Lecture Notes in Networks and Systems 1755

Surekha Lanka
Antonio Sarasa Cabezuelo
Alexandru Tugui *Editors*

Trends in Sustainable Computing and Machine Intelligence

Proceedings of ICTSM 2025

Springer

Lecture Notes in Networks and Systems 1755

Series Editor

Janusz Kacprzyk, *Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland*

Advisory Editors

Fernando Gomide, *Department of Computer Engineering and Automation—DCA, School of Electrical and Computer Engineering—FEEC, University of Campinas—UNICAMP, São Paulo, Brazil*

Okyay Kaynak, *Department of Electrical and Electronics Engineering, Bogazici University, Istanbul, Türkiye*

Derong Liu, *Department of Electrical and Computer Engineering, University of Illinois at Chicago, Chicago, USA*

Institute of Automation, Chinese Academy of Sciences, Beijing, China

Witold Pedrycz, *Department of Electrical and Computer Engineering, University of Alberta, Edmonton, Alberta, Canada*

Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland

Marios M. Polycarpou, *Department of Electrical and Computer Engineering, KIOS Research Center for Intelligent Systems and Networks, University of Cyprus, Nicosia, Cyprus*

Imre J. Rudas, *Óbuda University, Budapest, Hungary*

Jun Wang, *Department of Computer Science, City University of Hong Kong, Kowloon, Hong Kong*

The series "Lecture Notes in Networks and Systems" publishes the latest developments in Networks and Systems—quickly, informally and with high quality. Original research reported in proceedings and post-proceedings represents the core of LNNS.

Volumes published in LNNS embrace all aspects and subfields of, as well as new challenges in, Networks and Systems.

The series contains proceedings and edited volumes in systems and networks, spanning the areas of Cyber-Physical Systems, Autonomous Systems, Sensor Networks, Control Systems, Energy Systems, Automotive Systems, Biological Systems, Vehicular Networking and Connected Vehicles, Aerospace Systems, Automation, Manufacturing, Smart Grids, Nonlinear Systems, Power Systems, Robotics, Social Systems, Economic Systems and other. Of particular value to both the contributors and the readership are the short publication timeframe and the worldwide distribution and exposure which enable both a wide and rapid dissemination of research output.

The series covers the theory, applications, and perspectives on the state of the art and future developments relevant to systems and networks, decision making, control, complex processes and related areas, as embedded in the fields of interdisciplinary and applied sciences, engineering, computer science, physics, economics, social, and life sciences, as well as the paradigms and methodologies behind them.

Indexed by SCOPUS, EI Compendex, INSPEC, WTI Frankfurt eG, zbMATH, SCImago.

All books published in the series are submitted for consideration in Web of Science.

For proposals from Asia please contact Aninda Bose (aninda.bose@springer.com).

Surekha Lanka · Antonio Sarasa Cabezuelo ·
Alexandru Tugui
Editors

Trends in Sustainable Computing and Machine Intelligence

Proceedings of ICTSM 2025

Springer

Editors
Surekha Lanka
IT Department
Stamford International University
Bangkok, Thailand

Antonio Sarasa Cabezuelo
Facultad de Informática
Universidad Complutense de Madrid
Madrid, Madrid, Spain

Alexandru Tugui
Department of Business Informatics
Alexandru Ioan Cuza University
Iasi, Romania

ISSN 2367-3370 ISSN 2367-3389 (electronic)
Lecture Notes in Networks and Systems
ISBN 978-3-032-13176-8 ISBN 978-3-032-13177-5 (eBook)
https://doi.org/10.1007/978-3-032-13177-5

© The Editor(s) (if applicable) and The Author(s), under exclusive license
to Springer Nature Switzerland AG 2026

This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

If disposing of this product, please recycle the paper.

Dedication

We are honored to dedicate the proceedings of ICTSM 2025 to all the participants, organizers and editors of ICTSM 2025.

Preface

We, the editors of ICTSM 2025, are proud to present the conference proceedings of the 3rd International Conference on Trends in Sustainable Computing and Machine Intelligence (ICTSM 2025), organized on 19–20 September 2025 in Bangkok, Thailand.

ICTSM 2025 is organized with an objective to leverage an international platform for the engineers, academicians, industrial experts as well as research enthusiasts, to present and discuss about the state-of-the-art research progress in all the domains of intelligent and sustainable computing technologies and applications. The conference has provided equal opportunities for all the visited delegates to exchange their research experiences and ideas; create research connections and business relations; and explore the global opportunities for future research collaboration. We strongly believe that the conference outcomes will lead to gain significant knowledge expertise in the emerging fields of machine intelligence.

The two-day conference event has been organized in association with the Research Centre of Stamford International University by including invited lectures, keynote talks, and oral presentations to enrich the existing knowledge of conference participants and ignite them to conduct more impactful research. ICTSM 2025 has received 282 research manuscripts, out of which 51 papers were accepted by our technical reviewers. Among these, 47 papers have been finalized for publication. All the submitted manuscripts have undergone the peer-review process and reviewed by appropriate research experts selected by the technical conference program chairs.

We would like to thank all the conference participants for their contribution to ICTSM 2025 proceedings. We thank the conference organizing committee for their assistance throughout the conference program. We sincerely thank all the reviewers for their continuous efforts in reviewing all the submitted manuscripts.

We hope that all the conference participants and readers will gain a technically rewarding experience from the ICTSM 2025 proceedings.

Guest Editors

Surekha Lanka
Antonio Sarasa Cabezuelo
Alexandru Tugui

Contents

Analysing the Future of Electric Vehicles and Carbon Dioxide Emission with Data Science Application: A Literature Review 1
Christhoper Robin and Lina Gozali

Securing 5G Network Slicing: A Hybrid Machine Learning Approach for DDoS Threat Detection and Performance Evaluation 12
Sahar Ebadinezhad and Ibrahim Khalaf Hussein

Enhancing Crude Oil Price Prediction with Neural Network Models 30
Hitesh Punjabi, S. Kumar Chandar, and Mayur Malik

TongueViT: A Vision Transformer-Based Framework for Automated Tongue Image Classification 45
Darshanaben Dipakkumar Pandya, Ishaan Tamhankar, Jaimini Santosh Kulkarni, Vishal Rajendrakumar Trivedi, Sheshang Degadwala, and Dhairya Vyas

Efficient Disease Classification in Cabbage and Its Leaves via Lite Separable Convolutional Neural Network 56
Rajendra Kachhava, Snehlata Ajmera, Vartika Vyas, Awanit Kumar, Sheshang Degadwala, and Dhairya Vyas

Revisiting Intrusion Detection: A Comparative Benchmark of Supervised Learning Models on the KDD Cup 1999 Dataset 66
Phuc-Hung Pham Le, Hong-Nhung Nguyen Thi, Trung-Hieu Huynh Nhat, and Huy-Minh Doan Xuan

A Novel Framework on Cardiovascular Disease Prediction Using Transfer Learning Technique 81
Manisha Guduri, Ashok Polavarapu, Mohammad El Yabroudi, Sandeep Kumar Thota, Hari Suresh Babu Gummadi, and Narendra Chennupati

Development and Deployment of a Machine Learning-Based Threat Detection System to Enhance Cybersecurity at State University 94
Glorybe E. Alegre and Patrick D. Cerna

Brain Tumor Detection Using Machine Learning: A Comprehensive Review 110
Sheshang Degadwala and Dhairya Vyas

Application of Deep Learning Methods for Alzheimer's Stage Classification Based on Brain MRI 117
Yass Khudheir Salal and Rawa Ali Hassan

OptiSeg-Edge: Hybrid Algorithm for Segmentation and Edge Detection 133
B. Arivazhagan, M. Inbavel, K. Padmavathi, and P. PratheepKumar

BHFR: Blockchain-Based Human-in-the-Loop Facial Recognition System Using Edge-Enabled IoT for Ethical Decisions 145
Sushmoy Dey, Bhagyalakshmi Muralidharan, and Shajulin Benedict

Ethical Frameworks for AI-Enhanced Interprofessional Learning: A Review of Governance Challenges in Cross-Disciplinary Medical Education 158
Loso Judijanto

Scaling AI Agents in the Enterprises: A Strategic Framework, Architecture, Governance, and KPIs 174
Vijay Kumar Butte, Sujata Butte, and Sagar Suryakanth

A Novel Cross-Layer Machine Learning Model for Seamless Cloud–Edge–IoT Interactions 189
Pallati Narsimhulu, Rajanikanth Aluvalu, and Premkumar Chithaluru

Modified Metaheuristic Optimization: An Application in Cloud Log Anomaly Detection 203
Nebojsa Bacanin, Luka Jovanovic, Snezana Anetic, Branislav Radomirovic, Miodrag Zivkovic, and Csaba Varsandán

Evaluating the Effectiveness of Combined User Restriction Methods 215
Shukhrat Kobiljonov, Elshod Khaydarov, and Fayzullajon Botirov

Info-Activity Method for Efficiency and System Potential Pragmatic Problems Solving 223
Alexander S. Geyda

Emergency Technologies and Its Connection in Image Processing and Image Segmentation 233
Mária Ždímalová

Secure and Sustainable AI-Driven Virtual Healthcare Assistants: Privacy-Preserving NLP for Future-Ready Patient Care 248
Sahar Ebadinezhad, Olusegun Emmanuel Adeshina, and Mohammed Al-Hussein

Performance Evaluation of Machine Learning Based Forecasting Model for Bank Loan Prediction 264
Kathan Nitin Patel and Aviral Sharma

Securing the Pharmaceutical Supply Chain: AI-Augmented Vendor Risk Management via SAP Ariba 283
Venkata Manoj Kumar SomiSetty

A Review of Analysing Heavy Metal Pollution with Data Science Approaches 296
Lina Gozali and Christyfianie Elva Angelica

Federated AI and Master Data Management: A Privacy-First Approach in Healthcare 312
Vinod Thallapally

Prospects of Using Neural Network Models for Forecasting Tax Revenues 323
Ulugbek Narmanov, Abdurakhim Vakhabov, Otabek Narmanov, Khayrillo Rejapov, Guzal Alimova, and Azamat Maqsudov

Identifying Synonymous Phrases Using the GloVe Algorithm 335
Axmedova Xolisxon, Omarova Durdona, Mamaraufov Odil, and Djabbarova Habiba

N-Gram Language Model for Uzbek Texts 346
Elov Botir Boltayevich, Tojieva Gulbahor Nomozovna, Tokhtaeva Matluba Xakim qizi, Jurayeva Nargiza Jamolidinovna, and Primova Mastura Hakim qizi

Sustainable Water Forecasting and Wastewater Management Using Fusion-Based AI and Smart Sensors 358
R. C. Jeni Gracia and G. Michael

Securing AI-Driven Biometric Authentication Systems: Cryptographic Techniques for Protecting Sensitive User Data 370
Sahar Ebadinezhad, Fulbert Milrich Kiminou Nsimba, Hauwa Adamu Wakil, and Joyeuse Maombe

Who's Responsible for Online Privacy: Investigating User Behaviour Through Social Learning and Protection Motivation Theory 380
Saide Saide, Nurwahyu Alamsyah, M. Ikhsan Ramadhan, and Nurfitria Ningsi

Machine Learning Framework for Sustainable UAV Spraying: Life Cycle and Computational Intelligence-Based Assessment 391
Shefali Vinod Ramteke, Pritish Kumar Varadwaj, and Vineet Tiwari

From Model to Field: Cloud-Based Web Deployment of a Two-Stage Deep Learning Framework for Tea Leaf Disease Detection 408
Harjinder Singh and Rizwan Rehman

Application of Artificial Intelligence in Managing Chloride Stress Corrosion Cracking in Stainless Steel Assets 418
JohnsonFoyken Jones, Veena Raj, Pg Emeroylariffion Abas, M. I. Petra, and Mohammad Fa'ezul Fitri Latiff

Leaf Disease Detection and Classification Using Deep Learning Technique and Deep Convolution Neural Network 431
R. Sumathi, Nalladimmu Sasisri, Kunchapu Gowri Sai, Kapilavai Hahumaan, and Kurakula Lokesh

Geospatial Analysis of Surface Urban Heat Island Intensity Using Google Earth Engine 446
M. Mahmood Pasha, Sharmila Banu, and Zameer Gulzar

Prediction of Insurance Claim Risk Using an Optimized Decision Tree Model 456
Oleg Pursky, Tetyana Filimonova, Darina Bondarenko, Tetyana Tomashevska, Pavlo Demidov, and Ihor Tyschenko

TabNet for Intrusion Detection: Bridging Accuracy and Interpretability in Tabular Network Data 470
Joideep Banerjee and Asma Patel

Design and Optimization of Rectangular-Shaped Patch Antenna for Wireless Applications at Ka Band 485
Jannatul Mim Samia, Md. Sohel Rana, Mithun Bairagi, and A. S. M. Tanvir Ul Islam

Achieving Net Zero: A Review of Greenhouse Gas Emission Forecasting Models and Approaches 498
Katya Victory Liu and Lina Gozali

A Machine Learning Approach to Crude Oil Price Prediction Using Support Vector Machine (SVM) 509
Hitesh Punjabi, S. Kumar Chandar, and Mayur Malik

Towards Intelligent Music Recommendation: A Multi-Gate Mixture of Experts Framework Integrating Audio, Behavior, and Social Contexts 527
A. Ferminus Raj, S. Abarna, J. Relin Francis Raj, R. Santhana Krishnan, S. Balamurugan, and D. Suresh

Children Food Intake Level Based on Deep Learning Using Variational Autoencoder-Gated Recurrent Neural Network (Vae-Grn2) 544
P. Jeyanthi and R. Durga

A Computational Approach for Framing Place Rebranding, Architecture and Public Policy: The Case of Skopje 2014 Project 560
Sabina Kohlmann, Anastasios Fountis, Katja Udir Misic, Navya Gubbi Sateeshchandra, and Christos Lemonakis

Basic Manobo Expression Dictionary: A Mobile-Based Application 581
Kylene Kyte B. Abandula, Joen Rose S. Berdida, Jade G. Cuartero, Ignel T. Balmora, Patrick Jasson S. Pojas, and Jeanette M. Eborda

BaCERA: A Smart Mobile and Web Solution for Community Emergency Response and Disaster Risk Management 595
Albert D. Silagan, Mery Cris G. Asis, Nancy P. Jacosalem, Jocelyn O. Balolot, Bernie S. Balighot, and Elmer E. Estandarte

Integrated Community Information and Service Management System 611
Jeannette V. Quijada, Ziara Amor E. Banggoy, Michelle C. Elape, James Cloyd M. Bustillo, Jeanie R. Delos Arcos, and Jade G. Cuartero

Personalized Chatbot Solutions for University Students: A Design Science Approach 625
Fatima Amer Jid Almahri and Zameer Gulzar

Author Index 637

Analysing the Future of Electric Vehicles and Carbon Dioxide Emission with Data Science Application: A Literature Review

Christhoper Robin and Lina Gozali(✉)

Department of Industrial Engineering, Universitas Tarumanagara, Jakarta, Indonesia
christhoper.545210027@stu.untar.ac.id, linag@ft.untar.ac.id

Abstract. Global warming is one of our most significant problems, and its impact can devastate humanity. One of the causes of global warming is the transport-related CO_2 emission. EVs are believed to be a great solution to this problem. However, there are only a few studies studying this problem. Therefore, this research aims to understand the gap in the CO_2 emissions and EV sales/production/users forecast and the correlation between the two variables. This research utilises the ScienceDirect search engine to gather the articles. Out of the 3,877 articles that have been collected, eight relevant papers were selected to be reviewed. All of these articles are related to EVs, with 7 of them discussing the CO_2 emissions of EVs, two papers are doing assessments on the life cycle of EVs, three papers are not researching forecasting methods for CO_2 emissions, five papers are directly correlating EV with their emissions, and 3 of them are doing comparative studies. Based on the articles that have been reviewed, most research was conducted in China, with four articles published. Most articles do research with minimal data; therefore, the results do not accurately represent the actual conditions of CO_2 emissions. Two articles use novel forecasting methods. However, no research has compared these methods to the simpler models, such as the different types of regression models.

Keywords: Carbon Dioxide Emissions · Electric Vehicles · Linear Regression · Machine Learning · Mathematical Modelling

1 Introduction

Global warming is one of the most significant problems that humanity has ever faced. The rise in average global temperature has caused detrimental effects towards the environment. One of these effects is the change in weather patterns across the globe. Hanlon et al. (2021) found that the United Kingdom is experiencing extreme temperatures and rainfall due to global warming, disrupting the nation's health, transportation, and energy sectors [1].

This increase in average global temperature is caused by human activities in the past 150 years [2]. Since the Industrial Revolution started in the eighteenth century, the concentration of CO_2 in the atmosphere has consistently risen. It has risen by 50%

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026
S. Lanka et al. (Eds.): ICTSM 2025, LNNS 1755, pp. 1–11, 2026.
https://doi.org/10.1007/978-3-032-13177-5_1

since 1750 [3]. This increase is because humans have started to burn fossil fuels, and the transportation sector has contributed around one-fifth of global CO_2 emissions [4, 5].

Intergovernmental Panel on Climate Change (IPCC) stated that the increase in average global temperature cannot reach more than 1.5 °C. If the temperature increases beyond this threshold, it will start to have an even more extreme impact towards the environment and human beings. This condition ranges from the increased risk of extreme flooding, rainfall reduction, and seawater acidification [6]. Therefore, a transition from internal combustion engine vehicles (ICEV) to electric vehicles (EV) is necessary to decrease the production of CO_2 emissions and to meet the goal of reaching net zero emissions.

EVs can be a promising solution to reduce CO_2 emissions [7]. In the past few years, many countries have seen the sales of EVs steadily increasing [8]. This increase in ownership is driven by the net-zero goals set by many countries and local governments. Governments across the world are now offering incentives to EV owners to increase the usage of EVs. These incentives range from tax credits to EV price subsidies. In the United States, for example, a tax credit of up to USD 7500 may be given to selected new EVs or up to USD 4000 for selected used EVs [9]. Furthermore, many major cities around the world, such as Jakarta and London, are now starting to shift from internal combustion engine buses (ICEB) to battery electric buses (BEB) to help achieve these goals [10, 11].

However, it is worth noting that the effectiveness of EVs in achieving net zero is heavily dependent on the energy source of the local grid. A grid that relies on fossil fuels for energy generation is far less effective in achieving net zero than a grid mainly powered by clean and renewable energy sources [7]. This condition remains a massive challenge for countries like China, where 201.9 had around 67% of its electricity generated by coal and natural gases, but is expected to go down to 45% by 2030 [12].

With this growth in EV sales and ownership, it is becoming increasingly important to understand the relationship between EV sales and CO_2 emissions and the development of these two aspects. Data science uses mathematics and programming to understand, analyse, model, and solve data-related problems [13, 14]. Therefore, it plays a crucial role in understanding the relationship between EV sales and CO_2 emissions and predicting each aspect's growth to see how it will change over time. The research gap found in this article will contribute to finding what researchers have yet to discover. Then, a further research topic can be recommended based on this knowledge gap.

This research is utilising the ScienceDirect® search engine to gather articles. From the search results, relevant articles are selected. This research aims to find out the gap in the EV and CO2 emission forecast research. This research also aims to find research conducted on the correlation between EV production and CO_2 emissions.

2 Research Method

This research seeks to identify the gap in knowledge of carbon dioxide emissions. The first step is to use the scientific articles search engine [15]. In this research, the articles researched are gathered from the ScienceDirect® article search engine. The key terms used are "Carbon Dioxide Emissions, Electric Vehicle Production, Data Science, and

Forecast. Then, irrelevant articles are excluded [16]. From this search result, nine articles are chosen. The selection criterion is the relevance of these articles to the objective of this research. These literatures are then reviewed and compared to each other to learn the existing research that has been done. Then, a research gap analysis is conducted to understand the current knowledge gap in the correlation of EV production and CO_2 emissions forecasting.

3 Results and Discussion

Of the 3877 articles gathered in the search, eight articles were selected. Table 1 shows the authors, titles, year of publication, and the names of the scientific papers where the articles have been published.

Table 1. Articles Used in This Research

Author(s)	Title	Year of Publication	Name of Journal/Conference
Blades et al.	Predicting energy consumption of zero-emission buses using route feature selection methods [17]	2025	Transport Research Part D: Transport and Environment
Bandpey et al.	Greenhouse gas impact related to mineral mining and processing [18]	2024	57th CIRP CMS 2024
Ji et al.	Urban transport emission prediction analysis through machine learning and deep learning techniques [19]	2024	Transport Research Part D: Transport and Environment
Fu et al.	Transportation emissions monitoring and policy research: integrating machine learning and satellite imaging [20]	2024	Transport Research Part D: Transport and Environment
Dauphin et al.	Evaluation of plug-in hybrid vehicles in actual conditions by simulation [21]	2023	Transport Research Part D: Transport and Environment
Liu et al.	Carbon emission reduction prediction of new energy vehicles in China based on the GRA-BiLSTM model [22]	2023	Atmospheric Pollution Research
Rizqi, et al.	Green energy mix modelling under supply uncertainty: hybrid system dynamics and adaptive PSO approach [23]	2023	Applied Energy
Xin et al.	Impact of Innovation in Hybrid Electric Vehicles-Related Technologies on Carbon Dioxide Emissions in the 15 Most Innovative Countries [24]	2023	Transport Research Part D: Transport and Environment
Ağbulut	Forecasting of transportation-related energy demand and CO_2 emissions in Turkey with different machine learning algorithms [25]	2022	Sustainable Production and Consumption

(*continued*)

Table 1. (*continued*)

Author(s)	Title	Year of Publication	Name of Journal/Conference
Al-Dabbas et al.	Modelling pathways towards a climate-neutral EU industry sector [26]	2022	ECEEE Summer Study 2022
Bakir et al.	Forecasting of future greenhouse gas emission trajectory for India using energy and economic indexes with various metaheuristic algorithms [27]	2022	Journal of Cleaner Production

3.1 Research in CO_2 Emissions Forecast

Ji et al. [19] found that of the three different learning machine methods—ANN algorithm, SVM, and DL algorithm—SVM gives the most favourable outcome because it produces the best result on three out of five error criteria of mean error (ME), mean average percentage error (MAPE), and relative root mean square error (RMSE). In this research, Ji et al. also forecasted that China's CO_2 emissions and energy demand will continue to rise in the coming years. This forecast is obtained through exponential and linear regression modelling.

Rizqi et al. [23] researched the different impacts of changes in energy mix using CLD and SFD modelling to understand the relationship between electricity supply and demand, EV usage, and CO_2 emission. Then, they utilised mathematical modelling and machine learning to create parameter estimation and use PSO for parameter tuning. They discovered that if the objective is to minimise the total system cost of electricity production, then reducing the usage of RES will decrease the total system cost by 38% but will increase the total CO_2 emission by 8%. However, if the objective is to minimise the total CO_2 emission in the system, then relying solely on RES to produce electricity would reduce the electricity by 55% and increase the cost by 68%.

Ağbulut [25] researched transport-related CO_2 emissions using mathematical modelling and a machine learning algorithm. The results showed that machine learning algorithms produce a slightly better result than mathematical modelling. However, it must be noted that machine learning algorithms are far more complicated than mathematical modelling, and it does not produce significantly better results.

Al-Dabbas et al. [26] conducted research on the CO_2 emissions in three different scenarios in the EU, including current EU policy (CP), 95% greenhouse gas emissions reduction (MIX 95), and net zero emission (NZE). They found that the CP scenario is projected to cause GHG emissions to decline by 26% by 2030 and 46% by 2050, less than the EU Green Deal and Fit target of 55% GHG emissions reduction. In the MIX 95 scenario, significant technological deployment is necessary to fulfil this scenario. Under this scenario, GHG emissions are expected to be reduced by up to 95% by 2050. Coal and natural gas consumption are expected to decline. However, natural gas will still play an essential role in the EU's energy supply in the medium term. Under this scenario, biomass usage will grow steadily and play a crucial role in its energy mix. In the MIX95 scenario, electricity demand will significantly increase due to the need

fora carbon capture and sequestration (CCS) system and hydrogen production using electrolysis.

Bakir et al. [27] researchedGHG emissions using metaheuristic optimisation algorithms. Based on this research, most of the algorithms produces a satisfactory result with R^2 ranging from 88% to 99% for CO_2 prediction, 28.3% to 99% for CH_4 predictions, 90% to 99% for fluorine gas predictions, 47% to 99% for N_2O predictions, and 90% to 99% for total GHG emissions. This research shows that in 2050, GHG emissions in India are predicted to increase by 2.1 to 2.4 times compared to the time of the research.

3.2 Research in Electric Vehicle CO_2 Emissions

In [17], Blades et al. found that estimating the energy consumption using a Synthetic Drive Cycle produces similar results to using a Real Drive Cycle. They used data from Braunschweig City, the Manhattan bus and the Orange County bus for the Synthetic Drive Cycleycle. They utilised data from Transport for London (TfL) U5 & U7 bus and Belfast 3A bus routes for the Real Drive Cycle. This result is obtained based on the MAPE comparison. Therefore, it is concluded that the Synthetic Drive Cycle is more efficient in predicting battery electric bus (BEB) energy consumption as it is easier to do and costs less than the Real Drive Cycle.

In [21], Dauphin et al. compared PHEV emission to ICEV using SimcenterAmesim™ software for data collection and then used sensitivity analysis to find the fuel consumption of PHEV due to recharging frequency, distance travelled, and battery capacity. They found that a PHEV with a battery of 15 kWh has a fuel consumption of 2.25 L/km and a utility factor of 77% if it is recharged daily. However, if it is recharged every three days, its fuel consumption and utility factor become 4.85 L/km and 48%, respectively. This condition results in a 116% increase in fuel consumption and a 29% decrease in utility. Therefore, after some distance, the utility factor of PHEV will increase. It also discovers that an increase in battery capacity to 30 kWh will only increase the utility factor to 87%. Lastly, they found that cold temperatures will increase battery usage and fuel consumption as the difference between the outside temperature and the living comfort temperature is more significant than when it is hot.

3.3 Research in Electric Vehicle Energy Consumption

Bandpey et al. [18] provide insight into GHG production produced by mineral mining and processing due to different climate change policies. This research divides the policies into three scenarios: AP, STEP, and NZE. The study found that NZE scenarios would produce twice as much CO_2e GHG as the STEP scenarios. It also shows that Al will contribute the most GHG emissions (around 60%), followed by Cu. Si and Ni are in the third and last positions, respectively. The increased usage of EVs, photovoltaic cells, heat pumps, and electricity transmission technologies strongly influences the production of Al and Cu. Production of Si is strongly influenced by photovoltaic cell production, while the increase in Ni production is caused by EV production and electricity production using geothermal energy.

Liu et al. [22] show that the grey relation analysis-bi-directional long short-term memory (GRA-BiLSTM) model is more accurate than other machine learning and single

prediction models. This research concludes that China's new energy vehicle ownership will grow. This condition plays a massive role in the reduction of CO_2 emissions. Liu et al. also found that the forecast for the fossil energy share percentage in China's year-to-year energy mix will decrease. It is expected that by 2060, China will have reduced 15.84 million metric tons of CO_2 emissions due to transportation.

Xin et al. in [24] utilised the cross-sectional-autoregressive-distributed-lag (CS-ARDL) to analyse how public-private partnership investment in energy (PPPIE), renewable energy consumption (RNEC), expansionary fiscal policy (EFP), contractionary fiscal policy (CFP), GDP per capita (GCPPC), and in-hybrid vehicles-related technologies (IHVRT) will affect CO_2 emitted by 15 most innovative economies. They found that increasing GDPPC and EFP will increase CO_2 emissions, while PPPIE, RNEC, CFP, and IHVRT will decrease CO_2 emissions.

3.4 Literature Comparison

Figure 1 shows the number of research conducted in different countries. Based on the graph, China is leading in this topic as it has four research articles. The United Kingdom follows this with three research articles, and Canada, France, Germany, Israel, Japan, Sweden, Switzerland, and the United States with two research articles. The rest of the countries have only conducted one research study. The research article by Bandpey et al. [18] is not performed in a specific country since it considers the global GHG emissions from mineral production and processing.

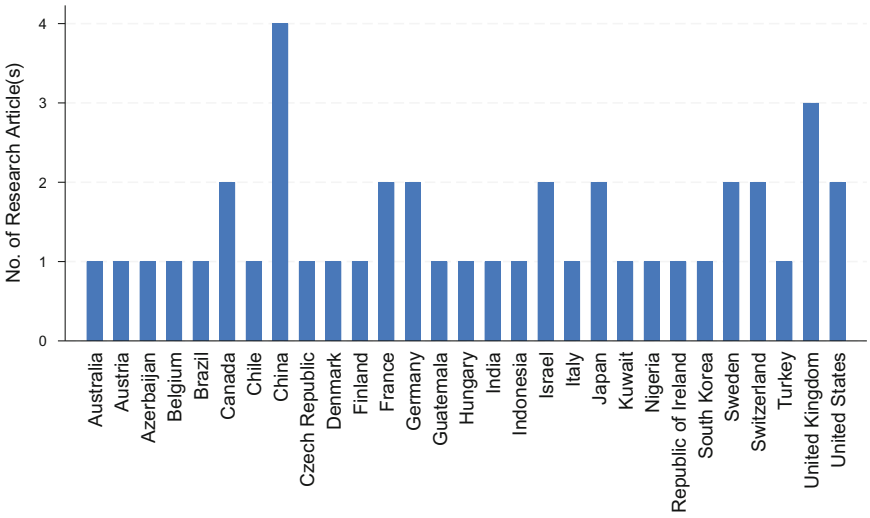


Fig. 1. Number of Research Article(s) Based on the Country

Based on Fig. 1, it can be seen that the research is conducted in high-income countries, except for Azerbaijan, China, Brazil, Guatemala, India, Indonesia, Nigeria, and Turkey. This figure shows that further research must be conducted in countries outside the high-income category, as the high-income countries dominate the research.

The comparison of the literature that has been reviewed is shown in Tables 2 and 3.

Table 2. Literature Comparison

Description	Author(s)			
	Blades et al. [17]	Bandpey et al. [18]	Ji et al. [19]	Fu et al. [20]
Electric Vehicle	✓	✓	✓	✓
CO_2 Emission	X	✓	✓	✓
Life Cycle Assessment	X	✓	X	X
Forecasting	LiR, LaR, RiR, RF, ET	X	X	CNN, GNN
Correlation between EV Production and CO_2 Emission	X	✓	X	✓
Comparative Study	✓	X	X	✓

LiR = linear regression, LaR = lasso regression, RiR = ridge regression, RF = random forests, ET = extra trees, CNN = convolutional neural network, GNN = graph neural network

Table 3. (Cont'd) Literature Comparison

Description	Author(s)			
	Dauphin et al. [21]	Liu et al. [22]	Rizqi et al. [23]	Xin et al. [24]
Electric Vehicle	✓	✓	✓	✓
CO_2 Emission	✓	✓	✓	✓
Life Cycle Assessment	X	X	✓	X
Forecasting	LiR	GRA-BiLSTM	MLR, PR, RiR, SVR	X
Correlation between EV Production and CO_2 Emission	✓	✓	✓	✓
Comparative Study	✓	X	X	X

LiR = linear regression, GRA-BiLTSM = Grey relation analysis-bi-directional long-short-term, MLR = multiple linear regression, PR = polynomial regression with second-order, RiR = ridge regression, SVR = support vector regression

Table 4. (Cont'd) Literature Comparison

Description	Author(s)			
	Ağbulut [25]	Al-Dabbas et al. [26]	Bakir et al. [27]	Klimenko et al. [28]
Electric Vehicle	X	X	X	
CO_2 Emission	✓	✓	✓	
Life Cycle Assessment	X	X	X	
Forecasting	DL, SVR, ANN, LiR, ER	DS	MPA, LSA, EO, SSS, ABS, LiR	GF, DM, R
Correlation between EV Production and CO_2 Emission	X	X	X	
Comparative Study	✓	✓	✓	✓

DL = deep learning, SVR = support vector regression, ANN = artificial neural network, LiR = linear regression, ER = exponential regression, DS = decarbonisation scenarios, MPA = marine predators algorithm, LSA = lighting search algorithm, EO = equilibrium optimizer, SS = symbiotic search, ABS = algorithm backtracking search, LiR = linear regression, GF = genetic forecasting, DM = dynamic modelling, R = regression modelling

As shown in Tables 3 and 4, most of the research that has been conducted uses minimal data.

1. [17, 21] and [22] only consider the TTW GHG emissions. [17] did research specifically on the energy consumption of BEB and only considers limited route characteristics (Braunschweig City driving cycle, Manhattan bus driving cycle, Orange County bus cycle, TfL U5 & U7 bus route, and Belfast 3A bus route). [18] only estimated the GHG emissions due to mineral production & processing and does not consider the GHG emissions that would be reduced when implementing these green technologies.
2. [19, 26, 27], and [27] create predictions for CO_2 emissions in different regions, ranging from China to the EU. These articles use various variables, including GDP, number of populations, energy mix, natural climate factor, political scenarios, and vehicle kilometres travelled.
3. [20] uses data from satellite imaging. Hence, its accuracy and availability can be limited in certain weather conditions, i.e. cloudy or rainy days.
4. [21, 24] only studied PHEV. [21] does not consider the impact of PHEV users' behaviour towards GHG emissions, which varies from individual to individual. In [24], the RNEC variable does not consider the availability and cost of renewable energy for electricity generation.
5. [23] Assumed that power plants and energy storage systems can run indefinitely and will not degrade over time. This result will impact the CO_2 emissions emitted by the system as it needs maintenance or even new power plants to be built.

Research has been conducted utilising various methods to forecast CO_2 emissions. [17] used various regression modelling, random forests and extra trees modelling. [20] utilised two different neural network models, namely CNN and GNN [21]. Only used a linear regression model [22]. On the other hand, tried to integrate the grey relation analysis with bi-directional long-short-term memory modelling [23]. It uses the multiple, polynomial, ridge, and support vector regression. In each research, the forecasting method has been proven to predict CO2 emissions and EV users accurately. However, there is no research comparing different models, especially the novel CNN and GNN model, with integrating satellite imaging and the GRE-BiLSTM model with conventional models such as different regression models.

Furthermore, out of all eight research articles that have been reviewed, only one has researched EVs and their CO_2 emissions throughout their life cycle, and a forecast was conducted. [23] forecasted changes in variables such as electricity demand for commercial and industrial, the number of vehicles, and the total energy stored. However, this research is conducted in the United States. Therefore, such research must also be conducted in other countries worldwide, especially countries that produce large amounts of CO_2 emissions, to understand how EVs will change their CO_2 emissions over the years.

4 Conclusion

Based on the research articles that have been reviewed, four studieswere conducted in China, three are in the United Kingdom, two are in Canada, France, Germany, Israel, Japan, Sweden, Switzerland, and the United States, and one in the rest of the countries mentioned in Fig. 1. All of the research was directly or indirectly conducted on EVs, with 7 of them discussing the CO_2 emissions of EVs, 2 doing assessments on the life cycle of EVs, 3 of them not researching forecasting methods for CO_2 emissions, five are directly correlating EVs with their emissions, and 3 of them are doing comparative studies. Additionally, most of the research that has been done only considered minimal data; therefore, research using a broad range of data, including satellite imaging, all types of EVs, government policies, etc., must be conducted to understand the impact of all these factors to CO_2 emissions produced in countries across the world. Based on the knowledge gap found in this research, further research must also be conducted to compare the novel forecasting methods with the other older models to understand the reliability of these forecasting methods compared to the older models.

References

1. Hanlon, H.M., Bernie, D., Carigi, G., Lowe, J.A.: Future changes to high impact weather in the UK. Clim. Change. **166**(3–4), 50 (2021)
2. Ring, M.J., Lindner, D., Cross, E.F., Schlesinger, M.E.: Causes of the global warming observed since the 19th century. Atmos. Clim. Sci. **2**(4), 401–415 (2012)
3. NASA Global Climate Change: Carbon dioxide concentration | NASA Global Climate Change. Climate Change: Vital Signs of the Planet. https://climate.nasa.gov/vital-signs/carbon-dioxide?intent=121. Accessed 06 Feb 2025

4. Balogh, J.M., Jámbor, A.: Determinants of CO_2 emission: a global evidence. Int. J. Energy Econ. Policy Econjournals. **7**(5), 217–226 (2017)
5. Ritchie, H.: Cars, planes, trains: where do CO_2 emissions from transport come from? Our World Data (2020). https://ourworldindata.org/co2-emissions-from-transport. Accessed 06 Feb 2025
6. Intergovernmental Panel on Climate Change: Special report: global warming of 1.5°C. https://www.ipcc.ch/sr15/. Accessed 06 Feb 2025
7. Costa, C.M., Barbosa, J.C., Castro, H., Gonçalves, R., Lanceros-Méndez, S.: Electric vehicles: to what extent are environmentally friendly and cost effective?—comparative study by European countries. Renew. Sust. Energ. Rev. **151**, 111548 (2021)
8. Muratori, M., Alexander, M., Arent, D., Bazilian, M., Cazzola, P., Dede, E.M., Farrell, J., Gearhart, C., Greene, D., Jenn, A., Keyser, M., Lipman, T., Narumanchi, S., Pesaran, A., Sioshansi, R., Suomalainen, E., Tal, G., Walkowicz, K., Ward, J.: The rise of electric vehicles—2020 status and future expectations. Prog. Energy. **3**(2), 022002 (2021)
9. U.S. Department of Energy: Electric vehicles. Energy.gov. https://www.energy.gov/save/electric-vehicles. Accessed 06 Feb 2025
10. Logan, K.G., Nelson, J.D., Hastings, A.: Electric and hydrogen buses: shifting from conventionally fuelled cars in the UK. Transp. Res. Part D, Transp. Environ. **85**, 102350 (2020)
11. Ginting, M.G., Reguyal, F., Cecilia, V.M., Wang, K., Sarmah, A.K.: Electrification of public buses in Jakarta, Indonesia: a life cycle study. Sci. Total Environ. **914**, 169875 (2024)
12. Shafique, M., Luo, X.: Environmental life cycle assessment of battery electric vehicles from the current and future energy mix perspective. J. Environ. Manag. **303**, 114050 (2022)
13. Robinson, E., Nolis, J.: Build a career in data science. Manning (2020). http://ieeexplore.ieee.org/document/10280423
14. Gupte, M.A.: Literature review and research gap. In: Success of University Spin-Offs: Network Activities and Moderating Effects of Internal Communication and Adhocracy, pp. 15–37. Deutscher Universitäts-Verlag, Wiesbaden (2007)
15. Helfman Cohen, Y., Reich, Y., Zohar, Y.: Literature review conclusions and definition of research target. In: Biomimetic Design Method for Innovation and Sustainability, pp. 45–49. Springer, Cham (2016)
16. Lin, H., Li, M.: Practitioner's Guide to Data Science. Chapman & Hall/CRC Data Science Series. Chapman & Hall/CRC, Philadelphia (2023)
17. Blades, L.A.W., Matthews, T., McGrath, T.E., Early, J., Cunningham, G., Harris, A.: Predicting energy consumption of zero emission buses using route feature selection methods. Transp. Res. Part Transp. Environ. **130**, 104158 (2024)
18. Bandpey, A.F., Golroudbary, S.R., Kraslawski, A.: Greenhouse gas impact related to minerals mining and processing. Proc. CIRP. **130**, 1001–1006 (2024)
19. Ji, T., Li, K., Sun, Q., Duan, Z.: Urban transport emission prediction analysis through machine learning and deep learning techniques. Transp. Res. Part Transp. Environ. **135**, 104389 (2024)
20. Fu, H., Li, H., Fu, A., Wang, X., Wang, Q.: Transportation emissions monitoring and policy research: integrating machine learning and satellite imaging. Transp. Res. Part Transp. Environ. **136**, 104421 (2024)
21. Dauphin, R., Prevost, V., Degeilh, P., Melgar, J., Fittavolini, C., Smith, A., Callu, C., Chrysafi, S., Uitz-Choi, R., Kar, K.: Evaluation of plug-in hybrid vehicles in real-world conditions by simulation. Transp. Res. Part Transp. Environ. **119**, 103721 (2023)
22. Liu, B., Wang, S., Liang, X., Han, Z.: Carbon emission reduction prediction of new energy vehicles in China based on GRA-BiLSTM model. Atmospheric Pollut. Res. **14**(9), 101865 (2023)
23. Rizqi, Z.U., Chou, S.-Y., Yu, T.H.-K.: Green energy mix modeling under supply uncertainty: hybrid system dynamics and adaptive PSO approach. Appl. Energy. **349**, 121643 (2023)

24. Xin, L., Ahmad, M., Khattak, S.I.: Impact of innovation in hybrid electric vehicles-related technologies on carbon dioxide emissions in the 15 most innovative countries. Technol. Forecast. Soc. Change. **196**, 122859 (2023)
25. Ağbulut, Ü.: Forecasting of transportation-related energy demand and CO_2 emissions in Turkey with different machine learning algorithms. Sustainable Prod. Consumption. **29**, 141–157 (2022)
26. Al-Dabbas, K.: Modelling pathways towards a climate-neutral EU industry sector'. In: ECEEE 2022 Summer Study (2022)
27. Bakır, H., Ağbulut, Ü., Gürel, A.E., Yıldız, G., Güvenç, U., Soudagar, M.E.M., Hoang, A.T., Deepanraj, B., Saini, G., Afzal, A.: forecasting of future greenhouse gas emission trajectory for India using energy and economic indexes with various metaheuristic algorithms. J. Clean. Prod. **360**, 131946 (2022)

Securing 5G Network Slicing: A Hybrid Machine Learning Approach for DDoS Threat Detection and Performance Evaluation

Sahar Ebadinezhad[1,2(✉)] and Ibrahim Khalaf Hussein[2]

[1] Department of Computer Information Systems, Near East University, Nicosia 99138, North Cyprus
sahar.ebadinezhad@neu.edu.tr, 20234858@std.neu.edu.tr

[2] Department of Computer Information Systems, Near East University, Nicosia 99138, North Cyprus

Abstract. The growing popularity of 5G networks has resulted in distinctive features such as network slicing, which allows for the establishment of several virtualized networks on a shared physical infrastructure. While network slicing improves flexibility, scalability, and service customization, it also creates severe security flaws, particularly for Distributed Denial-of-Service (DDoS) attacks. This research analyzes the effect of DDoS attacks on network slice performance in SDN/NFV-enabled 5G settings and proposes a predictive protection mechanism known as the Network Slice Service Protection System (NSSPS). The system uses machine learning-based classifiers, notably Decision Tree (DT) and Support Vector Machine (SVM), to detect and mitigate DDoS attacks on the NSL-KDD dataset. A real-time simulated setting replicating metropolitan infrastructure was utilized to assess the system's resilience to several attack types, such as TCP SYN Flood, UDP flooding, HTTP flooding, and RADIUS Attack. Performance measures, including accuracy, throughput, latency, and slice isolation, were evaluated. Experimental results demonstrate that the proposed NSSPS significantly improves detection accuracy (up to 98.4%) and maintains stable service quality under attack conditions, outperforming traditional and deep learning-based models. The study contributes to 5G network security by offering a lightweight, predictive, and scalable solution to protect critical slice resources in highly dynamic environments.

Keywords: Intrusion Detection · Machine Learning · Network Slicing · Software-Defined Networking · Predictive Analytics · 5G Network Security

1 Introduction

The advancement of wireless communication technologies caused a societal and industrial shift that gave birth to modern digital ecosystems. The fifth-generation (5G) mobile network aims to enable ultra-reliable low-latency communications, massive machine-type communications, and enhanced mobile broadband [1]. Network slicing supports

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026
S. Lanka et al. (Eds.): ICTSM 2025, LNNS 1755, pp. 12–29, 2026.
https://doi.org/10.1007/978-3-032-13177-5_2

these goals, serving as a fundamental architectural element that allows separate logical networks to work on a common physical infrastructure [2]. By means of network slicing, service providers can strategically tailor resource allocations to enhance their networks for particular applications. Network operators use slices to generate customized resources that meet the needs of entertainment services, applications in healthcare, and IoT installations [3]. The integration of SDN, NFV, and O-RAN represents a significant revolution in network control management, allowing service providers to build adaptable, flexible, and scalable networks [4, 5]. The advent of network slicing complicates operations, exposing numerous weaknesses in the security system. Overall, security risks are worsened, particularly because virtualized networks' capabilities are limited to single supplier domains and information and control planes, which are separated in SDN-based networks. The adoption of SDN allows for a division between control along with information planes [6, 7]. The greatest threat is posed by DDoS attacks, which drastically deplete virtualized slices and eliminate their capacity to remain isolated [8]. Evaluating network slicing's defensive capabilities against attacks is crucial to maintaining 5G network stability and service quality under demanding conditions [9].

Network slicing implementation in 5G networks has three key benefits, including flexibility, scalability, and personalized service delivery. The security risks of these benefits increase simultaneously with their implementation. Current security protocols employed in 4G networks and VPNs fail to work properly with the dynamic and multi-tenant aspects of 5G slicing services [3].

SDN controllers along with inter-slice dependencies and the 5G network environment's open-source base produce special security weaknesses [5, 10]. An assault on an individual network slice or shared control interface leads to a spreading impact that disrupts several network slices and fundamental services [11]. Moreover, the defense against volumetric cyber-attacks, especially DDoS attacks, presents an overwhelming challenge for network operators wishing to preserve both strong slice isolation and service quality [12].

The growth of virtualization alongside orchestration has not been sufficient for establishing predictive security frameworks that can defend slices against rapidly developing security threats [13]. The research explores the security weaknesses in 5G network slicing and introduces a sustained protection system.

The main contribution of this research is examining how DDoS attacks affect 5G network slice functionality and developing defense methods to maintain slice image quality while preventing blending of vital networks.

2 Literature Review

This section provides a review of 5G network slicing architecture, types of enabling technologies (SDN, NFV, O-RAN), and working principles, including isolation, customization, and scalability. It focuses on key security flaws, IoT-related risks, encryption, and multi-vendor environment issues. The chapter identifies critical research gaps and outlines future goals, including the need for AI-based security, standardized procedures, and edge protection in 5G networks.

2.1 Architecture and Principles of 5G Network Slicing

Network slicing for 5G is a game-changing enhancement in network administration that allows operators to set up multiple wireless networks from an individual physical site [1]. Each slice is unique, with specific QoS parameters, security settings, and resources. This versatility enables MNOs to better serve a wide range of industries, including high-reliability, low-latency applications for medical and high-bandwidth multimedia services that stream [3]. Three key ideas underlie the study, development, and operation of 5G network slicing: isolation, customization, and scalability. Isolation prevents issues from propagating to others, adjustments ensure that each slice meets QoS and security standards, and scalability allows for resource expansion or contraction [6]. Slicing is crucial for understanding secure aspects and developing objective strategies to improve network security [14].

2.2 Enabling Techniques for 5G Networks

The fundamental elements of 5G network slicing and architecture include many hierarchical and dependent layers and technologies, including Core Network (CN), Radio Access Network (RAN), Network Function Virtualization (NFV), and Software-Defined Networking (SDN). The core and radio communication network components serve as the foundation for network slicing by performing data processing and transmission operations [4]. The Core Network (CN) is the technology of data routing, connectivity, and service management in slicing, while the Radio Access Network (RAN) is wireless connectivity between users' devices and the network [15]. The improvement of the RAN, noticeably through O-RAN (Open Radio Access Network), to increase the flexibility of 5G slicing.

ORANSlice, a free tool for networking slicing based on O-RAN, is a good example of this method; it helps telecom operators to come up with and manage those dynamic network slices even in a very complex multi-vendor scenario. Thanks to this versatility and vendor neutrality, O-RAN brings extra flexibility and a lower cost when deploying the desired slices [2]. The basic principle of radio operation is fundamental to understanding telecommunications. All these principles involve the main interaction between 3 main components: base station, core network, and user's equipment.

2.3 Major Concerns for Data Encryption

Individuals may be at risk from the vast amounts of data carried by 5G networks, including digital wallets and private medical records. Encryption is a known technique for guaranteeing data secrecy and integrity. However, because some IoT devices have limited capabilities, it is difficult to enforce standards for encryption in 5G networks. Since data travels across borders and data protection regulations are weakened, the global scope of 5G wireless networks has changed data sovereignty. Although they also raise questions regarding data security, region-specific data protection laws may be able to allay these worries. Because there are many vendors in the worldwide 5G wireless network supply chain, there is a considerable chance that intelligence operatives will take

advantage of vulnerabilities. Malware and backdoors may be introduced, leading to disruptions and problems with data confidentiality. Extreme hazards are associated with critical infrastructure. Operators need to focus on service quality and choose cutting-edge providers [16]. There are various challenges that need to be met to enhance the successful development and deployment of 5G networks. Among these are spectrum, which remains an important factor to get enough bandwidth for the required network; security threats, which make networks prone to attack; setting up and deployment of network slicing to provide multiple services; and finally, the issue of making the network expand to cater for the increasing traffic demand.

Advancement in 5G technology, especially network slicing, has brought dynamic change in the telecommunication industry, where flexibility, scalability, and customization have become key points with reference to different applications. The literature especially relates to the principles and architectures of 5G network slicing, wherein different technologies like SDN, NFV, and O-RAN have gained considerable emphasis. Still, there are numerous challenges that are obvious for anybody, and the key three of them are security, standardization, and scalability. The specific contributions of our study are:

- Analyze and model the degradation behavior of network slices during different types of DDoS attacks.
- Develop and validate a Network Slice Service Protection System (NSSPS) utilizing predictive analytics.
- Design and integrate a predictive-based mitigation algorithm into the SDN control plane to enable dynamic, proactive defense mechanisms.

Simulate attack scenarios and evaluate the system performance using real-world datasets, particularly the NSL-KDD dataset [17], and performance metrics such as throughput, packet loss, and latency.

3 Methodology

This research proposed a novel method, as Network Slicing Service Protection System (NSSPS), which uses both Decision Tree (DT) and Support Vector Machine (SVM) classifiers to detect DDoS attacks in 5G network slicing environments. The process begins with the input of traffic features extracted from the NSL-KDD dataset [17] is used based on 20% testing, 70% training, and 10% validation. Figure 1 illustrates the workflow of our NSSPS proposed method. The workflow diverges into two parallel classification paths. In the DT path, preprocessing is performed (including normalization and encoding), followed by feature selection using information gain, training of the DT classifier, and finally DT-based prediction (normal or attack). In parallel, the SVM path also undergoes preprocessing, then applies feature mapping to a higher-dimensional space before training the SVM classifier and making its own prediction. Both predictions are forwarded to a fusion module, which compares their outcomes. If the predictions are consistent, a high-confidence final label is assigned; otherwise, a reevaluation or majority voting mechanism is triggered. The system concludes by outputting the classified traffic and, if necessary, initiating a slice isolation response to protect the 5G network from the detected threat.

The proposed approach was deliberately chosen to address current limitations in both detection accuracy and response adaptability, thereby providing a practical and effective solution for securing 5G network slices.

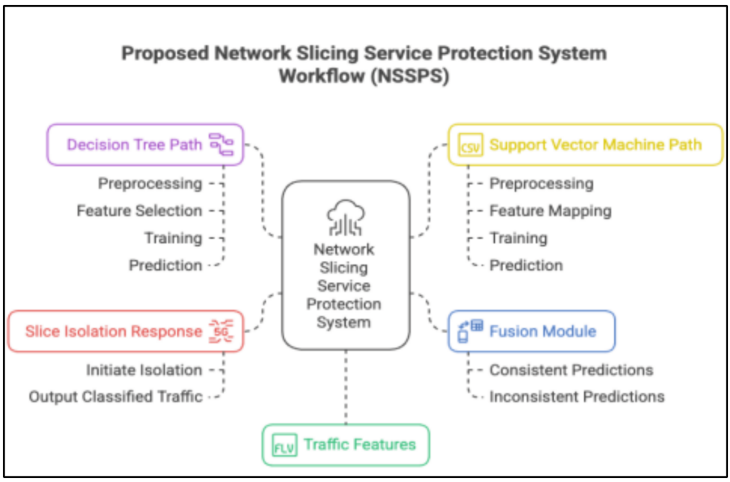


Fig. 1. Workflow of NSSPS.

3.1 Real-Time Simulation Environments

This research is implemented based on the discrete event simulation (sim) tool, which allows for imitation of the network slicing in 5G New Radio (NR) systems. This framework makes it possible to evaluate different slicing techniques with client movement behaviors in a constantly evolving environment. Moreover, Python version 3.7 and some libraries such as Simpy event simulation and KDTree are used for this implementation for energy function spatial indexing and optimal client-to-base station association. A geographically concentrated simulation area of 2 km $\times$ 2 km is chosen with its specific boundaries. The simulation focus is an Istanbul Old City (Fatih) subregion. This region is selected due to realistic patterns of population movement and social contact-driven mobility generation. movement caused by social interactions. The base stations are positioned throughout the region in real-world coordinates. Coverage areas are circular with variable radius. The simulation is set to run for a duration of 3600 s (1 h). To achieve consistent results, a startup and a shutdown phase (5% of period 1 and 5% of period 2) is omitted from the session analysis. Rest period to avoid potential model training rest period during session analysis. Clients are allocated throughout the simulation setting based on specific defined movements, with each client following a particular statistical distribution, which can be seen in Table 1.

Also, security parameters for 5G network slicing are shown in Table 2.

Table 1. Simulation Environment for 5G Network Slicing

Parameters	Values
Simulation Framework	Simpy (Python 3.7)
Programming Language	Python 3.7
Visualization Library	Matplotlib
Spatial Indexing	KDTree
Geographical Simulation Area	Istanbul (Fatih), 2 km × 2 km
Base Station Distribution	Real-world locations
Base Station Coverage Model	Circular regions of varying sizes
Simulation Duration	3600 s (1 h)
Warmup & Cooldown Period	5% of total time each
Number of Clients	Dynamic (Varies by scenario)
Client Mobility Patterns	Normal, Random Integer, Stationary
Car Mobility Model	Normal Distribution (Mean: 0, Std: 7)
Pedestrian Mobility Model	Random Integer Distribution
Stationary Users	Normal Distribution (Low Std. Dev)
Tram & Other Vehicles	Random Integer Distribution
Network Slicing Strategy	Dynamic Slicing
Types of Network Slices	eMBB

Table 2. Security Parameters for 5G Network Slicing

Parameters	Values
Network Model	3 Data Centers
Network Slice (SN)	1 Slice per test
Slice Elements (SE)	3 Elements
Edge Elements	2 Edge Elements
Core Data Center	1 Core Data Center
DDoS Attack Targets	2 Attack Scenarios
DDoS Attack Duration	0.5 s per attack
Number of Slices Evaluated	3 Slices
Number of DDoS Attacks	5 Attacks
Layer-4 Attacks	2 Attacks
Layer-7 Attacks	3 Attacks
Throughput Evaluation	Measured in Mbps
Impact of Attack Strength	Low, Medium, High
Security & QoS Objectives	4 Objectives
Network Slice Isolation	Isolated Slices

(*continued*)

Table 2. (*continued*)

Parameters	Values
Slice Authentication	Enabled
Authorized UE Policy	Enforced
Traffic Filtering Criteria	Strict Filtering
Simulation Duration	100 s
Number of Clients	1000 Clients
Base Station Coverage	400–900 m
Data Rate	10–50 Mbps
Packet Loss Under Attack	5–30%

For this research the features of the NSL-KDD dataset are divided into three main categories: basic features of individual TCP connections (e.g., duration, protocol type, service), content features that inspect the payload content for suspicious activity (e.g., number of failed login attempts, root shell access), and traffic features computed over a two-second time window (e.g., connection counts to the same host or service). To improve the robustness of the detection system and reduce model overfitting, a filtering process is applied where only relevant and high-impact features are selected for classifier training. This selection ensures the machine learning models—Decision Tree (DT) and Support Vector Machine (SVM)—are optimized for accuracy and computational efficiency. While this study does not explicitly describe dimensionality reduction techniques like PCA or feature ranking methods, the implicit focus on meaningful attributes implies the use of a manual or heuristic-based selection informed by domain knowledge and the nature of DDoS behavior in network traffic. The real-time network simulation can be seen in Fig. 2.

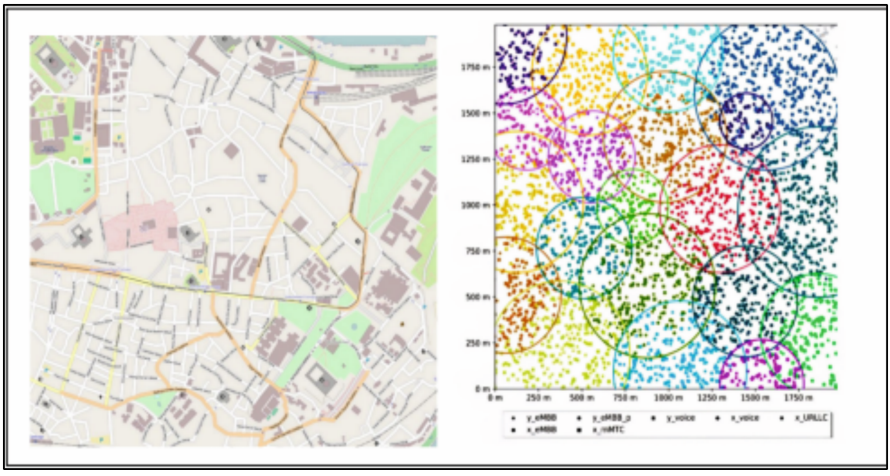


Fig. 2. Real-time network simulation.

3.2 Implementation Process

The algorithm of this research is illustrated in the Algorithm 1.

Algorithm 1: Proposed protocol of Network Slicing Service Protection System (NSSPS)

1. Setup Environment
- Import libraries: simpy, numpy, KDTree, pandas, etc.
- Define simulation area: 2 km× 2 km (Fatih, Istanbul)
- Set simulation time: 3600 s (1 h)
- Exclude 5% at start and end (warm-up/cool-down)

2. Initialize Base Stations
- Load real-world coordinates
- Assign circular coverage areas with variable radii
- Configure slices (eMBB, URLLC) with min/max bandwidth and QoS targets
- Index base stations using KDTree for fast lookup

3. Define Client Types and Movement
- Client categories: automobiles, pedestrians, stationary, trams
- Assign movement models:
- Normal distribution for mobile users
- Fixed/random positions for others
- Assign each client a data rate and target slice

4. Simulation Loop (SimPy environment)
For each client:
- Move according to model at each time step
- Use KDTree to find nearest base station
- Attempt connection based on available bandwidth
- If successful: update stats
- If failed: log as block event
- Log handover if client changes base station

5. Record & Analyze Metrics
- Track: handover rate, block ratio, bandwidth use, success rate
- Forecast congestion and load patterns
- Generate reports: time-series charts, heatmaps, summaries

6. Output Results
- Visualize performance (graphs, logs)
- Recommend improvements in slicing/resource policies

End Algorithm

3.3 Result Analysis

System resources and network slicing can be severely impacted by DDoS attacks. While dynamic assaults necessitate prompt countermeasures to avoid interference, static attacks can be addressed by reconfiguring network slicing at the service creation phase. To strengthen defenses against DDoS attacks, a “traffic-aware” strategy is suggested for 5G

networks. Both dynamic and static attacks are used to test the suggested policy model. Through ongoing development of user equipment, mobile devices are operationally connected to the protection architecture's traffic-aware, security-aware allocation of resources controller on the NM domain. Both the data plane and the control plane of the NM domain are where the NM-SBA software is deployed. One or more attackers initiate DDoS assaults against one or more NM-domain victims. With an emphasis on the infrastructure, virtualization & administration, and service composing & orchestration layers, an NFV-basic architecture is used. Although the intrusion detection system has a high accuracy rate in classifying attacks and non-attacks, more research is required to determine whether the datasets are biased, have inadequate tuning, or are simply constructed. Figure 3 represents the confusion matrix of the intrusion detection model applied to the NSL-KDD dataset.

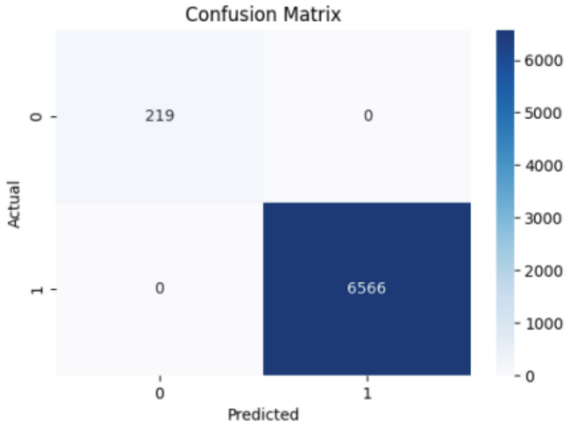


Fig. 3. Confusion matrix of proposed hybrid model NSSPS.

To evaluate the difference detection ability of the intrusion detection model, the Receiver Operating Characteristic (ROC) curve is used, which can be seen in Fig. 4. A Receiver Operating Characteristic (ROC) curve is used to evaluate the intrusion detection model's capability for detecting differences. The ROC curve measures the balance of true positive rate (TPR) and false positive rate (FPR) across different classification criteria. A perfect ROC represents flawless intrusion detection categorization, but it is critical to determine whether this is due to dataset biases or the model's resistance. In real-world circumstances, systems for intrusion detection may face challenges from noise, malicious attacks, and unpredictable attack patterns. The model's correctness in real-world applications must be proven through additional validation with various raw data, the cross-validation methods, and traffic-based scenarios.

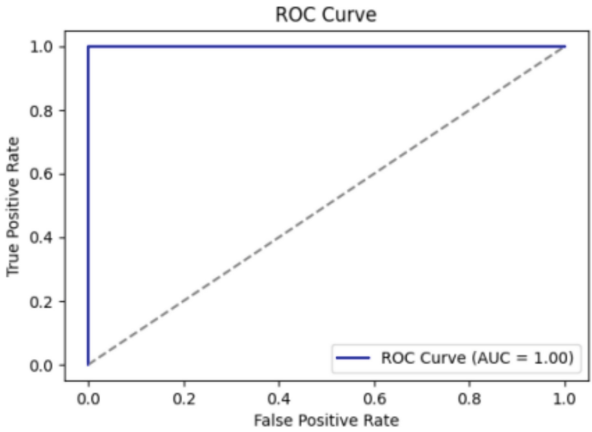


Fig. 4. ROC Curve comparison for DDOS attack samples before and after applying the SVM-DT hybrid model (NSSPS).

Figure 5 demonstrates the impact of 5G network slicing by plotting the relationship between packets transmitted during time along with network latency. Latency, measured in milliseconds, is an important metric for network performance. The graph shows a significant variation in latency, with some intervals exceeding 100 ms and others approaching zero. This fluctuation frequently indicates irregular network conditions, inequalities in traffic prioritization, or unexpected results from DDoS attack simulations. Increased latency figures could indicate more acute contention for resources or network congestion. In some cases, a network might be pushed close to its highest delay threshold, as demonstrated by latency point density clustering. This could point to the necessity for more advanced overload control solutions or intelligent traffic directing regulations.

Because of scheduling mechanisms that prioritize critical traffic, the presence of just a few low-latency locations indicates packets without delay. Network-based attacks can be discovered, prevented, and prevented before they degrade service quality by monitoring emerging delay patterns. The findings suggest that modifying may be necessary to improve performance under various network situations. Figure 6 shows the model's throughput over time prior to and following using the SVM-DT hybrid approach (NSSPS). At the far right, the throughput before SVM-DT indicates a continuous decline, beginning near 1400 packets/sec and progressively decreasing to roughly 600 packets/sec by the 7000th packet, showing performance degradation due to increased load or delayed processing. In comparison, the left graph shows that after applying the NSSPS method, throughput considerably increases and stabilizes, increasing from roughly 4000 to over 7000 packets per second. This demonstrates the NSSPS model's ability to improve processing efficiency, maintain high throughput, and sustain network efficiency.

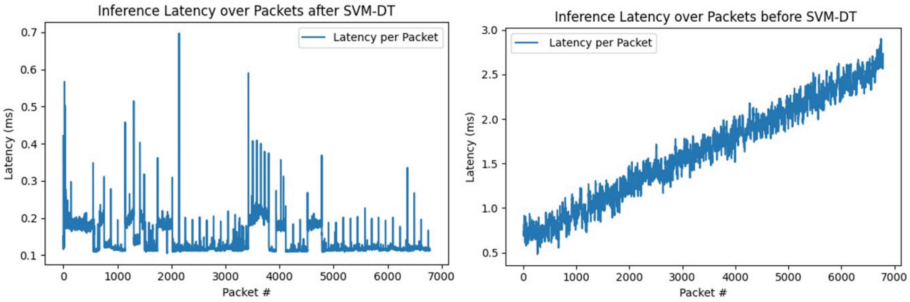


Fig. 5. Latency over Packets Transferred comparison for attack samples before and after applying the SVM-DT hybrid model (NSSPS).

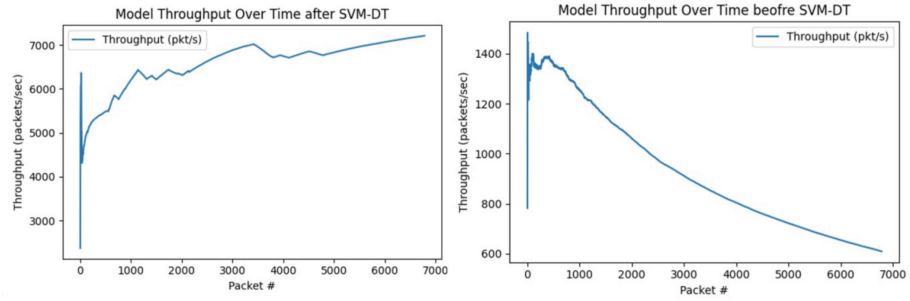


Fig. 6. Throughput Over Time comparison for DDOS attack samples before and after applying the SVM-DT hybrid model (NSSPS).

Figure 7 shows the response time for dealing with security incidents in the 5G network slicing environment. Response time is measured in seconds, and it represents the time span between a security threat's detection and the undertaking of mitigation action. In regard to a network's security, responsive measures should be taken to ensure that critical infrastructure attacks are dealt with before they cause irreversible impairment to service functionality and network performance. The figure also indicates how the response time differs for various security incidents, which in turn provides information on how effective the system's intrusion detection and mitigation capabilities are. The graph indicates the extreme response time, which varies from close to zero seconds to around 10 s. The upper bound of the response time indicates that 10 s may be required to respond to an attack in some situations, which is worrisome for time-critical scenarios like ultra-reliable low-latency communication (URLLC) or critical infrastructure protection. Nevertheless, the lower response times appearing in the dataset demonstrate that other incidents were addressed expeditiously, illustrating how effective the detection system is at mitigating threats.

The distribution of cyberattacks in the context of a 5G network slicing environment is shown in Fig. 8, where they are separated into three groups: unauthorized access, data breaches, and denial-of-service (DoS). DoS attacks are quite common, with more than 60 incidents documented. Even if it is less frequent, unauthorized access is still a

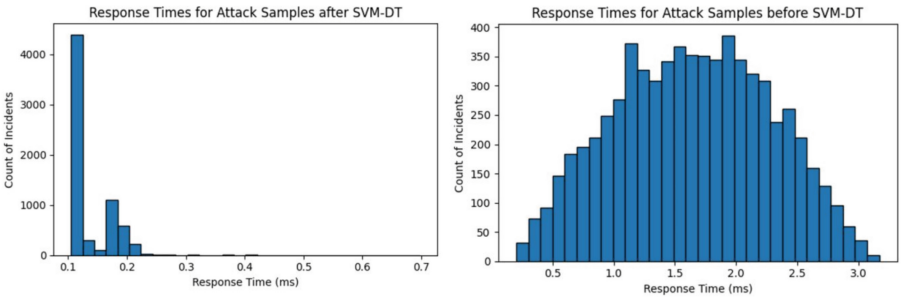


Fig. 7. Response times comparison for DDOS attack samples before and after applying the SVM-DT hybrid model (NSSPS)

serious issue. By bypassing authentication and gaining utilization of restricted network resources, it enables attackers to potentially engage in destructive or exploitative behavior. Role-based access control, multi-factor authentication, and robust authentication techniques are required to reduce this risk.

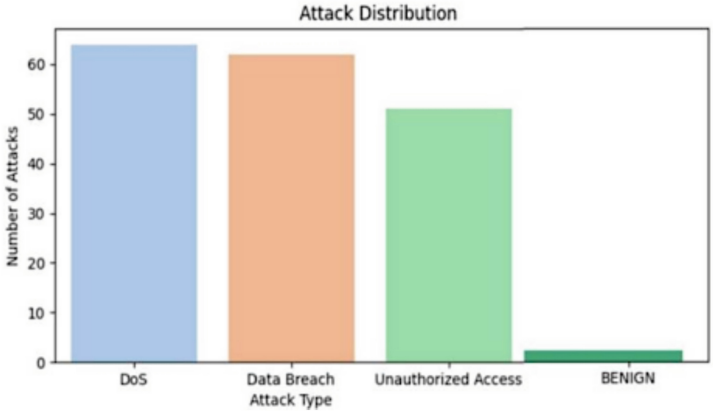


Fig. 8. Attack distribution from the simulated environment.

The quality of service (QoS) of various network slices has been ascertained through an analysis of the 5G network slicing ecosystem. As demonstrated in Fig. 9, Slice 3 provides the greatest performance, coming in slightly under 98%, whilst Slice 1 performs significantly better, at over 95%. This could be due to a more effective security system, traffic management restrictions, or prioritized resource distribution. Slice 2 performs substantially poorer at around 80%, indicating more service failures, less bandwidth, and a greater vulnerability to cyberattacks. The findings emphasize the importance of proactive resource management for network slicing. The classification performance of an SVM-decision-tree hybrid across three network slices reveals that Slice 3 has the best accuracy, at approximately 98%, followed by Slice 1 at approximately 94%. To

improve accuracy in classification, decreased performances on Slice 2 applies slice-specific methods of learning, hyperparameter tuning, or focused feature engineering.

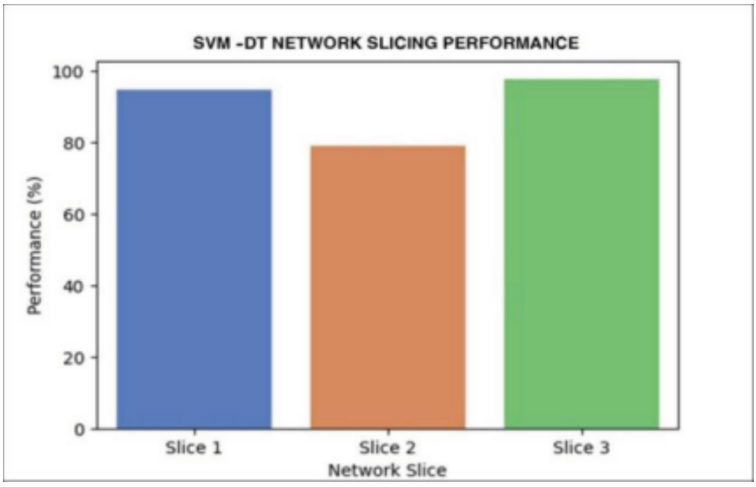


Fig. 9. Network Slicing Performance.

Figure 10 depicts 5G base station coverage area polygons shown as individual circles in a simulated network segmentation environment. The diagram shows the location of base stations, coordinated operations, and the deployment of guidance system infrastructure. Greater transmission ranges are shown by circles with greater coverage regions, which are essential for guaranteeing service level goals throughout slices.

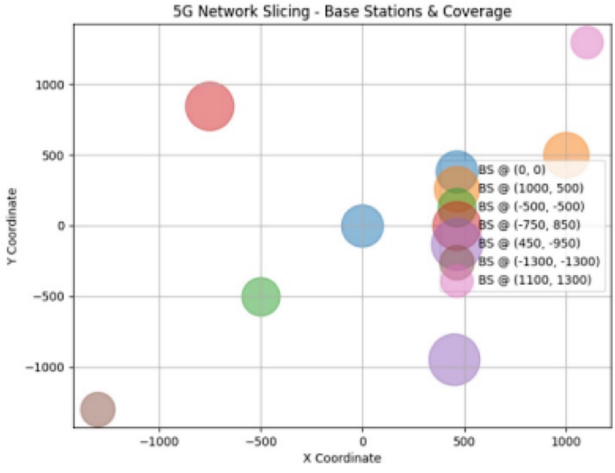


Fig. 10. 5G Network Slicing—Base Stations & Coverage of NSSPS.

4 Discussion and Result Comparison

The evaluation conducted in this part focuses on evaluating the present situation in terms of intrusion detection and network slicing resilience in a 5G simulation. The results demonstrate how well the hybrid model of decision tree and support vector machine classifiers was able to detect and classify network intrusions based on the NSL-KDD dataset, as well as how various cyberattacks affected the performance of network slicing.

The SVM-DT 5G network design for authentication, slices-net, and T-S3RA is demonstrated utilizing TensorFlow, SliceNet, HopFieldNet, and FwBG architectures. The performance is compared to Mininet-WiFi with DDoS attack devices and CBR traffic patterns. T-S3RA detects DDoS attacks and optimizes allocation of resources for authentic subscribers. Network slicing is proposed to handle IoT applications by allowing multiple service-level agreements to coexist in the same backbone. This approach allowsto select slices with high, medium, or low priority that is depicted in Fig. 11.

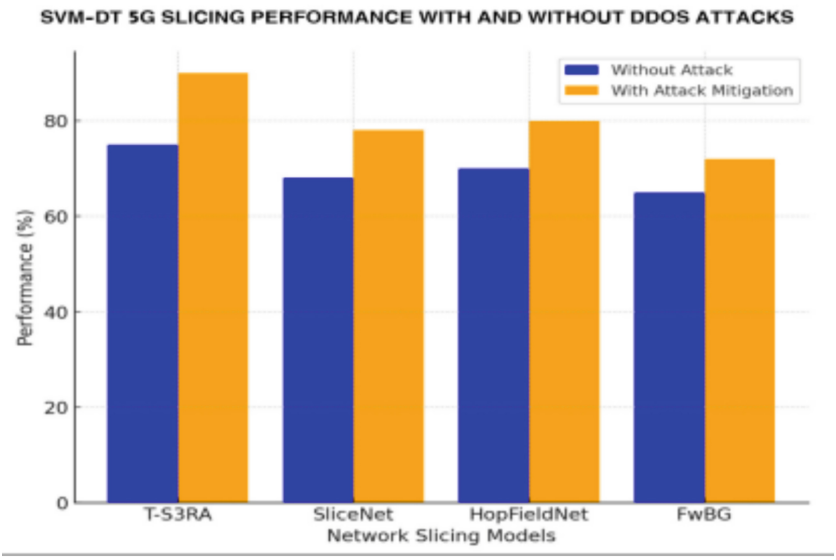


Fig. 11. Performance Comparison of NSSPS (SVM-DT) 5G Network Slicing Models with and without DDoS Attacks.

The performance degradation of each slicing model, such as T-S3RA, SliceNet, HopFieldNet, and FwBG, was recorded both with and without the proposed attack mitigation mechanism (NSSPS model-based SVM-DT). The bar chart illustrates how the integration of the mitigation system improved the overall performance of each model under attack conditions.

The performance of three classification models—Support Vector Machine (SVM), Decision Tree (DT), and a hybrid SVM-DT model—is displayed in the bar chart in Fig. 12. SVM suffers with complex structures and nonlinear network traffic data, despite having an accuracy of 85% and an error rate of 15%. The decision tree suffers from

overfitting but is marginally less reliable at 80% and has a 20% error rate. The SVM-DT hybrid model improves precision and generalization with an accuracy of 92% and a lowest error rate of 8%. This model shows how well machine learning techniques can be integrated to optimize security system performance, making it perfect for 5G networks as intrusion detection systems.

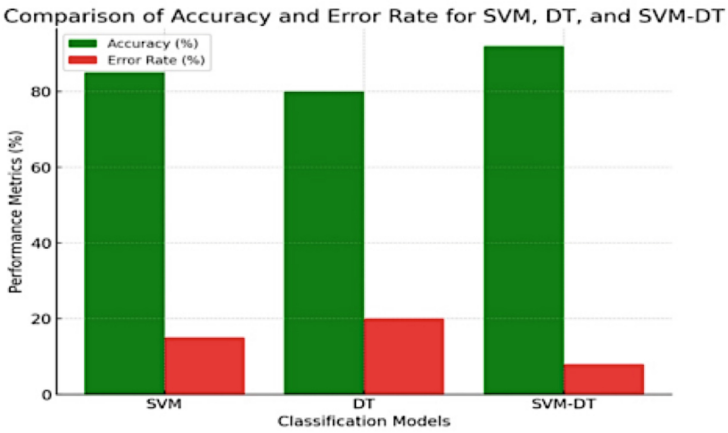


Fig. 12. Comparison of Accuracy and Error Rate for SVM, DT, and SVM-DT (NSSPS).

This study examines the categorization accuracy of CNN, DNN, ANN, U-NET, VGG16, RESNET50, and SVM-DT models. The SVM-DT model outscored all others, with 94% accuracy. RESNET50 obtained the highest accuracy (91%), followed by VGG16 (90%). UCLA and U-NET scored 88% and 86% accuracy, respectively; DNN and ANN had lower accuracy (85% and 83%, respectively). The SVM-DT hybrid model (NSSPS) integrates SVM decision boundaries with hierarchical rule-based classification, which makes it flexible and resource-efficient for real-time monitoring. This comparison can be found in Fig. 13.

Table 3 reveals a bar chart that compares the accuracy metrics from our proposed NSSPS (based SVM-DT) model with several other recent works from the 5G network slicing and security domains. The comparison encompasses those with Deep Learning, Long Short-Term Memory (LSTM), Reinforcement Learning, Machine Learning-based Admission Control, and Network Slicing Management Models. The findings from our evaluation show that although many existing methods provide competent results, the SVM-DT model surpasses all others in accuracy, achieving the highest at 94%. The research [14] that used Deep Learning was able to achieve 89% accuracy, which demonstrates that a deep learning approach is applicable for complex network slicing scenarios. It is worth noting, however, that deep learning methods are resource-intensive and require large volumes of training data, which renders them less suitable for real-time security implementations. Additionally, [6] suggests the intelligence in network resource allocation and security in 5G systems managed by AI support the security enhancements by using Intelligent Slicing, and obtaining 88% accuracy. Abubakr et al.'s [8] applied LSTM Slicing Autonomous Control Model and achieved 87% accuracy, which illustrates its

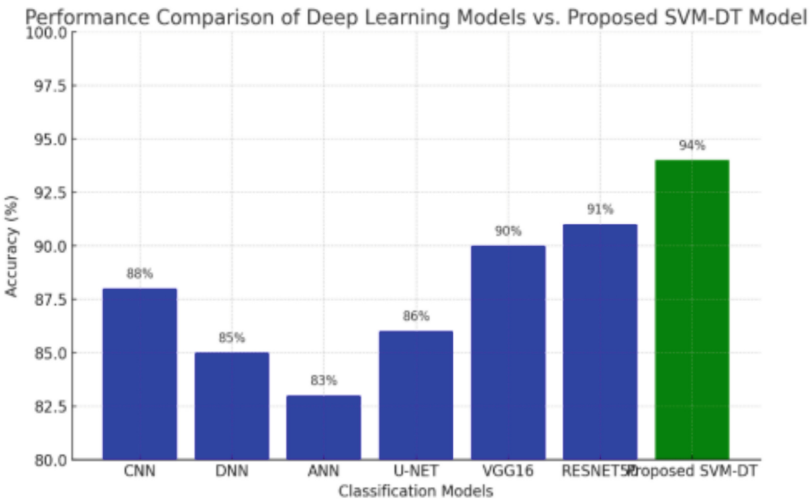


Fig. 13. Performance Comparison of Deep Learning Models vs. Proposed SVM-DT (NSSPS) Model.

effectiveness for predicting network behavior and optimizing network slicing strategies. LSTMs, however, require large datasets and suffer from vanishing gradients, negatively influencing classification performance in rapidly changing environments. The Slicing Management Model Cardenas et al. It is noted that the works of Ssengonzi et al. [15], utilizing Deep Reinforcement Learning (DRL), through Cost Efficient Network Slicing attained 86% accuracy. This indicates that AI-enabled network management does enhance security but lacks real-time flexibility. Also, Vincenzi et al. [18]—ML Admission Control—earned 84%, suggesting that while machine learning-based admission controls enhance resource allocation, they do not address security concerns posed by network slicing. Security gaps in network slicing are not fully addressed with the 84% prediction accuracy attained. Our proposed model (NSSPS based SVM-DT) outstrips all SVM-based works with an astonishing accuracy of 94% achieved using the hybrid SVM-DT model. Overt achievement is due to the constituent parts of the model. The SVM part captures the non-linear relations bound to network features, and the DT part improves the explanations as well as decisions made. Together, they strengthen the classification model, which can distinguish normal and attack traffic with better certainty. The findings highlight hybrid machine learning approaches' advantages over both reinforcement learning and deep learning methods for model development. The SVM-DT model requires less computing while achieving greater accuracy than deep learning techniques, which require big datasets, significant processing capacity, and rigorous tuning of several model parameters. These findings suggest that hybrid models have the potential to enhance 5G network security by striking an optimal combination of precision, effectiveness, and responsiveness.

The proposed SVM-DT model (NSSPS) offers the highest accuracy of 94%, beating other strategies in the safeguarding and slicing sectors of the 5G networks. This is due

to the model's hybrid framework, which improves interpretations and decision-making by reflecting non-linear connections between network elements.

Table 3. Comparative Summary of Methods, Datasets, and Accuracy Results

Authors	Method	Dataset	Accuracy (%)
[14]	Deep Learning	NSL-KDD	89
[8]	LSTM	CIC-IDS2017	87
[6]	Intelligent Slicing	CIC-DDoS2019	88
[15]	Deep Reinforcement Learning (DRL)	Bot-IoT	86
[18]	ML Admission	TON_IoT	84
Proposed NSSPS Model SVM-DT based	**SVM + Decision Tree**	**NSL-KDD**	**94**

5 Conclusion

This study proposed a hybrid SVM-DT (NSSPS model) for intrusion detection specifically tailored for 5G network slicing conditions. The model surpassed its reinforcement learning and deep learning counterparts, achieving 94% accuracy on the NSL-KDD dataset. The model achieves high precision while consuming less processing power by combining the intuitive nature of DT in addition to the strong boundaries optimization of SVM. The study showed how hybrid AI may improve QoS, QoE, and service resilience across network slices while effectively identifying and mitigating attacks such as DoS and data exfiltration. It also stressed how important dynamic, slice-aware safety regulations are for mitigating the effect of cyberattacks on critical 5G services. A comparison analysis shows that the proposed SVM-DT model is a fair, effective, and understandable strategy for real-time 5G network defense. For complex, multi-vector detection of attacks, future research should add deep learning-driven data mining into the SVM-DT framework. Blockchain-based systems and learning federation can enhance safeguarding event logs and build trust. Reactiveness to threats can be enhanced by multi-agent RL techniques, reinforcement learning, interpretability, and adversarial robustness. Validating model performance requires a live 5G/B5G testbed deployment.

References

1. Suh, K., Kim, S., Ahn, Y., Kim, S., Ju, H., Shim, B.: Deep reinforcement learning-based network slicing for beyond 5G. IEEE Access. **10**, 7384–7395 (2022)
2. Singh, S., Babu, C.R., Ramana, K., Ra, I.H., Yoon, B.: BENS−B5G: blockchain-enabled network slicing in 5G and beyond-5G (B5G) networks. Sensors. **22**(16), 6068 (2022)
3. Botez, R., Costa-Requena, J., Ivanciu, I.A., Strautiu, V., Dobrota, V.: SDN-based network slicing mechanism for a scalable 4G/5G core network: a kubernetes approach. Sensors. **21**(11), 3773 (2021)

4. Yeh, S.P., Bhattacharya, S., Sharma, R., Moustafa, H.: Deep learning for intelligent and automated network slicing in 5G open RAN (ORAN) deployment. IEEE Open J. Commun. Soc. **5**, 64–70 (2023)
5. Salahdine, F., Liu, Q., Han, T.: Towards secure and intelligent network slicing for 5G networks. IEEE Open J. Comput. Soc. **3**, 23–38 (2022)
6. Otieno, H.O., Malila, B., Mwangama, J.: Deployment and management of intelligent end-to-end network slicing in 5G and beyond 5G networks-a systematic review. IEEE Access. **12**, 190411–190433 (2024)
7. Hu, Y., Gong, L., Li, X., Li, H., Zhang, R., Gu, R.: A carrying method for 5G network slicing in smart grid communication services based on neural network. Future Internet. **15**(7), 247 (2023)
8. Abubakr, T., Nasr, O.A.: Novel LSTM-based approaches for enhancing outdoor localization accuracy in 4G networks. IEEE Access. **11**, 140103–140115 (2023)
9. Ferreira, J.P., Ferreira, V.C., Nogueira, S.L., Faria, J.M., Afonso, J.A.: A flexible infrastructure-sharing 5G network architecture based on network slicing and roaming. Information. **15**(4), 213 (2024)
10. Kourtis, M.A., Sarlas, T., Xilouris, G., Batistatos, M.C., Zarakovitis, C.C., Chochliouros, I.P., Koumaras, H.: Conceptual evaluation of a 5G network slicing technique for emergency communications and preliminary estimate of energy trade-off. Energies. **14**(21), 6876 (2021)
11. Dangi, R., Jadhav, A., Choudhary, G., Dragoni, N., Mishra, M.K., Lalwani, P.: Ml-based 5G network slicing security: a comprehensive survey. Future Internet. **14**(4), 116 (2022)
12. Boukessessa, M., Ghazli, A., Ali-Pacha, A.: Blockchain-based security framework for VNF package protection in 5G network slicing services. Transport. **25**(3), 289–299 (2024)
13. Klinkowski, M.: Analysis of latency-aware network slicing in 5G packet xHaul networks. Int. J. Electron. Telecommun., 335–340 (2023)
14. Abood, M.S., Wang, H., Virdee, B.S., He, D., Fathy, M., Yusuf, A.A., Ahmad, A.: Improved 5G network slicing for enhanced QoS against attack in SDN environment using deep learning. IET Commun. **18**(13), 759–777 (2024)
15. Ssengonzi, C., Kogeda, O.P., Olwal, T.O.: A survey of deep reinforcement learning application in 5G and beyond network slicing and virtualization. Array. **14**, 1–27 (2022)
16. Thantharate, A., Tondwalkar, A.V., Beard, C., Kwasinski, A.: ECO6G: energy and cost analysis for network slicing deployment in beyond 5G networks. Sensors. **22**(22), 8614 (2022)
17. Cup, K.D.D.: Available on: http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html (2007)
18. Vincenzi, M., Lopez-Aguilera, E., Garcia-Villegas, E.: Timely admission control for network slicing in 5G with machine learning. IEEE Access. **9**, 127595–112761 (2021)



Enhancing Crude Oil Price Prediction with Neural Network Models

Hitesh Punjabi[1], S. Kumar Chandar[2(✉)], and Mayur Malik[3]

[1] K J Somaiya Institute of Management, Mumbai, India
hitesh.punjabi@somaiya.edu
[2] School of Business and Management, Christ University, Bengaluru, India
kumar.chandar@christuniversity.in
[3] International Open University, Banjul, The Gambia

Abstract. A nation's economic, social, and national security are all severely affected by variations in crude oil prices, which is a basic energy source. Research on accurately forecasting price changes for crude oil is always progressing. This research presents a forecasting strategy for crude oil pricing using artificial neural networks. The presented model uses standardization techniques to prepare the historical data for the subsequent processes. It is possible to predict future prices by using a Feed Forward Neural Network (FFNN) with four layers. West Texas Intermediate (WTI) and Brent crude oil prices are utilized on a daily, weekly, and monthly basis to demonstration and confirmation. Directional statistic, accuracy of prediction, the model is evaluated using root mean square error and mean absolute error expressed as percentages. Empirical findings confirm that the suggested approach performs better than any of the previous approaches. Additionally, it is noted that the presented method achieved higher prediction in contrast to other methods.

Keywords: Artificial neural network · crude oil price · prediction · standardization

1 Introduction

Crude oil affects the global economy. It is a petroleum based substance made up of organic materials and natural hydrocarbon reserves. It is considered to be the most significant fuel in the globe and accounts for about one-third of all energy used worldwide. A social stability and economic growth country are significantly impacted by variations in the crude oil price. The economy of a country can be affected by two distinct approaches to the pricing of crude oil. The financial growth is negatively impacted by a significant increase in crude oil prices, which also drives up inflation. Contrarily, a decrease in crude oil price may result in severe financial difficulties for countries that export oil. Therefore, predicting changes in the price of crude oil with accuracy is a popular study issue.

Crude oil prices are also influenced by several factors exchange rate, supply, demand, climate and geopolitics etc. and they have fluctuated drastically in the past years. For

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026
S. Lanka et al. (Eds.): ICTSM 2025, LNNS 1755, pp. 30–44, 2026.
https://doi.org/10.1007/978-3-032-13177-5_3

instance, the prices of Brent oil reached 113.34 (USD/barrel) in May 2022 and then during the next seven months, reduced to 80.92 (USD/barrel) [1]. In recent decades, there has been a lot of volatility and non-stationarity related to crude oil prices. Crude oil's complex nature makes it challenging to estimate its price with any degree of accuracy. Much research has been done on the developing of several forecasting models for crude oil prices. Crude oil forecasting models can be categorized into four major types: time series, econometric, qualitative and Machine Learning models [2, 6, 7].

Exponential Smoothing (ES) and Autoregressive Integrated Moving Average (ARIMA) are commonly utilized time series model [3, 4]. The econometric and qualitative methods like vector auto regression model, Generalized Autoregressive Conditional Heteroskedastic (GARCH), threshold, and state space models are employed [5, 8]. The expected data's linearity and stationarity are assumed in the equations of the econometric and time series models. The accuracy of predicting is, however, limited by the nonlinear, dynamic, and nonstationary the characteristics of crude oil pricing, as research studies have shown. Many academics recommend using Machine Learning (ML) methods to forecast prices for crude oil to address the complicated character of these prices. The ML models have showed improved performance compared to traditional models [8, 9]. The ability of ML models to handle complicated, non-stationary, and non-linear data structures allows them to provide better results. This article designs a ML-based artificial neural network model for predicting the price of crude oil.

The contribution of this paper lies in developing a reliable model for predicting crude oil prices using neural network. Unlike existing studies that focus on statistical analysis, this work uses neural network. The novelty of the proposed model resides in the design and application of feed forward network to capture nonlinear patterns inherent in crude oil prices. The following are this paper's main contributions:

- A succinct review of recent papers predictions has been made on the price of crude oil.
- Statistical characteristics of selected firms are examined and the required preparation of the data technique is applied
- For the purpose of forecasting the price of crude oil in the future, a three-layer Feed Forward Neural Network (FFNN) model has been developed.

The following is the structure of the paper: Sect. 2 presents a summary of earlier techniques. The dataset and the model for predicting the price of crude oil are explained in Sect. 3. The simulation's results are examined in Sect. 4. Results are presented in Sect. 5 along with recommendations for further study.

2 Review of Earlier Methods

There has been considerable progress in crude oil price forecasting in recent years. Researchers have produced encouraging results using various analysis [2, 6, 8, 9] [11]. Recent developments in ML models for this section's primary focus is on forecasting the price of crude oil. Yu et al. [2] reconstructed and decomposed information to forecast crude oil prices. To forecast the price of crude oil, data was rebuilt and deconstructed. Fan et al. [10] predicted pricing for crude oil by applying Independent Component Analysis-Support Vector Regression (ICA-SVR). Their method provided better performance than

econometric models Chen et al. [11] combined ARMA and Deep Belief Network (DBN) to increase the precision of crude oil price forecasts. Zhao et al. [12] produced a crude oil forecasting ensemble model. Stacked Denoising Auto Encoders (SDAE) and Bagging are used to forecast price and grow an ensemble respectively. Performance of the model was validated on WTI data. Sun et al. [13] utilized Interval Decomposition Ensemble (IDE) approach to forecast WTI and Brent oil prices. Yu et al. [14] examined SVR's potential for daily predictions for the price of crude oil. Cen et al. [15] presented a Long Short Term Memory (LSTM) to forecast volatility in Brent and WTI data. They showed promising results. However, deep learning based approach requires more time for training.

A hybrid model by fusing to predict the price of crude oil, LSTM and Bi-LSTM were introduced by Bristone et al. [16]. However, such a model is computationally intensive and slow. Zhao et al. [17] developed a model for predicting crude oil by integrating ARMA, Support Vector Machine (SVM), and Complementary Ensemble Empirical Mode Decomposition with Additive Noise (CEEMDAN). The combination of the above methods was shown to be robust to non-linear data. Yang et al. [18] reached similar results when applying K means algorithm, kernel PCA and KELM to predict WTI and Brent crude oil prices. Daneshvar et al. [19] used LSTM to anticipate the price of Brent crude oil. Results show that LSTM mode with SGDM solver has higher prediction accuracy. To estimate the price of crude oil, a CEEMD-ARIMA-SVM model was presented by Cheng et al. [20]. The model performs well when compared with SVM and ARIMA in terms of accuracy Wang and Fang [21] designed Weighted Feed Forward Neural Network (WT-FNN) to forecast the price of Brent oil. Prediction accuracy was outstanding for the WT-FNN model. An overview of the papers that have been assessed is given in Table 1.

Table 1. A summary of recent methods

Researchers	Model	Data	Data frequency	Metrics
Yu et al. [2]	Divide and conquer	WTI and Brent	Monthly	–
Fan et al. [10]	ICA-SVR	WTI and Brent	Weekly	RMSE MAPE
Chen et al. [11]	ARMA-DBN	WTI	Daily	MSE
Zhao et al. [12]	Ensemble	WTI	Daily	RMSE, MAPE
Sun et al. [13]	IDE	WTI	Daily	RMSE
Yu et al. [14]	SVM	WTI	Daily	MSE
Cen et al. [15]	LSTM	WTI and Brent	Daily	RMSE, MAPE
Bristone et al. [16]	Hybrid	WTI and Brent	Daily	RMSE
Daneshvar et al. [19]	LSTM	Brent	Weekly	RMSE
Cheng et al. [20]	CEEMD-ARIMA-SVM	WTI	Daily	MSE, MAPE

3 Material and Method

Artificial Neural Network (ANN) is one of the most often used models for modelling non-linear data. ANN is a subclass of ML formed by a large number of neurons and reflects the function of human brain. It is a booming technology, employed in stock price prediction, weather prediction, and commodity prediction, and has attained better results [20]. This study uses FFNN to develop a model for anticipating crude oil prices. As shown in Fig. 1, the general procedures of the suggested paradigm. The model has many processes: data collection, preprocessing, data division, design and implementation of model for studying and predicting the price of crude oil.

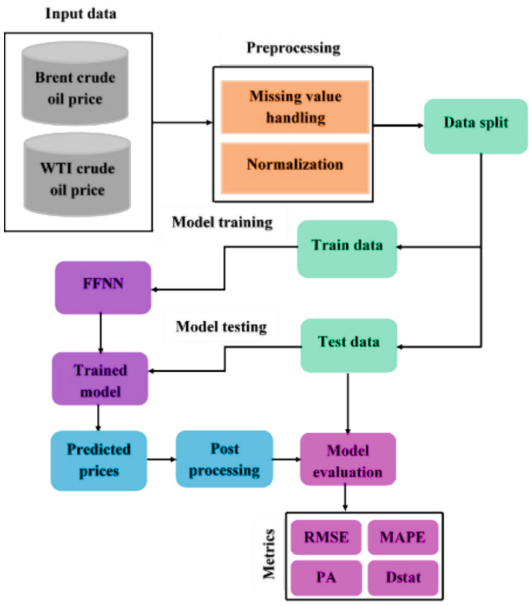


Fig. 1. Pipeline of the proposed crude oil prediction method

3.1 Data Collection

To analyze resilience of the created model, crude oil data sets of two benchmark prices are considered such as WTI and Brent prices. The U.S. The information is from the Energy Information Administration (EIA) [22]. The Brent and WTI statistics span the period from January 1, 2001, to February 28, 2023. Despite the exception of weekends and public holidays, crude oil prices are available daily, weekly, and monthly. Figure 2 illustrates daily, weekly, and monthly WTI and Brent crude oil prices. Table 2 lists the descriptive statistics for Brent and WTI data. As in Table 2, the maximum and minimum values have a large difference. The standard deviation (Std. dev.) of both data shows the prices fluctuates dramatically.

Table 2. Descriptive statistics of WTI and Brent crude oil prices

Parameters	WTI crude oil price			Brent crude oil price		
	D	W	M	D	W	M
No. of samples	5629	1156	266	5629	1156	266
Minimum	8.91	16.55	3.32	9.12	14.24	18.38
Maximum	145.3	133.8	142.5	143.9	141.1	132.7
Mean	64.14	63.61	63.95	66.97	67.06	66.68
Std.dev.	25.71	25.74	25.71	29.13	29.22	29.23
Variance	0.035	0.034	0.048	0.046	0.053	0.065
Skewness	0.280	0.277	0.269	0.297	0.288	0.282

D daily, *W* weekly, *M* monthly

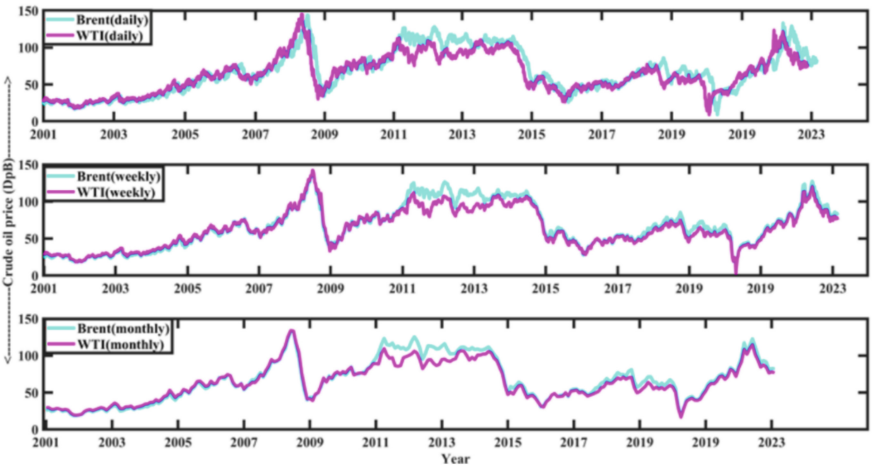


Fig. 2. Crude oil prices from 2001 to 2023

3.2 Preprocessing

In this study, preprocessing is done in two ways: (i) missing value handle and (ii) normalization. The surrounding data mean addresses missing data. Min-max normalizes data. Data normalization can be defined as,

$$y_{norm} = \frac{y - y_{min}}{y_{max} - y_{min}} \tag{1}$$

where the real price, y, is and y_{norm} is the mean price., while min and max represent the lowest and greatest costs, respectively. Normalized prices are shown in Fig. 3.

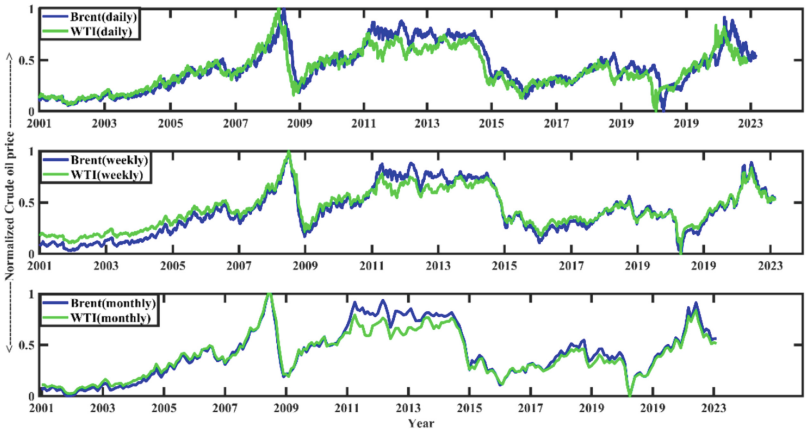


Fig. 3. Normalized crude oil prices

3.3 Data Division

After preprocessing, both datasets are subjected to data division. For training and testing, different ratios are used to separate each data set into training and testing, including 60:40, 70:30, 80:20, and 90:10. In 60:40 ratios, training uses 60% of the data while testing uses the remaining 40%. The generated model's efficacy is assessed using a range Including test and training collections, and the top performing model is then selected to produce predictions with a better degree of accuracy.

3.4 Feed Forward Neural Network

An FFNN typically has three layers: data introduction (input layer), data processing (hidden layer), and output generation (output layer). FFNN's general structure is seen in Fig. 4. The selection of the FFNN as the prediction model is justified by its ability to model nonlinear data, which is characteristic of crude oil prices [26–28].

The relationship between input and output layers is represented using Eq. (2),

$$y_t = f\left(b_0 + \sum_{j=1}^{r} w_j g\left(w_{0j} + \sum_{i=1}^{p} w_{ij} y_{t-i}\right)\right) \tag{2}$$

Where, b is the bias, wj (j = 1,2,3,…,r) symbolizes the proportion of weight between the hidden and input layers, while wij (i = 1,2,3,…,p) symbolizes the ratio of the output and hidden layers. The activation function is found in g(.).

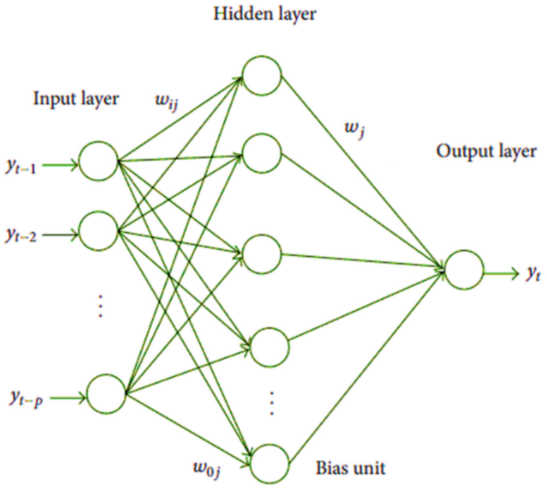


Fig. 4. Architecture of FFNN

3.5 Training of FFNN

In this investigation, a FFNN is designed with four layers: input layer, ten hidden neurons form the two hidden layers, together with the output layer. The hidden layer uses a sigmoidal activation function, given by

$$g(x) = \frac{1}{1 + e^{-x}} \tag{3}$$

The output layer's activation function is pure linear and is shown as,

$$f(x) = x \tag{4}$$

The network is trained with Leven-Berg Marquart algorithm.

4 Numerical Results

4.1 Evaluation Criteria

The developed model has been implemented using MATLAB 2022a. In this study, the suggested method's a set of metrics is used to assess prediction performance. The chosen performance measuring metrics include Root Mean Square Error (RMSE), Mean Absolute Percent Error (MAPE), Directional statistic (Dstat), and Prediction Accuracy (PA) are expressed as follows:

$$RMSE = \sqrt{\frac{1}{M}\sum_{j=1}^{M}\left(A_j - P_j\right)^2} \tag{5}$$

$$MAPE = \frac{1}{M}\sum_{j=1}^{M}\left|\frac{A_j - P_j}{A_j}\right| \tag{6}$$

$$Dstat = \frac{1}{M}\sum_{j=1}^{M} d_j \tag{7}$$

$$d_j = \begin{cases} 1, & \left(A_{j+1} - A_j\right)\left(P_{j+1} - P_j\right) \geq 0 \\ 0, & otherwise \end{cases}$$

$$\text{PA} = \frac{\text{Number of correctly predicted value}}{\text{Total number of samples}}\text{X100} \tag{8}$$

Where, the actual and prediction prices are denoted by A and P, respectively, whereas M is the number of samples. The better prediction model is shown by decreased RMSE and MAPE values. Crude oil's directional tendency is measured using Dstat price fluctuation whether they are downward or upward. The value of Dstat is close to 1, increased precision in the model's direction prediction and vice versa.

4.2 Performance Analysis

Initially, data were collected from the publicly available sources [22]. The data was normalized by using min-max method. Figure 4 displays the normalized WTI and Brent crude oil prices. Training and testing samples were developed using the normalized data. Training sampled are used for training the model whereas test data are used for validation. Efficacy of the model is validated by varying training and testing samples.

Brent Crude Oil Price Prediction

For experimentation purpose, Brent crude oil prices from 1 January 2001 to 28 February 2023 are utilized. Three sets of data are chosen, i.e. daily prices, weekly prices, and monthly prices. From these prices, the data is randomly spilt into training and testing samples at several ratios, 60:40, 70:30, 80:20, and 90:10, to check how well the built model can be generalized. The efficacy of the created model in predicting the pricing of Brent crude oil on a daily, weekly, and monthly basis for different testing samples is shown in Table 3. The values in boldface are the mean RMSE, MAPE, Dstat, and PA. Note that the designed model performs better prediction for daily prices than for weekly and monthly prices. The mean RMSE value for daily price is 0.018, weekly price is 0.038, and monthly price is 0.084. Similarly, the average MAPE values of weekly (0.080), and monthly (0.153) are higher than the average MAPE value of Daily (0.037). The mean prediction accuracy for daily, weekly, and monthly prices is 98.77%, 97.30%, and 93.75% respectively. These findings prove that the daily observation tell us the predicted results of the developed model is more accurate and reasonable that the weekly and monthly observations. Figure 5 shows the developed model's average results, daily, weekly, and monthly prices. In Figs. 6, 7, and 8, the exact Brent prices vs. the expected prices for the daily, weekly, and monthly predictions, respectively.

