advanced Intelligence Paradigms*, 2021,18(3):417-437.
- [12] Song J, Yu C, Li S. Continuous prediction of onsite PGV for earthquake early warning based on least squares support vector machine (in Chinese). *Chinese Journal of Geophysics- Chinese Edition*, 2021, 64(2):555-568.
- [13] Lai X, Li H, Pan Y. A combined model based on feature selection and support vector machine for PM2.5 prediction. *Journal of Intelligent and Fuzzy Systems*, 2021, 40(5):10099-10113.
- [14] Huang J, Li J, Li M, K Yan. Orthogonal Design-Based Grey Relational Analysis for Influence of Factors on Calcination Temperature in Shaft Calciner. *Journal of chemical engineering of Japan*, 2019, 52(11):811-821.
- [15] Chen C, Wang N, Chen M. Prediction Model of End-point Phosphorus Content in Consteel Electric Furnace Based on PCA-Extra Tree Model. *ISIJ International*, 2021,6(61):1908-1914.
- [16] Luo S, Dai Z, Chen T, H Chen, L Jian A weighted SVM ensemble predictor based on AdaBoost for blast furnace Ironmaking process. *Applied Intelligence*, 2020, 50(12):1997-2008.
- [17] Xu HM, Zhong WJ, Wang CL et al. Quantitative analysis and evaluation of manipulation comfort of tractor gear shifting based on combined methods. *Human Factors and Ergonomics in Manufacturing & Service Industries*, 2019, 29(4):285-292.
- [18] Liu X Y, Zhao Y X, Liu X D, et al. Design of Seamless Knitted Health Care Pants for Knee Arthritis Based on Grey Correlation Analysis. *Journal of fiber bioengineering and informatics*, 2020,13(2):79-87.
- [19] Li X, Zhuo B, Qi X, et al. Grey correlation analysis and path analysis between isoflavones content in Astragali Radix and climate factors. *Zhongguo Zhong yao za zhi = Zhongguo zhongyao zazhi = China journal of Chinese materia medica*, 2020, 45(14):3407-3413.
- [20] Chen Z, Tian K. Optimization of Evaluation Indicators for Driver's

- Traffic Literacy: An Improved Principal Component Analysis Method. *SAGE Open*, 2022, 12(2):242-252.
- [21] Yang X, Xiang Y, Jiang B. On multi-fault detection of rolling bearing through probabilistic principal component analysis denoising and Higuchi fractal dimension transformation. *Journal of Vibration and Control*, 2022, 28(9-10):1214-1226.
- [22] Barth J, Katumullage D, Yang C, et al. Classification of Wines Using Principal Component Analysis. *Journal of Wine Economics*, 2021, 16(1):56-67.
- [23] Bhat H F, Wani M A. Novel PSSM-Based Approaches for Gene Identification Using Support Vector Machine. *Journal of Information Technology Research*, 2021, 14(2):152-173.
- [24] Ren J, Zhang B, Zhu X, et al. Damaged cable identification in cable-stayed bridge from bridge deck strain measurements using support vector machine. *Advances in Structural Engineering*, 2022, 25(4):754-771.
- [25] Kang J S, Kim J T, Baek B J. Influence of boundary conditions on FDT prediction and improvement of FDT prediction accuracy in finishing mill process. *Asia Life Sciences*, 2019(3):1501-1512.
- [26] Mahareek E A, Desuky A S, El-Zhni H A. Simulated annealing for svm parameters optimization in student's performance prediction. *Bulletin of Electrical Engineering and Informatics*, 2021, 10(3): 1211-1219.