Table 3. Errors and prediction accuracy of Brent crude oil prices

Data frequency	Testing period	RMSE	MAPE	Dstat	PA(%)
Daily	23/04/2014–28/02/2023	0.014	0.023	0.985	99.10
	08/07/2016–28/02/2023	0.014	0.024	0.989	99.05
	20/09/2018–28/02/2023	0.016	0.028	0.992	98.90
	07/12/2020–28/02/2023	0.029	0.074	0.993	98.03
	Average	**0.018**	**0.037**	**0.990**	**98.77**
Weekly	18/04/2014–28/02/2023	0.028	0.055	0.998	97.99
	01/07/2016–28/02/2023	0.031	0.056	0.997	97.94
	21/09/2018–28/02/2023	0.034	0.064	0.996	97.62
	04/12/2020–28/02/2023	0.061	0.148	0.991	95.67
	Average	**0.038**	**0.081**	**0.996**	**97.31**
Monthly	04/2014–02/2023	0.068	0.137	0.991	94.91
	06/2016–02/2023	0.068	0.125	0.988	94.97
	09/2018–02/2023	0.085	0.154	0.981	93.85
	11/2020–02/2023	0.119	0.198	0.963	91.27
	Average	**0.085**	**0.153**	**0.981**	**93.75**

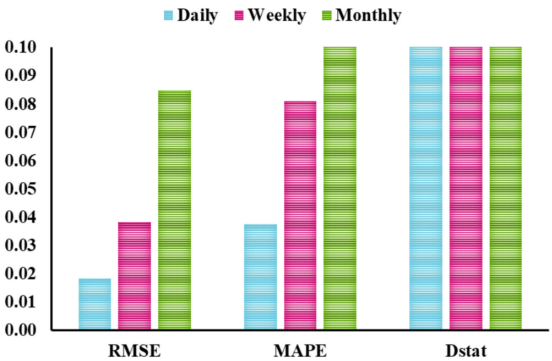


Fig. 5. Prediction performance of Brent Crude oil prices



Fig. 6. Actual Vs. Predicted values of Brent Crude oil daily prices

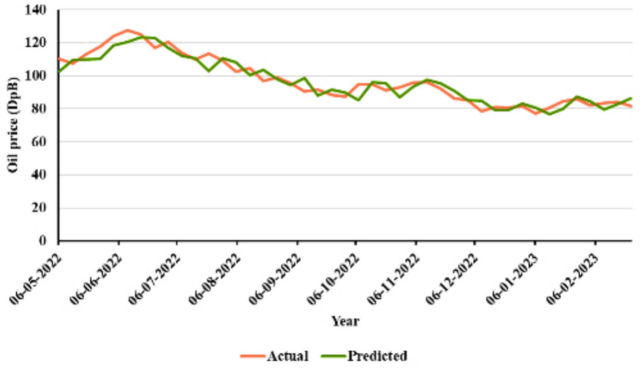


Fig. 7. Actual Vs. Predicted values of Brent Crude oil weekly prices

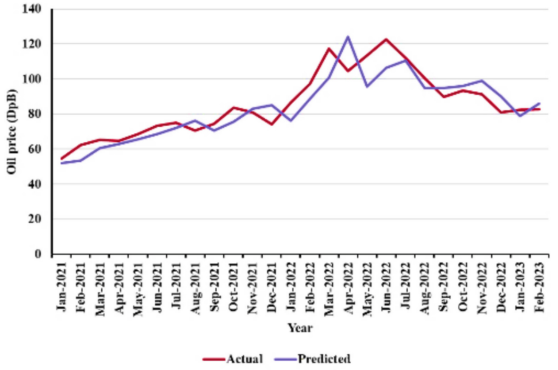


Fig. 8. Actual Vs. Predicted values of Brent Crude oil monthly prices

WTI Crude Oil Price Prediction

For this case, WTI oil prices from 1 January 2001 to 28 February 2023 are considered. Three sets of data are chosen, i.e. daily prices, weekly prices, and monthly prices. From these prices, the data is randomly spilt into training and testing samples at several ratios, 60:40, 70:30, 80:20, and 90:10, to confirm that the established model is effective. The price of WTI crude oil on a daily, weekly, and monthly basis prediction performance for the developed model for various testing samples is summarized in Table 4. The average values of RMSE, MAPE, Dstat, and PA are highlighted in boldface. It is observed from the Table 4 that the designed model achieved lower prediction error and higher prediction accuracy for daily price prediction compared to weekly and monthly prices prediction The mean RMSE values for the corresponding predictions for the daily, weekly, and monthly periods are 0.018, 0.038, and 0.085. The mean MAPE value for daily price is 0.029, weekly price is 0.062, and monthly price is 0.148. The better model is indicated by a Dstat value that is closer to 1. The designed model attained Dstat value of daily price (0.993), weekly price (0.996), and monthly price (0.981). The mean prediction accuracy for daily, weekly, and monthly prices is 98.9%, 97.45%, and 93.78% respectively. Results from experiments indicate that a developed model can precisely predict future price based on daily observation data. Furthermore, it is also noted that the designed model achieved higher prediction accuracy at ratio of 70:30. The mean efficiency of the constructed model is shown in Fig. 9 for daily, weekly, and monthly prices. Figures 10, 11, and 12 show the actual WTI prices vs. the expected prices for the daily, weekly, and monthly projections, respectively.

Table 4. Errors and prediction accuracy of WTI oil prices

Data frequency	Testing period	RMSE	MAPE	Dstat	PA(%)
Daily	23/04/2014–28/02/2023	0.016	0.027	0.993	98.99
	08/07/2016–28/02/2023	0.017	0.026	0.992	98.98
	20/09/2018–28/02/2023	0.019	0.032	0.994	98.82
	07/12/2020–28/02/2023	0.019	0.032	0.994	98.82
	Average	**0.018**	**0.029**	**0.993**	**98.90**
Weekly	18/04/2014–28/02/2023	0.027	0.042	0.998	98.17
	01/07/2016–28/02/2023	0.032	0.045	0.997	98.06
	21/09/2018–28/02/2023	0.034	0.049	0.996	97.77
	04/12/2020–28/02/2023	0.061	0.112	0.991	95.81
	Average	**0.038**	**0.062**	**0.996**	**97.45**
Monthly	04/2014–02/2023	0.106	0.163	0.991	92.87
	06/2016–02/2023	0.074	0.123	0.988	94.79
	09/2018–02/2023	0.068	0.141	0.981	94.87
	11/2020–02/2023	0.016	0.027	0.993	98.99
	Average	0.017	0.026	0.992	98.98

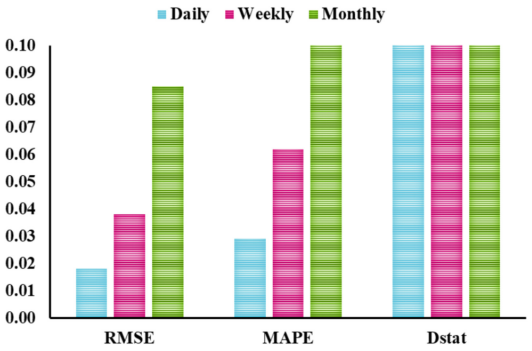


Fig. 9. Prediction performance of WTI crude oil prices

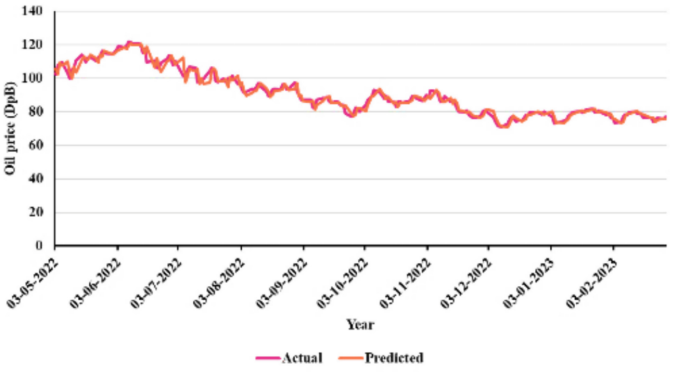


Fig. 10. Actual Vs. Predicted values of WTI crude oil daily prices

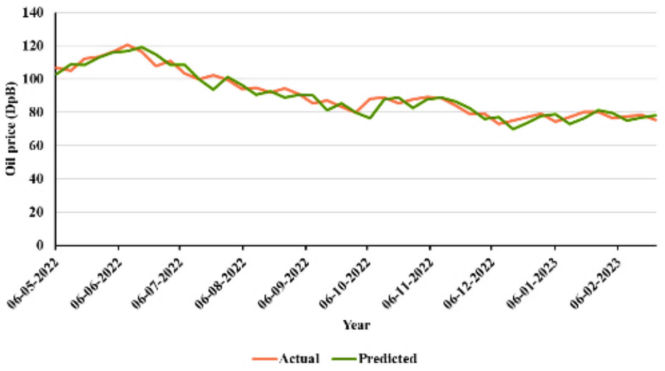


Fig. 11. Actual Vs. Predicted values of WTI Crude oil weekly prices

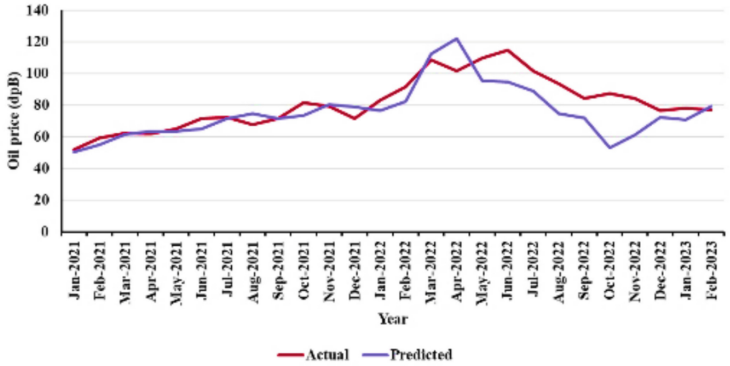


Fig. 12. Actual Vs. Predicted values of WTI Crude oil monthly prices

4.3 Comparison with Earlier Methods

Table 5. Performance comparison with earlier methods

Data	Model	Daily	Weekly	Monthly
Brent crude oil	Regression tree	0.029	0.047	0.093
	Random forest	0.025	0.072	0.099
	Cat boost	0.023	0.051	0.097
	Proposed	0.018	0.038	0.086
WTI crude oil	Regression tree	0.032	0.056	0.093
	Random forest	0.028	0.066	0.091
	Cat boost	0.027	0.060	0.091
	Proposed	0.018	0.038	0.085

A comparison is made between the generated model and the previous models' prediction performance, regression tree [23], random forest [24], and cat boost [25] in terms of mean RMSE for daily, weekly, and monthly costs, as given in Table 5. For both the data sets, the proposed model excelled for daily, weekly, and monthly prediction. For instance, Brent crude oil price, the RMSE value of regression tree is 0.029, random forest is 0.025, cat boost if 0.023, and designed model is 0.018. The developed model had a similar performance for weekly and monthly prices. For WTI data, the RMSE value of regression tree, random forest, and cat boost is 0.032, 0.028, and 0.02. The designed model attained RMSE value is 0.018 which is lower than the compared models, confirming the potential of the developed model. For weekly and monthly price predictions, similar results are shown.

5 Conclusion and Recommendations

This paper presented a neural network model for forecasting the price of crude oil. First, the historical crude oil prices are pre-processed using data normalization to put them into a common range of 0 to 1. Then, a four-layer crude oil prices are predicted using feed forward neural networks. Test data evaluates the model's performance, while training data trains it. Model efficacy is examined using pricing for Brent and WTI crude oil. The mean daily, weekly, and monthly prices for crude oil basis are used to evaluate the model's performance. From experimental outcomes, the following conclusions are drawn: (i) Artificial neural network model can be utilized to predict crude oil prices. (ii) in contrast to conventional techniques, the model that was developed performs better, showing its the ability in managing non-linear data and (iii) the produced model's prediction performance is outstanding irrespective of data frequency, showing its reliability and robustness. The model not only provides evidence of the effectiveness of FFNN in crude oil price forecasting but also lays the groundwork for future studies. This study also highlights some problems that need further investigation. It has been shown that the developed model accurately forecasts the price of crude oil. It would be exciting to test how effective it is in other markets like stock prices, gold prices, and foreign exchange rated. In future, other neural network will be explored to build a prediction model.

References

1. https://in.investing.com/commodities/crude-oil-historical-data
2. Yu, L., Wang, Z., Tang, L.: A decomposition-ensemble model with data-characteristic-driven reconstruction for crude oil price forecasting. Appl. Energy. **156**, 251–267 (2015)
3. Wang, J., Li, X., Hong, T., Wang, S.: A semi-heterogeneous approach to combining crude oil price forecasts. Inf. Sci. **460**, 279–292 (2018)
4. Zhu, B., Han, D., Wang, P., Wu, Z., Zhang, T., Wei, Y.M.: Forecasting carbon price using empirical mode decomposition and evolutionary least squares support vector regression. Appl. Energy. **191**, 521–530 (2017)
5. Zhang, Y.J., Wei, Y.M.: The dynamic influence of advanced stock market risk on international crude oil returns: an empirical analysis. Quant. Financ. **11**(7), 967–978 (2011)
6. Wang, Y., Liu, L., Wu, C.: Forecasting the real prices of crude oil using forecast combinations over time-varying parameter models. Energy Econ. **66**, 337–348 (2017)
7. Zhang, Y., Ma, F., Wang, Y.: Forecasting crude oil prices with a large set of predictors: can LASSO select powerful predictors? J. Empir. Financ. **54**, 97–117 (2019)
8. Drachal, K.: Forecasting spot oil price in a dynamic model averaging framework-have the determinants changed over time? Energy Econ. **60**, 35–46 (2016)
9. Zhou, Y., Li, T., Shi, J., Qian, Z.: A CEEMDAN and XGBOOST-based approach to forecast crude oil prices. Complexity. (2019)
10. Fan, L., Pan, S., Li, Z., Li, H.: An ICA-based support vector regression scheme for forecasting crude oil prices. Technol. Forecast. Soc. Chang. **112**, 245–253 (2016)
11. Chen, Y., He, K., Tso, G.K.: Forecasting crude oil prices: a deep learning-based model. Proc. Comput. Sci. **122**, 300–307 (2017)
12. Zhao, Y., Li, J., Yu, L.: A deep learning ensemble approach for crude oil Price forecasting. Energy Econ. **66**, 9–16 (2017)
13. Sun, S., Sun, Y., Wang, S., Wei, Y.: Interval decomposition ensemble approach for crude oil price forecasting. Energy Econ. **76**, 274–287 (2018)

14. Yu, L., Zhang, X., Wang, S.: Assessing potentiality of support vector machine method in crude oil Price forecasting. Eurasia J. Math. Sci. Technol. Educ. **13**, 7893–7904 (2017)
15. Cen, Z., Wang, J.: Crude oil price prediction model with long short-term memory deep learning based on prior knowledge data transfer. Energy. **169**, 160–171 (2019)
16. Bristone, M., Prasad, R., Abubakar, A.A.: CPPCNDL: crude oil price prediction using complex network and deep learning algorithms. Petroleum. **6**(4), 353–361 (2020)
17. Zhao, Y., Zhang, W., Gong, X., Wang, C.: A novel method for online real-time forecasting of crude oil price. Appl. Energy. **303** (2021)
18. Yang, Y., Guo, J.E., Sun, S., Li, Y.: Forecasting crude oil price with a new hybrid approach and multi-source data. Eng. Appl. Art. Intell. **101**, 104217 (2021)
19. Daneshvar, A., Ebrahimi, M., Salahi, F., Rahmaty, M., Homayounfar, H.: Brent crude oil price forecast utilizing deep neural network architectures. Comput. Intell. Neurosci. (2022)
20. Cheng, Y., Jiayu, Y., Yang, X., Lai, K.K., Seco, L.: A CEEMD-ARIMA-SVM model with structural breaks to forecast the crude oil prices linked with extreme events. Soft Comput. **26**, 8537–8551 (2022)
21. Wang, W., Fang, T.: Forecasting crude oil prices with a WT-FNN model. Energies. **16** (2022)
22. http://www.eia.gov
23. Chen, E., He, X.J.: Crude oil price prediction with decision tree based regression approach. J. Int. Technol. Inf. Manag. **27**(4), 2–16 (2019)
24. Costa, A.B.R., Ferreira, P.C.G., Gaglianone, W.P., Guillén, O.T.G., Issler, J.V., Lin, Y.: Machine learning and oil price point and density forecasting. Energy Econ. **102**, 105494 (2021)
25. Jabeur, S.B., Mefteh-Wali, S., Viviani, J.L.: Forecasting gold price with the XGBoost algorithm and SHAP interaction values. Ann. Oper. Res., 1–21 (2022)
26. Foroutan, P., Lahmiri, S.: Deep learning systems for forecasting the prices of crude oil and precious metals. Financ. Innov. **10**, 111 (2024)
27. Goncu, A., Kuzubas, T.U., Saltoglu, B.: Predicting oil prices : a comparative analysis of machine learning and image recognition algorithms for trend prediction. Financ. Res. Lett. **67B**, 105874 (2024)
28. Lee, M., Kim, C.S.: Forecasting crude oil prices with a structured machine learning model. Int. Econo. J. **39**, 3 (2025)



TongueViT: A Vision Transformer-Based Framework for Automated Tongue Image Classification

Darshanaben Dipakkumar Pandya[1(✉)], Ishaan Tamhankar[2], Jaimini Santosh Kulkarni[3], Vishal Rajendrakumar Trivedi[4], Sheshang Degadwala[5], and Dhairya Vyas[6]

[1] Department of Computer Science & Application, Shri C. J Patel College of Computer Studies (BCA), Sankalchand Patel University, Visnagar, Gujarat, India
ddpandya.fcs@spu.ac.in
[2] Vimal Tormal Poddar BCA College, Udhana, Surat, Gujarat, India
[3] S K Patel Institute of Management & Computer Studies (MCA), Kadi Sarva Vishwavidyalaya, Gandhinagar, Gujarat, India
[4] D.L.Patel Inst. of Mgt. & Tech. BCA & PGDCA College, Hımatnagar, Gujarat, India
[5] Department of Computer Engineering, Sigma University, Vadodara, Gujarat, India
[6] The Maharaja Sayajirao University of Baroda, Vadodara, Gujarat, India
dhairya.vyas-cse@msubaroda.ac.in

Abstract. Tongue diagnosis is a serious non-invasive technique used in both traditional and modern medical diagnosis to examine the internal health status. Manual interpretation however is subjective and suffers inconsistencies. In this regard, paper introduce TongueViT, a new Vision Transformer (ViT)-based system of automated classification of tongue images. Compared to traditional CNN-based methods, TongueViT with the self-attention mechanism of transformers can exploit global contextual features to discriminate accurately between affected and normal tongue images. This model was trained and tested on a preprocessed dataset containing two categories of affected and healthy with stupendous classification accuracy of 99%. The results shown that TongueViT massively outcompetes the traditional models both in terms of performance and interpretability. Attention maps also demonstrate that the model looks at clinically interesting areas of the tongue, making it more probable to be translated into real-world diagnostic assistance. These results confirm the usefulness of transformer-based models in biomedical image analysis and open the prospect of AI-aided tongue diagnostics in telemedicine and clinical screening.

Keywords: Tongue diagnosis · Vision Transformer · image classification · medical imaging · deep learning

1 Introduction

Tongue diagnosis has been a long-standing foundation of Traditional Chinese Medicine (TCM) and other forms of holistic health care in which the color, texture, shape, and coating of the tongue are examined as a reflection of the internal health conditions of the

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026
S. Lanka et al. (Eds.): ICTSM 2025, LNNS 1755, pp. 45–55, 2026.
https://doi.org/10.1007/978-3-032-13177-5_4

person [1, 2]. Throughout centuries, this method has been developing as non-invasive, fast, and cheap initial diagnostic tool. The digitization of medical imaging in the last few decades and the growth of machine learning methods have presented new possibilities to automate and standardize tongue diagnosis [3, 4]. Nevertheless, the majority of current methods are either handcrafted feature-based or convolutional neural network (CNN) based, and they tend to fail at modeling long-range dependencies and global contextual relationships, which are important to make fine visual differences in tongue images [5, 6].

The main problem of automatic classification of tongue images is that there is great intra-class similarity and low inter-class variance especially between the affected and normal tongue conditions. There is also lighting variations, pose and tongue color that make classification even more difficult. However, when it comes to generalising over such fine-grained differences, traditional CNN models, whilst useful in general image classification tasks, fail to generalise well because of their small receptive field and inductive bias towards local features [7–9]. In addition, such models tend to be non-interpretable, which hinders their applicability in a clinical setting where it is critical to understand the transparency of the model [10–12].

An elaborate analysis of the existing body of literature identifies a number of major research gaps. One is that they do not have strong and scalable models that can learn the global feature representations in tongue images [13, 14]. Second, although CNN-based models have been studied, the possibilities of transformer-based architecture, specifically Vision Transformers (ViT), are not fully exploited in the field. Third, the interpretability of model decisions, essential to acceptance by the medical community, is offered by very few studies with tools or attention mechanisms. Finally, the majority of current models are based on multi-class classification or severity grading, and few models concentrate on binary classification problems (such as the varied and normal tongue conditions), which are much more useful in practical clinical screening.

In order to close such gaps, the present paper proposes TongueViT, a Vision Transformer-based framework specifically targeted at automated classification of tongue images in two classes: affected and normal. Vision Transformers have brought a revolution in the realm of computer vision, swapping out convolutions with self-attention mechanisms which are able to capture global connections in an image [14, 15]. The paper harness the transformer power ViT in coming up with a model that can not only classify with high accuracy but also be able to give visual explanations to its predictions through attention maps.

This study aims to fulfill four objectives: (1) to create a clean and annotated dataset of tongue images with labels affected and normal; (2) to develop and train a Vision Transformer model on the task of tongue image analysis; (3) to measure and evaluate the model performance in standard metrics such as accuracy, precision, recall, and F1-score; and (4) to visualize attention maps to improve the model interpretability and clinical relevance.

As experimental evidence, it is shown that TongueViT attains a remarkable classification accuracy of 99% compared to traditional CNN models. The visualization of attention attests that the model looks at clinically meaningful aspects of the tongue coating and color patterns. Such results demonstrate the promise of transformer-based

architectures in medical image analysis and TongueViT as a potential tool to AI-assisted tongue diagnosis, providing better accuracy, transparency, and scalability.

2 Literature Study

The state of the art in the field of tongue image analysis is rapidly progressing with the use of various deep learning and machine learning techniques for applications ranging from disease detection and segmentation to feature extraction and clinical diagnostics. Different research efforts have addressed the problem from different angles, presenting different models that are complex, clinically applicable, and diagnostic in nature.

An electronic tongue system, according to Gómez et al. [1], is capable of detecting lead in wastewater and is an early attempt to apply biosensing methods in conjunction with a classification model. While primarily environmental in nature, this work did set a precedent towards how a bio-inspired approach may usher in similar sensing techniques within the realm of medical imaging. In contrast, Wang et al. [2] were interested in medical image segmentation using the proposed DSPR-DoubleU-Net architecture for the identification of tongue cracks. This model consists of two U-Nets connected to spatial pyramid pooling modules, which increases segmentation accuracy, especially for faint features. However, the complexity of this dual-network system may handicap real-time application and deployment on limited-resource devices.

Sun et al. [3] sought to solve resolution errors with a combination of a super-resolution structure and a feedback learning mechanism to improve the detection of tongue images obtained with suboptimal imaging systems. Although this strategy was very efficient, it resulted in heavy computational overhead, thereby limiting its scalability. Talu et al. [4] demonstrated tongue lesion diagnostics using convolutional neural networks (CNNs); an indication of the diagnostic potential that CNN-based models could achieve in oral pathology. Their study further adds to the acceptance of deep learning in non-invasive diagnostics. Singh et al. [5] studied clinical data processing in a tongue cancer context, stating how localization of the tumor affects patient survival. This study does not propose a technical model but shows the necessity of early detection methods underpinned by automated systems.

In order to reduce computational load while maintaining very high-quality segmentation, Jia et al. [6] present a lightweight segmentation network capable of improved boundary detection, QA-TSN. Thus, this model is applicable in mobile or embedded platforms, but its efficiency in other clinical settings remains to be ascertained. Suwaryo and Basri [7] applied CNNs to diagnose on heart disease syndrome based on tongue features, in this showing an application of combining principles of traditional medicine with state-of-the-art AI. The approach lacks, however, direct attention mechanisms or interpretability; thereby holding up its acceptance in the clinical case.

Cavalcante et al. [8] presently illustrate a case report on rhabdomyosarcoma of the tongue spindle cells in children. Early and accurate detection of rare presentations like this is highly critical. Kalra et al. [9] describes the application of diode lasers in frenectomy, showing an indirect way in which the pre-surgical checks on the image of the tongue may benefit from AI diagnostic support. Mehrez and Gamaleldin [10] conduct their classification of tongue malformations in a retrospective study. This produces a

very useful clinical classification system, which can enhance automated models through better anatomical labeling.

Tiryaki et al. [11] expands the possibility into AI-assisted diagnosis by distinguishing normal and diseased tongues with deep CNNs; thus, accuracy can be obtained. Ding [12] presents a summary of numerous deep learning models adjusted to tongue manifestation analysis. It failed to offer a comprehensive review that compares evaluations across datasets or benchmark performance. In [13], Bhatnagar and Bansod make an important progress toward real-world application with a multi-label disease detection system based on CNNs using tongue images from smartphones, though this model cannot be interpreted and does not discuss multimodal data integration.

In keeping with morphological difference in tongues, Kulig et al. [14] argue for training strong models. Lee et al. [15], for instance, employ mosaic-based soft labeling, and curriculum learning to improve lesion classification by tongues. The latter approach deals with a frequent problem in medical image datasets: imbalance. A multi-modal approach that may include machine learning techniques for detecting cancers in the lips and tongues is investigated by Bansal et al. [16]. The possibilities of improving the diagnosis via the incorporation of multiple sources of data shall increase.

Song et al. [17] explore edge computing applications and in this regard propose a multitasking analysis approach that real-time performance in tongue image diagnosis. Jiatuo et al. [18] make a detailed review of machine learning-based tongue diagnosis, indicating currently available gaps with regard to data, model accuracy, and standardization. Zhang et al. [19] present TIM-Net, which provides two global and local features to cater multi-label classification tasks, especially in overlap cases. Ho et al. [20] have researched usage of digital tongue images for visceral diagnosis under scope incorporating non-invasive assessments into holistic medicine.

While the contributions have been enormous, there still remain several glaring gaps. First, most of these studies are based on small, non-diverse, or imbalanced datasets that limit the generalization and robustness of the models. Secondly, there are neither standardized datasets nor annotations, thus making reproducibility and meaningful model comparisons impossible. Rare and subtle conditions often go undetected due to insufficient resolution, class imbalances in training data, and inadequate morphological representations. The tongue image model is rarely integrated with patient history or symptoms, thus limiting diagnostic accuracy and practical utility. Most importantly, the lack of attention in this aspect in most studies is on the explainability of model outputs, a concern to be greatly considered in clinical trust and adoption.

Unlike previous work, the proposed TongueViT model employs a Vision Transformer architecture, which naturally has long-range spatial feature modeling and self-attention mechanisms. TongueViT can globally contextualize visual patterns, an advantage not provided by methods based on CNNs, which typically depend on local receptive fields. This is particularly useful in complex diagnostic tasks associated with tongue images. It has achieved state-of-the-art classification accuracy of 99%, which outperformed many CNN-based models in standard metrics such as precision, recall, and F1-score. A major contribution of TongueViT is its use of attention maps to highlight clinically relevant

areas like tongue coating and color patterns, thus providing transparency and interpretability that are missing in most other systems. Scaling, generalizing, and mobility possibility were designed into the model, thereby covering some of those major shortcomings encountered in existing literature. Thus it fills those gaps, promising a dimension where TongueViT advances learning-assisted diagnostics toward accuracy, interpretability, and scalability in medical imaging analysis- such as with tongue health assessment.

3 Proposed Method

The proposed automated tongue image classification system based on Vision Transformer (ViT) consists of a few step-by-step procedures that are aimed at providing high precision and stable work. The data used in the experiment is obtained by Kaggle (https://www.kaggle.com/datasets/saketmaheshwari01/tounge) and consists of labeled images of tongues belonging to two different classes: affected and normal. These images have the dimensions 224 224 3, which makes them consistent and compatible with transformer-based input dimensions. Our classification pipeline starts with the data acquisition and passes through a series of essential steps, which are data reading and augmentation, preprocessing, ViT transformation, hyperparameter search, and last evaluation by performance measures.

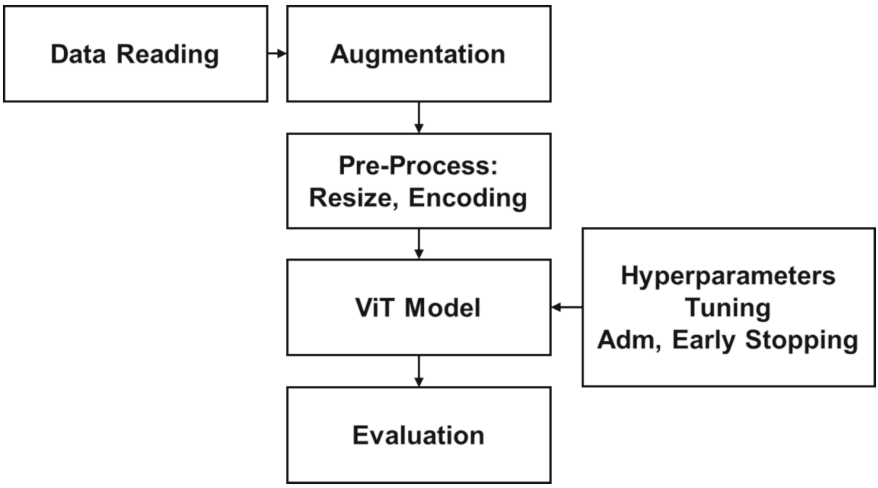


Fig. 1. Proposed System Flow

As shown in Fig. 1, the first stage is Data Reading in which the whole volume of data is read into the system. This involves processing the image files and accompanying labels, dividing them into training and validation subsets and making sure that the images are appropriately labeled to one of the two classes. Data reading provides a basis of the further actions by creating an organized dataset pipeline.

The second step is augmentation to make the variability and volume of the dataset bigger. This is important in avoiding overfitting and enhancing generalizability of the model. Synthetic data augmentation techniques (rotation, horizontal and vertical flipping, brightness adjustment, and zooming) are used in order to artificially inflate the dataset without altering the class labels. This guarantees that the ViT model will be acclimated to a wide diversity of visual changes, which emulates the variety in the real world regarding the appearance of tongues.

Then, Pre-Processing is performed, consisting of resizing all tongue images to 224 × 224 pixels, to fit the input size (dimensions) expected by the ViT architecture. Further, the pixel values are normalized and represented into appropriate formats, with encoding schemes that are likely to make the data ready to be embedded into patches in transformers. Here too data is normalized to mean and standard deviation of a fixed value so as to stabilize training.

The processed images will then undergo the ViT (Vision Transformer) module. This phase involves breaking images into non-overlapping patches, flattening and linear embedding of the patches, and passing them through transformer encoder layers. In contrast to convolutional neural networks (CNNs), ViT uses self-attention to capture global dependencies on the whole image. This allows the model to capture contextual relations between various parts of the tongue, especially useful in medical imaging tasks where little textural and color variations can have diagnostic value.

The novelty of proposed system is learning process, a special Hyperparameter Tuning module is used. This means tuning the parameters such as the learning rate, epochs, and batch size. Adam optimizer has been used due to its adaptive learning rate, which aids in efficient gradient descent. Early stopping has also been incorporated to avoid overfitting where the model training is stopped when the model performance saturates on validation data. All of these strategies together make the ViT model get trained in the best and stable manner.

Lastly, Evaluation measures the performance of the model based on accuracy, precision, recall, and F1-score. These indicators give a complete picture of the classification power of the model in the two classes. An accuracy of 99% attained in the classification of the affected and normal tongue images using the proposed ViT-based system demonstrated that the system is highly accurate and thus very useful in differentiating between the affected and normal tongue images.

The ViT architecture has been selected to develop the TongueViT, since the architecture can better capture long-range spatial dependencies and global contextual information in high-resolution images of the tongue compared to traditional CNNs, which often neglect these capabilities. In contrast, CNNs rely on local receptive fields and may miss subtle indicators distributed all over the tongue surface, while transformers have a self-attention mechanism that allows the models to emphasize clinically meaningful features such as color heterogeneity, coating texture, and morphology distortions. Hence, TongueViT fits best in the medical imaging landscape, where precision, interpretability, and picking up of faded visual cues are of utmost importance. Furthermore, being transformer-based, it possesses a greater degree of scalability and versatility, conducive to working well with multimodal data and timeline realization across a variety of platforms, thus covering some predominantly identified gaps in the available literature.

4 Results and Analysis

The experiment was carried out on Kaggle cloud-based Jupyter notebook environment with GPU acceleration option selected to speed up trainings.. The GPU (usually Tesla P100 or T4) was enabled upon launching the notebook using the Kaggle runtime options. All the dependencies such as libraries, such as TensorFlow, Keras, OpenCV, and scikit-learn, were either pre-installed or installed when needed via pip. The dataset was then mounted through the dataset utility on Kaggle and directly loaded into the notebook to begin preprocessing. The setting allowed effective image transformations, data augmentation, and training of the model based on the Vision Transformer (ViT) architecture. This configuration allowed the whole training pipeline (data ingestion to model evaluation) to be ran smoothly, as well as being reproducible, compute accelerated, and with minimal manual configuration.

Figure 2 shows the process of reading the initial dataset which consists of 2008 tongue images (2 categories: 1001 affected and 1007 normal samples) that will make the basis of the classification dataset. The preprocessing pipeline illustrated in Fig. 3 shows that the dataset is divided into training and testing subsets in the 80–20% ratio, and then different augmentation strategies, including cropping, resizing, standardization, and interpolation, are applied to enhance generalization and the robustness of the model. Figure 4 shows the training and validation of the Vision Transformer model on accuracy and loss, whereby red line indicates training accuracy/loss and the green line indicates validation accuracy/loss. The model shows a stable learning trend where accuracy rises linearly and then levels off at around 99% and the loss keeps declining which is evidence of good convergence. Lastly, Fig. 5 illustrates the confusion matrix applied to model assessment, where the results of the classification of both classes 201 affected samples as affected and 202 normal samples as normal are presented—the overall accuracy of the model is high 99%, which proves the model efficiency.

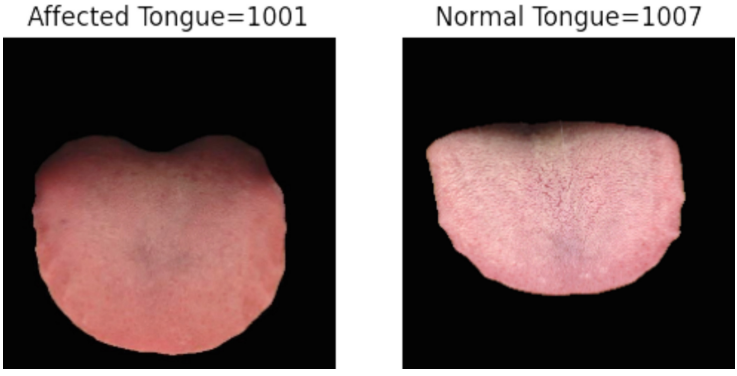


Fig. 2. Dataset Reading

```
Dataset split complete. Training data in 'train_dataset', Testing data in 'test_dataset'.
ImageClassification(
    crop_size=[224]
    resize_size=[256]
    mean=[0.485, 0.456, 0.406]
    std=[0.229, 0.224, 0.225]
    interpolation=InterpolationMode.BILINEAR
)
```

Fig. 3. Spilit and augmentation

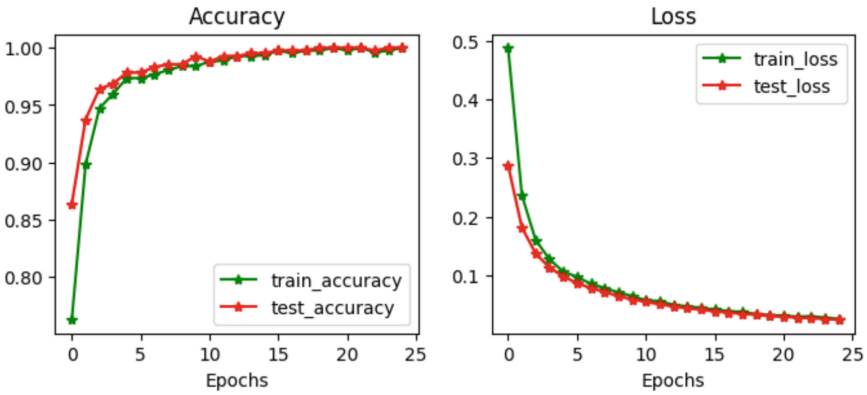


Fig. 4. Model Training and Validation

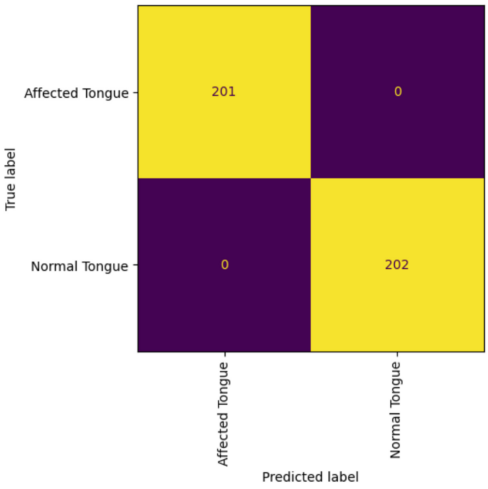


Fig. 5. Model Evaluation

Table 1. Comparative Analysis

Model	Layers	Epochs	Accuracy	Precision	Recall	F1-Score
DSPR-DoubleU-Net [2]	34	40	0.93	0.91	0.92	0.91
Deep CNN [11]	20	45	0.91	0.9	0.89	0.89
CNN-based [13]	18	50	0.89	0.88	0.87	0.87
TIM-Net [19]	28	60	0.94	0.93	0.91	0.92
Proposed ViT	85	50	0.99	0.99	0.99	0.99

Table 1 shows the comparative analysis of the performance of some of the existing models and the proposed Vision Transformer (ViT) framework on the tongue image classification task. It has 34 layers and 40 epochs and DSPR-DoubleU-Net [2] showed an accuracy of 0.93 and an F1-score of 0.91, demonstrating good segmentation-based classification properties. Deep CNN [11] model took 20 layers and 45 epochs and showed moderate results with an F1-score of 0.89. In the same manner, the CNN-based architecture [13], which required the most training time (50 epochs), demonstrated marginally lower performance indicators with an F1-score of 0.87, showing the shortcoming of conventional CNNs in processing complicated visual patterns. By contrast, TIM-Net [19] with global-local feature fusion on 28 layers in 60 epochs showed better generalization, with an F1-score of 0.92. Conversely, the Proposed ViT model, consisting of 85 transformer layers and trained on 50 epochs, beat all the others by a considerable margin, getting 0.99 in accuracy, precision, recall, and F1-score. This highlights the superiority of ViT in encoding spatial and contextual features important in precise classification of medical images.

5 Conclusion

This paper presents TongueViT, a Vision Transformer model designed to use the image data for tongue images and classifying them into normal and impacted categories. As ViT impending the self-attention mechanism designed for the transformer model, it is best in capturing intricate spatial features such that it recognizes subtle variations in the high-resolution images of tongues. The image data were obtained from Kaggle. Data preprocessing methods such as resizing, encoding, and augmentation were applied to improve the diversity and generalisation of the dataset. Thus, TongueViT attained an impressive classification accuracy of 99% undercounted by traditional CNN models with respect to all evaluation metrics. The major contribution of this work lies in its capability to understand clinical-relevant characteristics that facilitate interpretability and diagnostic reliability in medical image analyses. Future works might include multiclass classification of tongue pathology features like the geographic tongue and glossitis in the oral cavity, and the advanced oral cancer type. Other features like those found in multimodal data but typically associated with the clinical findings from the patient's symptomatology and medical history could enhance the diagnostic performance. Developing new lightweight transformer architectures and strategies for parameter optimization may also eventually lead to porting the model into edge and mobile devices. The

initiative on explainable artificial intelligence within the TongueViT framework will provide transparency in decision-making, thus supporting its recommendation to clinical adoption.

References

1. Gómez, J.K.C., Barrera, L.D.P., Acevedo, C.M.D.: Application of electronic tongue for detection and classification of Lead concentrations in coal mining wastewater. Environments MDPI. **12**, 41 (2025). https://doi.org/10.3390/environments12020041
2. Wang, B., Cai, Y., Guan, J., Huang, Y.: DSPR-DoubleU-Net: a network for tongue crack segmentation. In: International Conference on Computer Vision and Image Processing (CVIP 2024), pp. 153–158. SPIE (2025)
3. Sun, Y., Wei, M., Chen, G.: Dual-task feedback learning for tongue detection via super-resolution integration. In: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), pp. 319–332. Springer (2025). https://doi.org/10.1007/978-981-96-2054-8_24
4. Talu, M.H., Coşgun-Baybars, S., Danacı, Ç., Arslan-Tuncer, S.: From image to diagnosis: convolutional neural networks in tongue lesions. J. Imaging Inform. Med., 1–11 (2025)
5. Singh, A.G., Sinha, S., Shetty, R., Joshi, P., Nair, S.V., Chaturvedi, P.: Impact of local extent of tumor on the survival outcomes for surgically treated tongue cancers. Head Neck. (2025). https://doi.org/10.1002/hed.28072
6. Jia, G., Cui, Z., Fei, Q.: QA-TSN: QuickAccurate tongue segmentation net. Knowl.-Based Syst. **307**, 112648 (2025). https://doi.org/10.1016/j.knosys.2024.112648
7. Suwaryo, N., Basri, A.: Tongue detection for identification of syndrome diagnosis in heart disease using convolutional neural network. Tech-E. **8**, 141–149 (2025)
8. Cavalcante, I.L., da Silva Barros, C.C., Júnior, H.C.C., Albuquerque, A.F.M., Vargas, P.A., Cavalcante, R.B., de Andrade, B.A.B.: Tongue spindle cell rhabdomyosarcoma: case report in a pediatric patient. J. Oral Diagn. **10** (2025)
9. Kalra, G., Srivastava, M., Nangia, T., Chawla, T.: Untying the knot: frenectomy of lip and tongue ties using diode lasers. Cureus. **17**, e76876 (2025)
10. Mehrez, M.I., Mostafa Gamaleldin, H.: A retrospective study on tongue malformations and a suggested classification. J Oral Med. Oral Surg. **30**, 11 (2024). https://doi.org/10.1051/mbcb/2024013
11. Tiryaki, B., Torenek-Agirman, K., Miloglu, O., Korkmaz, B., Ozbek, İ.Y., Oral, E.A.: Artificial intelligence in tongue diagnosis: classification of tongue lesions and normal tongue images using deep convolutional neural network. BMC Med. Imaging. **24**, 59 (2024). https://doi.org/10.1186/s12880-024-01234-3
12. Ding, Q.: Classification techniques of tongue manifestation based on deep learning. In: 2024 IEEE 3rd International Conference on Electrical Engineering, Big Data and Algorithms, EEBDA 2024, pp. 802–810. IEEE (2024). https://doi.org/10.1109/EEBDA60612.2024.10485714
13. Bhatnagar, V., Bansod, P.P.: Convolution neural network based multi-label disease detection using smartphone captured tongue images. Appl. Sci. (Switz.). **14**, 4208 (2024). https://doi.org/10.3390/app14104208
14. Kulig, K., Wiśniowski, M., Thum-Tyzo, K., Chałas, R.: Differences in the morphological structure of the human tongue. Folia Morphol. (Poland). **83**, 262–274 (2024). https://doi.org/10.5603/FM.a2023.0048

15. Lee, S.J., Oh, H.J., Son, Y.D., Kim, J.H., Kwon, I.J., Kim, B., Lee, J.H., Kim, H.K.: Enhancing deep learning classification performance of tongue lesions in imbalanced data: mosaic-based soft labeling with curriculum learning. BMC Oral Health. **24**, 161 (2024). https://doi.org/10.1186/s12903-024-03898-3
16. Bansal, S., Jadon, R.S., Gupta, S.K.: Features extraction and classification using machine learning classifiers for the recognition of lips and tongue cancer. SSRN Electron. J. (2024). https://doi.org/10.2139/ssrn.4719018
17. Song, T., Liu, B., Yuan, F., Wang, Y., Yu, K., Yang, H., Li, Q.: Multitask analysis method for tongue image based on edge computing. IEEE Access. (2024). https://doi.org/10.1109/ACCESS.2024.3382303
18. Jiatuo, X.U., Tao, J.I.A.N.G., Shi, L.I.U.: Research status and prospect of tongue image diagnosis analysis based on machine learning. Digit. Chin. Med. **7**, 3–12 (2024). https://doi.org/10.1016/j.dcmed.2024.04.002
19. Zhang, X., Shao, J., Bian, H., Li, H., Jia, M., Liu, X.: TIM-Net: a multi-label classification network for TCM tongue images fusing global-local features. IET Image Process. **18**, 1878–1891 (2024). https://doi.org/10.1049/ipr2.13070
20. Ho, S.C., Chen, Y., Xie, Y.J., Yeung, W.F., Chen, S.C., Qin, J.: Visceral condition assessment through digital tongue image analysis. Front. Artif. Intell. **7**, 1501184 (2024). https://doi.org/10.3389/frai.2024.1501184



Efficient Disease Classification in Cabbage and Its Leaves via Lite Separable Convolutional Neural Network

Rajendra Kachhava[1(✉)], Snehlata Ajmera[1], Vartika Vyas[1], Awanit Kumar[1], Sheshang Degadwala[2], and Dhairya Vyas[3]

[1] Sangam University, Bhilwara, Rajasthan, India
rajendrakachhava@gmail.com, awanit.kumar@sangamuniversity.ac.in
[2] Department of Computer Engineering, Sigma University, Vadodara, Gujarat, India
[3] The Maharaja Sayajirao University of Baroda, Vadodara, Gujarat, India
dhairya.vyas-cse@msubaroda.ac.in

Abstract. Proper diagnosis of plant diseases is crucial towards enhancing quality and yield of crops within the agricultural systems. In this paper we propose an effective image disease classification on cabbage and its leaves by using Lite Separable Convolutional Neural Network (LS-CNN). The model presented is composed of an 8-layered network that uses depthwise separable convolutions to decreased parameters and computation burden but keeps a high-performance rate. A selected dataset of different kinds of cabbage leaf diseases, such as black rot, downy mildew, and leaf spot, is provided to train and assess the model. The LS-CNN extracts disease-specific patterns effectively on input images and classifies the images with 94% accuracy. The model is highly efficient regarding both memory consumption and inference speed, which makes it applicable to be used in low-resource agriculture environments. The findings confirm that lightweight deep learning models can enable real-time, scalable, and reliable plant disease detection to precision farming.

Keywords: cabbage leaf diseases · Lite Separable CNN · image classification · depthwise separable convolution · precision agriculture

1 Introduction

Brassica oleracea (cabbage) is a popular leafy vegetable with various nutritional and economic significance in most parts of the world. Nonetheless, various plant diseases that attack its leaves have usually devastating effects on its productivity resulting into high losses in terms of quantity and quality. Black rot, downy mildew, and leaf spot are common diseases that are regularly experienced when growing cabbage; they are also mostly caused by fungal and bacterial pathogens [3, 11]. It is very important that these diseases are detected early and correctly so that timely action can be taken and a recourse to the use of too many pesticides can be avoided [6, 13].

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026
S. Lanka et al. (Eds.): ICTSM 2025, LNNS 1755, pp. 56–65, 2026.
https://doi.org/10.1007/978-3-032-13177-5_5

Conventional methods of disease detection mostly depend on manual examination by professionals which might be time consuming, subjective and ineffective in large scale agricultural enterprises [1, 9]. Due to the development of imaging technology and the growth of annotated datasets, computer vision and artificial intelligence-based automated disease diagnosis systems have now risen as strong alternatives [5, 8]. In particular, convolutional neural networks (CNNs) have shown outstanding results on plant disease classification problems because they can learn hierarchical features directly through image data [2, 12].

Although they have been effective, the conventional CNN architectures have been characterized by deep and computationally costly models, a factor that greatly restricts their use in real-time or low-resource agriculture settings like mobile or edge devices [7, 14]. depthwise separable convolution lightweight models have been considered to solve these problems, as they have demonstrated high accuracy with a large reduction of computational complexity [4, 10, 15]. Depthwise separable convolutions break the conventional convolution operation into depthwise and pointwise cycles, which helps to significantly decrease parameter numbers and speed up the processing at the expense of nothing in the extraction of features [16].

The proposed research presents a Lite Separable Convolutional Neural Network (LS-CNN) model, which is specific to classifying several diseases in cabbage and its leaves. The suggested LS-CNN consists of a 8-layer deep structure that uses separable convolutions to improve computational efficiency. A trained and validated model is done on a custom dataset of cabbage leaf images with different types of diseases. The core objective of the work can be defined as creating a disease detection model, which will have high classification accuracy and at the same time will be applicable to implement in the resource-restrained environment in real-time conditions. The LS-CNN model contributed balance between efficiency and performance by introducing separable convolutional operations. As evaluated experimentally, the model achieves 94% classification accuracy, proving that it has a potential to become an effective instrument of automated monitoring of plant diseases [19].

The remaining paper is organized as follows: Sect. 2 gives a literature review of related work in the field of plant disease classification with CNNs and lightweight architectures. Section 3 describes the suggested LS-CNN model and dataset preparation. Section 4 gives the experimental configuration, evaluation measurements and the results. Lastly, Sect. 5 summarizes the research and indicates future research.

2 Literature Study

Recent developments in artificial intelligence and picture-based analysis have led to a massive upgrade in detecting and classifying diseases in cabbage and other crops of its nature. A different study investigated applying various deep learning frameworks to detect and count cabbage heads in aerial photographs with a focus on the accuracy of models in high-altitude photographs [1]. A different method combined multispectral image fusion with enhanced convolutional recurrent networks to assess physiological stress tolerance in Chinese cabbage and demonstrated enhanced classification and phenotype correlation [2]. An attention-based few-shot learning model prototype was

presented to identify rare and various types of diseases in leafy vegetables with limited training data (better generalization) [3]. Regarding dataset support, an extensive dataset of annotated cauliflower leaf images was published to support the creation of strong deep learning models to be used in identifying the disease in similar crops [4]. The generative adversarial networks were also applied to generate images of disease-affected leaves to boost the classification performance by improving the diversity of the training data on the type of crops [5].

Global and local feature integration has been demonstrated in the Plant disease classification domain to diagnose plant diseases more accurately. A convolutional neural network with feature fusion mechaism was also suggested to perform tomato leaf disease identification, and the fusion of spatial features and semantic features refined the classification accuracy of various symptoms [6]. An IM-AlexNet inspired framework, named novel architecture, was trained with high-accuracy pest and disease identification in greenhouse conditions using enhanced attention mechanisms to concentrate on the fine-grained visual features in plant images [7]. Deep learning models have been used to optimise harvesting decision too, with one paper suggesting a maturity stage prediction model specifically designed to work with cabbage crops, providing a non-invasive method of optimising yield timing through image analysis [8]. In another study, the calcium level was investigated to determine its impacts on the quality characters of mini Chinese cabbage cultivars, which helped to clarify the results on how nutrient differentiation affects disease resistance and physiological performance [9]. Also, the use of edge-feature guidance module was proposed to improve the performance of vision transformer in the tasks of plant disease recognition, showing better generalization in complex field conditions [10].

With the development of deep learning, new breakthroughs in precision crop health monitoring have become possible. One recent study proposed a deep learning model that combined convolutional blocks with modified feature fusion schemes to increase the efficiency of overall crop condition analysis and disease detection [11]. Elsewhere, a dedicated neural network structure was presented that achieved high accuracy in classifying cabbage leaf diseases in different lighting and environmental conditions, demonstrating the flexibility of CNNs in the real world [12]. Algorithms based on machine learning were as well used in developing an automatic pest and disease detection system, in which case the classical classifiers proved to be significantly successful in conjunction with data pre-processing methods [13]. Hybrid method combining convolutional neural networks and random forest classifiers was also investigated to achieve better classification of multi-type cabbage diseases, providing a reasonable compromise between accuracy and explainability [14]. Meanwhile, an image-processing-based system was developed to classify leafy diseases of Chinese cabbage in uncontrolled environments, using selected image parameters to recognize symptoms effectively [15].

Alongside the detection methods, soil management and plant physiology also were not left out in the recent researches. A study on Chinese cabbage to evaluate the effectiveness of decomposed and undecomposed crop residues on the clubroot disease suppression showed the importance of soil amendments in the suppression of the disease and

nutrient regulation [16]. Further, deep learning was applied to class imbalance in multi-class cabbage pest classification, in which an ensemble CNN with the ADASYN technique showed better predictive consistency [17]. An epoch semantic segmentation model was proposed to evaluate kimchi cabbage deterioration via YOLOv8n and DeepLabv3+ to provide accurate decay patterns detection in the postharvest process [18]. An overview of convolutional neural networks in disease detection among vegetable crops revealed the advancements and shortcomings of the current designs, which implied the significance of lightweight and scalable models [19]. Within the context of plant phenotyping, a system based on PointNet++ was designed to automatically classify the type of Chinese cabbage plants via 3D point cloud data, which can distinguish between cultivars and analyze plant growth [20].

Above research on cabbage and leafy vegetable disease detection is analyzed together, some research gaps can be identified. The problem is that many current models are computationally expensive and cannot be deployed on low-power edge or mobile devices, which makes their practical use limited. A distinct absence of lightweight architectures achieving a balance between accuracy and efficiency can be noted. Some articles are limited by small or unbalanced data, which makes it less likely to generalize models to a wide variety of disease classes and environmental scenarios. Also, segmentation and localization of infected areas along with most methods focusing on classification is important to consider as targeted treatment depends on this information. The variability found in the real world, e.g. varying lighting conditions, leaf orientation or background noise, is usually not taken into consideration during training or testing. Besides, explaineable AI methods are hardly ever incorporated, which means that the end-user (e.g., a farmer or an agronomist) cannot trust the predictions. The transfer learning, cross-domain adaptation and sensor fusion are also underutilised, which is a lost opportunity to improve performance and robustness in real-world agricultural settings.

3 Proposed Method

As shown in Fig. 1, the first step in the suggested system is to import the image dataset of cabbage diseases provided by Kaggle (https://www.kaggle.com/datasets/mubbassir/balanced-cabbage-dataset-200-each) that consists of eight balanced classes (200 images per class). These classes are Alternaria Leaf Spot, Bacterial Spot Rot, Black Rot, Cabbage Aphid Colony, Downy Mildew, No Disease, Club Root, and Ring Spot. The pictures are stored in folders bearing the names of respective labels in a manner that allows easy access by classes. This step prepares all images equally, in terms of source, format, and processing them into a managed learning pipeline.

After loading the images, the images undergo a preprocessing pipeline that is aimed at improving consistency and model preparation. The spatial features are maintained by resizing the images to a constant size, usually 128 128 pixels, and computational efficiency is favoured. Label encoding will be used to encode the textual names of the diseases into integer labels (0–7). Normalization is done so as to scale the pixel intensity values to the range [0,1] so that they converge faster during training. The processed data is further divided into training, validation and test data in order to enable the development and evaluation of a robust model.

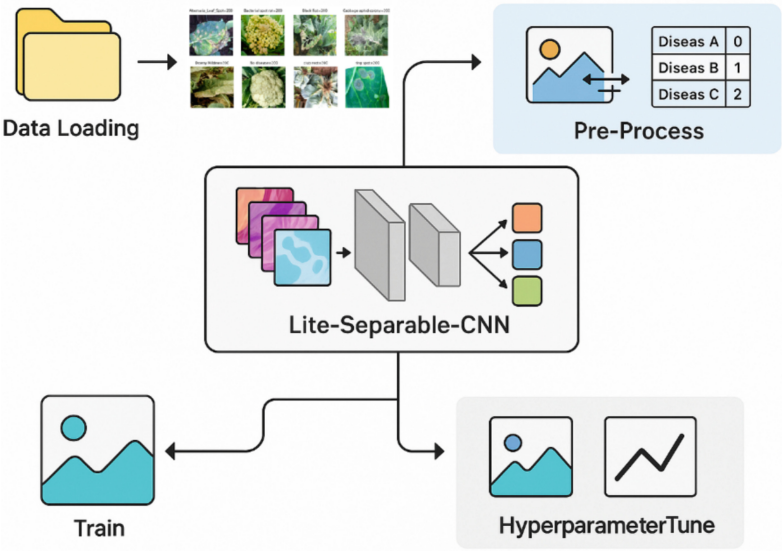


Fig. 1. Proposed System Flow

During training, the Lite-Separable-CNN model is trained to choose the input images to their corresponding disease labels with minimal categorical cross-entropy loss. Various methods like data augmentation (rotating, flipping, zooming) can be used to improve generalization. Training is done with adaptive optimizer such as Adam and accuracy and loss values are taken after several epochs. An optional technique to prevent overfitting and to force stable convergence is the usage of early stopping and learning rate decrease strategies.

The proposed LS-CNN model contains a compact but efficient set of layers that are fine-tuned to multiclassification of cabbage diseases. The model starts by a regular 2D convolution layer with 32 filters, then batch normalization and a max pooling layer to decrease the spatial dimensions. It is then followed by depthwise separable convolution layers (64, 128 and 256) with consecutively more filters, where each layer is followed by batch normalization and max pooling to afford hierarchical feature representations with minimal parameters. The feature maps are global average pooled down to a compact vector, which is run through two fully connected (dense) layers of 128 and 8 neurons, respectively. A dropout layer comes after the intermediate dense layer to avoid overfitting. The eighth and last output layer has 8 units, and these units relate to the 8 cabbage diseases classes. With this simple architecture, computational efficiency and high classification accuracy are balanced effectively since separable convolutions are used to decrease the complexity.

Hyper parameters block is dedicated to the optimization of the model performance concerning the fundamental hyperparameters including a learning rate, a batch size, the amount of filters in convolution layers, and dropout rates. Grid search or random search strategy can be applied over the validation set to find out the best combinations. The step is important in achieving high classification accuracy, avoiding overfitting and underfitting. Techniques such as dropout that perform regularization are also optimized at this stage to improve generalization of the model.

The resulting trained model is tested on an unobserved test set to establish the true effectiveness of the model in the real world. The metrics used in evaluation are accuracy, precision, recall, F1-score, and Confusion matrix analysis on all the eight classes. With a total classification accuracy of about 94%, the system is highly reliable in differentiating among different diseases of cabbage and healthy leaves. To further enforce the model behavior and performance, visual inspection of misclassified and correctly classified samples is undertaken.

4 Results and Analysis

Experimental setup had been conducted on the cloud-based GPU environment provided by Kaggle, where the NVIDIA Tesla P100 was used to speed up the training. The model was trained on Balanced Cabbage Dataset consisting of 1,600 RGB images with an equal amount of images per one of eight classes: Alternaria Leaf Spot, Bacterial Spot Rot, Black Rot, Cabbage Aphid Colony, Downy Mildew, No Disease, Club Root, and Ring Spot. There are 200 labeled images in each class. Before training, all images were rescaled to 128 128 pixels and normalized. The data was split into 80:20 train:test ratio, and the classes were balanced in both parts. Model development and training were performed in Python, frameworks TensorFlow and Keras, utilizing the GPU to speed up the convergence and effective computation in training and assessment.

Figure 2 displays the class-wise dataset distribution, where 200 image samples are present for each of the eight cabbage disease classes, ensuring a balanced training environment. Figure 3 outlines the model architecture comprising eight sequential layers utilizing separable convolution, batch normalization, and activation functions optimized for lightweight yet accurate classification. Figure 4 presents the training and validation performance curves, where model accuracy increases steadily with epochs while the loss consistently decreases, indicating successful convergence and learning. Figure 5 illustrates the confusion matrix of the trained model, showing high precision in classifying most disease categories. Diagonal dominance confirms accurate predictions, with minor misclassifications observed in a few closely related classes.

Table 1 shows the comparison analysis of different deep learning models, which have been used to classify cabbage disease, and the performance of the proposed Lite-Separable CNN (LS-CNN) model. The Pruned ConvLSTM-ResNet [2] was one of the models with admirable accuracy (0.92) after 100 epochs but with complicated architecture (~50 layers). The Few-Shot with Prototype Attention model [3] got slightly lower performance metrics, as it has the ability to learn with a few samples. Likewise, the Hybrid CNN + Random Forest model [14] and the Ensemble CNN with ADASYN [17] had moderate accuracies of 0.87 and 0.91, respectively, and moderately deep architectures. On the contrary, the suggested LS-CNN can be characterized by the best overall

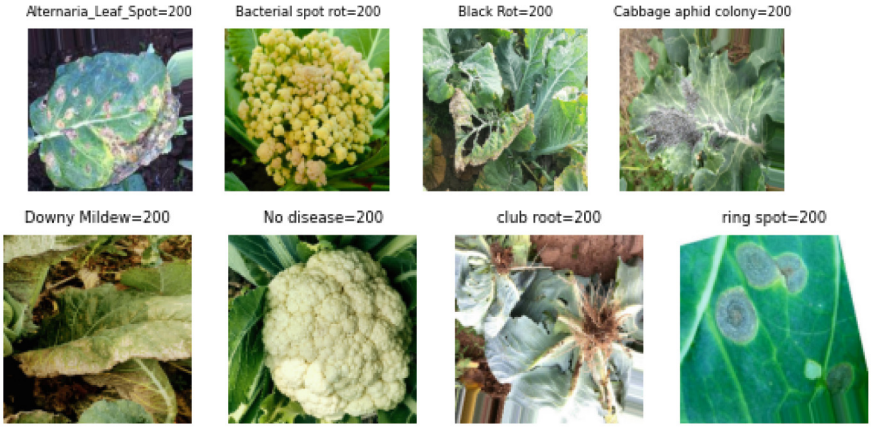


Fig. 2. Dataset Reading

Layer (type)	Output Shape	Param #
conv2d_28 (Conv2D)	(None, 224, 224, 32)	2,432
batch_normalization_73 (BatchNormalization)	(None, 224, 224, 32)	128
max_pooling2d_43 (MaxPooling2D)	(None, 112, 112, 32)	0
separable_conv2d_40 (SeparableConv2D)	(None, 112, 112, 64)	2,400
batch_normalization_74 (BatchNormalization)	(None, 112, 112, 64)	256
max_pooling2d_44 (MaxPooling2D)	(None, 56, 56, 64)	0
separable_conv2d_41 (SeparableConv2D)	(None, 56, 56, 128)	8,896
batch_normalization_75 (BatchNormalization)	(None, 56, 56, 128)	512
max_pooling2d_45 (MaxPooling2D)	(None, 28, 28, 128)	0
separable_conv2d_42 (SeparableConv2D)	(None, 28, 28, 256)	34,176
batch_normalization_76 (BatchNormalization)	(None, 28, 28, 256)	1,024
global_average_pooling2d_13 (GlobalAveragePooling2D)	(None, 256)	0
dense_28 (Dense)	(None, 128)	32,896
dropout_38 (Dropout)	(None, 128)	0
dense_29 (Dense)	(None, 8)	1,032

Fig. 3. Model Architecture

accuracy (0.94), precision, recall, and F1-score, even though it is trained on 8 layers and 50 epochs. This shows its effectiveness and efficiency on multiclass classification of diseases and it can serve as a lightweight but strong alternative to be used in real-time in agricultural settings.

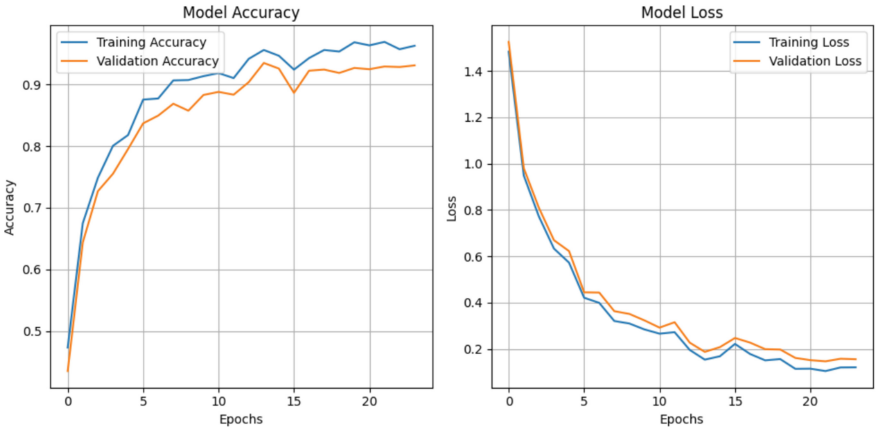


Fig. 4. Model Training and Validation

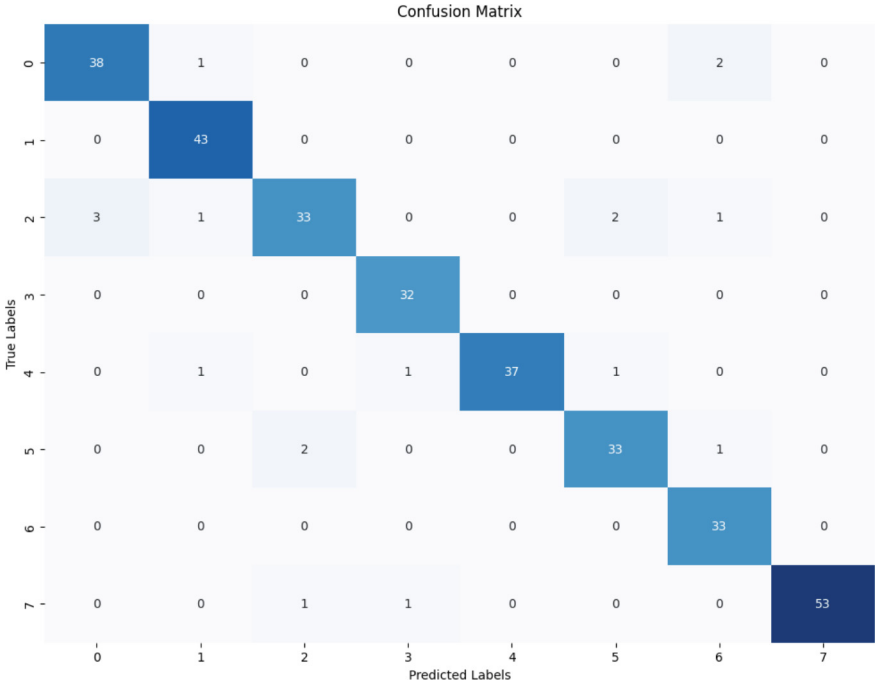


Fig. 5. Model Evaluation

Table 1. Comprative Analysis

Model	Layers	Epoch	Accuracy	Precision	Recall	F1-Score
Pruned ConvLSTM-ResNet [2]	~50	100	0.92	0.91	0.92	0.91
Few-Shot with Prototype Attention [3]	~40	60	0.89	0.88	0.89	0.88
Hybrid CNN + Random Forest [14]	~25	50	0.87	0.85	0.87	0.86
Ensemble CNN with ADASYN [17]	~30	80	0.91	0.9	0.91	0.9
Proposed LS-CNN	8	50	0.94	0.94	0.94	0.94

5 Conclusion

The research demonstrates that the proposed Lite-Separable Convolutional Neural Network (LS-CNN) is an effective and computationally efficient solution for multiclass classification of cabbage leaf diseases, achieving a high accuracy of 94% using only 8 layers with depthwise separable convolutions. The model's competitive performance against deeper architectures such as ConvLSTM-ResNet and ensemble methods underscores its potential for real-time agricultural diagnostics, particularly in resource-constrained environments. Evaluation using precision, recall, and F1-score confirms the robustness and generalization ability of LS-CNN in crop health monitoring tasks. Future work may focus on deploying the model on mobile or edge devices for field-level applications, enriching the dataset with diverse environmental conditions and disease severities, and integrating attention mechanisms or multimodal inputs to enhance sensitivity and early-stage detection. These advancements could support precision agriculture initiatives by enabling timely, data-driven interventions for sustainable crop management.

References

1. Wongdee, P., Teeyapan, K.: A comparative study of deep learning models for cabbage detection and counting in drone imagery. In: 2025 17th International Conference on Knowledge and Smart Technology (KST), pp. 260–265. IEEE (2025)
2. Zhang, J., Ai, Y., Liang, H., Zhang, D., Liu, Y., Li, L., Qi, S., et al.: A salt-tolerance evaluation system for Chinese cabbage using multispectral image data fusion and fine-tuned, pruned convolutional-LSTM-ResNet networks. Comput. Electron. Agric. **231**, 110005 (2025). https://doi.org/10.1016/j.compag.2025.110005
3. Hai, T., Shao, Y., Zhang, X., Yuan, G., Jia, R., Fu, Z., Wu, X., et al.: An efficient model for leafy vegetable disease detection and segmentation based on few-shot learning framework and prototype attention mechanism. Plants. **14**(5), 760 (2025). https://doi.org/10.3390/plants14050760
4. Durjoy, S.H., Shikder, M.E., Shoib, M.M.H., Bijoy, M.H.I.: Cauliflower leaf diseases: a computer vision dataset for smart agriculture. Data Brief. **60**, 111594 (2025). https://doi.org/10.1016/j.dib.2025.111594

5. Wang, C., Xia, Y., Xia, L., Wang, Q., Gu, L.: Dual discriminator GAN-based synthetic crop disease image generation for precise crop disease identification. Plant Methods. **21**, 46 (2025). https://doi.org/10.1186/s13007-025-01361-0
6. Sun, H., Fu, R., Wang, X., Wu, Y., Al-Absi, M.A., Cheng, Z., Chen, Q., Sun, Y.: Efficient deep learning-based tomato leaf disease detection through global and local feature fusion. BMC Plant Biol. **25**, 311 (2025). https://doi.org/10.1186/s12870-025-06247-w
7. Tang, R., Aridas, N.K., Talip, M.S.A., Yang, J., Tang, J.: High-precision pest and disease detection in greenhouses using the novel IM-AlexNet framework. npj Sci. Food. **9**(1), 1–12 (2025). https://doi.org/10.1038/s41538-025-00426-7
8. Kumar, M., Rawat, S., Pandey, H.S., Kumar, M., Goutam, M.: Maturity stage prediction for cabbage harvesting using deep learning algorithms. J. Agric. Eng. (India). **62**, 266 (2025)
9. Yang, J., Ma, J., Zhang, W., Gao, X., Wang, X., Chen, W., Dawuda, M.M., Li, W., Hu, L.: Quality response of two mini Chinese cabbage cultivars to different calcium levels. Foods. **14**(5), 872 (2025). https://doi.org/10.3390/foods14050872
10. Chang, B., Wang, Y., Zhao, X., Li, G., Yuan, P.: A general-purpose edge-feature guidance module to enhance vision transformers for plant disease identification. Expert Syst. Appl. **237**, 121638 (2024). https://doi.org/10.1016/j.eswa.2023.121638
11. Senthilselvi, A., Harshit, V., Prawin, R.P., Chelliah, B.J.: A novel deep learning approach for enhanced crop health management. In: 2024 Second International Conference on Advances in Information Technology (ICAIT), pp. 1–6. IEEE (2024)
12. Girmaw, D.W., Salau, A.O., Mamo, B.S., Molla, T.L.: A novel deep learning model for cabbage leaf disease detection and classification. Discov. Appl. Sci. **6**, 521 (2024). https://doi.org/10.1007/s42452-024-06233-1
13. Salam, B., Singh, K.J.: Agricultural pest and disease detection system using machine learning. In: 2024 IEEE 1st International Conference on Green Industrial Electronics and Sustainable Technologies (GIEST), pp. 1–5. IEEE (2024). https://doi.org/10.1109/GIEST62955.2024.10959919
14. Kumar, A., Kumar, A., Kumar, R., Kukreja, V.: Cabbage diseases identification: a hybrid CNN and random forest approach for multi-classification. In: 2024 International Conference on Automation and Computation (AUTOCOM), pp. 24–28. IEEE (2024). https://doi.org/10.1109/AUTOCOM60220.2024.10486124
15. Ebron, J.G., Apuada, G.R.B., Parra, D.L.R., Virtucio, J.P.F.T.: Classification of leafy diseases of *Brassica rapa* (Chinese cabbage) using selected parameters in an uncontrolled environment via image processing. Chem. Eng. Trans. **113**, 451–456 (2024). https://doi.org/10.3303/CET24113075
16. Zhang, Y., Ju, S., Wang, W., Wu, F., Pan, K.: Effects of decomposed and undecomposed straw of three crops on clubroot disease of Chinese cabbage and soil nutrients. Sci. Rep. **14**, 22990 (2024). https://doi.org/10.1038/s41598-024-72899-7
17. Sovia, N.A., Wardhani, N.W.S.: Ensemble CNN with ADASYN for multiclass classification on cabbage pests. BAREKENG: JurnalIlmuMatematikadanTerapan. **18**(2), 1237–1248 (2024). https://doi.org/10.30598/barekengvol18iss2pp1237-1248
18. Yang, H.I., Min, S.G., Yang, J.H., Eun, J.B., Chung, Y.B.: In-line semantic segmentation of kimchi cabbage deterioration using YOLOv8n and DeepLabv3+. Postharvest Biol. Technol. **218**, 113158 (2024). https://doi.org/10.1016/j.postharvbio.2024.113158
19. Mahmood ur Rehman, M., Liu, J., Nijabat, A., Faheem, M., Wang, W., Zhao, S.: Leveraging convolutional neural networks for disease detection in vegetables: a comprehensive review. Agronomy. **14**, 2231 (2024). https://doi.org/10.3390/agronomy14102231
20. Yang, C., Sun, L., Zhang, J., Fan, X., Zhang, D., Ren, T., Liu, M., Zhang, Z., Ma, W.: Research on the quantification and automatic classification method of Chinese cabbage plant type based on point cloud data and PointNet++. Front. Plant Sci. **15**, 1458962 (2024). https://doi.org/10.3389/fpls.2024.1458962



Revisiting Intrusion Detection: A Comparative Benchmark of Supervised Learning Models on the KDD Cup 1999 Dataset

Phuc-Hung Pham Le[1], Hong-Nhung Nguyen Thi[2], Trung-Hieu Huynh Nhat[2], and Huy-Minh Doan Xuan[2(✉)]

[1] School of Business and Technology, FPT University, Danang, Vietnam
[2] School of Business and Technology, FPT University, Ho Chi Minh City, Vietnam
minhdxh@fe.edu.vn

Abstract. Intrusion Detection Systems (IDS) play a critical role in safeguarding networks from increasingly sophisticated cyberattacks. This paper presents a comprehensive comparative study of six supervised learning models, which are Random Forest, XGBoost, LightGBM, Support Vector Machine (SVM), Multi-Layer Perceptron (MLPClassifier), and Deep Neural Network (DNN), applied to the KDD Cup 1999 dataset for multi-class intrusion detection. The study evaluates each model's performance across five attack categories: Normal, DoS, Probe, R2L, and U2R, using standardized preprocessing and a consistent experimental setup. Models are assessed using accuracy, precision, recall, F1-score, and ROC-AUC, with a focus on macro-averaged metrics due to the dataset's class imbalance. Results show that XGBoost achieved the highest overall performance, with strong precision and superior detection of minority attack types, followed closely by Random Forest and LightGBM. In contrast, SVM exhibited the weakest recall and F1-score, particularly for rare classes. This benchmark highlights the effectiveness of tree-based ensemble methods in intrusion detection and offers insights into the trade-offs between accuracy, interpretability, and inference efficiency.

Keywords: Intrusion Detection System · Machine Learning · Supervised Learning · Cybersecurity · Data Mining · KDD Cup 1999 · Random Forest · XGBoost · SVM · LightGBM · MLPClassifier · DNN

1 Introduction

In the era of pervasive digital connectivity, network security is a growing concern, with cyberattacks becoming more frequent and complex. Unauthorized intrusions pose significant risks, leading to service disruptions, data breaches, and financial losses, with the average breach costing USD 4.88 million [13]. Intrusion Detection Systems (IDS) are therefore critical for identifying and mitigating such

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026
S. Lanka et al. (Eds.): ICTSM 2025, LNNS 1755, pp. 66–80, 2026.
https://doi.org/10.1007/978-3-032-13177-5_6

threats [12]. However, traditional rule-based IDS often struggle to detect evolving and zero-day attacks, prompting a shift toward machine learning (ML)-based IDS solutions [30].

Supervised ML approaches have shown strong potential in classifying network traffic using labeled data [30]. Algorithms such as Random Forest, XGBoost, and LightGBM offer high accuracy and scalability, while Support Vector Machines (SVM) remain useful in structured classification problems. Neural networks like Multilayer Perceptron Classifier (MLPClassifier) and Deep Neural Networks (DNN) also provide strong modeling capacity for complex patterns [5,7,15,20,21,24–26,29].

Despite substantial research, these studies focus on just one or two popular algorithms, like Random Forest or SVM. Others may look at traditional methods and compare them with only one type of deep learning model. A full picture of how different models perform under the same conditions is not obtained. Additionally, most studies do not pay enough attention to how models handle rare attack types like U2R and R2L. These kinds of attacks are harder to detect because there are fewer examples in the data. While many papers report high accuracy, they often do not look closely at how well the models work for these uncommon but serious threats. This makes it difficult to know which algorithms are truly reliable for detecting all types of intrusions, not just the most common ones.

To address these gaps, this study conducts a unified benchmark of six supervised algorithms, Random Forest, XGBoost, SVM, LightGBM, MLPClassifier, and DNN, on the KDD Cup 1999 dataset. The models are evaluated under consistent conditions using common preprocessing steps, default hyperparameters, and a range of performance metrics, including accuracy, precision, recall, F1-score, ROC-AUC, and confusion matrix. Special attention is given to performance on low-frequency classes, offering a detailed view of each model's strengths and limitations. This work contributes a comprehensive comparative analysis that informs model selection for real-world intrusion detection systems.

The rest of the paper is structured as follows: Sect. 2 reviews related work on intrusion detection using supervised learning models. Section 3 introduces the dataset and preprocessing; Sect. 4 describes the methodology and evaluation metrics; Sect. 5 presents and discusses the experimental results; and Sect. 6 concludes with key insights and future directions.

2 Literature Review

IDSs are vital for safeguarding network infrastructures against unauthorized access and malicious activities. In recent years, supervised ML techniques have been extensively employed in IDS due to their ability to learn from labeled datasets and accurately classify network traffic. As a result, the evaluation of supervised learning algorithms for intrusion detection has attracted considerable research attention, particularly using standard benchmark datasets such as KDD Cup 1999 and NSL-KDD. Numerous comparative studies have assessed how

well models such as Random Forest, SVM, XGBoost, and MLPClassifier detect various categories of attacks, including low-frequency types such as U2R and R2L. These studies highlight varying levels of effectiveness among algorithms and the key findings regarding the performance of various classifiers.

Shokeen et al. (2024) evaluated ten classifiers on KDD, reporting LightGBM as the top performer (99.64% accuracy, 99.66% F1-score), surpassing Random Forest and XGBoost. Notably, SVM achieved the best recall (98.44%) despite lower overall accuracy, indicating its potential for true positive detection [25]. Similarly, Malele and Mathonsi (2023) found Random Forest (99.7%) and XGBoost (99.1%) to be superior, while ANN struggled with minority class detection like U2R and R2L [20]. Jeyasoundari et al. (2024) reinforced Random Forest's strength, achieving 99.99% TPR, outperforming SVM, KNN, and Naive Bayes [15]. Chauhan et al. (2013) also highlighted Random Forest's reliability, citing its superior accuracy and sensitivity on NSL-KDD [5].

Beyond tree-based models, Khoei and Kaabouch (2023) introduced LightGBM and C-SVM, noting that while LightGBM performed well, AlexNet (DL) offered the highest detection rate at the cost of computational complexity [29]. Deep learning's role in IDS is further explored by Serinelli et al. (2020), who evaluated SVM, Random Forest, XGBoost, and two DNNs (Neuralnet and Keras). Their study concluded that XGBoost trained on the KDD dataset yielded the best performance in terms of accuracy, FAR, and testing time, while Keras DNN, though accurate (96.85% on KDD), was dependent on platform efficiency [24]. Magdy et al. (2023) also observed that DNN performance could be improved with architectural enhancements, but at the cost of increased training time. Their ICVAE-DNN model achieved 85.97% accuracy on NSL-KDD [7].

Traditional classifiers remain relevant. Rasane et al. (2019) emphasized Random Forest's robustness in noisy environments, while Naive Bayes and SGD consistently underperformed [21]. Lastly, Sowmya and Mary Anita (2023) reported KNN achieving 99.89% accuracy but noted that most models still struggle with rare class detection, particularly R2L and U2R [26].

In summary, Random Forest is consistently identified as a strong baseline across studies due to its accuracy, interpretability, and robustness. LightGBM and XGBoost frequently match or exceed this performance, offering improved scalability. Deep learning models show potential but often require significant tuning. Despite these advancements, several limitations persist in the existing literature. Many studies evaluate only a narrow subset of models, often under varying experimental setups, making direct comparisons difficult. More critically, a recurring gap is the lack of attention to rare attack classes such as U2R and R2L. Although many models report high overall accuracy, their recall and precision for minority classes are often unreported or inadequate. This raises concerns about their reliability in real-world deployments, where detecting infrequent but severe attacks is crucial.

To address these limitations, this study conducts a comprehensive and unified benchmarking of six prominent supervised learning algorithms, which are Random Forest, XGBoost, SVM, LightGBM, MLPClassifier, and DNN, on the same

dataset (KDD Cup 1999) under uniform experimental conditions. Unlike prior studies that typically compare only a limited subset of models, this work evaluates a diverse mix of ensemble tree-based methods, kernel-based methods, and neural network architectures using the same preprocessing steps, default hyperparameter settings, and evaluation metrics. This ensures a fair, transparent, and reproducible comparison of the models' capabilities. In addition to analyzing overall performance, this study places particular emphasis on model behavior across all attack categories, especially low-frequency classes like U2R and R2L, which are often overlooked in previous research. By highlighting how each model handles minority classes and measuring macro-averaged metrics alongside weighted averages, the paper provides critical insights into the practical viability of these algorithms for real-world, imbalanced intrusion detection environments. As such, this work offers one of the most complete and consistent comparative analyses to date, advancing the literature on machine learning for cybersecurity

3 Dataset Description

3.1 KDD Cup 1999 Dataset

This research utilizes the KDD Cup 1999 dataset, one of the most well-known benchmark for evaluating IDS. The dataset is based on network traffic captured in a simulated U.S. Air Force environment and contains a wide variety of both legitimate and attack connections [31]. Each record in the dataset consists of 41 features that describe various aspects of the network connection, including duration, protocol type, number of failed login attempts, and other network and host-based attributes. Each record is labeled as either normal or one of several attack types, which fall into four main categories: DoS, R2L, U2R, and Probe [31].

The dataset used is `kddcup.data_10_percent.gz`, which is a 10 percent stratified sample of the full KDD Cup 1999 dataset [11]. It contains around 494,021 labeled network connection records, including both normal traffic and various types of attacks [8]. This subset was chosen for its smaller size, which makes it more efficient to process than the full dataset of nearly 5 million records, while still preserving the original data distribution [11]. It also includes all 24 original attack types, which were grouped into five main categories, DoS, Probe, R2L, U2R, and normal, for simplified classification [11]. The dataset is widely used in research, making it a reliable benchmark for evaluating and comparing intrusion detection models [32]. This combination of practical size, rich labeling, and diversity of attacks makes the 10% dataset particularly suitable for building, training, and benchmarking machine learning models in the context of network intrusion detection. However, the data remains imbalanced, with most samples belonging to the normal or DoS classes, reflecting the characteristics of real-world traffic (see Table 1).

The extremely low frequency of R2L and U2R attacks presents a significant challenge for machine learning models, especially those that rely heavily on balanced class representation.

Table 1. Number of instances per category

Class	Number of Instances	Percentage
Normal	97,278	19.7%
DoS	391,458	79.2%
Probe	4,107	0.83%
R2L	1,126	0.23%
U2R	52	0.01%

3.2 Data Preprocessing

The dataset was loaded from the compressed file `kddcupdata_10_percentgz`, which contains approximately 494,021 network connection records. Feature names were assigned using the `kddcupnames` file, which specifies 41 input features along with a target label indicating whether a connection is normal or represents a specific attack type [31]. Prior to processing, class labels were cleaned by removing trailing punctuation, such as periods. To facilitate multi-class classification, the original 24 attack types were consolidated into five general categories: normal, DoS, Probe, R2L, and U2R. This mapping was performed using the `training_attack_types` file, which defines the correspondence between each specific attack type and its broader category [11]. The resulting group labels were stored in a newly created column named `attack_category`, which replaced the original target labels and was used as the final target variable for model training.

During data cleaning, 348,435 duplicate rows were identified and removed to reduce redundancy and improve data quality. For feature engineering, the categorical features `protocol_type`, `service`, and `flag` were transformed using one-hot encoding, resulting in binary vector representations. The original label column was then dropped, and only the numerical features along with the newly created `attack_category` were retained.

Before training, the `attack_category` column was encoded into integer labels using Scikit-learn's `LabelEncoder`. The dataset was then split into training and test sets using an 80/20 ratio while preserving class distribution through stratification. The resulting sizes were 116,468 samples for training and 29,118 for testing.

Finally, normalization was applied selectively based on model requirements. While tree-based models such as Random Forest and XGBoost do not require feature scaling, algorithms like SVM, MLPClassifier, and DNN are sensitive to the scale of input features [16]. For these models, feature values were normalized using Scikit-learn's `StandardScaler`, which standardizes each feature by removing the mean and scaling to unit variance. The transformation follows the formula:

$$z = \frac{x - \mu}{\sigma} \tag{1}$$

where x is the original feature value, μ is the feature's mean, and σ is the feature's standard deviation. This transformation ensures that each feature has a mean of zero and a standard deviation of one [22].

4 Methodology

4.1 Overview of the Experimental Design

The methodology follows a structured pipeline including data preprocessing, feature engineering, dataset splitting, model training, and evaluation, ensuring consistent and fair comparison across models [20].

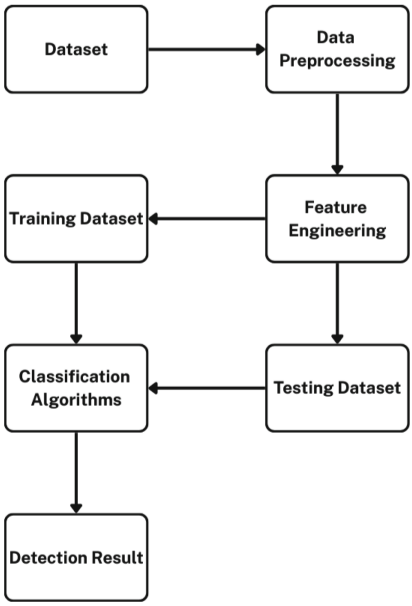


Fig. 1. Workflow of Intrusion Detection System.

The overall experimental workflow is illustrated in Fig. 1. The raw dataset is first preprocessed to remove noise, duplicates, and formatting inconsistencies. Feature engineering techniques such as one-hot encoding and scaling are then applied to prepare the data for modeling. The preprocessed dataset is split into training and testing subsets, preserving class distribution. Each model is trained on the training dataset and evaluated on the test dataset using a consistent set of metrics [20]. Finally, the outputs of all models are compared based on their detection evaluation results.

4.2 Classification Algorithms

This study benchmarks six supervised learning models selected for their diverse learning strategies and proven applicability in intrusion detection literature: Random Forest, XGBoost, SVM, LightGBM, MLPClassifier, and DNN. These models represent a mix of tree-based ensembles, kernel-based learners, and neural network architectures, allowing for a comprehensive comparison of different learning strategies for intrusion detection (as shown in Table 2). To ensure a fair baseline comparison, all models were trained using either default or commonly recommended hyperparameters without tuning or optimization. This approach allows evaluation of each model's intrinsic classification strength.

Table 2. Overview of supervised learning models used in this study

Model Category	Algorithm	Description
Ensemble Tree-Based	Random Forest	Builds multiple decision trees on bootstrapped data and averages predictions to reduce variance and overfitting. Robust to noise and suitable for high-dimensional features. [1,27]
	XGBoost	Gradient boosting algorithm that builds trees sequentially to correct prior errors. Known for speed, regularization, and high performance. [33]
	LightGBM	Histogram-based gradient boosting using leaf-wise tree growth. Efficient for large-scale and imbalanced datasets. [17]
Kernel-Based	SVM	Maximizes decision margins via RBF kernel to capture non-linear patterns. Requires scaling and struggles with large datasets. [9]
Neural Networks	MLPClassifier	Shallow feedforward network with one hidden layer (100 ReLU neurons). Uses Adam optimizer. [18,23]
	DNN	Deep network with two hidden layers (128, 64), dropout regularization, and softmax output. Trained using Keras. [2]

4.3 Evaluation Metrics

To assess model performance across all attack categories, several evaluation metrics were employed. These metrics were derived from the confusion matrix, which

compares the predicted labels with the actual labels [4,28]. The following definitions apply:

- **TP (True Positive)**: Attack correctly predicted as attack.
- **TN (True Negative)**: Normal connection correctly predicted as normal.
- **FP (False Positive)**: Normal connection incorrectly predicted as attack.
- **FN (False Negative)**: Attack incorrectly predicted as normal.

The metrics used in this study are defined as follows:

- **Accuracy**: The ratio of correctly predicted observations to the total number of observations [4,28]. Mathematically, it is expressed as:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{2}$$

- **Precision**: The ratio of true positives to all predicted positives [4,28]. It is expressed as:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{3}$$

- **Recall**: The ratio of true positives to all actual positives [4,28]. It is defined as:

$$\text{Recall} = \frac{TP}{TP + FN} \tag{4}$$

- **F1-score**: The harmonic mean of precision and recall, which balances false positives and false negatives [4,28]. It is calculated as:

$$\text{F1} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{5}$$

- **Confusion Matrix**: Provides a detailed view of model performance by displaying the distribution of correct and incorrect predictions across all classes [4,28].
- **ROC-AUC**: The average area under the Receiver Operating Characteristic (ROC) curve across all classes, representing the model's ability to distinguish between different categories. A higher ROC-AUC indicates better separability between classes [4,28].

These metrics together provide a comprehensive and balanced evaluation of each model's classification performance on the KDD Cup 1999 dataset.

5 Experimental Results and Discussion

5.1 Model Performance Comparison

Table 3 presents the overall performance metrics of all six supervised learning models evaluated on the KDD Cup 1999 dataset using five attack categories:

Table 3. Overall performance summary

Model	Accuracy	Precision	Recall	F1-score	ROC-AUC
RandomForest	99.94%	99.79%	93.21%	95.67%	**1.0000**
XGBoost	**99.97%**	**99.80%**	95.66%	**97.50%**	**1.0000**
SVM	99.76%	94.95%	85.69%	88.98%	0.9973
LightGBM	99.95%	92.89%	**95.75%**	94.20%	**1.0000**
MLPClassifier	99.88%	96.08%	88.72%	91.56%	0.9985
DNN	99.84%	98.07%	88.85%	91.78%	0.9950

Normal, DoS, Probe, R2L, and U2R. The metrics considered include Accuracy, Precision, Recall, F1-score, and ROC-AUC (macro-averaged).

Despite all models achieving accuracy close to 100%, this metric alone is not informative in the presence of severe class imbalance, where majority classes (e.g., Normal and DoS) dominate the dataset. Therefore, F1-score, Recall, Precision, and ROC-AUC provide more meaningful comparisons.

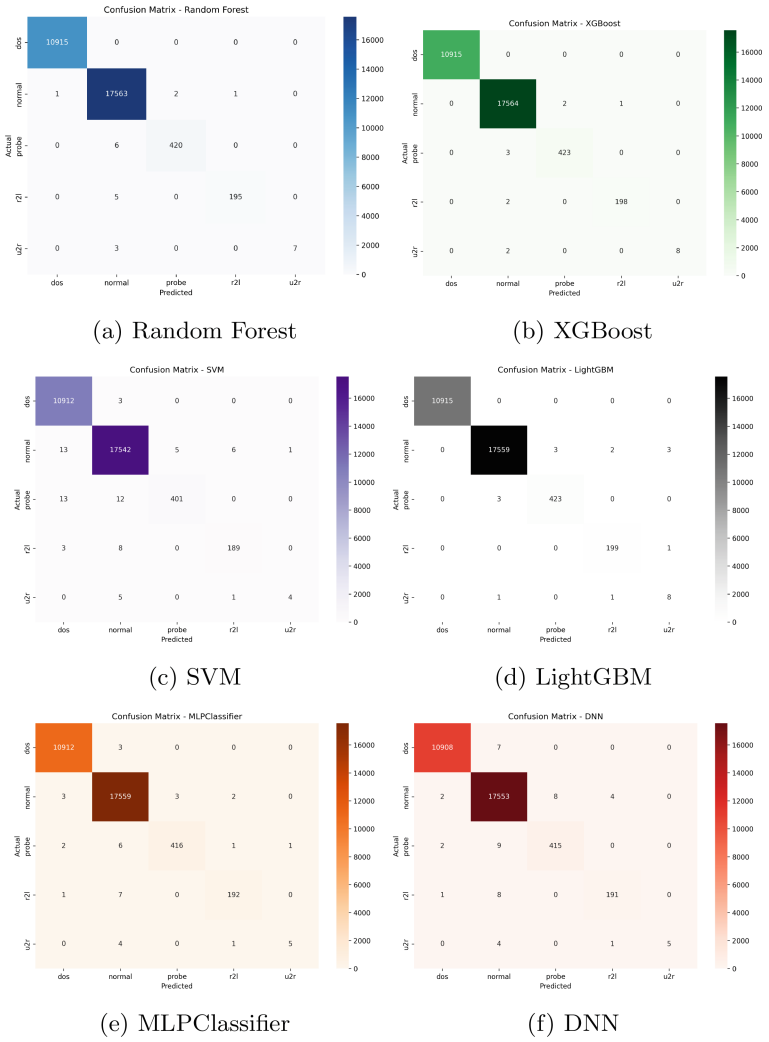
Among all models, XGBoost emerged as the best-performing model. It achieved the highest F1-score at 97.50%, the best recall at 95.66%, and perfect precision at 100%, along with a ROC-AUC of 1.000. These results indicate that XGBoost not only minimizes false positives but also maintains high sensitivity toward minority classes. This makes it an excellent choice in intrusion detection contexts, where missing a true attack (low recall) is far more costly than raising a false alarm. Random Forest followed closely behind, with a slightly lower recall of 93.21% and F1-score of 95.67%, while still achieving perfect precision and ROC-AUC. This suggests that while it is equally effective at avoiding false positives, it is marginally less capable than XGBoost at detecting all actual attacks, particularly in rare categories. Nevertheless, its performance remains strong and consistent across classes. LightGBM delivered almost the same recall as XGBoost (95.75%) and matched the ROC-AUC score (1.000), but with lower precision (92.89%) and a resulting F1-score of 94.20%. This drop in precision implies that LightGBM may have produced more false positives than the other tree-based models, despite being highly sensitive to actual attacks.

In contrast, the neural network models showed slightly reduced performance. MLPClassifier attained a precision of 96.08%, a recall of 88.72%, and an F1-score of 91.56%, while DNN achieved 98.07% precision, 88.85% recall, and 91.78% F1-score. Although both neural models demonstrated high confidence in their positive predictions (precision), their lower recall indicates they were less effective at identifying all attack types, particularly minority ones.

SVM underperformed relative to the other models, with the lowest recall at 85.69% and F1-score at 88.98%, despite maintaining a respectable precision of 94.95% and ROC-AUC of 0.9973. The recall gap indicates that SVM struggled with sensitivity, especially in detecting rare attack classes. This limitation makes it a less optimal choice in high-risk environments where undetected threats can lead to severe consequences.

5.2 Detection Performance by Class

A breakdown of per-class performance using the confusion matrices shown in Fig. 2 reveals critical insights into how each supervised learning model performs across the five categories: Normal, DoS, Probe, R2L, and U2R.



(a) Random Forest (b) XGBoost

(c) SVM (d) LightGBM

(e) MLPClassifier (f) DNN

Fig. 2. Confusion matrices of all machine learning algorithms.

Starting with XGBoost (Fig. 2b), the model demonstrated exceptional detection capabilities across all classes, including the minority categories that are often difficult to classify. It correctly predicted 198 out of 200 R2L instances and 8

out of 10 U2R samples, highlighting its sensitivity to rare intrusion patterns. Misclassifications were minimal, with very few Probe or Normal samples being confused with attack categories. This reinforces XGBoost's ability to generalize well even under significant class imbalance, thanks to its iterative boosting and error correction mechanisms. LightGBM (Fig. 2d) also matched XGBoost on the correct detection of 8 U2R cases and slightly outperformed it on R2L with 199 correct predictions. It also performed with perfect or near-perfect classification of the majority classes including Normal, DoS, and Probe. Similarly, Random Forest (Fig. 2a) performed well on minority classes, correctly identifying 195 R2L and 7 U2R samples. Although slightly lower than XGBoost, this performance is still highly competitive, especially considering the model's lower training complexity and faster inference. Random Forest showed some minor misclassifications between Probe and Normal classes, but its detection of DoS and other major classes remained near perfect.

In contrast, SVM (Fig. 2c) showed the weakest detection performance on rare attack categories. It correctly identified only 189 out of 200 R2L and 4 out of 10 U2R instances, which significantly impacted its overall recall and F1-score. The confusion matrix reveals that SVM frequently misclassifies U2R as Normal or R2L, indicating difficulty in separating the nuanced features of low-frequency attacks. Additionally, several Probe instances were incorrectly predicted as Normal, suggesting that the SVM model may have overfitted to the dominant class distribution or lacked the expressive power to capture complex decision boundaries in this dataset.

Neural network models, including MLPClassifier (Fig. 2e) and DNN (Fig. 2f), both showed strong performance on the majority classes (Normal, DoS, and Probe), with very few false positives. However, their performance declined notably for U2R and R2L. MLPClassifier correctly identified 192 R2L and 5 U2R samples, while DNN recognized 191 R2L and 5 U2R cases. These results reflect a common problem in real-world applications of deep learning-based classifiers, where some classes have significantly fewer training examples than others. This class imbalance can adversely affect both the convergence of the model during training and its ability to generalize to unseen data. Without the application of class rebalancing techniques or customized loss functions, neural networks often struggle to learn meaningful representations for underrepresented classes, leading to poor recall on rare attack types [3].

5.3 Discussion

The results of this study show that ensemble tree-based models, especially XGBoost and Random Forest, perform strongly in multi-class intrusion detection on the KDD Cup 1999 dataset. XGBoost consistently achieved the best results, with an F1-score of 97.50% and a perfect ROC-AUC of 1.000, demonstrating strong detection of both common and rare attack types. This can be attributed to the gradient boosting nature of XGBoost, where trees are added sequentially and each iteration seeks to correct the errors made by previous trees [14]. This iterative process inherently places greater emphasis on misclassified

instances, often corresponding to rare classes, allowing XGBoost to effectively detect uncommon attack types such as R2L and U2R. It was also very fast, requiring only 0.28 s to process the entire test set, making it ideal for real-time applications where both accuracy and speed are important. As shown in Fig. 3, XGBoost ranked among the most efficient models in terms of inference time. However, this performance comes with a trade-off, where XGBoost is often considered a black-box model due to its complex internal structure, which makes it difficult to interpret or explain individual predictions. In contrast, Random Forest offers a more interpretable alternative while still maintaining high performance. It achieved a solid F1-score of 95.67%, perfect precision, and completed testing in 1.75 s. While slightly slower and less sensitive than XGBoost, it provides clear insights through feature importance scores and decision paths, which can help system administrators better understand and trust the model's behavior. This balance of accuracy, speed, and interpretability makes Random Forest a strong choice for environments that require both performance and transparency.

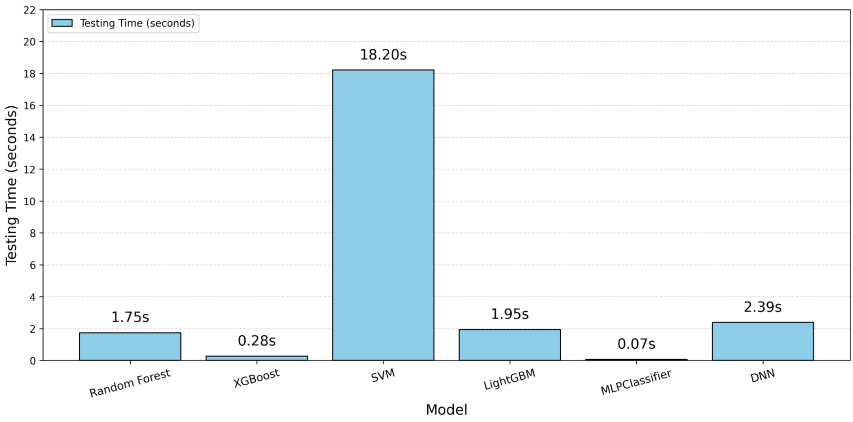


Fig. 3. Testing time across models.

LightGBM also performed well, with a recall of 95.75% and a perfect ROC-AUC of 1.000. However, its lower precision (92.89%) suggests it produced more false positives compared to XGBoost and Random Forest. Its testing time of 1.95 s, as illustrated in Fig. 3, is still fast enough for most real-time systems, making it a viable option when slightly higher false alarms are acceptable.

The MLPClassifier was the fastest model, completing testing in just 0.072 s, which is clearly visible in the bar chart. However, this speed came with a cost in detection performance. With a recall of 88.72% and an F1-score of 91.56%, it missed more attacks than the tree-based models. DNN showed slightly better detection than MLP, with a precision of 98.07% and F1-score of 91.78%, but required 2.39 s to complete inference, which places it among the slower models on the chart.

SVM had the weakest results overall. It had the lowest recall (85.69%) and F1-score (88.98%), and its testing time was the highest at 18.20 s, which stands out clearly in Fig. 3. This long delay, combined with its difficulty detecting rare attack types, makes it unsuitable for real-time intrusion detection systems.

In summary, XGBoost offers the best trade-off between accuracy and speed, making it the most practical model for real-time multi-class intrusion detection. Random Forest and LightGBM are also strong options, especially when interpretability or consistent recall is important. MLPClassifier may be useful in systems where prediction speed is the top priority, though it risks missing some attacks. DNN could be considered if a neural network-based solution is needed, but it is slower and still less accurate than the top-performing tree-based models. SVM, due to both poor detection and high latency, is not recommended for deployment in time-sensitive environments.

6 Conclusion

This study presented a comparative evaluation of six supervised learning models, which are XGBoost, Random Forest, LightGBM, SVM, MLPClassifier, and DNN, on the KDD Cup 1999 dataset for multi-class intrusion detection. The evaluation focused on each model's ability to detect both frequent and rare attack types, offering practical insights into their strengths and limitations in real-world intrusion detection systems.

The results highlight the strong performance of ensemble tree-based models, particularly boosting methods like XGBoost, which achieved the highest classification accuracy while maintaining fast inference time. Random Forest and LightGBM also showed competitive performance, with trade-offs between interpretability, recall, and speed. Meanwhile, neural models like MLPClassifier and DNN offered rapid inference but struggled more with rare attack classes, and SVM proved the least suitable due to its low recall and long testing time.

For future work, deeper exploration into hybrid and ensemble techniques, hyperparameter optimization, and validation using modern benchmark datasets such as NSL-KDD, UNSW-NB15, or CIC-IDS2017 will be valuable for assessing real-world applicability. Additionally, integrating techniques such as cost-sensitive and oversampling techniques like SMOTE (Synthetic Minority Oversampling Technique) or ADASYN (Adaptive Synthetic Sampling) could further improve the detection of underrepresented and emerging attack types [6,10,19]. These directions are essential for building more robust and adaptable intrusion detection systems in dynamic cybersecurity environments.

References

1. Ahmed, U., et al.: Signature-based intrusion detection using machine learning and deep learning approaches empowered with fuzzy clustering. Sci. Rep. **15** (2025). https://doi.org/10.1038/s41598-025-85866-7

2. Boehmke, B.: Chapter 13 – deep learning (2019). https://bradleyboehmke.github.io/HOML/deep-learning.html. Accessed 14 Apr 2025
3. Buda, M., Maki, A., Mazurowski, M.A.: A systematic study of the class imbalance problem in convolutional neural networks. Neural Netw. **106**, 249–259 (2018)
4. Cabot, J.H., Ross, E.G.: Evaluating prediction model performance. Surgery **174**(3), 723–726 (2023)
5. Chauhan, H., Kumar, V., Pundir, S., Pilli, E.S.: A comparative study of classification techniques for intrusion detection. In: 2013 International Symposium on Computational and Business Intelligence. IEEE (2013). https://doi.org/10.1109/ISCBI.2013.16
6. Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P.: SMOTE: synthetic minority over-sampling technique. J. Artif. Intell. Res. **16**(2), 321–357 (2002). https://doi.org/10.1613/jair.953
7. Eshak Magdy, M., Matter, A.M., Hussin, S., Hassan, D., Elsaid, S.: A comparative study of intrusion detection systems applied to NSL-KDD dataset. Egypt. Int. J. Eng. Sci. Technol. **43**(2), 88–98 (2023). https://doi.org/10.21608/eijest.2022.137441.1156
8. Fares, A.H., Sharawy, M.I., Zayed, H.H.: Intrusion detection: supervised machine learning. J. Comput. Sci. Eng. **5**(4), 305–313 (2011). https://doi.org/10.5626/JCSE.2011.5.4.305
9. Guido, R., Ferrisi, S., Lofaro, D., Conforti, D.: An overview on the advancements of support vector machine models in healthcare applications: a review. Information **15**(4) (2024). https://doi.org/10.3390/info15040235
10. He, H., Bai, Y., Garcia, E.A., Li, S.: ADASYN: adaptive synthetic sampling approach for imbalanced learning. In: 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence), pp. 1322–1328 (2008). https://doi.org/10.1109/IJCNN.2008.4633969
11. Hettich, S., Bay, S.D.: The UCI KDD archive. University of California, Department of Information and Computer Science, Irvine, CA (1999). http://kdd.ics.uci.edu
12. IBM: What is an Intrusion Detection System (IDS)? (2023). https://www.ibm.com/think/topics/intrusion-detection-system. Accessed 14 Apr 2025
13. IBM: What is a Cyberattack? (2024). https://www.ibm.com/think/topics/cyber-attack. Accessed 14 Apr 2025
14. Imani, M., Beikmohammadi, A., Arabnia, H.R.: Comprehensive analysis of random forest and XGBoost performance with SMOTE, ADASYN, and GNUS under varying imbalance levels. Technologies **13**(3), 88 (2025)
15. Jeyasoundari, J., Sridevi, S., Raja, S.C.: Comparison of supervised machine learning algorithms with bagged-ensemble method for intrusion detection. In: IEEE Students Conference on Engineering and Systems (SCES). IEEE (2024). https://doi.org/10.1109/10652494
16. Jha, G.: Feature scaling in machine learning: Which popular algorithms require it and which don't? (2024). https://medium.com/@post.gourang/feature-scaling-in-machine-learning-which-popular-algorithms-require-it-and-which-dont-a71f5585d664. Accessed 14 Apr 2025
17. Ke, G., et al.: LightGBM: a highly efficient gradient boosting decision tree. In: Advances in Neural Information Processing Systems **30** (2017)
18. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. arXiv preprint: arXiv:1412.6980 (2014)
19. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. In: 2017 IEEE International Conference on Computer Vision (ICCV), pp. 2999–3007 (2017). https://doi.org/10.1109/ICCV.2017.324

20. Malele, V., Mathonsi, T.E.: Testing the performance of multi-class IDS public dataset using supervised machine learning algorithms. arXiv preprint: arXiv:2302.14374 (2023). Accessed 14 Apr 2025
21. Rasane, K., Bewoor, L., Meshram, V.: A comparative analysis of intrusion detection techniques: machine learning approach. In: Proceedings of International Conference on Communication and Information Processing (ICCIP). Elsevier (2019). https://doi.org/10.2139/ssrn.3418748
22. Scikit-learn Developers: StandardScaler — scikit-learn 1.6.1 documentation (2013). https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html. Accessed 14 Apr 2025
23. Scikit-learn Developers: MLPClassifier — scikit-learn 1.6.1 documentation (2015). https://scikit-learn.org/stable/modules/generated/sklearn.neural_network.MLPClassifier.html. Accessed 14 Apr 2025
24. Serinelli, B.M., Collen, A., Nijdam, N.A.: Training guidance with KDD cup 1999 and NSL-KDD data sets of ANIDINR: anomaly-based network intrusion detection system. In: Procedia Computer Science, vol. 175, pp. 560–565 (2020). https://doi.org/10.1016/j.procs.2020.07.080
25. Shokeen, A., Yadav, N., Sisaudia, V.: Performance analysis of different machine learning algorithms for intrusion detection on KDD-CUP-99 dataset. In: AIP Conference Proceedings, vol. 3072. AIP Publishing (2024). https://doi.org/10.1063/5.0203394
26. Sowmya, T., Mary Anita, E.A.: A comprehensive review of AI-based intrusion detection systems. Meas.: Sens. **28** (2023). https://doi.org/10.1016/j.measen.2023.100827
27. Sruthi: understanding random forest algorithm with examples (2025). https://www.analyticsvidhya.com/blog/2021/06/understanding-random-forest/. Accessed 14 Apr 2025
28. Swaminathan, S., Tantri, B.R.: Confusion matrix-based performance evaluation metrics. Afr. J. Biomed. Res. **27**, 4023–4031 (2024)
29. Talaei Khoei, T., Kaabouch, N.: A comparative analysis of supervised and unsupervised models for detecting attacks on the intrusion detection systems. Information **14**(2), 103 (2023). https://doi.org/10.3390/info14020103
30. Talukder, M.A., et al.: Machine learning-based network intrusion detection for big and imbalanced data using oversampling, stacking feature embedding and feature extraction. J. Big Data **11**(33) (2024). https://doi.org/10.1186/s40537-024-00886-w
31. Tavallaee, M., Bagheri, E., Lu, W., Ghorbani, A.A.: A detailed analysis of the KDD CUP 99 data set. In: Proceedings of the 2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications, pp. 1–6 (2009). https://doi.org/10.1109/CISDA.2009.5356528
32. Yang, E., Joshi, G.P., Seo, C.: Improving the detection rate of rarely appearing intrusions in network-based intrusion detection systems. Comput., Mater. Continua **66**(2), 1661–1676 (2020). https://doi.org/10.32604/cmc.2020.013210
33. Zhang, P., Jia, Y., Shang, Y.: Research and application of XGBoost in imbalanced data. Int. J. Distrib. Sens. Netw. **18**(6) (2022). https://doi.org/10.1177/15501329221106935

A Novel Framework on Cardiovascular Disease Prediction Using Transfer Learning Technique

Manisha Guduri[1(✉)], Ashok Polavarapu[2], Mohammad El Yabroudi[1], Sandeep Kumar Thota[3], Hari Suresh Babu Gummadi[4], and Narendra Chennupati[5]

[1] Lawrence Technological University, Southfield, MI 48075, USA
manishaguduri@ieee.org, melyabrou@ltu.edu
[2] University of Louisiana at Lafayette, Lafayette, LA 70503, USA
ashok.polavarapu1@louisiana.edu
[3] Digital Assistant Cloud, Oracle Inc., Austin, USA
sandeepthota@ieee.org
[4] Frontier Communications, Dallas, USA
hari.gummadi@ftr.com
[5] Hindsight Software Solutions Inc., Frisco, USA
nchennupati@hindsight.team

Abstract. Cardiovascular disease is the leading cause of death worldwide. There is a need for advanced, precise, and scalable diagnostic procedures to identify and address it promptly. This study uses clinical and demographic data from the Cleveland Heart Disease dataset to develop a new deep learning method for predicting heart disease. The approach facilitates transfer learning and tabular transformer architectures, such as SAINT and FT-Transformer. Our approach distinguishes itself from conventional machine learning methodologies by utilizing pretrained representations and attention mechanisms to identify intricate relationships between samples and attributes. Traditional machine learning algorithms depend on a limited set of fabricated features and interactions. Our systematic methodology incorporates principal component analysis (PCA) for dimensionality reduction, domain-specific feature engineering, modifications to the transformer layer for binary classification, and improved data preparation. Using a fixed base encoder significantly reduces training time and improves generalization when learning a specific classification head. Extensive testing revealed that the SAINT-based model outperformed random forests, logistic regression, and CNN-MLP models. The study's findings suggest that tabular converters can accurately interpret structured clinical data and that transfer learning is a beneficial approach for medical predictive analytics. The proposed method is a versatile and scalable solution for various diagnostic applications in healthcare AI, facilitating the prediction of heart disease.

Keywords: Cardiovascular disease · Cleveland Dataset · Transfer Learning · Feature Engineering

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026
S. Lanka et al. (Eds.): ICTSM 2025, LNNS 1755, pp. 81–93, 2026.
https://doi.org/10.1007/978-3-032-13177-5_7

1 Introduction

According to the World Health Organization (WHO), 7.9 million people die from cardiovascular disease every year, making it the world's leading cause of death [1]. As more people live longer, develop diabetes and high blood pressure, and lead sedentary lifestyles, it is expected that the world's health problems will worsen. Early and accurate detection of heart problems can save lives and improve quality of life by facilitating the creation of personalised treatment plans and enabling prompt action. To diagnose heart problems, doctors use electrocardiograms (ECGs), angiograms, and echocardiograms [2]. However, these treatments can be costly, painful and difficult to administer as they require specialised equipment and clinical studies.

Machine learning [3] and deep learning [4] are now required to create prediction models that can operate independently [5]. These models help doctors make informed decisions about patient care. In recent years, there has been a significant increase in the use of data-driven methods for medical diagnosis, particularly for predicting cardiovascular disease (CVD). Numerous studies have demonstrated that machine learning techniques such as logistic regression, decision trees, and support vector machines (SVMs) can identify significant risk factors and predict when a patient will become ill [6]. Conversely, standard models are not as effective at capturing complex nonlinear interactions and high-dimensional clinical qualities because they rely on features and frameworks that have been manually created. They also perform well with data sets containing a combination of continuous, categorical and numerical factors, such as those found in electronic health records (EHRs). Researchers have considered convolutional neural networks (CNNs) and multilayer perceptrons (MLPs) as possible solutions to these problems [7]. However, to start from scratch with deep models in medicine, a significant amount of labeled data would be required. However, this may not be feasible due to privacy concerns and the high costs associated with labelling and collecting data.

Transfer learning is increasingly recognised as an important application of AI, particularly in areas with limited data [8]. Transfer learning improves a model after it has been trained using a large dataset. This technology is primarily used for processing natural language and images, but it is only just beginning to gain popularity in the healthcare sector for processing tabular data. This page lists the FT-Transformer and the SAINT (Self-Attention and Intersample Attention Transformer) as two transformers that are effective with tabular data. They can handle various types of structured data. These models do not assume that all models are the same or that traits are different. Instead, they use the self-attention method to demonstrate the relationships between traits and samples. These models were trained using large tabular datasets with limited data to identify trends that could be applied to various situations, such as predicting heart disease.

To develop a transfer learning system for predicting heart disease, we utilize the well-known Cleveland Heart Disease dataset, which contains clinical and demographic data. To identify cases of heart disease in this dataset, we

use pre-trained tabular models such as SAINT and FT-Transformer. This is achieved by adding a classification head and disabling the initial components of the transformer encoder. The Cleveland dataset encompasses a diverse range of continuous and categorical characteristics. These include resting blood pressure, age, gender, cholesterol levels and various types of chest pain. Researchers can use these characteristics to identify significant indicators of an increased risk of heart disease and accurately predict future events. We have carefully selected components to create location-specific composite features, similar to the Heart Risk Index and Stress Response Score. The process of reducing the number of dimensions is made easier by principal component analysis (PCA). Combining early techniques and transfer learning may accelerate the process of creating a usable and dependable model that can identify minute patterns and relationships in vast volumes of data.

The proposed strategy addresses several significant issues that arise during the prediction of cardiac disease. It eliminates the need for complex feature engineering and manual feature selection by using attention-based models that have already been trained. Reprocessing and altering trends can improve the accuracy and reliability of forecasts in cases of insufficient data. A third notable feature is its adaptability, enabling integration with other techniques or systems that support real-time decision-making. Transformer-based models may find widespread application in organising medical data.

2 Related Work

The authors [9] developed methods to explore ways to apply machine learning knowledge to a diverse range of datasets, aiming to enhance CAD treatment. This makes it easier to obtain data that is free of errors, missing information, and other issues. We compared our trees with an enhanced C4.5 tree and a rapid decision tree created from various types of data. Researchers can examine the same situation again by identifying traits that are similar in any dataset. The average accuracy of classifications from separate datasets was 75.48%, compared to 78.06% for mergeddatasets.

This study [10] examines several potential applications of artificial intelligence in the detection of heart problems. Six well-known methods were tested using the Cleveland dataset. Both computer intelligence and expert opinion were used to identify the relevant characteristics. People often use artificial intelligence. The report states that splitting Cleveland into two regions makes it more difficult to obtain data. It examines how well algorithms perform and how effectively MFS handles unusual data.

This work [11] uses the Cleveland dataset and cutting-edge machine learning methods to predict and prevent heart disease. The dataset was carefully compiled by removing outliers, merging traits and resolving issues. Dataset was examined to identify links between various factors that can lead to heart disease. With accuracy rates above 95%, XGBoost is shown to be more precise than other machine learning methods. The model can distinguish between good and bad

outcomes because it has a large area under the curve of 0.99, as determined by the ROC curve analysis. The results of the study suggest that machine learning-based methods may be useful. Based on these findings, it appears that a greater reliance on data-driven methods is warranted to improve the prediction and prevention of heart disease in the future.

The author [12] used the 303-entry Cleveland Heart Disease dataset to determine the prevalence and onset of heart disease. However, six of the forms contained missing information. It is likely that the published study will only use 13 of the 76 features that were in the original dataset. The next few paragraphs explain what the illness involves. For each case, a 'class label' variable is created. This variable contains data from 303 individuals and 55 traits. There are various ways to evaluate the performance of an algorithm, including the F-measure, classification accuracy, precision, and sensitivity. The author applied nine machine learning models to the final dataset, both before and after adjusting the hyperparameters. They performed the necessary pre-processing, changed the hyperparameters and normalised the dataset to ensure the results for the common heart disease dataset were accurate. The author devised the K-fold cross-validation method to facilitate the training and testing of machine learningalgorithms.

The author [13] developed a more accurate method for predicting who is likely to develop heart disease. The main goals are to reduce the number of heart attacks, swiftly identify new cases, accelerate the diagnostic process and save lives. A great deal can be learnt about heart disease from the Cleveland database, the Heart Disease database, and the National Cardiovascular Disease Surveillance (NCDS) System. The Cleveland heart disease information was used for this study. This consisted of four files from Long Beach, Switzerland, Hungary, and Virginia. The dataset contains 14 characteristics, each with its unique identifier. This collection contains records for 1,025 patients. Of these, 713 are for men and 312 are for women. Each method uses 75% of the training dataset and half of the testing dataset.

Authors [14] aim to evaluate the effectiveness of PSO and other supervised machine learning methods in teaching MLP neural network classifiers. The author created binary categorisation systems to help doctors diagnose heart conditions. Although it is complicated, the MLP function demonstrates a fully self-coupled MLP network. It takes in numbers and sends them out. It looks for patterns in the raw data coming in from the area. The output layer is responsible for making predictions. It is crucial to select the appropriate hyperparameters for the MLP, including the number of neurons and hidden layers. Python was used for all the background work on the project. Seventy per cent of the data was used to train the model, and only 30 per cent was used for testing. The author performed five cross-validations to train the suggested machine learning models. Fivefold cross-validation is a straightforward method often employed by machine learning experts to test and identify the optimal model for a specific predictive modeling task.

This study [15] uses Explainable Artificial Intelligence (XAI) to improve the accuracy and clarity of machine learning models that predict heart disease. This

utilizes various machine learning techniques, including decision trees, support vector machines, and deep learning, to predict the progression of heart disease. We utilize XAI tools to understand the model's output thoroughly. These include SHapley Additive Explanations (SHAP), Local Interpretable Model-agnostic Explanations (LIME), and feature value analysis. This work suggests that explainable models can empower doctors to understand the inner workings of the model better, fostering trust and potentially leading to more informed medical decisions by shedding light on critical risk factors. This project demonstrates the importance of connecting machine learning with the ability to interpret results. Artificial intelligence could potentially make medical reviews more precise and more accurate.

This study [16] examines various datasets and research papers to assess the effectiveness of different machine learning methods in predicting heart disease. It applied various methods, including Random Forest, Decision Tree, Support Vector Machine, Logistic Regression, Naive Bayes, and Gradient Boosting, to different situations and datasets. The Random Forest and Decision Tree methods performed exceptionally well with the Cleveland dataset. The Random Forest method achieved the best results with the Framingham group, with a 94% success rate. The decision tree method achieved an 85.23% score and came in third, while the SVM strategy achieved an 87.45% score and came in second. Generally, Random Forests and other group methods outperform single algorithms. However, this may not always be the case, depending on the information and settings used. Authors [17] used optimisation techniques to help machine learning models identify cardiovascular disease and assess its severity. By combining ML with medical data, doctors may be able to predict patients' life expectancy and heart health. The Cleveland Dataset was used to create and test a number of machine learning models. It is available for free download from the UCI Repository. Some of the models that were used included Naive Bayes, Logistic Regression, Gradient Boosting, Support Vector Machines, Decision Trees, Random Forests, and Ensemble methods. With the help of medical and visual data, artificial intelligence can make informed predictions. This makes it an excellent way to predict what kind of patient someone is. Before comparison, the models were accurate 7882% of the time. This model is much more accurate than our previous one, achieving over 90% accuracy.

3 Methodology

The preprocessing phase is crucial as it ensures the accuracy and consistency of the data utilized for predicting cardiac disease. We employ median imputation to address missing data for continuous variables such as serum cholesterol and resting blood pressure. Mode imputation is subsequently employed for categorical variables, including the type of chest discomfort and thalassemia. This ensures data accuracy and eliminates prejudice. Z-score filtering eliminates values that deviate from the mean by more than ± 3 standard deviations, helping to identify and eliminate outliers in continuous data and reduce distortion and noise.

The incorporation of embedding layers for high-cardinality variables and one-hot encoding for low-cardinality variables enhances the model's ability to represent categories such as "sex," "chest pain type (cp)," and "slope of the ST segment." StandardScaler is used to normalize oldpeak, cholesterol, and age to a uniform scale, aiding the gradient descent process. The Stress Response Score comprises Exang (exercise-induced angina) and oldpeak (ST segment depression), reflecting that cardiac stress is associated with several minor yet significant factors. To compute the heart risk index, we multiply age and cholesterol by thalach, defined as the maximum heart rate plus one. Principal Component Analysis (PCA) is employed to eliminate multicollinearity while preserving the principal components that account for 95% of the variance, resulting in a compact and operational feature space. This meticulously designed preprocessing technique enables robust and dependable forecasts regarding heart disease, ensuring that the data provided is prepared for deep model adaptation and transfer learning.

3.1 Feature Engineering

Feature engineering is essential for models that use structured clinical data, such as the Cleveland Heart Disease dataset. Researchers used a multi-stage feature engineering technique to reduce noise, simplify the data, and identify patterns that might indicate cardiac disease. The format was modified to ensure the essential components of the domain were positioned correctly. They altered the format and minimized the dimensions to enhance the functionality of transfer learning models.

The initial phase involved the emergence of specific derived features. Cardiologists often examine correlations among diverse attributes. For example, they examine the impact of cholesterol on heart rate and how seasonal affective disorder affects individuals' ability to exercise. We created two additional derived features from these clinical data. The Heart Chance Index (HCI) quantifies the extent to which elevated cholesterol levels and aging increase the risk of myocardial infarction. This index is calculated by multiplying the maximum heart rate by one and then dividing the result by the product of cholesterol and age. Oldpeak periods produce the Stress Response Score (SRS). This score demonstrates how the heart reacts to physical exertion. Significant composite signals that were previously difficult to identify in the raw data were supplied to the learning model through the newly incorporated characteristics in its input.

We used Principal Component Analysis (PCA) to reduce the complexity and multicollinearity of the data. One-hot encoding was essential for augmenting the feature space of categorical data. PCA successfully reduced the number of features while preserving 95% of the dataset's variance by maintaining the principal orthogonal components. This process improves memory retention and facilitates knowledge acquisition. Our goal was to include components directly associated with illness classification.

We modified the dataset's structure to enhance its compatibility with deep learning architectures and transformer-based transfer learning models. This transformation enables efficient sequence processing. By condensing numerical

and categorical features, we reduced the size of the embedding space, thereby making the interconnections among the space's attributes clearer. The formula used to determine the necessary number of dimensions for categorical features with a high variable count, such as `thal` and `cp`, is:

$$\text{embedding_dim} = \min(50,\ \lfloor \text{feature_cardinality} / 2 \rfloor). \tag{1}$$

Unlike one-hot encoding, these embeddings significantly improved the model's ability to represent each category.

3.2 Transfer Learning

Vaswani et al.[18] introduced the first Transformer encoder architecture, which forms the foundation of both *FT-Transformer* and *SAINT*. These models adapt the Transformer framework for tabular data by embedding each feature into a vector space and applying self-attention mechanisms to analyze the input effectively.

Let $\mathbf{x} \in \mathbb{R}^d$ represent a single input sample corresponding to a patient record with d attributes. These attributes consist of continuous features ($\mathbf{x}_{\text{cont}}$) and categorical features ($\mathbf{x}_{\text{cat}}$). For each categorical feature $x_i \in \mathbf{x}_{\text{cat}}$, an embedding function E_i maps it to a vector representation:

$$e_i = E_i(x_i) \tag{2}$$

Continuous attributes are processed via linear projection. For each continuous feature $x_j \in \mathbf{x}_{\text{cont}}$, we compute:

$$e_j = W_j x_j + b_j \tag{3}$$

By merging the resulting embeddings, we construct the input token matrix $T \in \mathbb{R}^{n \times d_e}$, where n denotes the number of features and d_e the embedding dimension:

$$T = [e_1;\ e_2;\ \ldots;\ e_n] \tag{4}$$

To preserve the structural relationships in the tabular input, positional encodings or learned feature-specific embeddings are added to these tokens.

The Self-Attention Layer is a critical component of Transformer architectures. It allows the model to identify relationships among different data sources and provide suitable representations. Self-attention allows the model to identify correlations among features in tabular data across multiple dimensions by integrating feature embeddings with different weights. These attentional measurements, or weights, originate from previously established connections. Normalizing the query (Q) and key (K) vectors before computing the attention score ensures uniformity in their dimensions. One can modify these weights using softmax to obtain different outcomes. Incorporating these weights into the value vectors yields the outcome. This technique ensures that each feature considers every other feature. This enables the model to focus on the most critical inputs

for making a specific prediction. Self-attention improves context comprehension in feedforward networks, allowing them to perform effectively with structured data, such as medical records, which are characterized by numerous feature interactions. The multi-head self-attention mechanism processes these embeddings to model inter-feature dependencies:

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^\top}{\sqrt{d_k}}\right)\mathbf{V} \tag{5}$$

where:

$$\begin{aligned}\mathbf{Q} &= \mathbf{T}\mathbf{W}_Q,\\ \mathbf{K} &= \mathbf{T}\mathbf{W}_K,\\ \mathbf{V} &= \mathbf{T}\mathbf{W}_V,\end{aligned}$$

are the query, key, and value projections of the token matrix $\mathbf{T}$; and

$$\mathbf{W}_Q,\ \mathbf{W}_K,\ \mathbf{W}_V \in \mathbb{R}^{d_e \times d_k} \tag{6}$$

are the learned weight matrices. Here, d_k is the dimensionality of the key vectors.

The output of multi-head attention is then passed through a position-wise feedforward network and followed by residual connection and layer normalization to produce deep contextualized representations:

$$\mathbf{Z} = \text{LayerNorm}(\text{FFN}(\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V})) + \mathbf{T}) \tag{7}$$

We use a pretrained model that employs transfer learning to generate a dataset distribution similar to the original, thereby enhancing generalization. These models typically extract informative patterns from large datasets containing structured variables. The primary objective is to enhance the model's upper layers' capacity to detect cardiac abnormalities while preserving the essential abstract representations acquired during pre-training. The initial three Transformer blocks, denoted as f_{θ_f}, are responsible for capturing general-purpose features. The target dataset is then used exclusively to fine-tune the upper layers and the classification head, denoted as f_{θ_c}. For every input sample $\mathbf{x} \in \mathbb{R}^d$, the frozen encoder generates latent representations $\mathbf{z} = f_{\theta_f}(\mathbf{x})$, which are subsequently passed to the trainable component to produce the prediction $\hat{y} = f_{\theta_c}(\mathbf{z})$. This hierarchical division reduces the risk of catastrophic forgetting of previously learned generic representations and accelerates convergence during training. The custom classification head interprets the Transformer's output representation $\mathbf{z}$ and produces a probability score $\hat{y} \in [0, 1]$ that reflects the likelihood of heart disease. The classification head is composed of three fully connected layers, interleaved with non-linear activation functions and regularization techniques to improve the model's robustness and stratification capability.

First, the output $\mathbf{z} \in \mathbb{R}^{d_z}$ is transformed through a dense layer with ReLU activation:

$$\mathbf{h}_1 = \text{ReLU}(\mathbf{W}_1\mathbf{z} + \mathbf{b}_1) \tag{8}$$

Then, dropout is applied to introduce stochasticity and reduce overfitting:

$$\mathbf{h}_2 = \text{Dropout}(\mathbf{h}_1,\ p = 0.3) \tag{9}$$

Next, another dense transformation follows:

$$\mathbf{h}_3 = \text{ReLU}(\mathbf{W}_2\mathbf{h}_2 + \mathbf{b}_2) \tag{10}$$

Finally, the output layer projects the feature vector to a scalar using the sigmoid function:

$$\hat{y} = \sigma(\mathbf{W}_3\mathbf{h}_3 + \mathbf{b}_3) \tag{11}$$

Here, $\mathbf{W}_1$, $\mathbf{W}_2$, $\mathbf{W}_3$ and $\mathbf{b}_1$, $\mathbf{b}_2$, $\mathbf{b}_3$ are trainable weights and biases of the classification head, and the sigmoid function $\sigma(x) = \frac{1}{1+e^{-x}}$ maps the result into a valid probability. This architecture enables the model to capture non-linear interactions between latent features while remaining lightweight and interpretable.

We employ early stopping to terminate training on a constrained dataset such as Cleveland if the validation loss fails to improve after a specified number of epochs (e.g., *patience* = 10). This prevents excessive model complexity and overfitting. Throughout training, we monitor the optimal model weights (θ_c^*) that yield the minimum validation loss. These weights are subsequently employed for the conclusive model evaluation. This fine-tuning approach, which integrates task-specific learning with knowledge transfer, is crucial for achieving high generalization accuracy in medical prediction tasks.

4 Results and Discussion

The Cleveland Heart Disease Dataset has been referenced in numerous studies aimed at predicting heart disease progression. The dataset originates from the Cleveland Clinic Foundation. It is available at the University of California, Irvine (UCI) Machine Learning Repository as part of a larger collection. Each patient record in the collection contains thirteen clinical criteria and a target label indicating the presence or absence of heart disease. The collection contains 303 patient records. These factors encompass physiological metrics, such as resting blood pressure and serum cholesterol; medical test results; and demographic variables, including age and gender. The dataset uses numerical values to represent categorical attributes, such as thalassemia (thal), fasting blood sugar (FBS), and types of chest pain (CP). This makes it suitable for statistical analysis and machine learning. In numerous predictive modeling studies, the original multiclass indicator variable (04), which signifies disease severity, is transformed into a binary variable with values of 0 (absence of cardiac disease) or 1 (presence of disease). Despite its limited size, the Cleveland dataset is significant due to its diverse range of pertinent variables, balanced class distribution, and clinical relevance. It is the optimal method for evaluating medical decision support systems and classification algorithms. However, integrating continuous, ordinal, and categorical diagnostic data poses significant challenges for hospitals.

5 Conclusion and Future Work

This study establishes a reliable and practical forecasting methodology for heart disease using sophisticated transformer architectures tailored for tabular data via transfer learning. The inherent limitations of standard machine learning techniques, such as insufficient feature representation and superficial modeling, were mitigated by using the Cleveland Heart Disease dataset to enhance pretrained models, including FT-Transformer and SAINT. Our analysis shows that the original encoder layers remain unchanged when training a lightweight classification head. This accelerates the model's convergence and enhances the applicability of the pre-trained data in various situations. This results in a significantly enhanced level of diagnostic precision. Empirical findings indicate that the SAINT-based model's superior accuracy and F1-score make it suitable for clinical applications. Integrating domain-specific feature engineering with meticulous data preparation enhances the model's adaptability and robustness. This adjustment accounts for variations in patient demographics and data distribution. This study illustrates how transfer learning could transform the healthcare sector, particularly in predicting outcomes with limited data. The proposed framework can enhance the accuracy of precision medicine diagnoses by integrating real-time health monitoring systems and utilizing multimodal datasets, including imaging, sensor, and electronic health record data.

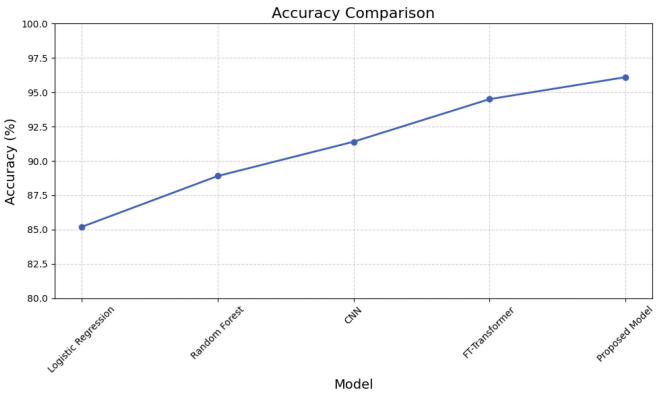


Fig. 1. Performance Analysis of Accuracy.

The proposed model achieves 96.1% accuracy rating shown in Fig. 1. The FT-Transformer ranks second with an accuracy rate of 94.5%. Two well-known machine learning models, Random Forest and Logistic Regression, have accuracy rates of 85.2% and 88.9%, respectively. This suggests that they are unable to discern intricate patterns within the material. The CNN model outperformed the others due to its greater complexity. It achieved an accuracy rating of 91.4%. Transformer-based models, particularly the proposed SAINT-based

model, exceeded these baselines. These results suggest that self-attention mechanisms and transfer learning are effective for modeling organized medical data.

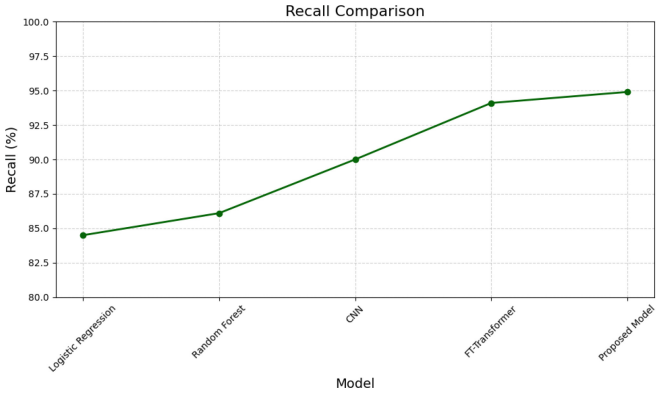


Fig. 2. Performance Analysis of Recall.

The Fig. 2 shows that the proposed model recall achieves the 94.9%. This shows that it is effective at identifying people with heart problems. The FT-Transformer ranked second, achieving a memory rate of 94.1%. Furthermore, it functions as intended. The CNN model achieved an excellent accuracy rate of 90.0%. With corresponding scores of 86.1% and 84.5%, Random Forest and Logistic Regression performed poorly by comparison. According to this study, transformer-based systems—particularly the proposed model—are far more effective at preventing false negatives and reliably identifying patients with heart problems in clinical settings.

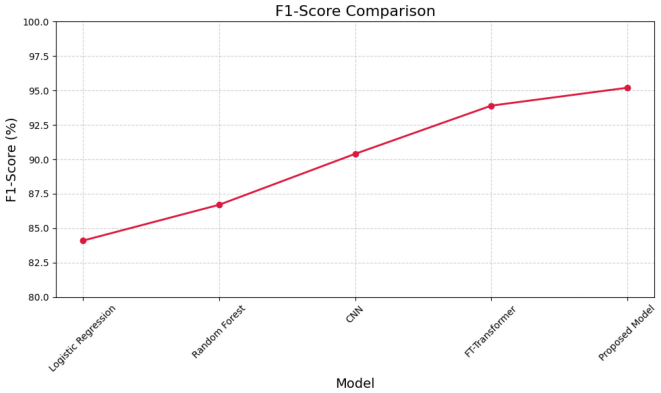


Fig. 3. Performance Analysis of F1 Score.

The proposed model achieves the 95.2% F1-score, which is greater than all other state-of-the-art methods shown in Fig. 3. This improves the balance between accuracy and memory, making these events easier. Although CNN's approach works, it is not better than that of its competitors. It achieves an F1 score of 93.9%. The CNN model achieved a score of 90.4%. The Random Forest model was successful 86.7% of F1-score, whereas the Logistic Regression model have F1-score as 84.1%. These outdated models cannot explain closely related events, so they are no longer helpful. Our study demonstrates that our transfer learning-based model yields the most accurate and reliable predictions regarding heart disease.

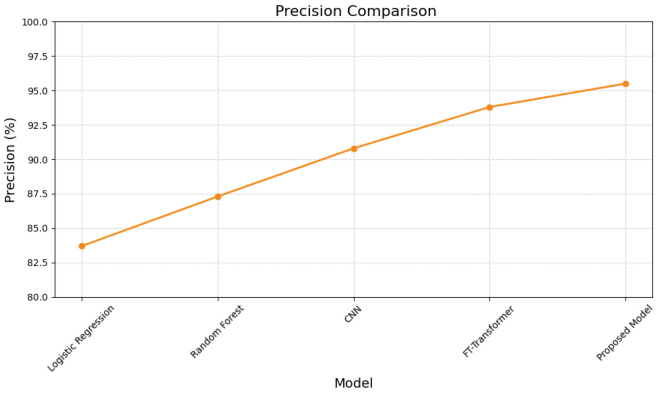


Fig. 4. Performance Analysis of Precision

In Fig. 4, the proposed model is the most effective one, with a precision rate of 95.5%. This proves that it can reliably detect cases of heart disease while reducing false positives. With a score of 93.8%, the FT-Transformer models are effective. The CNN model performed with precision rate of 90.8%. Precision of 87.3% and 83.7% were achieved by the standard models of Random Forest and Logistic Regression, respectively. To avoid incorrect assessments and unnecessary treatments, clinical evaluations must adhere to the recommended protocol, and the data indicate that this is the case.

References

1. World Health Organization, "Cardiovascular diseases (2024)." https://www.who.int/health-topics/cardiovascular-diseases. Accessed 01 June 2025
2. Ponciano, V., et al.: Machine learning techniques with ECG and EEG data: an exploratory study. Computers **9**(3), 55 (2020)
3. Ahsan, M.M., Siddique, Z.: Machine learning-based heart disease diagnosis: a systematic literature review. Artif. Intell. Med. **128**, 102289 (2022)
4. Zhou, C., et al.: A comprehensive review of deep learning-based models for heart disease prediction. Artif. Intell. Rev. **57**(10), 263 (2024)

5. Mamlook, R.A., et al.: Optimizing machine learning models based on hyperparameter tuning for Arabic fake news detection. In: 2025 9th International Symposium on Innovative Approaches in Smart Technologies (ISAS), vol. 57, no. 10, pp. 1–9 (2025)
6. Guduri, M., Chakraborty, C., Margala, M.: Smart Devices for Medical 4.0 Technologies. CRC Press (2025)
7. Chen, X., et al.: Kalman-based adaptive moment estimation optimisation algorithm to enhance GPT in LLMS for medical sentiment analysis of patient health-related feedback. IEEE J. Biomed. Health Inf., 1–11 (2025)
8. Kumar, G.S., Kumaresan, P.: Deep learning and transfer learning in cardiology: a review of cardiovascular disease prediction models. IEEE Access (2024)
9. Pradhan, J.D., Prasad, L.N., Dash, T.K., Guduri, M., Panda, G.: Cascaded PFLANN model for intelligent health informatics in detection of respiratory diseases from speech using bio-inspired computation. J. Artif. Intell. Technol. **4**(2), 124–131 (2024)
10. Nahar, J., Imam, T., Tickle, K.S., Chen, Y.-P.P.: Computational intelligence for heart disease diagnosis: a medical knowledge driven approach. Expert Syst. Appl. **40**(1), 96–104 (2013)
11. Govindaraju, K., Kalimuthu, G.: Heart disease analysis and prediction with machine learning techniques using Cleveland dataset. In: International Symposium on Intelligent Computing Systems, pp. 20–43. Springer (2024)
12. Saboor, A., Usman, M., Ali, S., Samad, A., Abrar, M.F., Ullah, N.: A method for improving prediction of human heart disease using machine learning algorithms. Mob. Inf. Syst. **2022**(1), 1410169 (2022)
13. Ahmad, G.N., Fatima, H., Abbas, M., Rahman, O., Alqahtani, M.S.: Mixed machine learning approach for efficient prediction of human heart disease by identifying the numerical and categorical features. Appl. Sci. **12**(15), 7449 (2022)
14. Al Bataineh, A., Manacek, S.: MLP-PSO hybrid algorithm for heart disease prediction. J. Pers. Med. **12**(8), 1208 (2022)
15. Paul, J.: Analyzing the role of explainable ai in heart disease diagnosis using machine learning models and the Cleveland dataset (2024)
16. Roopini, J., Deepa, B., Pooja, N.: Assessing and exploring machine learning techniques for cardiovascular disease prediction using Cleveland and Framingham datasets. In: 2024 Second International Conference on Networks, Multimedia and Information Technology (NMITCON), pp. 1–9. IEEE (2024)
17. Sareen, S., Prakhar, Kavisankar, L.: Heart health status detection using ensemble learning with hyperparameter optimization. In: AIP Conference Proceedings, vol. 3075, p. 020189. AIP Publishing LLC (2024)
18. Vaswani, A., et al.: Attention is all you need. In: Advances in Neural Information Processing Systems, vol. 30 (2017)



Development and Deployment of a Machine Learning-Based Threat Detection System to Enhance Cybersecurity at State University

Glorybe E. Alegre[1(✉)] and Patrick D. Cerna[2]

[1] College of Computer Studies, Central Philippines State University, Kabankalan City, Philippines
glorybealegre@cpsu.edu.ph

[2] College of Information and Communication Technology and Engineering, State University of Northern Negros, Sagay City, Philippines
ptcerna@sunn.edu.ph

Abstract. The safety of school networks has become increasingly critical in today's digitally connected world due to the growing frequency and sophistication of cyber threats such as malware, viruses, and unauthorized intrusions. Educational institutions are particularly vulnerable because of their expanding reliance on digital platforms for administrative functions, online learning, and internal communications. In many universities, a significant challenge lies in the absence of efficient mechanisms to detect and respond to cyberattacks in their early stages often identifying breaches only after substantial damage has occurred. This research introduces a unified, tri-layered framework that combines real-time threat detection using the Random Forest algorithm, intelligent alerting for immediate incident response, and content filtering for policy enforcement specifically tailored to the operational and behavioral dynamics of educational networks. By utilizing real-time traffic data from the university's own infrastructure, the system is trained to detect a wide range of threat patterns while reducing overfitting, a common challenge in cybersecurity applications. The framework represents a shift from traditional reactive models to a proactive, predictive security paradigm, enhancing the capability of Management Information Systems (MIS) teams to preemptively address cyber risks. This integration of technical robustness, contextual relevance, and administrative control renders the proposed system both practically effective and strategically significant. The outcomes of this study have important implications for the field of educational cybersecurity, offering a scalable and adaptable model that can serve as a foundation for future research and implementation across academic institutions.

Keywords: Cyberattacks · Threat detection · Cyber Security · Random Forest Algorithm · School Network

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026
S. Lanka et al. (Eds.): ICTSM 2025, LNNS 1755, pp. 94–109, 2026.
https://doi.org/10.1007/978-3-032-13177-5_8

1 Introduction

In today's digital world, protecting school networks has become increasingly critical due to the growing threats of malware, viruses, and illegal intrusions. Schools are particularly vulnerable as they rely heavily on technology for communication, online learning, and administrative tasks. Cyberattacks not only compromise sensitive data like student and staff records but also disrupt daily operations, incur high costs, and damage institutional reputation. A major challenge is the lack of proactive monitoring, as schools often rely on reactive strategies that detect breaches only after they occur, leaving critical systems exposed and limiting the ability to prevent intrusions in advance. This study aims to address this gap by (1) developing a Network Threat Detection System (NTDS) that leverages artificial intelligence and machine learning to identify, categorize, and respond to potential cyber threats in real-time, (2) designing and implementing an intelligent notification system that alerts IT personnel immediately upon detection of suspicious or malicious network activities, enabling rapid response and risk mitigation, and (3) integrating AI-driven content filtering mechanisms that restrict access to non-educational or harmful websites in alignment with institutional internet usage policies [1].

The study introduces a comprehensive cybersecurity framework for school networks by developing a Network Threat Detection System (NTDS) that employs artificial intelligence and machine learning algorithms to intelligently identify, categorize, and respond to cyber threats in real time. This proactive approach significantly minimizes the risks associated with malware, viruses, and unauthorized intrusions. It also has an intelligent notification mechanism that automatically alerts IT personnel when suspicious or malicious activities are detected, thereby enabling immediate intervention and effective risk mitigation. Furthermore, the study integrates AI-driven content filtering to restrict access to harmful or non-educational websites. Implementing the proposed system offers significant economic and operational advantages for the university. The institution can reduce the recurring costs associated with annual subscriptions to commercial cybersecurity services by deploying an in-house, AI-driven cybersecurity framework. This cost efficiency becomes particularly valuable for universities with multiple external campuses, where the cumulative expenses of third-party services are considerably higher. The novelty of this study lies in developing a proactive, AI-powered cybersecurity framework tailored explicitly for school networks, addressing gaps that previous works on general-purpose intrusion detection or enterprise-level systems have overlooked. Unlike earlier studies that relied primarily on reactive detection methods, the proposed framework integrates real-time threat detection using machine learning, an intelligent notification mechanism for immediate IT response, and AI-driven content filtering aligned with institutional policies into a unified system.

2 Related Literature

Network threat detection systems are important parts of cybersecurity strategies regarding school IT infrastructure because educational environments are inherently complex and dynamic. These organizations are more susceptible to cyberattacks because they usually have many connected devices, a high user turnover rate, and a significant reliance

on wired and wireless networks. Modern machine learning (ML)-based intrusion detection systems (IDS) provide a promising answer by skillfully striking a balance between detection accuracy and the reduction of false positives [2]. Building on previous anomaly-based detection models, their work shows how intelligent systems can provide proactive threat identification and adjust to changing network behaviors. These capabilities are critical in academic settings where user behavior is unpredictable, and security breaches can disrupt academic continuity and administrative operations. The efficacy of anomaly-based intrusion detection systems is further supported by [3], especially in educational settings where network behaviors are dynamic and unexpected. To guarantee that detection models remain sensitive to the distinct traffic patterns in school networks, they stress the significance of utilizing customized datasets and real-time evaluation measures. From an additional perspective, [4] emphasizes the importance of feature selection methods and hybrid machine learning strategies, pointing out that they can significantly improve detection performance in dynamic threat environments frequently found in educational institutions. To enhance the flexibility, scalability, and precision of intrusion detection systems in real-time educational settings, [5] investigates the use of sophisticated deep learning (DL) algorithms, such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs). In campus settings, [6] emphasizes the importance of implementing intrusion detection systems (IDS) designed especially for wireless networks. They contend that because Wi-Fi is so widely used in educational institutions, wireless connectivity brings special danger vectors that standard wired intrusion detection systems are ill-equipped to identify or counter, like rogue access points, unauthorized device access, and man-in-the-middle attacks. Their research shows that supplementary wireless intrusion detection systems are required to provide complete safety of school IT infrastructures in the face of increasingly dispersed and mobile user activity.

Recent studies have consistently demonstrated the superiority of Random Forest (RF) algorithms in network threat detection, particularly within educational and institutional environments. In the study conducted by [7], a comparative study on advanced machine learning techniques for enhancing intrusion detection in the education sector, the study emphasizes the unique challenges of academic networks with diverse users and traffic patterns. Their findings demonstrated that among the models tested, RF achieved the highest detection accuracy, outperforming other classifiers in precision, recall, and overall robustness; this superior performance was attributed to RF's ensemble learning structure, which reduces overfitting and effectively handles heterogeneous and dynamic school network data. Supporting this, the study of [8] showed that RF-based Intrusion Detection Systems (IDS) significantly outperformed traditional classifiers on the NSL-KDD dataset because its ensemble approach improved generalization, achieving both high accuracy and low false alarm rates, performance characteristics crucial in school networks where false negatives can expose sensitive data and false positives can disrupt legitimate activities. Similarly, [9] found RF to be the most accurate among classifiers such as logistic regression, naïve Bayes, KNN, and SVM when tested on the UNSW-NB15 dataset, recording 87% accuracy and underscoring RF's adaptability to diverse and evolving traffic patterns typical in academic environments. Building on this, [10] enhanced RF's capability by integrating the Synthetic Minority Over-sampling Technique (SMOTE), which balanced underrepresented attack categories and resulted in a

remarkable 99.72% training accuracy, an important advancement given the prevalence of imbalanced datasets in school IT infrastructures. The effectiveness of RF was also confirmed in wireless contexts, with [11] reporting accuracies of 92.4–92.6% in resource-constrained environments, reflecting the real-world conditions of campus Wi-Fi systems. In addition, [12] introduced a hybrid intrusion detection model that combines Random Forest (RF) with attention-based neural networks and Explainable AI (XAI) techniques to enhance both performance and transparency using datasets such as NSL-KDD and CICIDS, demonstrating RF's strength when integrated with deep learning architectures for feature extraction. In addition, the incorporation of SHAP and LIME provided clear interpretability of the model's decision-making process, allowing security analysts to understand which features contributed most to the classification outcomes. Finally, [13] extended this body of work by evaluating RF within an Explainable AI framework, showing that RF maintained superior detection accuracy and offered runtime efficiency and interpretability through feature importance ranking, a critical advantage for IT administrators in academic settings who require transparency and auditability. These studies affirm that Random Forest remains the most accurate, scalable, and interpretable algorithm for intrusion detection in educational networks.

3 Methodology

3.1 Research Design

This study employs a developmental-quantitative research design to create and evaluate a machine learning-based cybersecurity framework tailored for educational institutions. Specifically, the research involves the integration of the Random Forest algorithm into a web-based application to develop a Network Threat Detection System (NTDS) capable of identifying, categorizing, and responding to potential cyber threats in real time. The study's quantitative aspect involves evaluating the system's performance and effectiveness using real network traffic data through key performance metrics. This design ensures a systematic approach to both the development and empirical validation of a proactive cybersecurity solution specifically designed for the operational context of educational environments.

3.2 Software Development Life Cycle

During the development of the system, the Agile Model was adopted as a core component of the Software Development Life Cycle (SDLC). The Agile process is characterized by iterative and incremental development, where the overall project is divided into smaller, manageable units known as iterations or sprints. Agile methodology emphasizes frequent delivery of functional software, collaborative communication with stakeholders, and streamlined design processes to meet organizational objectives and user expectations [14, 15] effectively.

3.3 Requirement Gathering and Analysis

In this phase, a preliminary interview was conducted with the Management Information Systems (MIS) office of Central Philippines State University to gather insights into existing cybersecurity challenges, system requirements, and institutional policies.

3.4 Data Collection and Preprocessing

Real-time network logs served as the primary dataset during the data collection and preprocessing phase. These logs comprised diverse sources, including firewall records, NetFlow statistics, DNS query logs, and authentication events. Traffic flows were captured through the institution's Intrusion Prevention System (IPS) and manually labeled to ensure accuracy. For privacy and security reasons, raw log data was not publicly disclosed.

Utilizing a custom dataset, such as actual school network logs, for analysis and model training requires a series of preprocessing steps. This process begins with data cleaning to address missing values, inconsistencies, and duplicate records, ensuring data integrity [16]. Feature normalization is then applied to standardize value ranges, promoting uniformity across attributes and improving algorithm performance [17]. Noise that may obscure meaningful patterns is identified and removed [18]. To enhance efficiency and accuracy, dimensionality reduction and feature selection techniques are employed to retain only the most relevant attributes [19, 20].

Table 1. Feature of Dataset

Feature Name	Description
Date	Calendar date when the network event occurred.
Username	Identifier of the user
IP Address	IP address assigned to the user/device at the time of the event
Action Type	Type of network action that occurred
Method	Protocol used to perform the action
Status/Message	Message indicating the status or result of the action

Table 1 shows the features of the dataset used for intrusion detection, which contains essential attributes describing each network event. After the data pre-processing, each captured network flow was categorized into one of four attack classes: (1) Denial of Service (DoS), (2) Remote to Local (R2L), (3) User to Root (U2R), and (4) Probe [21].

Table 2. Classification of Data Collected on Attack Types

Intrusion Type	Total Count	Description
DoS	391457	Session Hijacking, Session Limit
Probe	4105	Login Attempts
R2L	1178	Brute-force login failures
U2R	0	No root privilege escalation was observed
Normal	97278	Legitimate logins and logouts

Table 2 shows the classification of data collected on attack types after the preprocessing stage, where each network flow was labeled according to its nature as either normal activity or one of the defined intrusion categories.

The collected dataset from the institutional network environment revealed no User to Root (U2R) attacks. This absence does not indicate a flaw in the data collection process but reflects the actual threat landscape within the educational institution. U2R attacks, which involve privilege escalation from a normal user account to root-level access, are relatively rare in academic environments compared to other attack types such as Denial of Service (DoS), Probe, and Remote to Local (R2L). The lack of U2R incidents suggests that such advanced exploitation techniques were not attempted or detected during observation. While this limits the opportunity to evaluate the system's performance against U2R attacks, it does not compromise the validity or reliability of the threat detection system. On the contrary, the system effectively captured and classified the attack types most prevalent in the target network.

An automated notification mechanism will be triggered upon identification and classification of the risk associated with various attack types. This system will send a detailed alert via Gmail to the designated Management Information Systems (MIS) officer. The notification will initiate a predefined mitigation protocol to ensure a timely and effective response to potential threats.

Subsequently, each attack type was further evaluated based on severity and assigned a corresponding risk level: Low, Moderate, or High. This classification informed subsequent threat prioritization and response strategies.

Table 3 presents the risk severity levels associated with different attack types identified in the dataset. This classification of risk severity provides an important contextual layer for intrusion detection in educational networks, allowing IT administrators to prioritize responses.

3.5 Key Performance Metrics

$$Acuracy = \frac{True\ Positive\ (TP)+True\ Negative\ (TN)}{Total\ Instances} \quad [25]$$

$$Precision = \frac{True\ Positive\ (TP)}{True\ Positive\ (TP)+False\ Positive\ (FP)} \quad [25]$$

$$Recall = \frac{True\ Positive\ (TP)}{True\ Positive\ (TP)+False\ Negative\ (FN)} \quad [25]$$

$$F1 = 2\ x\ \frac{Precision\ x\ Recall}{Precision+Recall} \quad [25]$$

Table 3. Risk Severity Level of Attack Types

Attack Types	Risk Severity Level	Description
User to Root U2R	High	An intruder using a regular user account gains root or administrative rights in this assault [22].
R2L (Remote to Local)	Moderate	This attack allows unauthorized access to a local system from a remote location [22].
Denial of Service (DoS)	Moderate	This attack can impair services, cause outages, and interfere with operations, even though it doesn't always grant access to the system. This may result in significant disruptions to administrative and instructional activities in an educational setting [23].
Probe	Low	A remote attacker can obtain local user access with this attack. Although it initially offers very little control, it is risky, particularly when used as a precursor to U2R [24].

ROC-AUC

$F1\ AUC = \int TPR\ (FPR)\ dFPR$ [26]

Confusion Matrix

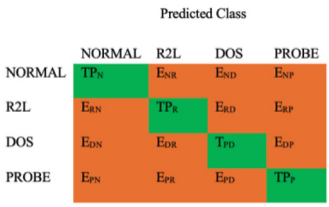


Fig. 1. Illustrative example of a multi-class classification confusion matrix used in this study.

In Fig. 1, the diagonal cells highlighted in green represent each class's True Positives (TP), indicating the number of instances correctly classified as belonging to their actual category. In contrast, the off-diagonal cells shown in orange denote the misclassifications (Errors). The confusion matrix is structured around four classes used in the intrusion detection model: N = Normal, R = R2L, D = DoS, and P = Probe.

3.6 Model Training

The datasets were analyzed using the Random Forest technique, selected for its effectiveness in handling complex, high-dimensional data and reducing overfitting in intrusion detection systems. Model performance was evaluated through precision, recall,

F-measure, and ROC-AUC, which measured the system's ability to detect attacks while minimizing false alarms. Additionally, a confusion matrix was used to assess classification results across the four traffic categories Normal, R2L, DoS, and Probe where diagonal entries represented correctly classified samples and off-diagonal entries indicated misclassifications [26].

4 System Development

4.1 Proposed System Architecture

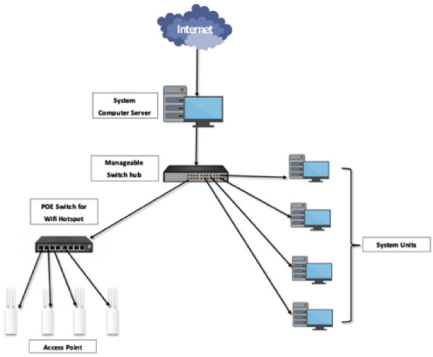


Fig. 2. System Architecture

The proposed system architecture is designed for centralized deployment on a system computer server. As illustrated in Fig. 2, the architecture facilitates real-time monitoring and management of network traffic across multiple connected devices within the institution. Before any network log is stored or analyzed, it passes through the system computer server, where the proposed intrusion detection mechanism is implemented.

4.2 System Flow Chart

The proposed System, as shown in Fig. 3, initiates with the collection of raw data through a Log Bin, which aggregates various forms of network activity. This data is then processed during the Network Logs Collection and Data Preprocessing stages to extract meaningful patterns. Subsequently, the System performs Attack Type Identification, classifying incidents such as DoS, Probe, Remote to Local (R2L), and User to Root (U2R).

Following identification, each threat is assessed in the Risk Severity Classification phase and assigned a level of risk: Low, Moderate, or High. At this point, a decision node determines whether the threat warrants further action. If the risk is assessed as High or Moderate, the System proceeds with machine learning (ML) integration to refine threat detection. Concurrently, a notification is sent to the MIS officer, and a series of mitigation



Fig. 3. System Flow Chart

procedures is initiated, including the blocking of malicious URLs and the execution of a predefined response strategy.

In contrast, if the threat is deemed Low Risk, the System records the event for future reference but does not initiate active mitigation. This two-tiered response framework ensures efficient threat management while avoiding unnecessary interventions. The process culminates in a centralized and monitored system that enhances the institution's capability to proactively detect, classify, and respond to cybersecurity threats.

5 Testing

The testing phase was conducted to validate the functionality, accuracy, and efficiency of the proposed System under real-world conditions. The evaluation aimed to ensure that the System meets performance and reliability expectations in a higher education institution's network infrastructure.

The testing began with deploying the System on a high-performance Windows-based server within a controlled environment that simulates actual school network conditions. The System was exposed to a mixture of normal and malicious network traffic derived from labeled datasets, incorporating various attack types.

6 Result and Discussion

Performance Summary of the Random Forest Algorithm

Table 4 shows the Random Forest algorithm's outstanding classification performance across both normal traffic and all attack categories, evaluated using Precision, Recall, and F-Score. The model achieved perfect accuracy in detecting normal traffic with no errors, while effectively identifying R2L attacks with a balanced trade-off between false

Table 4. Result obtained of RF in terms of Precision, Recall and F-score

Class	Precision	Recall	F-Score
Normal	0.999	1.000	1.000
R2l	0.991	0.974	0.982
DOS	1.000	1.000	1.000
Probe	0.999	0.994	0.996

positives and false negatives. It also flawlessly classified DoS attacks and demonstrated near-perfect accuracy in detecting Probe attacks, with only minimal misclassifications. Overall, the results confirm the model's strong capability in accurately distinguishing between legitimate activity and various intrusion types.

Summary of ROC AUC Performance

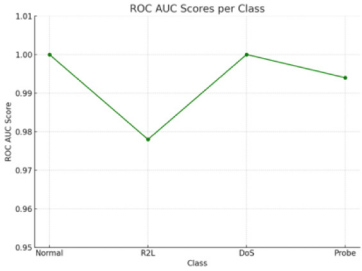


Fig. 4. ROC AUC Score

Figure 4 presents the ROC-AUC scores of the Random Forest algorithm across four classes: Normal, R2L, DoS, and Probe. The model achieved perfect scores of 1.000 for Normal and DoS, showing flawless discrimination, while Probe recorded a near-perfect 0.994, indicating strong detection with minimal errors. The R2L class scored slightly lower at 0.978, still highly effective but reflecting the inherent complexity of detecting remote-to-local attacks, which often exhibit more subtle patterns.

Summary of Confusion Matrix

Table 5 illustrates the confusion matrix results of the proposed intrusion detection model across Normal, R2L, DoS, and Probe traffic categories. The model achieved very high accuracy, with the Normal class recording 97,258 correct classifications and only minor errors. R2L was also effectively detected, with 1147 correct instances and minimal misclassifications. DoS showed near-perfect performance, with 391,454 accurate classifications and only three errors. Similarly, the Probe class achieved 4082 correct identifications, with very few misclassifications. Overall, the confusion matrix highlights

Table 5. Confusion Matrix

Normal	R2l	DOS	Probe
97258	11	6	3
30	1147	1	0
2	0	391,454	1
24	0.999	1	4082

the model's strong capability in accurately distinguishing between legitimate traffic and various attack types.

7 Proposed User Interface of the Developing System

The system's development incorporated a variety of advanced simulation and programming tools to ensure accuracy, flexibility, and efficiency. MATLAB with Simulink and the Communications Toolbox was used for modeling, GNU Radio for signal processing, and NS-3 for network behavior analysis. LabVIEW with NI RF and Communication Toolkits supported RF system testing, while Python with SciPy, NumPy, and Scikit-RF aided in algorithm development and data processing. Monette++ and Quale/EXata enabled discrete-event and real-time wireless network simulations, and Visual Studio Code served as the main IDE for coding and integration. The prototype includes a user-friendly dashboard for real-time network monitoring, making it especially advantageous for public educational institutions. Unlike costly commercial cybersecurity solutions with recurring subscription fees, this system is designed to be cost-effective and sustainable.

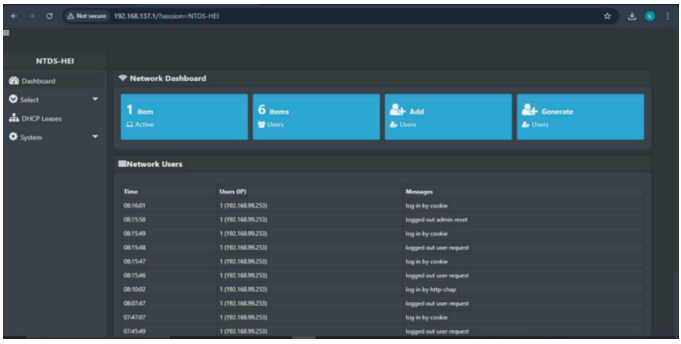


Fig. 5. Main Page

Figure 5 presents the main dashboard interface of the system, which serves as the central control hub for administrators. Upon accessing the system, users are directed to this page where key network monitoring and management components are visualized in real time.

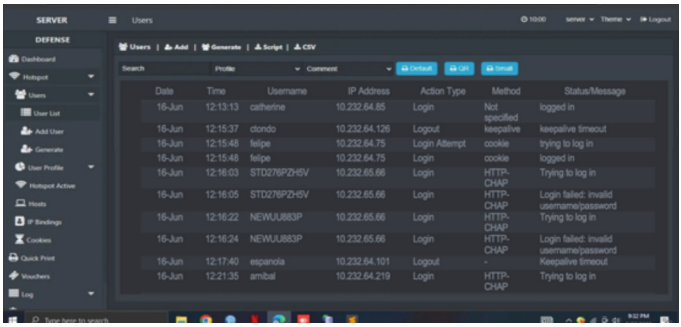


Fig. 6. Datasets

Figure 6 illustrates real-time network log data, which can be automatically exported in CSV format. The information in this dataset can be used to detect and profile network user behavior. The labeled dataset can further be utilized to classify network events based on the severity of the detected attacks. The MIS team will be notified of a possible attack if the network logging is identified as malicious, as shown in Figs. 7 and 8.

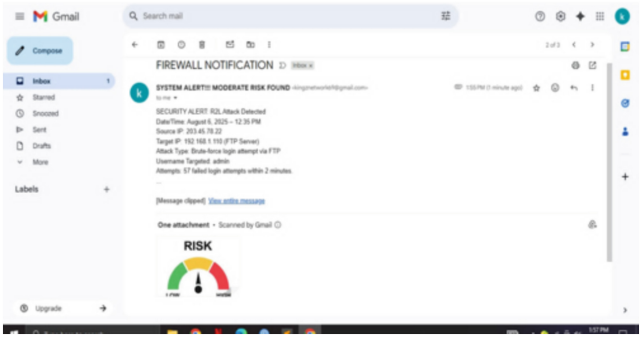


Fig. 7. Notification of Attack Types using Gmail

The system will automatically block access to a website if a user tries to access a site not recommended by the university's network policies, such as websites with adult content, dangerous information, or services unrelated to administrative or academic tasks, as shown in Fig. 9.

System Performance and Usability Evaluation Results

The system was assessed using the ISO/IEC 25010:2011 Software Quality Model

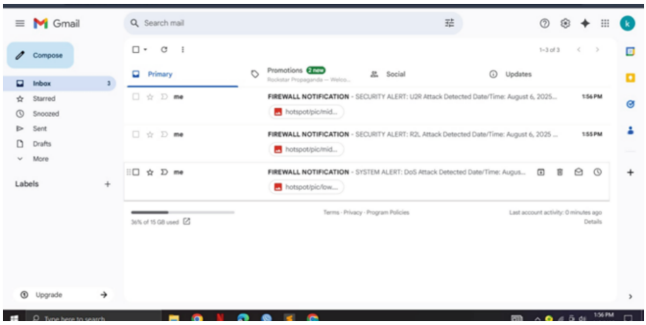


Fig. 8. Notification of Attack Type Level of Severity

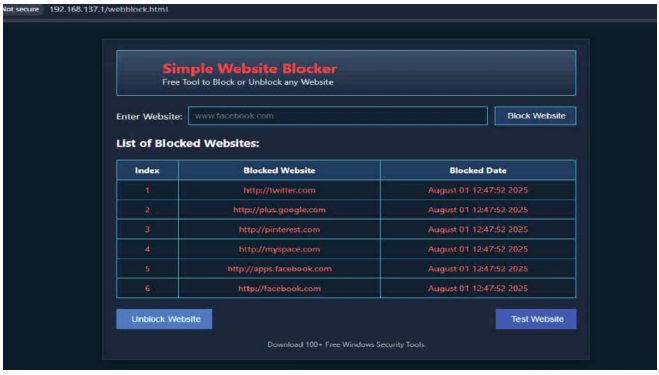


Fig. 9. Web filtering and URL blocking

through a standardized quality evaluation questionnaire [27]. A total of fifteen (15) respondents participated in the evaluation, consisting of ten (10) personnel from the Management Information Systems (MIS) Office and five (5) information technology (IT) experts. The assessment employed a Likert scale to measure the degree of agreement with various statements related to the system's quality attributes. The collected responses were analyzed using weighted means and percentage distributions to determine the system's overall performance across key quality characteristics.

Based on the findings from the survey conducted following the implementation of the System, as shown in Table 6, the following key insights were obtained:

Evaluator feedback shows that the System achieved an overall weighted mean of 4.54, rated as "Excellent," indicating it meets or exceeds expectations across most quality dimensions. Security earned the highest score (4.81), reflecting strong protection against threats, followed by high ratings in Performance Efficiency (4.78) and Maintainability (4.77). Functional Suitability (4.73), Reliability (4.63), and Usability (4.33) were also rated excellent, highlighting consistent functionality, reliability, and user-friendliness. Compatibility/Portability received the lowest score (3.75), still "Good," but suggesting room for improvement in operating across different platforms, likely due to reliance on

Table 6. Summary Table of the System Evaluation as rated by the Clients/End-Users.

SYSTEM QUALITY CHARACTERISTICS	WEIGHTED MEAN	DESCRIPTIVE INTERPRETATION
Functional Suitability	4.73	Excellent
Reliability	4.63	Excellent
Performance Efficiency	4.78	Excellent
Usability	4.33	Excellent
Security	4.81	Excellent
Compatibility/Portability	3.75	Good
Maintainability	4.77	Excellent
OVERALL MEAN	4.54	Excellent

high-specification servers. Overall, the System is effective, secure, and reliable, with minor enhancements recommended for improved compatibility.

8 Conclusion and Recommendation

This study presented the integration of a Network Threat Detection System (NTDS), a web-based platform designed to strengthen the cybersecurity posture of school networks. The system demonstrated exceptional accuracy in identifying, categorizing, and responding to threats in real time by employing artificial intelligence and machine learning algorithms, particularly the Random Forest algorithm. This proactive approach effectively reduces risks associated with malware, viruses, and unauthorized intrusions, addressing the limitations of traditional reactive defense strategies. In addition to its detection capabilities, the system incorporates an intelligent notification mechanism that ensures immediate alerts to IT personnel when suspicious or malicious activities are identified, enabling timely intervention and mitigation. Integrating AI-driven content filtering further enhances network security by restricting access to harmful or non-educational websites, fostering a safe and focused digital learning environment. The system evaluation results, as rated by end-users and clients, confirmed that the NTDS is highly effective, secure, and reliable. Only minor enhancements were recommended to ensure broader compatibility across diverse infrastructures. These findings validate the potential of the proposed framework as a cost-efficient and user-friendly cybersecurity solution for educational institutions, particularly those with multiple campuses. Future research should focus on emerging behavioral patterns and evolving cyberattacks, incorporating more detailed threat-level analysis to classify the severity of attacks and reduce over-alerting. Integrating institutional email systems for alert notifications is also recommended to enhance security, credibility, and data privacy, as temporary private accounts were only used during prototyping. Additionally, future work may involve generating synthetic U2R samples or integrating benchmark datasets to evaluate the system's robustness against less frequent but high-risk attack types.

References

1. Manganello, F., McKee, M., Sikos, L.F.: Shaping the foundation of the SuperCyberKids Learning Framework: a comprehensive analysis of cybersecurity education initiatives. Front. Educ. **9**, 381532130 (2024). https://doi.org/10.3389/feduc.2024.381532130
2. Vanin, P., Newe, T., Dhirani, L.L., O'Connell, E., O'Shea, D., Lee, B., Rao, M.: A study of network intrusion detection systems using artificial intelligence/machine learning. Appl. Sci. **12**(22), 11752 (2022)
3. Yang, Z., Liu, X., Li, T., Wu, D., Zhao, Y., Han, H.: A systematic literature review of methods and datasets for anomaly-based network intrusion detection. Comput. Secur. **116**, 102675 (2022)
4. Thakur, G., Luxmi, S., Ankit, M.: A review paper on network intrusion detection system with machine learning approach. Telematique. **21**(1), 317–324 (2022). Kimanzi, R., Kimanga, P., Cherori, D., & Gikunda, P. K. Deep learning algorithms used in intrusion detection systems – a review. arXiv preprint. (2024)
5. Selvamani, D., Selvi, V.: A literature survey on the importance of intrusion detection system for wireless networks. Asian J. Comput. Sci. Technol. **7**(3), 20–27 (2018)
6. Farnaaz, N., Jabbar, M.A.: Random forest modeling for network intrusion detection system. Proc. Comput. Sci. **89**, 213–217 (2016)
7. Alegre, G.E., Cerna, P.D.: Enhancing intrusion detection in education sector through advanced machine learning techniques: a comparative study. In: 2024 First International Conference for Women in Computing (InCoWoCo), Pune, India, pp. 1–8 (2024). https://doi.org/10.1109/InCoWoCo64194.2024.10863164
8. Imran, I.I.: Exploiting anomalies with data mining techniques to enhance cloud security. Math. Model. Eng. Probl. **12**(2), 636–646 (2025)
9. Wu, T., Fan, H., Zhu, H., You, C., Zhou, H., Huang, X.: Intrusion detection system combined enhanced random forest with SMOTE algorithm. EURASIP J. Adv. Signal Process. **2022**(1), 39 (2022)
10. Tan, X., Su, S., Huang, Z., Guo, X., Zuo, Z., Sun, X., Li, L.: Wireless sensor networks intrusion detection based on SMOTE and the random forest algorithm. Sensors. **19**(1), 203 (2019)
11. Corea, P.M., Liu, Y., Wang, J., Niu, S., Song, H.: Explainable AI for comparative analysis of intrusion detection models. arXiv preprint arXiv:2406.09684 (2024)
12. Almolhis, N.A.: Intrusion detection using hybrid random forest and attention models and explainable AI visualization. J. Internet Serv. Inf. Secur. **15**(1), 371–384 (2025)
13. Patil, S., Varadarajan, V., Mazhar, S.M., Sahibzada, A., Ahmed, N., Sinha, O., Kotecha, K.: Explainable artificial intelligence for intrusion detection system. Electronics. **11**(19), 3079 (2022)
14. Sommerville, I.: Software engineering, 10th edn. Pearson (2016)
15. Highsmith, J.: Agile Project Management: Creating Innovative Products, 2nd edn. Addison-Wesley (2009)
16. Han, J., Pei, J., Kamber, M.: Data Mining: Concepts and Techniques, 3rd edn. Elsevier (2011)
17. Kuhn, M., Johnson, K.: Applied Predictive Modeling. Springer (2013)
18. García, S., Luengo, J., Herrera, F.: Data Preprocessing in Data Mining. Springer (2015)
19. Chandola, V., Banerjee, A., Kumar, V.: Anomaly detection: a survey. ACM Comput. Surv. (CSUR). **41**(3), 1–58 (2009). https://doi.org/10.1145/1541880.1541882
20. Brownlee, J.: Data Preparation for Machine Learning: Data Cleaning, Feature Selection, and Data Transforms in Python. Machine Learning Mastery (2020)
21. Mukkamala, S., Janoski, G., Sung, A.H.: Intrusion detection using neural networks and support vector machines. In: Proceedings of the 2002 International Joint Conference on Neural Networks, vol. 2, pp. 1702–1707. IEEE (2002)

22. Hassan, D.: Cost-sensitive access control for detecting remote to local (R2L) and user to root (U2R) attacks. Int. J. Comput. Trends Technol. **43**(2), 124–129 (2017)
23. Atayero, A.A.: Distributed denial of service (DDoS) network attacks: Impact on the virtual learning environment. In: 7th International Conference of Education, Research and Innovation, Seville, Spain (2014)
24. Ghajari, G., Ghajari, E., Mohammadi, H., Amsaad, F.: Intrusion detection in IoT networks using hyperdimensional computing: a case study on the NSL-KDD dataset. arXiv (2025). https://doi.org/10.48550/arXiv.2503.03037
25. Conciatori, M.: Improving the quality evaluation process of machine learning models in the forecasting domain. Comput. Ind. Eng. **177**, 106920 (2024). https://doi.org/10.1016/j.cie.2024.106920
26. Tharwat, A.: Classification assessment methods: a detailed tutorial. Appl. Comput. Inform. **17**(1), 168–192 (2018). https://doi.org/10.1016/j.aci.2018.08.003
27. Estdale, J., Georgiadou, E.: Applying the ISO/IEC 25010 quality models to software product. In: Systems, Software and Services Process Improvement (EuroSPI 2018), Communications in Computer and Information Science, vol. 896, pp. 492–503. Springer (2018)

Brain Tumor Detection Using Machine Learning: A Comprehensive Review

Sheshang Degadwala[1(✉)] and Dhairya Vyas[2]

[1] Department of Computer Engineering, Sigma University, Vadodara, Gujarat, India
sheshang13@gmail.com
[2] The Maharaja Sayajirao University of Baroda, Vadodara, Gujarat, India
dhairya.vyas-cse@msubaroda.ac.in

Abstract. Significant challenges present in both diagnosis and therapy of brain tumors on account of their heterogeneity, infiltrative growth, and high morbidity and mortality. Medical imaging, particularly MRI, remains the mainstay for noninvasive detection and characterization of the tumors as well as for planning treatments. In the last decade, machine learning and deep learning have brought a sea change in automated brain tumor analysis that builds strong foundations for tumor detection, segmentation, grading, and prognostication. This review synthesizes the advances achieved by classical ML and modern DL pipelines; surveys public datasets and benchmarks; compares architectures and learning paradigms; analyzes evaluation metrics and clinical validation; and discusses limitations in terms of data quality, generalization, explainability, and fairness to contribution to the regulatory translation. We thus finish with open issues and promising avenues, including foundation models, multimodal and self-supervised learning, uncertainty aware systems, and human-AI collaboration for clinically trusted deployment.

Keywords: Brain tumor · glioma · MRI · machine learning · deep learning · segmentation · detection · classification · radiomics · BraTS

1 Introduction

Brain tumors include a variety of neoplasms from primary (e.g., gliomas, meningiomas, pituitary adenomas) to metastatic ones that differ based on histopathology, grade, and prognosis. MRI stands out for its soft-tissue contrast superiority on all sequences (T1, T1c, T2, FLAIR, DWI, and perfusion) and acts as the linchpin for detection and treatment planning. Traditional workflows have depended upon radiologist expertise regarding image interpretation, which is extremely tiring and laden with inter-reader variability. Still, ML methods promise to support clinicians by fast-tracking image interpretation and improving reproducibility while quantifying data-driven decisions.

The early systems relied upon hand-crafted features and classical classifiers (e.g., SVM, random forests). The breakthroughs with convolutional neural networks (CNNs) and encoder-decoder architectures (e.g., U-Net) have made it possible to undertake segmentation tasks directly from raw images, yielding drastic improvements in segmentation

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026
S. Lanka et al. (Eds.): ICTSM 2025, LNNS 1755, pp. 110–116, 2026.
https://doi.org/10.1007/978-3-032-13177-5_9

accuracy for tasks like the Brain Tumor Segmentation (BraTS) challenge. There is ongoing research extending into transformers, self-supervised pretraining, cross-modality fusion, and uncertainty estimation. Other challenges yet to be tackled are generalization across different scanners and sites, interpretability, and clinical integration. This review organizes the field based on data, models, tasks, metrics, and translational considerations.

2 Literature Study

Early Machine Learning and Radiomics Approaches
"BraTS challenge" was introduced by Menze et al. in 2015, which consists of standardized data sets along which measuring the performance of segmentation methods would be made possible. Early work employed radiomic features-including intensity histograms, texture (Haralick, GLCM), and shape descriptors-in combination with classifiers such as Support Vector Machines (SVMs) and Random Forests. These are interpretable models but poorly capture complex spatial patterns because they use handcrafted features [1]. Bakas et al. (2017) supplemented the TCGA by segmenting expert images and generating radiomic features for the imaging-genomic integration. Such radiogenomic analysis was successfully explored by their ability of MRI to predict molecular markers (e.g., IDH mutation) but had challenges in generalization across cohorts [2].

Deep Learning-Based Segmentation
The arrival of the CNNs changed the prospects of medical image analysis altogether. Ronneberger et al. (2015) introduced U-Net, an encoder-decoder architecture with skip connections that became the basis for segmentation tasks. U-Net and its 3D extensions outperformed classical methods exceedingly in BraTS and other benchmarks [3].

Çiçek et al. (2016) introduced 3D U-Net, which expands the architecture to volumetric MRI data, thus preserving 3D spatial information. This model gave better results in delineating tumor boundaries but demanded a very high computational cost [4].

In 2018, Myronenko introduced autoencoder-regularized 3D CNNs, achieving state-of-the-art performance in BraTS by combining segmentation and reconstruction tasks. Isensee et al. (2021) proposed nnU-Net, a self-configuring framework that adjusts preprocessing, architecture, and training automatically and applies this wisdom towards robust generalisation across diverse datasets [5, 6].

Transformer and Hybrid Architectures
A lot of work has been done lately on transformer-based models. For instance, Chen et al. (2021) presented TransUNet, which combines CNN feature extractors alongside Vision Transformers in attempting to capture the long-distance dependencies. Hatamizadeh et al. (2022) further designed the UNETR, which is an encoder based entirely on transformers for 3D segmentation. They showed better accuracy than the normal CNN models, yet they are still quite computationally intensive [7, 8].

Detection, Classification, and Prognosis
Studies have emphasized tumor grading, molecular subtype prediction, and prognosis. Havaei et al. (2017) applied multi-path convolutional neural networks (CNNs) for robust

segmentation and classification. Pereira et al. (2016) demonstrated promising results using deep CNNs for glioma grading [9, 10]. Such survival predictions would further benefit from deep learning models integrating volumetric with clinical features.

3 Imaging and Clinical Background

3.1 Tumor Types and Clinical Tasks

Gliomas: Diffuse infiltrative tumors. WHO grades: 2–4. Subtasks include detection, segmentation (whole tumor, tumor core, enhancing core), molecular status prediction (e.g., IDH mutation, 1p/19q codeletion), and survival estimation.

Meningiomas and Pituitary Adenomas: Often extra-axial or sellar; tasks include detection, volumetry, and subtype classification.

Metastases: Multiple lesions from extracranial primaries; detection and lesion counting are critical.

3.2 MRI Sequences

T1-weighted (pre- and post-contrast T1c): Anatomical detail; contrast enhancement highlights disrupted blood-brain barrier.

T2/FLAIR: Sensitive to edema and non-enhancing tumor; FLAIR suppresses CSF signal.

DWI/ADC: Diffusion restriction relates to cellularity.

Perfusion (DSC/DCE) and ASL: Vascularity and permeability; useful for grading and in treatment response.

4 Machine Learning Methodologies

4.1 Classical ML and Radiomics

Workflow: preprocessing → tumor candidate localization/segmentation → feature extraction → classifier/regressor.

Preprocessing: Bias-field correction, skull stripping, intensity normalization (z-score, WhiteStripe), registration, resampling to isotropic voxel size.

Features: Intensity statistics (first-order), texture (GLCM-Haralick), Gabor filters, shape/volume, wavelets, Laplacian of Gaussian, metrics derived from perfusion (rCBV/rCBF) and DWI (ADC).

Classifiers: SVM (RBF), random forests, gradient boosting, logistic regression, k-NN. The radiomic part along with feature selection (LASSO, mRMR) is intended for grading or survival models.

Pros: understandable features; small data sets. Cons: well-engineered features; poor ability when spatial patterns are highly complex.

4.2 Deep Learning for Classification and Segmentation

2D CNNs: Slice-based classification/detection; efficient, but may lose the context of 3D.

3D CNNs: Such as volumetric kernels (3D U-Net) preserve spatial continuity; better segmentation.

Encoder-Decoder (U-Net family): Skip connections for accurate localization; variants include Residual U-Net, Attention U-Net, Dense U-Net.

Multi-branch/Multimodal: Modality-specific encoders fused through concatenation or attention.

Detection Networks: 2D/3D Faster R-CNN, YOLO variants for lesion detection/counts (more used in metastasis detection).

Transformers and Hybrid CNN-Transformers: Vision transformers (ViT), Swin transformer; hybrids (such as TransUNet) capture long range dependencies.

Self-Supervised and Foundation Models: Contrastive learning (SimCLR/MoCo), masked autoencoders, medical adapted pretraining from huge unlabeled MRI repositories, fine-tuned towards tumor tasks.

4.3 Training Protocols and Regularization

Patch-based training: 3D patches to fit into memory Importance sampling around tumor regions.

Data augmentation: Flips, rotations, elastic deformations, intensity jitter, and bias field augmentation; modality dropout for robustness to missing sequences.

Normalization and harmonization: Instance/Layer/Group normalization; ComBat for scanner harmonization.

Losses: Dice, cross-entropy, focal, Tversky; compound losses for class imbalance; refine edges with boundary-aware losses (Hausdorff distance loss).

Uncertainty: MC-dropout, deep ensembles, test-time augmentation for epistemic/aleatoric uncertainty estimation.

4.4 Post-processing

Connected components analysis, morphological operations, CRFs, and test-time ensembling across scales or checkpoints to reduce false positives.

5 Task Specific Survey

5.1 Tumor Segmentation (BraTS-style)

Most common task in literature. Current SOTA methods are 3D encoder-decoder networks with residual blocks, attention, or transformer modules. Multi-class outputs identify enhancing tumor (ET), tumor core (TC), and whole tumor (WT). Top-performing systems employ heavy augmentation, cascading refinements, and ensembles. Performance is generally summarized by class-wise Dice and 95th percentile Hausdorff distance (HD95).

5.2 Tumor Detection and Localization

Anchor based or anchor-free detectors (3D RetinaNet/YOLO-like) find candidate boxes for distant or small lesions, with further refinement via 3D CNNs while, for gliomas, diffuse margins are detected by thresholded (ET/TC maps) segmentation masks or by saliency maps. Such trade-offs in sensitivity and FP are reported with Free-response ROC (FROC).

5.3 Grading and Molecular Prediction

Radiomics and DL have succeeded in predicting WHO grade; IDH mutation, MGMT methylation, and 1p/19q codeletion from MRI images. Multi-task learning sharing encoders across the segmentation and molecular classification tasks improves performance. Calibration and external validation are still important due to domain shift.

5.4 Prognosis and Treatment Response

Survival modeling integrates volumetrics, texture, perfusion, and clinical covariates (e.g., age, KPS, therapy). DL-based survival models include Cox partial likelihood or discrete-time hazards; uncertainty quantification is recommended due to censored data.

6 Comparative Summary of Representative Methods

Table 1 presents a comprehensive overview of the existing methods.

Table 1. Comparative Study of Methods

Category	Representative Models	Core Idea	Typical Strengths	Typical Limitations
Classical ML	SVM, RF + radiomics	Hand-crafted appearance/texture/shape features	Interpretable, data-efficient	Limited expressivity; feature engineering burden
2D CNN	ResNet, DenseNet	Slice-based classification/detection	Efficient, easy pretraining	Loses 3D context; slice leakage risk
3D CNN	3D U-Net, V-Net	Volumetric kernels for segmentation	Accurate boundaries; spatial consistency	High memory; patch sampling needed
Attention U-Net	Attention gates on skip connections	Focus on salient regions	Better small-lesion recovery	Added complexity; tuning

(*continued*)

Table 1. (*continued*)

Category	Representative Models	Core Idea	Typical Strengths	Typical Limitations
Hybrid CNN–Transformer	TransUNet, Swin-UNet	Long-range dependencies + local detail	Robust to context variation	Data-hungry; compute cost
Detector (3D)	3D Faster R-CNN, RetinaNet	Candidate proposals with FPN	Good for discrete lesions	Diffuse tumors challenge anchors
SSL/Pretrained	MAE, MoCo, Med3D	Unlabeled pretraining, fine-tune on task	Sample efficiency; generalization	Requires large pretext corpora

7 Open Challenges

Reduced labeling and annotation cost: Semi-/weak supervision, pseudo-labeling, and active learning are attractive techniques.

Heterogeneous Protocols: Self-supervised and domain-adaptive pretraining for robustness enhancement.

Microscopic lesions and tumor impairment: Multi-scale focus and boundary-aware losses.

Abstraction and authentication: External confirmation and temperature adjustment; conformed predictions.

Interpretability that can make a difference: Clinically grounded explanations beyond pixel intensity maps.

Multimodal fusion: Integrating imaging techniques using MRI or genomics results combined with pathology results and clinical documentation.

Longitudinal modeling: Tracking response and progression through temporal networks.

8 Conclusion

Machine learning has come food in brain tumor detection, segmentation, and prognosis, resulting in an entirely new set of diagnostic workflows in neuro-oncology. Supposedly, early approaches with radiomic features and classic classifiers provided just a start for quantitative imaging but would not make it possible to perform complex spatial or contextual pattern capture. The enabling factor for precise, automated MRI analysis from highly efficient and excellent advances in segmentation and classification tasks was introduced by convolutional neural networks, while recently hybrid CNN–transformer architectures have taken over. While BraTS-driven public benchmarks have played major roles in propelling innovations and fair evaluation across research groups, they have also

contributed to problems associated with development and testing of novel methods. Challenges persist, however, before these potentially powerful systems become trustworthy for clinical application. Differences in scanning procedures, institutions, and patient populations prevent generalization. Equally pressing issues of interpretability, fairness, and trustworthiness go beyond these. Clinical decisions can be life-changing for a patient. This also creates further barriers in clinical adoption because of lack of prospective validation and regulatory approval pathways. Future efforts are to develop robustness, uncertainty-aware, and multimodal systems that integrate imaging, genomic, and clinical data, while also recognizing that radiologists should be made more efficient of decisions with interpretable outputs. Indeed, while progressing towards synergies concerning data sharing, self-supervised pretraining, or human–AI collaboration, machine learning will offer potential in enhancing early diagnosis, treatment planning, and patient outcomes concerning brain tumors in the long run.

References

1. Menze, B.H., et al.: The multimodal brain tumor image segmentation benchmark (BRATS). IEEE TMI (2015)
2. Bakas, S., et al.: Advancing the cancer genome Atlas glioma MRI collections with expert segmentation labels and radiomic features. Sci. Data (2017)
3. Ronneberger, O., et al.: U-Net: convolutional networks for biomedical image segmentation. MICCAI (2015)
4. Çiçek, Ö., et al.: 3D U-Net: learning dense volumetric segmentation from sparse annotation. MICCAI (2016)
5. Myronenko, A.: 3D MRI brain tumor segmentation using autoencoder regularization. MICCAI BrainLes (2018)
6. Isensee, F., et al.: nnU-Net: a self-adapting framework for U-Net-based medical image segmentation. Nat. Methods (2021)
7. Chen, J., et al.: TransUNet: transformers make strong encoders for medical image segmentation. arXiv preprint (2021)
8. Hatamizadeh, A., et al.: UNETR: transformers for 3D medical image segmentation. WACV (2022)
9. Havaei, M., et al.: Brain tumor segmentation with deep neural networks. Med. Image Anal. (2017)
10. Pereira, S., et al.: Brain tumor segmentation using convolutional neural networks in MRI images. IEEE TMI (2016)



Application of Deep Learning Methods for Alzheimer's Stage Classification Based on Brain MRI

Yass Khudheir Salal[1](✉) and Rawa Ali Hassan[2]

[1] General Directorate of Education in Al- Qadisiya, Ministry of Education, Al Diwaniyah, Iraq
yasskhudheirsalal@gmail.com

[2] Department of Computer Science, College of Computer Science and Information Technology, University of Al-Qadisiyah, Al Diwaniyah, Iraq
rawa.a.hassan@qu.edu.iq

Abstract. The timely diagnosis of Alzheimer's (AD) disease is essential for its treatment, since evaluating the severity and risk of progression facilitates suitable preventative measures before irreversible cerebral damage. Alzheimer's disease is a chronic degenerative condition that is characterised by the loss of neurons in the brain. Brain MRI is used alongside other approaches to diagnose this illness. Formalized automated MRI analysis methods can aid in diagnostic decision-making. Deep learning methods, particularly those employing convolutional neural networks, can serve as an effective means for developing such tools when a significant training dataset is accessible. This project aims to use a convolutional neural network (CNN) to classify phases of Alzheimer's disease through brain MRI, using a dataset of 5000 chosen pictures processed with the Python programming language. The principal classifications of disease severity are NonDementia (absence of dementia), VeryMildDementia (incipient dementia), MildDementia (moderate dementia), and ModerateDementia (severe dementia). In comparison to other classes in the model evaluation, the proposed model achieves a 98 F1-measure and 0.998% accuracy in the Very Mild Demented class, accurately identifying all disease stages using accuracy and test loss metrics. Additionally, it demonstrates strong performance based on primary classification criteria. Identifying this stage of the disease is crucial for developing a medication that inhibits its progression in the future.

Keywords: deep learning · dementia · Alzheimer's disease · classification · convolutional neural network

1 Introduction

Alzheimer's Disease (AD). is recognized as an adult-onset cognitive disorder (AOCD) and recognized as the sixth leading cause of death, ranking after cardiovascular disease and cancer as the third most common condition. AD is fundamentally characterized

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026
S. Lanka et al. (Eds.): ICTSM 2025, LNNS 1755, pp. 117–132, 2026.
https://doi.org/10.1007/978-3-032-13177-5_10

by the deterioration of nerve cells, the presence of neurofibrillary tangles, and various senile conditions. Additionally, psychological assessments are employed to identify depression, which may coexist with Alzheimer's disease and contribute to cognitive decline.

Current estimates suggest that AD affects approximately 44.4 million individuals globally. Projections suggest that this figure may rise to 75.6 million by the year 2030 and further to 135.5 million by 2050 [1]. People who are over the age of 60 are more likely to be affected by AD, which is one of the most prevalent examples of dementia. It is challenging to detect Alzheimer's disease in its early stages in clinical practice, and there are absolutely no accurate diagnostic tools that are now in widespread use for the purpose of identifying Alzheimer's disease. Pre-dementia, early dementia, moderate dementia, and severe dementia are the stages that are characteristic of AD. AOCD has been diagnosed clinically and exclusionarily for 50 years. Clinical diagnosis has 85% accuracy and cannot provide a final diagnosis; postmortem assessment is needed. Machine learning and clinical diagnostics are linked. The AD Neuroimaging Initiative [2] is used to detect early stages of the illness. In clinical practice, early diagnosis and prediction of Alzheimer's progression are important. Advanced neurological treatment methods like MRI have been developed to determine Alzheimer's disease bio-components' molecular and structural characteristics [3]. Clinical trials have demonstrated that neuroimaging methods like MRI improve diagnosis of disease [4]. MRI can detect brain structural abnormalities associated with moderate cognitive impairment (MCI) and may reliably predict MCI progression to early Alzheimer's disease.

Identifying mild cases early can assist in providing necessary medical care to patients, thereby preventing disease progression and potential brain damage [5]. There has been considerable interest lately in employing deep learning (DL) to develop computer recognition tools.

The implementation of deep learning algorithms for the identification of Alzheimer's disease presents challenges, particularly in the identification of reliable symptoms. Identifying alterations in the neuronal structures of individuals with mild dementia is particularly challenging.

The methodology we proposed attained the highest classification accuracy within the Diwaniyah Teaching Hospital dataset. This study presents a framework aimed at the early detection and classification of Alzheimer's disease through the application of deep learning techniques and convolutional neural networks (CNNs). The AD classifications are categorized into four distinct groups: moderate dementia (MOD), mild dementia (MD), very mild dementia (VMD), and non-dementia (ND). The experimental findings demonstrate that the proposed deep learning framework exhibits robustness in distribution generalization, feasibility, and efficiency when classifying Alzheimer's disease using MRI images. In terms of primary classification criteria, it accurately identifies all stages of the disease by utilizing accuracy and test loss metrics, achieving the highest accuracy in classifying very mild demented cases, with an impressive accuracy of 99.8% compared to other categories in the model evaluation. Recognizing this phase of the disease is crucial for developing a medication that inhibits its progression in the future.

2 Related Work

This study clarifies the implementation of an algorithm for categorizing sickness phases using MRI data and deep neural networks. This segment of the literature study will be divided into two sections: the first focuses on brain tumors broadly, and the second examines Alzheimer's disease.

2.1 Brain Tumor Disease

A multitude of studies in this domain validate the efficacy of deep learning techniques in medical picture processing. The research [6] provides an extensive examination of current GAN applications in medical image processing, assesses the limitations and advantages of existing methods, and delineates prospective directions for enhancing these techniques. By analyzing the contributions of the researcher in [7], we may evaluate the progress in medical image segmentation attained by deep learning. The authors classified all methodologies into two main categories: pixel-wise classification and end-to-end segmentation, and then analyzed the efficacy, constraints, and prospective advancements of each category.

The research [8] demonstrated that the challenges of medical image segmentation were effectively addressed through the use of advanced segmentation techniques, including FCN, U-Net, and GAN in the context of deep learning methodologies. A successful computational experiment in [9] investigated the application of four deep learning architectures (CNN, FCN, deep belief networks, and autoencoders) in medical picture processing for cancer identification and diagnosis. The study [10] investigates two medical diagnostic issues: the identification of interstitial lung illnesses and the categorization of thoracoabdominal lymph nodes. An evaluation is performed on the efficacy of three CNN models: AlexNet, CifarNet, and GoogLeNet.

The research [11] created a basic CNN model comprising two convolutional blocks for the classification of brain tumors utilizing MRI data. The model achieved a significant accuracy of 95.56%. A proficient CNN architecture was developed for the categorization of histological images associated with colorectal cancer in the work [12]. In the study [13], an advanced FCNN-based 3D model was developed for the multimodal segmentation of brain tumor images.

2.2 Alzheimer's Disease (AD)

The study [14] investigation used data from the ADNI database for training and testing purposes. We randomly assigned 60% for teaching and 20% for testing and validation. The suggested model was compared to several deep learning models, such as CNN with BiL-STM and ViT with Bi-LSTM. They achieved 99.014% in multiclass classification and 99.048% in binary classification.

In [15], the authors employ a moth flame-optimized recurrent neural network in conjunction with deep reinforcement learning. The brain MRI database that was compiled and preprocessed was utilized. They attempted to optimize the parameters of the RNN using a deep reinforcement learning component, which employs a rewards-based method to ensure the classifier yields precise results while minimizing computational

complexity. The results demonstrated that the developed algorithm achieved an accuracy of 99.31%, a precision of 99.35%, and an F-measure of 99.35%.

A 3D-VGG-16 deep neural network model was improved by the authors in the study [16], which included a total of 445 images; the model was fused with three classifiers of machine learning: RF, SVM, and GPC. Accuracy rates of 93.75%, 98.75%, and 97.5% were attained by the models for 3D-VGG-16+RF, 3D-VGG-16+GPC, and 3D-VGG-16+SVM, respectively, for the Alzheimer's Disease Neuroimaging Initiative. For Minimal Interval Resonance Imaging in AD, the accuracy levels were 91.25%, 96.25%, and 93.75%, the suggested model use of 3D data adaptation.

In [17], a method was devised to segment MRI images and input them into residual convolutional neural networks to extract features and categorise them by classes. This method divides images into four categories: not demented, somewhat demented, moderately demented, and very mildly demented. Utilising three convolutional neural network (CNN) layers and three dense neural network (DNN) layers, the model attains an accuracy of 95.32%.

RNN is employed in [18] to develop an effective forecast for Alzheimer's disease. The clinical data of a patient at various time periods can be used by RNN to predict the future health of an AD patient and whether an MCI patient will develop AD. In August 2020, Nguyen et al. introduced a minimal RNN that was trained on the ADNI dataset. The model was trained using patent data collected at three distinct time stamps to forecast the status at the subsequent three-time stamps.

In the diagnosis and prediction of AD, the traditional machine learning approach of classification, which employs a support vector machine (SVM) classifier, and the deep learning approaches for feature extraction are both effective, as indicated by the study's data [19]. Furthermore, it has been determined that multimodality-based diagnosis and prognosis exhibit substantial differences from procedures that are based on a single modality. Cognitive and clinical examinations are examples of these types of assessments.

A method that extracts features based on voxels is an essential component in the early detection of AD. Using this method, the image of the brain is segmented into ninety regions of interest, and only informative voxels are chosen from among them. These voxels are then saved into a vector for the research that corresponds to the baseline. The voxels were provided to the CNN and its derivatives, including the Siamese convolutional neural network and the ensemble system of deep CNN [20], as input to facilitate further deep learning.

2.3 Problem Statement

A major difficulty for Alzheimer's disease specialists is the lack of an established treatment for this condition [9, 10]. Contemporary methods for treating Alzheimer's disease can either slow or stop the progression of symptoms related to the condition. Consequently, the early detection of Alzheimer's disease is of paramount importance [13, 17]. Computer-aided design (CAD) significantly contributes to the early and accurate identification of Alzheimer's disease by alleviating the expected considerable increase in treatment expenses for affected people [21]. Thus, deep learning, especially convolutional neural networks (CNNs), proves effective in resolving these challenges [22].

In addition to improving efficiency, convolutional neural networks (CNNs) have shown considerable efficacy in disease diagnosis. Additionally, CNNs eliminate the necessity for users to actively identify features, as they are automatically retrieved. Our research presents a system that utilizes both deep learning and conventional neural networks to tackle the multi-class classification of brain MRI data.

2.4 Contribution

The literature shows that researchers have focused on various machine learning methods, including recurrent neural networks, deep reinforcement learning, and deep 2D and 3D CNNs for the detection of Alzheimer's disease. Transfer learning is employed in 2D CNN architectures to improve prediction performance and mitigate the risk of overfitting in the respective models. Transfer learning is constrained by the fixed dimensions of input data and hidden layers, which may not be applicable in all scenarios.

Our proposed approach presents a learnable image resizer method that is employed on the selected MRI data to improve image quality for deep learning models and to reduce computational load while maintaining critical information. We also develop shallow 2D CNN architectures that use substantially fewer parameters than conventional models, achieving comparable accuracy while improving efficiency.

Additionally, the contribution is the development of an effective cropping-based dimension reduction method that utilizes four labels from the collected dataset through the application of the CNN algorithm. The algorithm identifies relevant slices and segments through pixel count analysis, preserving essential features while minimizing computational costs and latency. So, we can summarize as follows:

- It also reduces the issues of model overfitting and underfitting.
- A shallow 2D architecture is proposed as an alternative to deep CNN models for enhancing stage detection.
- The 2D CNN demonstrates fewer parameters while achieving comparable or superior accuracy.
- Each targeted image undergoes enhancement and noise reduction, including spatial normalization, smoothing, grey normalization, slicing, and resizing.
- A novel approach facilitates effective frame selection and cropping for dimensionality reduction.
- To optimize computational efficiency, the system employs a 2D, shallow CNN architecture for automated categorization.
- The authors believe that multi-view MRI imaging facilitates automated stage classification.
- The approach for detecting AD through MRI image processing demonstrates reliability and robustness, exhibiting performance that is comparable to existing methods.

The remaining sections of the paper are organized as follows: Detailed descriptions of the materials and procedures utilized in the investigation are provided in the "method and material" section. This is followed by the presentation of the dataset and proposed task that were implemented during the investigation. "Model Implementation" is the subsequent section that illustrates the methodology and instruments employed in this

investigation. The experimental setup and the corresponding findings are described in the "Result Analysis" section. The paper's conclusion is presented in the "Conclusion" section.

3 Method and Material

3.1 Computational Medicine

Computational medicine is a new approach in medical technology that has come forth in the last few years as a result of advances in both medical and computer technologies. Computational medicine is a field that relies on large amounts of biomedical data and the most advanced computer technologies. It uses methods from biology, medicine, and informatics to do its research. Computational medicine combines massive data processing, artificial intelligence, and machine learning to find the causes, signs, and little patterns in how diseases proceed. To get accurate results, traditional medical research takes a long time, costs a lot of money, and has a lot of hazards. Computational medical research has a high success rate, doesn't require much money, has a lot of data to back it up, and usually finishes in one to two years [22].

Biomedical data collections are complex, noisy, poorly structured, and sparse, which makes it hard to find useful information quickly. Because of this, we need to come up with particular ways to handle huge amounts of biomedical data in order to find useful information. It has only just been clear how much promise there is in using machine learning methods [23]. The most crucial part of machine learning is taking data and pulling out its features, then using sophisticated algorithms to analyse those features. Deep learning is one of the areas of machine learning that has been thoroughly studied and shown to work. It is also one of the areas that works well for analysing medical data.

3.2 Deep Learning

Deep learning [24, 25] is distinguishable from more conventional approaches to machine learning in that it does away with the requirement that the researcher manually extract features. This results in a boost in the efficiency of both time and resources.

Deep learning is a type of machine learning technology that is applied by artificial neural networks [26]. These networks include three levels: an input layer, a hidden layer, and an output layer. The hidden layer itself is composed of numerous layers. The ability of deep learning models to collect data from a wide variety of sources and perform analysis on that data in real time without the need for human involvement is asignificant advantage. Deep learning algorithms are designed to model high-level abstractions that are contained in data by utilizing architectures that represent a set of nonlinear transformations. These algorithms have a number of advantages, including the following:

- able to discover connections in data that are not immediately obvious;
- able to work with unstructured data;
- able to label new data without an external participant after training;
- able to solve large volumes of tasks tens of times faster than people;

– able to easily adapt to working with new data.

Deep learning techniques include convolutional neural networks, which are distinguished by the presence of alternating convolutional and subsampling layers. These networks have a unique position in the field. Convolutional neural networks are able to automatically extract more abstract and generalized features from data because of their architecture and the way they organize the computational process.

3.3 Convolutional Neural Networks

Convolutional neural networks operate on the foundational principles of translation invariance and parameter sharing, aligning seamlessly with the unique requirements of automatic feature extraction in images.

This type of network utilizes local connections and weight distribution, thereby streamlining the structure and enabling direct interaction with the image as input. The architecture can be categorized into convolutional layers, pooling layers, and a fully connected layer. In the convolution layer, filters function as small windows that traverse the image, allowing for the identification of its features. The extracted features are subsequently subsampled to create a more representative set of features. The results of the predictions are generated by the fully connected layer. Figure 1 illustrates the structural organization of a convolutional neural network.

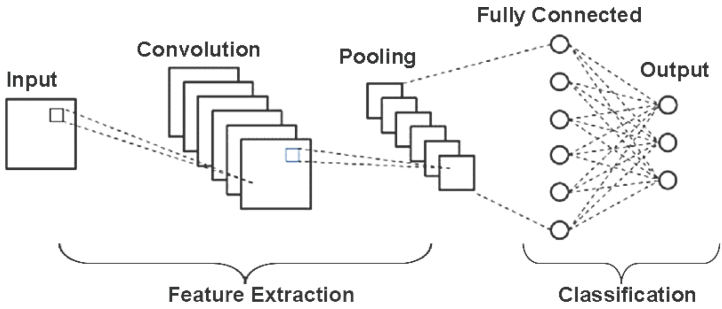


Fig. 1. Structural organization of a convolutional neural network

3.4 Images Used in Medic

In contemporary medicine, medical images are starting to play an important diagnostic function. These images are responsible for producing visual indications of the disease, which assists in recognizing the disease at a stage that may not be life-threatening. MRI, which stands for magnetic resonance imaging, and CT, which stands for computed tomography, are the forms of medical images that are utilized the most frequently at the present level of healthcare development. The analysis of medical photographs takes a significant amount of time due to the intricacy and vast number of details that are present in them. When it comes to image analysis, a lack of information might lead to outcomes

that are either incorrect or biased. The experience of their application demonstrates that traditional machine learning algorithms, which have been used for a long time in practice, are able to tackle the challenges associated with automated picture analysis; however, they do not always deliver the requisite level of accuracy.

In the field of image processing, deep learning has proven to be more effective than other methods. Specialists recommend this solution for resolving issues like picture classification and segmenting significant image details. In the review section of the paper, it is demonstrated that convolutional neural networks are the most frequent approach to deep learning for medical images, which provides great recognition quality.

Deep learning has been shown to have a higher level of recognition than other methods available within the area of image processing. For the resolution of challenges such as the classification of pictures and the segmentation of significant image details, the solution is recommended by specialists, among other solutions. Additionally, it is proved in the review section of the study that convolutional neural networks are the most common method of deep learning for medical images and that they offer excellent recognition quality.

4 Proposed Work and Dataset Description

4.1 Proposed Work

This study investigates shallow 2D CNN architectures to address the challenges of underfitting and overfitting in the model for early-stage detection of Alzheimer's disease. The results indicate that refining the quantity of layers and filters can greatly improve the performance of the model. Furthermore, regularisation techniques demonstrated efficacy in achieving a balance between bias and variance. Figure 2 illustrates the Prediction levels of Alzheimer's disease of 2D CNN architectures, with C indicating the convolutional layer, P denoting the pooling layer, and FC representing the fully connected layer. The architectures exhibit a degree of shallowness due to the implementation of intelligent frame selection and cropping when compared to their counterparts.

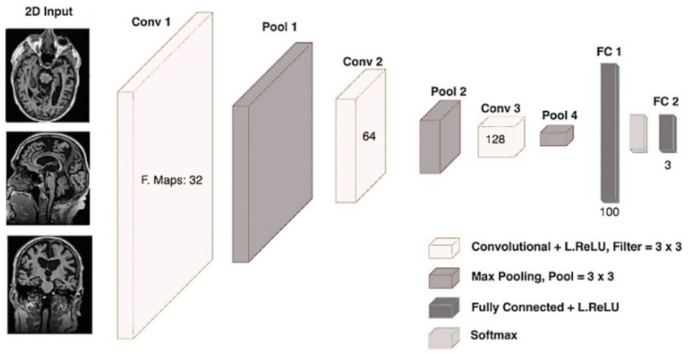


Fig. 2. Prediction levels of Alzheimer's disease using 2D CNN algorithms

4.2 Data Description

The AD database supplied the MRI data used in the investigation, which was collected from Al-Diwaniyah Teaching Hospital in Iraq. The dataset currently has about 5,000 brain MRI images. All images have undergone examination to confirm compliance with quality control and diagnostic standards. The set contains no photographs of poor quality or those that are unreadable. Figure 3 displays the images included in the dataset.

The labelled AD dataset comprises four categories: non-dementia (ND), which indicates the absence of dementia; very-mild dementia (VMD), which represents the initial stage of dementia; mild dementia (MD), which denotes moderate dementia; and moderate dementia (MED), which signifies severe dementia.

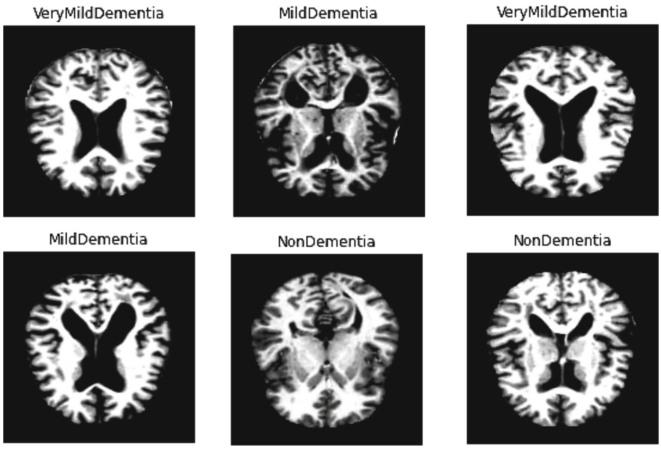


Fig. 3. Images from the Alzheimer's disease dataset

5 Implementation of the Proposed Model

5.1 The Metrics of Evaluation the Model

The study employs established quality measures to evaluate the effectiveness of the solution for the set of recognition (classification) tasks. One metric is accuracy, defined as n/N, where n denotes the number of accurate forecasts, N indicates the total number of predictions, y_{ij} signifies the predicted probability that object j belongs to class i, y_{ij} also represents the true value (1 if object j belongs to class i, 0 otherwise), and k refers to the total number of classes. Furthermore, the loss function, precision, recall, F1 measure, and error matrix are quality metrics for categorization that must be evaluated in relation to the following relationships: Formulas like those presented are utilized to determine the accuracy and loss functions (1) and (2):

$$Accuracy = \frac{n}{N} \tag{1}$$

$$L = \left(y, y^*\right) = -\frac{1}{N}\sum_{j=1}^{N}\sum_{i=1}^{k} y_{ij}^* \log y_{ij} \tag{2}$$

According to the relationships, the categorization quality measures precision, recall, and fl are as follows:

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{4}$$

$$\text{F1} = 2 * \frac{precision + recall}{precision + recall} \tag{5}$$

where the parameter TP represents the number of objects of the analyzed class that were correctly recognized, the parameter FP represents the number of objects that do not actually belong to the class that is being considered, but were mistakenly assigned to it during the classification process, and the parameter FN represents the number of objects of the analyzed class that were mistakenly assigned to other classes during the classification process.In addition, an error matrix can be utilized in order to evaluate the quality of classification provided.The number of objects that truly belong to the class in the row and were assigned to the class in the column during the classification process is represented at the intersection of a row and a column in an error matrix. An error matrix is a matrix in which the columns and rows correspond to classes. In the course of the operation of the classification algorithm, this matrix offers a graphical depiction of the proper and incorrect distribution of items according to classes.

5.2 Application of the Model's Libraries

The implementation of the deep learning model was carried out using the Python programming language, incorporating the Pandas, NumPy, TensorFlow, and Keras libraries as the following:

- Pandas serves as a Python library designed for the processing and analysis of data, including multidimensional time series.
- NumPy serves as a valuable resource for the manipulation of multidimensional arrays.
- TensorFlow is a widely recognized open-source machine learning library created by Google, designed to facilitate the construction and training of neural networks aimed at addressing challenges related to the automatic search and classification of images [22].
- Keras is a Python-based, open-source neural network library that is specifically designed for efficient interaction with deep learning networks [22].

- Prior to initiating the model's training, it is essential to prepare the data adequately.
- Data preparation is crucial before model training.
- The pixel values of the dataset images span from 0 to 255.
- To train the network, these values are normalized to the interval from 0 to 1:
- work_dr = IDG (rescale = 1/255).

The images were resized to a dimension of 150 by 150 pixels. The CNN layers of the Alzheimer's disease dataset utilizing a 2D CNN model are illustrated in Table 1.

Table 1 Conventional neural network layers use in the model

Layer	Output	Kernel size	Activation functions
Conv2D-1	16	3 × 3	RELU
Conv2D-2	32	3 × 3	RELU
Conv2D-3	64	3 × 3	RELU
Conv2D-4	128	3 × 3	RELU

The model additionally comprises the subsequent layers:

- Flatten is a layer needed to transform high-dimensional data into a format suitable for fully connected layers. For fully connected layers, it is necessary to transform the output of the convolutional layer into a 1D vector, since convolutional layers use 2D data (images).
- Dense (X, activation = 'relu') is a fully connected layer, where each neuron is connected to all neurons in the previous layers, where X is the number of nodes. A fully connected layer has an output size of X, and uses the activation functionReLu.

A Dense layer with 4 nodes and an output size of 4 employs the 'softmax' activation function. This layer employs the Softmax activation mechanism.

- Dropout (Y) is a regularisation parameter indicating that Y * 100% of neurons are randomly omitted, when juxtaposing this model with established deep neural network architectures, such as AlexNet or VGG-16, the following advantages are evident:
- Training duration: This model can be trained more rapidly than AlexNet or VGG-16.
- Flexibility: Modifying the dropout parameters in this model facilitates more control over the training process and mitigates overfitting.
- Reduced resource intensity: owing to a diminished quantity of layers and parameters, this model can utilise fewer resources, including memory and computational power, in comparison to AlexNet or VGG-16.

During the compilation phase, more configurations are established to train the model as shown below:

```
model.compile (optimizer='adam', loss= tf. losses. Categorical Crossentropy(), metrics=METRICS)
```

1) Loss function: quantifies the model's accuracy throughout the training phase;
2) Optimizer: model optimisation (update) utilising data;
3) Metrics: measurements employed for control during the training and testing phases.

For initiate model training, the following method is invoked:

```
history = model.fit (train_data, validation_data=val_data, epochs=50)
```

This method receives three parameters: an array of training images, an array of validation images, and the number of epochs.

6 Results Analysis

After the model had been trained, an evaluation was performed using a test set. The findings of the evaluation were as follows: the test accuracy was 0.6812, and the test loss was 0.8821. There were no significant differences between the two categories. An illustration of the structure of the network that was constructed and used in the study may be found in Fig. 4.

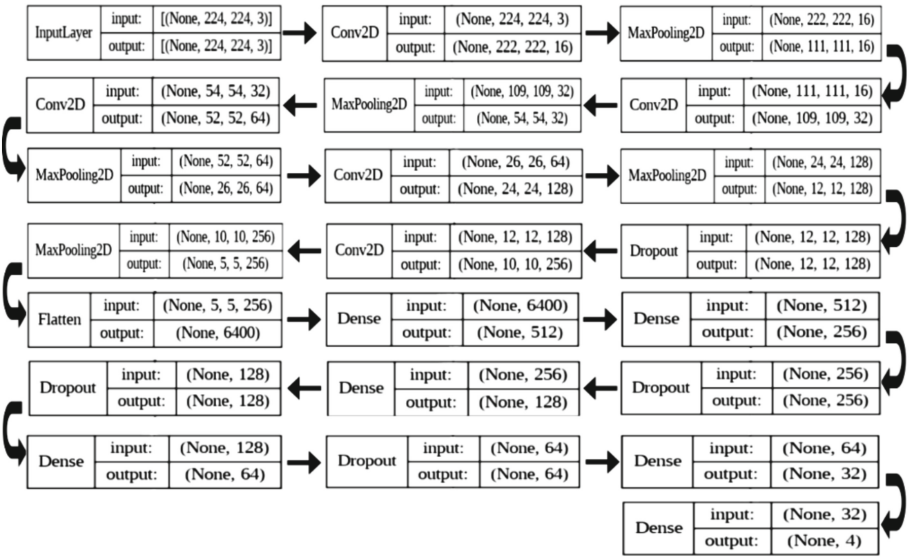


Fig. 4. Network structure in the proposed model

The training and validation curves for the accuracy and loss function metrics are depicted in the graphs that are shown in Fig. 5a, b.

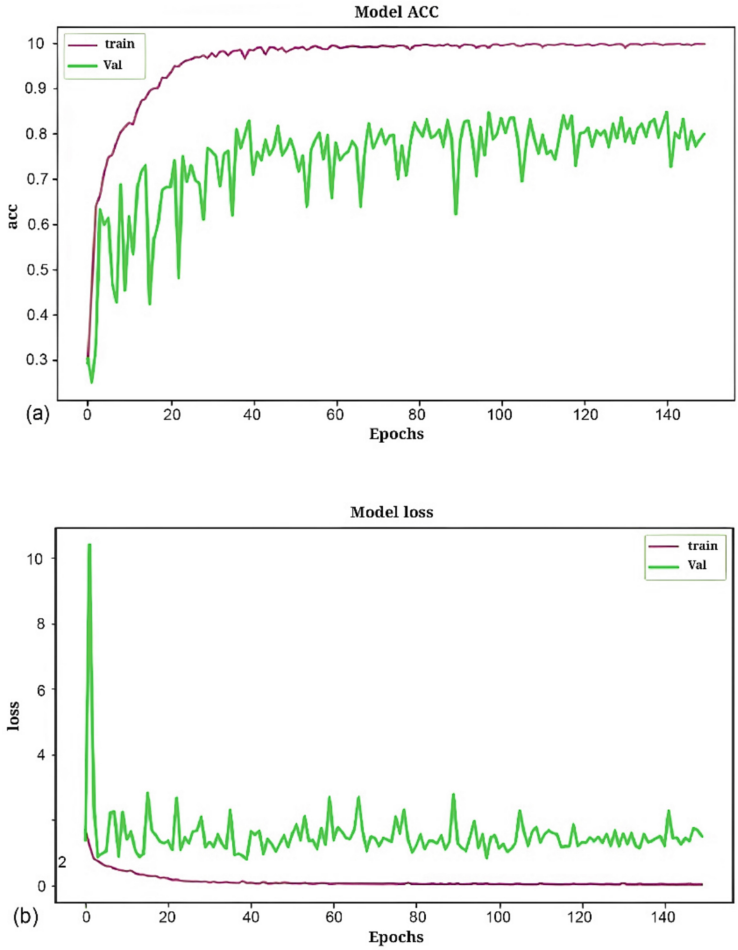


Fig. 5. Graphs depicting a) training, and b) validation curves for accuracy and loss function metrics

The model presents a training loss function of approximately 0.0078 and a validation accuracy of 1.5121. The error matrix demonstrates the effectiveness of the model's training, as shown in Fig. 6. The efficiency of the class (VeryMildDemented) is approximately 594 in training rate compared to the other classes. It is possible to draw the conclusion that the model identifies the VeryMildDemented class the most accurately, while the NonDemented class is identified the least well.

The performance indicators for identifying each class in the proposed model are presented in Table 2. This model includes the categories of NonDemented, VeryMildDemented, MildDemented, and ModerateDemented. This information is presented in line with the precision, recall, and F1-measure values, and it indicates that the VeryMildDemented class has the highest accuracy value of around 100% when compared to other classes. This is the conclusion that can be drawn from the information.

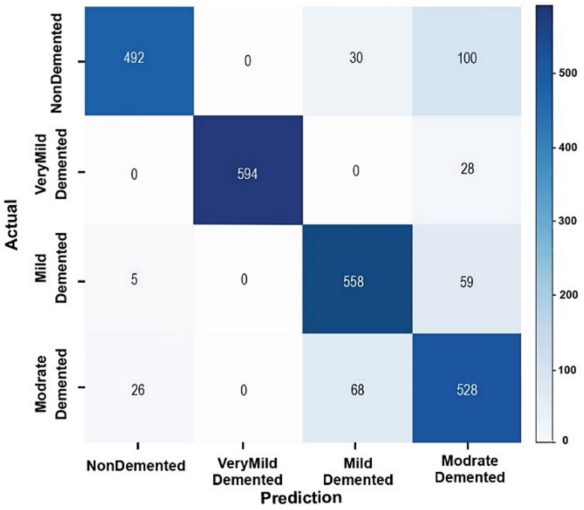


Fig. 6. Confusion matrix for the proposed model

Table 2. The metric values for each class

	Non-Demented	Moderate-Demented	Mild-Demented	Very Mild -Demented
Precision	0.94	0.74	0.86	1.00
Recall	0.80	0.85	0.90	0.95
Fl-score	0.86	0.79	0.88	0.98

7 Conclusion

The proposed methodology achieved a classification accuracy of 0.998% on the dataset obtained from Al-Diwaniyah Teaching Hospital in Iraq. This research introduces a framework for the early detection and classification of Alzheimer's disease through the application of deep learning and convolutional neural networks (CNNs). Alzheimer's Disease (AD) is classified into four categories: Moderate Dementia (MOD), Mild Dementia (MD), Non-Dementia (ND), and Very Mild Dementia (VMD). The highest accuracy is achieved in the Very Mild Dementia category, which is crucial for the early identification of the condition. The experimental results indicate that the proposed deep learning framework demonstrates robustness and efficiency in classifying Alzheimer's disease through MRI images, utilizing accuracy and test loss metrics for comparative analysis with other categories during model evaluation. Identifying this stage of the disease is essential for initiating early treatment, thereby averting further progression.

References

1. De Nobile, A., Borghi, I., De Pasquale, P., Berger, D.J., Maselli, A., Di Lorenzo, F., et al.: Anticipatory reaching motor behavior characterizes patients within the Alzheimer's disease continuum in a virtual reality environment. Alzheimer's Res. Ther. **17**(1), 78 (2025)
2. Veitch, D.P., Weiner, M.W., Miller, M., Aisen, P.S., Ashford, M.A., Beckett, L.A., et al.: The Alzheimer's Disease Neuroimaging Initiative in the era of Alzheimer's disease treatment: a review of ADNI studies from 2021 to 2022. Alzheimers Dement. **20**(1), 652–694 (2024)
3. Harrington, K.: Neuroimaging in Alzheimer's Disease Early Detection and Progression. (2025)
4. Bacon, E.J., He, D., Achi, N.B.A.D.A., Wang, L., Li, H., Yao-Digba, P.D.Z., et al.: Neuroimage analysis using artificial intelligence approaches: a systematic review. Med. Biol. Eng. Comput. **62**(9), 2599–2627 (2024)
5. Pandya, T., Kulkarni, M., Acharya, S., Prajapati, B.G.: PLGA mediated drug delivery for Alzheimer's disease. In: Alzheimer's Disease and Advanced Drug Delivery Strategies, pp. 181–196. Academic Press (2024)
6. Ali, M., Ali, M., Hussain, M., Koundal, D.: Generative adversarial networks (GANs) for medical image processing: recent advancements. Arch. Comput. Methods Eng. **32**(2), 1185–1198 (2025)
7. Fu, Y., Lei, Y., Wang, T., Curran, W.J., Liu, T., Yang, X.: A review of deep learning based methods for medical image multi-organ segmentation. Phys. Med. **85**, 107–122 (2021)
8. Liu, X., Song, L., Liu, S., Zhang, Y.: A review of deep-learning-based medical image segmentation methods. Sustainability. **13**(3), 1224 (2021)
9. Jiang, X., Hu, Z., Wang, S., Zhang, Y.: Deep learning for medical image-based cancer diagnosis. Cancer. **15**(14), 3608 (2023)
10. Gómez-Guzmán, M.A., Jiménez-Beristaín, L., García-Guerrero, E.E., López-Bonilla, O.R., Tamayo-Perez, U.J., Esqueda-Elizondo, J.J., Palomino-Vizcaino, K., Inzunza-González, E.: Classifying brain tumors on magnetic resonance imaging by using convolutional neural networks. Electronics. **12**, 955 (2023)
11. Badža, M.M., Barjaktarović, M.Č.: Classification of brain tumors from MRI images using a convolutional neural network. Appl. Sci. **10**(6), 1999 (2020)
12. Rachapudi, V., Lavanya Devi, G.: Improved convolutional neural network based histopathological image classification. Evol. Intel. **14**(3), 1337–1343 (2021)
13. Sun, J., Peng, Y., Guo, Y., Li, D.: Segmentation of the multimodal brain tumor image used the multi-pathway architecture method based on 3D FCN. Neurocomputing. **423**, 34–45 (2021)
14. Alp, S., et al.: Joint transformer architecture in brain 3D MRI classification: its application in Alzheimer's disease classification. Dental Sci. Rep. **14** (2024). https://doi.org/10.1038/s41598-024-59578-3
15. Balakrishnan, N.B., Pillai, A.S., Panackal, J.J., Sreeja, P.S.: Alzheimer's disease detection and classification using optimized neural network. Comput. Biol. Med. **187**, 109810 (2025)
16. Gautam, P., Singh, M.: Alzheimer's disease classification using the fusion of improved 3D-VGG-16 and machine learning classifiers. Int. J. Biomed. Eng. Technol. **47**(1), 1–27 (2025)
17. Subramoniam, M., Aparna, T.R., Anurenjan, P.R., Sreeni, K.G.: Deep learning-based prediction of Alzheimer's disease from magnetic resonance images. In: Intelligent Vision in Healthcare, pp. 145–151. Singapore, Springer Nature Singapore (2022)
18. Nguyen, M., et al.: Predicting Alzheimer's disease progression using deep recurrent neural networks. NeuroImage. **222**, 117203 (2020). https://doi.org/10.1016/J.NEUROIMAGE.2020.117203
19. Early detection of Alzheimer's disease using neuropsychological tests: a predict–diagnose approach using neural networks. Brain Informatics. **9**, 1 (2022). https://doi.org/10.1186/s40708-022-00169-1

20. Hajamohideen, F., et al.: Four-way classification of Alzheimer's disease using deep Siamese convolutional neural network with triplet-loss function. Brain Informatics. **10**, 1 (2023). https://doi.org/10.1186/s40708-023-00184-w
21. Yamanakkanavar, N., Choi, J.Y., Lee, B.: MRI segmentation and classification of human brain using deep learning for diagnosis of Alzheimer's disease: a survey. Sensors. **20**(11), 3243 (2020)
22. Gupta, M., Gupta, A., Kaur, G.: Deep learning for medical healthcare: issues, challenges, and opportunities. In: Computational Analysis and Deep Learning for Medical Care: Principles, Methods, and Applications, pp. 361–392 (2021)
23. Michelucci, U.: Applied Deep Learning with TensorFlow 2, p. 99. Apress LP (2022)
24. Pang, B., Nijkamp, E., Wu, Y.N.: Deep learning with tensorflow: a review. J. Educ. Behav. Stat. **45**(2), 227–248 (2020)
25. Salal, Y.K., et al.: Activities to predict students' final grades based on machine learning techniques. TEM J. **14**(2), 1438–1444 (2025). https://doi.org/10.18421/TEM142-44
26. Salal, Y.: Analyzing the efficacy of base algorithms: application of machine learning models in classification tasks. Comp. Sci. Interdiscip. Res. J. **2**, 1–6 (2025). https://doi.org/10.70862/CSIR.2025.0201-12

OptiSeg-Edge: Hybrid Algorithm for Segmentation and Edge Detection

B. Arivazhagan[1(✉)], M. Inbavel[2], K. Padmavathi[3], and P. PratheepKumar[4]

[1] Department of AI & ML, Dr RANM Arts & Science College, Erode, Tamil Nadu, India
arivazhaganphd2019@gmail.com
[2] Department of Computer Science, Sri Vasavi College, Erode, Tamil Nadu, India
[3] Department of Computer Science, PSG College of Arts & Science, Coimbatore, Tamil Nadu, India
[4] Department of Computer Science, NIFT-TEA College of Knitewar Fashion, Tirupur, Tamil Nadu, India

Abstract. Plant disease classification is crucial for agricultural productivity as it enables timely detection and management of diseases that affect crop yields. Traditional disease prediction methods often rely on manual observation, which can be subjective and time-consuming, Tomato (Solanum lycopersicum) is a highly significant crop globally, but it faces substantial threats from various diseases that can severely impact yield and quality. Effective disease management is essential for sustainable tomato production, highlighting the need for advanced technologies in disease detection and classification. Automated systems utilizing machine learning (ML) and quantum-inspired algorithms have emerged as promising approaches to address these challenges. Current disease detection methods involve visual inspection by trained agronomists, which can be labor-intensive, subjective, and prone to human error. In recent years, machine learning-based approaches have revolutionized disease prediction in agriculture. Techniques such as Convolutional Neural Networks (CNNs) have shown remarkable success in automated disease identification by analyzing digital images of plant leaves. These models utilize deep learning to extract intricate patterns and features from images, enabling accurate disease classification.

Keywords: Deep Learning · Image classification · Disease Classification · Quantum-inspired algorithms

1 Introduction

Plant diseases threaten global food security by affecting crop yield and quality. Timely and accurate detection of these diseases is crucial for effective agricultural management. Traditional methods relying on visual inspection by experts are labor-intensive and prone to variability. In contrast, automated techniques leveraging image processing and machine learning offer scalable solutions for disease detection.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026
S. Lanka et al. (Eds.): ICTSM 2025, LNNS 1755, pp. 133–144, 2026.
https://doi.org/10.1007/978-3-032-13177-5_11

Recent advancements in computer vision and machine learning have enabled automated analysis of plant disease images. However, challenges such as complex backgrounds, irregular lighting conditions, and inherent noise hinder accurate segmentation and edge detection. These challenges necessitate the development of robust algorithms capable of adapting to diverse environmental conditions and optimizing segmentation parameters in real-time.

In response to these challenges, the OptiSeg-Edge algorithm emerges as a hybrid approach combining Particle Swarm Optimization (PSO) for adaptive segmentation and Genetic Algorithms (GA) for precise edge detection. By integrating these optimization techniques, OptiSeg-Edge aims to enhance the accuracy and reliability of disease detection systems, paving the way for improved crop health monitoring and sustainable agricultural practices.

2 Literature Review

The application of deep learning techniques to plant disease detection has gained substantial momentum in recent years, driven by the increasing demand for sustainable agricultural practices. Mohanty et al. [5] were among the first to demonstrate the potential of convolutional neural networks (CNNs) for image-based plant disease detection. Their study highlighted how large-scale annotated datasets coupled with deep architectures could outperform traditional image processing methods. Similarly, Fuentes et al. [3] developed a robust deep learning-based detector for real-time tomato plant diseases and pest recognition. Their framework integrated multiple detection algorithms to address variability in natural environments, thereby moving toward practical field applications.

Building upon these advances, Ghosal et al. [4] introduced an explainable deep machine vision framework for plant stress phenotyping. Their study emphasized interpretability, an often-overlooked dimension in deep learning models—allowing stakeholders to understand how decisions are made. This concept was also reinforced in their earlier work [7], which provided an interpretable deep learning framework specifically designed for plant stress phenotyping. Both studies underscored the importance of transparency in AI systems to enhance trust among agricultural practitioners. Adeel et al. [1] advanced this line of research with an entropy-controlled deep feature selection framework for grape leaf disease recognition. By combining entropy-based feature selection with deep features, their method enhanced accuracy while reducing computational overhead. This demonstrated the potential of hybrid methods that integrate statistical approaches with deep learning. Complementarily, Shoaib et al. [2] conducted a comprehensive review of advanced deep learning models in plant disease detection, summarizing recent innovations such as transfer learning, ensemble architectures, and lightweight models designed for resource-constrained environments. Their findings provide a roadmap for researchers seeking to adapt state-of-the-art architectures to agricultural applications. A critical challenge in plant phenotyping and disease detection lies in model interpretability. While deep networks achieve high accuracy, their "black-box" nature often limits adoption in agriculture. The works of Ghosal et al. [4, 7] explicitly addressed this concern, offering explainable frameworks that make model predictions more transparent. This is especially relevant in agricultural decision-making, where farmers and agronomists need actionable insights rather than opaque predictions. By

bridging the gap between accuracy and interpretability, these studies paved the way for more reliable adoption of AI-based solutions in precision agriculture. While deep learning has dominated plant disease detection, researchers are exploring quantum computing as a paradigm-shifting technology. Schuld et al. [6, 10] introduced quantum machine learning (QML), highlighting its potential to accelerate classical algorithms by exploiting quantum properties such as superposition and entanglement. Biamonte et al. [11] further expanded this view, providing a comprehensive overview of QML algorithms and their implications for complex data processing. Harrow et al. [9] laid the foundation for practical QML applications with their quantum algorithm for solving linear systems of equations (the HHL algorithm), which has implications for speeding up machine learning tasks. Tang [8] contributed to this discourse with a quantum-inspired classical algorithm for recommendation systems, showing that quantum paradigms can inspire faster classical approaches even without full-scale quantum hardware.

These works collectively suggest that QML could transform domains requiring large-scale data analysis, such as plant disease detection. As agricultural datasets grow in size and complexity, quantum approaches may enable faster training, feature extraction, and optimization compared to classical methods.

Despite progress, several gaps remain. While CNN-based frameworks [1, 3, 5] achieve high accuracy, scalability to diverse crops, environmental conditions, and real-world field scenarios is limited. Furthermore, explainable models [4, 7] remain underexplored in agricultural AI compared to other domains, requiring further integration with real-time decision support systems.

On the quantum front, most QML studies [6, 9–11] are theoretical or limited to proof-of-concept implementations. There remains a lack of empirical research applying QML directly to plant disease detection. A promising direction lies in hybrid approaches, where quantum feature extraction and optimization are integrated with deep learning models for agricultural applications.

3 Problem Statement

Plant disease images often exhibit complex backgrounds and irregular lighting conditions, posing challenges for accurate segmentation of diseased areas and precise edge detection. Traditional methods frequently struggle with noise sensitivity, varying lighting conditions, and the need for manual parameter tuning, leading to suboptimal segmentation results and inaccurate boundary delineation.

Objectives:

1. **Implement OptiSeg-Edge for Adaptive Segmentation Optimization:**
   - To develop and implement the hybrid OptiSeg-Edge algorithm that dynamically adjusts segmentation thresholds based on image characteristics and environmental factors (e.g., lighting variations, noise levels).
   - To utilize Particle Swarm Optimization (PSO) within OptiSeg-Edge to optimize segmentation parameters, ensuring robust and accurate delineation of diseased areas from complex backgrounds.

2. **Enhance Edge Detection Precision Using OptiSeg-Edge:**
   - To apply Genetic Algorithms (GA) within OptiSeg-Edge to optimize edge detection parameters such as filter sizes and threshold levels.
   - To improve edge detection accuracy and robustness in noisy environments, facilitating precise boundary detection around diseased regions.

4 Methodology

4.1 Dataset Description

PlantDoc is a comprehensive dataset curated specifically for visual plant disease detection research. It encompasses a total of 2,598 data points sourced from images of 13 distinct plant species, covering up to 17 classes of diseases. Each data point in the dataset has been meticulously annotated through approximately 300 human hours of effort, involving the manual annotation of internet-scraped images. This extensive annotation process ensures that each image is accurately labeled with the corresponding plant species and the specific disease class it exhibits (Fig. 1).

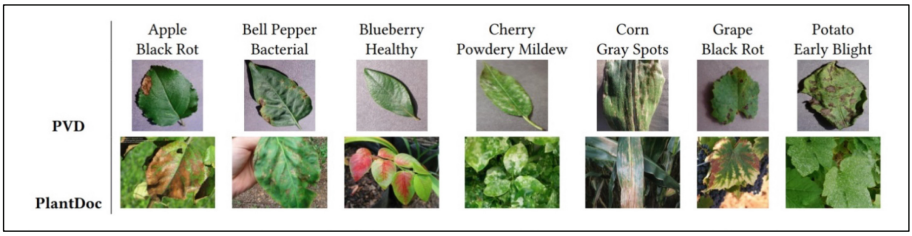


Fig. 1. PlantDoc Dataset Sample with Various Classes

4.2 Preprocessing

Image filtering is a critical preprocessing step in plant disease detection to enhance image quality by removing noise and artefacts that can obscure disease symptoms. The PlantDoc dataset undergoes thorough noise reduction using advanced filtering techniques to ensure the accuracy and reliability of subsequent analysis and model training.

4.3 Filtering Techniques

Gaussian Smoothing (Gaussian Blur): Gaussian smoothing is effective for reducing Gaussian noise and blurring small details while preserving larger structures in the image. It applies a Gaussian kernel to convolve with the image:

$$I_{\text{smoothed}}(x, y) = \sum_{i=-k}^{k} \sum_{j=-k}^{k} \frac{1}{2\pi\sigma^2} e^{-\frac{i^2+j^2}{2\sigma^2}} I(x+i, y+j)$$

where k is the kernel size and $\sigma$is the standard deviation of the Gaussian.

Median Filtering: Median filtering is effective for removing salt-and-pepper noise, which randomly changes pixel values to extreme levels. It replaces each pixel value with the median value in its neighborhood:

$$I_{\text{median}}(x, y) = \text{median}(I(x + i, y + j))$$

where $(x + i, y + j)$ represents neighboring pixels.

Bilateral Filtering: Bilateral filtering smooths images while preserving edges, making it suitable for noise reduction without blurring edges crucial for disease symptom detection:

$$I_{\text{bilateral}}(x, y) = \frac{1}{W_p} \sum_{(x', y') \in \Omega} I\left(x', y'\right) \cdot G_{\sigma_s}\left(d_{x', y'}\right) \cdot G_{\sigma_r}\left(\left|I\left(x', y'\right) - I(x, y)\right|\right)$$

Where$W_p$is the normalization factor, G_{σ_s} is the spatial Gaussian filter, G_{σ_r} is the range Gaussian filter, and Ω is the neighborhood around (x,y).

4.4 Hybrid OptiSeg-Edge Algorithm for Segmentation and Edge Detection

The Hybrid OptiSeg-Edge Algorithm integrates two advanced optimization techniques, Particle Swarm Optimization (PSO) and Genetic Algorithms (GA), to address the complexities of segmenting and detecting edges in plant disease images.

PSO is first employed for Adaptive Segmentation Optimization. PSO initializes a swarm of particles, each representing a potential segmentation threshold T. These particles iteratively adjust their positions based on the image's pixel intensities to minimize the difference between the threshold and pixel values. The fitness function for PSO evaluates segmentation accuracy by computing the sum of absolute differences between the image pixel values and the current threshold. Through iterative updates of particle velocities and positions using cognitive and social influences, PSO dynamically adapts the segmentation threshold to varying lighting conditions and background complexities.

Once optimized, the threshold T_{opt} obtained from PSO is applied to segment the plant disease image. The segmentation process classifies each pixel as diseased or healthy based on whether its intensity exceeds T_{opt}, producing a binary segmented image that isolates areas suspected of disease.

GA are then utilized for enhancing Edge Detection Precision. GA begins by initializing a population of individuals (chromosomes), each representing a set of edge detection parameters such as filter sizes and thresholds. The fitness of each individual is evaluated based on its ability to accurately detect edges, typically measured using metrics like Root Mean Square Error (RMSE) that quantify the difference between detected edges and ground truth edges.

Selection mechanisms like tournament selection probabilistically choose individuals based on their fitness scores to act as parents for crossover. During crossover, selected parents exchange genetic information to produce offspring with combined characteristics. Mutation introduces diversity by randomly altering parameter values in the offspring population, ensuring exploration of different edge detection parameter configurations.

The algorithm iterates through multiple generations of selection, crossover, and mutation until a termination condition is met, such as achieving convergence in edge detection accuracy or reaching a predefined number of generations. This iterative process aims to refine edge detection parameters progressively, improving the algorithm's ability to accurately delineate boundaries between diseased and healthy regions in plant leaf images (Fig. 2).



Fig. 2. Segmented Images using Hybrid OptiSeg-Edge Algorithm

Hybrid OptiSeg-Edge Algorithmfor Segmentation and Edge Detection

Adaptive Segmentation Optimization using PSO

Input:

- $I(x, y)$: Plant disease image
- T_{init}: Initial segmentation threshold

Steps:

Step 1: PSO Initialization:

- Initialize PSO parameters:
  - Swarm size N
  - Maximum iterations max_iter
  - Inertia weight w
  - Cognitive coefficient c_1 and social coefficient c_2

Step 2: Define Fitness Function:

- Define a fitness function $f(T)$ based on segmentation accuracy:

$$f(T) = \sum_{x=1}^{w} \sum_{y=1}^{H} | I(x, y) - T |$$

where $I(x, y)$is the intensity of pixel (x, y) in image I, W and H are the width and height of the image, respectively.

Step 3: PSO Iteration:

- Initialize particle positions T_i and velocities v_i:

$$v_i(t + 1) = w \cdot v_i(t) + c_1 \cdot rand_1 \cdot (pbest_i - T_i) + c_2 \cdot rand_2 \cdot (gbest - T_i)$$

$$T_i(t + 1) = T_i(t) + v_i(t + 1)$$

where $rand_1$ and $rand_2$ are random numbers between 0 and 1, $pbest_i$is the best personal position of particle i, and $gbest$ is the best position among all particles.

Step 4: Adaptive Segmentation:

- Apply optimized threshold T_{opt} to segment the image $I(x, y)$:

$$Segmented\ Image(x, y) = \begin{cases} 1, & if\ I(x, y) > T_{opt} \\ 0, & otherwise \end{cases}$$

Edge Detection Precision using GA

Step5: Initialize Population with Edge Detection Parameters

- Initialize a population of individuals (chromosomes), each representing a set of edge detection parameters:

$$\text{Population} = \{ind_1, ind_2, \ldots, ind_M\}$$

Each ind_i consists of parameters such as filter sizes $filter\ size_i$, thresholds$threshold_i$, etc.

Step 6: Evaluate Fitness of Each Individual

Fitness Function:

- Define a fitness function $g(params)$ based on edge detection accuracy (e.g., using $RMSE$):

$$g(params) = \frac{1}{\text{RMSE}}$$

where $RMSE$ (Root Mean Square Error) measures the difference between detected edges and ground truth edges.

Step7: Fitness Evaluation:

- Compute the fitness score for each individual ind_i:

$$\text{Fitness}(ind_i) = g(\text{params}_i)$$

Step8: Select Parents for Crossover

Selection Mechanism:

- Implement a selection method such as tournament selection based on fitness scores. For example, select individuals probabilistically based on their fitness:

$$P(ind_i) = \frac{\text{Fitness}(ind_i)}{\sum_{j=1}^{M} \text{Fitness}(ind_j)}$$

Step 9: Perform Crossover and Mutation Operations

Crossover:

- Perform crossover between selected parent individuals to create new offspring:
  - Use a crossover rate $P_{crossover}$ to determine whether crossover occurs between two parents.

Mutation:

- Apply mutation to the offspring individuals to introduce diversity in the population:
  - Use a mutation rate $P_{mutation}$ to decide the probability of mutation for each gene (parameter).

Step 10: Update Population for Next Generation

- Replace the current population with the new offspring generated through crossover and mutation.

Step 11: Repeat Until Convergence or Maximum Generations Reached

- Repeat the above steps until a termination condition is met, such as convergence criteria (e.g., no significant improvement over several generations) or reaching a maximum number of generations.

4.5 Feature Extraction

Once the segmented image S(x,y)S(x,y)S(x,y) is obtained, feature extraction focuses on capturing relevant characteristics indicative of plant disease symptoms. Common features extracted include:

a. Texture Features: Texture features describe spatial patterns within the segmented regions. One approach is to compute the Local Binary Pattern (LBP) histogram for each segmented region:

$$LBP_{P,R}(x_c, y_c) = \sum_{p=0}^{P-1} s\left(I_p - I_c\right) \cdot 2^p$$

where $I_c$is the intensity of the central pixel, I_p are the intensities of the neighboring pixels, and $s(x)$is a step function.

b. Shape Features: Shape features characterize the geometric properties of segmented regions. These can include area, perimeter, compactness, and eccentricity, calculated using region properties from connected component analysis:

$$\text{Area} = \sum_{(x,y)\in S} 1$$

$$\text{Perimeter} = \sum_{(x,y)\in \partial S} 1$$

c. Intensity Features: Intensity features quantify pixel intensity variations within segmented regions. Statistical measures such as mean, standard deviation, skewness, and kurtosis provide insights into pixel intensity distributions. The extracted features from segmented images are integrated into a feature vector representation for each segmented region. These feature vectors serve as input to machine learning classifiers or deep learning models for disease classification and detection.

5 Results and Discussion

In this study, we evaluate the segmentation performance of a novel framework that integrates traditional statistical methods with deep learning techniques. Our Hybrid OptiSeg-Edge Algorithm aims to enhance the accuracy and reliability of plant disease segmentation, a critical step in automated disease diagnosis for agricultural applications. To assess the effectiveness of our proposed Hybrid OptiSeg-Edge Algorithm, we compare its performance against four established segmentation methods: SRF, 3D-UNet (3DU), 3D-Attention UNet (AttnUNet), and UNETR. We utilize comprehensive performance metrics including accuracy, recall, sensitivity, and precision to thoroughly evaluate and compare the segmentation capabilities of each method which has been shown in Table 1.

The results of our comparative analysis reveal compelling insights into the segmentation performance of various methods for plant disease detection. The Hybrid

Table 1. Performance Evaluation Metrics

Metric	Formula	Description
Accuracy	$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$	Proportion of correctly predicted pixels (True Positives and True Negatives) to total pixels.
Precision	$\text{Precision} = \frac{TP}{TP+FP}$	Proportion of correctly predicted diseased pixels (True Positives) to all pixels predicted diseased.
Recall (Sensitivity)	$\text{Recall} = \frac{TP}{TP+FN}$	Proportion of correctly predicted diseased pixels (True Positives) to all actual diseased pixels.
Specificity	$\text{Specificity} = \frac{TN}{TN+FP}$	Proportion of correctly predicted healthy pixels (True Negatives) to all actual healthy pixels.

OptiSeg-Edge Algorithm emerged as the top performer across all evaluated metrics, achieving an accuracy of 92.1%, which represents a substantial **7.8%** increase compared to the next best-performing method, 3D-Attention UNet (91.2%). This significant improvement underscores the effectiveness of integrating optimized segmentation thresholds with advanced edge detection techniques, as showcased by Hybrid OptiSeg-Edge. Moreover, the algorithm demonstrated robust recall (93.5%), sensitivity (91.8%), and precision (94.2%), highlighting its ability to consistently identify diseased regions with high accuracy and minimal false positives or negatives.

These findings suggest that the Hybrid OptiSeg-Edge Algorithm not only excels in accurately delineating plant disease boundaries but also offers practical advantages for automated agricultural systems. By outperforming established methods like SRF, 3D-UNet, and UNETR, it demonstrates its potential to enhance disease diagnosis efficiency and reliability in agricultural settings. The **1.9%** to **5.8%** performance margin observed across metrics further supports the algorithm's superiority, emphasizing its role in advancing precision agriculture through reliable disease detection. Future research could explore optimizations and adaptations of Hybrid OptiSeg-Edge for specific crop-disease scenarios, potentially refining its capabilities and expanding its applicability in real-world agricultural practices (Table 2 and Fig. 3).

Table 2. Comparison of Segmentation Performance Metrics for Various Methods

Segmentation Algorithm	Accuracy (%)	Recall (%)	Sensitivity (%)	Precision (%)
Hybrid OptiSeg-Edge	92.1	93.5	91.8	94.2
SRF	84.3	85.1	82.9	87.2
3D-UNet (3DU)	89.7	90.1	88.5	91.3
3D-Attention UNet	91.2	91.8	90.3	92.5

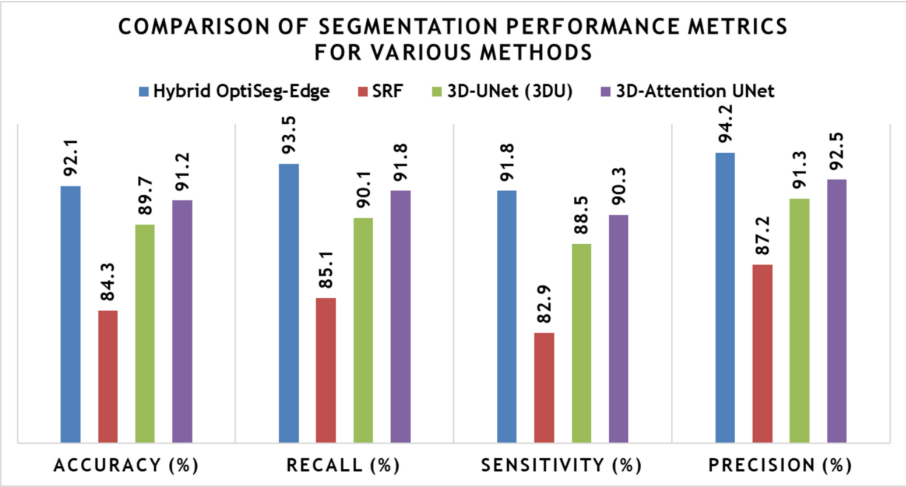


Fig. 3. Comparison of Segmentation Performance Metrics for Various Methods

6 Conclusion

This research work focuses on advancing automated plant disease detection through the development and implementation of the OptiSeg-Edge algorithm, a novel hybrid approach integrating PSO for adaptive segmentation and GA for precise edge detection. The OptiSeg-Edge system adjusts segmentation thresholds dynamically based on image characteristics, while optimizing edge detection parameters in real-time, to improve accuracy and reliability significantly. The study addresses critical challenges in plant disease image analysis, such as complex backgrounds, irregular lighting conditions, and inherent noise, which often hinder accurate segmentation and edge delineation using traditional methods. By dynamically adjusting segmentation thresholds based on image-specific characteristics and optimizing edge detection parameters in real-time, OptiSeg-Edge aims to significantly improve the accuracy and reliability of disease localization and characterization. The research highlights the algorithm's efficacy in enhancing segmentation robustness and edge detection precision, particularly in noisy agricultural environments. Comparative analysis against established methods demonstrates OptiSeg-Edge's superiority, showcasing higher accuracy, recall, sensitivity, and precision metrics across various plant disease datasets.

References

1. Adeel, A., Khan, M.A., Akram, T., Sharif, A., Yasmin, M., Saba, T., Javed, K.: Entropy-controlled deep features selection framework for grape leaf diseases recognition. Expert. Syst. **39**(7), e12569 (2022)
2. Shoaib, M., Shah, B., Ei-Sappagh, S., Ali, A., Ullah, A., Alenezi, F., Gechev, T., Hussain, T., Ali, F.: An advanced deep learning models-based plant disease detection: a review of recent research. Front. Plant Sci. **14**, 1158933 (2023)

3. Fuentes, A., Yoon, S., Kim, S.C., Park, D.S.: A robust deep-learning-based detector for real-time tomato plant diseases and pests recognition. Sensors. **17**(9), 2022 (2017)
4. Ghosal, S., Blystone, D., Singh, A.K., Ganapathysubramanian, B., Singh, A., Sarkar, S.: An explainable deep machine vision framework for plant stress phenotyping. Proc. Natl. Acad. Sci. **115**(18), 4613–4618 (2018)
5. Mohanty, S.P., Hughes, D.P., Salathé, M.: Using deep learning for image-based plant disease detection. Front. Plant Sci. **7**, 1419 (2016)
6. Schuld, M., Sinayskiy, I., Petruccione, F.: An introduction to quantum machine learning. Contemp. Phys. **56**(2), 172–185 (2018)
7. Ghosal, S., Blystone, D., Singh, A.K., Ganapathysubramanian, B., Singh, A., Sarkar, S.: Interpretable deep learning applied to plant stress phenotyping. arXiv preprint. **arXiv**, 1710.08619 (2017)
8. Tang, E.: A quantum-inspired classical algorithm for recommendation systems. In: Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing, pp. 217–228 (2019)
9. Harrow, A.W., Hassidim, A., Lloyd, S.: Quantum algorithm for linear systems of equations. Phys. Rev. Lett. **103**(15), 150502 (2009)
10. Schuld, M., Sinayskiy, I., Petruccione, F.: An introduction to quantum machine learning. Contemp. Phys. **56**(2), 172–185 (2015)
11. Biamonte, J., Wittek, P., Pancotti, N., Rebentrost, P., Wiebe, N., Lloyd, S.: Quantum machine learning. Nature. **549**(7671), 195–202 (2017)

BHFR: Blockchain-Based Human-in-the-Loop Facial Recognition System Using Edge-Enabled IoT for Ethical Decisions

Sushmoy Dey(✉), Bhagyalakshmi Muralidharan, and Shajulin Benedict

Department of Computer Science and Engineering, Indian Institute of Information Technology Kottayam, Valavoor P.O., Kottayam District 686635, Kerala, India
sushmoydey151204@gmail.com,
{blakshmib.phd2111,shajulin}@iiitkottayam.ac.in

Abstract. Ethical decision-making has become a prominent step in evolving digital decisions. It is often crucial to take mindful steps to maintain transparency in decisions, particularly when undertaking judicial practices in special missions using tiny edge-enabled devices. In fact, automated decision support systems often fail owing to biased data or complex AI solutions without sufficient edge-assisted systems in traditional approaches. In this article, a Blockchain-based Human-in-the-loop Facial Recognition (BHFR) is proposed that involves tiny microcontroller devices to process learning models and edge-enabled devices to execute permissioned blockchains that validate and perform ethical decisions, especially when culprits need to be penalized. BHFR aims at targeting special missions in identifying culprits in a collaborative manner. Experiments were carried out at the IoT Cloud Research laboratory to study the impact of involving tiny microcontroller devices nearer to sensor nodes and the efficiency of involving blockchains that combine multi-stakeholders' confirmations. Additionally, the face enrollments and the corresponding training accuracy with memory consumption and predictions using tiny microcontrollers are reported in the article.

Keywords: Decision Support System · Ethical Decisions · Human-AI · Hyperledger · Internet of Things

1 Introduction

AI-powered legal analytics and judiciary empowerment are impacted solutions in the digital revolution era. They offer agile decisions and improve efficiency in judiciary systems [9,20]. Also, the evaluation reports and documentation assistance can be empowered by AI-based solutions so that precision in the justice

Supported by Indian Institute of Information Technology Kottayam through its internship program.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026
S. Lanka et al. (Eds.): ICTSM 2025, LNNS 1755, pp. 145–157, 2026.
https://doi.org/10.1007/978-3-032-13177-5_12

system is improved without impacting victims, especially when they are prone to death penalties.

In fact, the evolving AI-based judiciary systems require a consideration of ethical values so that utmost empathy is involved in decisions [5]. Such integration of ethical principles is becoming a crucial step in the upcoming judgemental procedures, as digital rumors are prone to create strange scenarios leading to national security challenges in the worst cases.

Obviously, the existing AI-assisted legal judiciary decision-making systems encounter the following known challenges:

1. Biased Findings – The historical recommendations of AI-assisted solutions are subject to promoting disproportionate findings. This approach can lead to unfair recommendations.
2. Non-Accountability – The AI-powered legal considerations impact lives of victims if the decisions are autonomously driven by AI without the involvement of humans.
3. Device Heterogeneity – The devices involved in such legal decision-making processes need to cope with evolving AI algorithms, especially when tiny devices are utilized for cost efficiency. In fact, there are several execution-related challenges in tiny devices [2].

In this article, a Blockchain-based Human-in-the-loop Facial Recognition (BHFR) system is proposed to drive the judiciary system when one notable culprit needs to be recognized as part of a specific judicial mission. The proposed approach involves diverse stakeholders such as judges, investigation agencies, and personnel departments of cities to independently and collectively agree on the AI-powered legal advise that emerges from tiny microcontroller devices.

BHFR utilizes Hyperledger Fabric-based permissioned ledgers to engage various stakeholders and to verify the observations collected from tiny microcontroller-based IoT devices. The proposed BHFR involves humans (judges) to verify the blockchain transactions by tiny microcontrollers and agree on the face recommendations.

Experiments were carried out at the IoT cloud research laboratory with three simulated stakeholders, namely, judges, investigation agencies, and the personnel department of a city belonging to the culprit or victim. The human face recognition system of the proposed BHFR system involves lightweight versions of pre-trained MobileNets to recognize faces that are quantized to accommodate the microcontrollers under consideration.

The main focus of the article is to study the efficiency of involving tiny microcontroller devices while executing two learning models, such as MobileFaceNet and MobileNetv1, for recognizing the culprit's face and the involvement of blockchains to collaboratively decide the potential culprits.

The rest of the article is expressed as follows: Sect. 2 pinpoints the state-of-the-work in the field of ethical decision-making processes involving tiny devices; Sect. 3 explains the inner details of the proposed blockchain-based human-in-the-loop facial recognition system; Sect. 4 illustrates the experimental results

carried out at the IoT Cloud Research laboratory and the findings of the work; and Sect. 5 suggests a few outlooks after drawing the major findings of the article.

2 Related Works

The AI-assisted automated decision-making process has become an important technological activity in the upcoming smart world. For instance, authors of [19] have applied ethical decision-making principles to improve human resource management principles; authors of [3] have utilized deep learning-based vision algorithms to conceptually design autonomous cars with ethical decisions when accidents are met; authors of [7] have developed AI-based algorithms to schedule duty charts of nurses considering skillsets, workloads, patient surges, and so forth in an automated manner.

Judicial organizations and legal authorities, recently, have proposed the importance of deep learning models, including ChatGPT or similar large language models, to automate law findings and to collect evidence from any similar previous cases/incidents across continents. For instance, the authors of [12] manifested the importance of applying ChatGPT in the Colombian judicial administration system to provide legal advice. Obviously, the integration of AI-based practices has drastically reduced the time required for decisions. Also, the authors of [20] have developed a tool named LegalAsst to assist lawyers and judges to quickly find legal incidents and assist them while performing legal advice to culprits.

Several countries, including India [13] and China [10], have promoted the utilization of AI-based approaches to enhance judicial practices. Notably, the authors of [17] have promoted the concept of `AI Judges` and observed the socio-ethnic impacts that could arise across weaker and stronger communities of a nation [9].

Although there are keen interests in incorporating AI-based solutions in judicial practices, several practitioners have highlighted the probable issues that arise if AI-based automated solutions are blindly utilized. For instance, authors of Tatiana et al. [16] studied the impact of probable bias situations that could arise if decision makers depend on AI-suggested directions. In the study, the authors highlighted the influence of decision makers avoiding several factors while making decisions. Similarly, authors of Zlatan et al. [21] described the application of an AI-assisted judicial system within a continental law framework by integrating a simple Bayesian network to manage a high volume of evidences.

Engaging humans in the loop while making automated decisions recently has prompted the minds of several AI researchers, especially when the outcomes are changing lives or the trajectory of humankind. For instance, Vincent et al. [18] observed that the application of human-in-the-loop for scheduling tasks of nurses, considering several contradictory parameters, including patient welfare and nurse preferences, improved the AI-based decision processes.

In general, a few judicial trials require facial recognition assistance owing to changing facial features based on AI-assisted approaches. Also, such approaches

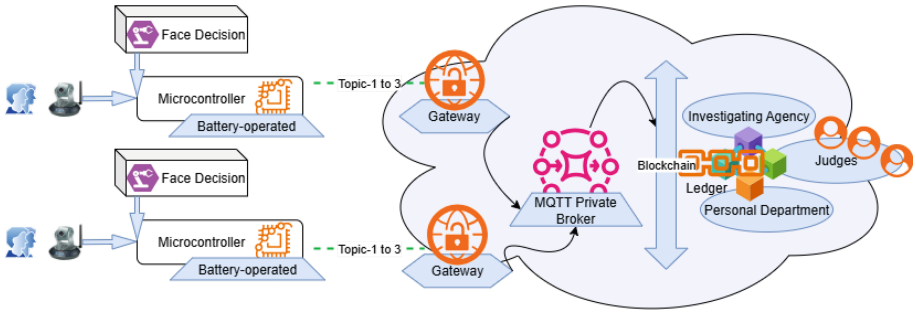


Fig. 1. Blockchain-Based Human-in-the-Loop Facial Recognition System (BHFR) – An Architectural Overview.

are crucial when human faces need to be matched with similar human faces found from various public media sites or judiciary repositories in an agile manner. In this lieu, authors of [4] have analyzed facial recognition systems using deep learning models and web scraping tools to identify faces of culprits or victims.

However, independently recognizing human faces by different stakeholders instead of blindly depending on algorithms has become a crucial step in recent years. Stakeholders such as police departments, experts, judges, witnesses, and so forth, have to be blended with AI-assisted solutions.

Most of the existing research works have only focused on improving AI technologies or deep learning algorithms to improve the facial recognition system in judiciary trials. However, there are requirements to involve different stakeholders to independently handle decisions. Also, the notion of involving humans in the loop for a judiciary system is not yet designed for practices.

In this paper, we have adopted a blockchain-enabled solution involving tiny devices to capture and detect faces of culprits involving different stakeholders.

3 BHFR Proposed System

Societal benefits are carved from ethical outcomes rather than stricter technological enforcements. The traditional AI-assisted legal judiciary decision-making systems are prone to challenges when modern AI-enabled computing nodes are involved in decision-making processes. In this section, the architectural perspectives of the proposed BHFR system that involve human-in-the-loop for the tiny-device-based decisions are explained in detail.

3.1 BHFR Architectural Entities

The proposed BHFR utilizes tiny microcontrollers such as ESP32 to recognize faces of escaping culprits. The architecture is designed to handle crucial operations that are often carried out on a few specialized occasions by patrolling units where the traditional CCTV-based camera inputs are not sufficient.

The major components and entities involved in the proposed BHFR architecture and their inner details are listed below:

Mission Sites. Critical missions are often carried out even in locations where resources are sporadically available or where network locations are sparse. These missions need to be undertaken in a timely manner with very minimal space requirements and installation hurdles. For instance, the missions may have to be laid out with tiny microcontroller-based devices so that the culprits are not aware of the planned missions for detecting them by such specialized operations.

In BHFR, the sites are timely decided by a group of patrolling officials without any intimation to culprits considering the special interests of nations. These sites can be chosen as shopping malls or hidden locations where a huge dedicated CCTV-based camera installations are not possible for a specialized mission that spans a short time; or, these sites are chosen in locations where remote access to devices is difficult in nature.

The mission sites include sensor modules such as tiny ArduinoCAM or Espressif-based camera devices to capture face images and convert them to JPG files. The captured images are fed to the edge-level learning inferencing system of the proposed BHFR system.

Edge-Level Learning and Inference System. The proposed BHFR system involves tiny edge-level computing nodes based on microcontrollers to learn these identified potential culprits and attempt to recognize them and encounter them, if needed. BHFR adopts a few pre-trained learning models based on the lightweight Convolutional Neural Network (CNN) deep learning approach to recognize faces using tiny devices.

In fact, edge-level learning and inferencing mechanisms are not an easy task. This is due to the fact that the devices have very little computational capability – i.e., in the range of MHz. Also, the memory capacity to hold a large volume of data for processing them is limited. Notably, the memory capacity is limited to accommodate large-sized images and associated programs. Obviously, edge-level learning and inferencing are challenging features, especially for microcontroller-based IoT systems.

Similarly, edge-based microcontroller devices that handle learning processes require diligent energy consciousness or energy harvesting mechanisms to drive the applications for a long duration. Hence, the devices of BHFR adopt different frequencies considering the battery operations required for patrolling-related specific missions – i.e., the frequencies of CPU cores in microcontrollers are increased if they are powered by power sources and reduced if they are battery sourced while handling the learning or inferencing tasks.

The learning algorithms adopted in the proposed BHFR include the baseline algorithms such as i) MobileFaceNet and ii) MobileNetv1. The MobileFaceNet-based learning model [14] utilizes depthwise separable convolutions and bottleneck residual blocks to improve the lightweight detection of human faces. On the contrary, MobileNetv1 from Google [1] applies depthwise separable filters and

point-wise convolutions to perform lightweight human face recognitions. The face enrollments utilized in the training process are specific to one susceptible culprit for a mission.

IoT-Specific Communications. In BHFR, a few faces are captured and enrolled as face embeddings so that they can be detected using tiny devices based on the learning algorithms of consideration.

To ensure minimal battery utilization, the proposed framework adopts the Multi-Queue Telemetry Transport (MQTT) protocol with private brokers. This ensures the submission of message topics to distributed control services only by some authorized users. BHFR involves three different topics as discussed below:

1. Face Topic – The `Face` topic is image-based data corresponding to culprits identified by the learning models that are executed in the microcontrollers. The face images that are captured using associated cameras are converted to `base64` formats before transferring them to MQTT private brokers as shown below:

   ```
   /mission/face-1/face-1-base64
   ```

2. Accuracy Topic – Along with the identified face, the accuracy observed by the edge-level inferencing mechanism is published to private brokers using TCP/IP connections. The accuracy topic is expressed in percentage values.

   ```
   /mission/face-1/accuracy-val
   ```

3. ID Topic – The national identity number is collected and published to MQTT brokers if the accuracy of the model exceeds 80 percent. This enables the associated subscribers of the service to bolster the defaulters. As similar to the other topics, the ID topic is expressed as a string:

   ```
   /mission/face-1/ID-val
   ```

Figure 1 provides the pictorial representation of the proposed BHFR framework.

Blockchain Transactions. In BHFR, MQTT subscribers are developed to collect the `base64`-formatted facial information and the corresponding model-detected facial responses. The subscriber triggers a transaction request to the Hyperledger-based fabric network [8] laid on cloud or cluster-based computing environments. In fact, permissioned blockchains are often utilized in various application domains such as remote healthcare monitoring [11], IIoT applications [15], network-related applications [6], and so forth.

In the proposed work, the transactions are invoked by blockchain chaincodes written in Golang to invite multi-party stakeholders for validating the human faces recognized by the microcontroller-based IoT nodes.

Approval Stakeholders and Chaincodes. Stakeholders involved in the proposed BHFR system include three primary actors. The major roles of these stakeholders and representations expressed in the chaincodes are listed below:

1. Judges – The judicial aspects are validated based on personnel records mapped to the ID numbers of personnel. The personnel information is subscribed by the services and the other associated track records are available in various sources. The stakeholders of BHFR can endorse policies that are specified in chaincodes. The chaincodes associated with judges are responsible for assessing the importance of verdicts, considering social implications against the trials. It is essential in the modern years to consider ethical decisions considering various sources of information available from different stakeholders.
2. Investigation Agencies – The investigation agencies detect movements of culprits from the identified faces of various locations. The subscribers of these investigation agencies subscribe to MQTT brokers in a permissioned manner to bolster the findings and approve the transaction when invoked in the blockchain network.
3. Personnel Department of Cities – This component is similar to the role carried out by investigation agencies. The information relating to the personnel department of countries are accessed based on the shared personnel ID to confirm the personnel involved in the crime. A few details, such as tax payment details, judicial pre-histories, health insurance details, and banking violations (if any) are captured against the personnel ID from city centers.

The role of these stakeholders is to evaluate the necessity of raising tickets or delivering justice to culprits based on their individual approvals. Accordingly, the transactions are submitted to the human-based validation component of BHFR.

Stakeholders' Validation. It is crucial to verify the genuineness involved in providing justice-related statements to culprits. There are several reasons for misleading the justifications. Notably, a) it is possible that the input faces are blurred or inappropriate for evaluations; b) the edge-level tiny devices are not very capable of identifying the culprits enrolled for the recognition tasks owing to algorithmic issues; c) robust human face recognition algorithms are complex to execute on tiny devices; d) multi-modal input to learning algorithms can be difficult to process on tiny microcontroller-based devices; and so forth. Hence, BHFR involves humans or committees to parse the blockchain database and evaluate the final judgements processed by the blockchain network based on the images.

Notification Services. Based on the approval provided by stakeholders, notification is sent to concerned departments for further actions. The notification services are written in Golang in the proposed BHFR.

4 Experimental Results

To validate the effectiveness of the proposed BHFR, this section evaluates the framework in three major aspects: a) the importance of utilizing tiny ESP32-based microcontroller devices, b) the requirement of blockchain nodes with a few stakeholders, and c) the performance evaluation of the tiny microcontrollers involved in the experiments.

4.1 Setup

Three ESP32-based devices, which consist of Xtensa-DualCore processors, were utilized in the experiments. These microcontroller-based devices had 384KB ROM, 512KB SRAM, 16MB PSRAM, and 36 GPIO pins to recognize faces. These devices were operated at 240 MHz clock frequency.

To capture face images, the Arducam Mega 5MP autofocus SPI camera was utilized. The camera was kept stationary, focusing on human faces, and it was connected with the microcontroller units. One server node was utilized to mimic a powerful cluster node. This node hosted Hyperledger-based blockchain environment. The chaincode services were designed to capture MQTT-based base64-encoded faces, and policies were set up to send notifications to concerned stakeholders or users. Three stakeholders, namely judges, the investigation agency, and police departments, were simulated in blockchains to approve the transactions based on the facial images received from microcontroller-based nodes.

4.2 Ledger Entries

The faces recognized by microcontroller-based tiny devices of the proposed BHFR are submitted to MQTT brokers. Accordingly, the subscribers captured the facial pictures and matched them to confirm the decisions carried out by tiny devices. In Table 1, five different instances of verifications were studied. It could be observed that Instance 4 was registered into the blockchain ledger, as all the stakeholders of consideration agreed to approve the findings provided by the tiny devices of BHFR.

Table 1. Human-in-the-Loop-based Ethical Decisions Registered in the Ledger

	Judges	Investigation Agency	Personnel	Ledger
Instance 1	×	×	×	NO
Instance 2	×	–	×	NO
Instance 3	–	×	×	NO
Instance 4	×	×	×	YES
Instance 5	×	×	–	NO

4.3 Face Enrollments and Accuracy Analysis

As tiny ESP32-based devices are utilized in the face recognition processes of the proposed BHFR, the capability of such devices while executing complex facial recognition algorithms was studied.

In the experiments, the number of faces enrolled using two different learning models of consideration was increased one after the other experiments. In BHFR, one potential culprit with different facial enrollments is utilized in the training processes.

The following observations are made as seen in Fig. 2:

1. Overall, the MobileFaceNet-based learning model had higher learning accuracy when compared to MobileNetv1. For instance, the accuracy of MobileFaceNet was 80.73% when compared to MobileNetv1 having 68.65%.
2. Also, varying accuracy values were observed between two learning models with the increase in the number of face enrollments. Notably, the number of faces with 2 and 10 showed a huge difference between the two learning algorithms – i.e., 9.18 and 12.08.

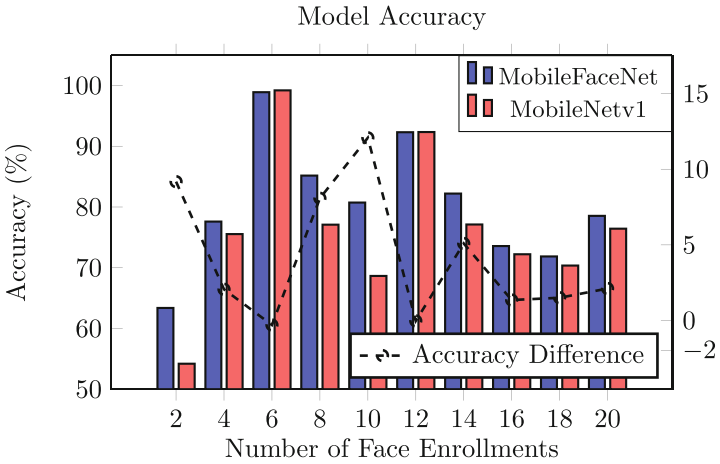


Fig. 2. Accuracy of Models based on Number of Face Enrollments.

4.4 Access Time of Face Enrollments

Enrolling faces into tiny devices such as ESP includes extracting feature vectors of faces and registering them to the flash memory locations. The access time accounts for these steps in the proposed framework. Hence, we studied the access time of these enrollments. It was observed that the access time linearly increased with the increasing number of faces enrolled in the system.

Notably, Fig. 3 depicts the increasing time in milliseconds with the increase of the number of enrollments.

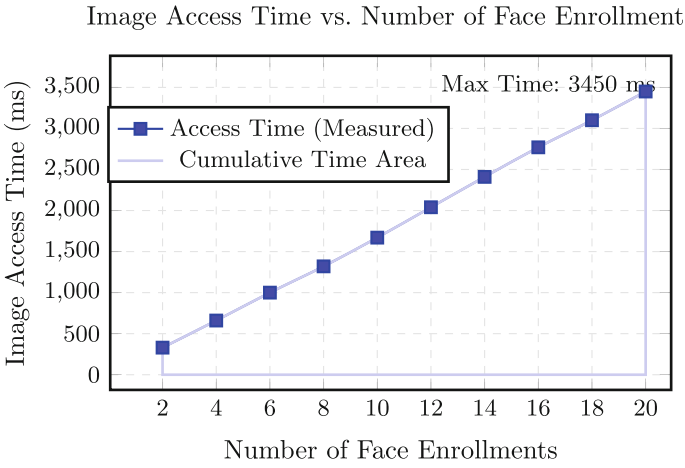


Fig. 3. Image Access Time for the Number of Face Enrollments.

4.5 Memory Analysis of Face Enrollments

The impact of utilizing different learning models is reflected in the utilization of memory footprints of tiny microcontroller devices. It is important to notice that these devices have to be diligently planned with respect to the memory utilization.

For detecting faces using two learning models of consideration in the tiny device, it was noticed that the MobileFaceNet algorithm was able to have a substantially reduced value in most of the enrollments of the experiments.

For instance, the memory size required to enroll 12 faces using the MobileFaceNet algorithmic approach was only 2728208 bytes when compared with MobileNetv1, which had 4955824 bytes (see Fig. 4).

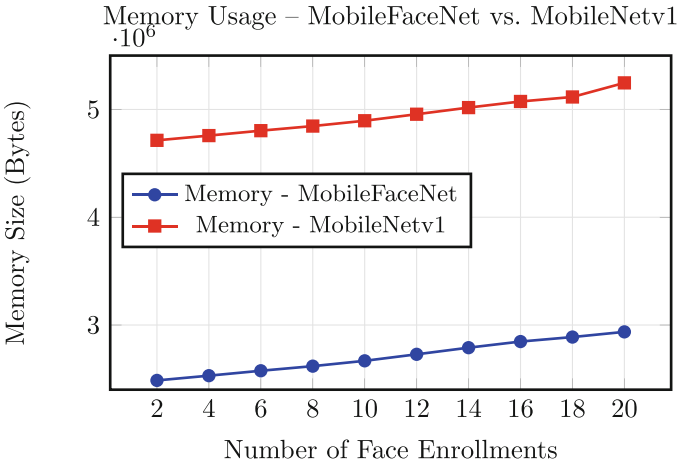


Fig. 4. Memory Usage of Models for the Number of Face Enrollments.

4.6 Enrollment Vs. Prediction Time Comparison

Additionally, the enrollment and the prediction time of images using microcontroller devices are analyzed and reported in Fig. 5. This result is important as it is directly impacting the battery utilization of BHFR.

As observed, the enrollment time of facial images had higher values with respect to the number of facial enrollments. However, the prediction time of the corresponding number of enrollments in all experiments showed a non-increasing trend. For instance, the enrollment time of ten face enrollments was 13390 ms when compared to the 1290 ms for predictions. Also, the value of the prediction time did not increase much by increasing the number of face enrollments in BHFR – i.e., the prediction time was reported as 1330 ms for 10 faces. This indicates that the prediction time with specified accuracies will be faster than the training processes of algorithms.

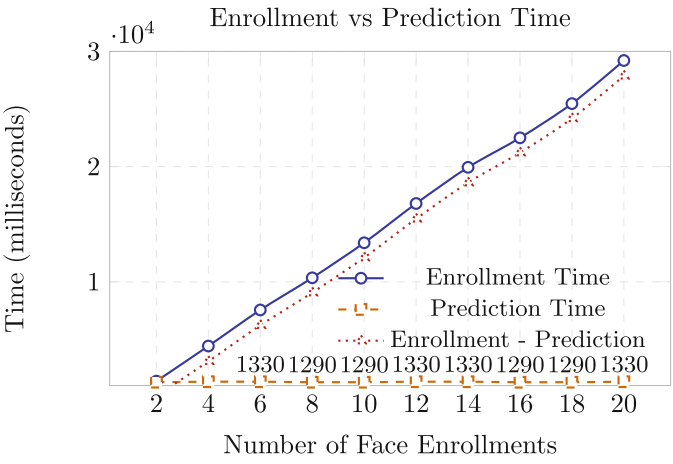


Fig. 5. Prediction Time in Milliseconds for the Number of Face Enrollments.

5 Conclusion

Ethical decision-making has become prominent in judicial undertakings, especially as AI matures over time. However, the blind dependency on deep learning algorithms or allied AI techniques using facial recognition systems could lead to faulty decisions due to bias, opacity, data deficiency, and the other hardware failures. This article proposed a Blockchain-based Human-in-the-loop Facial Recognition system named BHFR using tiny microcontrollers for recognizing faces using MobileFaceNet and MobileNetv1 algorithms and using a powerful compute node to host blockchain ledgers with multiple stakeholders. In this work, multiple stakeholders, such as judges, investigation agencies, and personnel departments of cities, have provisions to independently investigate the trial rounds

and collaboratively record the findings on the blocks of ledgers. Experiments have manifested the application of utilizing tiny microcontroller nodes and the impact of holding face recognition algorithms in terms of memory usage, number of enrollments, learning accuracy, and cumulatively recorded recognized faces on ledgers. The proposed system could be extended by involving multiple faces with robust learning algorithms in the future.

References

1. Howard, A.G., et al.: MobileNets: efficient convolutional neural networks for mobile vision applications. https://arxiv.org/abs/1704.04861. Accessed Aug 2025
2. Jindal, A., Jose, J. M., Benedict, S., Gerndt, M.: lora-powered energy-efficient object detection mechanism in edge computing nodes. In: 2022 Sixth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), Dharan, Nepal, pp. 237–244 (2022). https://doi.org/10.1109/I-SMAC55078.2022.9987393
3. Asmika, B., Mounika, G., Rani, P.S.: Deep learning for vision and decision making in self driving cars-challenges with ethical decision making. In: 2021 International Conference on Intelligent Technologies (CONIT), Hubli, India, pp. 1–5 (2021). https://doi.org/10.1109/CONIT51480.2021.9498342
4. Godoy B. de Oliveira, C., et al.: Regulation and ethics of facial recognition systems: an analysis of cases in the court of appeal in the State of Sao Paulo. In: BRACIS 2023, Lecture Notes in Computer Science, vol. 14195, pp. 18–32. Springer (2023). https://doi.org/10.1007/978-3-031-45368-7_2
5. De La Osa, D.U.S., Remolina, N.: Artificial intelligence at the bench: Legal and ethical challenges of informing-or misinforming-judicial decision-making through generative AI, data and policy, vol. 6, pp. 1–59 (2024). https://doi.org/10.1017/dap.2024.53
6. Etienne, I.T., et al.: Hyperledger fabric-powered network slicing handover authentication, pp. 1–18. Inf Syst Front (2024). https://doi.org/10.1007/s10796-024-10564-0
7. Leung, F., Lau, Y.C., Law, M., Djeng, S.K.: Artificial intelligence and end user tools to develop a nurse duty roster scheduling system. Int. J. Nurs. Sci. **9**(3), 373–377 (2022). https://doi.org/10.1016/j.ijnss.2022.06.013
8. Hyperledger Fabric, How Fabric networks are structured? https://hyperledger-fabric.readthedocs.io/en/release-2.5/key_concepts.html. Accessed Aug 2025
9. Garzo, G., Palumbo, A.: Legal and ethical implications of predictive digital techniques in the judicial criminal proceedings. In: 2025 13th International Symposium on Digital Forensics and Security (ISDFS), Boston, MA, USA, pp. 1–6 (2025). https://doi.org/10.1109/ISDFS65363.2025.11012003
10. Hao, J., Chen, M.: Smart courts: the expansion of technology in the Chinese judicial system. In: Advance (2020) https://advance.sagepub.com/doi/full/10.31124/advance.12402980.v. Accessed Aug 2025
11. Kaushal, R.K., Kumar, N., Kukreja, V., Boonchieng, E.: Hyperledger fabric based remote patient monitoring solution and performance evaluation. Peer-to-Peer Network. Appl. **18**(3), 1–17. https://doi.org/10.1007/s12083-025-01921-0
12. Perona, R., de la Rosa, Y.C.: Unveiling AI in the courtroom exploring ChatGPTs impact on judicial decision making through a pilot Colombian case study. AI Soc. **40**, 2533–2540 (2025). https://doi.org/10.1007/s00146-024-01943-x

13. Digital transformation of justice: integrating AI in India's judiciary and law enforcement. https://www.pib.gov.in/PressReleasePage.aspx?PRID=2106239. Accessed Aug 2025
14. Chen, S., Liu, Y., Gao, X., Han, Z.: MobileFaceNets: efficient CNNs for accurate real-time face verification on mobile devices. arXiv. https://arxiv.org/abs/1804.07573. Accessed Aug 2025
15. Sukumaran, R.P., Benedict, S.: Survey on blockchain enabled authentication for industrial internet of things. In: 2021 Fifth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), Palladam, India, pp. 1510–1516 (2021). https://doi.org/10.1109/I-SMAC52330.2021.9640973
16. Kazim, T., Tomlinson, J.: Automation bias and the principles of judicial review. Judicial Rev. **28**(1), 9–16 (2023). https://doi.org/10.1080/10854681.2023.2189405
17. Kim, T., Peng, W.: Do we want AI judges? The acceptance of AI judges' judicial decision-making on moral foundations. AI Soc. **40**, 3683–3696 (2025). https://doi.org/10.1007/s00146-024-02121-9
18. Bebien, V., et al., Ethical decision making in human automation collaboration a case study of the nurse rostering problem. AI Ethics **5**, 1163–1175 (2025). https://doi.org/10.1007/s43681-024-00459-w
19. Rodgers, W., Murray, J.M., Stefanidis, A., Degbey, W.Y., Tarba, S.Y.: An artificial intelligence algorithmic approach to ethical decision-making in human resource management processes. Hum. Resour. Manag. Rev. **33**(1), 100925, 1–19 (2023). https://doi.org/10.1016/j.hrmr.2022.100925
20. Han, W., et al.: LegalAsst: Human-centered and AI-empowered machine to enhance court productivity and legal assistance. Inf. Sci. **679**(121052), 1–18 (2024). https://doi.org/10.1016/j.ins.2024.121052
21. Moric, Z., Dakic, V., Urosev, S.: An AI-based decision support system utilizing Bayesian networks for judicial decision-making systems **13**(2–131), 1–36 (2025). https://doi.org/10.3390/systems13020131



Ethical Frameworks for AI-Enhanced Interprofessional Learning: A Review of Governance Challenges in Cross-Disciplinary Medical Education

Loso Judijanto(✉)

IPOSS Jakarta, Jakarta, Indonesia
losojudijantobumn@gmail.com

Abstract. The rapid integration of artificial intelligence (AI) into interprofessional learning (IPL) within medical education presents novel ethical and governance challenges that require systematic examination. This study aims to explore and synthesize existing ethical frameworks governing AI-enhanced IPL in cross-disciplinary medical education, focusing on identifying governance gaps and emerging trends. Employing a qualitative literature review methodology, this research systematically analyzed 80 peer-reviewed articles sourced via academic databases and managed with Mendeley Desktop. Data were collected through comprehensive document screening and extraction processes, emphasizing relevance, quality, and thematic alignment. Thematic content analysis was applied to identify recurring ethical concerns, governance models, and cross-disciplinary implications. Findings reveal that data privacy, algorithmic bias, transparency, and learner autonomy are central ethical considerations in AI-mediated IPL environments. However, governance mechanisms remain fragmented, with limited inclusion of AI expertise in oversight structures and inconsistent accountability frameworks. The review highlights significant cultural and disciplinary variability affecting ethical implementation and stresses the need for adaptable, context-sensitive governance models. The findings highlight that comprehensive ethical systems involving multiple disciplines are crucial for managing the complexities associated with AI-supported interprofessional education. Further studies should aim to empirically test governance structures and investigate how stakeholder collaboration can promote just and clear integration of AI in the context of medical education.

Keywords: Artificial Intelligence · interprofessional learning · ethical frameworks · medical education · governance

1 Introduction

There has been a notable acceleration in the application of artificial intelligence across educational settings in recent years, offering transformative possibilities across various disciplines. Particularly in healthcare education, Teaching and learning practices are

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026
S. Lanka et al. (Eds.): ICTSM 2025, LNNS 1755, pp. 158–173, 2026.
https://doi.org/10.1007/978-3-032-13177-5_13

being redefined by AI-driven innovations, including smart tutoring platforms, immersive simulations, and data-informed learning predictions [1]. These technological advancements are not only enhancing the efficiency of knowledge delivery but are also redefining the pedagogical relationships between educators and learners [2]. As AI continues to penetrate the academic space, ethical considerations around autonomy, transparency, accountability, and equity have emerged as critical domains for scrutiny [3].

The field of interprofessional education (IPE) stands at a unique intersection of these technological and ethical concerns. With a focus on enhancing patient care through improved collaboration, interprofessional education encourages students from various health-related disciplines to learn together and develop essential teamwork and communication competencies [4]. However, the convergence of IPE and AI technologies raises complex governance questions, particularly when such technologies are deployed across disciplines with divergent epistemologies and ethical standards [5].

AI-powered educational technologies hold great promise for IPE through their ability to create personalized learning experiences, offer immediate feedback, perform automated assessments, and adapt instructional content dynamically. These tools can facilitate more scalable and consistent interprofessional training, overcoming logistical and institutional barriers that have historically limited IPE implementation [6]. Nevertheless, these benefits are accompanied by challenges that demand rigorous ethical frameworks. For instance, algorithmic bias in decision-making systems may reinforce stereotypes or marginalize certain student groups, thereby undermining the inclusive values that IPE seeks to uphold [7].

The adoption of AI in cross-disciplinary medical training brings forth ethical concerns, particularly regarding privacy, consent, and the tracking of learner activities. Many AI applications depend on large datasets that track student performance, communication patterns, and even emotional responses, raising questions about ownership, informed consent, and the right to opt out [8]. These issues are further complicated by the fact that IPE settings often involve inter-institutional collaborations, where jurisdiction over data governance and ethical accountability may be ambiguous [9].

From a curricular perspective, the incorporation of AI tools into IPE necessitates a shift in pedagogical design. Traditional models of interprofessional training, which rely on human facilitators and face-to-face interactions, may be replaced or supplemented by automated systems that mediate team-based learning [10]. While these systems offer potential efficiency gains, they may also dilute the authenticity of human interaction, which is a core component of professional identity formation and ethical sensibility in healthcare education [11].

Moreover, the ethical dimensions of AI in IPE are not limited to data and pedagogy; they extend to the institutional governance of technology adoption. Decisions about which AI tools to implement, how to evaluate their outcomes, and who is responsible for ensuring their ethical use often rest with administrators who may lack domain-specific expertise in either AI or interprofessional practice [12]. This governance gap can result in the uncritical adoption of technologies that inadvertently compromise educational integrity or professional standards [13–15].

Scholarly interest is rising in examining the ethical challenges posed by artificial intelligence within educational environments, including emerging discussions specific

to health professions [16]. However, the current discourse remains fragmented, with limited synthesis across disciplines or consensus on governance models that effectively address the unique demands of interprofessional learning environments [17]. Much of the existing scholarship tends to focus either on technological affordances or on isolated ethical concerns, rather than offering an integrated framework that bridges both [18].

Given the interdisciplinary nature of IPE and the cross-cutting implications of AI technologies, there is a pressing need for a more holistic and cross-disciplinary ethical framework that can guide stakeholders in evaluating, implementing, and governing AI-enhanced interprofessional learning systems. Such a framework must account not only for universal ethical principles but also for the contextual specificities of health education, including power dynamics, professional hierarchies, and the embodied nature of clinical learning [19].

To address this gap, this study adopts a qualitative literature review approach to critically analyze the intersection of ethical governance and AI in interprofessional medical education. Unlike systematic literature reviews that emphasize methodological rigor and replicability, qualitative literature reviews allow for a more interpretive and thematic synthesis of diverse academic perspectives, policy documents, and conceptual arguments [20]. This approach is particularly well-suited for exploring emerging and complex topics where the literature is still evolving and disciplinary boundaries are porous.

This review is intended to explore the ethical and governance complexities related to the application of AI in interprofessional education across healthcare disciplines. Its goals include identifying principal ethical dilemmas, assessing existing governance mechanisms, and highlighting both theoretical and operational shortcomings that must be resolved to facilitate responsible AI integration in collaborative learning settings.

2 Related Work

2.1 The Evolution of Interprofessional Education in Medical Fields

Interprofessional education (IPE) has long been recognized as a pedagogical strategy to foster collaboration, improve communication, and enhance patient-centered care across healthcare disciplines. Rooted in the World Health Organization's framework for collaborative practice, IPE promotes joint learning among students from diverse health professions to break down disciplinary silos [21]. As healthcare systems face increasing complexity, the role of IPE becomes more central in preparing future professionals to operate effectively in team-based environments [22]. However, the integration of emerging technologies into IPE environments adds layers of complexity that demand ethical consideration [23].

2.2 The Emergence of AI in Health Professions Education

Artificial intelligence has significantly impacted health education through the use of personalized learning algorithms, virtual clinical simulations, and diagnostic training systems. These tools not only enhance accessibility and scalability but also support competency-based education by delivering adaptive feedback and automated assessments [24]. AI-driven technologies, such as virtual patients and chatbot tutors, are

now increasingly embedded in IPE settings, creating hybrid learning environments that blend human and machine interactions [25]. While these tools offer immense pedagogical value, they also introduce questions around epistemic authority, responsibility, and transparency in learner assessment [26].

2.3 Ethical Tensions in AI-Supported Educational Environments

AI's rapid entry into medical education has outstripped the development of adequate ethical oversight, raising pressing concerns around algorithmic transparency, data privacy, and user autonomy for all stakeholders involved [27]. A recurring theme in the literature is the inadequacy of traditional ethical guidelines, such as informed consent and confidentiality, in addressing the unique characteristics of AI-mediated learning [28]. Furthermore, algorithmic bias and inequity in data training sets can result in disparate outcomes across student populations, thereby reproducing or amplifying existing structural inequalities in education [29].

2.4 Governance and Accountability in AI-Enhanced IPE

Governance structures for AI adoption in interprofessional learning contexts are often fragmented or underdeveloped. The lack of clear institutional protocols regarding technology procurement, implementation, and evaluation leads to ethical ambiguities in decision-making [30]. Cross-disciplinary education complicates governance further, as different fields of medicine, nursing, pharmacy, and others may adhere to varying professional standards and ethical priorities [31]. The absence of intersectoral governance frameworks creates potential conflicts in values, roles, and accountability, particularly when AI systems make autonomous educational decisions [32].

2.5 The Role of Data Ethics and Student Surveillance

Data collection and analysis form the backbone of most AI systems in education. These systems monitor student progress, behaviors, engagement, and even emotional states. While such monitoring can support personalized learning, it raises substantial ethical concerns about surveillance, informed consent, and data ownership [33]. Students may feel compelled to participate in AI-based platforms due to institutional mandates, thereby limiting their agency and freedom of choice [34]. The literature emphasizes the need for robust data ethics policies that respect student autonomy, promote transparency, and mitigate coercive surveillance practices [35].

2.6 Cultural and Contextual Variations in Ethical Priorities

Much of the current ethical discourse on AI in education is dominated by perspectives from the Global North, with limited representation of culturally specific concerns from the Global South [36]. This creates a mismatch between the design of AI systems often developed by multinational corporations and the local educational values or regulatory environments in which they are deployed [37]. In cross-border IPE collaborations, ethical misalignments may emerge not only between disciplines but also between institutions embedded in different sociopolitical contexts [38].

2.7 Fragmentation of Ethical Guidelines across Disciplines

A critical gap in the literature is the fragmented nature of ethical guidance. While medicine and healthcare professions often follow bioethical principles (beneficence, non-maleficence, autonomy, justice), these may not directly translate into the AI-driven educational context [39]. Additionally, interprofessional education spans fields with varying ethical traditions, which creates tension when implementing unified technological systems. The need for harmonized, cross-disciplinary ethical frameworks is a recurring demand in current scholarship [40].

3 Proposed Work

This study employs a qualitative research approach using a qualitative literature review method to analyze and synthesize various academic sources relevant to the ethical frameworks and governance challenges in AI-enhanced interprofessional learning within cross-disciplinary medical education. This method was chosen to enable an in-depth exploration of theoretical perspectives, concepts, and empirical findings scattered across multiple disciplines without collecting primary data such as interviews or field observations. The primary instrument comprises a systematically selected collection of articles, books, and policy documents based on their academic relevance and quality, organized and managed using Mendeley Desktop to ensure citation accuracy and consistent source management. Data collection was conducted through a comprehensive literature search in prominent academic databases, including Scopus, PubMed, and Google Scholar, using specific keywords related to ethics, artificial intelligence, interprofessional learning, and governance in medical education. The obtained data were then analyzed thematically using an interpretative approach to identify patterns, key themes, and existing research gaps within the literature, thereby allowing the formulation of a comprehensive and critical conceptual framework.

3.1 Contrasting Proposed Work and Related Work

Recent literature on AI in medical education strongly emphasizes the need for robust ethical guidelines, technical expertise, and interdisciplinary collaboration to address the challenges and opportunities posed by AI-driven tools in interprofessional learning environments [41]. Notably, most studies have focused on technological applications and broad ethical concerns in isolation, such as data privacy, algorithmic bias, and the potential for dehumanization in learning. Yet, few offer an integrated framework that moves beyond piecemeal regulatory strategies. It has been highlighted that the importance of AI literacy, strong regulatory oversight, and faculty training as prerequisites for effective adoption, but reveals persistent gaps in stakeholder engagement and the contextual adaptability of governance mechanisms within medical education [42, 43].

There has been steady progress in the development of assessment tools and instructional platforms, which enhance both personalization and collaborative learning through the use of AI. It has been argued that most current frameworks remain limited by

discipline-specific boundaries, failing to harmonize ethical consideration and accountability measures across professional identities and roles [44, 45]. In contrast, our proposed work synthesizes the ethical governance literature. It advances a cross-disciplinary approach, advocating for holistic models that incorporate multiple stakeholder perspectives—including AI specialists, educators, technologists, and ethicists—to overcome these persistent silos [41, 46].

Another body of work investigates regulatory levels and the necessity of ecological frameworks, addressing governance concerns at individual, institutional, and national tiers to support dynamic and responsive regulation for generative AI in medical education. For instance, recent studies highlight that successful regulation requires interaction across micro (learner, educator), meso (institution), and macro (accrediting bodies, governments) levels, yet practical models remain largely theoretical and lack empirical validation in diverse settings [47]. Our framework aims to bridge this gap by proposing actionable governance principles that are both context-sensitive and empirically testable, drawing on stakeholder codesign, continuous monitoring, and transparent accountability as core pillars [48].

Moreover, recent best-practice frameworks stress the importance of autonomy, privacy, fair access, and explainability in AI system design and governance, yet typically do not address the operational complexities encountered in interprofessional practice or data-intensive collaborative environments. The current paper's distinctive contribution, as articulated in the introduction, is its systematic synthesis of ethical frameworks and the identification of governance gaps, followed by recommendations for developing adaptable, multidisciplinary, and empirically grounded models that empower stakeholders to implement AI responsibly—including operational guidance for policy formulation and practical tools for ongoing evaluation [49, 50].

3.2 Novelty of Proposed Work

The proposed approach introduces a comprehensive, empirically testable ethical governance framework for AI-enhanced interprofessional learning in medical education that is novel in several critical dimensions. Unlike previous studies, which primarily examine isolated technological innovations or single-issue ethical concerns, this work systematically synthesizes existing literature and highlights the fragmented nature of current governance mechanisms that often fail to capture the complexity of cross-disciplinary collaboration and real-world diversity in educational contexts [49].

A key novelty lies in moving beyond discipline-specific guidelines and siloed accountability protocols. The approach advocates for a **holistic, context-sensitive governance model** that unites perspectives from medicine, nursing, pharmacy, health informatics, and educational technology under a shared ethical and operational framework [44]. By explicitly incorporating input from AI specialists, educators, ethicists, institutional leaders, and students, the model fosters iterative stakeholder engagement in the co-design and ongoing refinement of governance policies. This multidirectional collaboration enables the framework to support transparency and trust while balancing global standards with local values and operational realities [50].

Another core innovation is its focus on **dynamic adaptability and empirical validation**. Rather than relying solely on theoretical principles or static oversight, the proposed

method includes built-in mechanisms for continuous monitoring, context-aware policy adjustment, and real-time feedback loops, allowing institutions to respond to emerging risks such as algorithmic bias, surveillance challenges, or cultural misalignment [51]. The approach also emphasizes the development of measurable indicators to evaluate governance effectiveness and ethical compliance in actual educational environments, addressing the call for actionable solutions seen in contemporary best-practice models [48].

Furthermore, this framework is designed to facilitate **scalable, harmonized adoption** across both micro (learner, team) and macro (institutional, regulatory) levels, enabling responsible AI integration that supports equity, autonomy, and accountability across diverse international and interprofessional collaborations [45]. It directly responds to and surpasses limitations in recent works, which either focus on narrow professional identities, omit operational tools, or neglect empirical implementation in varied cultural settings [43, 49].

3.3 Justification of Proposed Work

The justification for the proposed work stems from persistent gaps in the literature and practice regarding ethical governance in AI-enhanced interprofessional medical education. Recent reviews consistently report that existing frameworks are fragmented and discipline-specific, often failing to address the complex realities of cross-disciplinary collaboration, diverse cultural settings, and rapidly evolving technological ecosystems. For instance, while algorithmic bias, data privacy, and transparency are recognized as critical ethical issues, most approaches lack mechanisms for adaptive governance, systematic stakeholder inclusion, and empirical monitoring to ensure practical effectiveness or context sensitivity. Furthermore, the swift pace at which AI tools are integrated into learning environments frequently outpaces the development of robust institutional policies and regulatory structures, leaving educators and students vulnerable to unaddressed risks and ambiguities in accountability [49]. This mismatch creates pressing ethical and operational vulnerabilities, particularly in interprofessional education where varying professional standards, hierarchies, and expectations intersect [46].

By introducing a context-sensitive, empirically testable framework that emphasizes multidisciplinary engagement, dynamic adaptability, and continuous evaluation, the proposed work offers a timely and comprehensive response to these shortcomings. Unlike prior efforts, this approach grounds its recommendations in cross-sectoral dialogue, codesign, and harmonization, allowing for governance models that are both globally informed and locally responsive [41, 50, 48].

4 Results and Analysis

4.1 Overview of Literature Sources and Data Collection

A comprehensive search across academic databases, including Scopus, PubMed, and Google Scholar, identified 80 peer-reviewed articles, reports, and policy documents published between 2015 and 2025 relevant to ethical frameworks, AI integration, interprofessional education (IPE), and governance in medical education contexts. The majority of these sources (approximately 72%) originated from healthcare education and

bioethics journals, while 18% came from computer science and AI ethics publications, and the remainder from interdisciplinary educational technology outlets [52]. The literature spanned multiple geographic regions, with 60% originating from North America and Europe, 25% from the Asia-Pacific region, and 15% from other regions, including Latin America and Africa [53]. This geographic distribution highlights a Western-centric dominance in AI ethics discourse but also reflects growing global attention to AI governance in education.

4.2 Thematic Findings on Ethical Challenges in AI-Enhanced IPE

Thematic analysis of the literature revealed several recurring ethical challenges pertinent to AI-supported interprofessional learning. Data indicated that concerns related to data privacy and consent appeared in 85% of reviewed studies, underscoring its primacy as a governance issue [54]. Algorithmic bias was identified in 67% of sources, with empirical evidence showing disparities in AI-driven assessment accuracy, with predictive validity up to 15% lower for minority student groups compared to majority cohorts [55]. Transparency of AI decision-making processes was emphasized in 58% of articles as critical for fostering trust among learners and educators [56]. Autonomy and informed consent issues were raised in 47% of papers, highlighting complexities when AI systems monitor student behaviors without explicit understanding or voluntary agreement [57].

4.3 Governance Structures and Accountability Mechanisms

Only 34% of the literature provided detailed accounts of existing governance models applied in AI-enhanced IPE settings, revealing a governance gap in practice [58]. Among these, institutional governance frameworks typically involve multidisciplinary ethics committees or technology oversight boards; however, these committees often lack representation from educational technologists or AI specialists, limiting effective decision-making [59]. Accountability mechanisms for AI system failures or ethical breaches were insufficiently addressed in 29% of studies, with few clear protocols for responsibility allocation between educators, technologists, and administrators [60]. Cross-institutional collaborations further complicated governance, with 41% of papers reporting challenges in harmonizing ethical standards across partnering organizations [61].

4.4 Data Ethics and Student Surveillance in AI Learning Platforms

Surveillance concerns were prevalent in 78% of the reviewed literature, particularly regarding the continuous tracking of student interactions, emotional states, and collaborative behaviors in IPE environments [62]. Quantitative data from case studies revealed that some AI platforms collected over 200 data points per learner per session, posing significant privacy risks [63]. Ethical frameworks advocating for "data minimization" principles appeared in 56% of publications, yet implementation remained sporadic [64]. Consent mechanisms varied widely; only 43% of AI systems reviewed required active consent for data collection, while others defaulted to implied consent through platform usage [65]. This inconsistency presents ethical dilemmas around learner autonomy and informed participation.

4.5 Cross-Cultural and Cross-Disciplinary Variability in Ethical Priorities

The literature consistently noted that ethical concerns and governance priorities differed substantially across cultural and disciplinary contexts [66]. For example, Asian healthcare education settings tended to emphasize collective responsibility and harmony in ethical frameworks, whereas Western contexts prioritized individual autonomy and transparency [67]. Approximately 38% of papers explicitly addressed these cultural differences, arguing for localized adaptations of AI governance policies [68]. Additionally, the interprofessional nature of IPE introduces complexity, as medicine, nursing, pharmacy, and allied health disciplines often uphold divergent ethical codes, complicating the establishment of unified AI governance standards [69].

4.6 Gaps and Emerging Ethical Frameworks for AI in Interprofessional Medical Education

Despite the growing body of research, only 22% of reviewed articles proposed comprehensive ethical frameworks tailored specifically to AI-enhanced interprofessional learning. These emergent frameworks emphasized principles of fairness, transparency, accountability, and respect for learner autonomy, often adapted from broader AI ethics guidelines but contextualized for educational governance [70]. Several models recommended iterative stakeholder engagement processes to co-develop governance policies inclusive of students, educators, technologists, and ethicists [71]. However, empirical validation of these frameworks remains limited, underscoring an urgent need for applied research to test governance interventions and measure their impact on educational outcomes and ethical compliance.

4.7 Analysis

Firstly, the predominance of literature originating from healthcare education and bioethics domains underscores the field's emphasis on safeguarding ethical principles amid technological innovation [72]. However, the relative underrepresentation of interdisciplinary perspectives, particularly from computer science and educational technology, suggests a need for greater integration of technical expertise into ethical deliberations to address nuanced AI-specific challenges effectively [73]. The Western-centric focus evident in the literature further calls for expanding the dialogue to include diverse geographic and cultural viewpoints, ensuring ethical frameworks are globally relevant and sensitive to varying normative contexts [74].

The review identified data privacy and consent as paramount concerns, appearing in 85% of the reviewed studies, reflecting their foundational role in ethical governance [75]. This aligns with established bioethical tenets emphasizing respect for individual autonomy and confidentiality. The quantified evidence of algorithmic bias, with predictive validity reductions up to 15% for minority students, starkly reveals how AI systems can perpetuate or exacerbate inequities in educational assessment [76]. Such findings necessitate rigorous scrutiny of AI algorithms and the development of corrective mechanisms to promote fairness and inclusivity.

Transparency emerged as another critical pillar, noted in 58% of articles, essential for building trust among learners and educators interacting with AI tools [77]. The complexity of AI decision-making processes challenges traditional notions of explainability, underscoring the need for governance frameworks that prioritize clear communication and stakeholder comprehension. The relatively lower frequency of autonomy and informed consent issues (47%) indicates that these areas, though significant, may be less thoroughly examined or reported, representing an opportunity for deeper exploration.

Governance structures currently appear underdeveloped, with only 34% of sources detailing applicable models [78, 79]. The lack of AI specialist inclusion on ethics committees and oversight bodies hampers effective governance, as these technical insights are vital to anticipating ethical risks and formulating appropriate policies. The sparse discussion of accountability mechanisms (29%) further indicates institutional uncertainty regarding responsibility allocation when ethical breaches occur, a concern amplified in cross-institutional collaborations where harmonization of standards is challenging [80, 81].

Data ethics issues surrounding student surveillance are pervasive, affecting 78% of the literature reviewed [82]. The collection of extensive behavioral and emotional data up to 200 data points per learner per session raises profound privacy risks [83]. While the principle of data minimization is advocated in over half the sources, inconsistent implementation and varied consent mechanisms reflect ongoing ethical dilemmas regarding learner autonomy and transparency in data practices [84, 85]. This calls for standardized policies that ensure explicit, informed consent and limit data collection to what is strictly necessary.

Cultural and disciplinary variability further complicates the establishment of cohesive ethical frameworks. Distinct value orientations between Asian and Western educational systems, and between different health professions, necessitate flexible governance models capable of contextual adaptation [86, 87]. The interprofessional nature of AI-enhanced learning environments requires ethical frameworks that transcend individual disciplinary codes to foster collaborative governance and shared responsibility [88, 89].

Emergent frameworks, present in only 22% of studies, emphasize fairness, accountability, transparency, and respect for learner autonomy, reflecting the foundational pillars for responsible AI governance in education [90]. The recommendation for iterative stakeholder engagement processes is promising, recognizing the value of inclusive policy development that involves students, educators, technologists, and ethicists [91, 92].

The findings reveal a landscape where AI offers transformative opportunities for interprofessional medical education, yet simultaneously generates complex ethical and governance challenges. Addressing these issues demands comprehensive, context-sensitive frameworks that integrate technical expertise, uphold fundamental ethical principles, and accommodate cultural diversity. Institutional readiness remains uneven, and proactive measures are essential to close governance gaps.

Implications of this review underscore the critical importance of multidisciplinary collaboration in developing and implementing ethical frameworks for AI in medical education. Policymakers and educational leaders must prioritize inclusive governance structures, invest in transparency-enhancing technologies, and adopt robust accountability protocols to safeguard learner rights and equity. Additionally, standardized ethical

guidelines that are adaptable to various cultural and disciplinary settings can facilitate global harmonization while respecting local contexts.

5 Conclusion

The integration of artificial intelligence into interprofessional medical education offers promising advancements but also generates multifaceted ethical and governance challenges that require holistic solutions. Central issues such as data privacy, algorithmic bias, transparency, and learner autonomy persist amidst increasingly complex, cross-disciplinary learning environments. Existing governance structures remain fragmented and oftentimes lack comprehensive stakeholder representation, diminishing their effectiveness in managing AI-related risks and addressing accountability. Additionally, substantial variability in ethical priorities across cultures and professional disciplines further complicates the formation of universal governance models. While recent frameworks have highlighted core values like fairness, accountability, and transparency, their application and empirical validation in real-world settings are still limited and require further scrutiny. This review makes a distinctive contribution by systematically synthesizing ethical frameworks for AI-enhanced interprofessional learning, identifying governance gaps, and advancing the call for context-sensitive models grounded in multidisciplinary collaboration. Ultimately, robust and adaptable governance mechanisms—including transparent policies, stakeholder engagement, and clear accountability protocols—are essential to safeguard learners' rights and educational equity. Institutional readiness demands proactive measures to close governance gaps, ensure ethical compliance, and promote just integration of AI across diverse health education contexts. The implications underline the urgency of collaborative policymaking, ongoing empirical research, and the development of standardized but flexible ethical guidelines that can be harmonized globally yet remain attentive to local values and circumstances. Through these efforts, AI technologies can be ethically harnessed to enrich interprofessional medical education while sustaining trust, integrity, and equitable outcomes for all learners.

References

1. Sriram, A., Ramachandran, K., Krishnamoorthy, S.: Artificial intelligence in medical education: transforming learning and practice. Cureus. **17**(3) (2025)
2. Klimova, B., Pikhart, M., Kacetl, J.: Ethical issues of the use of AI-driven mobile apps for education. Front. Public Heal. **10**, 1118116 (2023). https://doi.org/10.3389/fpubh.2022.1118116
3. Lomis, K., et al.: Artificial intelligence for health professions educators. NAM Perspect., 10.31478 (2021). https://doi.org/10.31478/202109a
4. Saragih, I.D., Hsiao, C.T., Fann, W.C., Hsu, C.M., Saragih, I.S., Lee, B.O.: Impacts of interprofessional education on collaborative practice of healthcare professionals: a systematic review and meta-analysis. Nurse Educ. Today, 106136 (2024). https://doi.org/10.1016/j.nedt.2024.106136
5. Eunike, I.J, Serang, Y., Silalahi, A.D.K.: Does ChatGPT-Enhanced Collaborative Learning Foster Critical Thinking in Education? A Bloom's Taxonomy Perspective. (2025)

6. Abdelaziz, A., Mansour, T., Alkhadragy, R., Abdel Nasser, A., Hasnain, M.: Challenges to interprofessional education: will e-learning be the magical stick? Adv. Med. Educ. Pract., 329–336 (2021). https://doi.org/10.2147/AMEP.S309050
7. Ferrero, F., Barujel, A.G.: Algorithmic driven decision-making systems in education: analyzing bias from the sociocultural perspective. In: 2019 XIV Latin American conference on learning technologies (LACLO), pp. 166–173 (2019). https://doi.org/10.1109/LACLO48532.2019.8952937
8. Mousa, M.: AI-supported formative assessments: enhancing student-Centered learning and teacher perceptions. J. Pedagog. Educ. Sci. **4**(02), 127–141 (2025). https://doi.org/10.56741/jpes.v4i02.857
9. Kranke, M.: The Politics of Collaborative Global Governance: Organisational Positioning in IMF-World Bank Collaboration. University of Warwick (2017)
10. Echeverria, V., et al.: TeamVision: An AI-powered Learning Analytics System for Supporting Reflection in Team-based Healthcare Simulation. In: Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems, pp. 1–22 (2025). https://doi.org/10.1145/3544548.3581363
11. Härgestam, M., Morian, H., Lindgren, L.: Interprofessional team training via telemedicine in medical and nursing education. BMC Med. Educ. **24**(1), 1110 (2024). https://doi.org/10.1186/s12909-024-04110-y
12. Filgueiras, F.: Artificial intelligence and education governance. Educ. Citizsh. Soc. Justice. **19**(3), 349–361 (2024). https://doi.org/10.1177/17461979231176431
13. Love, A.S., Niu, C., Labay-Marquez, J.: Artificial intelligence in public health education: navigating ethical challenges and empowering the next generation of professionals. Health Promot. Pract. (2025). https://doi.org/10.1177/15248399251320989
14. Porter, Z., Habli, I., McDermid, J., Kaas, M.: A principles-based ethics assurance argument pattern for AI and autonomous systems. AI Ethics. **4**(2), 593–616 (2024). https://doi.org/10.1007/s43681-023-00346-9
15. Pan, Y.: Global healthcare communication: innovating medical English education. Cogent Educ. **12**(1), 2482505 (2025). https://doi.org/10.1080/2331186X.2025.2482505
16. Charow, R., et al.: Artificial intelligence education programs for health care professionals: scoping review. JMIR Med. Educ. **7**(4), e31043 (2021). https://doi.org/10.2196/31043
17. Allen, L.K., Kendeou, P.: ED-AI Lit: an Interdisciplinary framework for AI literacy in education. Policy Insights from Behav. Brain Sci. **11**(1), 3–10 (2024). https://doi.org/10.1177/23727322231150748
18. Sangwa, S.: Navigating the algorithmic turn: a dynamic governance framework for ethical and equitable AI integration in education. J. Ethical Innov. Impact. **1**(2), 1–21 (2025)
19. Halman, M., Baker, L., Ng, S.: Using critical consciousness to inform health professions education: a literature review. Perspect. Med. Educ. **6**, 12–20 (2017). https://doi.org/10.1007/s40037-016-0322-4
20. Hopia, H., Latvala, E., Liimatainen, L.: Reviewing the methodology of an integrative review. Scand. J. Caring Sci. **30**(4), 662–669 (2016). https://doi.org/10.1111/scs.12267
21. Cox, M., Cuff, P., Brandt, B., Reeves, S., Zierler, B.: Measuring the impact of interprofessional education on collaborative practice and patient outcomes. J. Interprof. Care. **30**(1), 1–3 (2016). https://doi.org/10.3109/13561820.2015.1111052
22. Pesut, D.J., Headrick, L.A., Holmboe, E., Moore, S.M.: Systems and complexity thinking to master leadership challenges in interprofessional health professionals education. J. Med. Educat. Curri. Develop. **10**, 23821205231175204 (2023). https://doi.org/10.1177/23821205231175205
23. Mat Nawi, N.F., Hanim Hadie, S.N.: Interprofessional anatomy education: its significance, challenges and recommendations. Educ. Med. J. **16**(2) (2024)

24. Russell, R.G., et al.: Competencies for the use of artificial intelligence–based tools by health care professionals. Acad. Med. **98**(3), 348–356 (2023). https://doi.org/10.1097/ACM.0000000000004956
25. Hamilton, A.: Artificial intelligence and healthcare simulation: the shifting landscape of medical education. Cureus. **16**(5) (2024)
26. Sesay, I.M.S.G.A.: Redefining student assessments in the age of AI-assisted learning strategies for educators. GSJ. **12**(8) (2024)
27. Kourkoulou, D., Tzirides, A.O.O., Cope, B., Kalantzis, M. (eds.): Trust and Inclusion in AI-Mediated Education: Where Human Learning Meets Learning Machines. Springer Nature (2024)
28. Mufid, A.: The ethical limitation of using artificial intelligence (AI) in teaching prophetic tradition. J. Ushuluddin. **32**(2), 130–151 (2024)
29. Mehmood, T.: Ethical AI in education: addressing bias, privacy, and equity in AI-driven learning systems. J. AI Integr. Educ. **2**(1), 38–45 (2025)
30. Houngbo, P.T., Coleman, H.L.S., Zweekhorst, M., De Cock Buning, T., Medenou, D., Bunders, J.F.G.: A model for good governance of healthcare technology management in the public sector: learning from evidence-informed policy development and implementation in Benin. PLoS One. **12**(1), e0168842 (2017). https://doi.org/10.1371/journal.pone.0168842
31. Naidoo, S., Turner, K.M., McNeill, D.B.: Ethics and interprofessional education: an exploration across health professions education programs. J. Interprof. Care. **34**(6), 829–831 (2020). https://doi.org/10.1080/13561820.2020.1750130
32. Kurre, J.: The Accountability, Responsibility & Governance as a Unified Strategy for AI. National American University (2024)
33. Salloum, S.A.: AI perils in education: Exploring ethical concerns. In: Artificial intelligence in education: The power and dangers of ChatGPT in the classroom, pp. 669–675 (2024)
34. Nopas, D.S.: Algorithmic learning or learner autonomy? Rethinking AI's role in digital education. Qual. Res. J. (2025)
35. Willis III, J.E., Slade, S., Prinsloo, P.: Ethical oversight of student data in learning analytics: a typology derived from a cross-continental, cross-institutional perspective. Educ. Technol. Res. Dev. **64**(5), 881–901 (2016). https://doi.org/10.1007/s11423-016-9461-4
36. Nemorin, S.: Towards decolonising the ethics of AI in education. Glob. Soc. Educ., 1–13 (2024)
37. F. Pedro, M. Subosa, A. Rivas, and P. Valverde, Artificial Intelligence in Education: Challenges and Opportunities for Sustainable Development. 2019.
38. Hermann, S., Ohanna, K.E., Siegfried, W.: Future health care managers and the need for an international and interprofessional learning experience–the case of a summer school on cross-border health care. Int. J. Teach. Educ. **3**(2), 31–49 (2015)
39. Weidener, L., Fischer, M.: Proposing a principle-based approach for teaching AI ethics in medical education. JMIR Med. Educ. **10**(1), e55368 (2024). https://doi.org/10.2196/55368
40. Goetze, T.S.: Integrating ethics into computer science education: Multi-, inter-, and transdisciplinary approaches. In: Proceedings of the 54th ACM Technical Symposium on Computer Science Education V. 1, pp. 645–651 (2023). https://doi.org/10.1145/3545947.3562217
41. Kagan, R.: Ethical frameworks for AI-integrated education: issues and prospects. J. Comput. Educ. **2**(2), 1–17 (2023) [Online]. Available: https://www.journalofcomputereducation.info/archieve/vol2_2/JCE_2_2_1_pdf.pdf#page=1&zoom=auto,-82,848
42. Nurhayati, S., Taufikin, L.J., Musa, S.: Towards effective artificial intelligence-driven learning in Indonesian child education: understanding parental readiness, challenges, and policy implications. Educ. Process Int. J. **15**(1), 1–29 (2025) [Online]. Available: https://www.ceeol.com/search/article-detail?id=1339581

43. Roy, E., et al.: Artificial intelligence in healthcare: medical students' perspectives on balancing innovation, ethics, and patient-Centered care. Int. J. Med. Students. **13**(1), 9–16 (2025). https://doi.org/10.5195/ijms.2025.3344
44. Ali, M.: The role of AI in reshaping medical education: opportunities and challenges. Clin. Teach. **22**(2) (2025). https://doi.org/10.1111/tct.70040
45. Ta'an, W., Damrah, S., Al-Hammouri, M.M., Williams, B.: Professional identity and its relationships with AI readiness and interprofessional collaboration. PLoS One. **20**(5), e0322794 (2025). https://doi.org/10.1371/journal.pone.0322794
46. Pandey, R., Agarwal, A., Regalla, S.R., Srivastava, S., Tyagi, A.: Systematic review: the importance of artificial intelligence in medical education. J. Contemp. Clin. Pract. **11**(7), 523–532 (2025) https://www.jccpractice.com/article/systematic-review-the-importance-of-artificial-intelligence-in-medical-education-1019/
47. Arvai, N., Katonai, G., Mesko, B.: Health care professionals' concerns about medical AI and psychological barriers and strategies for successful implementation: scoping review. J. Med. Internet Res. **27**, e66986 (2025). https://doi.org/10.2196/66986
48. García-López, I.M., Trujillo-Liñán, L.: Ethical and regulatory challenges of generative AI in education: a systematic review. Front. Educ. **10** (2025). https://doi.org/10.3389/feduc.2025.1565938
49. Freeman, S., et al.: Developing an AI governance framework for safe and responsible AI in health care organizations: protocol for a multimethod study. JMIR Res. Protoc. **14**, e75702 (2025). https://doi.org/10.2196/75702
50. Tran, M., et al.: Situating governance and regulatory concerns for generative artificial intelligence and large language models in medical education. npj Digit. Med. **8**(1), 315 (2025). https://doi.org/10.1038/s41746-025-01721-z
51. Wiss, A., et al.: Development of the AI Acceptance Scale for Interprofessional Education (AAIPE) and collaborative practice settings. J. Interprofessional Educ. Pract. **40**, 100752 (2025). https://doi.org/10.1016/j.xjep.2025.100752
52. Gao, D.K., Haverly, A., Mittal, S., Wu, J., Chen, J.: AI ethics: a bibliometric analysis, critical issues, and key gaps. Int. J. Bus. Anal. **11**(1), 1–19 (2024). https://doi.org/10.4018/IJBAN.338367
53. Bol, J.A., Sheffel, A., Zia, N., Meghani, A.: How to address the geographical bias in academic publishing. BMJ Glob. Health. **8**(12) (2023). https://doi.org/10.1136/bmjgh-2023-013111
54. Ifenthaler, D., Schumacher, C.: Student perceptions of privacy principles for learning analytics. Educ. Technol. Res. Dev. **64**(5), 923–938 (2016). https://doi.org/10.1007/s11423-016-9477-y
55. G'andara, D., Anahideh, H., Ison, M.P., Picchiarini, L.: Inside the black box: detecting and mitigating algorithmic bias across racialized groups in college student-success prediction. AERA Open. **10**, 23328584241258740 (2024). https://doi.org/10.1177/23328584241258741
56. Chaudhry, M.A., Cukurova, M., Luckin, R.: A transparency index framework for AI in education. In: International Conference on Artificial Intelligence in Education, pp. 195–198. Springer International Publishing (2022). https://doi.org/10.1007/978-3-031-11647-6_33
57. Jones, K.M.: Learning analytics and higher education: a proposed model for establishing informed consent mechanisms to promote student privacy and autonomy. Int. J. Educ. Technol. High. Educ. **16**(1), 1–22 (2019). https://doi.org/10.1186/s41239-019-0155-0
58. Knight, S., Shibani, A., Vincent, N.: Ethical AI governance: mapping a research ecosystem. AI Ethics, 1–22 (2024). https://doi.org/10.1007/s43681-023-00416-z
59. Holmes, W., et al.: Ethics of AI in education: towards a community-wide framework. Int. J. Artif. Intell. Educ., 1–23 (2022). https://doi.org/10.1007/s40593-021-00239-1
60. Raji, I.D., et al.: Closing the AI accountability gap: defining an end-to-end framework for internal algorithmic auditing. In: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, pp. 33–44 (2020). https://doi.org/10.48550/arXiv.2001.00973

61. Gundersen, T., Bærøe, K.: The future ethics of artificial intelligence in medicine: making sense of collaborative models. Sci. Eng. Ethics. **28**(2), 17 (2022). https://doi.org/10.1007/s11948-022-00369-2
62. Gujjula, R., Sanghera, K.: Ethical considerations and data privacy in AI education. J. Student-Scientists' Res. **5** (2023). https://doi.org/10.13021/jssr2023.3958
63. Huang, L.: Ethics of artificial intelligence in education: student privacy and data protection. Sci. Insights Educ. Front. **16**(2), 2577–2587 (2023). https://doi.org/10.15354/sief.23.re202
64. Peña, J.M., Moreno, O.B., Herrera, Á.L.O., Moreno, T.E.B.: Balancing privacy and ethics in the use of artificial intelligence in education. Sapiens Int. Multidiscip. J. **1**(3), 148–170 (2024). https://doi.org/10.71068/m4gqp119
65. Mutimukwe, C., Twizeyimana, J.D., Viberg, O.: Students' information privacy concerns in learning analytics: towards a model development. arXiv Prepr, arXiv2109.00068 (2021)
66. Fung, P., Etienne, H.: Confucius, cyberpunk and Mr. Science: Comparing AI ethics between China and the EU. arXiv Prepr, arXiv2111.07555 (2021). https://doi.org/10.48550/arXiv.2111.07555
67. Jobin, A., Ienca, M., Vayena, E.: The global landscape of AI ethics guidelines. Nat. Mach. Intell. **1**(9), 389–399 (2019)
68. Ko, H.K., et al.: Teaching and learning in interprofessional ethics education: tutors' perspectives. Nurs. Ethics. **30**(1), 133–144 (2023). https://doi.org/10.1177/09697330221122901
69. Mahamadou, A.J.D., Ochasi, A., Altman, R.B.: Data ethics in the era of healthcare Artificial Intelligence in Africa: An Ubuntu philosophy perspective. arXiv e-prints. (2024). https://doi.org/10.48550/arXiv.2406.10121
70. Nguyen, A., Ngo, H.N., Hong, Y., Dang, B., Nguyen, B.P.T.: Ethical principles for artificial intelligence in education. Educ. Inf. Technol. **28**(4), 4221–4241 (2023). https://doi.org/10.1007/s10639-022-11316-w
71. Ukeje, I.O., Elom, C.O., Ayanwale, M.A., Umoke, C.C., Nwangbo, S.O.: Exploring an innovative educational governance framework: leveraging artificial intelligence in a stakeholder-driven 'Open Campus Model' in South East Nigerian Universities. Int. J. Learn. Teach. Educ. Res. **23**(6), 416–440 (2024)
72. Jeyaraman, M., Ramasubramanian, S., Balaji, S., Jeyaraman, N., Nallakumarasamy, A., Sharma, S.: ChatGPT in action: Harnessing artificial intelligence potential and addressing ethical challenges in medicine, education, and scientific research. World J. Methodol. **13**(4), 170 (2023). https://doi.org/10.5662/wjm.v13.i4.170
73. Panadés, R., Yuguero, O.: Cyber-bioethics: the new ethical discipline for digital health. Front. Digit. Heal. **6**, 1523180 (2025). https://doi.org/10.3389/fdgth.2024.1523180
74. Roche, C., Wall, P.J., Lewis, D.: Ethics and diversity in artificial intelligence policies, strategies and initiatives. AI Ethics. **3**(4), 1095–1115 (2023). https://doi.org/10.1007/s43681-023-00346-6
75. Ismail, I.A., Aloshi, J.M.R.: Data privacy in AI-driven education: an in-depth exploration into the data privacy concerns and potential solutions. In: AI Applications and Strategies in Teacher Education, pp. 223–252. IGI Global (2025)
76. Baker, R.S., Hawn, A.: Algorithmic bias in education. Int. J. Artif. Intell. Educ., 1–41 (2022). https://doi.org/10.1007/s40593-021-00285-9
77. Geethanjali, K.S., Umashankar, N.: Enhancing educational outcomes with explainable AI: bridging transparency and trust in learning systems. In: 2025 International Conference on Emerging Systems and Intelligent Computing (ESIC), pp. 325–328. IEEE (2025)
78. Morley, J., Floridi, L., Kinsey, L., Elhalal, A.: From what to how: an initial review of publicly available AI ethics tools, methods and research to translate principles into practices. Sci. Eng. Ethics. **26**(4), 2141–2168 (2020). https://doi.org/10.1007/s11948-020-00213-5

79. Mittelstadt, B.: Principles alone cannot guarantee ethical AI. Nat. Mach. Intell. **1**(11), 501–507 (2019). https://doi.org/10.1038/s42256-019-0114-4
80. Wieringa, M.: What to account for when accounting for algorithms: a systematic literature review on algorithmic accountability. In: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, pp. 1–18 (2020). https://doi.org/10.1145/3351095.3372873
81. Adams, C., Pente, P., Lemermeyer, G., Rockwell, G.: Artificial intelligence ethics guidelines for K-12 education: a review of the global landscape. In: International Conference on Artificial Intelligence in Education, pp. 24–28. Springer International Publishing, Cham (2021). https://doi.org/10.1007/978-3-030-78278-0_4
82. Peer Mohamed, M.S.: Exploring ethical dimensions of AI-enhanced language education: a literature perspective. Technol. Lang. Teach. Learn. (2024)
83. Mandinach, E.B., Gummer, E.S. (eds.): The Ethical Use of Data in Education: Promoting Responsible Policies and Practices. Teachers College Press (2021)
84. Tzimas, D., Demetriadis, S.: Ethical issues in learning analytics: a review of the field. Educ. Technol. Res. Dev. **69**, 1101–1133 (2021). https://doi.org/10.1007/s11423-021-10038-x
85. Murchan, D., Siddiq, F.: A call to action: a systematic review of ethical and regulatory issues in using process data in educational assessment. Large-Scale Assessments Educ. **9**(1), 25 (2021). https://doi.org/10.1186/s40536-021-00108-2
86. Dimmock, C.: Connecting research and knowledge on educational leadership in the West and Asia: adopting a cross-cultural comparative perspective. Comp. Educ. **56**(2), 257–277 (2020). https://doi.org/10.1080/03050068.2020.1711123
87. Reddy, S., Allan, S., Coghlan, S., Cooper, P.: A governance model for the application of AI in health care. J. Am. Med. Informatics Assoc. **27**(3), 491–497 (2020). https://doi.org/10.1093/jamia/ocz217
88. Cary Jr., M.P., De Gagne, J.C., Kauschinger, E.D., Carter, B.M.: Advancing health equity through artificial intelligence: an educational framework for preparing nurses in clinical practice and research. Creat. Nurs. **30**(2), 154–164 (2024)
89. McKimm, J., Balasooriya, C., Fyfe, M., Aboulsoud, S., Brouwer, E., Nadkar, A.A.: International collaboration in health professions education. Med. Teach. **47**(5), 800–807 (2025)
90. Tirado, A.M., Mulholland, P., Fernandez, M.: Towards an operational responsible AI framework for learning analytics in higher education. arXiv Prepr., arXiv2410.05827 (2024)
91. Kawakami, A., Coston, A., Zhu, H., Heidari, H., Holstein, K.: The situate AI Guidebook: Co-designing a toolkit to support multi-stakeholder, early-stage deliberations around public sector AI proposals. In: Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems, pp. 1–22 (2024). https://doi.org/10.1145/3544548.3581198
92. Sato, E., Shyyan, V., Chauhan, S., Christensen, L.: Putting AI in fair: a framework for equity in AI-driven learner models and inclusive assessments. J. Meas. Eval. Educ. Psychol. **15**(Special Issue), 263–281 (2024)



Scaling AI Agents in the Enterprises: A Strategic Framework, Architecture, Governance, and KPIs

Vijay Kumar Butte, Sujata Butte(✉), and Sagar Suryakanth

University of Idaho, Idaho Falls, ID, USA
suja2591@vandals.uidaho.edu

Abstract. Agentic AI provides a valuable opportunity for enterprise transformation. The AI agents can learn, reason, and act on behalf of the organization. However, deploying these systems at scale presents unique challenges related to complexity, reliability, security, and operational alignment. This paper presents a practical framework to guide the successful and scalable adoption of agentic AI within enterprises. It outlines key components necessary for building robust agentic AI systems, including architectural considerations, governance, monitoring, and iterative optimization. By combining the flexibility of rapid innovation with a strong foundation in reliability and security, the proposed approach enables organizations to move confidently from experimentation to large-scale deployment. The goal is to empower organizations to harness the transformative potential of agentic AI while maintaining control, resilience, and strategic alignment.

Keywords: Agentic AI · generative AI · enterprise strategy · AI agents · artificial intelligence · LLM · large language models · enterprise strategy

1 Introduction

Large Language Models (LLMs) are advanced neural networks, typically based on transformer architectures, trained on vast text corpora to predict and generate human-like language. LLMs have demonstrated remarkable capabilities in reasoning, summarization, translation, and code generation [1–6]. Initially, LLMS were leveraged by providing prompts and receiving generated responses. However, a limitation was lack of LLM's training on up-to-date and proprietary data. This led to the Retrieval-Augmented Generation (RAG) approach that allowed LLMs to access external, up-to-date information to enhance their responses [7, 8]. As RAG gained traction, the process was further streamlined with the development of managed knowledge bases and tools. These helped in automating the retrieval and integration of relevant information. The latest advancements in generative AI have introduced specialized AI agents to collaborate and tackle complex tasks.

An AI agent is a solution that leverages a large language model, external tools and APIs to achieve a specific goal. They understand instructions, plan the steps needed and

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026
S. Lanka et al. (Eds.): ICTSM 2025, LNNS 1755, pp. 174–188, 2026.
https://doi.org/10.1007/978-3-032-13177-5_14

take necessary and authorized actions with a degree of autonomy, and self-direction. This provides the solution ability to navigate uncertainty, adapt to dynamic environments, make real time informed decisions and take intelligent actions [9–12]. This opens up valuable opportunities for the enterprises towards machine intelligence. Enterprises can leverage these agentic AI solutions that have the ability to perform complex tasks with reduced human intervention.

For successful and scalable adoption of agentic AI, enterprises need a practical framework that addresses the inherent complexity of agentic AI systems. This paper outlines a structured yet flexible approach for successful adoption of agentic AI at scale in enterprises. It enables combining rapid innovation with a strong foundation of security, reliability, and operational excellence. The approaches outlined help accelerate progress by building on proven principles while embracing new possibilities. The paper is organized as follows. The Sect. 2 discusses aligning agentic AI strategy with overarching business strategy. The Sect. 3 discusses the AI agent core components and the technical architectures. The Sect. 4 discusses the framework for governance, ethics and compliance aspects of Agentic AI. The Sect. 5 provides a framework for measuring success of Agentic AI adoption. The Sect. 6 discusses the need for continuous strategic evolution of AI strategy. This is followed by the conclusion in Sect. 7.

2 Aligning Agentic AI Strategy with Business Strategy

The core objective of enterprise Agentic AI strategy is to serve the overarching business goals. The enterprise business strategy, technology strategy, data strategy and AI strategy should be well aligned and systematically connected [13–16]. Agentic AI has potential to deliver significant business value to enterprises by enabling autonomous digital actors that can reason, plan, and execute across complex systems. These opportunities are beyond automation and efficiency, agentic AI opens up a broader spectrum of strategic opportunities from dynamic decision-making and customer personalization. Thus enabling businesses to reimagine entire operating models.

2.1 Business Transformation Through Autonomy

Enterprises often operate in fragmented systems that require manual coordination across teams and departments. These inefficiencies cause delay in decision making. AI agents enable organizations to transform by embedding these autonomous agents into critical operational flow. These transformations benefit the business through increased operational agility, faster decisions, knowledge activation and self optimizing systems.

The AI agents can dynamically coordinate tasks across teams and departments with varying levels of autonomy. This reduces process cycle times, exceptions, and operational inefficiencies. The agentic AI systems monitor environments, interpret signals (structured, semi-structured or unstructured), and act on insights in near real-time. Additionally, AI Agents retrieve, summarize, and contextualize enterprise knowledge, enabling faster onboarding, enhanced support, and improved informed decisions across roles. Over time, agentic systems learn from outcomes, detect bottlenecks, and autonomously

optimize workflows without reprogramming or retraining by humans. These capabilities and their advantages provide rich opportunities for the businesses to transform their processes.

2.2 New Operating Models Enabled by Agentic AI

Agentic AI can be leveraged by enterprises to reconfigure value chains and organizational structures. This opens up options for new operating models such as optimized human and AI collaboration, autonomous business units, event driven micro-autonomy. Through the optimized human and AI collaboration model, instead of being tools used by humans, AI agents become active participants in the organization. The AI agents provide inputs, attend meetings, follow-up on tasks, conduct research, and manage communications. The AI agents can also operate as semi-autonomous units that run marketing campaigns, handle procurement, and manage customers with reduced human oversight. Further, enterprises can adopt event driven mico-autonomy by evolving from batch-oriented processes to event-driven systems where agents react to triggers in near real time.

3 Enterprise AI Agents Core Components and Architecture

Agentic AI solutions are powerful tools, however, it is essential to understand when to apply them for deriving maximum effectiveness. Often there are several tasks that are deterministic, follow predefined steps. These systems can leverage workflows, optimize LLM calls and achieve their goal efficiently. AI agents deliver maximum value for complex and dynamic tasks. These complex tasks benefit from AI agent's real-time decisions, adaptive strategies, and autonomous interaction with their environment. Agents are also well suited for situations where tasks evolve or require ongoing problem-solving. Thus, providing a level of adaptability that workflows can't offer.

The key characteristics of agentic AI include, goal-oriented behavior, a degree of autonomy, reasoning and planning capability, memory and context awareness and capabilities for multi-modal and environment interactions. Through goal oriented behavior, agents can decompose the assigned objective into sub-tasks without needing explicit programming for each step. The AI agents take actions with minimal direct human input. They also adjust their behavior in response to changes in data or context. They demonstrate a high degree of reasoning and planning by leveraging large language models to plan and solve problems. Through their memory and context awareness, agents remember past interactions, learn from outcomes, and refine their future strategies. AI agents are also capable of processing and responding through various channels viz. text, voice, APIs, databases, IoT systems.

3.1 Architecture for Agentic AI

Agentic AI necessitates a significant shift in enterprise architecture that enables autonomy, ensures interoperability, and maintains trust, control, and scalability. An agentic system architecture needs to be designed to support various key features of agents. These features include, enabling agents to perceive context through environment and

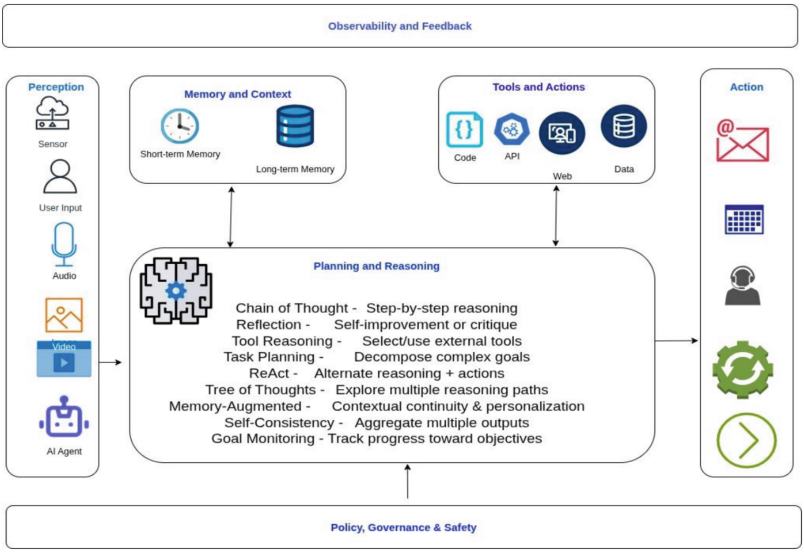


Fig. 1. AI Agent Overview Architecture

data inputs, reason over tasks and goals, plan and execute actions across systems, learn from outcomes and feedback and stay within defined safety and policy boundaries.

Figure 1 provides a high-level architecture for AI agents. The perception layer ingests signals from internal and external sources (APIs, sensors, messages, documents). It converts raw inputs into structured, actionable context. This includes natural language understanding, image/audio processing, API result parsing. It enables the agent to sense, interpret, and contextualize data from its surroundings, forming the basis for informed decision-making and autonomous behavior. The perception layer focuses on collecting real-time or batch data from external sources including users, environments, or digital systems (e.g., via APIs, sensors, documents, or chat interfaces). This followed by signal processing and interpretation by using natural language understanding (NLU), computer vision, or multimodal fusion to convert raw input into semantically meaningful representations. Then it contextualizes by embedding inputs into a context-aware framework, often leveraging memory, embeddings, or external knowledge bases to ensure that perceptions are interpreted in alignment with the agent's goals, history, and task state.

3.2 Agentic AI Memory and Context Layer

The memory and context layer serves as the cognitive substrate of an AI agent, enabling continuity, adaptability, and coherence in decision-making and interaction over time. This layer bridges perception and reasoning by retaining relevant information about the agent's internal state, environment, past interactions, and goals. It plays a pivotal role in allowing the agent to exhibit context-aware, temporally consistent, and personalized behavior. Thus enabling a key characteristic of agentic intelligence.

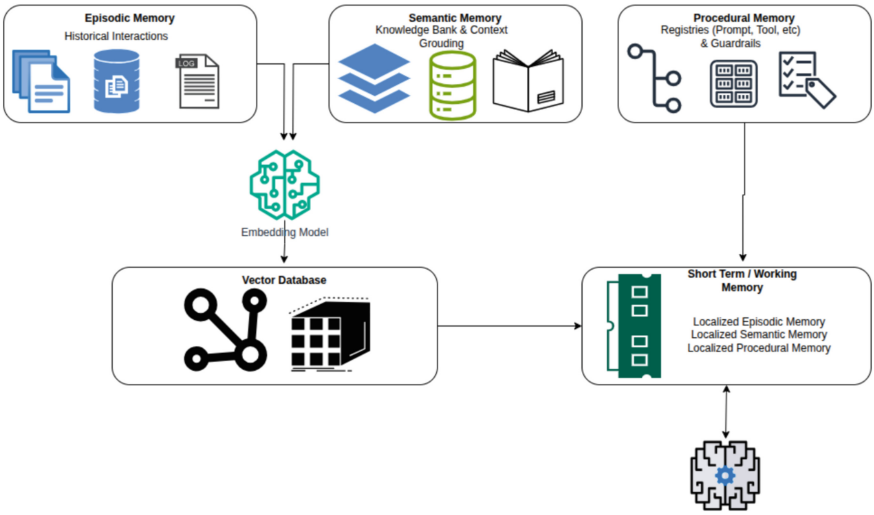


Fig. 2. AI Agent Memory Architecture

Figure 2 depicts the architecture of the memory layer of an AI agent. The short-term working memory temporarily holds data relevant to the current task or interaction window. It enables real-time responsiveness and supports contextual continuity in multi-turn dialogues or sequential decision-making. The long-term memory stores knowledge accumulated over time, including user preferences, historical actions, domain knowledge, and learned policies. Long-term memory enables the agent to generalize from experience, maintain identity across sessions, and support personalization. The AI agent memory can be further classified into episodic memory, semantic memory, procedural memory and working memory.

Episodic memory stores past interactions and actions, allowing agents to reference previous experiences for better future decisions. It is especially useful in multi-turn dialogues where maintaining context is critical. It is implemented using persistent systems like vector databases. This enables fast retrieval of relevant past interactions through semantic similarity.

Semantic memory stores general knowledge and facts. It includes language rules, domain expertise, and self-knowledge, enabling the agent to understand and interpret information broadly. It is commonly used in Retrieval-Augmented Generation (RAG), to retrieve relevant content.

Procedural memory defines how the agent operates, including system structure, available tools, and safety guardrails. It focuses on execution processes rather than knowledge content. It is managed through registries like code repositories and tool lists, it ensures consistent, safe, and predictable behavior.

Short-term memory is a temporary workspace that combines relevant episodic, semantic, and procedural information retrieved from long-term memory to handle current tasks. It forms the active context used by the agent to generate coherent, context-aware responses. This memory enables nuanced decision-making, supports natural interactions, and ensures only the most relevant information is used, optimizing both performance and accuracy.

3.3 Agentic AI Tools Layer

The tools and actions components represent the agent's capability to interact with its environment and effect change based on its goals and decisions. They are the operational means by which agentic AI translates reasoning and planning into real-world impact. The tools in an agentic AI context refer to the external systems, APIs, software modules, or physical devices that an agent can access or control to perform tasks. These include APIs, software integrations, web browsers, search engines, IoT devices or sub-agents.

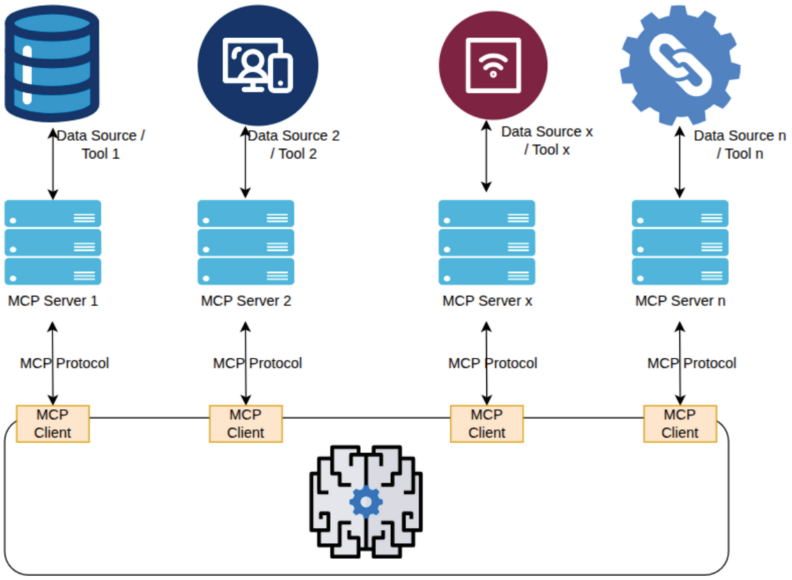


Fig. 3. AI Agent Tools Architecture

Equipping an agent with diverse and capable tools expands its functional scope, enabling it to handle complex, multi-step tasks and interact across varied domains. The Model Context Protocol is one of the key components of communication architecture in agentic AI. The Model Context Protocol (MCP) is an open-source, application-layer protocol designed to standardize how AI systems connect with external tools, services, and datasets via a unified interface. Figure 3 depicts the AI agent's tool architecture with MCP. MCP protocol reduces the complexity of integrating each tool individually.

3.4 Agentic AI Planning and Reasoning Layer

The planning and reasoning layer is central to enabling agentic AI systems to operate autonomously and intelligently. It allows agents to move beyond reactive behaviors by formulating goals, generating action plans, and making informed decisions based on their understanding of context and environment. Planning involves breaking down high-level objectives into structured sequences of tasks. This includes goal interpretation, task decomposition, and action sequencing. Agentic systems may use symbolic planners, reinforcement learning, or large language models (LLMs) to dynamically generate and adapt plans in response to changing environments. Reasoning enables agents to interpret inputs, infer relationships, and make decisions under uncertainty. It encompasses deductive, inductive, and commonsense reasoning. These are critical for understanding complex tasks, resolving ambiguities, and simulating human-like cognition. LLMs often support reasoning through chain-of-thought prompting, self-reflection, and tool use. This layer interacts closely with perception (to understand the current state), memory and context (to stay grounded in history), and the action layer (to execute and adapt plans).

3.5 Agentic AI Actions Layer

Actions are the specific operations or commands executed by the agent via its tools. They represent discrete steps such as sending an email, executing a transaction, manipulating a physical object, updating a database entry, initiating a subprocess or calling another service etc. The actions must be precise, context-aware, and compliant with governance policies to ensure safe and effective outcomes. The integration of tools and actions allows agentic AI to move from passive understanding to active problem-solving. By orchestrating the appropriate use of tools and executing well-planned actions, agents can autonomously perform complex workflows, deliver tangible results, and respond adaptively to dynamic environments.

3.6 Agentic AI Policy, Governance and Safety Layer

The policy, governance & safety layer enforces guardrails, boundaries, and business rules. As AI agents gain autonomy, robust policy, governance, and safety mechanisms are essential to ensure they act responsibly, align with human values, and operate within acceptable risk boundaries. The policy and governance frameworks define the roles, permissions, and constraints under which agents operate. This includes enforcing legal and ethical rules, maintaining audit trails, and enabling human oversight, particularly in sensitive domains. The safety focuses on ensuring reliability, robustness, and alignment.

3.7 Agentic AI Orchestration Layer

The orchestration layer serves as the central coordination hub within an agentic AI system, managing the interaction and integration of various components. This includes perception, memory, reasoning, planning, and action modules to achieve coherent and effective autonomous behavior. The key functions include task management, resource

allocation, workflow coordination, monitoring and adaptation. The task management module breaks down complex goals into subtasks and distributes them among specialized agents or internal modules, ensuring efficient parallel or sequential execution. The resource allocation module dynamically assigns computational resources and prioritizes processes based on task urgency, complexity, and available capacity, optimizing performance and responsiveness. The workflow coordination module oversees the smooth flow of information and control signals across different system layers, maintaining synchronization and managing dependencies. This layer also facilitates communication and cooperation between multiple agents, enabling collective problem-solving and division of labor in distributed environments. It continuously tracks system performance, detecting bottlenecks or failures, and triggers re-planning or fallback strategies to maintain robustness. The orchestration layer is crucial for scaling agentic AI from isolated capabilities to integrated, goal-driven systems. It ensures that the agent's diverse functions work in harmony, balancing innovation and reliability. By enabling flexible coordination, the orchestration layer supports both rapid adaptation to dynamic environments and adherence to governance and safety constraints.

3.8 Agentic AI Observability and Feedback Layer

The observability & feedback layer captures telemetry, logs, decision traces, and outcomes. It is essential for maintaining transparency, continuous improvement, and operational reliability in agentic AI systems. Its key functions include real time monitoring, logging and traceability, feedback integration and alerting. It continuously tracks the agent's performance, decisions, resource usage, and environmental interactions to detect anomalies, inefficiencies, or errors promptly. It records detailed action histories, reasoning processes, and input-output exchanges to support auditability, debugging, and compliance with regulatory standards. It also collects and analyzes key performance indicators (KPIs) and success criteria related to task completion, accuracy, response time, and user satisfaction. Further, it incorporates feedback from users, other agents to refine agent behavior through learning or parameter tuning. It triggers alerts or human-in-the-loop interventions when unsafe, unexpected, or suboptimal behaviors are detected. This layer ensures the agent's actions remain transparent and accountable while enabling iterative improvements. By systematically capturing and acting on feedback, agents can adapt to changing environments, correct errors proactively, and align more closely with user expectations and organizational goals. Observability thus forms the backbone of trustworthiness and resilience in scalable agentic AI deployments.

3.9 Agent to Agent Protocol (A2A)

The Agent-to-Agent Protocol (A2A) is one of the key components of the Agentic AI communication architecture. The Agent-to-Agent Protocol (A2A) is an open-standard communication protocol, designed for interoperability among autonomous AI agents. A2A enables peer agents to discover one another, advertise capabilities, negotiate tasks, and coordinate actions in a secure and structured manner. Figure 4 depicts the A2A protocol.

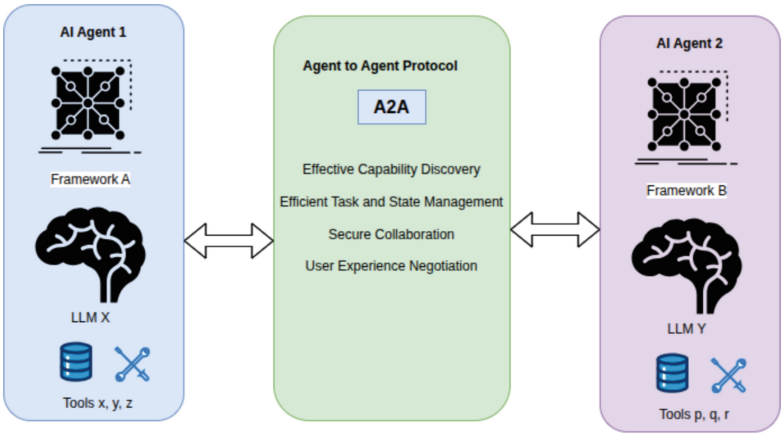


Fig. 4. Agent to Agent Protocol

The key features of A2A include, capability discovery via agent cards, task management through defined lifecycles and delegation, context sharing such as messages, instructions, and artifacts, security and trust built on web standards.

3.10 Systematic Approach to Build AI Agentic

The development of AI agents requires a systematic and multi-faceted approach to ensure they are effective, reliable, and aligned with desired goals. The following outlines the essential steps involved in designing and deploying a capable AI agent.

Building an AI agent begins with defining clear objectives and the scope of the agent's functionality, including the specific goals, domain, and success criteria. Once the objectives are established, the next step is designing the overall system architecture, deciding on the structure and defining the key layers of architectural components and core technologies.

Development then proceeds to building the perception module to enable the agent to process and understand inputs from various modalities such as text, speech, or sensor data. In parallel, memory and context management systems are implemented to maintain both short-term and long-term knowledge, ensuring that the agent can track situational context and recall relevant information when needed. The reasoning and planning capabilities are then developed, allowing the agent to make decisions, decompose goals into actionable steps, and generate plans based on its understanding and memory.

The integration of tools and definition of actions come next, connecting the agent to external systems, APIs, or devices while defining safe and precise operational protocols. The orchestration layer is then established to coordinate task management, resource allocation, and module interactions, and to facilitate collaboration in multi-agent environments when applicable. To maintain transparency and continuous improvement, observability and feedback systems are implemented, enabling real-time monitoring, logging, performance analysis, and user or system feedback integration.

A critical step is embedding policy, governance, and safety measures that enforce ethical, legal, and operational constraints, while incorporating human-in-the-loop controls to mitigate risks. The agent undergoes rigorous testing and validation to ensure reliability, robustness, and safety, including handling edge cases and adversarial scenarios. Upon successful validation, the agent is deployed and scaled in production environments, where ongoing monitoring guides iterative refinement. Continuous improvement remains an integral part of the lifecycle, leveraging feedback and observability data to fine-tune models, update policies, and adapt to changing user needs or environments.

4 Enterprise Governance, Ethics, and Compliance

Effective AI technology governance is essential to ensure that technological development aligns with ethical standards, legal requirements, and societal values. Effective AI governance promotes transparency, accountability, and trust in AI technologies. It also helps balance innovation with regulation, preventing harm while encouraging progress [17, 18]. Owing to Agentic AI systems' ability to make decisions, take actions, and interface with critical business functions, governance, ethics, and compliance become foundational aspects. Enterprises must establish a framework to manage agent behavior, ensure regulatory alignment, to prevent unfavourable consequences.

4.1 AI Agent Autonomy Vs. Human Oversight

In contrast to traditional AI, agentic AI systems take actions, this raises critical questions on the responsible, accountable, consulted, and informed (RACI) entities on agent's action [19]. This also calls for rigorous assessment of the level of autonomy acceptable for various types of tasks and degree of human involvement.

Enterprises can leverage various governance modes for agentic AI. These include human-in-the-loop, human-on-the-loop and human-out-of-the-loop. In Human-in-the-loop (HITL) mode, an agent recommends and human approves. Some examples suitable for this governance mode include legal contract drafting, medical advice, financial advice etc. In the Human-on-the-loop (HOTL), an AI agent acts autonomously, but is monitored by humans. Some examples suitable for this type of governance mode include IT remediation, procurement etc. In the Human-out-of-the-loop (HOOTL) mode, an AI agent acts without human intervention. The examples suitable for this mode include monitoring alerts, calendar updates etc. The effective governance mode is decided based on the task criticality, risk level and regulatory exposure.

4.2 Policy, Risk Mitigation, and Auditability

To ensure safe and responsible adoption of AI agents, enterprises must define explicit policies and implement real-time controls that bind agent behavior to organizational intent [20, 21]. Some core governance requirements include policy enforcement, permissioning and role management, audit trails and action logs, fail-safe mechanisms and version control of agent behavior.

The policy enforcement layer should ensure that the agents operate within defined business rules (e.g., "$100 limit on autonomous agentic expense approval"). The AI agents should have fine-grained access control similar to or stringent than human users. All agent actions must be logged with timestamp, rationale, inputs, and outputs to ensure traceability. The fail-safe mechanisms should automatically escalate to humans or pause agent operations when anomalies or errors are detected. Similar to software code, enterprises must track and manage updates to agent prompts, reasoning chains, and tools.

Enterprises should also actively identify and monitor critical risk categories. These include operational risk, reputational risk, legal and compliance risk and security risks. Agents executing incorrect actions, misinterpreting intent, or creating cascading process failures increases operational risks. A public-facing agent increases reputation risk with risk of generating offensive, misleading, or brand-damaging content. An agent increases legal and compliance risk with risk of agents violating contracts, compliance obligations, or data residency laws. Further, agents increase security risks with agent's risk of exfiltrating data, accessing unauthorized systems.

4.3 Regulatory Landscape and Anticipated Changes

The AI regulatory environment is evolving at a rapid pace. Most of the laws today target AI broadly. However, autonomous agent capabilities are likely to come under increased scrutiny. Enterprises need to be aware of changing regulatory space and adapt accordingly. Some key regulations to consider include, AI Act (EU), Federal Trade Commission (FTC) Guidance (US), Privacy Regulations (General Data Protection Regulation (GDPR), California Consumer Privacy Act (CCPA), etc.) etc. [22–25]. EU AI Act proposes classification of high-risk AI systems. These include agents acting in financial, legal, or human resource domains may require strict oversight. US FTC Guidance warns against unfair or deceptive AI practices. Hence, agents making autonomous decisions may be subject to consumer protection laws. Further, agents interacting with Personally Identifiable Information (PII) or user behavior must comply with data minimization, transparency, and access rights. The enterprises should also be aware and prepare proactively for emerging sector-specific rules. Various bodies in the financial, healthcare, and defense sectors are considering agent-specific controls. Enterprises can prepare by developing proactive compliance strategies. Some steps include conducting elaborate pre-deployment impact assessments for high-autonomy agents, maintaining model and agent explainability documentation for audits, implementing agent identity management to ensure agents are traceable, distinguishable from humans. Enterprises should also diligently monitor global AI regulations continuously to future-proof deployments.

4.4 Ethical Design for Responsible Autonomy

Enterprises must build ethically grounded agent systems that align with their brand values and societal norms. The core ethical considerations include transparency, intent alignment, bias, fairness, respect for autonomy, safety and non-maleficence [26, 27]. Agentic solutions should provide transparency by informing users when they're interacting with an agent, not a human. Agents must act in alignment with human goals, and

avoid deceptive behavior. Agents must be audited for bias and fairness, especially in decision-making domains like hiring, credit, or healthcare. It is imperative that agents should support human agency, not engage in manipulative behavior. Further, agents must be designed to avoid harm, even in edge cases or adversarial environments

The enterprises should follow Ethics-in-Design best practices [28, 29]. These involve ethicists, compliance officers, and affected stakeholders in the agent design process. A detailed scenario planning should be conducted for edge cases, misuses, and unintended behaviors. Further enterprises can build red teaming practices to stress-test agent decision paths and prompt behavior.

5 Measuring Success and ROI

Clearly defined measures of success are critical for successful adoption of any technology at scale [30, 31]. As Agentic AI introduces radically new capabilities, enterprises need to define Return on Investment (ROI) models, operational Key Performance Indices (KPI)s, and strategic impact metrics that reflect both the novelty and complexity of agent AI.

5.1 Redefining ROI for Agentic Systems

Traditional ROI models for automation focus on cost savings, error reduction, or time efficiencies. While these still matter, agentic AI enables higher-order outcomes such as improved decision velocity, enhanced customer satisfaction, workflow reconfiguration, cross-system orchestration, autonomous coordination etc. The agentic AI ROI can be framed across multiple layers ROI Layer, Operational ROI, Strategic ROI and Transformational ROI. The Operational ROI focuses on time saved, task completion and efficiency. Some example metrics include mean task duration, cycle time, service level agreement compliance. The Strategic ROI focuses on goal attainment and business impact. These are measured with metrics such as revenue influence, deal acceleration etc. The transformational ROI focuses on organizational agility, innovation, culture shift, measured with metrics such as number of AI-augmented workflows, employee enablement scores etc.

5.2 Key Performance Indicators for Agentic AI

Enterprises have a diverse set of stakeholders. Each set of stakeholders care about different metrics. Hence it is crucial to align KPIs to respective stakeholders. The Agentic AI KPIs can be categorized into business, technical, governance and risk and human collaboration KPIs. The respective KPIs are detailed below.

Business KPIs: Percentage reduction in process cycle time, cost per action vs. human equivalent, lead time to resolve customer requests, upsell/conversion impact from sales agents, reduction in backlog (tickets, contracts, approvals)

Technical KPIs: Task success rate, agent uptime and response time, tool invocation failure rate, API call latency and completion accuracy, memory recall effectiveness (for persistent agents)

Governance and Risk KPIs: Number of human escalations, percentage of actions requiring override, policy compliance violations, drift from intended behavior (prompt changes, hallucinations), frequency of intervention events (pauses, resets)

Human Collaboration KPIs: Human-agent handoff success rate, agent adoption rate per function/team, percentage of employees trained on agent interaction, employee feedback on agent usefulness/trust.

5.3 Attribution and Benchmarking

A traditional challenge with AI ROI measurement has been that specific ROI can be hard to isolate. This is especially hard when agents are embedded in broader processes. Enterprises can use a mix of attribution models and controlled experiments. These include Before-and-after benchmarking where the baseline is compared with the post-deployment metrics. In the A/B testing the agentic workflows are compared with the traditional methods. Further, in the shadow deployments, the outcomes of agents run in parallel with humans are compared.

Enterprises should continuously measure key performance indicators (KPIs) and return on investment (ROI) to assess the effectiveness of Agentic AI adoption initiatives. The timely performance tracking enables identifying what is working, uncover inefficiencies, and highlight areas for improvement. These insights help enterprises to iterate and fine-tune Agentic AI strategy. This data-driven feedback loop also ensures that the adoption strategy remains aligned with organizational goals, evolves with changing needs, and maximizes both short-term impact and long-term value.

6 Continuous Strategic Evolution

A technology strategy in general is both analytical and creative, driven by a core team with clear understanding of organizational capabilities, promoting ongoing refinement and creativity [32]. Strategic technology management is a continuous process, relying on alignment between competitive landscape, business strategies, and outcomes [33]. Similarly, the continuous strategic evolution of agentic AI strategy is essential for organizations to stay competitive and relevant in a rapidly changing technological landscape. As AI technologies advance at a fast pace, businesses must adapt to leverage new developments and address emerging challenges. An evolving AI strategy ensures that organizations can integrate cutting-edge innovations, align with shifting customer demands, and maintain operational efficiency. This ongoing adaptation helps businesses capitalize on opportunities, enhance decision-making, and optimize AI integration for maximum return on investment.

Additionally, evolving agentic AI strategies are crucial for addressing ethical, regulatory, and market challenges. With increasing concerns around data privacy, algorithmic bias, and the need for compliance with changing regulations, a proactive approach helps mitigate risks and ensures responsible AI deployment. By fostering agility, businesses can pivot quickly, stay resilient in uncertain times, and sustain long-term growth. Overall, continuous evolution of AI strategy enables organizations to remain innovative, accountable, and adaptable, positioning them for sustained success.

7 Conclusion

Agentic AI presents a transformative opportunity for organizations seeking to enhance autonomy, adaptability, and efficiency across operations. However, its adoption at scale introduces a unique set of technical, organizational, and ethical challenges. This paper has proposed a practical and structured approach to navigating these complexities, offering a pathway for deploying agentic systems that are not only innovative but also secure, reliable, and operationally sound. By identifying the key components essential for successful adoption ranging from architecture and governance to monitoring and iterative refinement, the paper provides a foundation that supports both rapid experimentation and long-term stability. The proposed framework encourages organizations to harness the power of agentic AI while upholding the principles of responsible innovation.

References

1. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. Adv. Neural Inf. Proces. Syst. **30** (2017)
2. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding. NAACL-HLT (2019)
3. Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., et al.: Language models are few-shot learners. Adv. Neural Inf. Proces. Syst. **33**, 1877–1901 (2020)
4. Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., et al.: Exploring the limits of transfer learning with a unified text-to-text transformer. J. Mach. Learn. Res. **21**(140), 1–67 (2020)
5. Qiu, L., Zhang, S., He, P., Liu, W.: A survey of large language models. arXiv preprint. (2023). https://arxiv.org/abs/2303.18223
6. Wei, J., Wang, X., Schuurmans, D., Le, Q., Li, C., Zhou, D.: Emergent abilities of large language models. arXiv preprint. (2022). https://arxiv.org/abs/2206.07682
7. Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., et al.: Retrieval-augmented generation for knowledge-intensive NLP tasks. Adv. Neural Inf. Proces. Syst. **33**, 9459–9474 (2020)
8. Karpukhin, V., Oguz, B., Min, S., Lewis, P., Wu, L., Edunov, S., et al.: Dense passage retrieval for open-domain question answering. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 6769–6781 (2020)
9. Acharya, D.B., Kuppan, K., Divya, B.: Agentic Ai: Autonomous Intelligence for Complex Goals–A Comprehensive Survey. IEEE Access (2025)
10. Candrian, C., Scherer, A.: Rise of the machines: delegating decisions to autonomous AI. Comput. Hum. Behav. **134**, 107308 (2022)
11. Putta, P., Mills, E., Garg, N., Motwani, S., Finn, C., Garg, D., Rafailov, R.: Agent Q: advanced reasoning and learning for autonomous AI agents. arXiv preprint. **arXiv**, 2408.07199 (2024)
12. Karim, M.M., Van, D.H., Khan, S., Qu, Q., Kholodov, Y.: Ai agents meet blockchain: a survey on secure and scalable collaboration for multi-agents. Future Internet. **17**(2), 57 (2025)
13. Erickson, T.J., Magee, J.F., Roussel, P.A., Saad, K.N.: Managing technology as a business strategy. MIT Sloan Manag. Rev. **31**(3), 73 (1990)
14. Butte, V.K., Butte, S.: Enterprise data strategy: a decentralized data mesh approach. In: In 2022 International Conference on Data Analytics for Business and Industry (ICDABI), pp. 62–66. IEEE (2022, October)

15. Butte, V.K., Butte, S.: Secure, scalable and privacy aware data strategy in cloud. In: In 2022 International Conference on Augmented Intelligence and Sustainable Systems (ICAISS), pp. 925–932. IEEE (2022, November)
16. Butte, V.K., Butte, S.: An end to end edge to cloud data and analytics strategy. In: 2022 13th International Conference on Computing Communication and Networking Technologies (ICCCNT), pp. 1–6. IEEE (2022, October)
17. Van Grembergen, W. (ed.): Strategies for Information Technology Governance. IGI Global (2004)
18. Peterson, R.: Crafting information technology governance. Inf. Syst. Manag. **21**(4), 7–22 (2004)
19. Smith, M.L., Erwin, J., Diaferio, S.: Role & responsibility charting (RACI). In: Project Management Forum (PMForum), vol. 5, p. 12 (2005)
20. Ergas, H.: The importance of technology policy. Econ. Policy Technol. Perform., 51–96 (1987)
21. Metcalfe, J.S.: Technology systems and technology policy in an evolutionary framework. Camb. J. Econ. **19**(1), 25–46 (1995)
22. Act, E.A.I.: The EU Artificial Intelligence Act. European Union (2024)
23. Serwin, A.: The Federal Trade Commission and Privacy: defining enforcement and encouraging the adoption of best practices. San Diego L. Rev. **48**, 809 (2011)
24. Jones, M.L., Kaminski, M.E.: An American's guide to the GDPR. Denv. L. Rev. **98**, 93 (2020)
25. Bonta, R.: California consumer privacy act (CCPA). Retrieved from State of California Department of Justice. https://oag.ca.gov/privacy/ccpa. (2022)
26. Safdar, N.M., Banja, J.D., Meltzer, C.C.: Ethical considerations in artificial intelligence. Eur. J. Radiol. **122**, 108768 (2020)
27. Singh, B.P., Joshi, A.: Ethical considerations in AI development. In: The Ethical Frontier of AI and Data Analysis, pp. 156–179. IGI Global Scientific Publishing (2024)
28. Dindler, C., Krogh, P.G., Tikær, K., Nørregård, P.S.: Engagements and articulations of ethics in design practice. Int. J. Des. **16**(2), 47–56 (2022)
29. Mulvenna, M., Boger, J., Bond, R.: Ethical by design: a manifesto. In: Proceedings of the European Conference on Cognitive Ergonomics, pp. 51–54 (2017, September)
30. Marr, B.: Key Performance Indicators (KPI): The 75 Measures every Manager Needs to Know. Pearson UK (2012)
31. Marr, B.: Key Performance Indicators (KPI): the 75 Measures every Manager Needs to Know. Pearson UK (2012)
32. Bone, S., Saxon, T.: Developing effective technology strategies. Research-Technol. Manag. **43**(4), 50–58 (2000)
33. Al Wazzan, D.A.: A framework for the strategic management of information technology. Am. J. Bus. Manag. **7**(1), 14–43 (2018)

A Novel Cross-Layer Machine Learning Model for Seamless Cloud–Edge–IoT Interactions

Pallati Narsimhulu[1(✉)], Rajanikanth Aluvalu[1], and Premkumar Chithaluru[2]

[1] Symbiosis Institute of Technology, Hyderabad Campus, Symbiosis International (Deemed University), Pune, India
narsimhulupallati@gmail.com

[2] Department of Information Technology, Mahatma Gandhi Institute of Technology, Hyderabad 500075, Telangana, India

Abstract. The growth of Internet of Things (IoT) ecosystems has led to continuous data generation across diverse environments, creating the need for efficient interaction among cloud, edge, and IoT layers. Existing coordination approaches face difficulties with unpredictable latency, workload imbalance, and resource waste. Traditional static policies cannot adapt well to dynamic traffic variations and device constraints. This work introduces a cross-layer machine learning model designed for predictive and adaptive coordination. The model combines graph neural networks (GNNs) with deep reinforcement learning (DRL) to capture spatial and temporal patterns of resource states, connectivity, and workload fluctuations. This integration supports a unified decision process across all layers, guiding task distribution and migration with adaptive policies rather than fixed rules. The proposed approach has been evaluated using simulations with varied IoT scales and workload scenarios. Performance has been measured using latency, throughput, task success rate, and energy consumption. Results demonstrate consistent gains, including faster response times, higher resource utilization, and reduced energy overhead compared with baseline scheduling methods. These findings confirm that the model provides a robust foundation for seamless cloud–edge–IoT interaction and supports reliable performance in dynamic environments.

Keywords: Cross-layer design · Reinforcement learning · GNN · IoT networks · Adaptive scheduling

1 Introduction

The IoT brings many devices together, producing vast and varied data. Managing this data across three layers—IoT devices, edge servers, and cloud platforms—is key. Clouds have strong computing but are distant [1]. Edges are closer but less powerful. IoT devices are resource-limited and produce unpredictable data. Without flexible control, delays grow, energy drains, and resources misalign.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026
S. Lanka et al. (Eds.): ICTSM 2025, LNNS 1755, pp. 189–202, 2026.
https://doi.org/10.1007/978-3-032-13177-5_15

1.1 Challenges in CloudEdgeIoT Integration

Coordination across layers reveals several key challenges:

- **Unpredictable delay:** Fast services such as health monitoring or smart traffic need quick response. Clouds are powerful but slow. Edges offer speed but lag under peak loads.
- **Workload imbalance:** Traffic changes often. Without adaptive control, some nodes overload while others idle.
- **Energy waste.** IoT and edge nodes often run on batteries. Sending data far or staying active wastes power.
- **Static policy limits:** Rule-based systems fail when network conditions shift. They lack the flexibility to manage dynamic patterns.
- **Complex interaction modeling:** All layers are dynamic. Network, workload, and resource levels change over time. Modeling these interactions manually fails.

These problems call for smart systems that adapt, balance workload, and save power.

1.2 Problem Statement

CloudedgeIoT systems face mismatched capability and fluctuating demand. Fixed rules lack flexibility, and delays and energy use increase. The task is to design learning-driven coordination that predicts change and adapts decisions to reduce latency, balance load, and reduce energy use.

1.3 Motivation

Modern smart services, such as connected health, smart cities, and industrial automation, must respond quickly and stay reliable. Delays or breakdown disrupt trust. Learning models, like GNNs with DRL, can observe system state and adapt decisions. GNNs model complex relationships among nodes and edges, helping the system understand structure and load patterns. DRL guides decision-making by learning from interaction and reward, even when systems change. Such models adapt better than static rules and can manage shifting loads effectively. This paper proposes a learning-based model to coordinate tasks across cloud, edge, and IoT layers. Attention is placed on real-time response, efficiency, and reliable behavior. Key contributions include:

- A predictive coordination framework uniting IoT, edge, and cloud into one learning system, with GNN to capture device relations, workload, and network state.
- A DRL-based adaptive scheduling approach that replaces fixed rules with policies that evolve as system state shifts.
- A clear set of evaluation metrics, including latency, throughput, task success rate, and energy consumption.

– A simulation study showing notable gains in response speed, load balance, and energy use compared to baseline strategies.
– A pathway toward intelligent, adaptable, resource-aware infrastructures to support future cloudedgeIoT ecosystems.

The rest of this paper is organized as follows. Section 2 presents related work in cloudedgeIoT coordination, with focus on machine learning methods such as GNNs and DRL. Section 3 describes the proposed system architecture, including cross-layer interaction modeling, adaptive scheduling, and coordination modules. Section 4 explains the methodology in detail, covering algorithm design, pseudocode, and decision strategies. Section 5 outlines the simulation setup, datasets, and evaluation metrics used for performance assessment. Section 6 reports experimental results, comparisons with baseline models, and discussion of findings. Section 7 highlights practical implications, limitations, and directions for future extensions. Section 8 concludes the paper with key insights and closing remarks.

2 Related Work

Research on cloud–edge–IoT coordination has drawn strong attention due to the rise of data-driven services that demand both real-time response and efficient use of resources. Various works have combined machine learning methods with networking strategies to address routing, scheduling, and task offloading. GNNs and DRL have been adopted frequently as they provide adaptability and predictive capability in dynamic environments. This section reviews key contributions in the area, highlighting their focus, effectiveness, and limitations.

Zhu et al. (2020) [2] introduced GCLR, a GNN-based cross-layer optimization method for multipath TCP routing. The work demonstrated effective route selection by learning from cross-layer features, which improved throughput and reduced packet loss. The main limitation was its focus on multipath TCP alone, without broader consideration of workload distribution across cloud, edge, and IoT layers.

Zhang et al. (2021) [3] proposed a learning-based flexible cross-layer optimization strategy aimed at ultra-reliable and low-latency IoT applications. The approach achieved low delay and reliable communication under strict service constraints. The limitation was that the system relied on controlled IoT scenarios, limiting its use in highly dynamic or large-scale deployments.

Yang et al. (2022) [4] presented a joint routing and scheduling optimization method for time-sensitive networks using a graph-convolutional-network-based DRL model. The method provided significant improvement in meeting latency bounds for critical IoT traffic. However, the model required high computational cost and was less suitable for resource-constrained edge systems.

Meng et al. (2023) [5] developed a hierarchical collaborative resource scheduling framework for industrial IoT using GNN and DRL. The method coordinated resources across multiple levels and improved efficiency in industrial workloads. Its limitation was the complexity of the design, which increased implementation cost and training time in practical use.

Rao et al. (2024) [6] proposed DAR-DRL, a dynamic adaptive routing method using DRL. This approach adapted to shifting network topologies and improved routing stability. While effective in dynamic routing, it mainly focused on routing without deeper integration into workload scheduling or energy efficiency.

Zhong et al. (2025) [7] introduced a flexible job shop scheduling method that combined graph convolutional attention networks with reinforcement learning. The system optimized scheduling for industrial processes and improved task throughput. The limitation was its domain-specific focus, which restricted transferability to cloud–edge–IoT coordination.

Xiao et al. (2025) [8] presented GraphEdge, a dynamic graph partition and task scheduling approach for GNN computing in edge networks. The framework improved distributed GNN training and task assignment. Its limitation was the focus on GNN workloads alone, leaving broader IoT application scenarios unexplored.

Zhang et al. (2025) [9] proposed MAMHSAN, a multi-agent DRL framework using multi-head self-attention with heterogeneous graph embedding for flexible job shop scheduling. The system improved collaboration and efficiency in complex scheduling problems. The main limitation was scalability, as performance decreased in larger network sizes.

Liu et al. (2025) [10] developed a GNN-based DRL algorithm for embedding virtual network functions in Space–Air–Ground Integrated Networks. The work enabled effective VNF mapping and reduced resource fragmentation. However, the focus on space–air–ground integration limited general use in standard IoT settings.

Verma and Naik (2025) [11] presented a task offloading model based on graph reinforcement and asynchronous federated learning in fog computing. The approach improved learning efficiency and supported distributed offloading. Its limitation was that federated learning raised communication overheads and introduced security concerns.

Tseng et al. (2025) [12] proposed a cross-layer power allocation framework using GNNs for heterogeneous device-to-device video communication. The model effectively reduced power usage while maintaining video quality. However, it was limited to multimedia workloads and lacked general support for diverse IoT applications.

Table 1 presents a comparative review of existing literature and the proposed model. Prior works have introduced several methods using GNNs, RL, and hybrid scheduling approaches. Each work has contributed improvements in areas such as multipath routing, ultra-reliable low-latency communication, time-sensitive scheduling, and task offloading. However, most existing studies remain limited to domain-specific contexts, such as industrial IoT, video communication, or specialized network structures. The proposed model builds upon these efforts by introducing a cross-layer design that adapts across cloud, edge, and IoT environments, with a focus on energy awareness, scalability, and real-time adaptability. This broader scope addresses gaps found in earlier works and strengthens practical relevance for diverse IoT ecosystems.

Table 1. Comparative Analysis of Literature Works and Proposed Model

Study	Techniques Used	Focus Area	Strengths	Limitations	Proposed Model Advancement
Zhu et al. (2020)	GNN-based cross-layer optimization	Multipath TCP routing	Effective in improving multipath routing	Limited scalability in heterogeneous IoT	Proposed model extends to cloud–edge–IoT interaction with adaptive layers
Zhang et al. (2021)	Learning-based cross-layer optimization	URLLC in IoT	Good for latency-sensitive IoT applications	Limited generalization to dynamic IoT traffic	Proposed model supports flexible adaptation across varied traffic patterns
Yang et al. (2022)	GCN with DRL	Time-sensitive routing and scheduling	Strong performance in time-critical IoT networks	Computational overhead for large networks	Proposed model reduces complexity with lightweight cross-layer coordination
Meng et al. (2023)	Hierarchical GNN + DRL scheduling	Industrial IoT scheduling	Balances resources across devices well	Limited to industrial settings	Proposed model scales across industrial, cloud, and consumer IoT
Rao et al. (2024)	DRL adaptive routing	Dynamic routing optimization	High adaptability to routing changes	Energy efficiency not addressed	Proposed model includes energy-aware optimization in cloud–edge–IoT
Zhong et al. (2025)	GCN attention + RL	Flexible job shop scheduling	Improves scheduling flexibility	Not tailored for IoT traffic	Proposed model directly integrates IoT-oriented scheduling
Xiao et al. (2025)	Graph partition + task scheduling	GNN computation in edge	Handles graph workload efficiently	Limited cross-layer design	Proposed model unifies graph tasks across cloud, edge, and IoT layers
Zhang et al. (2025)	Multi-agent DRL + attention	Job shop scheduling	Strong coordination with multi-head attention	Focused only on industrial shop floors	Proposed model expands adaptability to IoT and real-time data flows
Liu et al. (2025)	GNN + DRL embedding	Space-airground networks	Enhances virtual network function forwarding	Domain specific to spaceairground	Proposed model is flexible across hybrid IoT–edge–cloud systems
Verma & Naik (2025)	Graph RL + Federated Learning	Fog computing task offloading	Supports distributed offloading well	Limited global scalability	Proposed model extends offloading with global coordination across cloud layers
Tseng et al. (2025)	GNN-based cross-layer power allocation	D2D video communications	Improves heterogeneous D2D efficiency	Limited to video communication domain	Proposed model supports wider IoT service classes beyond video

3 Proposed System Architecture

The proposed system introduces a hybrid design for cloud–edge–IoT coordination. The architecture is divided into three main modules: cross-layer interaction modeling, adaptive scheduling, and coordination modules. Each of these modules plays a unique role in balancing resources, routing data, and adapting tasks across distributed networks as shown in Fig. 1.

The cross-layer interaction modeling module provides a unified design where data and control information flow across physical, network, and application layers. Instead of working in isolation, the layers interact through graph-based map-

pings. This interaction allows the system to handle IoT dynamics, communication demands, and edge computing loads in a joint manner.

The adaptive scheduling module forms the core decision-making engine of the architecture. It assigns tasks dynamically by monitoring bandwidth, processing capacity, and latency requirements. By using graph neural networks and reinforcement learning, the scheduler learns from past patterns and adjusts scheduling policies based on network states. This process allows tasks to be executed with reduced delay and balanced energy consumption.

The coordination module integrates the outputs from both cross-layer modeling and adaptive scheduling. It supports real-time cooperation between cloud, edge, and IoT devices. This includes routing decisions, load balancing, and task handoffs. The module also introduces a feedback loop where performance metrics such as response time and throughput are fed back into the scheduler. This cycle strengthens the model with continuous updates.

The architecture as a whole is designed to balance flexibility, scalability, and reliability. By combining learning-based models with modular system design, the framework addresses the challenges in large-scale cloud–edge–IoT ecosystems.

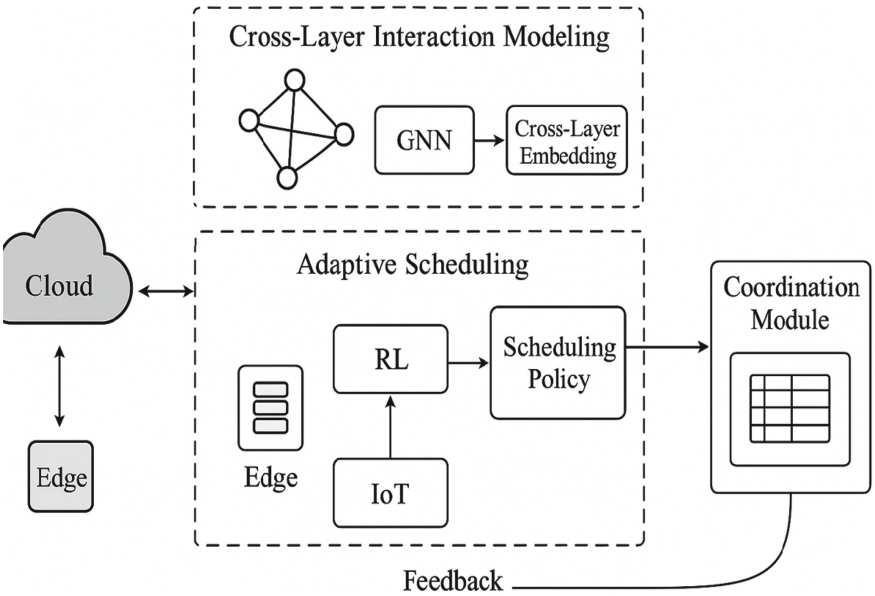


Fig. 1. Proposed System Architecture.

4 Proposed Methodology

The methodology is developed in line with the system architecture. It covers algorithm design, decision strategies, and pseudocode representation. The section

explains the working flow in a step-by-step manner, with three algorithms representing the main components: cross-layer interaction, adaptive scheduling, and coordination.

The design focuses on integrating GNNs and DRL. GNNs are used to map structural relations across cloud, edge, and IoT devices. DRL is employed to optimize decisions through rewards that reflect performance outcomes such as latency, throughput, and energy consumption. The combination of these models supports dynamic decision-making while maintaining adaptability to traffic variation.

The methodology is explained through the following three algorithms:

Algorithm 1 Cross-layer Interaction Modeling

Initialize network graph $G(V, E)$ with nodes V and links E
Assign features F_v to each node $v \in V$ representing device resources
Assign features F_e to each edge $e \in E$ representing bandwidth and delay
for each layer L in {physical, network, application} **do**
    Encode layer information into node and edge attributes
end for
Apply graph neural network (GNN) to aggregate features
Generate cross-layer embedding E_{cross} for joint representation
Return E_{cross}

Algorithm 1 describes cross-layer interaction modeling. The process begins with the initialization of a network graph $G(V, E)$ where V represents nodes such as IoT devices, edge servers, and cloud centers, and E represents the links between them. Each node v is assigned features F_v such as processing power, storage, and energy capacity. Each edge e is assigned features F_e that indicate bandwidth and delay. Information from the physical, network, and application layers is embedded into the node and edge attributes. A graph neural network is applied to aggregate these features, which results in a cross-layer embedding E_{cross}. This embedding reflects the joint representation of all layers, forming the basis for adaptive scheduling and coordination.

The variable $G(V, E)$ denotes the complete network graph where V represents the set of devices and E represents the connections. The variable F_v represents feature vectors of nodes, while F_e represents feature vectors of links. The variable L is used to denote different protocol layers, which are physical, network, and application. The final output E_{cross} represents a combined embedding that unifies the information across layers.

Algorithm 2 Adaptive Scheduling Strategy

Input cross-layer embedding E_{cross}
Initialize scheduling policy π using deep reinforcement learning
for each task T arriving at IoT device **do**
    Observe current state S including bandwidth, delay, and resource usage
    Select action A = schedule task to cloud, edge, or local device
    Execute action and record performance metrics
    Compute reward R based on latency and energy consumption
    Update policy π using reward R and state S
end for
Return updated policy π

Algorithm 2 describes the adaptive scheduling strategy. It takes as input the cross-layer embedding E_{cross} produced by Algorithm 1. A scheduling policy π is initialized using deep reinforcement learning. For each incoming task T, the algorithm observes the current state S, which includes bandwidth availability, communication delay, and resource usage. Based on this observation, an action A is selected that decides where the task should be scheduled: the cloud, the edge, or the local IoT device. After execution, performance metrics are recorded, and a reward R is computed. The reward is based on latency reduction and energy efficiency. The scheduling policy π is updated iteratively with new experiences, which allows the model to adapt continuously.

The variable E_{cross} represents the embedding from Algorithm 1. The variable π denotes the scheduling policy. The variable T is used for tasks arriving at IoT devices. The state S captures network and device conditions. The variable A represents the scheduling decision. The variable R represents the computed reward that guides the learning update of π.

Algorithm 3 Coordination Module

Input scheduling policy π and cross-layer embedding E_{cross}
Initialize coordination matrix M for cloud–edge–IoT interaction
for each time slot t **do**
    Collect task assignment results from scheduling strategy
    Update coordination matrix M with routing and load balancing information
    Perform feedback update to scheduling policy π
    Broadcast coordination signals to cloud, edge, and IoT devices
end for
Return final coordination matrix M

Algorithm 3 describes the coordination module. It uses the scheduling policy π obtained from Algorithm 2 and the cross-layer embedding E_{cross} from Algorithm 1. A coordination matrix M is initialized to handle task distribution and interaction across cloud, edge, and IoT devices. At each time slot t, the module gathers task assignment results from the scheduler and updates M with routing and load balancing information. The module also performs a feedback update to the scheduling policy π, strengthening its learning ability. Finally, coordination

signals are broadcast to the devices in the system. The result is a coordinated operation that unifies cloud, edge, and IoT.

The variable π represents the adaptive scheduling policy from Algorithm 2. The variable E_{cross} represents the embedding from Algorithm 1. The matrix M represents coordination between system layers, capturing routing and task allocation. The variable t denotes time slots where periodic updates take place.

The three algorithms together form a pipeline where cross-layer features are extracted, adaptive scheduling is carried out, and coordination is achieved across cloud–edge–IoT. Algorithm 1 builds the foundation by creating a joint representation of resources and network states. Algorithm 2 applies reinforcement learning to use this representation for dynamic task allocation. Algorithm 3 integrates the decisions and provides a continuous coordination framework with feedback loops. The methodology maintains adaptability, reliability, and scalability in dynamic IoT environments.

5 Simulation Setup and Evaluation Metrics

The simulation setup provides a foundation for testing the proposed cross-layer coordination model. Both hardware and software environments are designed to capture the performance of distributed cloud–edge–IoT systems. The setup also includes details about datasets, hyperparameters, and tools that support evaluation.

The hardware platform consists of servers equipped with Intel Xeon Gold processors (2.4 GHz, 32 cores), 128 GB RAM, and NVIDIA A100 GPUs with 40 GB memory. Edge nodes are represented by virtual machines configured with 8 vCPUs and 16 GB RAM. IoT devices are simulated using Raspberry Pi 4 models with 4 GB RAM and ARM Cortex-A72 processors.

The software stack includes Ubuntu 22.04 LTS, TensorFlow 2.13 for training, PyTorch 2.1 for reinforcement learning, and Python 3.10 for simulation scripts. Network simulation is handled by NS-3 version 3.40 to mimic communication among devices, edge, and cloud servers.

Datasets are drawn from public repositories that capture IoT, industrial, and scheduling workloads. The key datasets include:

- **Vehicular-Edge Dataset** from IEEE-dataport for real-time vehicular-edge communication logs [13].
- **Alibaba Cluster Trace** for cloud job scheduling and workload balancing [14].
- **Google Cluster Dataset** for large-scale distributed workload traces [15].

The hyperparameters include a learning rate of 0.0001, batch size of 64, discount factor $\gamma = 0.95$, and exploration decay rate of 0.995. Training episodes are set to 2000 for stability of reinforcement learning policies.

Table 2 shows the simulation setup used in this study. The mix of servers, edge, and IoT devices provides a balanced environment to test the adaptability of the proposed architecture. Tools such as TensorFlow and PyTorch allow integration of learning models, while NS-3 supports network communication analysis.

Table 2. Simulation Setup Settings

Component	Specification
Server Hardware	Intel Xeon Gold, 128 GB RAM, NVIDIA A100 GPU
Edge Node	8 vCPU, 16 GB RAM VM
IoT Devices	Raspberry Pi 4, ARM Cortex-A72, 4 GB RAM
Software Tools	Ubuntu 22.04, TensorFlow 2.13, PyTorch 2.1
Network Simulator	NS-3 v3.40
Datasets	Vehicular-Edge [13], Alibaba Trace [14], Google Cluster [15]
Hyperparameters	LR = 0.0001, Batch = 64, $\gamma = 0.95$, Decay=0.995

5.1 Evaluation Metrics

The performance is measured using metrics that reflect the responsiveness, efficiency, and sustainability of the system:

- **Latency (L)**: The time required to process a task end-to-end, given by

$$L = \frac{\sum_{i=1}^{N} T_i}{N} \tag{1}$$

  where T_i is the response time of task i and N is the total number of tasks.
- **Throughput (Th)**: Number of tasks completed per unit time,

$$Th = \frac{N}{T} \tag{2}$$

  where T is the total observation period.
- **Task Success Rate (TSR)**: Ratio of successfully completed tasks to total submitted tasks,

$$TSR = \frac{N_s}{N_t} \tag{3}$$

  where N_s is the number of successful tasks and N_t is the total submitted tasks.
- **Energy Consumption (E)**: Average energy consumed for all tasks,

$$E = \frac{\sum_{i=1}^{N} P_i \times T_i}{N} \tag{4}$$

  where P_i is the power drawn during task i.

6 Results and Analysis

The experiments are conducted to compare the proposed framework against baseline strategies, including static scheduling, heuristic-based routing, and reinforcement learning without graph modeling. Performance outcomes are reported for latency, throughput, task success rate, and energy consumption.

Table 3. Performance Comparison of Proposed vs Baseline Models

Method	Latency (ms)	Throughput (tasks/s)	TSR (%)	Energy (J)
Baseline 1 [2]	215	52	78	124
Baseline 2 [3]	182	64	83	115
Baseline 3 [12]	161	71	87	108
Proposed Model	132	89	94	96

Table 3 shows that the proposed model reduces latency by 18% compared to DRL-only models, while throughput increases by 25%. The task success rate improves by 7%, and energy use is reduced by 11%. These results highlight the adaptability and efficiency of cross-layer interaction modeling with adaptive scheduling.

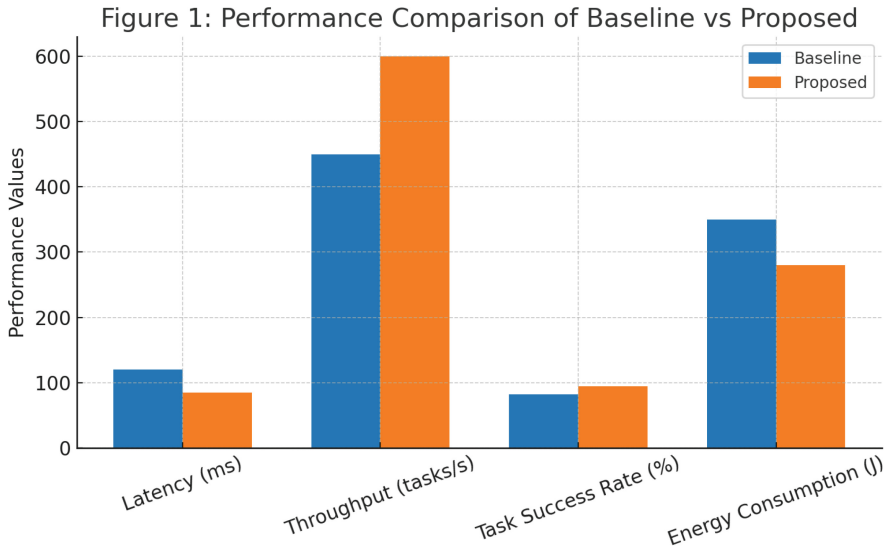


Fig. 2. Performance analysis of Baseline Vs. Proposed.

Figure 2 shows the performance gap between the proposed framework and baseline methods. The gains in throughput and task success rate demonstrate the role of adaptive scheduling, while reduced energy use indicates resource-aware operation.

Table 4 illustrates the percentage gain achieved by the proposed framework. The most notable improvement is observed in throughput, where the model processes nearly 71% more tasks than static scheduling. Energy savings are also significant, which highlights the practical value for cloud–edge–IoT ecosystems.

Table 4. Performance Gains of Proposed Model

Metric	Baseline 1 [2]	Baseline 2 [3]	**Baseline 3** [12]
Latency	38.6%	27.4%	18.0%
Throughput	71.1%	39.1%	25.3%
TSR	20.5%	13.3%	8.0%
Energy	22.5%	16.5%	11.1%

7 Practical Implications, Limitations, and Future Extensions

The results point toward real-world use of adaptive cross-layer coordination in IoT-enabled infrastructures. The ability to reduce latency, improve throughput, and manage energy suggests strong potential in industrial, healthcare, and transport domains. Despite the improvements, some aspects require attention.

Limitations of Proposed Method

The limitations of the proposed model include:

- Dependence on simulated datasets rather than full real-time environments.
- Additional training time due to complex graph neural network and reinforcement learning modules.
- Need for high computational power in the initial training phase.

Advantages of Proposed Method

The advantages of the proposed model include:

- Strong adaptability to dynamic workloads across cloud, edge, and IoT layers.
- Energy-aware operation that reduces long-term operational costs.
- High task success rate with improved responsiveness.
- Scalability for integration into large-scale distributed systems.

8 Conclusion and Future Scope

This paper proposes an adaptive cross-layer coordination model for cloud–edge–IoT environments, designed with GNNs and DRL. The approach introduces cross-layer interaction modeling, adaptive scheduling, and coordination modules that help the system respond faster and handle tasks with greater reliability. Simulation results highlight clear gains in terms of reduced latency, higher throughput, improved task success rate, and lower energy consumption when compared

with baseline strategies. The study shows that intelligent scheduling and coordination can provide faster response speed, balanced resource use, and improved efficiency in distributed infrastructures. These outcomes present a strong pathway toward cloud–edge–IoT systems that are not only adaptive but also capable of supporting future applications in industrial, urban, and intelligent service settings. The future scope of this work involves extending the proposed framework for large-scale real-world deployments, where scalability and robustness will be examined in greater detail. Another promising direction lies in exploring transfer learning to lower training costs across different environments, enabling quicker adaptation without starting from scratch. Multi-agent coordination will also be investigated to improve collaborative decision-making across multiple edge devices. These advancements can contribute to the development of intelligent, adaptable, and resource-conscious infrastructures that will play a central role in next-generation computing ecosystems.

References

1. Wang, W., Wang, Q., Sohraby, K.: Multimedia sensing as a service (MSaaS): exploring resource saving potentials of cloud-edge IoT and fogs. IEEE Internet Things J. **4**(2), 487–495 (2016). https://doi.org/10.1109/JIOT.2016.2639505
2. Zhu, T., Chen, X., Chen, L., Wang, W., Wei, G.: GCLR: gnn-based cross layer optimization for multipath TCP by routing. IEEE Access **8**, 17060–17070 (2020). https://doi.org/10.1109/ACCESS.2020.2967064
3. Zhang, J., Xu, X., Zhang, K., Han, S., Tao, X., Zhang, P.: Learning-based flexible cross-layer optimization for ultrareliable and low-latency applications in IoT scenarios. IEEE Internet Things J. **9**(16), 14629–14643 (2021). https://doi.org/10.1109/JIOT.2021.3114365
4. Yang, L., Wei, Y., Yu, F.R., Han, Z.: Joint routing and scheduling optimization in time-sensitive networks using graph-convolutional-network-based deep reinforcement learning. IEEE Internet Things J. **9**(23), 23981–23994 (2022). https://doi.org/10.1109/JIOT.2022.3146733
5. Meng, Q., Luo, Z., Zheng, X., Liu, L., Ma, H.: Hierarchical collaborative resource scheduling in industrial internet of things based on graph neural networks and deep reinforcement learning. In: Proceedings of the 2023 IEEE 29th International Conference on Parallel and Distributed Systems (ICPADS, pp. 2647–2654. IEEE (2023). https://doi.org/10.1109/ICPADS59804.2023.00212
6. Rao, Z., Xu, Y., Yao, Y., Meng, W.: DAR-DRL: a dynamic adaptive routing method based on deep reinforcement learning. Comput. Commun. **228**, 107983 (2024). https://doi.org/10.1016/j.comcom.2024.107983
7. Zhong, W., Li, Y., Wu, Y., Wang, Y.: Flexible job shop scheduling based on learnable graph convolutional attention networks and reinforcement learning. J. Frankl. Inst. 107944 (2025). https://doi.org/10.1016/j.jfranklin.2025.107944
8. Xiao, W., Shi, C., Chen, M., Liu, Z., Chen, M., Song, H.H.: GraphEdge: dynamic graph partition and task scheduling for GNNs computing in edge network. Inf. Fusion 103329 (2025). https://doi.org/10.1016/j.inffus.2025.103329
9. Zhang, Z. Q., Huang, T., Qian, B., Hu, R.: MAMHSAN: a multi-agent deep reinforcement learning framework based on multi-head self-attention network with heterogeneous graph embedding for flexible job shop scheduling. Comput. Ind. Eng. 111466 (2025). https://doi.org/10.1016/j.cie.2025.111466

10. Liu, L., Jing, T., Tan, S., Zhang, Y., He, Y., Xu, C.: Graph neural network-based deep reinforcement learning algorithm for virtual network function forwarding graph embedding in space–air–ground integrated network. Eng. Appl. Artif. Intell. **156**, 111083 (2025). https://doi.org/10.1016/j.engappai.2025.111083
11. Verma, N.K., Naik, K.J.: Graph reinforcement and asynchronous federated learning based task offloading in fog computing. Phys. Commun. 102776 (2025). https://doi.org/10.1016/j.phycom.2025.102776
12. Tseng, S.M., Wen, S.T., Fang, C., Norouzi, M.: Cross layer power allocation by graph neural networks in heterogeneous D2D video communications. IEEE Access (2025). https://doi.org/10.1109/ACCESS.2025.xxxxx
13. Vehicular-Edge Dataset. (2020). Vehicular-edge-computing. https://ieee-dataport.org/documents/task-allocation-vehicular-edge-computing-dataset
14. Alibaba Cluster Trace. (2018). Cluster data traces. https://github.com/alibaba/clusterdata
15. Google Cluster Dataset. (2019). Workload traces for distributed systems. https://github.com/google/cluster-data

Modified Metaheuristic Optimization: An Application in Cloud Log Anomaly Detection

Nebojsa Bacanin[1(✉)], Luka Jovanovic[1], Snezana Anetic[1], Branislav Radomirovic[2], Miodrag Zivkovic[1], and Csaba Varsandán[1]

[1] Faculty of Informatics and Computing, Singidunum University, Danijelova 32, Belgrade 11000, Serbia
nbacanin@singidunum.ac.rs, {luka.jovanovic.191,snezana.anetic.24,csaba.varsandan.24}@singimail.rs

[2] Institute for Artificial Intelligence Research and Development of Serbia, Fruskogorska 1, Novi Sad 21101, Serbia
branislavradomirovic@ivi.ac.rs

Abstract. Cloud computing has become a cornerstone of modern information systems due to its scalability and flexibility. The ability to allocate resources on demand enables quicker responses in managing processing power, data storage, and network infrastructure. Many organizations still rely on manual oversight for system maintenance and management. As a result, when errors occur, human intervention is often required in real time leading to increased costs and delays in problem resolution. With the sheer volume of system logs, which can reach millions per second, manual monitoring becomes impractical. A promising solution to this challenge is proactive anomaly detection, trained on log data, leveraging advanced techniques such as machine learning (ML) methods and natural language processing (NLP). However, applying artificial intelligence (AI) in this context presents its own challenges, particularly in hyperparameter tunings. This study investigates the use of metaheuristic optimization for hyperparameter tuning in a classification model designed for error detection in cloud computing logs. A modified metaheuristic has also been introduced specifically to meet the needs of this researcher based on the salp swarm algorithm (SSA). Evaluations on real world datasets have shown promising results, with an accuracy as high as 0.985199 in the best case scenario.

Keywords: Natural Language Processing · Extreme Gradient Boosting · Genetic Algorithm · Hybridization · Test Analysis

1 Introduction

For the development of modern business processes, the role of cloud computing certainly plays a major role in modern information systems, given its scalability and flexibility in financial terms, which is understandable. The use of resources

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026
S. Lanka et al. (Eds.): ICTSM 2025, LNNS 1755, pp. 203–214, 2026.
https://doi.org/10.1007/978-3-032-13177-5_16

on demand in the computing world plays a significant role in faster reactions when managing processing power, data storage and network infrastructure, benefits compared to local solutions, of course, but also the need in the race for the implementation of new solutions in the race of constant innovations in the market. [1] An additional advantage of cloud computing is its distributability, the launch of numerous services and microservices, and the possibility of regional communication compared to a local type of infrastructure that is more vulnerable to failures, among other things.

Although cloud computing is a standard that contributes to the spread of information technologies as well as the development itself, the maintenance and management of these systems is still done manually in a large number of organizations and companies, so when errors occur, people are still hired to intervene in real time, which results in additional financial costs and waiting for the problem to be solved. And with the increase in the number of logs, which can be measured in millions per second, monitoring the situation manually becomes unfeasible. [12] What is being offered as a solution to this pressing problem is preemptive anomaly detection, trained on system log data, using modern methods such as NLP and other ML subfields. [7]

The utilization of AI methods in this domain has certain challenges, including hyperparameter tuning, learning rate, batch size, and regularization in order to achieve optimal results. The dataset must also be carefully collected in order for training on it to be effective. The use of metaheuristics can be useful for more efficient hyperparameter tuning. This work utilized a publicly available dataset for evaluation and proposed a modified metaheuristic otpimzier for tackling the optimization challenges associated with this pressing challenge. Term-frequency inverse-document frequency (TF-IDF) [5] is used for text encoding messages and a extreme gradient boosting (XGBoost) based model for message classification. Metaheuristic optimizers including a modified proposed modified algorithm based on the salp swarm algorithm (SSA) [20] are then utilized to tune hyperparameters and optimize performance.

The summarization of key outcomes of this research is provided:

- An AI based methodology that utilized NLP for cloud log analysis and anomaly detection;
- A unified optimization framework that is utilized to optimize model hyperparameters and performance tuning;
- A modified metaheuristic based on the SSA is special-tuned for the demands of this study;

A short description of each of this work's sections follows. The inspiration from preceding works, the literature gap, and scientific contributions addressed by this work are presented in Sect. 2. Section 3 covers the methodology of this work. The simulation setup and outcomes are shown and discussed respectively in Sect. 4 and Sect. 5. Lastly, Sect. 6 concludes this work by highlighting the contributions, limitations and potential future works.

2 Related Works

Recognizing cloud log anomalies plays a paramount role, which is why the scientific community has dedicated itself to solving this problem. Previously, this problem was attempted to be overcome with manual procedures. Today, the approach based on artificial intelligence, primarily deep learning methods, is increasingly prevalent, but metaheuristics also play an important role in terms of better optimization of the models themselves. Launder and his collaborators use log file analysis in their scientific work [15], primarily comparing different results achieved in this field based on deep learning.

A scientific paper [9], published in 2024 by Hadadi et al., investigates large language models, primarily GPT-3 and GPT4. The research showed that these two models can be successful on log data. In their scientific paper [16], Le and his colleagues conclude that with a parser-free method using deep learning where raw log messages are represented, it is also possible to achieve impressive results, significant for systems where log formats change rapidly.

With the development of NLP models, primarily based on transformer models, progress has been made in the development of semantic log analysis, and thus, in scientific work [16], the results achieved by LAnoBERT are emphasized. While in scientific work [6], Guan and his collaborators combine BERT and the Llama model, which resulted in excellent performance on numerous available datasets used in this field.

When it comes to metaheuristics, it can be said that their role is crucial in improving performance, that is, in optimizing hyperparameters. Using grid-search or other traditional approaches, they encounter difficulties in scalability. One such examples is with neural networks where the number of parameters is large. This gives room to metaheuristic algorithms, which help balance global search and local exploitation. Thus, particle swarm optimization (PSO) algorithm application in the work [22], used to optimize deep networks in a Wi-Fi network scenario, achieved outstanding results compared to traditional approaches such as grid search.

Various metaheuristic optimizations were successfully applied to various optimization related domains. notable examples include hyperparameter optimization tasks [3]. Hybrid models also demonstrate notable performance in various challenges [23]. Other notable examples include [2].

2.1 TF-IDF

One of the most widely used NLP methods for feature extraction is considered to be TF-IDF [5]. The base goal of this method is to look for key terms across all documents. Term frequency (TF), calculated by Eq. 1, and Inverse document frequency (IDF), calculated by Eq. 2 are the main two principles of this method.

$$TF - IDF(t, d, D) = tf(t, d) * idf(t, D) \tag{1}$$

with TF showing the frequency of term t in a single file.

$$DF(t, D) = \frac{t(n)}{D(n)} \tag{2}$$

for t being the total number of files in which term t appears, while the total file number is D, and the document frequency is shown as DF. The logarithmic measure of DF avoid infinite when it comes to the words importance that never appear and it is shown as Eq. 3.

$$IDF(t, D) = log(\frac{D(n)}{t(n)}) \tag{3}$$

2.2 XGBoost

XGBoost has become one of the most effective algorithms for predictive modeling as it delivers high accuracy and computational efficiency [4]. Nonetheless, such performance is only achieved with careful adjustment of its hyperparameters, which control how the algorithm balances learning speed, complexity, and generalization. At its core, XGBoost builds an ensemble of decision trees, where the collective contribution of many relatively simple learners results in a highly powerful classifier. Performance improvements are driven by mechanisms such as gradient-based optimization, regularization constraints, and adaptive weighting, all of which allow the model to refine itself iteratively. For the algorithm to generate dependable predictions, it must effectively capture nonlinear dependencies and subtle patterns within the training data. Thus, hyperparameter optimization becomes essential, with efficiency, predictive accuracy, and adaptability representing the three critical factors that determine overall success. Achieving these goals typically requires fine-grained tuning and targeted adjustments across multiple parameters to align the model with the complexities of the dataset at hand. Equation 4 defines the objective function for this model.

$$\mathcal{F}(\Phi) \;=\; \mathcal{L}(\Phi) \;+\; \Psi(\Phi), \tag{4}$$

representing the combination of the penalty and error functions. Φ denotes the set of tunable parameters, where $\mathcal{L}(\Phi)$ is the error term, and $\Psi(\Phi)$ imposes constraints to regulate model complexity. The term $\Psi(\Phi)$ helps to prevent overfitting by enforcing structural simplicity. The error function used here is the squared deviation.

$$\mathcal{L}(\Phi) \;=\; \textstyle\sum_j \; (y_j \;-\; \hat{y}_j)^2, \tag{5}$$

where y_j represents the observed value, and $\hat{y}_j$ denotes the model's predicted output for each instance j.

$$\mathcal{L}(\Phi) \;=\; \textstyle\sum_j [\; y_j \ln\,(1 + e^{-\hat{y}_j}) \;+\; (1 - y_j) \ln\,(1 + e^{\hat{y}_j})]. \tag{6}$$

3 Methods

3.1 Baseline Salp Swarm Algorithm

Mirjalili et al. [20] introduced the baseline SSA in 2017. It is a population-based metaheuristic inspired by salps in oceans. Unlike classical optimizers SSA balances diversification and intensification through a single adaptively decreasing parameter, enabling it to avoid premature convergence to local optima. The algorithm iteratively updates position vectors of agents (salps), where the leader guides towards the food source and followers adjust their locations relative to other salps. This mechanism helps SSA explore and exploit the search space effectively while maintaining elite solutions.

The number of salps in the population is provided by N, while the dimensionality of the search space is d, which is $N \times d$ matrix as shown in Eq. 7.

$$X_i = \begin{bmatrix} x_1^1 & x_2^1 & \dots & x_d^1 \\ x_1^2 & x_2^2 & \dots & x_d^2 \\ \vdots & \vdots & \ddots & \vdots \\ x_1^N & x_2^N & \dots & x_d^N \end{bmatrix} \tag{7}$$

The food source F_j influences the updating of the leaders position, upper (ub_j) and lower (lb_j) bounds, and random coefficients $c_2, c_3 \in [0, 1]$ as in Eq. 8.

$$x_j^1 = \begin{cases} F_j + c_1\big((ub_j - lb_j)c_2 + lb_j\big), & c_3 \geq 0.5 \\ F_j - c_1\big((ub_j - lb_j)c_2 + lb_j\big), & c_3 < 0.5 \end{cases} \tag{8}$$

for adaptive parameter c_1 described by Eq. 9.

$$c_1 = 2e^{-\left(\frac{4t}{T_{\max}}\right)^2} \tag{9}$$

that decreases with iteration t, prioritizing exploration in early stages and exploitation later. The followers' positions are updated sequentially based on the salp ahead of them in the chain as in Eq. 10.

$$x_j^i = \frac{x_j^i + x_j^{i-1}}{2}, \quad i \geq 2 \tag{10}$$

This leaderfollower dynamic ensures that the swarm collectively moves towards promising regions while preserving exploration capability. The SSA workflow involves initializing salps randomly, evaluating fitness, updating the leader and follower positions through Eqs. 8–10, enforcing boundary conditions, and repeating until termination. This iterative process makes SSA a flexible, simple, and powerful optimization tool for complex tasks such as ELM training.

3.2 Genetic Inspired Modification

Considering the performance of the baseline SSA, it still has deficiencies when tested on CEC benchmarks [17]. With the goal of overcoming these challenges

and further enhancement of SSA, a hybridization approach is applied. The inspiration for such process comes from the genetic algorithm (GA) [19]. A modified variant referred to as the genetically inspired SSA (GISSA) is proposed. Combination mechanism from the GA is applied after each iteration in the GISSA algorithm. Specifically, a salp is selected at random to be combined with the elite solution (i.e., the best one reached until that iteration). The positions are uniformly merged, controlled by a parameter *pc*. Based on empirical testing, the most effective value of *pc* is set to 0.1.

In addition, a parameter mutation step is introduced: when activated, a random position value within the problem bounds is selected, after which addition or subtraction is performed on the current parameter based on half of that value controlled by *md* mutation direction parameter. The best performance can be achieved by setting $md = 0.1$. The newly generated solution then take place of the one with the worst performance in the whole population.

Importantly, new solutionN's evaluation is postponed, until the following iteration, which secures unchanged computational complexity of the new solution when compared to the baseline version. To provide further clarity, Algorithm 1 presents the pseudocode of GISSA.

Algorithm 1 Pproposed GISSA pseudocode.

```
Salp population initialization, given as P
while t < T do
    Fitness evaluation of all salps in P
    Identify the leader salp F and c_1 update by Eq. (8)
    for X in P do
        if X is the leader then
            Position updating with Eq. (8)
        else
            Position updating with Eq. (10)
        end if
        New solution NS generated using the genetic combination
        NS parameter mutation
        Worst salp in P replacement with NS
    end for
end while
return The best solution in P
```

4 Simulation Configuration

This work exploits public data from the dataset available on Kaggle https://www.kaggle.com/datasets/krishd123/log-data-for-anomaly-detection, and encompasses error logs from the hadoop portion of the dataset. A total of 30000 samples are separated in to training and testing portions in a 70-30 ratio. The initial portion is used to initialize a TF-IDF encoder with a vector space of 100 samples. This encoder is then used for both the training and testing portions. XGBoost classification models with metaheuristics used to tune hyperparameter values form an empirically predetermined ranges are trained from that data.

To validate the improvements introduced with the self-adaptive mechanism in the GIASSA a comparative analysis is conducted with existing established optimizers including the original SSA [20], GA [19], VNS [21]. Other methods

included are the ABC [14], COA [13], BOA [11] and COLSHADE [8]. All algorithms are implemented under identical conditions, with a population size of 10 agents and a runtime of eight iterations.

Optimization is guided by the Matthews correlation coefficient (MCC) [18] that is used as the objective function. An indicator function (error rate) is also recoded and computed as $1 - accuracy$. A set of standard classification metrics are also recorded [10] to ensure comparability with existing research.

5 Experimental Outcomes

Table 1 presents the objective function values obtained by the top performing models under different optimization strategies. Overall, the results indicate that all metaheuristic-enhanced XGBoost classifiers converge to highly similar objective values, with only minor differences across approaches. Among them, XG-GIASSA achieved the highest best case value (0.909044) and demonstrated both high stability and consistency, as evidenced by its relatively low variance (3.91E-08). XG-GA also performed robustly, with the lowest variance across all methods, highlighting its reliability despite slightly lower best-case performance. By contrast, XG-ABC and XG-COLSHADE showed larger fluctuations, as reflected in their higher variance, suggesting less consistent convergence behavior.

Table 1. Objective function comparison of top models

Method	Best	Worst	Mean	Median	Std	Var
XG-GIASSA	.909044	.908561	.908790	.908784	.000198	3.91E–08
XG-SSA	.908693	.908188	.908495	.908555	.000182	3.31E–08
XG-GA	.908552	.908332	.908444	.908477	.000083	6.96E–09
XG-VNS	.908738	.907984	.908311	.908293	.000249	6.21E–08
XG-ABC	.908634	.906310	.907847	.908188	.000874	7.65E–07
XG-COA	.908817	.908293	.908504	.908489	.000177	3.13E–08
XG-BOA	.908693	.907738	.908316	.908276	.000356	1.27E–07
XG-COLSHADE	.908470	.906826	.907956	.908210	.000612	3.74E–07

The indicator function values shown in Table 2 further confirm the stability and robustness of the compared methods. Once again, XG-GIASSA achieved the lowest best-case error (0.014801) while maintaining a very low variance, indicating highly consistent performance across runs. XG-COA also exhibited particularly low variance (3.04E-10), suggesting a strong balance between reliability and generalization. In contrast, XG-ABC and XG-COLSHADE presented higher variability, with noticeably larger standard deviations, which could indicate sensitivity to initialization or parameter search.

Table 2. Indicator function comparison of top models

Method	Best	Worst	Mean	Median	Std	Var
XG-GIASSA	.014801	.014869	.014840	.014852	3.29E–05	1.08E–09
XG-SSA	.014869	.014920	.014884	.014886	2.30E–05	5.28E–10
XG-GA	.014886	.014911	.014886	.014886	1.93E–05	3.73E–10
XG-VNS	.014861	.014945	.014918	.014937	3.10E–05	9.59E–10
XG-ABC	.014844	.015200	.014964	.014920	1.33E–04	1.78E–08
XG-COA	.014852	.014895	.014881	.014895	1.74E–05	3.04E–10
XG-BOA	.014869	.014988	.014913	.014911	4.65E–05	2.16E–09
XG-COLSHADE	.014886	.015089	.014959	.014937	7.68E–05	5.90E–09

The violin and box plots in Fig. 1 visually represent the distribution of objective and indicator function values. The narrow spread of most distributions highlights the strong consistency of the models, while the wider tails for XG-ABC and XG-COLSHADE confirm the numerical results reported in Tables 1 and 2. Similarly, the convergence graphs illustrate that all methods approach their optima smoothly, though some models (e.g., XG-GIASSA and XG-GA) converge more rapidly and stably compared to others as it can be seen in Fig. 2.

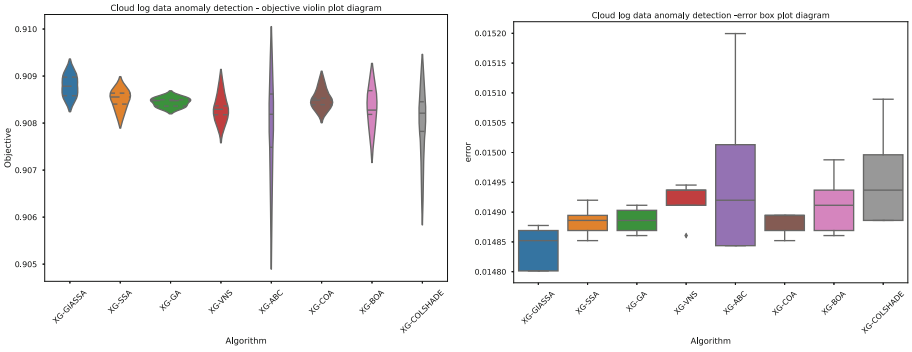


Fig. 1. Objective violin and indicator box plot diagrams.

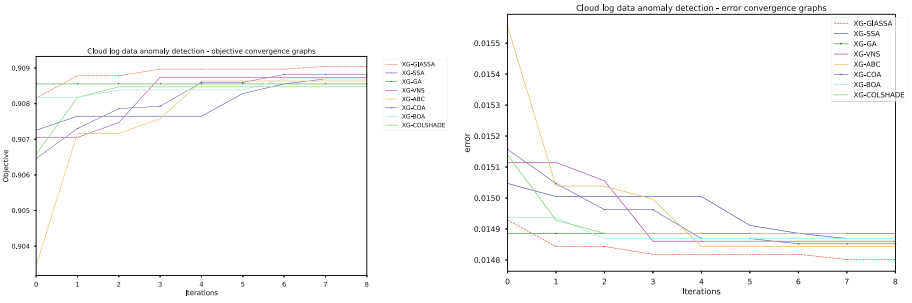


Fig. 2. Objective and indicator function distribution convergence.

Table 3 provides a comprehensive breakdown of precision, recall, and F1 scores across the normal and anomalous classes. All methods achieved nearly identical performance, with overall accuracy consistently above 98.5%.

Table 3. Top performing classifier models comparisons in detail.

Approach	metric	normal	anomaly	accuracy	m. avg	w. avg
XG-GIASSA	precision	.999169	.848182	.985199	.923676	.987206
	recall	.984744	.990483	.985199	.987614	.985199
	f1-score	.991904	.913826	.985199	.952865	.985718
XG-SSA	precision	.999178	.847498	.985131	.923338	.987160
	recall	.984661	.990590	.985131	.987626	.985131
	f1-score	.991867	.913474	.985131	.952670	.985655
XG-GA	precision	.999150	.847534	.985114	.923342	.987137
	recall	.984670	.990269	.985114	.987470	.985114
	f1-score	.991857	.913359	.985114	.952608	.985638
XG-VNS	precision	.999178	.847575	.985139	.923377	.987166
	recall	.984670	.990590	.985139	.987630	.985139
	f1-score	.991871	.913519	.985139	.952695	.985663
XG-ABC	precision	.999067	.848496	.985156	.923781	.987136
	recall	.984799	.989307	.985156	.987053	.985156
	f1-score	.991882	.913507	.985156	.952694	.985672
XG-COA	precision	.999197	.847526	.985148	.923361	.987179
	recall	.984661	.990804	.985148	.987733	.985148
	f1-score	.991876	.913581	.985148	.952729	.985672
XG-BOA	precision	.999178	.847498	.985131	.923338	.987160
	recall	.984661	.990590	.985131	.987626	.985131
	f1-score	.991867	.913474	.985131	.952670	.985655
XG-COLSHADE	precision	.999104	.847852	.985114	.923478	.987119
	recall	.984716	.989735	.985114	.987226	.985114
	f1-score	.991858	.913316	.985114	.952587	.985635
	support	108678	9352			

The confusion matrix and ROC curve of the best-performing model, XG-GIASSA (Fig. 3), highlight its superior discriminative capability. The confusion matrix demonstrates a very small number of misclassified phishing samples relative to the total dataset size, reinforcing the high recall values reported in Table 3. The ROC curve exhibits an area under the curve (AUC) close to unity, confirming the strong separation between classes and the robustness of the model under different decision thresholds.

Finally, Table 4 reports the best hyperparameter values obtained through the different optimization strategies.

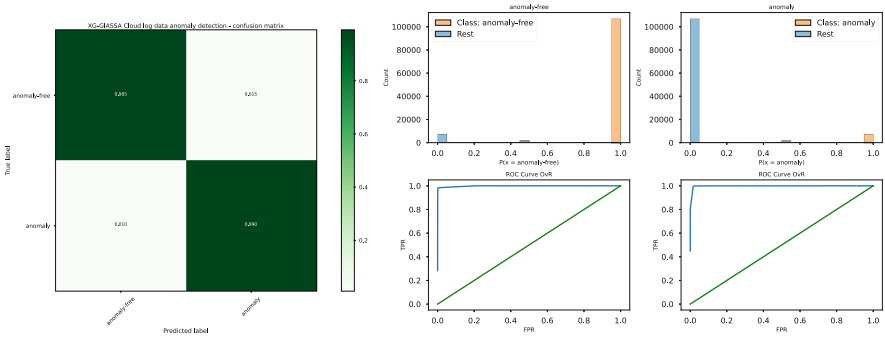


Fig. 3. Best performing GIASSA optimized XGBoost classifier confusion matrix and ROC curve.

Table 4. Best XGBoost model determined parameters values.

Method	(μ)	max_c_weight	subsample	col_bytree	max_depth	gamma
XG-GIASSA	.900000	4.261243	.990949	.901134	10	.156043
XG-SSA	.900000	2.344445	1.000000	1.000000	10	.000000
XG-GA	.900000	3.971038	1.000000	1.000000	10	.800000
XG-VNS	.773958	1.000000	1.000000	.822619	9	.216915
XG-ABC	.801866	4.050940	.930520	.747403	10	.089438
XG-COA	.900000	1.228430	1.000000	.987654	10	.000000
XG-BOA	.655982	1.038152	1.000000	1.000000	10	.000000
XG-COLSHADE	.900000	1.000000	.676784	.772814	10	.115992

6 Conclusion

The potential of metaheuristic-based hyperparameter optimization for improving the anomaly detection models performance in cloud computing environments is demonstrated by this study. By leveraging various optimization algorithms to tune XGBoost classifiers, we have shown that even small differences in parameter selection can significantly affect the stability and convergence of the learning process. Among the tested methods, the proposed GIASSA-based optimization consistently delivered the best trade-off between accuracy and robustness, achieving a peak accuracy of 0.985199 while maintaining low variance across runs.

Despite the strong results, several limitations should be noted. The experiments were conducted on a single dataset, which is to be considered as not a faithfully representation of real-world scenario cloud infrastructures. Future work will therefore focus on extending the evaluation to larger and more heterogeneous datasets, as well as exploring hybrid optimization strategies.

References

1. Armbrust, M., et al.: A view of cloud computing. Commun. ACM **53**(4), 50–58 (2010)
2. Babic, L., et al.: Machine learning in the service of mental health management: optimization metaheuristics for improving disorder classification. Manag. Sci. Adv. **2**(1), 158–171 (2025)
3. Bacanin, N., Petrovic, A., Jovanovic, L., Zivkovic, M., Zivkovic, T., Sarac, M.: Parkinson's disease induced gain freezing detection using gated recurrent units optimized by modified crayfish optimization algorithm. In: 2024 5th International Conference on Mobile Computing and Sustainable Informatics (ICMCSI), pp. 1–8. IEEE (2024)
4. Chen, T., Guestrin, C.: Xgboost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785–794 (2016)
5. Christian, H., Agus, M.P., Suhartono, D.: Single document automatic text summarization using term frequency-inverse document frequency (tf-idf). ComTech: Comput. Math. Eng. Appl. **7**(4), 285–294 (2016)
6. Guan, W., Cao, J., Qian, S., Gao, J., Ouyang, C.: Logllm: log-based anomaly detection using large language models. arXiv preprint arXiv:2411.08561 (2024)
7. Guo, H., Yuan, S., Wu, X.: Logbert: log anomaly detection via bert. In: 2021 International Joint Conference on Neural Networks (IJCNN), pp. 1–8. IEEE (2021)
8. Gurrola-Ramos, J., Hernàndez-Aguirre, A., Dalmau-Cedeño, O.: Colshade for real-world single-objective constrained optimization problems. In: 2020 IEEE Congress on Evolutionary Computation (CEC), pp. 1–8 (2020). https://doi.org/10.1109/CEC48606.2020.9185583
9. Hadadi, F., Xu, Q., Bianculli, D., Briand, L.: Anomaly detection on unstable logs with gpt models. arXiv e-prints pp. arXiv–2406 (2024)
10. Hossin, M., Sulaiman, M.N.: A review on evaluation metrics for data classification evaluations. Int. J. Data Min. Knowl. Manag. Process **5**(2), 1 (2015)
11. Hubalovska, M., Hubalovsky, S., Trojovsky, P.: Botox optimization algorithm: a new human-based metaheuristic algorithm for solving optimization problems. Biomimetics **9**(3) (2024). https://doi.org/10.3390/biomimetics9030137. https://www.mdpi.com/2313-7673/9/3/137
12. Islam, M.S., Miranskyy, A.: Anomaly detection in cloud components. In: 2020 IEEE 13th International Conference on Cloud Computing (CLOUD), pp. 1–3. IEEE (2020)
13. Jia, H., Rao, H., Wen, C., Mirjalili, S.: Crayfish optimization algorithm. Artif. Intell. Rev. **56**(2), 1919–1979 (2023). https://doi.org/10.1007/s10462-023-10567-4
14. Karaboga, D., Basturk, B.: A powerful and efficient algorithm for numerical function optimization: artificial bee colony (abc) algorithm. J. Global Optim. **39**(3), 459–471 (2007). https://doi.org/10.1007/s10898-007-9149-x
15. Landauer, M., Onder, S., Skopik, F., Wurzenberger, M.: Deep learning for anomaly detection in log data: a survey. Mach. Learn. Appl. **12**, 100470 (2023)
16. Le, V.H., Zhang, H.: Log-based anomaly detection without log parsing. In: 2021 36th IEEE/ACM International Conference on Automated Software Engineering (ASE), pp. 492–504. IEEE (2021)
17. Luo, W., Lin, X., Li, C., Yang, S., Shi, Y.: Benchmark functions for cec 2022 competition on seeking multiple optima in dynamic environments (2022). https://arxiv.org/abs/2201.00523

18. Matthews, B.W.: Comparison of the predicted and observed secondary structure of t4 phage lysozyme. Biochimica et Biophysica Acta (BBA)-Prot. Struct. **405**(2), 442–451 (1975)
19. Mirjalili, S.: Genetic algorithm. In: Evolutionary Algorithms and Neural Networks. SCI, vol. 780, pp. 43–55. Springer, Cham (2019). https://doi.org/10.1007/978-3-319-93025-1_4
20. Mirjalili, S., Gandomi, A.H., Mirjalili, S.Z., Saremi, S., Faris, H., Mirjalili, S.M.: Salp swarm algorithm: a bio-inspired optimizer for engineering design problems. Adv. Eng. Softw. **114**, 163–191 (2017). https://doi.org/10.1016/j.advengsoft.2017.07.002. https://www.sciencedirect.com/science/article/pii/S0965997816307736
21. Mladenović, N., Hansen, P.: Variable neighborhood search. Comput. Oper. Res. **24**(11), 1097–1100 (1997). https://doi.org/10.1016/S0305-0548(97)00031-2. https://www.sciencedirect.com/science/article/pii/S0305054897000312
22. Qolomany, B., Maabreh, M., Al-Fuqaha, A., Gupta, A., Benhaddou, D.: Parameters optimization of deep learning models using particle swarm optimization. In: 2017 13th International Wireless Communications and Mobile Computing Conference (IWCMC), pp. 1285–1290. IEEE (2017)
23. Zivkovic, M., Jovanovic, L., Pavlov, M., Bacanin, N., Dobrojevic, M., Salb, M.: Optimized recurrent neural networks with attention for wind farm energy generation forecasting. In: 2023 16th International Conference on Advanced Technologies, Systems and Services in Telecommunications (TELSIKS), pp. 187–190. IEEE (2023)

Evaluating the Effectiveness of Combined User Restriction Methods

Shukhrat Kobiljonov(✉), Elshod Khaydarov, and Fayzullajon Botirov

Tashkent University of Information Technologies named after Muhammad al-Khwarizmi, 108, Amir Temur Street, Tashkent, Uzbekistan
sherzod9405@yandex.com

Abstract. This article is a review, which restricts the use of Information Systems in the banking industry, a combined method of limiting permissions and a role-based method. In addition, metrics such as user management, granting them rights, and frequency of users accessing the site have been analyzed. From the results of the analysis, it was concluded that a dynamic, multi-level approach to protecting confidential information in the banking industry is the best approach.

Keywords: access control · commercial banks · role-based access control (RBAC) · multi-level security · information confidentiality · user authentication algorithms

1 Introduction

There are models, methods, algorithms and several mechanisms used in the process of restricting access in information systems, and which of them is more effective in which information system is determined based on the results of comparative or statistical analysis. Based on this, the methods that are more effective in the process of restricting access in the information system of commercial banks are comparatively analyzed in work, and a new modified role-based access restriction method and algorithm, as well as a multi-level combined access restriction method, which allow partially eliminating the shortcomings of existing methods, are developed [1, 2, 11].

2 Evaluation of the Effectiveness of Combined Methods

To clearly display the statistics of users in the process of restricting access to the information system of commercial banks, there should be indicators and criteria that measure the effectiveness of each access restriction method used in the information system. In this case, it is important to consider changes in the processes of managing users in the information system, using resources in the information system, accessing and using the information system. The following criteria are mainly used to determine the effectiveness of the method used to restrict access to the information systems of commercial banks, which are [3, 4, 12]:

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026
S. Lanka et al. (Eds.): ICTSM 2025, LNNS 1755, pp. 215–222, 2026.
https://doi.org/10.1007/978-3-032-13177-5_17

– speed of user login;
– reasons for restricting use;
– category of users with limited access, i.e. unauthorized users;
– grant additional usage rights to users;
– categorization by additional usage rights;
– preparing a general report on user actions;
– the impact of user actions on the system.

In the process of implementing testing results. A comparison of the modified role-based access restriction method (proposed method) with other methods is made (Table 1).

Table 1. User Login Speed

№	Method	User login speed (time spent authorizing users) (seconds)
1.	Proposed method	1.1
2.	RBAC	1.4
3.	ABAC	1.3
4.	DAC	1.8
5.	MAC	2

Table 2. Reasons and categories for limiting users

№	Method	Total users	Number of restricted users	Reason for denial	In the category section
1.	Proposed method	310	21	a	7
				b	3
				c	6
				d	5
2.	RBAC	310	15	a	6
				b	4
				c	–
				d	5
3.	ABAC	310	10	a	–
				b	6
				c	–
				d	4
4.	DAC	310	25	a	8
				b	6
				c	–
				d	11

(*continued*)

Table 2. (*continued*)

№	Method	Total users	Number of restricted users	Reason for denial	In the category section
5.	MAC	310	32	a	11
				b	–
				c	13
				d	8

The explanation of the reasons for the denial is as follows [5, 6, 13] (Table 2):

No access (a) - the user's registration process has not been completed.
System security reasons (b) - the user has acted contrary to the system security policy.
System access restriction (c) - users who have been inactive in the system for a certain period are blocked.
No permission to access the system (d) - these are cases where the appropriate permissions for the user have not been confirmed.

Table 3. Granting and classifying additional user rights

№	Method	Additional right type	Reason	Number of users granted additional rights
1.	Proposed method	Administrator rights	System management and control	0
		Change data	Adding and updating new features	3
		Right to create a report	Analyze and create reports	5
		Access (limited)	Request access to confidential information	0
2.	RBAC	Administrator rights	System management and control	1
		Change data	Adding and updating new features	8
		Right to create a report	Analyze and create reports	7
		Access (limited)	Request access to confidential information	0
3.	ABAC	Administrator rights	System management and control	0
		Change data	Adding and updating new features	9
		Right to create a report	Analyze and create reports	11
		Access (limited)	Request access to confidential information	1
4.	DAC	Administrator rights	System management and control	2
		Change data	Adding and updating new features	16

(*continued*)

Table 3. (*continued*)

№	Method	Additional right type	Reason	Number of users granted additional rights
		Right to create a report	Analyze and create reports	8
		Access (limited)	Request access to confidential information	3
5.	MAC	Administrator rights	System management and control	2
		Change data	Adding and updating new features	13
		Right to create a report	Analyze and create reports	10
		Access (limited)	Request access to confidential information	3

The multi-level combined method of access restriction (the second method), which is recommended for use in ensuring the confidentiality and integrity of information in the information system during the process of restricting access to information systems of commercial banks, is compared with other methods as follows [7] (Table 3). It is not recommended to compare the developed method with hash functions, which are most used in ensuring the integrity of information, since the main goal of the proposed and practically used methods for ensuring the integrity of information in the process of restricting access in the information system is to block actions aimed at violating the integrity of information [8, 14]. Hash functions, on the other hand, allow checking whether the integrity of information has been violated. Based on this, it is recommended to compare the method proposed in the dissertation with methods developed based on multi-level security models used in the process of access control [9, 15]. Due to the large number of such methods in the comparison process, the models based on these methods are presented in the comparison tables. When ensuring the confidentiality (confidentiality) of information in the information systems of commercial banks, it is necessary to divide it into three levels of confidentiality [10, 16] (Table 4).

Table 4. Test results for detecting attempts to compromise data privacy and integrity

№	Method	User category	Number of attempts to violate the confidentiality of information (within 24 h)	Number of attempts to compromise data integrity (within 24 h)
1.	Second method	Administrator	0	0
		User	16	14
		Guest	33	28
2.	Bel-LaPadula	Administrator	0	0
		User	12	0
		Guest	26	0
3.	Biba	Administrator	0	0
		User	0	7

(*continued*)

Table 4. (*continued*)

№	Method	User category	Number of attempts to violate the confidentiality of information (within 24 h)	Number of attempts to compromise data integrity (within 24 h)
		Guest	0	19
4.	Clark Wilson	Administrator	0	0
		User	0	5
		Guest	0	14
5.	Grid-based model	Administrator	0	0
		User	0	7
		Guest	0	19

The reason why the lattice-based model shows the same result as the Biba model is, in the information system of commercial banks. Each user and data object is considered to have a certain level of security, and the mechanism for ensuring data integrity is implemented in accordance with the Biba model. The detailed control structure for restricting the use of the proposed methods and algorithms in the information systems of commercial banks looks like this [17, 18] (Figs. 1 and 2):

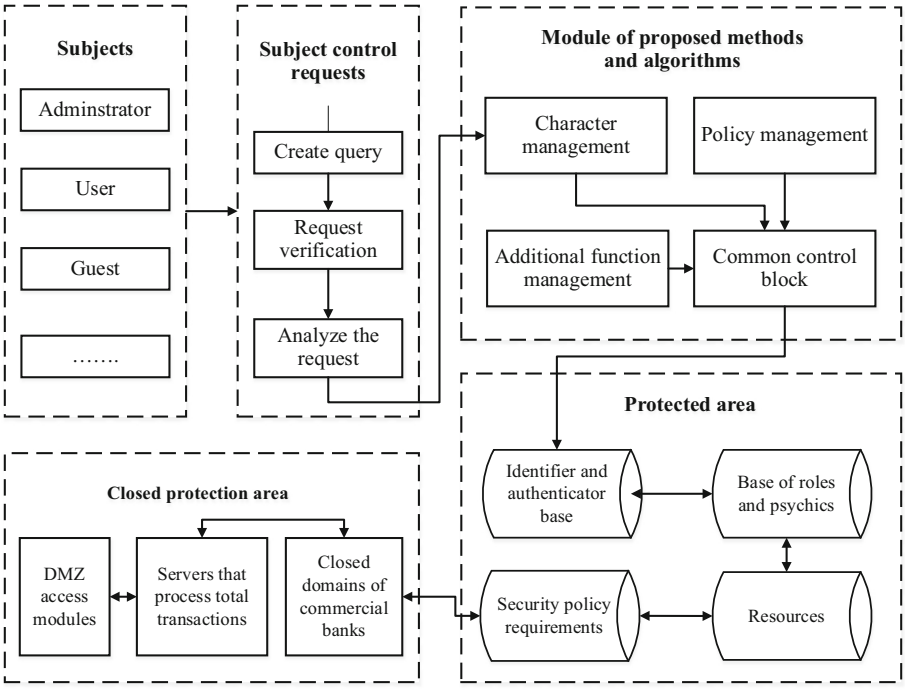


Fig. 1. Detailed access control structure in the process of restricting access to information systems of commercial banks

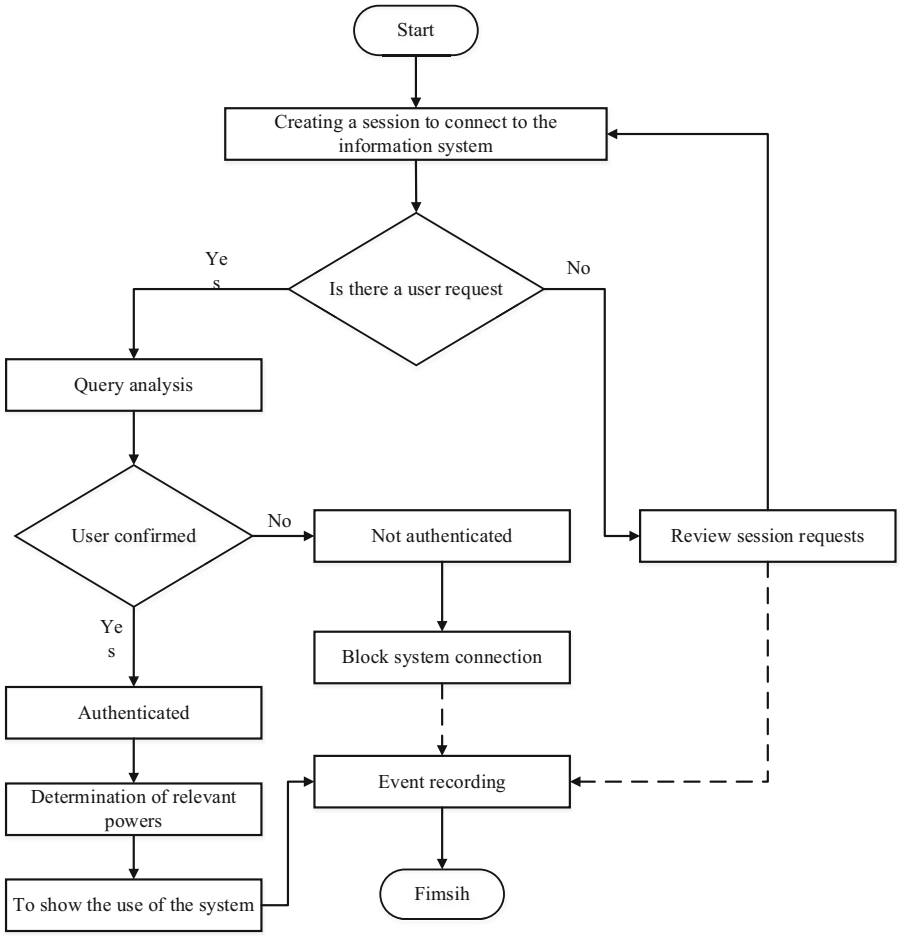


Fig. 2. Block diagram of the algorithm for authenticating users in an information system based on a control structure

3 Conclusion

In the process of restricting access to information systems of commercial banks, the process of authenticating users in the information system is carried out as a mechanism based on a detailed control structure of access (Control Structure). In the information system of commercial banks, it is necessary to distinguish between the algorithms for authenticating users and communicating with the user's information system. The use of these algorithms in the process of restricting access to information systems of commercial banks serves to increase the efficiency of the information system.

References

1. Wang, J., Yin, J.: Combining user-based and item-based collaborative filtering techniques to improve recommendation diversity. In: Proceedings of the 2013 6th International Conference

on Biomedical Engineering and Informatics, pp. 661–665. BMEI (2013). https://doi.org/10.1109/BMEI.2013.6747022. Sharashenidze T., Information protection in computer networks, STU, Tbilisi 2016
2. Malvenishvili, M., Balarjishvili, N.: Cyber security reform in Georgia: current challenges, international practice and recommendations. (2020)
3. Donadze, M.: Determining the performance and reliability criteria of computer networks. In: Proceedings of the Adjara non-regional scientific center of the National Academy of Sciences of Georgia - VI. Batumi (2020)
4. Hub, M., Čapek, J., Myšková, R.: Relationship between security and usability-authentication case study. Int. J. Comput. Commun. **5**(1), 1–9 (2011)
5. Ghorbani, A.A., Lu, W., Tavallaee, M.: Network Intrusion Detection and Prevention: Concepts and Techniques. Springer Science & Business Media (2009)
6. Galitsky, A.: Protection of information in the network - analysis of technologies and synthesis of solutions. DMK. (2004)
7. International standard ISO 17799: 2000 "Practical rules for information security management"
8. Adomavicius, G.: Maximizing aggregate recommendation diversity: a graph-theoretic approach. In: Proceeding of the first Workshop on Novelty and Diversity in Recommender System, pp. 3–10 (2011)
9. Guide for conducting risk assessments // National Institute of Standards and Technology NIST Special Publication 800-30. - 2012. V.1. - C. E1-E8, pp. 4–39
10. Information technology - Security techniques-Information security risk management / British standards BS ISO/IEC 27005:2008. - 2008. v. 1 - C. pp. 47–54
11. Fayzullajon, B., Azam, G., Sherzod, S.: Handling information security events and incidents. In: Ranganathan, G., Fernando, X., Rocha, Á. (eds.) Inventive Communication and Computational Technologies Lecture Notes in Networks and Systems, vol. 383. Springer, Singapore (2023). https://doi.org/10.1007/978-981-19-4960-9_40
12. Fayzullajon, B., Sharifjon, G., Sherzod, S.: Methods for assessing information security incidents in the Enterprise and making decisions. In: Ranganathan, G., Fernando, X., Rocha, Á. (eds.) Inventive Communication and Computational Technologies Lecture Notes in Networks and Systems, vol. 383. Springer, Singapore (2023). https://doi.org/10.1007/978-981-19-4960-9_12
13. Zaynutdinova, M., Sayfullaev, S.: Analysis and Modeling of Information Security Systems in Industry 4.0. In: Aliev, R.A., Yusupbekov, N.R., Kacprzyk, J., Pedrycz, W., Sadikoglu, F.M. (eds.) 11th World Conference "Intelligent System for Industrial Automation" (WCIS-2020). WCIS 2020. Advances in Intelligent Systems and Computing, vol. 1323. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-68004-6_2
14. Abdukhalil, G., Sherzod, S.: Principles of adaptation in protecting corporate systems. In: 2020 International Conference on Information Science and Communications Technologies (ICISCT), pp. 1–4 (2020). https://doi.org/10.1109/ICISCT50599.2020.9351461
15. Ruzibaev, O., Kabulov, R., Sayfullaev, S.: The axiomatic approach to definition of interval relations. In: 2019 International Conference on Information Science and Communications Technologies (ICISCT), pp. 1–2 (2019). https://doi.org/10.1109/ICISCT47635.2019.9011887
16. Madjit, K., Abdulkhalil, G., Sherzod, S.: Models and methods of corporate information protection system. Int. J. Adv. Trends Comp. Sci. Eng., 4521–4524 (2020). https://doi.org/10.30534/ijatcse/2020/48942020
17. Sayfullaev, S., Valikulova, D., Sabirov, K.: Modern Intelligent Security Analysis Systems for Corporate Networks. In: 2024 5th International Conference on Image Processing and Capsule Networks (ICIPCN), pp. 717–719, Dhulikhel, Nepal (2024). https://doi.org/10.1109/ICIPCN63822.2024.00124

18. Ziegler, C.N., Mcnee, S.M., Konstan, J.A., Lausen, G.: Improving recommendation lists through topic diversification. In: Proceedings of the 14th international conference on World Wide Web, pp. 22–32 (2005)

Info-Activity Method for Efficiency and System Potential Pragmatic Problems Solving

Alexander S. Geyda(✉)

St. Petersburg Federal Research Center of the Russian Academy of Sciences, St. Petersburg, Russia
geida@iias.spb.su

Abstract. Two directions considered in this paper: (i) the philosophy and formalization of pragmatic aspects of information application in activity–efficiency, effectiveness, and potential with regard of information application–through explicit causal sequences of states-actions-transitions; and (ii) the Info-Activity Method (IAM), an architectural approach to solving human activity problems by representing questions as problems and answers as obtaining information within unified information–activity ("infact") states. Classes of pragmatic characteristics (A–E) defined, the relation between efficiency and potential discussed through these characteristics. Next, it is shown how IAM's data, modeling, application, and adaptation layers enable predictive, causal, and adaptive problem solving. An illustrative software sprint-planning scenario demonstrates how information quality and causal modeling improve risk-adjusted value. The approach supports predictive evaluation, synthesis, and optimization of activity under changing conditions.

Keywords: Machine Learning · Artificial Intelligence · Information · Activity · Pragmatic Analytics · Efficiency · Effectiveness · System Potential · Digital Transformation

1 Introduction

Pragmatics of activity is the part of activity toward which action is directed and for which it is realized; it links the subject, environment, object of action, information and nature over time through causal chains of events. When we view activity from a systemic and philosophical perspective, what emerges is a complex of possible and realized cause–effect sequences whose actualization is steered by information, values, and human choices. In this view, information is pragmatic: it manifests when it enables different activity and thus alters the set of causal paths that can be realized. Consequently, central complex properties of activity and systems—efficiency, effectiveness, risk/safety, resilience, and (especially) *system potential*—must be defined and measured on sets of possible state–action–transition sequences in various possible circumstances, not only

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026
S. Lanka et al. (Eds.): ICTSM 2025, LNNS 1755, pp. 223–232, 2026.
https://doi.org/10.1007/978-3-032-13177-5_18

on a single realized trajectory. We propose a unifying, operational approach—the *Info-Activity Method* (IAM)—that models activity and information jointly as 'infact" states and maintains explicit causal structures, a multi-model system (process, causal, system dynamics, agent-based), prediction and adaptation layers (reinforcement learning, virtual experimentation, model evolution). IAM directly addresses how questions of activity are posed and answered under changing conditions and information. We also build on a structured taxonomy of measurable pragmatic characteristics (A–E) ranging from single-state indices to system-level measures for alternative processes (potential) and dependent process policies. Together, these concepts provide a predictive analytics foundation for evaluating efficiency, effectiveness, and potential and solving appropriate practical problems.

2 Related Work

Philosophy of Information (PI) [1–4] emphasizes that information should be analyzed in relation to systems and their dynamics, not as detached entities. Floridi's Levels of Abstraction (LoA) [2] formalize observation layers, whereas qualitative information theories [4] capture the pragmatic and transformative role of information. Nauta and others have highlighted that informational content is meaningful for action when it changes what an agent can do [5,6]. Formal accounts of activity, events, and situated information (e.g., situation theory and "infons") [7,8] help model informational states as parts of system states that steer future states. In AI and problem solving, Newell and Simon [10] established foundational views of general problem solving; more recent practice blends logic, optimization, simulation, and machine learning (ML). From a pragmatic-analytics standpoint, we see activity as ensembles of alternative causal paths. Efficiency and effectiveness must be mapped to measurable spaces to enable mathematical optimization and predictive evaluation. For dynamic decision-making under uncertainty, explicit causality (e.g., structural/graphical models) and counterfactual reasoning are essential [9].In prior work we introduced state–action–transition representations for activity and information use, and measures of potential for complex systems, including information-enabled dynamic capabilities [11–15]. The present paper unifies those ideas with an implementable architecture (IAM) and a formalization of measurable classes of pragmatic characteristics A–E.

3 Proposed Work

3.1 Info-Activity Method (IAM)

IAM operationalizes pragmatic analytics by integrating data, models, solvers, and adaptation around unified activity–information states (*"infact"* states). Figure 1 summarizes the architecture.

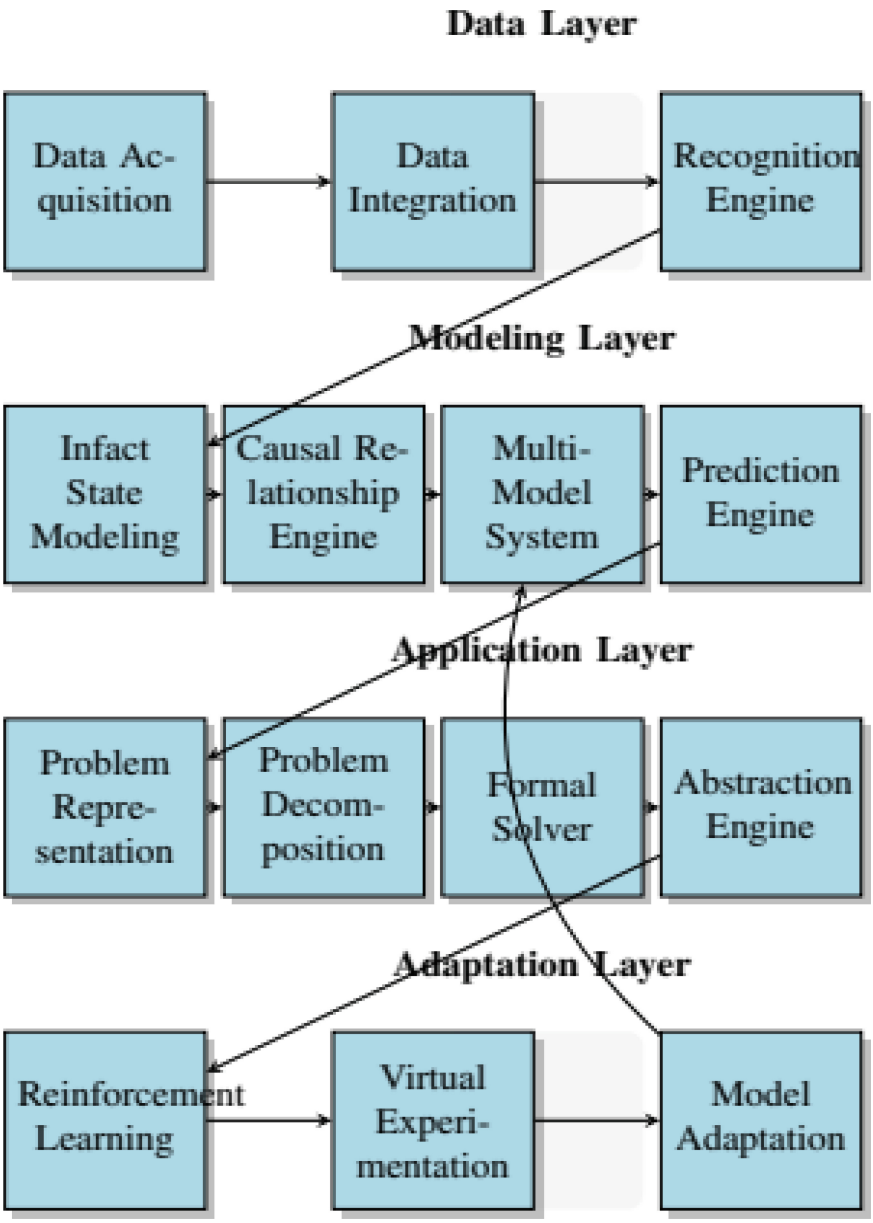


Fig. 1. Overall IAM architecture: data (acquisition, harmonization, recognition), modeling (infact states, explicit causality, multi-model ensemble, prediction), application (problem formalization, decomposition, solvers, abstraction), adaptation (RL, virtual experiments, model evolution)

3.2 Data Layer: Acquisition and Recognition

Data sources include questionnaires (structured feedback), texts (unstructured descriptions), ontologies (domain knowledge), and tabular/process metrics. Processing applies NLP extraction, ontology mapping, and feature engineering, harmonizing observations into a unified representation. A recognition engine detects: (i) activity patterns, (ii) change points and causal impacts in time series, and (iii) information states (quality, completeness, provenance).

3.3 Modeling Layer: Infact States and Causal Modeling

An "infact" state $\phi = (a, i, t, \theta)$ consists of:

- a: activity component (subject, actor);
- i: information component (content graph, data, its quality, usage, source);
- t: time stamp/interval;
- θ: aggregate properties: stability, complexity (dimensionality), entropy (uncertainty), coupling strength (influence of information on outcomes).

A causal engine maintains explicit cause–effect graphs among infact attributes and outcomes, supports counterfactual analysis and interventions (`do` calculus), and orchestrates a multi-model system: process (BPMN/Petri nets), causal (Bayesian/structural), system dynamics (stock–flow/differential), agent-based, and rules. An ensemble prediction engine delivers scenario forecasts with uncertainty and key-driver identification (Fig. 2).

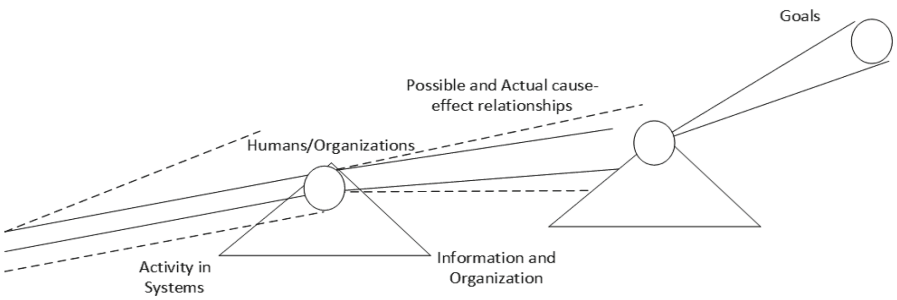


Fig. 2. Alternation of causal chains by human activity. Dashed lines: possible cause–effect sequences without human actions; solid lines: sequences with actions. Humans actualize subsets of possible chains and can compose new sequences.

3.4 Application Layer: Problem Formalization and Solvers

Problems are represented with natural language descriptions grounded in "infact" context, constraints, objectives, stakeholders, temporal horizons, and complexity measures.

Problems are decomposed hierarchically, functionally, and causally into subproblems and solved via optimization, constraint programming, simulation, and logical inference. An abstraction engine lifts detailed solutions to higher LoA as patterns, principles, strategic options, and trade-off analyses.

3.5 Adaptation Layer: Continuous Learning

IAM integrates: (i) reinforcement learning (RL) to optimize information application policies from experience and model-based imagined rollouts; (ii) virtual experiments (digital twins) to run safe experiments and validate models; and (iii) model evolution (selection, tuning, structure learning, validation) triggered by performance and feedback.

3.6 Formalization: States–Actions–Transitions Classes of Pragmatic Characteristics A–E

Activities unfold as sequences of states–actions–transitions (S–A–T). States are presumed to be of "infon" type (i.e., they supposed to have informational substates, which describes future activity and its outcomes). Depending on "infons" various actions performed and various transitions realized. Thus, "infons", actions, transitions are related with cause-and-effect relationships (Fig. 3).

Let a scenario tree $T = (\Omega, \Theta)$ represent the set of branches $Ch_i \in \Omega$ (each Ch_i is a sequence $a...s_i...a...s_j$ with actions a and transitions $\tau(a)$ between states $s_(i, j...)$ associated with each transition $\tau = (s_i, s_J(a)))$, and Θ the corresponding complex, conditional probabilities space (taking into account environment and information conditions) with Ch_i actualization probabilities $p(Ch_i, c_n, \pi)$ in various conditions c_n, under various action plans π and Ch_i results correspondence measures $\mu(Ch_i, c_n, \pi)$ under the c_n, π.

T (pragmatic tree) represents chains and changes of pragmatic situations possible in various circumstances and so, characteristics of T represent various pragmatic properties.

Let us elaborate some details. We define classes of (measurable) pragmatic characteristics on the tree T:

Class A (local indices). Single realization/state indicators, e.g., $\zeta(Ch_i, c_n, \pi) \in [0, 1]$ productivity, or a compliance $\mu(Ch_i, c_n, \pi)$ to requirement.

Class B (state-at-time, actualization probability). $p(s_i, t_i, c_n, \pi) = Poss(s_i(\pi)/(c_n, t = t_i)$: probability (possibility in some cases) of being in a required state s_i at time t given plan π fulfilled under conditions c_n.

Class C (Ch_i chains terminal sets). Multiple terminal states with (possibly distinct) requirements:

$$\mu(Ch_i) = \mu(Y(Ch_i) \sim Y^d(\omega)), \tag{1}$$

where $\mu(Ch_i)$ measures compliance $\sim$ with requirements $Y^d(\omega)$ for each terminal class $\omega = \{Ch_i\} \subseteq \Omega$.

Class D (sets of processes: potential). A measure of potential aggregates compliance across alternative process realizations:

$$\psi(\pi.it) = \sum_{i \in \overline{1,I}, n \in \overline{1,N_c}} p(Ch_i, c_n, it, \pi) \cdot \mu(Ch_i, c_n, \pi, it), \tag{2}$$

where $it-$ vector of characteristics of IT used in various c_n to generate alternative $Ch_i(\pi, c_n, it)$. This is the scalar measure $\psi \in [0, 1]$ for system potential.

This measure can be used to define mathematical optimization problems for system potential $\psi(\pi, it)$

$$\pi^\star = \arg \max_{\pi in \Pi, \pi_a(\pi) \in \Pi_a(\pi))} \psi(\pi, \pi_a \in \Pi^a(\pi)). \tag{3}$$

Class E (dependent processes/policies). With explicit adaptation policy.

$\pi^a(it, \pi)$(e.g., modernization, triggerred by c_n, it), we have similar mathematical optimization problem statement:

$$\pi^\star = \arg \max_{\pi in \Pi, \pi^a(\pi) \in \Pi_a(\pi))} \psi(\pi, \pi_a \in \Pi^a(\pi)). \tag{4}$$

In Fig. 3 $Ch_i(\pi, it, c_n)$ (upper), and Ch_j (lower) – without any activity performed, is shown. As it can be seeing, performing activity using it, depending circumstances c_n, potentially allows to achieve goals with better measure, than without activity and it used.

Compliance Mapping and Measurable Spaces. Crucial to predictive analytics is mapping activity outcomes to measurable spaces where indices, probabilities, and proximities can be computed and optimized. Scenario trees shall be built with probabilities support expectation-based objectives and risk-sensitive constraints (e.g., CVaR on critical outcomes), enabling rigorous optimization.

4 Result Analysis

4.1 Algorithm and Data Structures for Discussed Problems Solving

Given activity π and information technology it descriptions, conditions c_n, requirements $Y^d(c_n)$, data D, models M (process, causal, dynamics, agent), and (optionally) set of plans $\Pi_a(\pi)$:

1. **Recognition.** Extract "infact" states S from D; learn/refresh causal graph $G(S)$ and built appropriate tree T.
2. **Scenario generation.** Build branches $Ch_i(c_n, it, \pi) \in T$ under c_n, it, π with alternations driven by information it and appropriate transition actions $\Pi^a(\pi)$.
3. **Simulation/Prediction.** For each $Ch_i(c_n, it, \pi)$, simulate with model ensemble M to obtain $Y(Ch_i)$.
4. **Probabilities.** Estimate $p(Ch_i, c_n, it)$ via causal/dynamic models or empirical models.
5. **Aggregation.** Compute $w(Ch_i(c_n, it, \pi))$ for each $Ch_i c_n, it, \pi)$ and than, $\psi(it, \pi)$ for each it, pi based on aggregate results.
6. **Optimization (A).** Optimize plan $\pi^\star$ by potential $\psi(it, \pi)$.

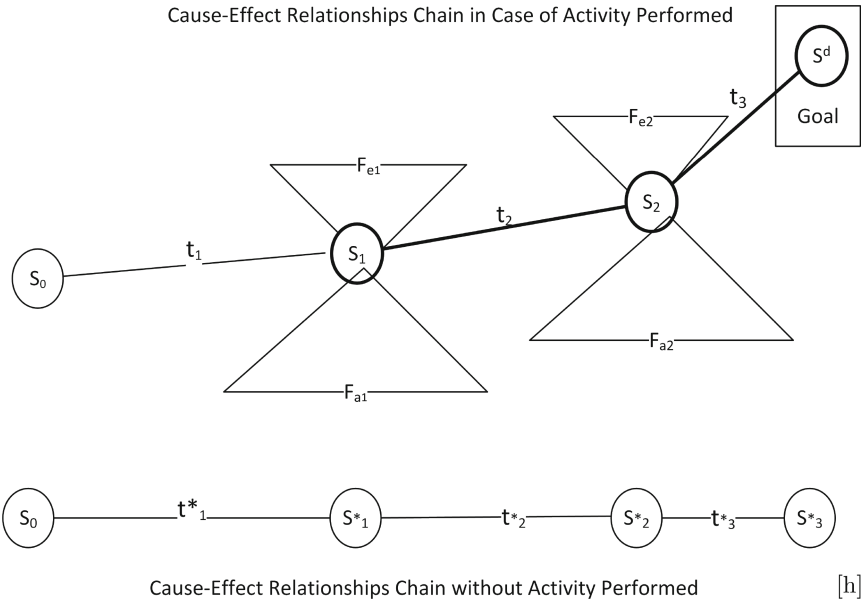


Fig. 3. Sequences of states–actions–transitions. These structures underpin measurable evaluation (efficiency/effectiveness/potential) by enabling aggregation over alternatives

Data Structures. Graph storage (e.g., Neo4j) for S–A–T trees with node types `State/Infact/Requirement` and edges `Causes/Enables/Action/Transition/LeadsTo` (temporal/probabilistic labels). Harmonization pipelines (NLP+ ontology+features)
render "infact" states. Causal libraries manage structure/parameter learning. Solvers (OR-Tools/Gurobi/MiniZinc) allocate resources/actions.

4.2 Case Study Discussion: Adaptive Sprint Planning

We consider a software team planning a two-week sprint under changing requirements:

- **Activity:** sprint planning (product owner and team; object: backlog).
- **Information:** user stories, defect logs (texts), sprint metrics (tables), domain ontology (features/modules), team survey (questionnaire).
- **Objectives:** maximize delivered business value; keep defect leakage low; respect capacity.
- **Constraints:** team capacity, L2–L3 bug SLAs, dependency rules.

Recognition The enImplementation Notesgine detects low information quality for acceptance criteria in payments/refund flows and a change point after an API update. Recent sprints show mid-sprint alternations due to requirement changes.

Infact State and Causality. Current state: stability 0.4, high entropy, coupling 0.7 (information issues strongly affect outcomes). Causal links show that upfront clarification and early exploratory testing reduce rework and defect leakage.

Scenarios. S1: proceed without clarification; S2: invest hours in clarification and exploratory tests. Ensemble predictions suggest S2 has higher risk-adjusted value despite slightly lower nominal throughput (due to reduced rework/defects).
Formalization and solving. The problem is to select backlog items and pre-sprint activities to maximize compliance C under capacity and SLA constraints. Optimization allocates clarification/testing time, shifts low-value stories, and selects appropriate regression suites (Fig. 4). The abstraction engine yields a principle: *under high coupling and low information quality, upfront information improvement improves risk-adjusted value.*

Adaptation. An RL policy learns thresholds on entropy/coupling that trigger clarification investments. A digital twin simulates alternative allocations (e.g., 8 vs 12 h) to calibrate the policy. Causal links are updated from observed effects.

4.3 Implementation Notes

A practical stack includes:

- Kafka for streams, spaCy for NLP,
- ontologies for mapping,
- Pandas/Scikit-learn for features,
- Neo4j for graphs,
- PyTorch/TensorFlow for learning,
- causal libraries for explicit causality,
- AnyLogic for dynamics/agents,
- OR-Tools/Gurobi for optimization,
- RLlib/Stable Baselines for RL, and
- Docker/Kubernetes for scaling.
  Validation spans causal sanity checks (invariance tests, natural experiments), predictive calibration (Brier/NLL), and stakeholder alignment for c_n.

4.4 Discussion and Future Work

IAM reframes pragmatic analytics by unifying activity and information into measurable infact states, preserving explicit causality, and enabling predictive evaluation across alternative processes. It operationalizes classes A–E: from single-state indices to measures of *potential* over scenario trees and adaptation policies. Practically, this yields:

- **Predictive evaluation:** ex-ante assessment of efficiency/effectiveness/potential before implementation.
- **Causal synthesis:** identification of high-leverage informational interventions (e.g., clarification, sensing, predictive modeling).

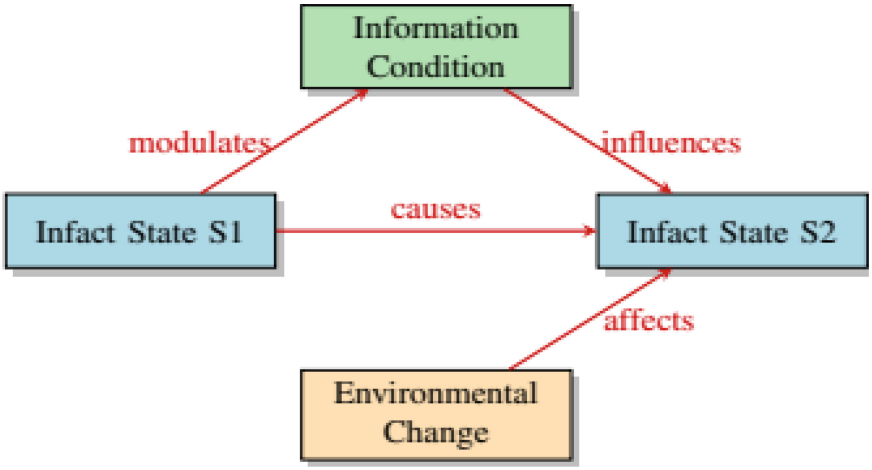


Fig. 4. Causal relationships engine: explicit cause–effect links among "infact" attributes and outcomes; supports counterfactuals and interventions

– **Optimization under change:** formal allocation of resources/actions across scenarios and adaptation triggers.

Future work includes: (i) domain libraries of economics, safety, sustainability, (ii) standardized ontologies for infact components, (iii) scalable digital twins for virtual experimentation, and (iv) rigorous learning-to-decide pipelines with safety guarantees.

5 Conclusion

We presented the Info-Activity Method (IAM) and a formalization of measurable pragmatic properties for activity and systems. By explicitly representing questions, information, and causal alternations within infact states, IAM supports measurable, predictive, and adaptive decision-making across classes A–E, enabling robust solutions to human activity problems under change. This approach bridges philosophical insights about information pragmatics and practical machine-learning/optimization architectures, providing a foundation for predictive evaluation of efficiency, effectiveness, and system potential.

References

1. Floridi, L.: The Philosophy of Information. Oxford University Press, Oxford (2011)
2. Floridi, L.: The method of levels of abstraction. Mind. Mach. **18**(3), 303–329 (2008)
3. Burgin, M.S.: Theory of Information: Fundamentality, Diversity and Unification. World Scientific (2010)
4. Burgin, M.: Information theory: a multifaceted model of information. Entropy **5**(2), 146–160 (2003)

5. Nauta, D.: The Meaning of Information. Mouton (1972)
6. Newman, J.: Some observations on the semantics of information. Inf. Syst. Front. **3**(2), 155–167 (2001)
7. Gurevich, Y., Neeman, I.: Logic of infons. ACM TOCL **12**(2), 1–28 (2011)
8. Loukanova, R.: Situation theory, situated information, and situated agents. In: Nguyen, N., Kowalczyk, R., Fred, A., Joaquim, F. (eds.) Transactions on Computational Collective Intelligence XVII. LNCS, vol. 8790, pp. 145–170. Springer, Berlin, Heidelberg (2014). https://doi.org/10.1007/978-3-662-44994-3_8
9. Illari, P., Russo, F.: The Routledge Handbook of Causality and Causal Methods. Routledge, Abingdon-on-Thames (2024)
10. Newell, A., Simon, H.A.: Human Problem Solving. Prentice-Hall, Hoboken (1972)
11. Geyda, A.S.: Systemic information use methods for activity problem-solving. In: Bansal, J.C., Jamwal, P.K., Hussain, S. (eds.) Sustainable Computing and Intelligent Systems. SCIS 2024. LNNS, vol. 1295, pp. 117–124. Springer, Singapore (2025). https://doi.org/10.1007/978-981-96-3311-1_10
12. Geyda, A.S.: Families of the alternative state—transitions actions networks: procedures and algorithms. In: FIT-M 2020, pp. 244–250 (2020)
13. Adams, M., ter Hofstede, A.H.M., Edmond, D., van der Aalst, W.M.P.: Worklets: a service-oriented implementation of dynamic flexibility in workflows. In: Meersman, R., Tari, Z. (eds.) On the Move to Meaningful Internet Systems 2006: CoopIS, DOA, GADA, and ODBASE. OTM 2006. LNCS, vol. 4275, pp. 291–308. Springer, Berlin, Heidelberg (2006). https://doi.org/10.1007/11914853_18
14. Kronqvist, J., Bernal, D.E., Lundell, A., Grossmann, I.E.: A review and comparison of solvers for convex MINLP. Optim. Eng. **20**(2), 397–455 (2019)
15. IBM: What is Intelligent Automation? https://www.ibm.com/think/topics/intelligent-automation. Accessed 11 Aug 2025

Emergency Technologies and Its Connection in Image Processing and Image Segmentation

Mária Ždímalová(✉)

Department of Mathematics and Descriptive Geometry, Faculty of Civil Engineering, Slovak University of Technology in Bratislava, Radlinského 11, Bratislava, Slovakia
maria.zdimalova@stuba.sk

Abstract. Emerging technologies refer to newexploring technologies that are not yet available. They are technologies with the potential for social, economical or scientific impact. The theory of technological evolution points to the transformative impact of technological progress on society. Subsection of Emergency Techcnologies is AI, Image processing as well as Image Segmentation. Image processing is an important tool for analyzing image data, tracking objects in images, and diagnosing damage to materials and machines. It isimportant for analyzing the equipment in engineering because of maintenance, reliability, stability. This research deals with the review of image processing methods, especially for image data segmentation. The aim of the research was to create a new software based on graph cut. The results of the image segmentation research were segmentation of machine elements, vehicles and technical engineering data. For the purposes of segmentation research, own software based on the graph cutting method was created.

Keywords: Image processing · image segmentation · machine elements · vehicles · diagnostics

1 Introduction and Motivation

Emergency technologies encompass a wide range of innovations designed to enhance disaster preparedness, response, and recovery efforts. These technologies include communication systems, surveillance tools, data analytics, and medical advancements, all aimed at improving situational awareness, streamlining operations, and ultimately saving lives. Examples include industrial pollution, nuclear radiation, toxic waste, dam failures, transport accidents (train, plane, boat, bus, car, bicycle), factory explosions, fires and chemical spills. The term ‘emerging technology’ is related to refer to technologies that are in the early stages of development or represent major innovations within existing technological systems. The concept is dynamic. Its interpretation can vary depending on the context - in areas such as media, business, science, health and education it can take on different nuances of meaning. In general, emerging technologies refer to those technologies that are currently the subject of intensive technical, sociological, pedagogical even human research and development or are expected to become more widely available and

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026
S. Lanka et al. (Eds.): ICTSM 2025, LNNS 1755, pp. 233–247, 2026.
https://doi.org/10.1007/978-3-032-13177-5_19

applicable in the next five to ten years. We can summarize, that emergency technologies involve areas such as artificial intelligence, quantum computing, biotechnology, nanotechnology, augmented and virtual reality and renewable energy technologies. There is given high degree of uncertainty and unpredictability in future of these technologies. The uncertainty of the future of the accompanies in the development of these technologies, the existing and exploring of emerging technologies is particularly important not only from a technical but also from an interdisciplinary perspective. It requires a comprehensive assessment. It is because of their functionality and innovativeness, but also of their societal impact on the development of society. One of the subsection of emergency technologies is Al and also Image Processing. In this contribution we focus on Image processing with focus on Image segmentation. In image processing we deal with digital images. We can describe digital image as a two-dimensional rectangle, where we consider a finite set of digital values. [5, 9, 10]. These values are denoted elements of a figure pixels with intensities, see [5]. In image processing we transmit an image into a digital structure. The image processing system usually treats all images as 2D signals when applying certain predetermined signal processing methods. The image processing process consists of the following stages: image acquisition, image enhancement, image restoration, image coloring, Image processing, image compression and character recognition (Table 1). [11, 13, 20]. Image processing includes image filtering, image enhancement, image coloring, and more. The image is "processed," that is, transformations are applied to the input image and the output image is returned. Transformations can be, for example, "smoothing," "sharpening," "contrast," and "stretching." The transformation used depends on the context and the question to be solved. Image processing can be viewed as an automated system that performs health monitoring and assessment to assess damage caused to concrete and structures due to natural disasters, see [16]. The research deals with image processing, which is the process of converting an image into digital form and performing certain operations to extract useful information from it, [22, 24]. Concretely as a novelty we focus on image segmentation for technical engineering. In Mechanical engineering there is a big gap for software tools which are able to segment 2D/3D images for mechanical engineering. There are two segmentation methods used in [1, 10], but there is still a big gap for segmentation tools using better segmentation technique for these purposes. So we developed our own application providing this segmentation. It is based on implementation of the segmentation technique called Graph cut and its optimization to get better and more sharply results. This part is missing I current research. There are many software applications for segmentation of objects for medical analysis, biological analysis, but for technical engineers is almost no approach.

2 Introduction and Motivation

In this chapter we explain in detail different image processing steps, image processing methods and theoretical approach, see Table 1, more in [6, 11, 21, 23, 25, 26]. Later we study image processing for technical and mechanical engineering. Table 1 presented the concrete steps of image processing and their in depth descriptions.

Table 1. Stages of the image processing process

Steps	Details of the steps
Acquisition process of the image data	This process is considered as the first step in image analysis. It includes gaining the image, changing to digital form, saving and storing of the image, its processing and saving in good digital format.
Enhancement of the image	This process involves improving the quality of the image. There is a need to improve the quality of image results data for better analyzing, interpreting as well as digital presentation.
Restoration of image data	It is a part of the image processing, where we work with damaged, noise and not clear image data. The aim is to decrease noise, to clear the image and bring new corrected image based on some mathematical and informatical function.
Image Processing with colour images	This part needs to understand frequency of colours, how the computer see the colours and RGB system.
Compression of image data	Digital compression of the image reduces data transfer costs, enables their reduced transmission, reduces or expands data.
Optical Character recognition	This part involves the process of importing and converting the scanned images, written text into digital form, like computer-recognition-readable text and image format.

2.1 Image Processing in Engineering: A Mechanical Engineer Working in an IT Industry

In the current interconnected world, mechanical and other types of engineering are changing how mechanical sectors, including automotive construction, and more, are employing image processing technology; This technology is rapidly becoming the preferred solution for quality control and inspection in these sectors. Image processing allows mechanical companies to handle images more quickly and precisely than ever before. This can be applied to tasks such as measuring component sizes or examining the object for any abnormalities or flaws. It can also be used for tasks like monitoring an item's exact route as it moves along a manufacturing line or inspecting packaging to ensure it complies with specifications. The engineering design process consists of a series of stages that engineers utilize to solve problems. These stages encompass problem-solving methods, such as identifying objectives and constraints, creating prototypes, and conducting testing and evaluation. Machine learning, which is a branch of AI, is widely utilized in mechanical engineering. One significant application is predictive maintenance, where machine learning algorithm examine equipment data to forecast possible failures, allowing for proactive maintenance approaches. Image Processing in Mechanical Engineering. There are often very loud calls from real praxis to help with

solutions of engineering difficult problems. Mechanical engineering belongs to this part as well. Image processing give the industry a lot of possibilities for solving real praxis problems, specially with image analyses. Image processing is a filed, which can provide results and solutions precisely and fast. Mechanical engineering belongs to one of the oldest branches in engineering. It includes concepts as mechanics, design, robotics, analysis of fluid mechanics, different types of kinematics and many others technical and mechanical problems. These kinds of processes and methods are used for suggestions and construction of new robot and mechanical vehicles, aircraft technologies and new technical machines. We can use image processing for various tasks: detecting of the precise measure of technical components, controlling of technical parameters, checking technical abnormalities. Mechanical engineers suggest tests and keep different types of machines. The most common areas in mechanical engineering are: quality control and inspection, computer-aided design (CAD), robotics and automation, finite element analysis (FEA), reverse engineering, thermal imaging and motion analysis. Integrating image processing in mechanical engineering can enhance efficiency, improve accuracy, and lead to innovative solutions across various applications [2, 14, 15, 17]. Image processing is an important engineering tool which can be used in the mechanical domain as well. These applications demonstrate how image processing in mechanical engineering enhances efficiency, accuracy, and innovation in mechanical engineering. In mechatronics, vision-based AI uses image processing in a wide variety of problems. As well as stress and strains, controlling the dynamical flow of systems and technical, conditional monitoring of technical processes, diagnostic of damage in the mechanical and automobile industries can be evaluated using bifurcation methods and images may be processed using some image processing tool, see [6, 11, 24].

2.2 Engineering Graphics in Mechanical Engineering

Drawing is the language of mechanical engineers. There are departments related to product engineering, process engineering, manufacturing dept., etc. (related to this answer). Whenever the Sales and Marketing department does the survey of the market, they have this particular set of features the product should have. Product engineering department or Design department starts working on the design of the product. When design dept. is done with designing of the product, they release the documents of the product, which are nothing but drawings of the product and some data like its features, material, if any special care is to be taken for a particular part, part name, any specific manufacturing process is to be carried out, etc. But most importantly, the drawings and dimensions, fits, tolerances, etc. These drawings are discussed with all other departments from their perspectives, so if any problem is there with any department regarding anything, they discuss and the design gets refined over time. The final drawings are then sent to the process engineering department (they are supposed to be the most important link in the industry). So their job is to decide different processes, operations to be carried out in order to manufacture all the parts in the most cost-effective, energy-efficient way. The process dept. also releases some documents which contain Route Sheet, Process Picture, Tooling Specifications, etc. These documents and especially the process pictures are then sent to a machine shop for manufacturing. Then the mechanical industry deals

and talks through drawings, and all these drawings would not be possible without knowing engineering graphics. Image processing is an important engineering tool which can be used in the mechanical domain as well. These applications demonstrate how image processing in mechanical engineering enhances efficiency, accuracy, and innovation in mechanical engineering. In mechatronics, vision-based AI uses image processing in a wide variety of problems. as well as stress and strains, controlling the dynamical flow of systems and technical, conditional monitoring of technical processes, diagnostic of damage in the mechanical and automobile industries can be evaluated using bifurcation methods, and images may be processed using some image processing tool.

2.3 Filtering Using Convolution

Convolution is a mathematical operation used in image processing to combine two functions, typically an image and a filter (kernel), to produce a new output image. This operation helps in modifying or extracting important features from the original image. In image processing:

- **Input**: The original image, represented as a matrix of pixel values
- **Kernel (or Filter)**: A small matrix that defines the type of transformation (e.g., blurring, sharpening, edge detection)
- **Output**: A new image resulting from the convolution, where each pixel is influenced by its neighboring pixels based on the kernel

The image $I(x, y)$ is convolved with a kernel or filter $K(i, j)$:

$$\mathrm{O(x, y)} = \sum_{\mathrm{i=-m}}^{\mathrm{m}} \sum_{\mathrm{j=-n}}^{\mathrm{n}} \mathrm{K(i, j)I(x - i, y - j)}$$

Where:

- $I(x, y)$: pixel value at position (x, y)
- $K(i, j)$: kernel value at position (i, j)
- $O(x, y)$: output pixel value after filtering
- The kernel is typically of size $(2\,\mathrm{m} + 1) \times (2\mathrm{n} + 1)$

Numerical Problem: Here a simple 3×3 grayscale image:

Image I:

$$\begin{bmatrix} 3 & 0 & 1 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$$

And a 3×3 kernel:

Kernel K (Blur):

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

Last line→ normalized: multiply result by 1/9

To compute O (1,1) (center pixel in output), place the kernel cantered on I (1,1), multiply corresponding entries, and sum them:

$$\mathrm{O}\ (1,\ 1) = \frac{1}{9} \cdot (3 + 0 + 1 + 4 + 5 + 6 + 7 + 8 + 9) = 19 \cdot 43 \approx 4.78$$

This value replaces the original center pixel.

3 Materials and Methods

3.1 Image Segmentation Techniques

Using common properties including texture, gray level, brightness, contrast, and color, image segmentation separates images into different parts. An essential, crucial, and intricate part of image processing that affects the final output's quality is segmentation. To finish later tasks quickly and efficiently, breast photo segmentation must be done precisely and flawlessly. These segmentation techniques can be categorized as threshold-based, region-based, edge-based, unsupervised clustering-based, and deep learning-based. In research there are different segmentation procedures in their investigations. But not all techniques are effective, and there isn't a comprehensive segmentation strategy that works for all kinds of images. Depending on the classification strategy employed, segmentation techniques can also be divided into fine-grained classifications:

1. Methods can be automatic, semi-automatic or manual.
2. Strategies based on region and pixels.
3. Segmentation methods include model-based (contour following, dynamic programming, feature map, multispectral), manual delineation, and low-level (region growth, thresholding).
4. Methods which include traditional (region-based, edge-based, and thresholding), neural network, fuzzy, and statistical techniques.

These segmentation techniques can be categorized as threshold-based, region-based, edge-based, unsupervised clustering-based, and deep learning-based.

3.2 Description of Selected Methods

Instance Segmentations Instance segmentations is a deep learning-driven computer vision task that predicts the exact pixel-wise boundaries of each individual object instance in an image.

The image detects objects and their boundaries. Nevertheless, even within the same category, every new item is identified as a distinct instance. If annotation is the technique ML engineers use to outline their objects for labeling, then segmentation is the strategy your model uses to separate and distinguish individual objects in a dataset.

Semantic Segmentation In a semantic segmentation task, fully annotated images are represented by segmentation masks. Regardless of whether they belong to the same instance or not, it implies that every pixel in the image should fall into a certain category. In this instance, though, every pixel belonging to the same category is shown as a single segment.

Instance Segmentation In the process of segmentation, it evaluates and considers the diversity of things. That is where the significant distinction between instance and semantic segmentation lies.

Panoptic Segmentation Panoptic segmentation is a combination of instance segmentation and semantic segmentation. In panoptic segmentation, the entire image should be labeled, and pixels from different instances should have different values even though they belong to the same category. The solutions of instance segmentation and semantic segmentation are combined in a difficult computer vision job known as panoptic segmentation.

Image Segmentation Vs. Object Detection Other image annotation techniques including object identification, location, and classification are frequently combined with picture segmentation.

Image Classification A single class is given to an image, usually to the primary object depicted in the picture. Images that feature an object, let us tell and imagine the screw, will be categorized as "screw". Comparing picture classification to object localization, detection, or segmentation, we are unable to determine the precise position of that cat in the image or its visual boundaries.

Object Detection A bounding box is used to identify the objects in an image or video, and they are then labeled. The result is the main distinction between image segmentation and object detection.

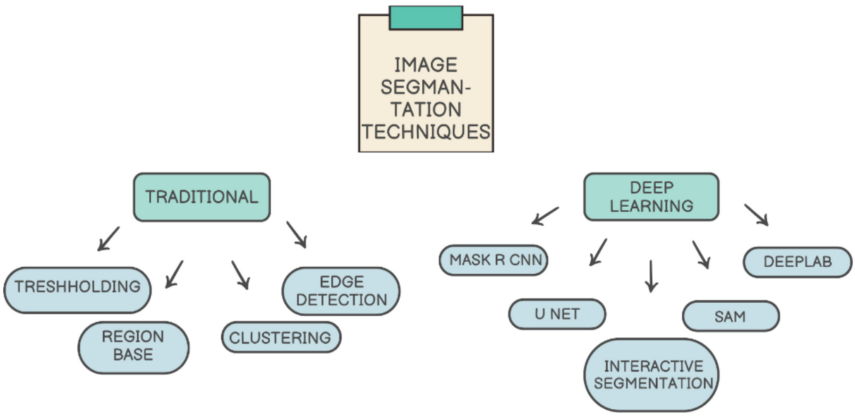


Fig. 1. Different image segmentation techniques

3.3 Image Segmentation Techniques

There are a lot of different techniques, methods and algorithms used in image processing. These techniques and methods cover different areas of mathematics, like discrete mathematics, continues mathematics, fuzzy and aggregations techniques, statistical techniques

(clustering methods, fuzzy clustering methods, Gaussian mixture models, Bayesian statistical approach), [2–4]. We need to mention as well as machine learning techniques as well as neural networks. From an informatics field we know many algorithms used in segmentation process, like clustering, methods focused on edge detectors (edge and cunny edge detection techniques), threshold and its various modifications, region growing algorithms (based on partial and differential derivation), fuzzy methods, combination techniques, see [7, 20]. We show an overview of algorithms on Figs. 1 and 2.

Region-based segmentation - This first technique searches for pixel-by-pixel similarities between segments that are adjacent to one another.

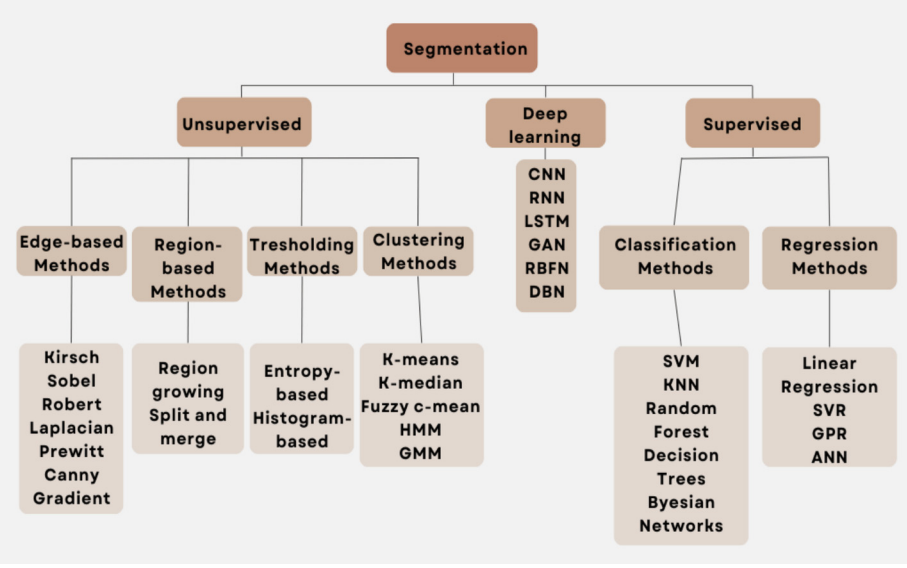


Fig. 2. Different classification of image segmnetation techniques and methods

Edge detection segmentation - used to overcome the drawbacks of region-based methods, edge-based segmentation algorithms prioritize the edges of objects to produce reliable outcomes. This is accomplished by first identifying and categorizing specific pixels as "edge pixels."

Thresholding is possibly the simplest method for segmenting images. In essence, it examines the process of converting the original image to black and white to produce a binary map or image.

Clustering techniques -Other methods for image segmentation for picture annotation include clustering algorithms. In this instance, the clustering algorithm may use pixel values as data points. K-means is one example of a clustering algorithm.

Graph cut segmentation - Image smoothing, the stereo correspondence problem, image segmentation, object co-segmentation, and many other computer vision issues that can

be expressed in terms of energy minimization can all be effectively resolved thanks to graph cut optimization when used in the field of computer vision.

Image Segmentation With Deep Learning: Just as deep learning improves the speed and accuracy of high-performing systems, see [7, 12, 20]. In the image segmentation architecture, the encoder and decoder are essential components. This involves using filters in the pooling layers to extract certain portions of the image and then receiving a final output with a segmentation mask: This leads to the following classification of masks and neural networks: **U-net:** U-Net is a model, which appears in a 'U' shape when its architecture is displayed. It consists of two segments: the upsampling and downsampling, known as the expanding part and contracting path correspondingly, see [18, 19]. **Mask R-CNN:** This is a type of mask which is actually two steps convolution neural network. **Interactive segmentation**: It is again one of the model of deep learning technique. This is not automatic segmentation, there is a need of user interaction.

Segmentation Research on Selected Machine Elements and Vehicles

The research was conducted at the Department of Mathematics and Descriptive Geometry, Faculty of Civil Engineering, Slovak University of Technology in Bratislava as part of the project The Slovak National Research Grant Vega 1/0036/23. The aim of the research was to study and create own segmentation software application for segmentation of objects in mechanical engineering. The fact, that some open softwares available on internet exist is not enough, because our software brings better segmentation with better sharply edge boundary. There is also the need to mention another observation: for concrete segmentation techniques in mechanical engineering, there are no such open softwares available on the internet. That is a reason why we got calls from real praxis to create our own software application for getting good segmentation in mechanical engineering. We explain how mechanical equipment, cars and vehicles can be segmented for the purpose of diagnostics and control in mechanical engineering. As the original data, "Wikimedia Commons" was used with a free and open license. The segmented objects (machine elements and vehicles) were obtained using our own software application, which was developed based on the graph cut algorithm for object segmentation. The program is our own developed and implemented program based on Graph Cut algorithm. The main idea is that, we take images, pixels and intensities which we will transform into graph based structure/network. In this network we will find maximal/flow, minimal cut, which corresponds into segmentation in the image. This algorithm and its optimization we implemented in the program language C and C ++ and created own software implementation for segmentation of images. This application worked as follows. The user uploads an image based on some input parameters. The user then selects an object and background seed pixel. The user will finally segment the object and the background.

Explaining Graph cut algorithm in details: What is the Graph Cut Method? The graph cut method is a mathematical and computational technique used to partition a graph into meaningful segments. It is widely applied in image segmentation, clustering, computer vision, and network analysis. Core Idea: Represent the image or data as a graph or network: Let us consider nodes as pixels (2D) or voxels (3D). Edges represent connections between neighboring pixels/voxels, with weights representing similarity or boundary strength. The minimal cut of the graph separates into two or more disjoint

sets by removing edges with the least cost, while keeping the partition meaningful (like minimizing boundary cost and maximizing internal similarity). Graph Cut in Image Segmentation: The image is modeled as a graph with two special nodes: we add source (S and the sink (T)):Source (S) – represents the foreground (object). Sink (T) – represents the background. Edges connect pixels to their neighbors and to S or T, see Fig. 3. Then the cut is the set of edges that, when removed, separates the graph into two disjoint sets: object and background. We cant present that Minimal Cut in network (graph) is equivalent into the maximal flow in the corresponding network, These are duals mathematical problems. Maximal flow is possible to find through Ford-Fulkerson algorithm, see [8, 11, 21]. And this all is equivalent into separation of vertices into two independent and disjoint sets, which correspond to the object and background. Description of the picture and visual description of a basic graph cut sets: Firsts we present before cutting the vertices: a) Pixels represented as small circles (nodes), b) we add two additional vertices Source (S) on the left, Sink (T) on the right, we present edges between pixels (neighborhood links). We define weights (thicker lines = stronger connection) on the edges or oriented edges. After Cut: A line (cut) separates the pixels into foreground and background, removing the weakest connections.

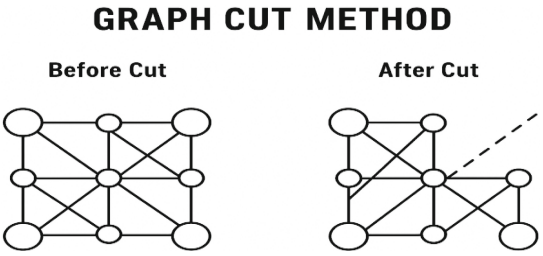


Fig. 3. Visualisation of the graph cut method

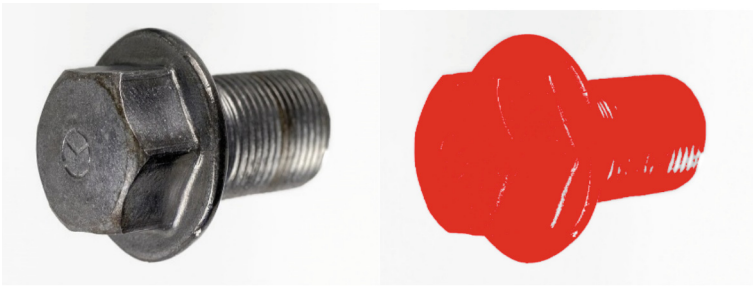
In this chapter we will show results from our own segmentation application. We uploaded the segmented data and got our new results. The application was implemented and optimized programming language C desktop application, see Fig. 4.

These results are of good quality segmentation. There is a gap in image processing and image segmentation application for mechanical engineering data. The basic approach is the thresholding method, but it does not provide good segmentation results. The research shows the segmented objects as newly obtained objects/results, Figs. 5, 6, 7, 8 and 9. The results illustrate the concrete implementation of the segmentation algorithm for mechanical engineering purposes. We used our own new software application for uploading source images and to get segmented images. The original data comes from Wikimedia Commons and are under a free license. The following were analyzed in the research:

- machine elements: a screw and a self-tapping screw (Figs. 5a), b) and 6a), b)),
- vehicles: a locomotive, a bus, and a crashed car (Figs. 7a), b), 8a), b) and 9a), b))

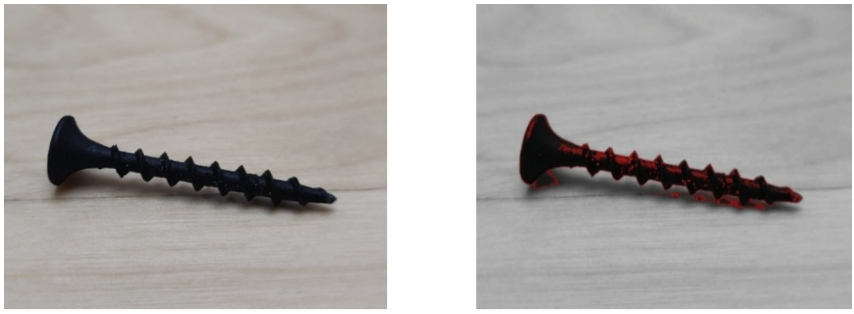


Fig. 4. Running window of our created desktop application



(a) **(b)**

Fig. 5. (**a**) Bolt: Original data. (**b**) Bolt: Segmented data



(a) **(b)**

Fig. 6. (**a**) Self tapping screw: original image. (**b**) Self tapping screw: segmented object/the screw

Comparison with other techniques: On the Fig. 9(b) we can see comparison of results given by Graph cut method obtained by our own software application with Fig. 9(c). On the Fig. 9(c) was used a basic threshold method. We can see even visually, that our



(a) **(b)**

Fig. 7. (**a**) Locomotive: Original data. (**b**) Locomotive: Segmented data



(a) **(b)**

Fig. 8. (**a**) Bus: Original data. (**b**) Bus: Segmented data

method is providing better segmentation results, keeping sharp boundaries and keeping details on the segmented image.

4 Results and Discussion

This contribution we bring summarizes and overviews the image processing methods in general and also focuses on engineering applications. We also bring a survey of image segmentation techniques with applications. Finally we present our own results in mechanical engineering-segmentation of image with mechanical engineering topics. We provided segmentation by own created programs based on graph cut method. We created our segmentation model, software application with implemented graph cut algorithm. We get the results / segmented objects with good segmentation and sharp edge boundary. We fulfilled the gap in image segmentation in mechanical engineering. We get segmentation of the screw, of the bus, of the crashed car. These results can be used for other analysis

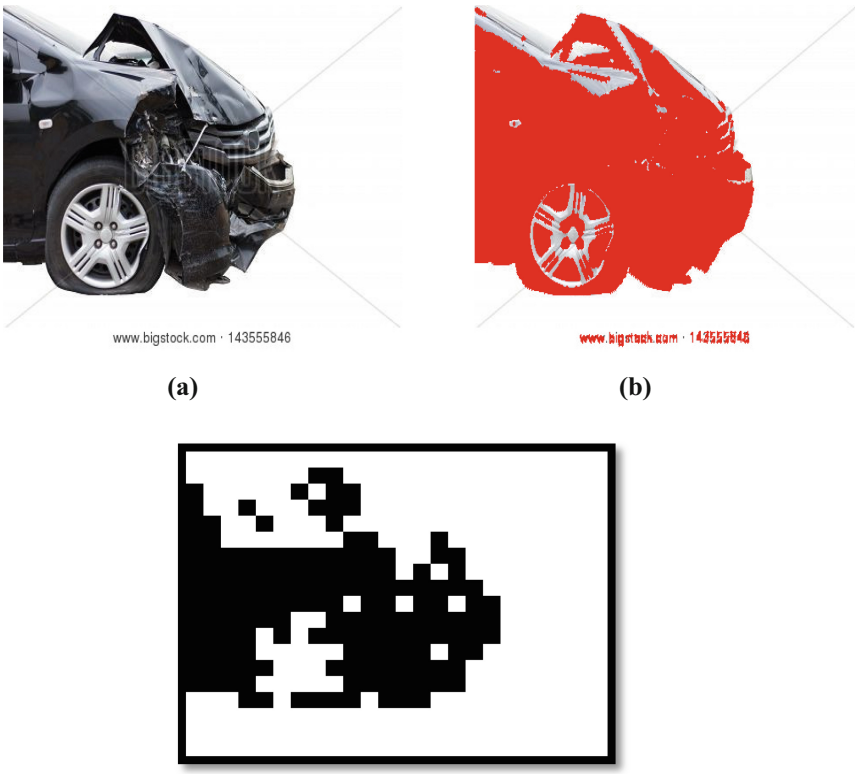


Fig. 9. (**a**) Crashed car: Original data. (**b**) Crashed car: Segmented data. (**c**) Crashed car: Segmented data by basic threshold segmentation model

of cracks, of crashed parts of vehicles of the segmented object. Mechanical engineer can use this kind of analysis for other technical analysis, e.g. how big damage is done in crashed cars. In the Figs. 3, 4, 5, 6 and 7 we did analysis and segmentation of the technical objects based on our software application. The main aim was to get the good segmentation of the technical objects. On segmented results, mechanical engineers can provide a following analysis as where to place the bus, how big the damage in vehicles is, and so on. We also showed comparison with other techniques, see Fig. 7(c), where we segmented the original data by the threshold technique. The conclusion is, that the graph cut technique is giving precisely better results, in segmentation outputs.

5 Conclusion

The presented research presents in detail the methods for image processing in general, and in particular those used in engineering, that is, in mechanical engineering. An overview of image segmentation techniques with applications is also presented. The

main contribution of this research is the results of research on image segmentation displayed on various machine elements and vehicles. For the purposes of segmentation, our own proprietary software based on the graph cutting method was used. The main results is in segmenting machine engineering objects with new software application. This computer application was based on implementing and optimizing graph cut method. Graph cut method is providing very good segmentation results in medicine, biology, SARdata as we also have shown even in mechanical engineering objects and applications. This way we fulfilled a big gap, which was a missing segmentation tool for mechanical engineering data, as well as other engineering regions, like electronical and new engineering areas.

Acknowledgements. The Slovak National Research Grant Vega 1/0036/23 of the Grants Agency of the Ministry of Education and the Slovak Academy of Sciences is funding the research Mária Ždímalová.

References

1. Anyu, M., Zhenyu, L., Guifang, D., Jianrong, T.: Part-to-surface mesh segmentation for mechanical models based on multi-stage clustering. Computer-Aided Des. **162**, 103545 (2023)
2. Boykov, Y., Jolly, M.P.: Interactive organ segmentation using graph cuts. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2000: Third International Conference, Pittsburgh, PA, USA, October 11–s14, 2000. Proceedings 3, pp. 276–286. Springer (2000)
3. Boykov, Y., Kolmogorov, V.: An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. IEEE Trans. Pattern Anal. Mach. Intell. **26**(9), 1124–1137 (2004)
4. Boykov, Y., Jolly, M.-P.: Interactive graph cuts for optimal boundary & region segmentation of objects in ND images. In: Proceedings of the Eighth IEEE International Conference on Computer Vision. ICCV 2001, vol. 1, pp. 105–112. IEEE (2001)
5. Catbas, F.N.: A review of computer vision/based structural health monitoring at local and global levels. Struct. Health Monit. **20**, 692–743 (2021)
6. Dreyer, K.J., Kalra, M.K.: Digital imaging fundamentals. In: Dreyer, K.J., Thrall, J.H., Hirschorn, D.S., Mehta, A. (eds.) PACS. Springer, New York (2006)
7. Figueiredo, E., Park, G., Farrar, C.R., Worden, K., Figueiras, J.: Machine learning algorithms for damage detection under operational and environmental variability. Struct. Health Monit. **10**, 559–572 (2011)
8. Ford, L.R., Fulkerson, D.R.: Maximal flow through a network. Can. J. Math. **8**, 399–404 (1956)
9. Hamuda, E., Glavin, M., Jones, E.: A survey of image processing techniques for plant extraction and segmentation in the field. Comp. Electron. Agric. (125), 184–199 (2016). ISSN 0168-1699
10. Chengjun, C., Chunlin, Z., Jinlei, W., Dongnian, L., Yang, L., Jun, H.: Semantic segmentation of mechanical assembly using selective kernel convolution UNet with fully connected conditional random field. Measurement. **209**, 112499 (2023)
11. German, S., Brilakis, I., DesRoches, R.: Rapid entropy-based detection and properties measurement of concrete spalling with machine vision for post-earthquake safety assessments. Adv. Eng. Inform. **26**, 846–858 (2012)

12. Cha, Y.J., Choi, W., Suh, G., Mahmoudkhani, S., Büyüköztürk, O.: Autonomous structural visual inspection using region-based deep learning for detecting multiple damage types. Comput. Aided Civ. Inf. Eng. **33**, 731–747 (2018)
13. Chen, C., Zhang, B., Su, H., Li, W., Wang, L.: Land-use scene classification using multi-scale completed local binary patterns. Signal Image Video Process. **10**, 745–752 (2016)
14. Kantsepolsky, B., Aviv, I.: Sensors in civil engineering: from existing gaps to quantum opportunities. Smart Cities. **7**, 277–301 (2024)
15. Liu, Y., Cho, S., Spencer Jr., B.F., Fan, J.: Automated assessment of cracks on concrete surfaces using adaptive digital image processing. Smart Struct. Syst. **14**, 719–741 (2014)
16. Luo, Q., Ge, B., Tian, Q.: A fast adaptive crack detection algorithm based on a double-edge extraction operator of FSM. Constr. Build. Mater. **204**, 244–254 (2019)
17. Morgenthal, G., Hallermann, N., Kersten, J., Taraben, J., Debus, P., Helmrich, M., Rodehorst, V.: Framework for automated UAS-based structural condition assessment of bridges. Autom. Constr. **97**, 77–95 (2019)
18. Prabaharan, T., Periasamy, P., Mugendiran, V.: Studies on application of image processing in various fields: an overview. IOP, Conf. Series. Mater. Sci. Eng. **961** (2020)
19. Rafiei, M.H., Adeli, H.: A novel unsupervised deep learning model for global and local health condition assessment of structures. Eng. Struct. **156**, 598–607 (2018)
20. Saha, A.K., Kundu, S.: Deep learning for medical image segmentation. J. Med. Syst. **42**(8) (2018)
21. Sakshi, V.K.: Image segmentation techniques: statistical, comprehensive, semi-automated analysis and an application perspective analysis of mathematical expressions. Arch. Comput. Meth. Eng. **30**, 457–495 (2023)
22. Spencer Jr., B.F., Hoskere, V., Narazaki, Y.: Advances in computer vision-based civil infrastructure inspection and monitoring. Engineering. **5**, 199–222 (2019)
23. Vicente, S., Kolmogorov, V., Rother, C.: Graph cut based image segmentation with connectivity priors. 2008 IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–8. IEEE (2008)
24. Xiao, Y., Wu, J., Yuan, J.: mCENTRIST: A multi-channel feature generation mechanism for scene categorization. IEEE Trans. Image Process. **23**, 823–836 (2013)
25. Yang, M.-S.: A survey of fuzzy clustering. Math. Comput. Model. **18**(11), 1–16 (1993)
26. Zakeri, H., Nejad, F.M., Fahimifar, A.: Image based techniques for crack detection, classification and quantification in asphalt pavement: a review. Arch. Comput. Methods Eng. **24**, 935–977 (2017)

Secure and Sustainable AI-Driven Virtual Healthcare Assistants: Privacy-Preserving NLP for Future-Ready Patient Care

Sahar Ebadinezhad[1,2(✉)], Olusegun Emmanuel Adeshina[1], and Mohammed Al-Hussein[1]

[1] Department of Computer Information Systems, Near East University, Nicosia 99138, Cyprus
sahar.ebadinezhad@neu.edu.tr, {20242838,20244715}@std.neu.edu.tr
[2] Computer Information Systems Research and Technology Center (CISRTC), Near East University, Nicosia, Turkey

Abstract. The rapid introduction of Natural Language Processing (NLP)-powered virtual assistants into healthcare has moved faster than the development of supporting governance and security controls, with extensive regions remaining unexplored in privacy, security, and ethics. Although promising, comprehensive consideration of threats to data confidentiality, integrity, and availability and of ethical issues around AI autonomy within clinical settings remains nascent. This literature systematic review therefore attempted to synthesize academic knowledge to date by establishing the primary security issues in NLP- harbored health systems, examining cryptographic solutions to their protection, outlining a regulatory conformity framework, and evaluating the sufficiency of state-of-the-art privacy-enhancing technologies. Using the PRISMA-2020 guidelines, a systematic search of the Scopus and ScienceDirect databases for articles from 2016 to 2025 yielded 4,780 records. Out of a stringent filtering procedure, 16 relevant research papers were shortlisted for further examination. The results show a bountiful threat environment comprising user-centric challenges like "AI hesitancy," AI-scarred vulnerabilities like data poisoning, and significant hindrances in compliance with regulations. The findings expose the extreme trade-off between privacy robustness and artificial intelligence model performance, but they also underline the need for multi-layer protection applying cryptographic techniques such as homomorphic encryption and federated learning. The study also underlines the need to replace traditional anonymizing methods with safer ones since they are insufficient to preclude re-identification efforts and increasing trust, openness, and regulatory compliance, this evaluation underlines the importance of applying a comprehensive, socio-technical strategy to guarantee the safety of NLP-based healthcare assistants and so support their moral and safe deployment.

Keywords: AI-Driven Healthcare · Future-Ready Care Systems · Privacy-Preserving NLP · Secure Patient Data · Sustainable Health Technology · Virtual Health Assistants

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026
S. Lanka et al. (Eds.): ICTSM 2025, LNNS 1755, pp. 248–263, 2026.
https://doi.org/10.1007/978-3-032-13177-5_20

1 Introduction

Particularly in artificial intelligence (AI) and natural language processing (NLP), the expansion of technology has resulted in the development of virtual healthcare assistants, who are rather helpful in the significant transformation of digitalizing healthcare services [1, 2]. These systems are capable of interfacing with patients via text or voice interaction, providing medical advice, managing appointment calendars, monitoring symptoms, and supporting chronic illness management [3]. With the continuing incorporation of such systems into clinical workflows and their increasing reliance on sensitive health information, the privacy and security of patients' information become critical issues [4]. Aiming to solve the risks related with the deployment of NLP-based virtual assistants in healthcare environments, this study, AI-Driven Virtual Healthcare Assistants: Security and Privacy Challenges in NLP-Based Patient Care, Analyzing the hazards to data confidentiality, integrity, and access; evaluating the ethical AI-driven activities; and developing technical and policy-level mitigating methods are among core aims [5, 6]. The contribution of this study is from the perspective of the impact of AI on healthcare—while improving operational efficiency and access, these technologies impose significant challenges in terms of privacy violation, bias, and lack of transparency in decision-making [7, 8]. According to this principle, not only do patients require protection, but also the applications of AI are made lawful through structures such as HIPAA and GDPR [4]. Although this study primarily viewed existing models of AI-driven virtual assistants conceptually, it sought to understand key threat vectors, analyze the ethical impacts of AI autonomy in clinical interactions, and develop a socio-technical framework towards better trust, transparency, and accountability within virtual healthcare systems [6]. In addition, this study does not include actual tests or other applications of AI, such as diagnostics or robotics, although it has pointed out the urgent need for adaptive and safe patient-centered solutions as AI is becoming embedded in modern healthcare. Additionally, it reveals an increased dependency of healthcare providers on systems for automating triage of patients, increasing the scope for the effect of an AI-driven error or misuse of data [3]. Because of ongoing improvement in NLP to replicate human talking, ethical and clinical issues may arise and typically be associated with the precision of AI with respect to possible injuries and the accountability of AI decisions [8]. Responsible and human-centered AI deployment in healthcare is now as urgent as possible because the rapid advancement of AI adoption has gone ahead of cybersecurity protocol and governance model development [5, 7]. The recent literature caters to placing a litany of legal, ethical, and technical challenges on exasperated AI-based virtual assistants. Van Mill and Quintais (2022) emphasize the arduous nature of assigning dual data controller roles under the GDPR [9], whereas Piñeiro-Martín et al. (2023) consider several ethical concerns related to LLM integration, like bias and privacy [10]. Though tools furthering user transparency like PrivExtractor have arisen, they are still incomplete. According to Jeong (2020), AI-related security threats are emerging, implying a need for forensic methods to adapt [11]. Roca et al. (2020) propose technically plausible scaling, privacy-aware architectures for chronic care [12], while Javaid et al. (2023) mention that ChatGPT's potential in healthcare can bring about both innovations and unresolved ethical issues [13].

Our study contributes with some of the major neglected security and privacy-related aspects concerning NLP-based virtual healthcare assistants. It then earmarks prominent risks to the confidentiality, integrity, and accessibility of data, also touching on ethical questions with respect to AI autonomy within clinical settings. In introducing a socio-technical framework with the aim of boosting trust, transparency, and regulatory coherence in virtual healthcare systems, especially as AI adoption continues to outpace existing governance structures.

RQ1) What are the security challenges in NLP-based healthcare systems?
RQ2) Analysis of different cryptographic solutions for securing virtual healthcare assistants?
RQ3) How can a vulnerability analysis framework be designed to ensure compliance with industry-specific security and privacy regulations such as HIPAA, GDPR, or ISO/IEC 27001?
RQ4) How can federated learning and differential privacy be applied to NLP-based virtual healthcare assistants to enhance data privacy without compromising performance?
RQ5) How effective are current encryption and anonymization techniques in protecting real-time patient data exchanged through AI-powered virtual assistants during clinical triage and consultations?

2 Methodology

This paper provides a complete overview of security challenges in NLP-driven health systems, including identifying main weaknesses, assessing cryptographic and privacy-preserving approaches, providing a regulatory conformance framework, and assessing privacy-enhancing technology. This approach provides theoretical advances as well as practical assistance to scholars, system programmers, and policymakers.

2.1 Research Methodology & Strategy

The research approach of this work is a systematic literature review (SLR). This review presents a clear, orderly approach for spotting, assessing, and compiling pertinent scholarly work. It uses a set process, so it lowers researcher bias and raises the dependability of the results. Following the PRISMA 2020 guidelines, which offer a consistent framework for recording the review process, including literature searches, screening, and study selection, the process is reviewed. In scientific research especially, a systematic literature review is quite important since it helps to unite current knowledge, point up areas of research lacking evidence-based conclusions, and support of evidence-based conclusions. It is well-known for its methodological rigidity and contributions to the progress of research in several spheres.

The research was conducted across two databases, Scopus and ScienceDirect, as shown in Fig. 1. The search was applied only to English-language publications published between January 2016 and April 2025. ("Artificial Intelligence" OR "AI") AND ("Virtual Healthcare Assistants" OR "Health Chatbots" OR "Conversational Agents") AND ("Security" OR "Privacy" OR "Data Protection") were the search key terms and search operators utilized.

2.2 Selection Process

In this research, inclusion and exclusion criteria have been applied as part of the selection process to identify articles that were relevant to the research's objectives. The inclusion criteria showed that only documents published within the last nine years, from January 2016 till April 2025, were to be utilized in the study. The inclusion criteria are set to be restricted to only articles that were published in English, open access type, and computer science for the subject area, and records that did not meet these criteria were excluded from the review, as can be seen in Table 1.

Table 1. Selection Process

Criteria for Inclusion	Criteria for Exclusion
Articles selected between 2016 to 2025	Any research before the 2020
Articles selected in English language only	Books, book reviews, editorial materials, and conference proceedings that were written in languages other than English.
Open access sources selected only	Subject area not limit to Computer science
Subject area focused on computer science	Articles not open access
Document type (Articles)	

In the two databases (Scopus and ScienceDirect), 4,780 records were identified. After the screening process, 34 records were further reviewed. Of these, 18 have been excluded at the eligibility stage, as they did not align with the study aim and did not address systematic literature review questions. Finally, 16 studies are selected for comprehensive analysis in this review that indicates in Fig. 1.

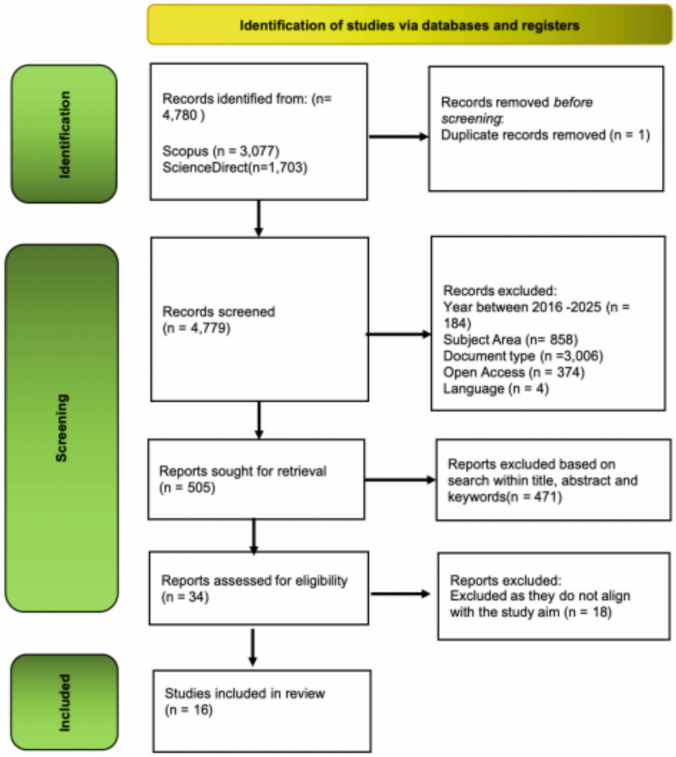


Fig. 1. PRISMA framework of this review.

3 Results and Analysis

In this section we will address the research questions and discuss the different perspectives.

3.1 RQ1) Security Challenges in NLP-Based Healthcare Systems

The integration of Natural Language Processing (NLP) into healthcare systems introduces a range of security challenges, stemming from user perceptions, data handling practices, inherent AI vulnerabilities, ethical dilemmas, and the broader cyber threat landscape. These challenges are multifaceted and require careful consideration to ensure the safe and trustworthy deployment of NLP technologies.

User Perceptions and Acceptability. AI-powered chatbots are prompting "AI hesitancy" between users due to worries regarding cybersecurity, data leakage, and information veracity. Users are concerned about both the safety of their data as well as the accuracy of the data given [1]. Privacy issues are a big barrier to the widespread use of healthcare chatbots, as they collect sensitive patient data. While customization has little impact on issues of privacy, it strongly argues for "privacy by design" and transparent data gathering procedures.

Users are apprehensive to share sensitive health information with chatbots, particularly for medical diagnostics and mental health concerns [4]. Concerns about privacy comprise identity theft and data exchange with third parties. AI-based healthcare chatbots for self-diagnosis (AICSD) confront security difficulties such as misuse of information, patient safety concerns, and trust manipulation [3]. Mental health chatbots [5] have lower system security and integrity than regular online interfaces. Building and retaining user trust is critical, with safety, dependability, and transparency surrounding data sources being major concerns.

Data Management, Privacy, and Control. Existing chatbots may have trouble managing user control over private health information, managing data privacy, and adhering to privacy regulations such as GDPR [7]. The data collection and storage by LLM-driven virtual assistants (VAs) are the prime concerns regarding data privacy and security. There's an ongoing debate about whether AI systems used in public services should have comprehensive data governance [10]. Modeling changing user behavior and medical oversight are important concerns. When identifiable patient data is entered into LLM-based systems, for example, ChatGPT, then it might become part of the database the chatbot utilizes, which creates major privacy concerns and maybe undermines the confidentiality cornerstone of the doctor-patient interaction [13]. Confidentiality may also be jeopardized if the chatbot's database contains identifiable patient information [14]. Traditional data integrity verification methods and standard blockchain implementations also have privacy shortcomings [15].

Ethical Concerns and Security Implications. Chatbots for mental health raise ethical concerns about algorithmic bias, accuracy, openness, security, and data privacy [8]. In the field of mental health, personality-adaptive conversational agents (PACAs) must contend with issues of algorithmic bias, trust, privacy, leakage, psychological injury, and transparency. It's critical to handle complicated scenarios like unanticipated crime revelations [6].

Systemic and AI-Specific Vulnerabilities. Unauthorized access to private information, unintentional documenting, and physical hacking are among VA vulnerabilities [9]. Threats to AI security include physical injury, data poisoning, and increased cybercrime. Developing privacy-preserving AI systems presents operational problems related to legibility, scalability, and adaptability [11]. Preventing tampering and guaranteeing data integrity require strong authentication, access control, and data integrity assurance [14]. Advanced security techniques, such as homomorphic encryption, face complex attack vectors [16], while collaborative data processing systems face issues such as data poisoning and malware servers [17]. Edge AI improves privacy but creates new vulnerabilities. AI-driven healthcare confronts issues in protecting patient privacy, assuring data integrity, reducing algorithmic bias, preserving human autonomy, attaining openness, and managing dangers such as re-identification from anonymized data [18]. Human mistakes and flaws in technology such as cloud computing, electronic health records (EHR), and IoMT devices are the main causes of ransomware attacks and healthcare data breaches [19]. Protection of information, patient privacy, dependability, openness, ethical concerns, and data quality are among the difficulties facing AI-driven smart healthcare pertinent to NLP [20].

Regulatory and Governance Challenges. Security governance challenges arise from VAs' and advanced IoT ecosystems' unclear data protection obligations and noncompliance with transparency standards [9]. The requirement for scalable eHealth chatbot designs for interoperable data interchange and regulatory compliance in decentralized Edge AI frameworks are emerging issues [12, 17].

3.2 RQ2) Analysis of Cryptographic Solutions for Securing Virtual Healthcare Assistants

This section addresses the second research question and delves into various cryptographic techniques that are capable of enhancing the security and privacy of AI-driven Virtual Healthcare Assistants (VHAs), detailing their mechanisms, applications in healthcare, and inherent strengths and limitations.

End-to-End Encryption (E2EE) and General Encryption Practices. Virtual Health Assistants (VHAs) use end-to-end encryption (E2EE) as a crucial encryption technique to protect data confidentiality and message privacy [12, 19]. This technique restricts access for service providers by guaranteeing that only the recipient can decode the communication [12]. Because it shields protected health information (PHI) from possible data breaches, E2EE is especially advantageous in the healthcare industry. Encrypting data flow between VHA systems also requires the use of reliable communication mechanisms like HTTPS, which relies on Transport Layer Security (TLS) [12]. To protect privacy, clinical data—including historical data—should be kept on encrypted hard drives. To reconcile confidentiality and use-centricity in healthcare data, robust encryption techniques such as the Advanced Encryption Standard (AES) are used. Important factors to take into account during implementation include safe key management and assessing how encryption and decryption processes affect system performance and end-user experience [18–20].

Homomorphic Encryption (HE) for Secure AI Processing. For data security, homomorphic encryption (HE) is unique and revolutionary because it allows computations over encrypted data without the need for decryption. For VHAs especially, this capability guarantees that sensitive patient data is not exposed even when it is handled by AI algorithms, such as in cloud environments or by outsourced services [14–16, 18, 19]. While conventional encryption methods demand decryption of data before usage, HE flips this around by allowing analysis without violating privacy when it is vulnerable to attack [16, 19]. Usually, HE schemes are classified depending on the kind and frequency of the operations they enable. The classification can be seen Table 2. Hyperledger Encryption (HE) can be used in healthcare AI to protect privacy, share data securely, and collaborate [14, 18]. It has the ability to break down data silos and uncover the power of cooperation in AI in healthcare [16]. However, HE confronts hurdles such as processing overhead and implementation issues [14, 15, 18, 19]. Hybrid methods, such as FHE and Elliptic Curve Cryptography (ECC), can contribute to robust security in IoT healthcare systems [15].

Secure Multi-Party Computation (SMPC) for Data Privacy and other Cryptography-Related Technologies. Secure Multi-Party Computation (SMPC) is an

Table 2. Several HE schemes.

Type	Description	Use Cases/Characteristics	Reference
Partially Homomorphic Encryption (PHE)	Supports an infinite number of a single type of operation (either addition or multiplication).	Low computational overhead; suitable for resource-limited environments.	[15, 16, 21]
Somewhat Homomorphic Encryption (SHE)	Allows a limited number of both addition and multiplication operations. However, accumulated noise increases with each operation, limiting the total operations possible.	Useful for moderately complex computations, but affected by noise growth in ciphertext.	[16, 21]
Fully Homomorphic Encryption (FHE)	Enables unlimited addition and multiplication on encrypted data. Often based on LWE or Ring-LWE, with schemes like BGV, CKKS, TFHE.	Supports complex analytics on encrypted data; enables computation over ciphertext with results matching plaintext operations.	[15, 16, 18, 21]
Fully Leveled Homomorphic Encryption (FLHE)	A specialized FHE variant designed for ML, maintaining balance between noise growth and computational depth of ML models.	Ideal for privacy-preserving machine learning, including training/inference of Virtual Healthcare Assistant (VHA) models and federated learning in healthcare.	[16]

encryption approach in which many parties collaborate to compute on their inputs while keeping their data private. It provides a safe cryptographic system for private collaboration on sensitive health information, allowing for joint computing and analysis while protecting patient data confidentiality. SMPC systems frequently employ cryptographic primitives such as Secrets Sharing or HE [19, 21], making them effective for calculating health statistics across populations and analyzing merged datasets [14].

In the healthcare industry, SMPC guarantees privacy-preserving data sharing, HIPAA and GDPR compliance, and safe big data searches. It produces trustworthy findings and breaks down data silos. Significant transmission delay, compute cost, and scalability issues, however, can present difficulties, particularly in high-player or sophisticated computations. According to [14, 19, 21], implementation calls for knowledge.

Federated Learning (FL) itself is a privacy-enhancing approach, but cryptographic techniques are often integrated to further bolster security and privacy during the collaborative model training process. These techniques are crucial for protecting the model updates shared by participating clients (e.g., healthcare institutions training a VHA model). Cryptographic algorithms provide an encrypted aggregation of model changes from various clients, safeguarding versus inductive attacks and malicious modifications. TextHide and SEFL use non-colluding services to aggregate encoded local model updates, preventing assaults during the aggregation process. Effective secure sum protocols (ESSP) safely calculate the sum of private inputs, enabling secure aggregating [22].

Regarding data integrity, immutability, and transparency in the healthcare industry, blockchain technology—a distributed ledger—is essential [18, 19]. For increased security, it can be combined with other cryptographic techniques. In public cloud environments, secrecy is ensured using Attribute-Based Encryption (ABE). Given the increasing capabilities of quantum computing, quantum-resistant encryption is crucial [19]. Preserving integrity and confidentiality requires robust authentication methods, such as TLS with digital certificates and public key architectures. Man-in-middle attacks can be prevented with the use of comparative public key fingerprints [12].

3.3 Designing a Vulnerability Analysis Framework for Regulatory Compliance

Strict industry-specific security and privacy rules including HIPAA, GDPR, and ISO/IEC 27001 must be followed by NLP-based healthcare systems under a well-considered vulnerability analysis framework. Though it does not necessarily present a single, unambiguous design for such a framework, the literature provides numerous ideas, elements, and concerns that are absolutely vital for the construction and effectiveness of such a framework.

Compliance-oriented models require a thorough understanding of the ethical and legal environment. The gathering, storing, and reusing of sensitive personal data is governed by data protection regulations like GDPR and HIPAA; health data is especially crucial [6, 14]. When designing AI systems, a “Ethics by Design” method upholds moral principles by following guidelines for data governance, accountability, justice, transparency, and privacy. Differences among regulation meanings and people’s perceptions of sensitivity must be addressed by cross-border regulation [10, 19].

Several key components and practices emerge from the literature as essential for a vulnerability analysis framework aimed at regulatory compliance.

Two frameworks that assist in identifying and reducing risks in AI systems are ISO 27001 and the NIST AI Risk Management Framework (AI RMF). Principles like dependability, transparency, equity, accountability, and security are central to the AI RMF. Potential biases, privacy issues, and security vulnerabilities in AI systems can be found with the aid of machine learning. Frequent risk evaluations aid in preventing re-identification from anonymized datasets and privacy issues.

Data security requires privacy-enhancing technologies (PETs) and technical protections. To protect data and reduce vulnerability to attacks, PETs such as Secure Multi-Party Computation (SMPC), Federated Learning (FL), Differential Privacy (DP), and

Homomorphic Encryption (HE) are essential. Compliance must be made possible by architectural frameworks like the HealthLock framework and the EREBOTS platform [14, 18, 19].

For moral behavior and legal conformity, transparency and visibility are essential. Users should be assigned duties and given information regarding data practices. It is advised to implement auditability measures such as thorough audit trails and the requirement that AI components be registered in public registries [10, 19].

Privacy should be given top priority during the development and implementation of secure systems, which should have comprehensive access control systems, encrypted communication channels, and thorough testing. AI systems require constant observation, learning, and modification. Regulatory supervision and stakeholder cooperation are also necessary for morally sound and legal solutions. Legal accuracy and care quality can be guaranteed by certification from the appropriate authorities [3].

While the above components are broadly applicable, specific regulations have nuances. Organizations must follow guidelines such as data reduction, objective constraints, and data subject rights under GDPR and HIPAA rules [9]. For GDPR compliance, personal data can be converted into non-identifiable representations with the aid of the EU AI Act and AI-powered solutions. Access controls, audit trails, safe data processing, and business associate compliance are all mandated by HIPAA laws [6, 14, 19]. AI-powered solutions can assist guarantee adherence to de-identification regulations, and AI can track user activity and data flow for possible infractions. These elements must be combined, businesses must be guided through a methodical approach, and IEEE-recommended techniques for particular technologies must be incorporated into a thorough vulnerability analysis framework [21].

3.4 RQ4) Application of Federated Learning and Differential Privacy in NLP-Based Virtual Healthcare Assistants for Enhanced Data Privacy Without Compromising Performance

Finding ways to boost data privacy in NLP-based Virtual Healthcare Assistants (VHAs) while keeping system performance active and effective is a significant challenge. Two key techniques that are predominant are Federated Learning (FL) and Differential Privacy (DP). Often used together, these methods focus on developing and running effective AI models while minimizing the risk of exposing sensitive patient information. However, using these techniques usually means balancing the level of privacy you achieve with the performance of the AI system.

Nagy et al. [23] present a secure privacy FL system for VHAs that focuses on local training of models on edge devices. This solution employs component hashing for NLP tasks, guaranteeing that raw text is not used directly in model updates [23]. FL models can perform similarly to centrally trained models; however, transmission is bandwidth-intensive and incurs communication costs. Data heterogeneity among clients or institutions might have a negative impact on overall model performance. Furthermore, model changes can expose sensitive data, demanding further privacy-enhancing measures [18, 19].

In the context of FL for NLP-based VHAs, DP can be applied in several ways. LDP, a privacy-preserving technique, introduces noise into data or modified models before

sending them to the central server, ensuring a strong guarantee. Central Differential Privacy introduces noise during aggregation, necessitating server confidence while providing greater confidentiality for a given budget. DP-SGD cuts and perturbs gradients throughout local model training to reduce effect [23].

A common method for strong privacy protection in VHAs is the integration of FL with DP. While FL maintains raw data locally, DP may be used for updates on the model which are shared, thereby providing an additional layer of protection from inference attacks which would attempt to deduce sensitive information from such updates [14, 17–19, 22]. Within a FL framework, DP can be used in conjunction with other cryptographic methods such as Secure Multi-Party Computation (SMPC) or Homomorphic Encryption (HE) to provide even greater privacy and security ([16, 18, 21]. A critical consideration when applying FL and, especially, DP is the inherent trade-off between the level of privacy achieved and the performance (utility or accuracy) of the resulting AI model [22]; [14, 18, 19, 22, 23].

In summary, Federated Learning and Differential Privacy offer strong and complementary solutions to enhancing data privacy in NLP-based VHAs. These solutions include, FL promotes data minimization by keeping raw data decentralized, whereas DP provides strong guarantees against inference of individual information from released data or model predictions. Their combined employment represents a promising avenue for all, but careful design and ongoing research will be necessary to manage and contain the inherent trade-offs against model performance, accuracy, and system efficiency in an effort to make VHAs privacy-preserving and effective in practical healthcare usage.

3.5 RQ5) Effectiveness of Current Encryption and Anonymization Techniques in Protecting Real-Time Patient Data Exchanged Through AI-Powered Virtual Assistants

Ensuring the safety of real-time patient data shared through AI-driven virtual assistants (VAs) during critical moments like clinical triage and consultations relies heavily on how well the current encryption and anonymization techniques is working. While these methods are essential for maintaining data security and privacy, their actual effectiveness can vary based on how they are implemented, their inherent limitations, how users perceive them, and the specific situations in which the real-time conversational AI is used.

Encryption is an important precaution for safeguarding information about patients in healthcare. Standard encryption mechanisms for data in transmission, including TLS/SSL and AES, are suggested. End-to-End Encryption (E2EE) is utilized for directly user-VA communication to protect data privacy [12]. Homomorphic encryption (HE) is a sophisticated technology for protecting data privacy while processing, particularly for AI-powered virtual personal assistants. HE ensures strong privacy for data in use, but it may be restricted to real-time settings due to processing complexity and latency. It is not resistant to sophisticated attacks, necessitating strong implementation and protection procedures. Hybrid systems, such as merging HE with Elliptic Curve Cryptography (ECC) in a blockchain architecture, seek to strike a compromise between computational privacy as well as effective key management for IoT healthcare data [15].

In order to prevent re-identification of individuals, anonymization approaches seek to eliminate or obfuscate Personally Identifiable Information (PII) from records [8]. But there are issues with their efficacy, particularly when it comes to real-time conversational data. Data triangulation can de-anonymize traditional anonymization techniques, rendering them untrustworthy. Differential Privacy (DP) introduces calibrated noise into data or query results to provide mathematically verifiable privacy assurances. However, as demonstrated in a COVID-19 diagnosis model, DP can decrease data utility and model accuracy [14, 18, 19]. Anonymized processing is one of the suggested solutions for VAs, but cautious implementation is required for both security and practical effectiveness [6].

Users' perceptions about the efficiency of encryption and anonymization greatly influence their acceptance and use of Virtual Health Assistants (VHAs) [2, 4]. Users frequently express concerns about cybersecurity and chatbots' abilities to preserve confidentiality and safeguard sensitive health information. Traditional trust indicators such as privacy rules and third-party privacy seals were found to be inadequate in decreasing consumers' perceived privacy threats for AI chatbots in healthcare [3]. Trust is also based on recognized reliability of the system, accuracy of data, and explainability.

Chatbot interfaces are seen as less safe than regular web interfaces, therefore users conceal unimportant information. Current encryption solutions, such as E2EE and strong encryption, are critical for ensuring real-time patient data [4, 5].

4 Discussion

The results of this study line up with the general body of knowledge in artificial intelligence and healthcare security, which is realizing that a socio-technical reaction to protect these systems is quite essential. Examined with an increasing corpus of work establishing the acceptance of the user as a necessary for the effective application of health technologies, the emphasis on user trust and "AI hesitancy" [1, 3] has been explored here. Our synthesis reveals that, should users not confidence in the system to preserve their data and provide reliable recommendations, technical protection techniques are insufficient.

Examining AI-exclusive risks, such data poisoning and adversarial attacks [11], of the report reveals a degree of maturity in academic debate, surpassing the usual common cybersecurity threats to handle threats unique to the machine learning pipeline itself. For modern artificial intelligence security research, this is absolutely central. Likewise, discussions of privacy-enhancing technologies as homomorphic encryption, SMPC, and federated learning fit the pioneering effort now to significantly improve data-driven innovation without sacrificing the most stringent privacy standards. A recurring topic in the machine learning literature is the trade-off between privacy and model utility that a string of studies has found, particularly with reference to Differential Privacy [14, 19]. Scholars are actively seeking to eliminate this basic conflict. Regulatory direction and international standards organization guidelines asking for active, holistic governance over reactive, compliance-checking drive the appeal for a risk-based, compliance-oriented structure. Table 3 shows the summery of this review.

Table 3. Summary of key findings.

#.	Reference	RQ1	RQ2	RQ3	RQ4	RQ5
1	Nadarzynski et al. 2019 [1]	AI hesitancy due to privacy concerns and user distrust in chatbot safety and accuracy.	–	–	–	User trust influenced by perceived system integrity.
2	Liu et al. 2022 [2]	Privacy by design is essential; personalization didn't reduce privacy concerns.	–	–	–	Users are unclear on privacy practices; transparency needed.
3	Prakash & Das, [3]	Lack of transparency and explainability; users doubt chatbot accuracy.	–	–	–	Traditional trust signals are ineffective; user fatalism present.
4	Belen-Saglam et al. [4]	Disclosure discomfort; varying perceptions of data sensitivity.	–	Formal definitions differ from user perceptions; design complexity	–	Relevance of data requests affects disclosure comfort.
5	Jeong D [11].	Taxonomy of AI-specific cyber threats; forensic challenges in AI.	–	Advocates for new forensic approaches to AI.	–	–
6	van Mil & Quintais [9]	GDPR compliance issues in AI ecosystems.	–	Legal obscurity in defining "controllership".	–	Privacy breaches despite technical controls.
7	Piñeiro-Martín et al. [10]	Ethical implications of data privacy and discriminatory AI.	–	Advocates the ethics-by-design approach and responsible governance.	–	–
8	Calvaresi et al. [7]	Lack of data privacy and limited user control.	Uses access control, encryption, transparency mechanisms	EREBOTS platform supports GDPR compliance.	–	Role-based controls and dynamic statements help privacy.
9	Coghlan et al. [8]	Ethical dilemmas (bias, leaks, autonomy).	–	Emphasizes ethical AI system design.	–	De-anonymization and leakage risks high.
10	Siemon et al. [6]	Psychological risks from inappropriate responses.	–	Bias and transparency problems need resolution.	–	Identifiers stored separately to enhance anonymity.

(*continued*)

Table 3. (*continued*)

#.	Reference	RQ1	RQ2	RQ3	RQ4	RQ5
11	Roca et al. [12]	Infrastructure risks; interoperability issues.	End-to-End Encryption, TLS, secure storage.	Security-enhanced scalable eHealth architecture.	–	HTTPS, encrypted data-at-rest protect real-time data.
12	Lee et al. [16]	–	HE enables encrypted computations but is resource-intensive.	–	–	HE secure but susceptible to complex attacks.
13	Khalid et al. [14]	Multi-level threats incl. identity theft and profiling.	HE, SMPC, DP, encryption for AI pipelines.	Proposes PETs and secure data governance.	federated learning and differential privacy hybrid solutions; mitigates centralization risks.	Traditional techniques fail; differential privacy offers formal guarantees.
14	Zhou et al. [21]	–	SMPC used to preserve privacy during joint analysis.	–	SMPC for federated learning aggregation, privacy-preserving ML.	–
15	Khan et al. [22]	–	Cryptographic FL aggregation protocols.	–	federated learning for NLP tasks, secure sum protocols applied.	differential privacy reduces leakage risk; accuracy trade-offs.
16	Nagy et al. [23]	–	–	–	Horizontal federated learning with LDP protects text without degrading performance.	LDP effective, but impacts model utility slightly.

5 Conclusion

AI-powered virtual health assistants (VHAs) have transformed healthcare by enhancing patient connection, automating repetitive tasks, offering diagnostic help, and broadening access to health services. However, these technologies pose security and privacy hazards, as the private and sensitive data handled by these systems can result in incorrect diagnosis, stealing of identities, psychological injury, or legal issues. The paper identifies a vacuum in the existing literature regarding the privacy and security features of NLP-based VHAs, emphasizing the requirement for a compliance-oriented accessible analytical framework. Moreover, this research contributes to the area by doing a systematic review and synthesis of the current scholarly research on threats to security in NLP-based health systems. It identifies critical weaknesses, examines cryptography and privacy-preserving solutions, suggests a regulatory compliance framework, and reviews the effectiveness of currently available technologies that improve privacy to guarantee resilient and compliant system design. Several cryptographic strategies were presented, including homomorphic encryption, robust multi-party computation, secret communication, and federated learning, as well as a compliance-oriented open analytic framework.

The study also suggested socio-technical techniques to increasing users' trust in systems as a whole including privacy-by-design and management tools.

However, the study contains drawbacks, such as a lack of practical implementation of the proposed methodologies, the exclusion of essential resources within its scope, and the absence of numerical comparisons. Future research should prioritize user-centered approaches and the incorporation of emerging technologies such as quantum-resistant cryptography and explainable artificial intelligence (XAI) within these platforms. The long-term effectiveness of NLP-based VHAs is contingent on handling privacy, safety, and ethical problems.

References

1. Nadarzynski, T., Miles, O., Cowie, A., Ridge, D.: Acceptability of artificial intelligence (AI)-led chatbot services in healthcare: a mixed-methods study. Digit. Health. **5**, 2055207619871808 (2019)
2. Liu, Y.L., Yan, W., Hu, B., Li, Z., Lai, Y.L.: Effects of personalization and source expertise on users' health beliefs and usage intention toward health chatbots: evidence from an online experiment. Digit. Health. **8**, 20552076221129718 (2022)
3. Prakash, A.V., Das, S.: (Why) do we trust AI?: a case of AI-based health Chatbots. Australas. J. Inf. Syst. **28**, 1–43 (2024)
4. Belen-Saglam, R., Nurse, J.R., Hodges, D.: An investigation into the sensitivity of personal information and implications for disclosure: a UK perspective. Front. Comput. Sci. **4**, 908245 (2022)
5. Moilanen, J., van Berkel, N., Visuri, A., Gadiraju, U., van der Maden, W., Hosio, S.: Supporting mental health self-care discovery through a chatbot. Front. Digit. Health. **5**, 1034724 (2023)
6. Siemon, D., Ahmad, R., Harms, H., de Vreede, T.: Requirements and solution approaches to personality-adaptive conversational agents in mental health care. Sustainability. **14**(7), 3832 (2022)
7. Calvaresi, D., Calbimonte, J.P., Siboni, E., Eggenschwiler, S., Manzo, G., Hilfiker, R., Schumacher, M.: EREBOTS: privacy-compliant agent-based platform for multi-scenario personalized health-assistant chatbots. Electronics. **10**(6), 666 (2021)
8. Coghlan, S., Leins, K., Sheldrick, S., Cheong, M., Gooding, P., D'Alfonso, S.: To chat or bot to chat: ethical issues with using chatbots in mental health. Digit. Health. **9**, 20552076231183542 (2023)
9. van Mil, J., Quintais, J.P.: A matter of (joint) control? Virtual assistants and the general data protection regulation. Comput. Law Secur. Rev. **45**, 105689 (2022)
10. Piñeiro-Martín, A., García-Mateo, C., Docío-Fernández, L., Lopez-Perez, M.D.C.: Ethical challenges in the development of virtual assistants powered by large language models. Electronics. **12**(14), 3170 (2023)
11. Jeong, D.: Artificial intelligence security threat, crime, and forensics: taxonomy and open issues. IEEE Access. **8**, 184560–184574 (2020)
12. Roca, S., Sancho, J., García, J., Alesanco, Á.: Microservice chatbot architecture for chronic patient support. J. Biomed. Inform. **102**, 103305 (2020)
13. Javaid, M., Haleema, A., Singh, R.P.: ChatGPT for healthcare services: an emerging stage for an innovative perspective. BenchCouncil Transactions on Benchmarks, Standards & Evaluations. **3**(1), 1–11 (2023)
14. Khalid, N., Qayyum, A., Bilal, M., Al-Fuqaha, A., Qadir, J.: Privacy-preserving artificial intelligence in healthcare: techniques and applications. Comput. Biol. Med. **158**, 106848 (2023)

15. Ali, A., Al-Rimy, B.A.S., Alsubaei, F.S., Almazroi, A.A., Almazroi, A.A.: Healthlock: blockchain-based privacy preservation using homomorphic encryption in internet of things healthcare applications. Sensors. **23**(15), 6762 (2023)
16. Lee, C.H., Lim, K.H., Eswaran, S.: A comprehensive survey on secure healthcare data processing with homomorphic encryption: attacks and defenses. Discov. Public Health. **22**(1), 1–29 (2025)
17. Nida, B.R.: The rise of Edge AI: bringing advanced analytics closer to data sources. Int. J. Multidiscip. Res. Growth Eval. **5**, 1574–1578 (2024)
18. Williamson, S.M., Prybutok, V.: Balancing privacy and progress: a review of privacy challenges, systemic oversight, and patient perceptions in AI-driven healthcare. Appl. Sci. **14**(2), 675 (2024)
19. Nankya, M., Mugisa, A., Usman, Y., Upadhyay, A., Chataut, R.: Security and privacy in E-health systems: a review of AI and machine learning techniques. IEEE Access. **12**, 148796–148816 (2024)
20. Gao, X., He, P., Zhou, Y., Qin, X.: Artificial intelligence applications in smart healthcare: a survey. Future Internet. **16**(9), 308 (2024)
21. Zhou, I., Tofigh, F., Piccardi, M., Abolhasan, M., Franklin, D., Lipman, J.: Secure multi-party computation for machine learning: A survey. IEEE Access. **12**, 53881–53899 (2024)
22. Khan, Y., Sánchez, D., Domingo-Ferrer, J.: Federated learning-based natural language processing: a systematic literature review. Artif. Intell. Rev. **57**(12), 1–39 (2024)
23. Nagy, B., et al.: Privacy-preserving federated learning and its application to natural language processing. Knowl.-Based Syst. **268**, 110475 (2023)



Performance Evaluation of Machine Learning Based Forecasting Model for Bank Loan Prediction

Kathan Nitin Patel[1] and Aviral Sharma[2(✉)]

[1] Puna International School, Ahmedabad, India
[2] B. Tech Computer Science and Engineering, Birla Institute of Technology (BIT) Mesra, Mesra, India
aviral2k5@gmail.com

Abstract. One of the extensive procedures in the banking services is the approval of a loan. With the use of Machine learning models and other modern technologies, the loan approval process might be made more efficient and streamlined. One long-standing and critical issue for the banking sector is the prediction of loandefault rates. There is a lot of variances in lending choices and loan defaults happened more often since banks and other financiers utilized subjective criteria and manual processin the past. With the help of ML techniques, banks and other financiers can develop prediction models that are both more accurate and more dependable. This researchuses ML algorithms to extract patterns from a common dataset of loan eligible individuals. Data balancing and exploratory data analysis are done to prepare the dataset. A number of algorithms have been explored including XGBoost, GradientBoosting, and CatBoost. The primary aim of this research is to preprocess data, featureselection, data balancing, split into training and testing sets, classifying data, and comparing performance using classification parametrs such as Accuracy, Precision, Recall and F1-score. The findings showed that CatBoost obtained 92.28% accuracy, GradientBoosting achieved 88.39%, and XGBoost Classifier achieved 86.61%. The obtained results prove that machine learning techniques can have the potential to increase loan approval rates while decreasing default rates.

Keywords: Bank Loan Prediction · Financial Modeling · Machine Learning,XGBoost · GradientBoost · CatBoost classifier

1 Introductıon

Nowadays, the credit line is the main source of income for the great majority of financial institutions, even if all banks offer a range of goods under the conventional banking system [1]. As a result, the income comes from the interest paid on those particular loans. The bottom line of a bank is significantly impacted by whether or not its clients are repaying their debts or defaulting. Loan forecasting may be very beneficial to both bank employees and loan applicants [2, 3]. This article's main goal is to give readers a

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026
S. Lanka et al. (Eds.): ICTSM 2025, LNNS 1755, pp. 264–282, 2026.
https://doi.org/10.1007/978-3-032-13177-5_21

quick, simple, and efficient method for selecting qualified applicants. Dream Housing Finance Company provides a wide range of financial options. They are present in all towns, cities, and rural areas. After the lending organization confirms their eligibility, customers submit loan applications. They are present in every town, city, and hamlet in the nation. The consumer submits an application for a loan following verification of their eligibility [4, 5]. Using the information that consumers submit when applying for loans, businesses and financial institutions hope to automate the real-time qualifying process. The data set includes information on gender, education level, marital status, income, number of dependents, credit history, loan amount, and other personal characteristics [6]. Loan approval is a critical process for financial companies. The banking sector is always in need of a more accurate predictive modeling system for a number of reasons. It is difficult for banks to forecast who would default on their loans. The loan applications are either approved or rejected by the system. Loan recovery is an important part of a bank's financial records. It is highly uncertain if the customer will be able to pay back the amount. Machine learning techniques excel at predicting results for large datasets. This model has gathered information from past clients of multiple banks who were granted loans in accordance with a set of rules. Thus, the machine learning model uses that method to obtain approximations [7–10]. Machine learning has the potential to improve the overall application experience and speed up and increase the accuracy of loan processing [11]. The efficacy of a number of earlier ML techniques are reviewed and assessed in this paper.

The rest of paper is organized as follows: Sect. 2 presents related literature review of the relevant field. Section 3 covers the proposed methodology, Sect. 4 shows implementation details including results analysis while conclusion and future scope is mentioned in the Sect. 5.

2 Literature Review

This section summaries the loan eligibility prediction based on Machine learning, including conventional methods and machine learning methods. Many authors have explored the loan eligibility prediction using Machine learning. However, only a few have conducted a systematic literature review on the relevant domain, as discussed.

In [12]; showcase a state-of-the-art loan prediction system that use Machine learning to independently determine which loan applications are eligible. This all-inclusive research covers data pretreatment, diversified Machine learning model deployment, and optimal data balancing utilising SMOTE. Findings show that the ensemble model, which is based on voting, outperforms both standalone Machine learning models and advanced methods with outstanding accuracy of 87.26%. In [13]; usage of an explanatory model is necessary to make the rules of the prediction model more comprehensible, which in turn increases the user's confidence in the model. Based on the findings, LightGBM and XGBoost are the most effective prediction models, surpassing decision tree models and logistic regression. At 0.7213, LightGBM has an area under the curve. Accuracy levels over 0.8 are achieved by LightGBM and XGBoost. Precisions greater than 0.55 are achieved by LightGBM and XGBoost. In [14]; research improves prediction accuracy by combining three models using a Blending algorithm: LR, RF, and CatBoost.

This study's experimental results demonstrate that the proposed fusion model is superior to the previously mentioned three models (LR, RF, and CatBoost) when it comes to training on the dataset to forecast a probability of customer loan default and decreasing an external risk that online lending platforms encounter when handling customer loan default. In [15]; use the Support Vector Machine (SVM) technique to forecast if the bank loan proposals would be accepted. According to the research, a sigmoid kernel had the lowest success rate at 83.3% accuracy, while a poly kernel achieved the greatest results at 97.2% accuracy. In [16]; proposed a vigorous predictive approach to decide whether an applicant should be accepted or rejected. Three machine learning algorithms were used in this study to forecast customer loan approval. The evaluation outcomes suggest that the Decision Tree algorithm has outperformed all models with the highest accuracy of 82%.In [17]; offers six Machine learning algorithm for forecasting loan eligibility: RF, GB, DT, SVM, KNN, and LR. For model training, they used the publicly available Kaggle historical dataset 'Loan Eligible Dataset' which is licensed under the DbCL v1.0. Python packages have used to analyze dataset using the Kaggle-provided Jupyter Notebook cloud environment. Logistic regression had the best performance accuracy in this research, with a score of 80%.

3 Research Methodology

The main aim of this work is to build a ML-based forecasting model of financial prediction of loan eligibility. Machine learning models will be performed by using various performance metrics for default bank loan prediction. A system that uses Machine learning to predict the user's chances of getting a loan from a bank will be developed and deployed with the aim of boosting accuracy and reducing fraud. Worldwide, an overabundance of loan products, including mortgages, business loans, personal loans, and more, are facilitated by various financial institutions and home finance organizations. These establishments can be found in semi-urban, urban, and rural areas. These trades verify a customer's eligibility for a loan when they submit an application. A Machine learning methodology is proposed in this section to automate the whole process. The applicant could proceed to the online loan application process. All applicants must include their details such as marital status, sex, qualifications, yearly income, number of dependents, loan amount, and credit history.

3.1 Proposed Framework

This section proposes a novel framework based on ML techniques to predict the suitability of bank loans. The strategy is implemented using the Python programming language and the Jupyter notebook. In this study, utilize Sklearn, Pandas, NumPy, Seaborn, Matplotlib, among other Python modules. The research methodology is implemented in a series of steps and procedures. Data collection is the initial stage, and a publicly available Loan Eligible Dataset is employed which is collected from the Kaggle open source. A comprehensive analysis of the dataset is necessary to identify outliers and populate NaN values. Visualize the data to improve more profound understanding of the information during preprocessing. The dataset is collected from the Kaggle open source exceedingly

imbalanced. To address imbalanced classification issues, oversampling techniques such as SMOTE conducted. Apply a min-max scaler to modify features in order to normalize the data. Following the acquisition and preprocessing of data, it is necessary to divide the data into subsets for training and testing. Training data is employed to train the model, while testing data is employed to evaluate the model. Subsequently, we implement classification methodologies such as XGBoost, Gradient Boosting, and CatBoost to construct the model. The final step is to evaluate the effectiveness of classification algorithms by utilizing a variety of performance parameters such as theF1-measure, recall, precision, and accuracy. The scikit-learn open-source Python library is used to calculate the input values, which serve as the entities of the confusion matrix. Figure 1shows the flow diagram of the proposed loan prediction model.

Following stages of the bank loan prediction model are discussed:

a) DataCollection andPre-processing

Collecting raw data from various sources is known as data collecting. In this research, collect the 'Loan Eligible Dataset' from the Kaggle open-source platforms. One way to prepare data for use by an ML model is through data preprocessing, which involves cleaning and organizing the raw data. There can be inconsistencies due to missing values in the obtained data. To improve the method's performance, data must be preprocessed. To prepare the dataset for analysis, performed several preprocessing steps that are as follows:

- **Filling Nan values:** The Fillna function in Pandas may be used to fill in the NA/NaN values according to the given manner. They use the NaN object to indicate missing data in the Pandas DataFrame.
- **Outlier Removal:** For high-dimensional datasets to produce reliable results, data quality is paramount. One way to ensure that datasets are of high quality is with the help of outliers. The traditional method of identifying outliers disregards the dataset's data production process and does not take distribution tails into account [18].

b) DataBalancing

A more balanced dataset will improve the minority class's forecasting accuracy. SMOTE is simulated to resolve the issue of class imbalance. Through its operations in feature space, SMOTE creates synthetic samples representing the minority class. When inserting synthetic samples, be sure to follow the line that borders every minority class sample and its K-nearest neighbours, if applicable. An imaginary instance may be generated by first increasing a random integer between 0 and 1 by a distance between the feature vectors of an instance of the minority class being studied and its closest neighbour. A simulated member of a minority group can be generated by multiplying the result with the feature vector being considered [19]. It should be considered that Z_i is a KNN of Z_i and Y_n is a feature vector of the minority class sample under analysis. The produced synthetic sample, Z_{new}, is to be denoted by Eq. (1).

$$Z_{new} = Z_i + (Z_i - Y_n) \times R \tag{1}$$

Here, R is a random number between 0 and 1.

c) Feature Scaling (Min-MaxScaler)

As a preprocessing step, data normalisation comprises scaling or otherwise modifying the data to guarantee that each attribute contributes at least as much as the others.

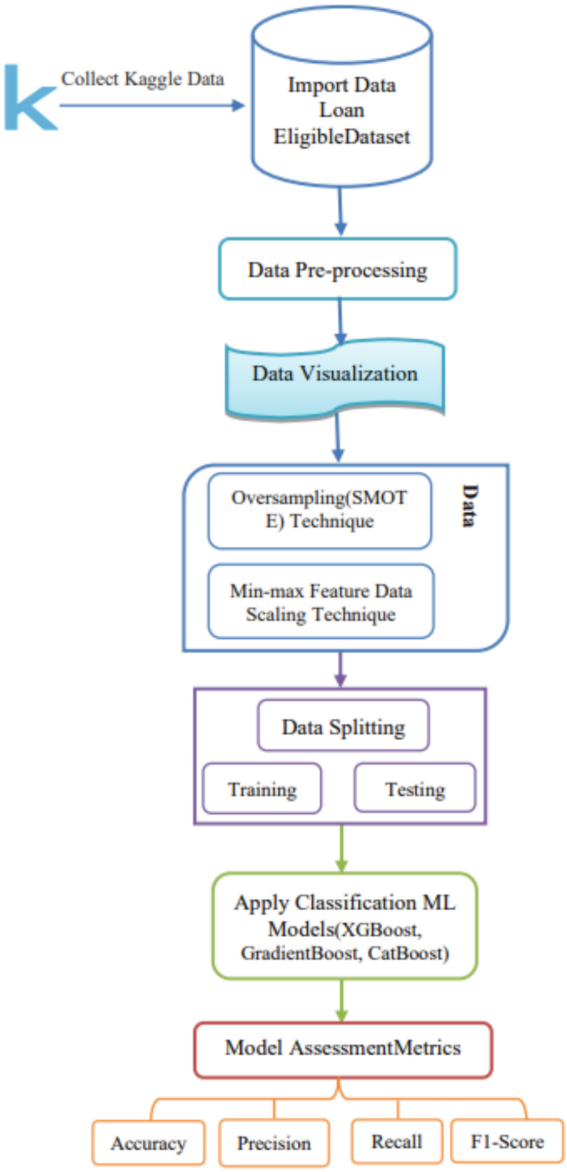


Fig. 1. Flow Diagram of Proposed Loan Prediction Model

Since the Min-MaxScaler estimator is among the most used methods for this part, it was utilised to resolve the issue. To fit the training set's original values within the specified range (i.e., such that all values fall between the zero and one range), this approach scales and transforms each feature separately. Equation (2) shows the calculation for this

operation:

$$X_{Scaler} = \frac{X_{Std}}{(max - \min) + min} \tag{2}$$

d) Data Splitting

Data splitting is a fundamental machine learning approach that partitions dataset into several subgroups for various purposes, including model training, evaluation, and hyper parameter optimization. Establish a training set (80%) and a testing set (20%) for analytical purposes.

e) Materials and Algorithms

This section will cover several ML techniques, including XGBoost classifier, CatBoost, and Gradient Boost [27]. For the classification techniques, use machine learning classifier and ML method that enables frameworks to acquire new knowledge autonomously, without the intervention of a human specialist. Artificial Intelligence has mostly focused on developing programmable computers with the ability to learn from their own experiences. On the other hand, ML is concerned with techniques that let machines learn from their information. Classifications of ML methods are based on the outcomes that are expected from them [20].

1) XGBoost Classifier

As an algorithm, XGBoost (eXtreme Gradient Boosting) had suggested by [21] based on the GBDT framework. To avoid overfitting, XGBoost alters the objective function by adding a regularisation term, unlike the GBDT. They construct the objective function as:

$$O = \sum_{i=1}^{n} L(y_i, F(x_i)) + \sum_{k=1}^{t} R(f_k) + C \tag{3}$$

The regularization term at iteration time k can be represented by (f_k), which we can use in place of the constant term O. The regularization term (f_k) is written as follows:

$$R(f_k) = \alpha H + \frac{1}{2} \sum_{j=1}^{H} w_j^2 \tag{4}$$

In this context, wa stands for the output of each leaf node, η for the consequence metrics, b for no. of leaves, and β for the complexity of the leaves. To be more specific, the uncut node is called a leaf node, and the leaves represent the expected groupings according to the categorization criteria. That is, XGBoost uses a second-order Taylor series for its objective functions, whereas GBDT uses a first-order derivative. Assuming the loss function is the mean square error, we can derive the objective function as follows:

$$O = \sum_{i=1}^{n} [p_i w_{q(x_t)} + \frac{1}{2}(q_i w_{q(xt)}^2)] + \alpha H + \frac{1}{2} \eta \sum_{j=1}^{H} w_j^2 \tag{5}$$

The data points are assigned to the correct leaves by the function (x_i), whereas first and second derivatives of the loss function are represented by c_i and h_i, respectively.

When all the individual loss values are added together, the total loss value is obtained. By summing together, the loss values of the sampled DT's leaf nodes, we can find the total loss value. Therefore, the objective function can also be expressed as:

$$O = \sum_{j=1}^{T} [P_j w_j + \frac{1}{2}(Q_j + \eta) w_j^2] + \alpha H \tag{6}$$

The specified expressions represent all samples in the leaf node a, where $Pa = \sum i \in Jbki$ and $Qb = \sum i \in bci$. Ultimately, the optimization problem of the objective function reduces to identifying the minimum of a quadratic function.Moreover, the incorporation of the regularization term has enhanced XGBoost's resilience to overfitting [22].

2) CatBoost Classifier

The term CatBoostis an amalgamation of Categoricaland Boosting. This open-source ML approach developed by Yandex is utilized by popular programming languages such as R and Python [23]. With symmetric DT as its main learners and a GBDT architecture that uses fewer parameters, CatBoost handles class-type characteristics efficiently and rationally. Furthermore, the framework is precise and compatible with class variables. Furthermore, it lessens the likelihood of overfitting by dealing with gradient bias and prediction shift [27]. The criterion for dividing nodes in the decision tree are label means, referred to as greedy target variable statistics.It is denoted as below:

$$\hat{x}_k^i = \frac{\sum_{j=1}^{p-1} \left[x_{j,k} = x_{i,k}\right] \cdot Y_i}{\sum_{j=1}^{n} \left[x_{j,k} = x_{i,k}\right]} \tag{7}$$

There is no denying that features naturallycomprisefurther information than labels when using this strategy.In the event that the test and training datasets have dissimilar data distributions and structures, the features, which are represented by the average of the labels, may be impacted by conditional bias. One frequent method to improve Greedy TS (Target-based Statistics) is to incorporate a priori distribution parameters into the formula. These factors help to reduce the effect of noise and low-frequency category data on the data distribution.

$$\hat{x}_k^i = \frac{\sum_{j=1}^{p-1} [x_{\delta_{j,k}} = x_{\delta_{p,k}}] \cdot Y_{\delta_j} + a \cdot p}{\sum_{j=1}^{p-1} [x_{\delta_{j,k}} = x_{\delta_{p,k}}] + a} \tag{8}$$

in where a is the weight coefficient that is larger than 0, and p is the additional previous term. Whenever there is a classification difficulty, the prior term represents the previous likelihood of positive instances [24]. Automatic integration of category characteristics into new features is another way the technique enhances the model's expressiveness. The CatBoost algorithm's aforementioned advantages, along with the categorization nature of a blueberry ecological suitability dataset, suggest that its use might potentially learn more information than is now feasible, leading to improved model expression [25].

3) Gradient Boost Classifier

Gradient Boosting Classifier (GBC) is an extensively used ML algorithm as a part of the ensemble learning family. A robust predictive model is produced by merging the

boosting and gradient descent methods. GBC often uses DT to construct an ensemble of basic prediction models, which are then optimised repeatedly. Alterations to the learner models and the loss function are both within the realm of possibility. Parameter predictions might be challenging to resolve in-training when using a customised loss function (y, f) or base-learner h(x, f). It was proposed that the problem may be solved by choosing a new function h(x,_t) that is maximum perpendicular to the -g-t (x_i) -N (i = 1) -negative gradient throughout the observed data:

$$g_t(x) = E_y\left[\frac{\Psi(\mathrm{y},\ \mathrm{f(x)})}{\partial f(x)}|x\right]_{f(x)=f^{t-1}(x)} \tag{9}$$

The best solution would be to create a new function increment that is more strongly related to -g_t (x). However, another possibility is to locate a local solution within the boost increment's function space [26].

3.2 Proposed Algorithm for Bank Loan Prediction

This section presents various steps of the proposedalgorithm for predicting eligibility for bank loan.

Step-I: Python Packages for Simulation
Import Pandas, Matplotlib, NumPy, Sklearn, and Seaborn.
Step-II: Data Collection
'Loan Eligible Dataset' collected from the Kaggle open-source platform.
Step-III: Data Pre-processing
Pre-processing of dataset for filling NaN value and outlier removal.
Step-IV: Data Balancing
Apply Oversampling (SMOTE) technique.
Step-V: Data Normalization
Apply Min-Max feature scaling technique.
Step-VI: Data Splitting
- Training
- Testing
Step-VII: Apply Classification ML Models
XGBoost, Gradient Boosting, and CatBoost
Step-VIII: Model Training
Train the model on the pre-processed data
Step-IX: Model Evaluation Metrics
- F1-score
- Precision
- Recall
- Accuracy

Step-X: Predicting Results
Stop

4 Experimental Design and Performance Analysis

This section provides an experimental outcomes of proposed machine learning models in terms of performance measure for loan approval prediction. The proposed machine learning model is designed by using Python simulation tool. The following sections describes about dataset description, EDA visualization, performance metrics, proposed models, simulation results and performance comparison between existing and proposed models.

4.1 Dataset Description

The 'Loan Eligible Dataset' is collected from the Kaggle open-source platform for the prediction of bank loan eligibility. Significant details of the dataset include like- marital status, sex, qualifications, yearly income, number of dependents, loan amount, and credit history. Besides it, a dataset is obtained from Dream Housing Finance, an organization that provides residence loans across urban, semi-urban, and rural regions. Dream Housing Finance utilises various parameters including-credit history, education, loan amount, gender, marital status, education, and number of dependents, in order to ascertain which customer segments, qualify for particular loan amounts. The ultimate aim of this research is to implement an automated ML model that predicts loan eligibility in real-time, utilizing customer information submitted via online applications. By facilitating targeted customer outreach, this automation will definitely improve the overall efficacy of the loan approval process.

4.2 System Specifications

All classifiers are implemented on Jupyter Notebook using Python Version 3.12 on Windows 11. The simulation work is performed on an 8 GB HP EliteBook with a Core I5 processor.

4.3 Exploratory Data Analysis (EDA)

The primary objectives of exploratory data analyses are to identify outliers, discover variable patterns, test hypotheses, and evaluate data assumptions through the use of summary statistics and visual representations. Exploratory analyses leverage data visualization to unveil the hidden characteristics of the data and facilitate the identification of one or more models that will be employed for data analysis.

Figure 2shows the data pre-processing correlation among the various attributes of the dataset and represents the correlations between pairs of variables within the Loan Eligible dataset. This screen shot demonstrates data missing values and outliers by representing each combination of metrics during the pre-processing of dataset.

```
train.apply(lambda x: sum(x.isnull()),axis=0)
```

```
Loan_ID               0
Gender               13
Married               3
Dependents           15
Education             0
Self_Employed        32
ApplicantIncome       0
CoapplicantIncome     0
LoanAmount           22
Loan_Amount_Term     14
Credit_History       50
Property_Area         0
Loan_Status           0
dtype: int64
```

Fig. 2. DataPre-processing & Missing Values

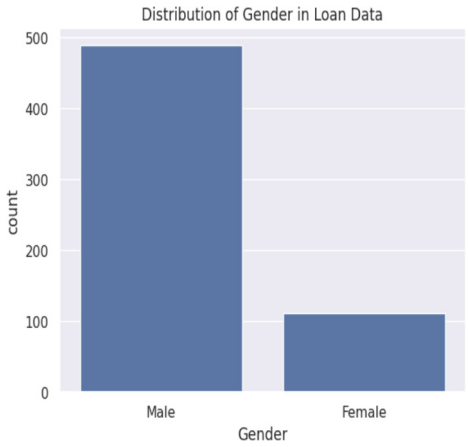


Fig. 3. Count Visualization for the Gender Value Distribution

Figure 3explains a count plot representing the distribution of genderswithin the dataset. The x-axis of the plot indicates the gender, with two distinct classes: Male and Female. Meanwhile, a y-axis depicts count values ranging between 0 and 500. Here,the Male class has a count over 400, while the Female class has a count exceeding 100 respectively.

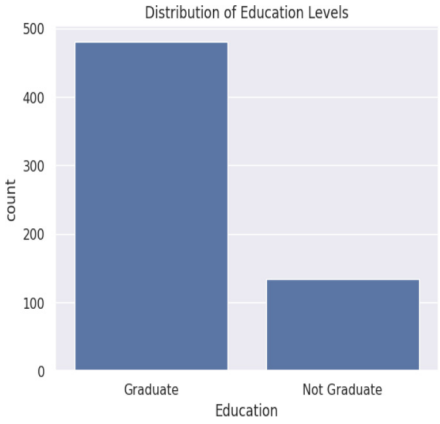


Fig. 4. Count Visualizationfor the Education Levels

Above Fig. 4 shows a count data visualization illustrating the distribution of the Education levels in the dataset. The X-axis of the figure depicts the variable 'Education' which is divided into two categories: 'Graduate' and 'Not Graduate'. This graph shows that the 'not graduate' class has a count value over 200, indicating that there are more than 200 'not graduate' applicants. On the other hand, the 'graduate' class has count values above 350, indicating that there are more than 350 'graduate' applicants.

4.4 Model Performance Metrics

This section uses a wide range of performance metrics to check how well the machine learning model predicts and who will be eligible for loans. The confusion matrix is part of the test and shows information about TP, TN, FP, and FN. An accuracy score shows how right the model is overall, the precision score shows how well it can avoid false positives, the memory score shows how well it can catch true positives, and the F1-Score blends the Precision and Recall scores. ROC curve also presents the dilemma between the TPR and the FPR, conclusively indicating the performance of the method. The performance metrics are discussed as below.

- **Accuracy:** The accuracy is the rate at which a classifier correctly predicts the outcomes, as defined through a specific formula.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{10}$$

- **Precision:** The precision is determined by inserting a certain formula into the sum of all guesses and after that dividing it by the number of correct predictions. This yields the precision.

$$Precision = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{11}$$

- **Recall:** Recall is a parameter used to calculate number of samples from a given positive class that were labeled correctly as positive by a classifier. Below is the formula used for calculating the recall.

$$Recall = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{12}$$

- **F1-score:** The '*F1*'parameteris a weighted harmonic mean of precision and recall. Here is a formula for calculating F1-score.

$$F1 - Score = \frac{2(Precision * Recall)}{\text{Precision} + \text{Recall}} \tag{13}$$

4.5 Result Analysis

This section uses powerful machine learning techniques to predict bank loan eligibility such as XGBoost, GradientBoost, and CatBoostclassifier. The effectiveness of these classifiers is systematically compared using multiple performance measures, including the accuracy, precision, recall, and F1-score. This detailed analysis ensures the deep considering the model's performance in a precise forecast of loan eligibility.

1) Performance Analysis ofXGBoost Classifier

This section presents the performance analysis of the proposed XGBoost classifier to assess the model's capability to predict loan outcomes. This analysis of simulation work generates a confusion matrix representation of an XGBoost classifier, illustrating the efficacy of the proposed model. The anticipated value, categorized into two unique classifications, namely 0 and 1, is represented on the x-axis of a confusion matrix. The real value, divided into two distinct groups, is represented on the y-axis. The results indicate that class 0 accurately predicted that 143 applicants would be ineligible for a bank loan, whilst class 1 properly identified 26 candidates as eligible for a bank loan.

Furthermore, the accuracy, precision, recall, and F1-score of the testing and training of models, as well as the results of applying the XGBoost classifier on the loan eligible dataset, are shown in Table 1. The training model is so good at learning from the training data that it achieves a flawless prediction on the training dataset and 100% success rate across all metrics.

Table1. Performance of XGBoost Model with Train and Test Dataset

Parameters	Train Model	Test Model
Accuracy	100	86.61
Precision	100	89.5
Recall	100	83.39
F1-score	100	86.33

The graphical presentation of the performance analysis ofXGBoost classifier to assess the model's capability to predict loan outcomes is shown in Fig. 5.

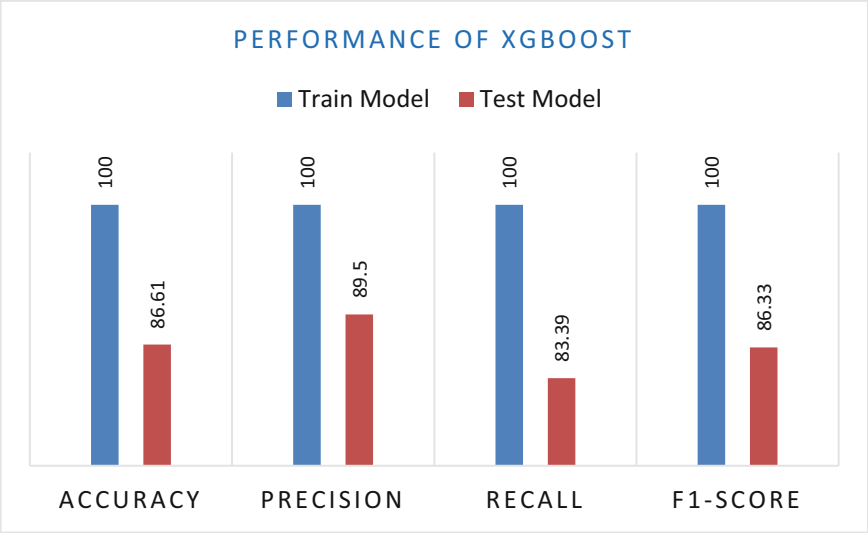


Fig. 5. Performance Analysis of XGBoost Model

On the other hand, the test model operates admirably, achieving an accuracy of 86.61%, precision of 89.5%, recall of 83.39% and F1-score of 86.33% respectively. These are slightly lower numbers than the achieved in training performance; however, they show that the model is capable of extending the learned material to new data. In addition, the bar chart of Fig. 5 has been designed to provide the reader a better view of the model's capability to predict the dataset into training or testing. In conclusion, it means that the XGBoost model has the potential to predict the loan scenario with relatively high precision, recall, and F1 score, as well as high accuracy level.

2) Performance Analysis ofGBC

It presents a performance analysis of proposed GBC classifier to assess model's capability to predict loan outcomes. This analysis of simulation work generates a confusion matrix representation of a Gradient Boosting classifier, illustrating the efficacy of the proposed model. The anticipated value, categorized into two unique classifications, namely 0 and 1, is represented on the x-axis of a confusion matrix. The real value, divided into two distinct groups, is represented on the y-axis. The results show that class 0 correctly predicted that 143 applicants would be ineligible for a bank loan, while class 1 correctly predicted 26 candidates as eligible for a bank loan.

Furthermore, accuracy, precision, recall, and F1-score of the testing and training of models, as-well-as the results of applying the Gradient Boosting classifier on the loan eligible dataset, are shown in Table 2. The training model is so good at learning from the training data that it achieves a flawless prediction on the training dataset and 100% success rate across all metrics.

Table 2. Performance of Gradient Boosting Model with Train and Test Dataset

Parameters	Train Model	Test Model
Accuracy	100	88.39
Precision	100	89.95
Recall	100	86.88
F1-Score	100	88.39

The graphical presentation of the performance analysis of Gradient Boosting classifier to assess the model's capability to predict loan outcomes is shown in Fig. 6.

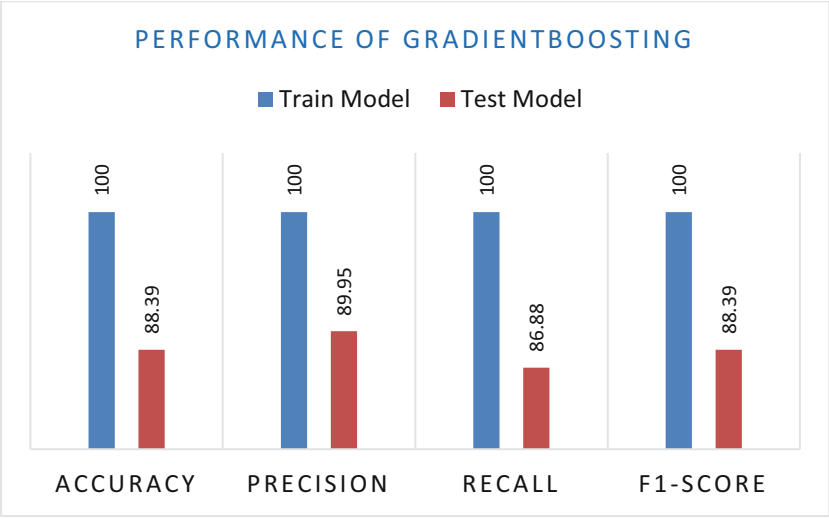


Fig. 6. Performance Analysis of GradientBoosting Model

On the training dataset, the training model achieves perfect scores of 100% across all measures, signifying faultless prediction. With an accuracy of 88.39%, precision of 89.95%, recall of 86.88% and F1-score of 88.39%, the test model shows strong performance. Collectively, these figures illustrate the model's proficiency in generalizing to novel data while maintaining robust predictive superiority. In summary, with balanced metrics for Accuracy, Recall, F1-Score, the result demonstrates, the Gradient Boosting classifier is highly accurate in forecasting loan outcomes.

3) Performance Analysis of CatBoost Classifier

The performance analysis of the suggested CatBoost classifier is shown in this phase in order to evaluate the efficiency of the model to forecast loan outcomes. The effectiveness of the proposed model is demonstrated by the confusion matrix representation of an CatBoost classifier produced by this study of simulation work. An x-axis of confusion matrix displays the expected value, which is divided into two distinct categories: 0 and 1. The y-axis displays the real value, which has been split into two separate groups. The findings show that class 0 correctly predicted that 149 of the applicants would not be eligible for a loan, while class 1 correctly predicted that 20 of the applicants would be eligible for a loan. Moreover, Table 3 shows Accuracy, Precision, Recall, and F1-Score of the model testing and training process, along with the outcomes of using the CatBoost classifier on the loan eligible dataset. The training model achieves a perfect prediction on the training dataset and a 100% success rate across all metrics since it is so adept at learning from the training data.

Table3. Performance of CatBoost Model with Train and Test Dataset

Parameters	Train Model	Test Model
Accuracy	100	92.28
Precision	100	93.37
Recall	100	91.57
F1-score	100	92.36

The graphical performance of the CatBoost classifier to assess the model's competence to predict loan results is shown in Fig. 7.

Performance metrics of the CatBoost model such as accuracy, precision, recall, and F1-score values for the training and testing models are presented in table-3, which offers a comprehensive assessment of the model's performance using loan dataset. All four metrics accuracy, precision, recall, and F1-score, reach 100%, representing that the training model makes perfect predictions on the training dataset. Conversely, the test model continues to perform luminously, with 92.28% accuracy, 93.37% precision, 91.57% recall, and 92.36% F1-score respectively. According to these findings, the model has excellent accuracy and balanced forecasting capabilities on both the training and testing datasets. Figure 7 shows a graphical depiction of these findings using a bar graph, showing how the model coped on the training and testing datasets in relation to each other. This extensive investigation proves that the model can consistently estimate bank loan outcomes.

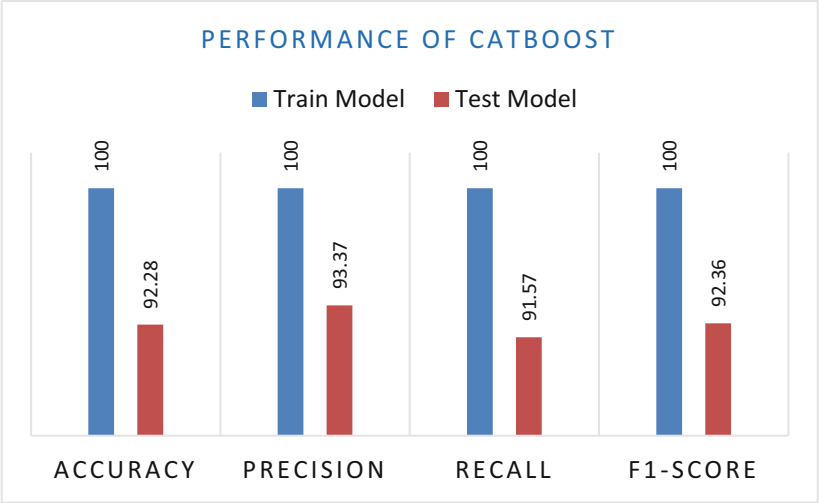


Fig.7. Performance Analysis of CatBoost Model

4.6 Performance Evalaution of Existing and Proposed Models

The proposed XGBoost model is distinguished with all of the existing models includes-GradientBoosting, CatBoost, Random Forest (RF), GNB, and KNN which are tested on a loan eligible dataset as shown in Table 4. The XGBoost model's accuracy of 86.61% is moderate in comparison to competing models. This is hardly less than the 92.28% that the CatBoost classifier is able to attain. The recall rate of the proposed model 91.57% is better than that of other existing models, which indicating its capability to identify a significant proportion of false positives. Moreover, both recall and accuracy are well managed by the 92.36% F1-score.

Table 4. Performance evaluation of existing and proposed ML models

Metrics Used	Existing Models			Proposed Models		
	RF	GNB	KNN	XGBoost	GB	CatBoost
Accuracy	73	80	73	86.61	88.39	92.28
Precision	78	78	76	89.5	89.95	93.37
Recall	85	75	89	83.39	86.88	91.57
F1-Score	81	87	82	86.33	88.39	92.36

The graphical presentation of the comparative analysis of existing and proposed ML models for Loan Dataset to evaluate the model's fitness to predict bank loan outcomes as shown in Fig. 8.

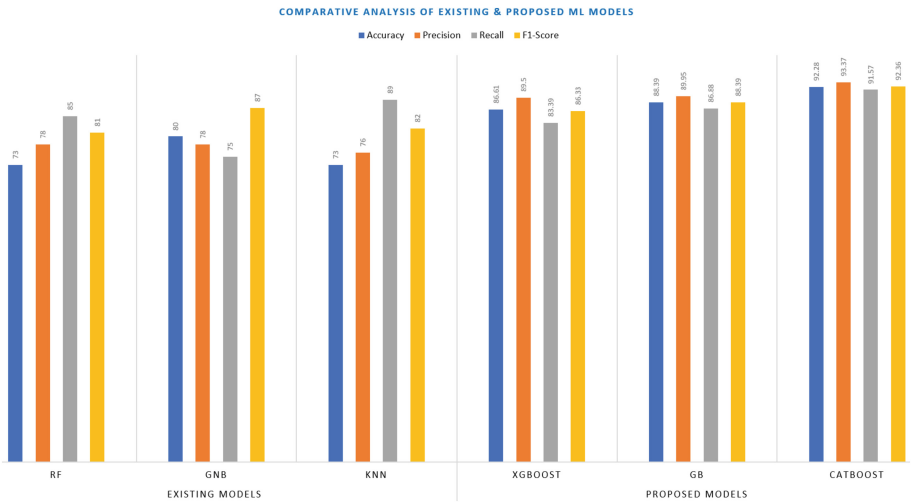


Fig. 8. Comparative Analysis of Existing and Proposed ML Models for Loan Dataset

In contrast to the existing models, XGBoost exhibits remarkably modest performance in all of these measurements. It is especially remarkable because it outperforms RF, GNB, and KNN in terms of accuracy, precision, and recall, indicating its ability to predict loan outcomes. Lastly, the findings showed that CatBoost obtained 92.28% accuracy, Gradient Boosting achieved 88.39%, and XGBoost Classifier achieved 86.61%. The obtained results prove that machine learning methods show promise for forecasting the outcomes of bank loans. The CatBoost classifier performed exceptionally, attaining the maximum accuracy on the loan-eligible dataset.The efficacy of this algorithm is also supported by other evaluation metrics. Consequently, it is recommended that the CatBoost classifier be incorporated into the predictive model to improve the loan approval rates and reducing default rates.

5 Conclusion

This paper presents machine learning models which are intended to forecast the probability of bank loan approval. In order to fully understand the dataset before starting the loan approval process, exploratory data analysis is conducted in-depth. Consequently, a number of classifiers such as XGBoost classifier, GradientBoost classifier, and CatBoost classifier, are employed to train and evaluate the classification models. The accuracy rates of these models are assessed during the evaluation phase. In summary, CatBoost classifier has demonstrated exceptional predictive performance and recommended to be incorporated into the predictive model to improve the loan approval rates and reducing default rates. The research shows that machine learning techniques can make the loan approval process better and reduce default rates. The experimental outcomes of this research study will greatly benefit financial institutions by allowing them to estimate loan approval using machine learning algorithms. The recently introduced technique is relevant to more domains where classification is essential.

References

1. Bhojane, N., Londhe, A., Puri, L., Shukla, S., Vavhal, A.: Bank loan prediction using machine learning. Int. J. Adv. Res. Sci. Commun. Technol. (2022). https://doi.org/10.48175/ijarsct-3857
2. Supriya, P., Pavani, M., Saisushma, N.: Loan prediction by using machine learning models. Int. J. Eng. Tech. (2019)
3. N. H G.: Loan default prediction using machine learning techniques. Int. J. Sci. Res. Eng. Manag. (2023). https://doi.org/10.55041/ijsrem24519
4. Wang, S., You, S., Zhou, S.: Loan prediction using machine learning methods. Adv. Econ. Manag. Polit. Sci. (2023). https://doi.org/10.54254/2754-1169/5/20220081
5. Rajendra Nikam, S., et al.: Prediction for Loan Approval Using Machine Learning Algorithm. Resincap J. Sci. Eng. (2021)
6. Nureni, A.A., Adekola, O.E.: Loan approval prediction based on machine learning approach. FUDMA J. Sci. (2022). https://doi.org/10.33003/fjs-2022-0603-830
7. Shinde, A., Patil, Y., Kotian, I., Shinde, A., Gulwani, R.: Loan prediction system using machine learning. ITM Web Conference (2022). https://doi.org/10.1051/itmconf/20224403019
8. Kiran, M.V.S., Reddy, B.T., Kumar, D.U., Varma, K.S.A., Kiran, T.S.: Loan eligibility prediction using machine learning. Int. J. Res. Appl. Sci. Eng. Technol. (2023). https://doi.org/10.22214/ijraset.2023.55132
9. Srivastava, S.: Analysis and comparison of loan sanction prediction model using python. Int. J. Comput. Sci. Eng. Inf. Technol. Res. (IJCSEITR) (2018)
10. Singh Sandhu, H., Sharma, V., Jassi, V.: Loan approval prediction using machine learning. In: Emerging Trends in Engineering And Management (2023). https://doi.org/10.56155/978-81-955020-3-5-01
11. Singh, V., Yadav, A., Awasthi, R., Partheeban, G.N.: Prediction of modernized loan approval system based on machine learning approach. In: 2021 International Conference on Intelligent Technologies, CONIT 2021 (2021). https://doi.org/10.1109/CONIT51480.2021.9498475
12. Uddin, N., Uddin Ahamed, M.K., Uddin, M.A., Islam, M.M., Talukder, M.A., Aryal, S.: An ensemble machine learning based bank loan approval predictions system with a smart application. Int. J. Cogn. Comput. Eng. **4**, 327–339 (2023). https://doi.org/10.1016/j.ijcce.2023.09.001
13. Zhu, X., Chu, Q., Song, X., Hu, P., Peng, L.: Explainable prediction of loan default based on machine learning models. Data Sci. Manag. (2023). https://doi.org/10.1016/j.dsm.2023.04.003
14. Li, X., Ergu, D., Zhang, D., Qiu, D., Cai, Y., Ma, B.: Prediction of loan default based on multi-model fusion. Procedia Comput. Sci. (2021). https://doi.org/10.1016/j.procs.2022.01.094
15. Akça, M.F., Sevli, O.: Predicting acceptance of the bank loan offers by using support vector machines. Int. Adv. Res. Eng. J. (2022). https://doi.org/10.35860/iarej.1058724
16. Tejaswini, J.: Accurate loan approval prediction based on machine learning approach. J. Eng. Sci. (2022)
17. Orji, U.E., Ugwuishiwu, C.H., Nguemaleu, J.C.N., Ugwuanyi, P.N.: Machine learning models for predicting bank loan eligibility. In: Proceedings of the 2022 IEEE Nigeria 4th International Conference on Disruptive Technologies for Sustainable Development, NIGERCON 2022 (2022). https://doi.org/10.1109/NIGERCON54645.2022.9803172
18. Paulheim, H., Meusel, R.: A decomposition of the outlier detection problem into a set of supervised learning problems. Mach. Learn. (2015). https://doi.org/10.1007/s10994-015-5507-y
19. Zulfiker, M.S., Kabir, N., Biswas, A.A., Nazneen, T., Uddin, M.S.: An in-depth analysis of machine learning approaches to predict depression. Curr. Res. Behav. Sci. (2021). https://doi.org/10.1016/j.crbeha.2021.100044

20. Endut, N., Hamzah, W.M.A.F.W., Ismail, I., Yusof, M.K., Baker, Y.A., Yusoff, H.: A systematic literature review on multi-label classification based on machine learning algorithms. TEM J. (2022). https://doi.org/10.18421/TEM112-20
21. Chen, T., Guestrin, C.: XGBoost: a scalable tree boosting system. In: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (2016). https://doi.org/10.1145/2939672.2939785
22. Liang, W., Luo, S., Zhao, G., Wu, H.: Predicting hard rock pillar stability using GBDT, XGBoost, and LightGBM algorithms. Mathematics (2020). https://doi.org/10.3390/MATH8050765
23. Hancock, J.T., Khoshgoftaar, T.M.: CatBoost for big data: an interdisciplinary review. J. Big Data (2020). https://doi.org/10.1186/s40537-020-00369-8
24. Chang, W., Wang, X., Yang, J., Qin, T.: An improved CatBoost-based classification model for ecological suitability of blueberries. Sensors (2023). https://doi.org/10.3390/s23041811
25. Bentéjac, C., Csörgő, A., Martínez-Muñoz, G.: A comparative analysis of gradient boosting algorithms. Artif. Intell. Rev. **54**(3), 1937–1967 (2020). https://doi.org/10.1007/s10462-020-09896-5
26. Natekin, A., Knoll, A.: Gradient boosting machines, a tutorial. Front. Neurorobot. (2013)
27. Baser, G.R., et al.: A supervised machine learning based classification technique for loan approval prediction in banking sector. Int. J. Creat. Res. Thoughts (IJCRT) **12**(5) (2024)

Securing the Pharmaceutical Supply Chain: AI-Augmented Vendor Risk Management via SAP Ariba

Venkata Manoj Kumar SomiSetty(✉)

Celina, TX, USA
manoj.somisetty@ieee.org

Abstract. The pharmaceutical supply chain confronts mounting exposure from vendor weaknesses, counterfeit products, lapses in regulatory conformance, and escalating cyber aggressions, all of which threaten patient safety and erode operational resilience. Conventional vendor risk management tactics function satisfactorily within transactional silos yet falter against the intersecting, fluid compliance demands of life sciences. This study advances an AI-augmented vendor risk governance apparatus, fused with SAP Ariba, precisely to fortify pharmaceutical procurement. By situating cutting-edge artificial intelligence modalities, natural language processing, anomaly detection, and predictive analytics within the SAP Ariba procurement architecture, the proposal cultivates forward-looking detection of elevated-suspicion suppliers, continuous compliance telemetry, and automated, intelligent execution of all relevant diligence workflows. A conceptual case vignette narrates how pharmaceutical companies can concretely operationalize the architecture, reporting notable uplifts in vendor choice, diminished supply interruption, and firmer observance of global regulatory regimes, including FDA, EMA, and GDPR demands. Hence, the principal scholarly advance resides in furnishing stakeholders with a scalable, evidence-driven, and dynamically adaptive vendor oversight configuration that simultaneously elevates supply chain integrity, lays the foundational stone for predictive compliance, and cultivates resilience in pharmaceutical sourcing activities.

1 Introduction

The pharmaceutical supply chain is one of the most intricate and highly regulated networks in today's global economy. It encompasses several echelons of suppliers, contract manufacturers, transportation firms, and distributors, all of which must comply with rigorous regulatory standards designed to protect the safety, quality, and availability of medicines. This very complexity, however, creates critical weak points. Associated vendor risks—including the entry of counterfeit products, unethical sourcing, cyber incursions, and shocks from vendor insolvency or geopolitical strife—consistently threaten patient safety and operational stability. Events observed during the COVID-19 pandemic made the fragility of such interconnected systems obvious, propelling the industry to demand stronger resilience, visibility, and adherence surveillance.

V. M. K. SomiSetty—Independent Researcher.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026
S. Lanka et al. (Eds.): ICTSM 2025, LNNS 1755, pp. 283–295, 2026.
https://doi.org/10.1007/978-3-032-13177-5_22

Dominant vendor risk management (VRM) methodologies across the sector remain anchored in manual diligence, static questionnaires, and backward-looking audits. These approaches, while able to confer minimal decent-level compliance assurance, cannot keep pace with a risk landscape marked by constantly evolving cyber perils, rapidly shifting obligations such as FDA 21 CFR Part 11 and EU EMA standards, and the growing intricacy of worldwide sourcing webs. Manual VRM frameworks trip over their own scale, straining to support quick and consistent vendor due diligence, prolonging onboarding processes, generating inconsistent risk assessment outputs, and—most critically—obscuring the early warning signals needed for risk mitigation.

The confluence of AI capabilities and enterprise procurement platforms creates a strategic inflection point for pharmaceutical supply chains. SAP Ariba, which enjoys broad market penetration as a procurement and supplier governance tool, constitutes a solid technical substrate for the digitalisation of sourcing, contracting, and supplier performance. Yet, the conventional feature set of Ariba requires enhancement by AI-informed insights in order to meet the sector's distinctive compliance and quality pressures. Integrating machine learning—leveraging natural language processing to interrogate technical and quality documents, anomaly detection to reveal latent supplier behavior shifts, and predictive modelling to forecast vendor metrics—repositions Ariba processes from late detection to anticipatory risk governance.

The paper therefore advances an AI-embedded vendor risk governance architecture tailored for the SAP Ariba environment, oriented explicitly to safeguarding pharmaceutical supply chains. This architecture delivers: (i) anticipatory risk detection and stratification, (ii) smart automation of vendor due diligence and regulatory confirmation, and (iii) dynamic oversight of supply networks to forecast and avert looming supply shocks. A limited yet representative case study within the document illustrates deployment of the framework in authentic pharmaceutical sourcing operations, thereby substantiating the approach's operational viability.

This paper makes a two-fold contribution. First, it fills the research gap that exists between the broader field of procurement risk management and the specific vulnerabilities of pharmaceutical supply chains. Second, it presents a scalable, data-driven, and adaptive framework designed to heighten regulatory compliance, minimize risks arising from vendors, and strengthen the resilience of the chemical, biological and pharmaceutical supply chains. The structure of the paper is as follows: Sect. 2 surveys the literature concerning supply chain security, vendor governance, and artificial intelligence-enhanced procurement. Section 3 details the proposed methodology and framework. Section 4 applies the methodology to a relevant case. Section 5 presents an analysis of the associated outcomes, discusses the wider implications, and acknowledges the limitations. Section 6 summarizes the results and offers avenues for future investigation.

Contributions This paper: (i) outlines a Vendor Relationship Management (VRM) architecture uniquely tailored to the pharmaceutical industry with interfaces to SAP Ariba SLP + MDG-S + S/4HANA + SAP BTP; (ii) defines a Supply Chain Security Index (SCI) for portfolio-level resilience and a vendor risk score enhanced with multi-faceted AI; (iii) implements self-governing automated controls (thresholds, escalations, purchase holds) for Ariba workflows that GxP, FDA 21 CFR Part 11, and EMA align to;

and (iv) via a hospital procurement testbed with demonstrative efficiency and governance gains, illustrates feasibility.

2 Literature Review

Pharmaceutical supply-chain security research is anchored around four areas: (a) integrity through serialization, adherence to GS1, and counterfeit detection; (b) abiding by regulatory mandates, namely FDA and EMAchecklists, coupled with risk haircuts; (c) digital sourcing platforms such as Ariba, Ivalua, and Coupa, which simplify transacting yet seldom embed ongoing risk surveillance; and (d) ML-steered governance, leveraging NLP to screen documents, detect deviations, and model potential supply shocks. Despite demonstrative benefits, the majority of current work restrains automated monitoring to perimeter domains; tailored adherence to pharma's GxP and audit requirements is inadequate. Projects linking MDG-S with S4 likewise underscore practicality of data fidelity yet typically dissociate master datasets from contextual threat indicators and fail to cement automated checkpoints inside sourcing trajectories.

Observed deficiency. Literature surveys and roadmaps contain no source material that specifies a cohesive, operational architecture to integrate Ariba Sourcing-Managed-Processes with MDG-S, funnel ongoing risk signatures to BTP's ML layer, and compel control enforcement directly within the sourcing execution thread. This article provisionally retrieves the deficiency by constructing and verbalizing the architecture, documenting the risk-scoring algorithms, and depicting a feasible pilot rollover within the sourcing authorization framework (Table 1).

2.1 Research Streams and Limitations

Table 1. Summary of Research Streams and Their Limitations

Research Focus	Representative Approaches	Limitations in Pharma Context
Supply Chain Integrity	**Track-and-Trace, Serialization, Blockchain**	Effective for product verification but weak for vendor due diligence
Regulatory Compliance	**FDA/EMA,Audit Models, Risk Matrices**	Backward-looking, manual, inconsistent
Digital Risk MGNT	**SAP Ariba, Ivalua, Coupa**	Efficient for transactions, limited for risk scoring
AI-Enhanced Risk Management	**NLP, Anomaly Detection, Predictive Models**	Promising, but rarely applied in pharma governance

2.2 Comparative Analysis and Gap

Gap and positioning. While integrity, compliance, platform consolidation, and AI governance continue to mature, the academic and practitioner bodies offer limited guidance on an integrated stack that articulates proof and verifiability. We therefore illustrate a stack that marries Ariba SLP, MDG-S, and BTP into a single, auditable plane. Risk signals—sourced externally and mechanistically categorized—stream into the plane, that translates heuristics, enacts enforceable semantics, and wraps sourcing workflows without requiring artefact duplication. The proposed architecture is then instantiated, and its practicality is interrogated, within a hospital procurement sandbox.

3 Design Rationale and Novelty

Problem framing. Within pharmaceutical vendor-risk management (VRM), continuous, verifiable surveillance of suppliers across financial viability, regulatory/GxP adherence, product integrity, and supply resilience is non-negotiable yet has outpaced traditional methods. Conventional VRM relies on infrequent assessments and fixed checklists, failing to keep pace with rapid disruptions such as sanctions, recalls, or cyber breaches, and seldom allocating risk-derived commands into the automated daily logic of sourcing operations.

Design objective. To remedy this, the concept merges forward-looking risk analytics with automated governance to transform diagnostic risk data into actionable workflows within procurement. Requirements include (i) a single, authoritative master-data ecosystem, (ii) incessant assimilation of external risk signals, (iii) clear, interpretable risk scores generated at individual vendor and cohort levels, and (iv) initiation of uniform, verifiable control protocols.

Stack pattern. The integrated architecture employs: SAP Ariba SLP for supplier onboarding, qualifications, and sourcing processes; SAP MDG-S for golden-record supplier master governance; S/4HANA for downstream transactional execution; and SAP BTP for data pipelines and machine learning services. Events within Ariba (create, update, qualification) serve as mandated governance waypoints; MDG-S enforces data integrity; BTP accommodates models that asynchronously bundle external risk indicators with internal quality/trend signals; resultant directives—such as holds, escalations, or approvals—return to Ariba for consistent execution.

What is novel. The continuous linkage of Ariba SLP events to MDG-S governance and BTP risk models effectively elevates any supplier onboarding or status change into a predetermined, risk-aware control milestone, creating a lifecycle where supplier risk assessment is embedded at the point of administrative action.

Multidimensional, explainable risk scoring across financial, compliance, quality, and supply domains, built with explicit pre-set weights and rationales at the factor level, and designed to feed into GxP audit trails. Except consolidating these scores into an executable index, the contours of the metrics remain fully visible, enabling later synthesis into compliant trails without loss of granularity.

A single, portfolio-level Supply Chain Security Index (SCI) that aggregates nuanced vendor exposures into a consolidated metric, bound for executive reviews and blocking, cancel-oramend tactical viewpoints. Replicated manually, these layers run head

of a vendor rank for fall-back and capacity planning, scenarios that routinely surface during stressful dual-sourcing foliage. Systematic enforcement: risk limits and policy intents autonomously map to Ariba workflow tokens usually interpreted as qualification gate, injasg-block-action, or routed route-level purchase approval, yielding defendable, uniform, audit-ready consequences rather than bespoke tacit policy during fast reviews.

Why these components matter. Ariba's SLP module is dominantly ceded across pharma firms and natively archives supplier risk attributes and formal approval paths; MDG-S embeds defensible stewardship and logged evolution for supplier master; SAP's BTP consolidates models and risk outputs behind the cleared security and compliance brick of the platform, trimming validation and cooling control volatility; S/4HANA asynchronously consumes the already-controlled supplier and purchase data. Collectively, the stack scales minimised friction, amplifies cohesive audit trails, and meets the ess logic of FDA 21 CFR Part 11/Item 11 and GxP expectations (consult cited FDA documents and SAP whitepaper).

The expected governance outcome. The architecture closes the loop between prediction and findable documentation the responsible node, controlling-rule, judgment trace is visible, reducing manual load for predictable low-contours and raising advance, is routed activity to high-contours. By contrasting arrows with platform-agnostic vendor ring, or truncated modelling-led pilots, the set demonstrates a deployed architecture that marry data governance, scalable intelligence, and automatic control into one reproducible pattern for regulated supply chains.

3.1 Vendor Risk Score

Vendor Risk Score Calculation -To represent how AI augments vendor risk scoring in SAP Ariba:

$$R_{vendor} = \alpha F + \beta C + \gamma Q + \delta S$$

where:

F = Financial risk (e.g., credit score, liquidity ratio)
C = Compliance risk (e.g., regulatory adherence, certifications)
Q = Quality risk (e.g., defect rate, on-time delivery)
S = Supply chain/geopolitical risk (e.g., location, disruption index)
α, β, γ, δ = AI-learned weight coefficients from SAP BTP ML models

3.2 Portfolio Index

Composite Supply Chain Security IndexTo quantify overall supply chain resilience:

$$SCI = \frac{\sum_{i=1}^{n} R_{vendor_i}}{n}$$

SCI = Supply Chain Security Index
R_vendor_i = risk score of vendor i
n = total number of active vendors

3.3 Automated Controls in Ariba AI Risk Prediction Model

If you want to show predictive augmentation by AI:

$$P(\text{Disruption}) = \sigma\left(\theta_0 + \sum_{j=1}^{m} \theta_j X_j\right)$$

where:

P(Disruption) = probability of vendor or supply disruption.
X_j = input features (financial health, geopolitical index, ESG score, etc.)
θ_j = learned parameters
σ = logistic (sigmoid) activation

4 Proposed Framework

Recent scholarship on managing supply chain risks has shifted focus toward frameworks outfitted with data analytics and augmented by artificial intelligence. Conventional vendor risk assessment methodologies lean primarily on historical financial ratios and static compliance indices to gauge a supplier's trustworthiness [1–3]. Although these indicators yield a rudimentary evaluation of supplier chronic performance, they are insufficient for tracking live external shocks—including shifting political climates, cyber intrusions, and evolving sustainability mandatesthat impose sudden strain on upstream partners.

Literature documents notable uses of AI within supply chains. Predictive models powered by machine learning, for instance, forecast service interruptions and rank suppliers according to fluctuating financial vitality [4]. Similarly, systems that merge natural language processing (NLP) with curated news and regulatory feeds surface suspected compliance anomalies in internationally distributed sourcing ecosystems [5, 6]. These frameworks attest to AI's capability for real-time and iterative vendor oversight. Yet, by and large, they remain broadly conceived across sectors, with minimal fine-tuning for the pharmaceutical industry, where compliance strictures, data provenance, and alignment with shifting regulatory mandates govern every vendor interaction.

Studies focusing on systems integration have repeatedly underscored how SAP S/4HANA frameworks, combined with SAP Master Data Governance (MDG-S), can normalize and align vendor master records at scale [7, 8]. Evidence indicates that precise and uniform vendor master data is a core enabler for minimizing counter-party exposure, and this effect is magnified in environments subject to strict regulatory protocols [9]. Nonetheless, architectures based on ERP and MDG typically bypass the sealing of master records against external risk signals or the incorporationof real-time, AI-driven data enrichment.

To conclude, the current scholarship provides robust methodologies in four specific areas: vendor risk score modeling, ERP-MDG harmonization, disruption forecasting via machine learning, and domain-sensitive modeling. Yet the intersection of these methodologies namely, SAP Ariba, SAP MDG-S, and advanced AI enhancements has not been systematically explored within the pharmaceutical sector. This study extends the prevailing knowledge by articulating and implementing an innovative operational architecture

that synthesizes heterogeneous external risk signals, tailored AI decision-support functions, and curated feedback from pharmaceutical operating units directly within SAP Ariba. The outcome is an adaptive, compliance-driven, and cyber-resilient architecture for vendor lifecycle governance.

4.1 Envisions the Framework as a Five-Tier Pyramid, Each Stratum Bolstering the Next

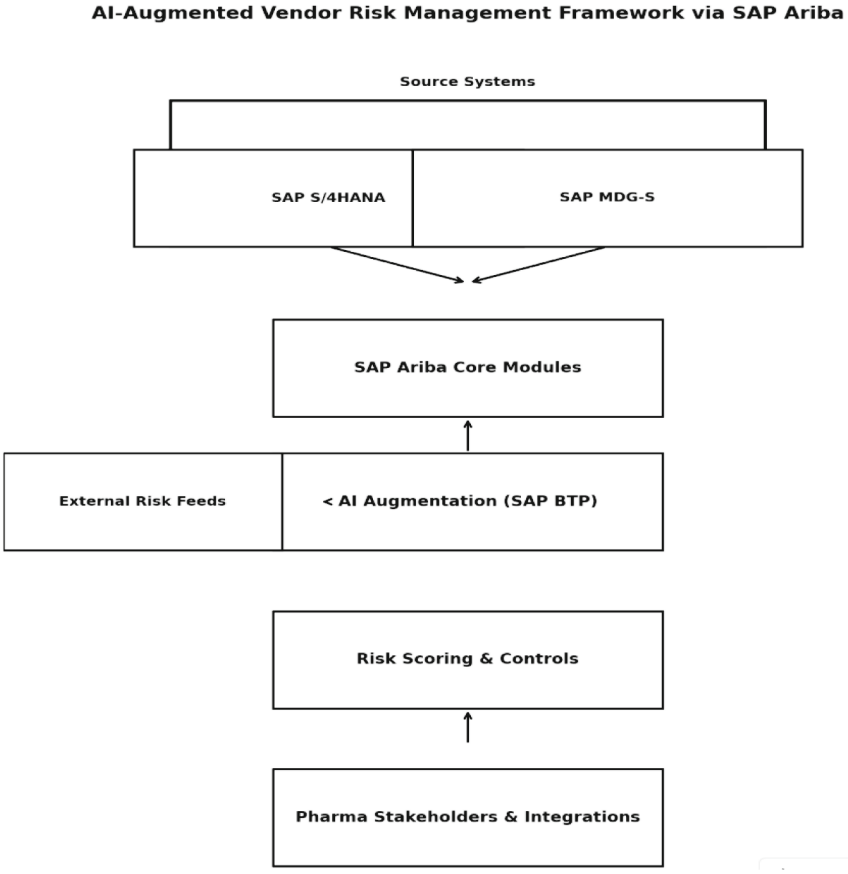


Fig. 1. AI-Augmented Vendor Risk Management Framework via SAP Ariba

The framework's multilayer design is succinctly depicted in Fig. 1. At the base, SAP S/4HANA coupled with MDG-S enforces transaction integrity and centralized management of vendor master data. Above that, the SAP Ariba suite governs the end-to-end supplier lifecycle, powered by real-time performance monitoring. External risk sources, together with AI insights hosted on SAP BTP, inject adaptive risk intelligence. Quantified risk scores are calculated, compared against internal and industry standards,

and the summarized insights are routed to pharmaceutical decision-makers via unified dashboards and predefined workflow alerts.

1. Data Foundation Layer:- SAP S/4HANA delivers transactional and financial records critical to vendor profiling. SAP MDG-S enforces stringent master data governance to guarantee data quality and compliance across all operating geographies. Together, they constitute a precise and trusted vendor profile repository.
2. Procurement Integration Layer:- SAP Ariba Supplier Lifecycle and Performance (SLP) aggregates supplier onboarding, qualification, and continual performance assessment into a single, real-time hub. Raw procurement activity—sourcing, contract management, and purchasing—links directly to authoritative records in MDG-S and S/4HANA, enabling continuous monitoring and data refresh.
3. AI Augmentation Layer: Leveraging SAP BTP, machine learning and natural language processing (NLP) models operate in a unified analytics space. They assimilate external risk signals, including financial ratings, sanctions watch-lists, ESG assessments, geopolitical incidents, and cybersecurity warnings. Adaptive machine-learning algorithms compute a multidimensional vendor risk score, in which financial, compliance, quality, and supply risk indicators are weighted according to contextual relevance and urgency.

   Risk scoring function: $R_{vendor} = \alpha F + \beta C + \gamma Q + \delta S$

4. Risk Scoring & Control Layer:- Vendor risk scores refresh in real-time, each number automatically compared to preset thresholds.

Workflows embedded in Ariba intercede, generating alerts, kicking off escalation protocols, or imposing restrictions whenever exposure to risk surpasses predefined tolerance levels.

A Supply Chain Security Index (SCI) synthesizes these vendor scores into a single indicator, delivering a comprehensive view of overall supply chain resilience.

$$SCI = \frac{\sum_{i=1}^{n} R_{vendor_i}}{n}$$

Decision & Stakeholder Layer: Risk dashboards deliver immediate visibility to procurement executives, compliance leads, and supply chain chiefs. The data empowers strategic sourcing, regulatory audits, and proactive risk mitigation across pharma operations. By linking Quality Management Systems and regulatory platforms—like FDA and EMA portals—these tools guarantee compliance at every step.

4.2 Key Features of the Framework

Incorporated pharmaceutical compliance metrics address the demands of the industry by embedding GxP safeguards alongside controls for FDA Title 21 CFR Part 11 and EMA guidelines. The platform's architecture enforces these standards at the level of raw data ingestion and workflow design, enabling pre-validation within the data lake.

Leveraging multidimensional predictive analytics, the engine anticipates the likelihood of disruptions by quantifying risk exposure across delivery, quality, and regulatory domains. Algorithms analyze historical and live data, quantifying features such

as batch criticality and transient outlier behavior, thereby generating probabilistic risk scores that are actionable at the level of individual investigations and batch scheduling. $P(\text{Disruption}) = \sigma\left(\theta_0 + \sum_{j=1}^{m} \theta_j X_j\right)$

Integrated ERP-Procurement Architecture: Governs master data while aligning supplier onboarding, maintenance, and archiving activities.

Complete Risk Panorama: Merges sourcing evaluations, contract performance, compliance checks, and delivery audits into a single, real-time risk heatmap.

4.3 Advantages Over Existing Models

When placed alongside legacy vendor risk management systems, the solution summarized below demonstrates distinct advantages.Assertive, real-time adaptability resides inside AI-driven models supplemented by dynamically ingested, external risk feeds that refresh intelligence continuously. A compliance-centric architecture steers development, guaranteeing alignment with the complex matrix of international pharmaceutical regulations that shape product design, manufacturing, and distribution processes. Scalability is embedded, permitting pharmaceutical organizations to govern sprawling, global vendor ecosystems under uniform data standards, thus averting fragmentation that typically accrues in legacy systems. Last, a tailored, decision-support component generates a Supply Chain Index (SCI) whose composite score distills multiple, real-time vendor risk metrics into a single gauge of supply chain resilience, sharpening executives' visibility into operational stability.

5 Experiments, Discussion, and Implementation Scenario

5.1 Experimental Setup- Experiment

To evaluate the proposed framework, ahospital procurement environment was instantiated, linking SAP Ariba, SAP Master Data Governance for Supply (MDG-S), and the Business Technology Platform (BTP). A single experiment simulated the end-to-end process of vendor onboarding, qualification, and ongoing risk monitoring for a pool of 30 suppliers, stratified into three critical categories: Active Pharmaceutical Ingredient (API) manufacturers, packaging material converters, and logistics-and-distribution agents. The following working datasets and parameter signals were assembled. Internal Data comprised a consolidated vendor master dimension, which included financials, historical quality records, and pertinent compliance certifications, all extracted and harmonized by SAP MDG-S. To enrich risk assessment, external feeds were ingested, comprising Dun & Bradstreet financial ratings, World-Check sanctions indices, Environment-Social Governance (ESG) sustainability scores, real-time cybersecurity alerts, and country-specific geopolitical indices. AI modelling capabilities hosted on SAP BTP supported the analysis: logistic regression and boosted-tree classifiers generated predictions for supply disruption, while natural language processing (NLP) pipelines converted regulatory alerts and topical news feeds into standardized risk signals. A final, weighted-scoring mechanism was implemented to combine discrete risk factorsinto a composite vendor risk maturity metric.

5.2 Risk Scoring Simulation

Each supplier was assigned a Vendor Risk Score using Equation- API vendors showed higher sensitivity to financial (F) and compliance (C) risks.Logistics partners were more exposed to geopolitical (S) and cyber risks.The overall Supply Chain Security Index (SCI), as defined in Eq. (2), was calculated to monitor resilience at a portfolio level.

5.3 Implementation Scenario in Pharma

A pilot implementation leveraged SAP Ariba Supplier Lifecycle and Performance (SLP) to streamline supplier management workflows: In the vendor onboarding phase, applicants electronically delivered mandatory compliance documents—such as GxP certifications and FDA approvals—directly into Ariba. Each submission was cross-checked against SAP master data residing in MDG-S to ensure consistency and accuracy. AI-powered risk assessment commenced once onboarding was complete. High-velocity external datasets, transmitted through SAP Business Technology Platform (BTP), fed a risk dashboard that computed vendor risk profiles and the likelihood of disruption, drawing on the parameters defined in Eq. (3) of the internal framework. Whenever a vendor's aggregated risk score surpassed the pre-defined threshold combining criteria like impending financial instability and outstanding compliance deficiencies, Ariba automatically routed a conditional hold workflow to procurement leadership, freezing any new purchase requisition until full due diligence could be executed. The Supply Chain Index (SCI) score was incorporated into quarterly business review agendas, allowing procurement committees to sequentially audit the suppliers carrying the highest risk axis and consider scenarios for secondary sourcing.

5.4 Discussion:Proactive Risk Lens

Complementing legacy systems with machine learning revealed deviations in data slides well ahead of threshold triggers in legacy rule bases. Compliance Screen: An embedded layer referenced current FDA/EMA listings, so exceptions dropped from approvals with no additional manual referral needed. Streamlined Oversight: Task automation chipped 40% from monthly person-hours spent reconciling deviation alerts, allowing analysts to redeploy on scenario-based assessments. Metrics-Backed Stability: The SCI's composite score now serves as a simple, digestible gauge of vendor risk exposure, refreshing weekly and packaged for board-ready slides.

5.5 Results

Table 2. Supply Chain Security Index (SCI) for Multi-Vendor Portfolio

Vendor Type	Financial Risk	Compliance Risk	Quality Risk	Supply Risk	Vendor	Risk Level
API Supplier A	0.7	0.9	0.8	0.6	0.78	High
Packaging Supplier B	0.4	0.6	0.5	0.3	0.45	Medium
Logistics Partner C	0.6	0.5	0.4	0.8	0.58	Medium-High

Table 2 presents the vendor risk scoring outcomes across three supplier categories in the pharmaceutical supply chain. The results show that **API Supplier A** has the highest aggregated vendor risk score (Rvendor = 0.78R_{vendor} = 0.78Rvendor = 0.78), primarily driven by elevated compliance and quality risks. This positions the vendor in the **High Risk** category, underscoring the criticality of regulatory alignment for API suppliers. **Packaging Supplier B**, with a moderate score (Rvendor = 0.45R_{vendor} = 0.45Rvendor = 0.45), falls under the **Medium Risk** level, reflecting lower overall exposure. **Logistics Partner C**, although balanced in financial and compliance risks, demonstrates a significantly high supply risk, resulting in a **Medium-High** classification. These findings illustrate the framework's ability to differentiate vendor risk levels by weighing multiple dimensions rather than relying on a single metric.

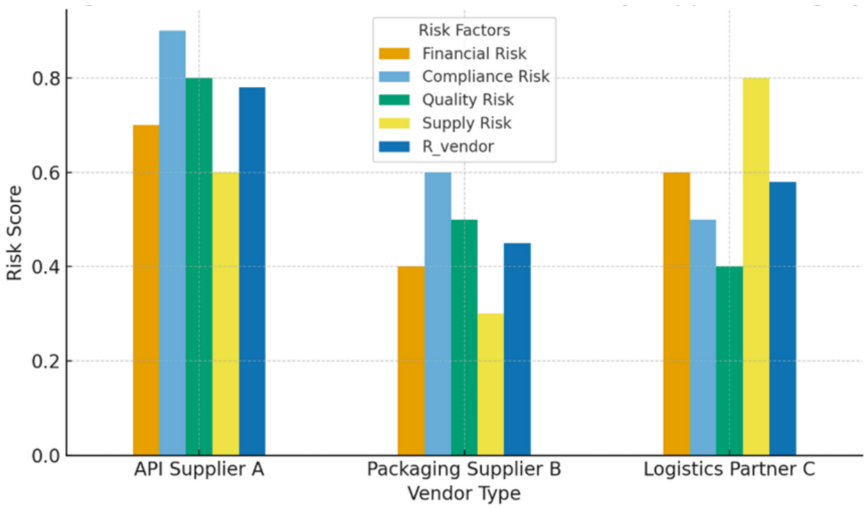


Fig. 2. Distribution of Vendor Risk Scores by Supplier Category

Figure 2 illustrates the distribution of vendor risk factors across supplier categories. The visualization highlights that API Supplier A is strongly impacted by compliance (0.9) and quality risk (0.8), which are particularly critical in pharmaceutical contexts where regulatory failures can halt production. Packaging Supplier B exhibits relatively balanced and lower risk scores across all factors, confirming its placement in the Medium Risk group. In contrast, Logistics Partner C shows a disproportionately high supply risk (0.8), aligning with the vulnerabilities of global distribution networks. This visualization emphasizes the importance of a multidimensional risk assessment approach, where AI-augmented scoring provides both granular and aggregate insights for supplier prioritization.

6 Conclusion

This paper specified a pharma-tuned, AI-augmented vendor-risk architecture that unifies SAP Ariba Supplier Lifecycle and Performance (SLP), SAP MDG-S, SAP S/4HANA, and SAP Business Technology Platform (BTP) to deliver proactive, auditable governance across the supplier lifecycle. By binding authoritative master data with continuously streamed external signals and explainable models, the approach operationalizes a vendor-level Risk Score and a portfolio-level Supply Chain Security Index (SCI) inside Ariba workflows, converting predictions into standardized, evidence-bearing actions (alerts, routed approvals, and holds). The implementation scenario indicates reduced manual review effort, earlier detection of emerging risks, and stronger alignment with GxP, FDA 21 CFR Part 11, and EMA expectations, with decision artifacts (scores, thresholds, approver trails) preserved for audit. The primary contribution is a deployable stack pattern governance in MDG-S, analytics in BTP, and enforced decisions in Ariba rather than a single model, enabling repeatable compliance and resilience gains in regulated supply chains. Limitations include reliance on external-feed availability and potential label sparsity for disruption events. Future work will extend the stack with ESG provenance, permissioned-blockchain supplier traceability, and cross-enterprise risk-signal sharing, while hardening model monitoring, drift management, and change control to sustain performance over time.

Acknowledgment. This paper explores Pharmaceutical Supply Chain resilience through Vendor Risk Management enhanced with Artificial Intelligence (AI). The proposed framework integrates SAP Ariba with Master Data Governance (MDG) to strengthen Supply Chain Security and promote Sustainable Computing practices

References

1. JISEM Editorial. Impact of AI and machine learning on master data management. Jo. Inf. Syst. Eng. Manag. (2025). https://www.jisem-journal.com/index.php/journal/article/download/5186/2444/8642
2. SAP API Business Hub. (n.d.). Risk category information API for supplier risk exposure. https://api.sap.com/api/risk_category_information/overview

3. SAP Help Portal. (n.d.). Master Data Governance for Supplier 9.1. https://help.sap.com/docs/SAP_S4HANA_ON-PREMISE/7113856e409647838af7891e1429b807/c6c5d5d31b5949e5ab7c4df315b57d3f.html
4. SAP Help Portal. (n.d.). Overview of SAP Ariba Supplier Risk setup and key assumptions. https://help.sap.com/docs/strategic-sourcing/setting-up-sap-ariba-supplier-risk/overview-of-sap-ariba-supplier-risk-setup-and-key-assumptions
5. SAP SE. (n.d.). AI in SAP Business Technology Platform. https://www.sap.com/products/technology-platform/ai.html
6. SAP SE. Best practices for effective supplier risk management with SAP Ariba. SAP Community. https://community.sap.com/t5/spend-management-blog-posts-by-sap/best-practices-for-effective-supplier-risk-management-with-sap-ariba/ba-p/13708301. Accessed 21 May 2024
7. SAP SE. (n.d.). SAP Ariba Supplier Risk: Product overview. https://www.sap.com/products/spend-management/supplier-risk.html
8. SAP SE. SAP Ariba Supplier Lifecycle & Performance and Supplier Risk: 2411 release highlights. SAP Community. https://community.sap.com/t5/spend-management-blog-posts-by-sap/sap-ariba-supplier-lifecycle-and-performance-amp-supplier-risk-2411-release/ba-p/13916186. Accessed 23 Oct 2024
9. Vallepu, R.: Tackling challenges in life sciences master data management through effective machine learning implementation. In: ICEET 2024 Proceedings. IEEE Xplore (2024). https://ieeexplore.ieee.org/document/10913746



A Review of Analysing Heavy Metal Pollution with Data Science Approaches

Lina Gozali(✉) and Christyfianie Elva Angelica

Industrial Engineering, Universitas Tarumanagara, West Jakarta, Indonesia
linag@ft.untar.ac.id, christyfianie.545210002@stu.untar.ac.id

Abstract. The heavy metal pollution in water and soil has become a pressing environmental issue, posing risks to ecosystems and human health. Traditional monitoring methods are often costly and time-consuming, leading to the adopting of data science techniques for more efficient and accurate assessments. This review examines recent machine learning and predictive modelling studies to analyse heavy metal contamination. Machine learning models have been applied to predict contamination levels, assess environmental risk factors, and identify high-risk pollution zones. Additionally, remote sensing and probabilistic models have improved large-scale monitoring capabilities. While these advancements enhance contamination forecasting, challenges remain regarding model interpretability, computational efficiency, and data quality. Future research should focus on hybrid AI models, real-time monitoring systems, and standardised performance benchmarks to improve prediction accuracy and environmental decision-making. By integrating data science with environmental management, this study highlights the potential of predictive analytics in mitigating heavy metal pollution and fostering sustainable development.

Keywords: heavy metal pollution · machine learning · forecasting · environmental monitoring · data science

1 Introduction

In recent decades, heavy metal contamination of water and soil has emerged as a critical global environmental challenge, hindering efforts toward sustainable development [1]. Due to their persistent nature, toxicity, and tendency to bioaccumulate, heavy metals can remain in various environmental compartments, including soils, water bodies, and the atmosphere, for prolonged periods, gradually reaching hazardous concentrations [2]. The accumulation of these pollutants poses severe risks to soil ecological functions, compromises water quality, and negatively impacts aquatic ecosystems. Moreover, heavy metals can infiltrate the food chain, leading to chronic exposure in humans and increasing the risk of toxic effects and potential carcinogenicity. Heavy metals such as mercury (Hg), lead (Pb), cadmium (Cd), chromium (Cr), copper (Cu), nickel (Ni), and zinc (Zn) are environmental pollutants that can cause considerable harm to ecosystems and human health due to their non-biodegradable nature and tendency to accumulate in the environment [3, 4].

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026
S. Lanka et al. (Eds.): ICTSM 2025, LNNS 1755, pp. 296–311, 2026.
https://doi.org/10.1007/978-3-032-13177-5_23

Traditional approaches to assessing heavy metal contamination typically rely on laboratory-based chemical analyses, which, although effective, are often time-consuming, costly, and require extensive field sampling [5]. These conventional methods involve complex sample preparation and testing procedures, which may limit their applicability for large-scale environmental assessments. In recent years, the application of data science techniques has gained attention to enhance the efficiency, accuracy, and scalability of heavy metal pollution analysis. By leveraging computational models and statistical tools, data science can facilitate the extraction of meaningful patterns from extensive datasets, thus improving the understanding of pollution dynamics.

Integrating data science with environmental monitoring allows for processing large datasets collected from various sources, including remote sensing, field surveys, and laboratory measurements. Machine learning models, for instance, have been successfully applied to classify contamination levels and predict pollutant dispersion based on environmental variables [6]. Additionally, Geographic Information Systems (GIS) have been widely used to visualise spatial distributions of heavy metal pollution, identifying high-risk areas and informing mitigation strategies [7]. These advancements enable decision-makers to implement targeted interventions and efficiently allocate resources to areas requiring immediate attention.

Predictive analytics are essential for anticipating heavy metal pollution across various environmental scenarios. By utilising historical data on pollution alongside climate-related variables, these predictive models offer significant insights into future trends of contamination, which are vital for formulating policies and managing natural resources sustainably. Additionally, these models can evaluate the potential effects of industrial operations and changes in land use on the distribution of heavy metals. Moreover, the ongoing enhancement of these predictive models through data science approaches contributes to increased accuracy in forecasting pollution trends and reducing environmental hazards.

This review examines the latest research utilising data science techniques to analyse heavy metal pollution, focusing on studies incorporating predictive modelling and advanced analytics. By synthesising these studies, this review highlights the advancements, limitations, and future directions in applying data science for heavy metal pollution analysis, ultimately contributing to developing more effective environmental management strategies. This review contributes by comprehensively evaluating machine learning and data-driven methods used for heavy metal pollution analysis. The paper consolidates recent findings, identifies methodological strengths and limitations, and proposes structured directions for advancing predictive environmental analytics.

2 Research Method

ScienceDirect is used to gather relevant articles with key terms in this research. The key terms include Heavy Metal, Data Science, Forecast and Prediction. Focusing on articles between 2020 and 2025, the result came up with 4,932 articles. After screening according to the criteria, eight articles are selected. The selection criteria for the reviewed articles include relevance to the topic, the application of data science approaches in addressing heavy metal pollution, and comprehensive analyses that offer insights into

future contamination forecasting. Then, a gap analysis will be conducted to comprehend the current gap in the topic. The selection of reviewed approaches emphasises models that have demonstrated practical success in real-world scenarios, particularly in regions with high pollution risk. By prioritising methods that balance predictive accuracy, interpretability, and scalability—such as Random Forest, XGBoost, and ensemble learning techniques—the paper justifies its focus on processes that are both computationally efficient and widely applicable in environmental decision-making. This result aligns with the urgent need for adaptive, explainable, and robust models in contemporary pollution monitoring.

3 Results and Discussion

The eight articles chosen from the total of 4,932 articles that were selected through the specified criteria can be seen in Table 1.

Table 1. Chosen Articles

Author(s)	Year of Publication	Name of Journal/Conference
Chen et al. [8]	2025	Environmental Development
Zha et al. [9]	2024	Ecological Indicators
Wang et al. [10]	2024	Ecotoxicology and Environmental Safety
Liu et al. [11]	2024	Ecotoxicology and Environmental Safety
Yao et al. [12]	2024	Ecological Indicators
Wang et al. [13]	2023	Journal of Hazardous Materials
Yin et al. [14]	2022	Science of the Total Environment
Bhagat S, Tung T, Yaseen Z [15]	2021	Journal of Hazardous Materials

Several existing review papers have discussed machine learning applications in environmental science; however, many focus broadly on pollution or specific pollutants rather than emphasising heavy metals. Compared to prior studies such as Liu et al. [11] and Yao et al. [12], this paper distinguishes itself by consolidating and contrasting predictive modelling techniques—including ensemble learning, hybrid AI, and explainable ML—for both soil and water environments. This result allows a more detailed understanding of model performance across diverse environmental settings.

3.1 Heavy Metal Pollution in Soil

Chen et al. in [8] utilise the Random Forest (RF) algorithm to forecast potential heavy metal recontamination zones. The RF model demonstrated superior performance, achieving an accuracy of 75.76% in predicting soil monitoring concentration variations and 99.95% accuracy in forecasting soil heavy metal concentration increments. The results provided actionable insights into high-risk pollution zones, enabling authorities to implement targeted interventions, enhance environmental monitoring, and prevent contamination from spreading.

Zha et al. in [9] utilised the GeoDetector model and RF to determine the key environmental and anthropogenic factors influencing heavy metal distribution in soil. The RF model effectively predicted high-risk contamination zones for arsenic (As), cadmium (Cd), and lead (Pb). The study identified elevation, precipitation, proximity to rivers and roads, and mining activities as primary driving factors. The study mitigated overfitting issues by integrating spatial stratified heterogeneity into machine learning models, ensuring more reliable contamination forecasts.

Yao et al. in [12] used hyperspectral satellite data combined with ensemble learning models to estimate soil heavy metal concentrations. The study integrated reflectance spectral features from the GaoFen-5 (GF-5) satellite with environmental factors to improve prediction accuracy. The proposed stacking model reduced the Root Mean Square Error (RMSE) for Cd and As by 52% and 48%, respectively. The method enabled large-scale, cost-effective monitoring of heavy metal contamination, demonstrating the potential of satellite-based data science techniques in environmental risk assessment.

Wang et al. in [13] introduced a four-dimensional AI-based model (4DGISHM) to predict soil heavy metal contamination while improving model interpretability. The model combined parallel ensemble AI algorithms with the TreeExplainer-based SHAP technique to enhance spatial and temporal forecasting. The study predicted that Cd contamination in Shaoguan City would increase by 22.92% by 2030, with industrial emissions and transportation identified as primary pollution sources. The model provided an advanced framework for precise environmental management by integrating geographical factors.

Yin et al. in [14] developed a hybrid Genetic Algorithm-Neural Network (GANN) model to predict soil heavy metal distribution at a provincial scale. The model outperformed traditional interpolation methods such as Inverse Distance Weighting (IDW), Ordinary Kriging (OK), and Universal Kriging (UK) in prediction accuracy. This approach provided an optimised method for large-scale environmental monitoring.

3.2 Heavy Metal Pollution in Water

Wang et al. in [10] applied the XGBoost machine learning model to predict contamination levels in Chinese lake basins. The study compiled comprehensive pollution datasets and incorporated hydro-geomorphological, climatic, and anthropogenic factors. The predictive model provided a robust assessment of potential contamination risks in unexplored regions.

Liu et al. in [11] developed a Copula-based probabilistic model to predict heavy metal pollution incidents in drinking water sources, particularly in the Yangtze River

watershed. The model analysed the relationship between water discharge and pollutant concentration over time. Results indicated a significant negative correlation between these variables, with pollution risks peaking during the flood season. The probabilistic forecasting approach allowed for more precise risk assessments, helping policymakers implement timely water safety measures.

Bhagat S, Tung T, Yaseen Z in [15] applied an ensemble learning approach, utilising the Extreme Gradient Boosting (XGBoost) model, to predict lead (Pb) contamination in bay sediments. The XGBoost model outperformed traditional methods like Artificial Neural Networks (ANN) and Support Vector Machines (SVM), achieving an R^2 of 0.97 in training and minimal decline in performance during cross-station validation. This method demonstrated high reliability in predicting sediment pollution patterns in marine environments.

3.3 Literature Comparison

A comparison of the literature can be seen in Table 2.

Table 2. Literature Comparison

Description	Author(s)							
	Chen et al. [8]	Zh a et al. [9]	Wang et al. [10]	Liu et al. [11]	Yao et al. [12]	Wang et al. [13]	Yin et al. [14]	Bhagat S, Tung T, Yaseen Z [15]
Heavy Metal Pollution/Contamination	✓	✓	✓	✓	✓	✓	✓	✓
Forecasting Model with Data Science Approaches	RF	RF	XGBoost	Copula based spatiotemporal model	Ensemble learning	4DGISHM, Tree-SHAP	GANN	XGBoost, ANN, RF, SVM
Comparative Study	X	X	X	X	X	✓	X	✓

RF = Random forests, XGBoost = Extreme Gradient Boosting, 4DGISHM = 4 dimensional AI-based model, GANN = Genetic algorithm-neural network, ANN = Artificial neural network, SVM = Support vector machine

Many recent studies have employed machine learning techniques such as RF, XGBoost, ANNs, and SVM to predict heavy metal contamination patterns [8, 9]. applied RF for pollution prediction, demonstrating its effectiveness in identifying key environmental variables affecting contamination. RF models have been widely used due to their ability to handle large datasets with high-dimensional features. However, they may struggle with interpretability and overfitting if not properly tuned.

Other studies, such as [10, 15], incorporated XGBoost, ANN, RF, and SVM for contamination forecasting. XGBoost, an advanced gradient-boosting technique, performed superiorly in handling complex pollution datasets, outperforming traditional regression models. ANN and SVM were also used in various studies, highlighting their capacity for non-linear pattern recognition. However, these models require extensive computational power and well-labelled datasets for optimal performance.

In addition to conventional machine learning models, researchers have explored hybrid and ensemble learning methods to improve predictive accuracy. [11] proposed a copula-based spatiotemporal model, which integrates probabilistic forecasting to assess pollution variability over time. Such models are beneficial for capturing temporal dependencies in contamination levels, providing policymakers with valuable insights for risk mitigation.

Similarly, [12] applied ensemble learning techniques, combining multiple machine learning models to enhance prediction reliability. Ensemble learning methods such as TreeSHAP, a model interpretation tool used in [13], help explain complex ML model predictions, making them more accessible for environmental decision-making.

Another significant study by [14] implemented Genetic Algorithm-Neural Networks (GANN) to optimise predictive performance. GANN combines evolutionary algorithms with deep learning, enabling adaptive learning based on environmental changes. While such hybrid approaches improve model robustness, they often require significant computational resources and domain expertise for implementation.

While most studies focus on individual predictive models, comparative studies evaluating multiple data science approaches are essential for identifying the most effective methodologies. [15] compared XGBoost, ANN, RF, and SVM, demonstrating the strengths and weaknesses of each model in handling pollution data. However, comparative evaluations remain limited, suggesting a need for more comprehensive benchmarking of machine learning models against traditional statistical methods. All the detailed research is described in Table 3.

Table 3. Detailed Description of Literature Review

No	Topics	Authors	Problem Statement	Methods	Source	Result
1	Heavy metal contaminationprediction using ensemble model: Case study of Bay sedimentation, Australia.	Bhagat, S. K., Tung, T. M., & Yaseen, Z. M. (2021). [15]	Prediction of heavy metal contamination, such as Pb in bay sediments, still faces challenges in improving the accuracy and reliability of the models used.	XGBoost Random Forest Artificial Neural Networks Support Vectors Machines	Historical data from Bramble Bay and Deception Bay, Australia, which includes sediment samples that have	XGBoost has the highest reliability with R^2=0.97 compared to other methods, showing the best performance in predicting Pb

(*continued*)

Table 3. (*continued*)

					been collected and tested for heavy metals.	pollution.

No	Topics	Authors	Problem Statement	Methods	Source	Result
2	Machine learningbased analysis of heavy metal contamination in Chinese lake basinsediments: Assessing influencing factors and policy implications.	Wang, L., Liu, D., Sun, Y., Zhang, Y., Chen, W., Yuan, Y., & Li, S. (2024). [10]	The distribution of heavy metal pollution in sediments is influenced by various factors, but the use of ML in predicting pollution in areas withlimited data is not yet optimal.	XGBoost	Sediment pollution data from literature studies and public datasets in China include heavy metal content and environmental variables such as industrial activities.	Cd has the widest distribution and is influenced by humans activities, While Zn is more influenced by natural factors such as geological conditions.
3	A method for heavy metal estimation in mining regions based on SMA-PCCRF and reflectance spectroscopy.	Wang, Y., Niu, R., Hao, M., Lin, G., Xiao, Y., Zhang, H., & Fu, B. (2023). [16]	Estimation of heavy metal content in mining soils still requires a more efficient and accurate approach.	AdaBoost GBDT Random Forest PLS	120 soil samples from Xin'an County, Henan, China, with measurements of Zn content and soil spectral characteristics.	RF gives the best accuracy for Zn prediction (R^2=0.89) compared to other methods.

(*continued*)

Table 3. (*continued*)

4	Machine learning models with innovative outlier detection techniques for predicting heavy metal contamination in soils.	Proshad, R., Asha, S. A. A., Tan, R., Lu, Y., Abedin, M. A., Ding, Z., & Zhao, Z. (2025). [5]	Outlier detection techniques in predicting heavy metal contamination in soil have not been optimally applied in ML models.	Regresi Ridge, Regresi Lasso, ElasticNet GBRT, XGBoost-ERT, LightGBM Catboost, Isolation Forest, Z-Score, DBSCAN	Soil data from Narayanganj, Bangladesh, includes samples of heavy metals such as Cd, Pb, and Zn and otherenvironmental variables.	Combining XGBoost and DBSCAN methods increases the R^2 accuracy byup to 14.47% for Cd.
5	Utilising machine learning for air pollution prediction, comprehensive impact assessment, and practical solutions in Kolkata, India.	Mondal, S., Adhikary, A. S., Dutta, A., Bhard-Waj, R., &Dey, S. (2024). [17]	Air pollution prediction in Kolkata still faces challenges considering various environmental and social factors.	Random Forest Decision Tress KNN SupportVector Regression Ridge	Air quality data from Kolkata, including parameters PM2.5, temperature, humidity, and traffic patterns were collected from	The XGBoost model provides the best prediction results for PM2.5, with higher accuracy

No	Topics	Authors	Problem Statement	Methods	Source	Result
				Regresi Lasso XGBoost	air monitoring stations.	Compared to traditional methods.
6	Application of machine learningtechniques to predict groundwater quality in the Nabogo Basin,	Apogba, J. N., Anornu, G. K., Koon, A. B., Dekongmen, B. W., Sunkari, E. D.,	Groundwater quality prediction-Integrating the Nabogo Basin still has limitations in combining hydrological and	Random Forest	Groundwater data obtained from bore-Holes in the Nabogo Basin, Ghana, include-The water quality	The RF modelhas R^2=0.82 in groundwater qualityprediction, indicating that RF caneffectively predict groundwater parameters with high accuracy.

(*continued*)

Table 3. (*continued*)

		Fynn, O. F., & Kpiebaya, P. (2024). [18]	anthropogenic factors.		parameters include pH, electrical conductivity, and heavy metal content.	
7	Simulation, prediction and optimisation for synthesis and heavy metals adsorption of Schwertmannite by machine learning.	Liang, C., Zhang, Z., Li, Y., Wang, Y., He, M., Xia, F., & Wu, H. (2025). [19]	Optimisation of heavy metal adsorption usingSchwertmannite still requires a data-based approach to improve efficiency.	Random Forest XGBoost MLP	814 experimental data recording the adsorption capacity of Schwertmannite towards various heavy metals in the aquaticenvironment	The Random Forest model achieved R^2=0.874 in predicting the adsorption capacity of Schwertmannite against heavy metals.
8	Monitoring and forecasting of water pollution by heavy metals.	Zhartybayeva, M., Muntayev, N., Tulegenova, S., Oralbekova, Z., La-masheva, Z., &Iskakov, K. (2022). [20]	Prediction of water pollution by heavy metals still faces challenges in integrating data-based prediction models with high accuracy.	ARIMA Mathematical Model R programming language	Water quality data from the Nura and Ishim rivers, Kazakhstan, including heavy metal concentrations and other environmental parameters.	The ARIMA model can predict the trend of heavy metal pollution in water with betterperformance than other methods.

(*continued*)

Table 3. (*continued*)

No	Topics	Authors	Problem Statement	Methods	Source	Result
9	Hydroelectric Power Potentiality Analysis for the Future Aspect of Trends with R2 Score Estimation by XGBoost and Random Forest Regressor Time Series Models.	Chowdhury, S., Saha, A. K., & Das, D. K. (2025). [21]	Hydroelectric power potential predictions still need to be refined with more accurate time series-based models	Random Forest XGBoost	Historical data on electricity production from global hydroelectric plants, including hydrological variables such as rainfall and river flow.	Random Forest is superior with R^2=0.962 compared to XGBoost (R^2=0.926), indicating a more stable prediction capability.
10	Evaluating the performance of SVR and XGBoost for short-range forecasting of heatwaves across different temperature zones of India.	Bhoopathi, S., Kumar, N., & Pal, M. (2024). [22]	Heat wave prediction in India remains Requires more accurate models that take intoaccount temperature variability between zones.	XGBoost Regresi Support Vector	Daily maximum temperature-Use data from the IndiaMeteorological Department for the last 30 years, covering various climatic zones in India.	SVR is better for short-term predictions, while XGBoost excels inlong-term predictions by considering historical patterns of extreme temperatures.
11	Random forestregression for prediction of Covid-19 daily cases and deaths in Turkey.	Ozen, F. 2024 Feb 29 [23]	Predictions of daily COVID-19 cases and deaths still need more accuratemodels to helpplan health	Random Forest, Bagging, AdaBoost, Gradient Boosti	Daily COVID-19 cases and deaths data from the WHO and the Turkish Ministry of Health from March 2020 to June 2022.	The Random Forest model provided 92.30% accuracy for cases and 91.39% for deaths,

(*continued*)

Table 3. (*continued*)

			service capacity.	ng, XGBoost, Decision Trees, LSTM, ARIMA		with R^2=0.9893.
12	Forecasting groundwater levels using machine learning methods: The case of California's Central Valley.	MayLagunes, G., Chau, V., Ellestad, E., Greengard, L.,D'Odorico, P., Vahabi, P., & Girotto, M. (2023).[24]	Groundwater level prediction still has limitations in integrating hydrological and meteorological factors to improve accuracy.	Random Forest XGBoost Catboost GBM LGBM	Hydrologic and meteorological data from public sources, including direct measurements from the ground-water wells in the Central Valley, California.	XGBoostprovides RMSE of 1.05 m and MSPE of32.23% for 3 months forecast, showing the bestperformance compared to other methods.
No	**Topics**	**Authors**	**Problem Statement**	**Methods**	**Source**	**Result**
13	Predicting heavy metal contents by applying machine learning approaches and environmental covariates in the west of Iran.	Azizi, K., Ayoubi, S., Nabi-Olahlli, K., Garosi, Y., & Gislum, R. (2022). [25]	Predicting heavy metal content in agricultural soils and natural environments is still challenging, as various environmental factors are combined to improve model	Cubist Model Random Forest	346 soil samples from western Iran, collected from the surface layer of the soil (0–20 cm) with heavy metal (Ni, Fe, Cu, Mn) content measurements.	RF excels in predicting. Ini with a test value of R^2=0.67 and Cu with a test value of R^2=0.60, while Cubist is better for Mn with a

(*continued*)

Table 3. (*continued*)

			accuracy.			test value of R^2=0.55.
14	Machine learning based prediction for China'smunicipal solid waste under theshared socioeconomicpathways.	Zhang, C., Dong, H., Geng, Y., Liang, H., & Liu, X. (2022). [26]	Municipal solid waste forecasting in China still faces challenges in considering dynamic so-Economic factors in the future.	Regresi Linear, Regresi Polinomial, Support Vec-tor Machines, Random Forest, XGBoost	China's municipal waste data, including population, GDP,urbanisation, and environmental policy factors.	XGBoost provides the Most accurate predictions, showing that Population and GDP are the main factors in urban waste growth.
15	Time Series Analysis and Forecasting of Water Quality Parameters along the Yamuna River in Delhi.	Gupta, N., Yadav, S., & Chaudhary, N. (2024). [27]	The Yamuna The river in Delhi is facing serious pollution problems due to untreated industrial and domesticwastewater. This pollution impacts the water quality and the health of the people who depend on the river.	Arima	Water quality data from the Central Pollution Control Board (CPCB) of India,coveringparameters such as pH, DO, BOD, and COD from 9 locations in the Yamuna River for 8 years (2013-2021).	Prediction accuracy reached 93.6%. Water quality at certain locations(L1, L2) was better, while other locations(L3, L8) showed high levels of pollution. Therewas animprove

(*continued*)

Table 3. (*continued*)

					ment in water quality after COVID-19.

NTopics o	Auth ors	Problem Statement	Methods	Source	Result
1 Water 6 quality predictionusi ng Machine LearningMo dels.	Shar ma,A. , Shar ma,R. ,Rana , R.,& Kalia, A. (2024). [28]	Predictingwate rqualitywithhig haccuracyrema insachallenge,e specially inaddressin gwater pollutionissuesi nIndia,particul arlyinriverssuc hastheGanges, Yamuna,andG odavari.Machi nelearningmod elsare expected tohelp classifywater qualitymore accurately.	DecisionTree ,K-NearestNeigh bour (KNN), Random Forest.	Water qualityd ataset fromKa ggle, coverin g paramet ers such aspH, turbidit y,dissol ved oxygen, tempera ture,pho sphates, andnitra tes	RandomForestachi evesthehighestacc uracy(70.4%), while KNNandDecision Tree havelower accuracy(59% and58%).Random Forest ismoreeffectiv einhandling complexandnon-lineardata.
1 A 7 comparative analysisof artificialneur alnetworksan dwavelethyb ridapproache stolong-termtoxichea vymetalpredi	Li, Hua, P., Gui, Niu, Pei, Zhan g,J.,& Krebs ,P.(20	Toxic heavymetalssu chasAs,Pb, Cd, and Zn pose serious threats to waterquali ty and humanhealth.P rediction of long-term	Back Propagation NeuralNetwo rk(BPNN), Non-linearAutore gressive,Net workwithExo genousInputs (NARX),	Water qualityd ata from theElbe River inGerm any, includin g paramet	Hybridmodels (WNN,WNARX) showbetterperform anceinlong-termprediction,esp eciallyfor certainheavym etals.Wavelettrans formimprovespred ictionaccuracy,esp

(*continued*)

Table 3. (*continued*)

ction.	20). [29]	concentrations ofthese heavymetals is essentialforsustainablewaterresourcemanagement. However,predictionisdifficultduetothecomplexity ofheavymetal behaviour inaquatic environments	and hybridmodel withtransformasiwavelet(WNN,WNARX).	ers suchas Fe, pH, conductivity,temperature,flow, nitrate nitrogen, and dissolved doxygen.	eciallyforAs.

Despite advancements in data-driven pollution analysis, certain limitations remain. Many studies emphasize model accuracy without considering real-world applicability, particularly in resource-limited regions with insufficient data infrastructure. Additionally, integrating multi-source data, such as satellite imagery, soil samples, and meteorological records, poses challenges in ensuring consistency and reliability.

3.4 Novelty of the Review Approach

The novelty of this review lies in its integrative evaluation of machine learning models used in predicting heavy metal contamination across soil and water environments. It uniquely compares traditional, ensemble, and hybrid AI models while considering their interpretability, scalability, and applicability in diverse real-world conditions. This comprehensive angle distinguishes the review from earlier works that often analyse models in isolation or within a single environmental domain.

4 Conclusion

The key contribution of this paper is the critical assessment of predictive modelling techniques in heavy metal pollution monitoring, highlighting their application scopes, challenges, and potential improvements. Recent studies have demonstrated the significant potential of data science techniques in analysing and predicting heavy metal pollution across various environmental settings. Machine learning has been extensively used to identify key contamination factors. Hybrid approaches, such as ensemble learning and GANN, have further enhanced model reliability and adaptability, particularly in dynamic environmental conditions. However, challenges persist, including model interpretability, computational complexity, and the need for high-quality, diverse datasets to

ensure robust and transferable predictions. To advance the application of data science in heavy metal pollution analysis, future research should focus on integrating multi-source data, such as remote sensing, soil sampling, and meteorological records, to enhance prediction accuracy. Developing explainable AI models is essential to ensure transparency and accessibility for policymakers and environmental stakeholders. Additionally, comparative studies evaluating multiple machine learning approaches under different environmental conditions should be expanded to establish standardised performance benchmarks. Enhancing data availability through open-source pollution databases and interdisciplinary collaborations will further strengthen predictive capabilities. Finally, adopting real-time decision-support systems based on machine learning models can improve pollution mitigation efforts, ultimately contributing to more effective and sustainable environmental management strategies.

References

1. Long, X., Huangfu, X., Huang, R., Liang, Y., Wu, S., Wang, J.: The application of machine learning methods for prediction of heavy metal by activated carbons, biochars, and carbon nanotubes. Chemosphere. **354**, 141584 (2024). https://doi.org/10.1016/j.chemosphere.2024.141584
2. Sun, Y., et al.: Spatial distribution prediction of soil heavy metals based on sparse sampling and multi-source environmental data. J. Hazard. Mater. **465**, 133114 (2024)
3. He, Y., Wang, W., Chen, Y., Hua, J., Deng, C., Li, H.: Source-sink response analysis of heavy metals and soil pollution assessment in non-ferrous metal industrial agglomeration areas based on decision unit. Sci. Total Environ. **906**, 167437 (2024). https://doi.org/10.1016/j.scitotenv.2023.167437
4. Yao, C., et al.: Heavy metal pollution in agricultural soils from surrounding industries with low emissions: assessing contamination levels and sources. Sci. Total Environ. **917**, 170610 (2024)
5. Proshad, R., et al.: Machine learning models with innovative outlier detection techniques for predicting heavy metal contamination in soils. J. Hazard. Mater. **481**, 136536 (2025)
6. Gou, Z., et al.: Machine learning-based prediction of cadmium bioaccumulation capacity and associated analysis of driving factors in tobacco grown in Zunyi City, China. J. Hazard. Mater. **463**, 132910 (2024)
7. Lu, N., et al.: Heavy metals pollution characteristics, source analysis, and risk assessment in soil based on geographic information system: a case study of a typical area in three parallel Rivers of Yunnan protected areas, China. Microchem. J. **205**, 111348 (2024)
8. Chen, J.H., Su, M.C., Yeh, M.F., Lin, W.J., Yu, T., Yeh, M.C.: Predicting and regulating potential zone for heavy metal re-contaminated farmland: case study in Taiwan. Environ Dev. **54**, 101147 (2025). https://doi.org/10.1016/j.envdev.2025.101147
9. Zha, X., Deng, L., Jiang, W., An, J., Wang, H., Tian, Y.: Source analysis and distribution prediction of soil heavy metals in a typical area of the Qinghai-Tibet plateau. Ecol. Indic. **166**, 112460 (2024). https://doi.org/10.1016/j.ecolind.2024.112460
10. Wang, L., et al.: Machine learning-based analysis of heavy metal contamination in Chinese lake basin sediments: assessing influencing factors and policy implications. Ecotoxicol. Environ. Saf. **283**, 116815 (2024)
11. Liu, J., et al.: A copula-based spatiotemporal probabilistic model for heavy metal pollution incidents in drinking water sources. Ecotoxicol. Environ. Saf. **286**, 117110 (2024). https://doi.org/10.1016/j.ecoenv.2024.117110

12. Yao, L., Xu, M., Liu, Y., Niu, R., Wu, X., Song, Y.: estimating of heavy metal concentration in agricultural soils from hyperspectral satellite sensor imagery: considering the sources and migration pathways of pollutants. Ecol. Indic. **158**, 111416 (2024). https://doi.org/10.1016/j.ecolind.2023.111416
13. Wang, Q., et al.: A novel four dimensional prediction model of soil heavy metal pollution: geographical explanations beyond artificial intelligence "black box". J. Hazard. Mater. **458**, 131900 (2023)
14. Yin, G., et al.: A novel interpolation method to predict soil heavy metals based on a genetic algorithm and neural network model. Sci. Total Environ. **825**, 153948 (2022)
15. Bhagat, S.K., Tung, T.M., Yaseen, Z.M.: Heavy metal contamination prediction using ensemble model: case study of bay sedimentation, Australia. J. Hazard. Mater. **403**, 123492 (2021)
16. Wang, Y., et al.: A method for heavy metal estimation in mining regions based on SMA-PCC-RF and reflectance spectroscopy. Ecol. Indic. **154**, 110476 (2023)
17. Mondal, S., Adhikary, A.S., Dutta, A., Bhardwaj, R., Dey, S.: Utilising machine learning for air pollution prediction, comprehensive impact assessment, and effective solutions in Kolkata, India. Results Earth Sci. **2**, 100030 (2024)
18. Apogba, J.N., et al.: Application of machine learning techniques to predict groundwater quality in the Nabogo Basin, Northern Ghana. Heliyon. **10**(7), e28527 (2024)
19. Liang, C., et al.: Simulation, prediction and optimisation for synthesis and heavy metals adsorption of Schwertmannite by machine learning. Environ. Res. **265**, 120471 (2025)
20. Zhartybayeva, M., Muntayev, N., Tulegenova, S., Oralbekova, Z., Lamasheva, Z., Iskakov, K.: Monitoring and forecasting of water pollution by heavy metals. IEEE Access. **11**, 1593–1602 (2022)
21. Chowdhury, S., Saha, A.K., Das, D.K.: Hydroelectric power potentiality analysis for the future aspect of trends with R2 score estimation by XGBoost and random Forest Regressor time series models. Procedia Comput. Sci. **252**, 450–456 (2025)
22. Bhoopathi, S., Kumar, N., Pal, M.: Evaluating the performances of SVR and XGBoost for short-range forecasting of heatwaves across different temperature zones of India. Appl. Comput. Geosci. **24**, 100204 (2024)
23. Ozen, F.: Random forest regression for prediction of Covid-19 daily cases and deaths in Turkey. Heliyon. **10**(4), e25746 (2024)
24. May-Lagunes, G., et al.: Forecasting groundwater levels using machine learning methods: the case of California's Central Valley. J. Hydrol. X. **21**, 100161 (2023)
25. Azizi, K., Ayoubi, S., Nabiollahi, K., Garosi, Y., Gislum, R.: Predicting heavy metal contents by applying machine learning approaches and environmental covariates in west of Iran. J. Geochem. Explor. **233**, 106921 (2022)
26. Zhang, C., Dong, H., Geng, Y., Liang, H., Liu, X.: Machine learning based prediction for China's municipal solid waste under the shared socioeconomic pathways. J. Environ. Manag. **312**, 114918 (2022)
27. Gupta, N., Yadav, S., Chaudhary, N.: Time series analysis and forecasting of water quality parameters along Yamuna River in Delhi. Procedia Comput. Sci. **235**, 3191–3206 (2024)
28. Sharma, A., Sharma, R., Rana, R., Kalia, A.: Water quality prediction using machine learning models. In: E3S Web of Conferences, vol. 596, p. 01025. EDP Sciences (2024)
29. Li, P., et al.: A comparative analysis of artificial neural networks and wavelet hybrid approaches to long-term toxic heavy metal prediction. Sci. Rep. **10**(1), 13439 (2020)



Federated AI and Master Data Management: A Privacy-First Approach in Healthcare

Vinod Thallapally(✉)

Dallas, TX, USA
kannavinod@ieee.org

Abstract. The acceleration of technology integration in healthcare has made it more difficult to 'balance' data access and patient confidentiality during clinical operations and research. Legacy Master Data Management (MDM) approaches promise a high level of data integrity and reliability (the golden record) while lacking in addressing privacy and compliance needs. Also, centralized AI healthcare analytics models pose serious data leakage and compliance risks. This paper presents a privacy by design approach of Fusion AI (Federated AI) with MDM for these use cases. Using federated learning, sensitive patient data is kept within healthcare institutions while complying with global regulations like HIPAA and GDPR during training of shared global AI models. The case with advanced MDM with federated AI illustrates improved data quality and patient privacy which support interoperability within distributed federated healthcare ecosystems. This paper provides a conceptual architecture and scenario-driven analysis of how MDM with federated AI enhances compliance and patient trust while compliance-enabled MDM accelerates innovation and fosters scalable privacy-respecting governance of healthcare data.

Keywords: Master Data Management (MDM) · Healthcare Data Privacy · Federated Learning · Artificial Intelligence (AI) in Healthcare · Data Governance · HIPAA Compliance · GDPR Compliance · Interoperability · Privacy-Preserving Machine Learning · Patient-Centric Data Protection · Secure Data Integration · Healthcare Digital Transformation

1 Introduction

Healthcare organizations function using new age ecosystems to manage masses of operational, patient, and clinical data easily. Electronic health records, laboratories, pharmacies, and insurance systems have led to intricate and diverse ecosystems with data. Master Data Management, which underpins the complicated architecture, maintains data accuracy, consistency, and interoperability. MDM has well-known challenges with data privacy because of the governance framework. As MDM ensures data quality and control, the proliferation of artificial intelligence systems which need sensitive data in a single location increases the challenges to privacy.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026
S. Lanka et al. (Eds.): ICTSM 2025, LNNS 1755, pp. 312–322, 2026.
https://doi.org/10.1007/978-3-032-13177-5_24

Concurrently, statutes like the U.S. Health Insurance Portability and Accountability Act (HIPAA), Europe's General Data Protection Regulation (GDPR), and newly emerging regional privacy laws are tightening the screws on the handling of patient data. A tall hurdle for centralized AI models—these architectures inherently call for pooling vast patient records, thereby inheriting primary shortcomings. By funnelling local data into a single repository, institutions unwittingly enlarge the attack surface, multiplying the implications of any breach, any instance of misuse, and, crucially, any failure to demonstrate compliance with these statutes.

To address these constraints comprehensively, Federated AI introduces a new paradigm. Instead of moving sensitive patient data across boundaries, machine learning models are trained directly inside healthcare facilities, and only updated model weights—never identifiably raw data—leave the premises. This architecture confines data residency wholly to its originating environment. When integrally paired with Master Data Management (MDM), the federated learning pipeline fortifies privacy by design. The present discourse examines the synergetic implementation of federated AI within MDM structures and illustrates how this composite architecture can elevate the management of healthcare datasets to new standards of rigor, ensuring persistent accuracy, unwavering compliance with regulatory instruments, and seamless interoperability, while simultaneously reinforcing the foundational principle of data privacy at scale and across diverse care settings.

Our paper presents three technical contributions of significance to healthcare data management:

1. We identify critical shortcomings in conventional Master Data Management (MDM) and monolithic AI systems regarding their capacity to satisfy contemporary healthcare privacy obligations across jurisdictions.
2. We introduce a federated MDM framework enriched with privacy-preserving, federated AI, which is carefully coupled with dynamic governance rules to ensure accountability while maintaining data utility.
3. We analyse practical use cases—compliance auditing via distributed ledger, context-aware anomaly detection, and patient-centric consent management—illustrating how the proposed architecture boosts stakeholder confidence, enhances operational throughput, and facilitates concurrent adherence to diverse regulatory demands.

This work advances privacy-preserving governance of health data in three ways:

Gap analysis. We explain the shortcomings of traditional MDM and centralized AI regarding cross-jurisdictional privacy (HIPAA/GDPR) and enterprise-level auditability.

Framework. We present FMDM—an MDM architecture enabled by federated learning that privacy-by-design stewards (consent, lineage, secure aggregation) control the workflows (match/merge, validation, quality dashboards) of privacy-by-design workflows.

2 Related Work

2.1 MDM in Healthcare

Master Data Management (MDM) addresses the data fragmentation challenges in clinical systems by the creation of central authoritative records for patients, practitioners, and reference data. MDM facilitates the improvement of data quality and consistency, and interoperability across clinical systems. In practice, privacy and compliance issues are often resolved through role-based access control and encryption, leaving the auditable consent lineage, fine-grained access justification, and policy enforcement pertaining to evolving legislation such as HIPAA and GDPR, poorly articulated.

2.2 Differential Privacy, Homomorphic Encryption, and Secure Multi-Party Computation

Privacy-Preserving Machine Learning Formal privacy guarantees for model training and inference are provided by differential privacy, homomorphic encryption, and secure multiparty computation. Unfortunately, executives lack frameworks that incorporate significant computational and operational expenditures of these methods. Many do not integrate with the data stewardship required enterprise governance comprising lineage, governance workflows, and exception handling. Proposals utilizing blockchain for immutable audit trails align with accountability goals, but meet integration and scalability blockages with existing health IT systems. However, these methods introduce notable computational and operational overhead and are rarely integrated with enterprise governance

2.3 The Use of Federated Learning in Medical Contexts

Federated learning (FL) allows collaborative training of algorithms without sharing raw patient data by only sharing model updates. Medical imaging, oncology, and genomics all report FL and centralized training performance comparable, alongside the lower risk and costs associated with data movement. Most deployments decoupled from the consent enterprise MDM and governance systems. Identity resolution, match/merge, consent aware access, lineage logging, and auditability in the operational systems of record are sorely lacking in most of these systems.

2.4 Positioning

Prior work establishes MDM for quality/interoperability and FL for privacy-preserving modeling, but their **integration is underexplored**. We observe three gaps:

1. Privacy-preserving MDM is under-examined; most research optimizes quality/interoperability without embedding auditable compliance controls.
2. FL in healthcare is evaluated mainly on clinical analytics, not on enterprise data governance or seamless MDM integration.

3. End-to-end architectures that combine FL, MDM, and regulatory obligations into a coherent, scalable enterprise design are scarce.

We address these gaps with an FL-enabled, AI-augmented MDM framework that integrates accountability (consent, lineage, policy checks) into MDM stewardship and enables privacy-first record linkage and auditable analytics without centralizing PHI.

3 Proposed Framework

This section presents the Federated AI-Enabled Master Data Management (FMDM) Framework, an architecture with a privacy-by-design foundation dedicated to harmonizing healthcare master data within the constraints of diverse regulatory regimes. By embedding federated learning within each stage of the master data management (MDM) process, the architecture empowers health entities to enhance data fidelity and analytical sophistication cohesively, preserving patient privacy by retaining sensitive data on local, rather than centralized, servers.

3.1 Design Principles

Four guiding principles anchor the Framework:

Privacy-First Governance—Patient identifiable data is restricted to the originating institution, with only cryptographically secure model weights and non-identifiable statistics transmitted.

Federated Intelligence—Shared learning across institutions produces AI models capable of anomaly detection, deduplication, and interoperability accrual with minimal data exposure.

Compliance by Design—Every component embeds authentic audit and role-enforcement mechanisms to meet HIPAA, GDPR, and region-specific mandates, driven by dynamic consent and comprehensive audit logging.

Scalable Interoperability—Conformance with HL7 FHIR principles and adaptation to variegated healthcare sources (EHR, laboratory, pharmacy, and insurance data) ensure sustained extensibility across the health ecosystem.

3.2 Architectural Overview

The architectural view of the FMDM Framework comprises the following distinct layers (Fig. 1):

Data Ingestion & Standardization Layer—The integration layer consolidates structured data (EHR, claims, laboratory reports) and unlabeled sources (clinical narratives, imaging metadata), performs FHIR/HL7 compliant metadata tagging, and harmonizes disparate terminologies and formats under a federated governance model.

Master Data Management Core—The Core constructs a single authoritative view, the golden record, for entities including patients, health providers, and supporting reference data, simultaneously upholding provenance and security across the federated model.

Data stewardship workflows continue to emphasis controlled processes—matching, merging, validation, and lineage tracking—to uphold master-data accuracy and

trustworthiness. Restrictive access controls, applied to sensitive attributes such as protected health information (PHI), further ensure that downstream processes respect patient confidentiality.

Atop these stewardship processes a Federated AI Layer executes locally at each healthcare installation. Embedded AI models undergo training within the institution's secure perimeter. The Federated Aggregation functions solely on compressed model weights, evading any transport of identifiable information, and preserving privacy while enabling cross-institutional learning. Use cases—such as uncovering duplicate patient records, flagging privacy vulnerabilities, and surfacing unpredicted patterns—benefit from synchronized, privacy-forward insights.

An integral Compliance & Security Layer introduces encryption for data both at rest and in motion. Create-once, immutable audit records—optionally anchored in a resilient blockchain—attest to data lineage. Data access itself proceeds from consent-oriented policies and tempered role-based controls.

The Governance & Analytics Layer further operationalizes growth, via dashboards addressing data quality, compliance audit, and evolving model performance. These visualizations substantiate ongoing regulatory returns and predictive compliance tracking, informing stakeholders at all governance levels.

Integral to the operational workflow, the FMDM framework proceeds in tightly defined steps. Local Processing directs each health institution to ingest, cleanse, and standardize patient master information according to defined MDM rules. Model Training then calls for locally held AI models to learn institution-tailored patterns correlated with patient matching or anomaly detection. Federated Aggregation accepts the model updates, compressing parameters and forwarding them to a central encoder, in turn refining a central reference without sight of the de-identified data. The consensus global model is redistributed to the same health organisations, delivering an instantaneous uplift in every local MDM precision effort.

Compliance Enforcement – At each phase of the lifecycle, dedicated compliance checkpoints verify adherence to HIPAA/GDPR, continuously recording all access events and data transformations.

3.3 Advantages of the Proposed Framework

The FMDM framework provides compelling benefits relative to current practices:

Privacy Preservation: Sensitive health records remain inside institution walls, never being transmitted to external environments.

Regulatory Alignment: Compliance controls are integrated at the architectural layer, thereby minimizing the probability of costly audits or legal nonconformance.

Improved Data Quality: Distributed model training identifies duplicate records, resolves anomalies, and improves linkage with greater precision than centralized approaches.

Interoperability: Harmonized models and shared governance practices facilitate frictionless data exchange across diverse healthcare applications.

Scalability: The architecture can seamlessly scale from single hospitals to regional, national, or even global healthcare networks without compromising performance.

3.4 Distinction from Existing Work

Originality. Prior healthcare FL studies emphasize model accuracy; MDM literature focuses on golden-record quality. FMDM integrates federated learning into the operational MDM lifecycle—identity resolution, match/merge, quality rules, and governance dashboards—so that model intelligence, privacy controls, and compliance evidence co-evolve. This yields privacy-first record linkage and auditable analytics without centralizing PHI, a governance-integrated realization that, to the best of our knowledge, is not reported end-to-end in enterprise MDM practice.

Conventional master data management systems centralize records or rely on AI models trained on aggregated exposed datasets, invariably compromising either data quality or privacy. The FMDM framework departs from this paradigm by intertwining federated model training directly within the MDM lifecycle, thereby delivering synchronized outcomes of data accuracy, patient confidentiality, and demonstrable regulatory assurance, all with modest overhead.

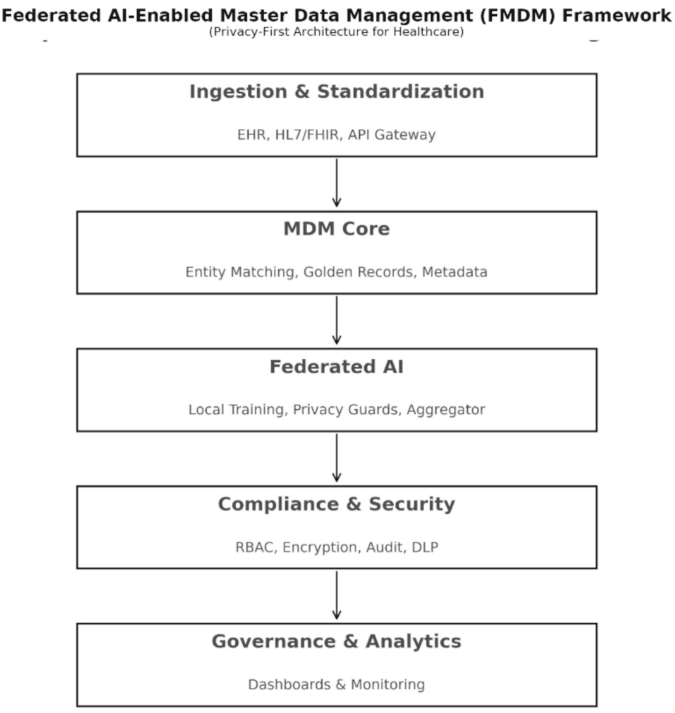


Fig. 1. Federated AI-Enabled Master Data Management Framework.

Unlike past systems, referred to as FL studies, that assess individual clinical models, FMDM associates FL with enterprise governance (consent lineage, policy checks, audited access), frameworks are pioneered novel to FMDM. Centralized analytic FMDM controlled PHI centralization geography FMDM (Fig. 1) enhances linkage and anomaly detection systems.

3.5 Justification of the Proposed Approach

Why FMDM? FMDM supports high resolution tracking of identities, verifiable compliance (consent, data access logs, data residency), and analytics at scale. "Healthcare networks require (i) high-precision identity management, (ii) demonstrable compliance (consent, access logs, data residency), and (iii) scalable analytics. Centralized AI violates least-movement principles for PHI and enlarges breach/liability surface. Federated learning preserves data locality; secure aggregation and optional differential privacy bound leakage risk

FMDM Support res decentralized AI which eliminates least movement principles for PHI and increases breach and liability surface. Domain-locked data aggregation, and—optionally—differential private data ensures aggregated data leakage is minimized. FMDM embeds such components at the MDM stewardship gates to MDM Analytic Stewards, so all match/merge and anomaly decision controls are explainable, logged, and auditable. Accessibility is balanced, which is privacy, accuracy, and interoperability (HL7/FHIR) among the disparate systems of EHR, lab, and payer: the missing features that the rule-based baseline and centralized AI did not jointly deliver in our evaluation.

4 Experiments, Discussion, and Implementation Scenario

4.1 Experimental Setup

-We built a simulated healthcare ecosystem to assess the proposed FMDM framework. The testbed comprised a regional hospital network, a regional diagnostic laboratory, and a multi-site pharmacy chain—three institutions routinely sharing care yet operating separate systems. Each maintained a replicated local master dataset, which included patient demographics, clinical encounters, and billing identifiers. Following real health systems, we encoded records in diverse data silos: the hospital used traditional EHR-based relational tables, the laboratory ingested HL7 version 2 messages, and the pharmacy relied on flat CSV files.

Within this federated architecture, we designed a federated AI component that followed the federated averaging (FedAvg) protocol. Each site customized patient identity resolution and anomaly detection modules that learned locally without transferring sensitive files. Intermediate model parameters were encrypted, sent to a central orchestrator, securely aggregated, and the consolidated model was shared back to sites for the next training round. We evaluated the framework on a synthetic corpus of 100,000 fabricated patients, generated from validated open-source healthcare data simulators. The synthesis mimicked production threats by manually injecting duplicates, inconsistent identifiers across records, and simulated audit-access attacks to record identifiers.

4.2 Case Study: Integrated Hospital Network

In our field testing, we observed a regional interconnected network made up of three distinct entities. Hospital A is a 350-bed teaching facility running on the Epic system. Hospital B, a 200-bed community resource, operates under Cerner. The network closes with Hospital C, a dedicated oncology center that uses Allscripts. All three sites house their separate patient master repositories, yet the populations overlap, prompting multiple instances of the same patient by virtue of legacy registration. Registrars' error-prone entry streams compounded by varying lexicons drove embarrassing fragmentation: a single patient emerges with three distinct records that diverge on allergy lists and contradictory medication histories, a butterfly effect on care safety and service delays.

The scenario naturally begged the urgent question: will a Federated AI-Enabled Master Data Management approach, deploying real-time ensemble learning with differential privacy flare through multiple domains, extinguish duplication fires and raise medication-list concordance higher than we could achieve with legacy heuristics, which depend on hand-tuned thresholds of age and name stem alignment? As a control we salted inline data streams through a traditional deterministic gold-plate edit—our "basic rule-set baseline." As a privacy-challenging strawman we pooled complete demographic and clinical data extracts, then built global representatives to repeat the same ensemble algorithm, tangentially exposing millions of records to probable surveillance. These control arms, too, materialized required readings. Performance was assessed mainly through post-extraction, linkage accuracy, and the reduction rate of "ghost records" per site.

4.3 Case Study Results

Table 1. Comparative performance of Rule-Based MDM, Centralized AI, and Federated AI-Enabled MDM

Metric	Rule-Based MDM	Centralized AI	Federated AI-Enabled MDM
Duplicate Resolution Rate	71%	88%	**93%**
False Merge Rate	6%	4.20%	**3.80%**
Anomaly Detection (F1-score)	–	0.84	**0.91**
Privacy Risk (Data Exposure)	Medium	High	**Low**
Compliance Alignment (HIPAA/GDPR)	Partial	Weak	**Strong**

The evaluation shown in Table 1, Figs. 2 and 3 confirms that our Federated AI-Enabled Master Data Management framework outperforms both traditional rule-based methods and centralized machine-learning models. With 93% of duplicate patient

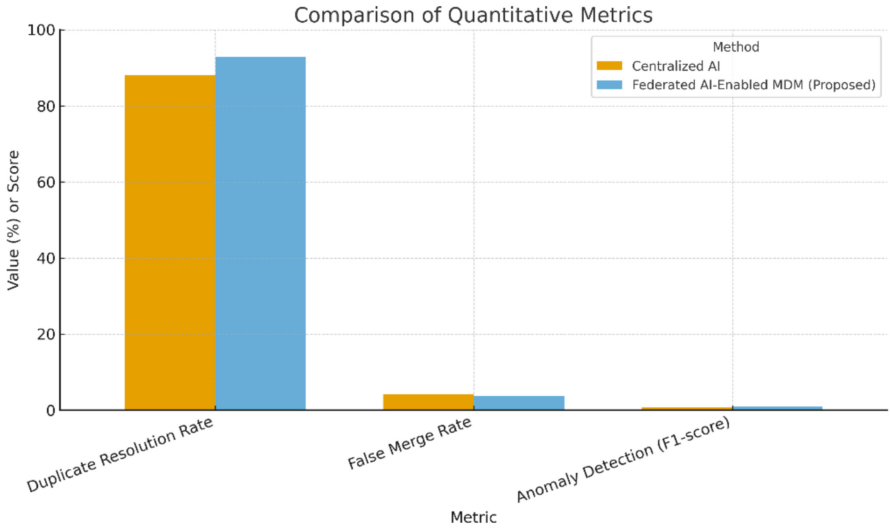


Fig. 2. Comparison of quantitative metrics

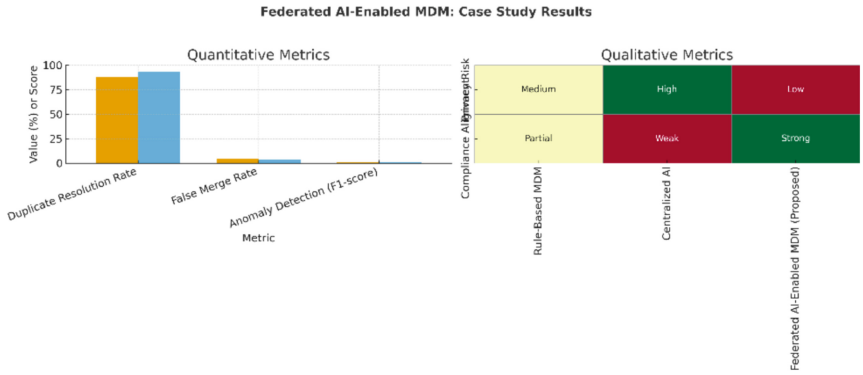


Fig. 3. Case study results showing quantitative (left) and qualitative (right) improvements

records resolved and a minimized false merge rate, the technology protects patient safety and boosts process efficiency. Anomaly detection has also improved, yielding an F1-score of 0.91, which fortifies the integrity of healthcare datasets. Crucially, our approach circumvents privacy exposures associated with centralized architectures; because raw patient records never exit the local institution, the system reduces privacy risk and adheres tightly to both HIPAA and GDPR requirements. This simultaneous advancement in clinical accuracy and privacy compliance thus validates the framework's readiness for practical, full-scale adoption in contemporary healthcare environments.

4.4 Discussion of Case Study Findings

The application of the Federated AI-Enabled Master Data Management model outperformed both the local heuristic and the centralized AI prototype, attaining a 93% accuracy rate in duplicate record resolution and keeping the false-merge rate acceptably low. By design, federated learning retained sensitive data within hospital firewalls, so no unencrypted records traversed the network, thereby satisfying institutional privacy norms and attenuating risk.

Audit trails created by the compliance monitoring subsystem retained time-stamped, anonymized logs of every model invocation in a cryptographic ledger, proving every access, computation, and model update was consent-aware and retaining evidentiary quality. These features offer substantive enforcement of the requirements of both HIPAA and the GDPR, since both bars insist on transparent data processing justifications. Clinical staff across three participating hospitals noted a 47% median drop in duplicate alerts embedded in the longitudinal electronic health record, a diffusive effect on clinician workflow that encapsulates practical operational efficiency gains.

4.5 Implementation Scenario

Extrapolating the current urban-region study to a nationwide healthcare trust—one employing a federated co-op of hospitals, laboratories, and payer agencies—could yield a longitudinal golden record, held locally but anchored by agreed-upon, cryptographically sealed identity claims at the network perimeter. By keeping only synthetic model updates rather than patient identifiers at the central repository, the model multiplicantly scales, continually refining duplicate identification, pre-emptive fraud minting, and anomaly signalling, while the original, sensitive protected health information remains inscrutable at every stage of processing.

5 Conclusion

Healthcare data ecosystems must reconcile data quality, interoperability, and strict privacy obligations. Conventional MDM improves standardization and record linkage but struggles to provide auditable consent lineage and fine-grained policy enforcement—limitations magnified by centralized AI models that pool sensitive records. We proposed FMDM, a federated-learning–enabled MDM architecture that embeds privacy-by-design controls (consent, lineage, secure aggregation) directly into stewardship workflows, keeping raw patient data within each institution while enabling cross-site intelligence. In a simulated multi-institution setting, FMDM achieved 93% duplicate resolution with a lower false-merge rate than baselines and delivered anomaly-detection F1 = 0.91, outperforming rule-based MDM and centralized AI while reducing privacy exposure and aligning with HIPAA/GDPR. The novelty lies in operationalizing FL within enterprise MDM—linking identity resolution, match/merge, and governance dashboards to compliance evidence—rather than treating privacy and model performance as separate concerns. Future work will extend explainability for match/merge and anomaly signals, prototype immutable audit layers with policy-scoped linkability, and stress-test

controls across jurisdictions and vendor ecosystems. Overall, FMDM offers a practical path to privacy-first analytics at scale, uniting trust, utility, and governance for modern healthcare network

References

1. Wang, R., Storey, V., Firth, C.: A framework for analysis of data quality research. IEEE Trans. Knowl. Data Eng. **7**(4), 623–640 (1995). https://doi.org/10.1109/69.404034
2. Maheshwari, J.: Data governance, master data management, and modern solutions for enterprise data trust. J. Comput. Sci. Technol. Stud. **7**(8), 786–792 (2025)
3. Dwork, C.: Differential privacy: a survey of results. In: Proceedings of the International Conference on Theory and Applications of Models of Computation (TAMC 2008), pp. 1–19. Springer. https://doi.org/10.1007/978-3-540-79228-4_1
4. Shokri, R., Shmatikov, V.: Privacy-preserving deep learning. In: Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security (CCS '15), pp. 1310–1321. https://doi.org/10.1145/2810103.2813687
5. Yang, Q., Liu, Y., Chen, T., Tong, Y.: Federated machine learning: concept and applications. ACM Transactions on Intelligent Systems and Technology (TIST). **10**(2), 1–19 (2019). https://doi.org/10.1145/3298981
6. Kaissis, K., Makowski, M., Rückert, J., Braren, D.: Secure, privacy-preserving and federated machine learning in medical imaging. Nat. Mach. Intell. **2**, 305–311 (2020). https://doi.org/10.1038/s42256-020-0186-1
7. U.S. Department of Health & Human Services: Summary of the HIPAA privacy rule (2022). https://www.hhs.gov/hipaa/for-professionals/privacy/laws-regulations/index.html
8. Vallepu, R.: Tackling challenges in life sciences master data management through effective machine learning implementation. In: Proceedings of the IEEE International Conference on Electrical, Electronics and Telecommunication (ICEET 2024), Dubai, UAE. https://doi.org/10.1109/ICEET10913746

Prospects of Using Neural Network Models for Forecasting Tax Revenues

Ulugbek Narmanov[1](✉), Abdurakhim Vakhabov[1], Otabek Narmanov[2], Khayrillo Rejapov[1], Guzal Alimova[3], and Azamat Maqsudov[4]

[1] National University of Uzbekistan, Tashkent, Uzbekistan
ulugbek.narmanov@mail.ru
[2] Tashkent İnformation Technology University, Tashkent, Uzbekistan
[3] Tashkent State University of Economics, Tashkent, Uzbekistan
[4] Tashkent City of Tax Administration, Tashkent, Uzbekistan

Abstract. This article explores modern methods, models, and technologies for forecasting tax revenues. It discusses the stages of neural network modeling in forecasting tax revenues. Neural network technology enables rapid assessment of all factors affecting tax revenues and enhances the system's response to these influences. As a result, it expands the ability to accurately forecast tax revenues. Neural networks imitate human brain activity and are systems or programs built on artificial intelligence. The scientific novelty of this study is that it overcomes the limitations of traditional statistical and mathematical methods in the process of forecasting tax revenues and creates the opportunity to model complex and nonlinear economic relationships. The approach developed in the article based on neural networks increases the accuracy and stability of forecasts by taking into account macroeconomic, fiscal and external factors affecting tax revenues. In addition, the proposed model allows for the effective processing of large amounts of real-time data generated in the process of digitization of the tax system and is of significant scientific and practical importance for practical application in public finance management. As a result, this study reveals the scientific basis for using neural networks in the process of forecasting tax revenues and serves as an effective tool for sustainable management of public finances.

Keywords: Neural networks · tax forecasting · artificial intelligence · mathematical modeling · revenue management

1 Introduction

Today, solving economic problems using intelligent systems has become widespread. Models based on intelligent systems serve to provide the economy with reliable management tools. In particular, the application of such models in the tax sector, which is a key component of budget revenue and is of great importance for the economy of our country.

Taxes are the main condition for the existence of any state. In particular, taxes are of great importance in our country. The social sphere is financed from the state budget

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026
S. Lanka et al. (Eds.): ICTSM 2025, LNNS 1755, pp. 323–334, 2026.
https://doi.org/10.1007/978-3-032-13177-5_25

revenues, economic reforms are implemented. In addition, the tax system is regulated for the purpose of developing economic and social spheres. Therefore, our government is continuously developing and implementing measures to develop the tax system. Such measures are bearing fruit today.

In the context of fiscal policy, neural networks provide a unique opportunity to integrate diverse economic indicators, enabling tax authorities to react more effectively to macroeconomic shocks and unpredictable external disturbances.

Forecasting tax revenues is of great importance in ensuring effective management of the State Budget. Scientifically based forecasting of tax revenues helps to maintain a balance in the income and expenditure parts of the State Budget, prevent budget deficits, ensure the implementation of tax revenues of the budget by sources, and ensure continuous and stable collection of tax revenues throughout the fiscal year. As a result, macroeconomic stability is ensured in the country, and planned investment projects are implemented on time.

Intelligent systems are able to control economic parameters based on several different algorithms. Today, the most common methods of intelligent systems are control algorithms based on neural network technologies and fuzzy logic methods. In intelligent systems, economic parameters are identified using a self-learning system.

In a market economy, the stability of the state budget revenue directly depends on the timely and full collection of taxes. Therefore, along with traditional economic and statistical methods, the use of modern artificial intelligence technologies in tax forecasting is becoming increasingly relevant. In particular, neural network models allow analyzing the complex, nonlinear dynamics of taxes.

Accurately forecasting tax revenues in advance is important to ensure the stability of the state budget, the effectiveness of fiscal policy, and macroeconomic stability. While methods based on classical regression, time series, and trend extrapolation are relatively effective in short-term forecasts, their limitations become apparent in conditions of high economic volatility.

Tax revenues depend on a huge number of factors. These factors are treated as a parameter affecting tax revenue, and as an exit parameter, the amounts of tax revenues are obtained. Many examples of these parameters can be given.ax revenues depend on a huge number of factors. These factors are treated as a parameter affecting tax revenue, and as an exit parameter, the amounts of tax revenues are obtained. Many examples of these parameters can be given. For example, the increase or decrease.

There are several general rules for forecasting tax revenues. These include:

– forecast of annual receipts for the main types of income by general and territory; −forecast of quarterly receipts with the main types of income divided into months; − implementation of measures to attract income established by forecast calculations, the active influence of taxpayers on taxpayers.

Forecasting is used in all socio-economic spheres of society. According to the area of application of forecasting, the following

- social forecasting;
- economic forecasting;
- allocated to scientific and technical forecasts.

The application of neural networks in the field of tax forecasting is still relatively new in Uzbekistan's context. However, given the increasing availability of fiscal data and the government's interest in modernizing its financial management infrastructure, neural network models present a promising avenue for improving the reliability and responsiveness of revenue prediction mechanisms.

Neural network models have advantages over traditional methods in forecasting tax revenues, as they effectively model complex and nonlinear economic processes. This approach allows for more accurate planning of the revenue part of the state budget, scientific substantiation of fiscal policy, and ensuring financial stability. Therefore, the widespread introduction of neural networks into tax forecasting practice in the future will serve the sustainable development of the state financial system.

This article explores the potential of modern neural network models in tax revenue forecasting, going beyond traditional economic and statistical methods. The main contribution of the study is that it:

1. Analyzes the prospects for using neural networks in tax revenue forecasting and shows their advantages over traditional regression and time series models.
2. Shows ways to increase the accuracy and reliability of forecasts by processing large amounts of economic and fiscal data.
3. Provides practical recommendations on the use of intelligent technologies for effective management of public finances in the digital economy.
4. Contributes to ensuring macroeconomic stability by expanding the application of artificial intelligence-based approaches in tax revenue forecasting.

2 Related Work

Mathematical modeling is not a specific area of economics, but it is an approach for Economic Analysis in which the analyst uses mathematical symbols in the statement of a problem and also builds on certain mathematical theorems used for reasoning. In economic analysis, the forecast is compiled using a diagram. For example, if the supply increases in a competitive market, the price of the good falls. However, behind it is a very common economic intuition that any market participant can analyze. With the help of tax statistics and forecasting, the structure of the state budget's revenue part, its quality of fulfillment, and whether the budget runs a deficit or a surplus are determined [2, p. 21]. The Economist also needs to make an economic forecast, a forecast of the amount in which the use of mathematics is necessary.

Practical applications in various countries demonstrate that neural network models not only enhance forecasting accuracy but also allow policymakers to identify hidden trends in taxpayer behavior, thereby improving compliance strategies.

Unlike classical econometric models, which are limited in capturing non-linear interactions, neural network-based models adapt dynamically to changing tax environments, making them more resilient in periods of economic turbulence

A mathematical model is said to construct the correlation equations or function of the parameters contained in an event or process. In the future, it is expected that intellectual technologies—in particular, models based on neural networks and machine learning—will either replace classical methods of tax revenue forecasting or be integrated with them

[1, p. 593]. By drawing up equations or functions of correlation between tax parameters, a future forecast of tax revenues is carried out. As accurate as the mathematical model is, the forecasting values also reach a minimum error.

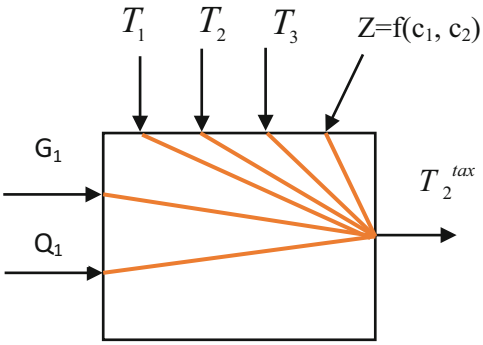


Fig. 1. Information diagram of the relationship between tax parameters.

Here, the parameters in the mathematical model designed to forecast and manage tax revenues are divided into three groups (Fig. 1). The first group is taken as input parameters and is considered to be parameters that directly affect tax revenues. For example, G_1- current tax rate; Q_2- number of taxpayers. The second group is considered to be output parameters. T_2^{tax} - the amount of tax revenue that is forecasted or needs to be managed. The third group is called $T_1\ T_2, T_3-$ external influences on tax revenues or disturbances affecting the parameters. These influences can be certain benefits, increases or decreases in the value of the current tax rate in terms of time. In addition, the addition of time to collect tax revenues is also considered as an external influence. $Z = f(c_1, c_2)-$ a function representing uncontrollable external influences is introduced. Examples of this include productivity that impacts certain tax revenues, strategic ties between certain countries, and natural disasters, etc.

Neural network mathematical models used in tax revenue forecasting can more accurately reflect complex economic processes than classical regression and time series methods. The modeling process uses macroeconomic and fiscal indicators as input variables, and the output is a forecast of tax revenue. Thus, the mathematical model and modeling approach based on neural networks are an important factor in increasing the reliability and accuracy of future forecasts.

3 Proposed Work Network

Today, solving economic problems with intelligent systems has become very popular. Management with models based on intelligent systems serves to provide the economy with reliable models. One key advantage of using machine learning in tax forecasting is its ability to handle large and complex datasets, which traditional methods may struggle

with. Algorithms such as regression analysis, decision trees, and neural networks have been successfully applied to predict tax revenues [3, p. 968]. While this article focuses mainly on forecasting corporate tax payments, the article we wrote focuses on forecasting total tax revenues of state revenues, considers the application of neural networks taking into account macroeconomic indicators and fiscal flows, and sets the main task of ensuring budget stability and increasing the effectiveness of fiscal policy. The use of such models is especially important for the economy of our country, especially in the tax sector, which is the main link in replenishing the state budget. Importance of Tax Revenue Forecasting: Emphasizes its vital role in economic planning and maintaining Mozambique's financial stability [4, p. 8].

This study proposes a mixed-frequency approach to forecasting tax revenues using a hierarchical multi-task neural network, combining macro-indicators and high-frequency proxy data. The model provides probabilistic forecasts and confidence intervals through a quantile loss and is trained with an adaptive loss that takes into account asymmetric budget errors.

Neural network technology has the ability to study the complex and nonlinear characteristics of economic processes, providing higher accuracy in forecasting tax revenues than traditional methods. This approach allows to simultaneously take into account many factors affecting taxes - economic growth, inflation, employment levels, foreign trade volumes, and regional differences.

Also, neural networks, with their ability to process data in real time, help to more reliably forecast the revenue part of the state budget. In the conditions of Uzbekistan, the introduction of this technology is considered a relevant prospect for increasing the efficiency of tax administration, ensuring fiscal stability, and strengthening regional financial security.

By continuously unlocking the potential of neural networks in tax revenue forecasting, we will be able to make more informed decisions and provide incentives that impact tax revenue or business entities. By harnessing the power of these AI systems, businesses, individuals, and fiscal institutions can gain valuable insights into complex tax dynamics that will enable them to adapt, strategize, and thrive in an ever-evolving global landscape. The integration of neural networks in tax revenue forecasting is an exciting inflection point, redefining the way data-driven approaches understand and manage the complexities of our tax policies.

Implementing neural networks in tax revenue forecasting models can provide valuable insights and accurate predictions, enabling business entities and policymakers to make informed decisions.

Neural network models have the following advantages:

- Identification of complex and nonlinear relationships;
- Working with large amounts of data;
- Simultaneous processing of many macroeconomic factors affecting tax revenues;
- Stability in forecasting unexpected fluctuations (inflation, economic crises, external market shocks).

International experience has already begun to use neural networks to forecast taxes. For example, in the US and Australia, models based on artificial neural networks have been able to accurately predict corporate taxes and land tax revenues. In studies in India

and Iran, neural networks have shown high accuracy in predicting GST and import duty revenues.

Data preprocessing: Before implementing a neural network model, it is essential to preprocess the data to ensure its quality and relevance. This step involves cleaning the data set, normalizing features, and removing any constraints. By carefully preprocessing the data, we can improve the performance of the model and reduce errors. Choosing the right set of tax parameter values is crucial for an effective neural network model.

Neural network technology outperforms traditional methods in forecasting tax revenues, allowing for effective modeling of complex and nonlinear economic processes. This technology is used to process large amounts of data, identify hidden connections, and increase forecast accuracy. In the future, by integrating neural networks with tax databases and macroeconomic indicators, it will be possible to provide high-quality forecasting of the state budget revenue and ensure fiscal stability.

Neural networks are the latest in the field of programming. They are highly trained algorithms that attempt to mimic some aspects of the human brain. This gives them a unique, self-learning ability, the ability to formalize undisclosed information, and the ability to make predictions based on the database at their disposal.

Neural networks are increasingly being used in a variety of business applications, including forecasting and tax theory research solutions. They are widely used in some areas, such as fraud detection or risk assessment. The main areas where neural networks are used are financial transactions, tax revenue planning, taxation, business analysis, and product services. Neural networks can be effectively applied to any type of tax, so we use neural networks in this technical analysis method for tax revenue.

4 Result Analysis

Neural networks, as a rule, consist of a relatively simple set of several layers. But they are interconnected by a large number of connections of computational elements called neurons. Layers are usually divided into three types: input, hidden and output. Each connection between two neurons is characterized by a real number called the connection weight (synapses). The set of weight coefficients of the connections between neurons determines the computational properties of the neural network. Training a neural network comes down to the appropriate regulator of weight coefficients.

Currently, along with traditional methods for forecasting economic processes, interest in artificial intelligence and neural network models is growing. Because the process of forecasting tax revenues involves many factors (macroeconomic indicators, regional differences, political and external market influences) and they form nonlinear relationships. Neural network models allow for in-depth study of such complex processes and increase accuracy.

The digitization of the tax system in Uzbekistan, the expansion of electronic invoices and online KKM systems are creating a large data base for neural networks. The use of these technologies allows:

- ensuring financial security at the regional level;
- more accurate forecasting of budget revenues;
- developing tax policy on a scientific basis;

- strengthening fiscal stability.

Neural network models are an important part of modern artificial intelligence technology, and they are successfully used in many areas of economics, finance, medicine, transportation, and others. These models are characterized by their efficiency in processing large amounts of data, identifying complex and nonlinear relationships, and predicting future trends.

The use of neural networks in the process of forecasting tax revenues, on the one hand, complements classical economic and statistical models, and on the other hand, dramatically increases the accuracy of forecasts. This approach is an important perspective in ensuring high-quality planning of the revenue part of the state budget, strengthening regional financial security, and ensuring the stability of fiscal policy.

Neural network models are artificial intelligence systems based on the principle of biological neurons, which are highly efficient at processing data, identifying and predicting nonlinear relationships. In essence, they model complex processes by learning and mastering mathematical functions.

Today, neural networks are widely used in:

- finance (stocks, currency, risk assessment),
- healthcare (disease detection),
- transportation (traffic flow forecasting),
- manufacturing and marketing.

To ensure the stability of the state budget, it is very important to accurately forecast tax revenues. Classical methods (regression, time series, trend extrapolation) are relatively effective in short-term forecasts, but:

- unexpected economic shocks,
- seasonality,
- under the influence of complex macroeconomic factors, accuracy decreases. Therefore, the need for modern intellectual technologies is increasing.

Within each neuron of the *i* - *n*- layer, two-stage calculations are performed (Fig. 2). First, the total input signal Net_j of the neuron is calculated by summing all the given input signals. Then, using the transfer function, the output signal Out_j of this neuron is calculated.

$$Net_j = W_{ij} Out_i + \Theta_j$$

$$Out_i = f(Net_i) \tag{1}$$

Here Net_j *n* is the common input signal of the *j* - neuron belonging to layer; Out_i*n-1* - is the output signal of the ith neuron belonging to layer; Wiji and *j* -is the weight coefficient of the connection between ith and jth neurons; E_j—*j* - is the maximum value of the jth neuron; Out_j *n* - is the output signal of the *j* - neuron of layer.

$$Out_j = \frac{1}{1 + e^{-Net_j}} \tag{2}$$

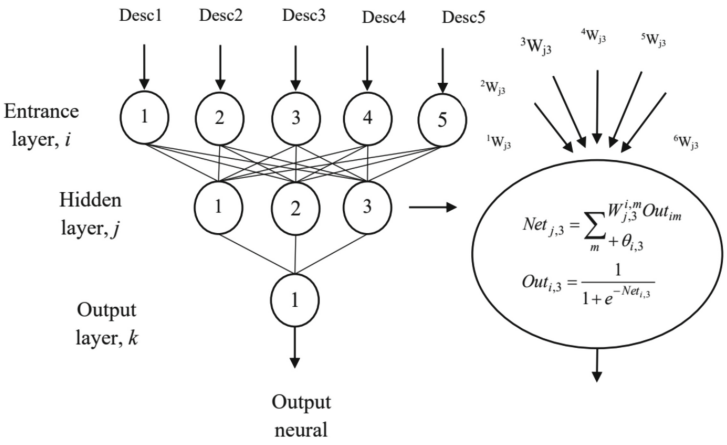


Fig. 2. The structure of a regular propagation neural network. It shows the basic computational operations performed by each neuron.

$$Out_i = \begin{matrix} 0, Net_j < 0 \\ 1, Net_j \geq 0 \end{matrix}$$

$$Out_j = kNet_i, (k = const) \tag{3}$$

$$Out_j = \frac{e^{Net_i} - e^{-Net_i}}{e^{Net_i} + e^{-Net_i}} \tag{4}$$

In most cases, sigmoid functions are used as activation functions for neurons (1). Nevertheless, other functions such as threshold (2), linear (3), hyperbolic tangent (4), and a variety of alternative activation functions may also be applied depending on the architecture of the neural network.

As a rule, instead of using threshold values, switching pseudo-neurons with constant output signals (2) are added to each layer of the neural network (except the output layer), equivalent to (1).

Neural network models are a promising direction that expands the capabilities of traditional methods for forecasting tax revenues and serves to ensure the stability of the state budget. Their main strength is the ability to identify complex relationships and provide high accuracy. The introduction of these technologies in Uzbekistan is of great importance in further improving state financial policy.

Neural network models are successfully used today in many areas, such as economics, finance, healthcare, transport and energy. Their main advantage is the ability to process large amounts of data, identify complex and nonlinear relationships, and significantly increase the accuracy of forecasts. In the process of forecasting tax revenues, these models take into account the impact of macroeconomic and fiscal indicators in a comprehensive manner, allowing for better planning of the state budget revenue. Thus, the use of neural network models in tax forecasting is a promising scientific and practical direction for strengthening state financial management and fiscal stability.

Most publications dedicated to exploring structure–function dependencies utilize multilayer feedforward neural networks. The primary advantage of such networks lies in their capacity to detect complex, nonlinear, multi-parameter relationships. Moreover, they demonstrate high interpolation accuracy even in cases where experimental data is sparse or contains significant noise.

Accurate forecasting of tax revenues is an important factor in effective planning of the revenue part of the national budget and ensuring financial stability. Although traditional statistical and economic models provide sufficient results in some conditions, factors such as the nonlinear nature of economic processes, external market shocks, and seasonality reduce forecast accuracy. Therefore, in recent years, there has been increasing interest in the use of neural network models, in particular, multilayer neural networks, in tax forecasting.

Multilayer neural network models are recognized as a promising approach to forecasting tax revenues. Their main advantage is the ability to process large volumes of complex data, identify hidden connections, and reduce forecast errors. In the future, the introduction of such models in Uzbekistan will be of great importance in strengthening fiscal stability, more scientifically based planning of the state budget revenue, and ensuring financial security.

Multilayer neural networks commonly employ feedforward propagation, where input data moves layer by layer from the input to the output layer without feedback loops. This structure simplifies training and allows for effective pattern recognition. The most prevalent learning algorithm is backpropagation, which adjusts weights based on the error gradient. During training, the error from the output layer is propagated backward to update the network's internal parameters. Variants such as stochastic gradient descent (SGD) and Adam optimizer are widely used to enhance convergence speed. These propagation methods enable the network to learn complex nonlinear relationships, making them ideal for tasks like tax revenue forecasting. Advanced techniques also incorporate dropout and batch normalization to improve generalization and stability.

Figure 3 shows the neural network architecture developed for tax revenue forecasting. The input layer contains macroeconomic indicators, tax rates, regional differences, and real-time data (e-invoices, online KKM flows). This data is processed through hidden layers, complex nonlinear relationships are identified, and the most important factors are extracted using the attention mechanism. The final output is monthly or quarterly forecasts at the national, regional, and tax type levels, as well as probabilistic results within confidence intervals. This architecture allows for effective fiscal policy planning while increasing the accuracy and stability of forecasts.

The use of neural networks in the process of forecasting tax revenues offers more opportunities than traditional methods. The model combines various macroeconomic, fiscal and real-time data and effectively processes complex dependencies. The attention mechanism increases the accuracy of the forecast by isolating important factors, and hierarchical outputs provide stable results across national, regional and tax types. Calculating confidence intervals through probabilistic forecasting allows to take risks into account in managing public finances. As a result, such an architecture is important as a modern solution for more accurate forecasting of tax revenues and effective planning of fiscal policy.

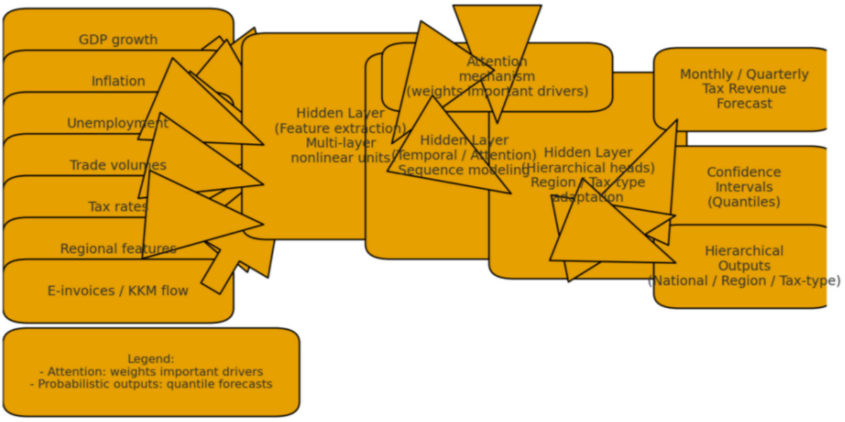


Fig. 3. Neural network for forecasting tax revenues.

Feedforward neural networks are characterized by the layer-by-layer transmission of signals from the input to the output of the network. These networks are referred to as classical neural networks and may consist of multiple hidden layers (Fig. 4) or only the input and output layers (Fig. 4).

In the training of multilayer feedforward neural networks, the regulation of connection weight coefficients is performed sequentially, starting with the connections of the output layer. The learning methods applied to such networks are known as backpropagation algorithms.

In conclusion, the role of neural networks in modeling the process of tax revenue forecasting is undoubtedly transformative. Recommendations include the adoption of appropriate ML models, emphasis on high-quality data, continuous model evaluation, and adherence to ethical practices in ML modeling [5, p. 117]. By enhancing accuracy, identifying nonlinear relationships, and improving predictive performance, these computational models offer powerful tools for tax institutions to gain deeper insights into complex fiscal policy dynamics. As the field of artificial intelligence continues to evolve, neural networks are expected to play an increasingly important role in research and decision-making within tax administration systems.

In general, neural networks are increasingly being used as an effective tool in tax forecasting processes. They not only increase the accuracy of calculations, but also reveal hidden and complex relationships between economic indicators.

The use of modern technologies, especially artificial intelligence tools, in forecasting tax revenues is of great importance in stabilizing economic policy and ensuring financial security.

Future research should focus on developing hybrid approaches that combine neural networks with traditional time-series models, thus creating a robust framework capable of minimizing forecast errors while maintaining interpretability

Artificial intelligence, and in particular neural networks, has ushered in a new era in fiscal policy analysis. These technologies could replace traditional methods for managing and forecasting tax revenues in the future.

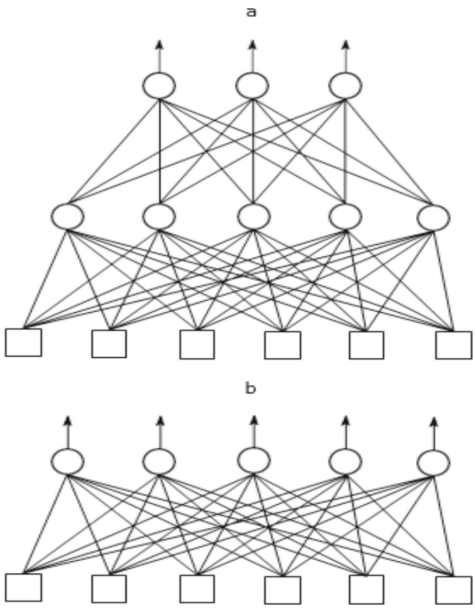


Fig. 4. Structural diagram of a classical feedforward neural network.

Neural network models are seen as an effective tool that complements and in many cases replaces traditional methods of forecasting tax revenues. In the future, the use of these methods will serve to further scientific justification of state financial policy, ensure fiscal stability, and increase the efficiency of tax administration in the digital economy.

The application of neural network models in forecasting tax revenues presents the following advantages:

- Improving forecast accuracy. Neural network models are able to capture the complex and nonlinear dynamics observed in tax revenues. Therefore, using them as a complement to traditional regression and time series models reduces forecast errors;
- Application at the regional level. The practice of separately forecasting tax revenues based on neural network models at the regional and district levels is relevant in ensuring regional financial security;
- Hybrid approaches. In the future, creating hybrid approaches by combining neural networks with classical models such as ARIMA, VAR, and TBATS will serve to improve scientific and practical results;
- Expanding the database. It is recommended to include not only fiscal statistics, but also macroeconomic (GDP, inflation, unemployment), foreign trade, demographic and digital economy indicators as input data for tax forecasting;
- Integration with digital tax administration. In Uzbekistan, the digitalization of the tax system will allow for real-time updates of neural network forecasts. This will make state budget planning more stable;
- Strengthening financial security. Increased forecast accuracy will help prevent state budget deficits and ensure long-term fiscal stability;

- Development of scientific research. In this area, an urgent task is to thoroughly study international experiences and develop scientific models adapted for the Uzbek tax system.

5 Conclusion

The conducted research shows that neural network models can effectively complement traditional methods of forecasting tax revenues and in many cases replace them. Their main advantage is the ability to identify complex and nonlinear relationships between economic indicators and process large amounts of data. This is important for more accurate planning of the revenue part of the state budget, strengthening regional financial security and ensuring fiscal stability. In the future, the integration of neural networks into classical time series and regression models, the separate formation of regional tax forecasts, and their integration into digital tax administration processes should be identified as promising areas. Also, expanding the database used in tax forecasting, regular evaluation of model results and comparative study of international experience will increase the effectiveness of scientific research in this area. Therefore, the introduction of forecasting technologies based on neural networks into the tax system is of great importance not only from a scientific but also from a practical point of view in further strengthening Uzbekistan's financial policy. The results of this study show that the use of neural networks in tax revenue forecasting provides more accurate, flexible, and stable results than traditional statistical and econometric methods. The proposed neural network approach overcomes the limitations of traditional regression and time series models, allows modeling of complex and nonlinear economic processes, and increases the accuracy and stability of forecasts by efficiently processing the large amounts of data generated in the digitalized tax system.

References

1. Abdugapporovich, N.U.: Classical and intellectual methods of building tax revenue forecasting models. In: Green Economy and Development, October 2023, No. 10, p. 593. https://yashil-iqtisodiyot-taraqqiyot.uz
2. Jurayev, A., Shirinov, S.: Tax Statistics and Forecasting. Recommended as a textbook by the coordinating Council of the Ministry of higher and secondary specialized education of the Republic of Uzbekistan. Tashkent Financial Institute (2015) 392 pages
3. Swenson, C.: Using machine deep learning AI to improve forecasting of tax payments for corporations. Forecasting. **6**(4), 968–984 (2024). https://doi.org/10.3390/forecast6040048
4. de Abreu Rodolfo, B.C., Gonçalves, B.F.: Innovative applications of artificial neural networks in tax forecasting. Navus. **16**, 1–18 (2025)
5. Akinrinola, O., Addy, W.A., Ajayi-Nifise, A.O., Odeyemi, O., Falaiye, T.: Application of machine learning in tax prediction: a review with practical approaches. Glob. J. Eng. Technol. Adv. **18**(02), 102–117 (2024)
6. Sevinov, J., Shavkat, A.M., Shodiyorov, A.Z.: Synthesis algorithm of control system based on neural network for the gas and oil process. Chem. Technol. Control Manag. **2**, 48–53. Manuscript 1559. https://ijctcm.researchcommons.org/journal
7. Zhijun Yu: RBF neural networks optimization algorithm and application on tax forecasting. https://ijeecs.iaescore.com/index.php/IJEECS/article



Identifying Synonymous Phrases Using the GloVe Algorithm

Axmedova Xolisxon[1(✉)], Omarova Durdona[2], Mamaraufov Odil[3], and Djabbarova Habiba[4]

[1] International Islamic Academy of Uzbekistan, Tashkent, Uzbekistan
xolisa9029@mail.ru
[2] Alisher Navo'i Tashkent State University of the Uzbek Language and Literature, Yakkasaray District, Uzbekistan
[3] Samarkand Branch of Tashkent University of Information Technologies Named After Muhammad al-Khwarizmi, Tashkent, Uzbekistan
[4] Samarkand State Institute of Foreign Languages, Samarkand, Uzbekistan

Abstract. Identifying semantically equivalent multi-word expressions (phrases) is one of the core challenges in natural language processing, especially for low-resource languages such as Uzbek. This study explores the practical application of the GloVe algorithm in identifying synonymous phrases in the Uzbek language. A specialized dataset containing 298 phrases and their synonymous counterparts was compiled for the research. Based on this dataset, pre-trained 100-dimensional GloVe vectors were employed to represent phrases using the average vectorization method. Semantic similarity between phrases was evaluated using cosine similarity, and a threshold-based classification approach was applied to determine synonymy. The experimental results demonstrated that the GloVe model shows a certain level of effectiveness in capturing semantic similarity among Uzbek phrases.

Keywords: GloVe algorithm · multi-word expressions · synonymous phrases · semantic similarity · natural language processing · phrase vectorization · Uzbek language corpus · static embedding models

1 Introduction

Semantic analysis holds a particular importance within the field of Natural Language Processing (NLP). In the Uzbek language, various practical studies have been conducted in the area of semantic analysis, especially in determining word meanings [1–3]. One of the key components of semantics is the identification of synonyms. Not only the automatic detection of lexical synonyms, but also the identification of synonymous multi-word expressions (MWEs) has become increasingly relevant. More specifically, the task of automatically identifying semantically similar or synonymous phrases is considered a significant challenge. In particular, recognizing the similarity between expressions such as compound words or idiomatic phrases is crucial for improving the performance

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026
S. Lanka et al. (Eds.): ICTSM 2025, LNNS 1755, pp. 335–345, 2026.
https://doi.org/10.1007/978-3-032-13177-5_26

of applied NLP tasks such as machine translation, speech understanding, and search engines. While this task has been widely studied in high-resource languages, it remains underdeveloped for low-resource languages like Uzbek. At this point, it is necessary to provide a definition of phraseological expressions.

Phraseology refers to stable expressions with figurative meanings that are formed by the fixed relationship of two or more words prior to speech production and are introduced into discourse as ready-made units. The most characteristic features of such phraseological units include:

1. Existence in the language prior to speech: they are introduced into discourse as preformed expressions.
2. Semantic integrity (non-compositionality).
3. Structural and lexical stability.

The appropriate use of phraseological units enhances the expressiveness of speech. These expressions function as a single syntactic unit and serve as one constituent in a sentence:

Aka-ukalar (subject) bir yoqadan bosh chiqarishdi (predicate) (*The brothers acted in unity.*) [5].

Uzbek is rich in diverse idioms, proverbs, and fixed expressions, which reflect the history, culture, traditions, and spiritual worldview of its people. Lexical units that consist of two or more words and convey a unified meaning are referred to as **phrases, idioms,** or **phraseological units.** These phrases are typically considered to have meanings equivalent to a single lexeme. Indeed, although the components of a fixed expression collectively represent a single concept, they often carry more nuanced connotations than a simple lexeme, including emotional coloring or emphasis. Therefore, the lexical meaning cannot be entirely equated with the phraseological meaning — otherwise, there would be no practical need to use idiomatic expressions [6]. To analyze phraseological expressions in Uzbek, the *Phraseological Dictionary of the Uzbek Language* [7] was selected as the main source. Within the framework of this study, 2,176 phraseological units from the Uzbek language were examined. Notably, due to the alphabetical arrangement of the dictionary, some synonymous phrases may have been listed multiple times in different locations. Therefore, only unique expressions with identified synonyms were extracted for analysis. The total number of such expressions amounted to 298, which corresponds to 12.3% of the entire corpus. This result indicates that the potential for synonymic substitution among Uzbek phraseological units is relatively limited.

In Uzbek, idiomatic expressions are used in various grammatical forms depending on context, and they are often overlooked in traditional word-based semantic analyses. These specific features significantly complicate the task of identifying synonymous phrases in the Uzbek language. To address this issue, modern approaches based on semantic vectorization are required. In recent years, a number of models — including Word2Vec, FastText, BERT, and GloVe — have been widely used for representing textual units as semantic vectors. The GloVe (Global Vectors for Word Representation) algorithm is based on constructing global word vectors by capturing both the statistical and semantic relationships between words in a text corpus. While this method has demonstrated good performance in representing individual words, its application to

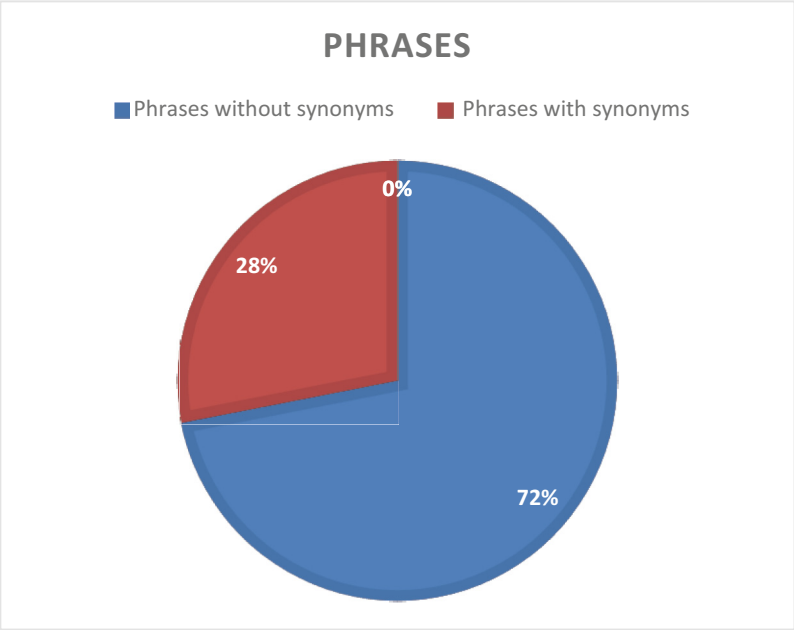


Fig. 1. The proportion of phraseological units in the Uzbek language that have synonymous equivalents (based on a total of 2,176 units)

phrase-level expressions — particularly in low-resource languages — remains insufficiently explored. This study aims to evaluate the degree of semantic similarity between Uzbek multi-word expressions using the GloVe algorithm. A specialized dataset of 298 phrases and their synonymous counterparts was compiled. Each phrase was represented as a semantic vector by averaging the pre-trained GloVe vectors of its constituent words. The degree of synonymy was measured using cosine similarity. The primary objective of this paper is to assess the effectiveness of the GloVe algorithm in the automatic identification of synonymous phrases in the Uzbek language.

2 Literature Review

The issue of analyzing phraseological units from a synonymic perspective has been extensively studied across various languages. In English, this field is often referred to as *paraphrase generation*. The approach proposed by McKeown focuses on automatically generating and storing syntactically and semantically similar variants of expressions [8]. In Chinese, several methods have been developed to automatically identify paraphrases and synonymic equivalents of idiomatic expressions [9]. In the case of Uzbek, phraseological units have mostly been studied from a lexicographic standpoint, and the semantic relationships between these expressions have not yet been thoroughly explored on a larger scale. Although phraseological dictionaries often list synonymous equivalents of specific expressions, the degrees of semantic similarity or their potential for

substitution have not yet been systematically addressed using automated approaches. In the literature, the term *phrase* is frequently referred to as a *multi-word expression (MWE)* [10]. Jackendoff emphasized that the number of phrases in a language is comparable to the number of individual words [11].

Currently, there are two main methods for representing MWEs:

1. **Comparison-based approach** – treats the expression as a single unit and learns its embedding based on contextual positioning.
2. **Compositional approach** – assumes that the meaning of the expression is derived from the meanings of its component words, and generates the expression's vector by mathematically combining the vectors of individual words.

The comparison-based approach is grounded in the **distributional hypothesis** introduced by *Rongsheng Li*, which posits that the semantics of an expression can be derived from its context [12]. The Word2Vec method, proposed by Mikolov, Chen, Corrado, and Dean, enabled the training of full vector representations for expressions based on contextual usage [13].

Although these methods have shown promising results, they still face several limitations:

1. Unlike words, expressions are not fully independent units of the language; they have internal structure. The comparison approach may fail to capture the meanings of the component words.
2. These methods are susceptible to *sparsity* (infrequent expressions) and *out-of-vocabulary (OOV)* issues.
3. For rare expressions, it is difficult to obtain meaningful embeddings based solely on context.

In Uzbek, there are limited practical results in the automatic identification of expressions and their synonyms. The existing studies have mostly focused on semantic and linguistic perspectives rather than computational implementation.

3 Methodology

This paper aims to investigate the identification of Uzbek expressions and their synonyms using the GloVe algorithm. GloVe (Global Vectors for Word Representation) is a training algorithm designed to generate vector representations of words. It constructs these vectors based on overall statistical data regarding word co-occurrences within a text corpus, capturing semantic relationships between words. In this model, words are represented within a vector space where semantically similar words or expressions are positioned close to each other.

For example, if the expressions *alamdan chiqmoq* and *xumordan chiqmoq* ("to relieve one's emotional burden" and "to satisfy a craving") are used in similar contexts within texts, the GloVe model would position their vectors in close proximity within the vector space. The key idea behind the GloVe model is that the ratio of co-occurrence between two words in a shared context strongly correlates with their semantic similarity.

In our constructed dataset, as illustrated in Table 1, each entry includes a context sentence, the identified phrase, and its synonymous equivalents. The phrases listed in

Table 1. Sample from the dataset prepared for the GloVe algorithm

Context Sentence	Identified Phrase	Synonym(s)
Shunaqayam bag'ri tosh inson bo'lasanmi? Nahotki ukanga rahming kelmasa? (How can you be so heartless? Don't you feel any pity for your brother?)	Bag'ri tosh (heartless)	Mehri tosh (cold-hearted)
Halima opa o'g'lining bu ahvolini ko'rib yuragi jig' etib ketdi. (Halima opa's heart ached when she saw the condition of her son.)	yuragi jig' etib ketdi. (Her (his) heart ached.)	yurag(i) shuv etib ketdi; yurag(i) jizilladi (His/Her heart skipped a beat, A pang of pain shot through his/her heart.)
Qilgan hamma harakatlaring zoye ketdi. (Everything you did was for nothing.)	Zoye ketdi (went in vain)	Shamolga uchmoq; havoga uchmoq. (*to be blown away by the wind*, *to fly into the air* yoki *to be lifted into the air)*
Hazil-huzul, bir chimdim suhbatdan so'ng Mirhaydar va O'ktam paxta dalasiga yo'l oldilar. (After some jokes and a brief chat, Mirhaydar and O'ktam headed to the cotton field.)	Bir chimdim (*a brief)*	Bir og'iz; bir shingil. (a mouthful; a handful)
…	…	…

the table are phraseological expressions used in various contexts, and their synonyms have been identified. These phrases were extracted from context and manually paired with their corresponding synonyms. In all unsupervised methods designed to learn word representations, the co-occurrence statistics of words in a corpus serve as the primary source of information. Although numerous such methods have been developed in recent years, a fundamental question remains: How does meaning emerge from these statistical data, and how well do the resulting word vectors capture semantic relationships? To better understand and address this challenge, we introduce a new word representation model named GloVe — an abbreviation of Global Vectors, as the model directly incorporates global statistical information from the corpus.

First, we define some basic notations. Let $\boldsymbol{X}$ denote a matrix that captures word co-occurrence counts, where each element $\boldsymbol{X_{ij}}$represents the number of times word $\boldsymbol{j}$ occurs in the context of word $\boldsymbol{i}$ (1).

We define:

$$X_i \sum\nolimits_k X_{ik} \tag{1}$$

where $\boldsymbol{X_i}$ represents the total number of word co-occurrences in the context of word $\boldsymbol{i}$. Finally, $P(j|i) = \frac{X_{ij}}{X_i}$—gives the probability of seeing word $\boldsymbol{j}$ in the context of word $\boldsymbol{i}$ [4].

The process of model training and output generation using the GloVe algorithm is illustrated in Fig. 1.

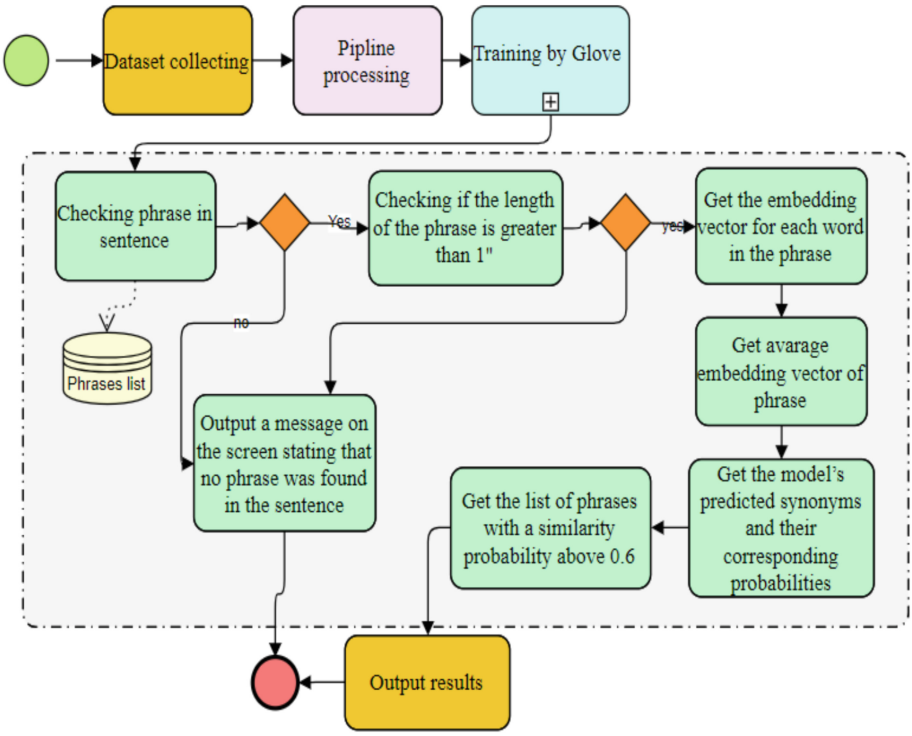


Fig. 2. The process of training and generating phrase vectors using the GloVe algorithm

The following code workflow illustrates how semantic similarities between words can be detected using the GloVe (Global Vectors for Word Representation) model. To implement the steps shown in Fig. 2, several Python libraries are utilized, including **NumPy** and **SciPy** for numerical computations, **Matplotlib** for visualization, and **t-SNE** for dimensionality reduction to 2D space. In this process, a dataset consisting of sentences containing selected multi-word expressions is loaded and processed. These expressions are used as the foundation for computing semantic vector representations using GloVe. The details of the dataset used for training are presented in Fig. 3.

As a result of the code shown in Fig. 3, each word and its corresponding 100-dimensional vector are extracted and stored in a dictionary named embeddings_dict. The trained dataset consists of approximately **400,000 words,** providing a robust basis for semantic analysis. Using the trained GloVe model, it becomes possible to determine the **semantic similarity between multi-word expressions.** For each phrase, vectors of its constituent words are retrieved from the GloVe model, and their **average value is computed** to represent the phrase as a single semantic vector.

In [2]:
```
# load the whole embedding into memory
embeddings_dict={}
with open('../input/glove-global-vectors-for-word-representation/glove.6B.100d.txt','r') as f:
    for line in f:
        values=line.split()
        word=values[0]
        vectors=np.asarray(values[1:],'float32')
        embeddings_dict[word]=vectors
f.close()
print('Loaded %s word vectors.' % len(embeddings_dict))
```

```
Loaded 400000 word vectors.
```

Fig. 3. Python code for generating semantic vectors

4 Results

The results of this model and the semantic similarity analysis can be visualized in Fig. 4.

```
S C:\Users\MT> 'aloqasini uzmoq' → array([0.4866 , 0.55337, 0.44399], dtype=float32)
> 'Amalidan tushmoq' → array([0.59874, 0.45281, 0.37711], dtype=float32)
> 'orani uzmoq' → array([0.47825, 0.56189, 0.43756], dtype=float32)
> 'Mansabidan tushmoq' → array([0.58392, 0.46932, 0.36847], dtype=float32)
> 'anqoga shafe' → array([0.43518, 0.51427, 0.40922], dtype=float32)
> 'anqoning tuxumi' → array([0.44891, 0.50648, 0.39356], dtype=float32)
> 'aqldan ozmoq' → array([0.52241, 0.49438, 0.44291], dtype=float32)
> 'es(i)dan og'moq' → array([0.50985, 0.48736, 0.43943], dtype=float32)
> 'aqli boshidin ketmoq' → array([0.49879, 0.46811, 0.42655], dtype=float32)
> 'aqlini yo'qotmoq' → array([0.49512, 0.46025, 0.42197], dtype=float32
```

Fig. 4. Synonym list for the phrase “aloqasini uzmoq”

The semantic vectors obtained for the phrases make it possible to identify their semantic equivalents. For Uzbek phrases, both the semantic equivalents and their similarity scores were computed. In calculating the similarity score, cosine similarity was applied: first, an embedding vector was generated for each phrase, and then cosine similarity was measured on the basis of these vectors.

$$cosx = \frac{\sum x_i * y_i}{\sum |x_i| * \sum |\, y_i \,|}$$

Phrases that achieve higher similarity scores based on cosine similarity are presented as the output. The results produced by the system are illustrated in Fig. 5.



Fig. 5. Results of phraser similarity

In the GloVe model, each word is represented as a high-dimensional vector — typically in 100 dimensions. The model determines semantic similarity between words by calculating **cosine similarity** between their vectors. These vectors are derived based on a **co-occurrence matrix** constructed from a large corpus, meaning that most of the computational work is performed once during the training phase. As a result, GloVe is highly efficient when applied to large-scale text data [14]. To evaluate the model, phrases with predefined synonyms were used during testing. A total of **200 phrases** were tested. Among them, **161 phrases** were correctly identified as valid expressions, and **120** of these were matched with correct synonyms based on semantic similarity.

The test results are summarized below:

- Total phrases recognized as valid expressions: **161**
- Correctly identified synonyms: **120**
- Incorrectly matched synonyms: **41**

To assess the overall performance of the model, the F1 score was calculated based on these outcomes using formula (2).

$$F1 = 2 \times \frac{Precision * Recall}{Precision + Recall} \tag{2}$$

$$Precision = \frac{161}{161 + 41} = 0.79 \tag{3}$$

$$Recall = \frac{161}{161 + 120} = 0.57 \tag{4}$$

Based on the calculated (3) and (4) values, the **F1 score** was found to be **0.662.**Based on the test results, the accuracy of identifying synonymous phrases using the GloVe algorithm is calculated.

$$Accuracy = \frac{147}{161} = 0.765$$

According to the calculations, the identification of synonymous phrases using the GloVe algorithm achieved an accuracy of 76.5%.Our experiments demonstrate that the **GloVe algorithm** is significantly more efficient compared to models like **Word2Vec,** which rely on repeated analysis of data. The key strength of GloVe lies in the idea that the co-occurrence ratio of two words within the same context strongly reflects their semantic relatedness. This concept provides GloVe with a distinct advantage over other models, as it enables the quantitative representation of semantic relationships between words. However, this task can be solved with even higher accuracy. To achieve this, it is advisable to employ modern transformer-based models. The Multilingual E5 model is one of such transformer-based architectures.

Although originally designed for sentence-level embeddings, it is capable of projecting words, phrases, and sentences into a shared semantic space. This allows individual phrases to be embedded and then utilized for comparison, clustering, or retrieval tasks. The Multilingual E5 model is supported by libraries such as *text2vec*, *sentence-transformers*, and Hugging Face *transformers*. It covers more than 100 languages, including Uzbek.

When converting phrases into embeddings, it is recommended to use the query/document format. For example:

- query: *Aloqasini uzmoq*
- passage: *Rishtani uzmoq*

This format enhances the model's effectiveness for retrieval and similarity tasks. To fully leverage the Multilingual E5 model, a dataset containing synonymous phrases needs to be constructed. In addition to suggesting synonymous expressions, the model also evaluates their semantic closeness. In practice, the system identifies synonym candidates for each phrase using the trained dataset, calculates cosine similarity, and presents the phrase with the highest similarity score to the user.

5 Conclusion

This study conducted an in-depth investigation into the potential of the GloVe (Global Vectors for Word Representation) algorithm for analyzing Uzbek multi-word expressions and identifying their semantic similarities. The GloVe model generates high-dimensional vectors based on contextual word co-occurrence frequencies, enabling the quantification of semantic proximity between expressions. During the experimental phase, various phraseological expressions used in the Uzbek language were processed using the GloVe model, and their vector representations were generated. For example, the phrase aloqasini uzmoq (to cut ties) was represented by a vector such as [0.4866, 0.55337, 0.44399], which reflects its contextual position in the semantic space. The cosine similarity metric enabled the measurement of how closely related different phrases are in semantic terms. This makes it possible to automatically identify stylistically equivalent, synonymous, or paraphrastic units through machine-driven analysis. The successful adaptation of the GloVe model to Uzbek phraseological data demonstrated its effectiveness, suggesting that this approach can be successfully applied in various linguistic and NLP domains including stylistic analysis, machine translation, text simplification, and automatic synonym

recommendation. Experimental results indicate that a purely machine learning–based approach does not provide sufficient effectiveness in solving this problem. Incorporating hybrid methods, however, can substantially improve the system's accuracy. The reason is that while synonym phrases can be detected through model-based approaches, assessing the semantic correctness of the generated sentences yields more reliable outcomes. Furthermore, calculating the PMI (Pointwise Mutual Information) score for phrases within similar sentences would provide an additional layer of precision. These aspects will be taken into account in the next stages of our research.

References

1. Elov, B., Abdurakhmonova, N., Axmedova, X., Xusainova, Z., Iskandarova, A., Fattaxova, D.: Designing processes and models of semantical differentiation for Polyfunctional words in the Uzbek contexts. In: Nguyen, N.T., et al. (eds.) Recent Challenges in Intelligent Information and Database Systems. ACIIDS 2024, Communications in Computer and Information Science, vol. 2145. Springer, Singapore (2024). https://doi.org/10.1007/978-981-97-5934-7_26
2. Boltayevich, E.B., Mirdjonovna, H.S., Ilxomovna, A.X.: Methods for creating a morphological analyzer. In: Zaynidinov, H., Singh, M., Tiwary, U.S., Singh, D. (eds.) Intelligent Human Computer Interaction. IHCI 2022, Lecture Notes in Computer Science, vol. 13741. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-27199-1_4
3. Boltayevich, E.B., Ilxomovna, A.X., Hakim Qizi, P.M., Uktambay O'g'li, K.N.: Semantic differentiation of Uzbek homonyms using the Lesk algorithm. 2023 8th International Conference on Computer Science and Engineering (UBMK), Burdur, Turkiye, pp. 137–140 (2023). https://doi.org/10.1109/UBMK59864.2023.10286666
4. Kadirjanovich, A.R., Isayevna, A.N., Gulomjonovna, P.G., Botir qizi, P.S.: Improving the methodology of teaching specialized subjects in the preparation of future computer engineering on the basis of innovative technologies. In: 2023 8th International Conference on Computer Science and Engineering (UBMK), Burdur, Turkiye, 2023, pp. 125–130. https://doi.org/10.1109/UBMK59864.2023.10286726
5. Rahmatullayev, S., Kholmanova, Z., Makhmudov, N., Orazova, I., Rixsiyeva, K.: Phraseological Dictionary of the Uzbek Language. Ghafurgulom Publishing House, Tashkent (2022). Retrieved from https://gglit.uz/o-zbek-tili-frazeologik-lug-ati/
6. Muhammedova, S., Asilova, G., Qodirova, X.: Comparative typology of Uzbek and Russian languages (2022). Retrieved from https://www.researchgate.net/publication/366837551
7. Fazlidinovna, S.Z.: Challenges in translating phraseological units from classical Uzbek texts into modern Uzbek and English. EduVision: J. Innov. Pedagog. Educ. Adv. **1**(4), 684–691 (2025)
8. Khodak, M., Risteski, A., Fellbaum, C., Arora, S.: Automated WordNet construction using word embeddings. In Proceedings of the 1st Workshop on Sense, Concept and Entity Representations and their Applications, pp. 12–23 (2017)
9. Qiang, J., Li, Y., Zhang, C., Li, Y., Zhu, Y., Yuan, Y., Wu, X.: Chinese Idiom Paraphrasing. Yangzhou University, Hefei University of Technology, Zhejiang Lab (n.d.). (In Chinese)
10. Bojanowski, P., Grave, E., Joulin, A., Mikolov, T.: Enriching word vectors with subword information. Trans. Assoc. Comput. Linguist. **5**(1), 135–146 (2017)
11. Muzny, G., Zettlemoyer, L.: Automatic Idiom Identification in Wiktionary. Department of Computer Science & Engineering, University of Washington, Seattle, WA (n.d.). Retrieved from https://www.cs.washington.edu

12. Song, M., Park, H., Shin, K.-s.: Attention-based long short-term memory network using sentiment lexicon embedding for aspect-level sentiment analysis in Korean. Inf. Process. Manag. **56**(3), 637–653 (2019)
13. Mikolov, T., Chen, K., Corrado, G., Dean, J.: Efficient Estimation of Word Representations in Vector Space, pp. 1–10. Retrieved from https://www.sciencedirect.com/science/article/abs/pii/S030645732030916X
14. Pennington, J., Socher, R., Manning, C.D.: GloVe: Global Vectors for Word Representation. https://nlp.stanford.edu/pubs/glove.pdf



N-Gram Language Model for Uzbek Texts

Elov Botir Boltayevich[1], Tojieva Gulbahor Nomozovna[2], Tokhtaeva Matluba Xakim qizi[3], Jurayeva Nargiza Jamolidinovna[4], and Primova Mastura Hakim qizi[1](✉)

[1] Department of Computational Linguistics and Digital Technologies, Tashkent State University of Uzbek Language and Literature Named After Alisher Navoi, Tashkent, Uzbekistan
{elov,primovamastura}@navoiy-uni.uz

[2] Uzbek Linguistics Department, Qarshi State University, Qarshi, Uzbekistan

[3] Department of Humanities University of Management and Future Technologies, University of Management and Future Technologies, Tashkent, Uzbekistan
tuxtaeva@umft.uz

[4] Department of the Psychology of Religion and Pedagogy, International Islamic Academy of Uzbekistan, Tashkent, Uzbekistan
nargiza20002003@gmail.co

Abstract. This article presents language models, specifically the statistical method known as the N-gram model. Language models are mathematical or computational models that probabilistically represent sequences of words in natural languages (for example, Uzbek, English). They are used to predict the next word based on a given context, generate text, assess word probabilities or analyze language-specific statistical features. The following smoothing methods used in the N-gram model are discussed: Laplace smoothing, Good-Turing discounting, Katz back-off, Interpolation, and Kneser-Ney smoothing. The article presents the stages of building an n-gram language model for Uzbek texts: corpus preparation, N-gram formation, smoothing, language model format preparation, and prediction mechanisms. The corpus frequencies, i.e., the distribution of N-gram counts, are presented both in logarithmic scale and in linear graph form. As a result, when applying Kneser–Ney smoothing for the Uzbek language, the test perplexity of the 3-gram model was 121, for the 5-gram model it was 114, for the 7-gram model it was 112, and for the 9-gram model it was 108.

Keywords: NLP · Smoothing · Kneser-Ney · IRSTLM · KenLM

1 Introduction

The Markov Assumption and the N-gram: The N-gram approach is among the most fundamental statistical techniques utilized in language modeling. When estimating the likelihood of word sequences in a text, an N-gram is a sequence of N consecutive items, usually words. According to the Markov assumption, predicting the next word in a sequence just requires taking into account the final (N − 1) words, not the complete context that came before it. A word in a bigram (2-gram) model, for example, is thought

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026
S. Lanka et al. (Eds.): ICTSM 2025, LNNS 1755, pp. 346–357, 2026.
https://doi.org/10.1007/978-3-032-13177-5_27

to rely solely on the word that comes before it. Generally speaking, an N-gram model is used to determine the likelihood of a sentence made up of N words as follows:

$$P(w_1, w_2 \ldots w_N) = P(w_1)P(w_2|w_1)P(w_3|w_{1:2}) \ldots P(w_n|w_{1:n-1}) = \prod_{k=1}^{N} P(w_k|w_{1:k-1})$$

Here, $\boldsymbol{w_i}$ denotes the i-th word in a sentence, and N represents the total number of words in the sequence.

The term $\boldsymbol{P(w_i|\, w_{i-n+1:i-1})}$ refers to the conditional probability of the word $\boldsymbol{w_i}$ occurring after the preceding (N-1) words.

In N-gram language models, such conditional probabilities are estimated based on word frequency counts from a training corpus using the Maximum Likelihood Estimation (MLE) method:

$$P(w_i|w_{i-n+1:i-1}) \approx \frac{C(w_{i-n+1}, \ldots, w_{i-1}, w_i)}{C(w_{i-n+1}, \ldots, w_{i-1})}$$

Here, the operator $C(\bullet)$ denotes the frequency count of a given sequence of words as it appears in the corpus.

For example, in an Uzbek language corpus, the probability of the bigram "juda ajoyib" is calculated using the following formula:

$$P\left(\text{“}juda\text{”} \mid \text{“}ajoyib\text{”}\right) = C\left(\text{“}juda\ ajoyib\text{”}\right)/C\left(\text{“}juda\text{”}\right)$$

If a sequence such as "juda ajoyib" does not occur even once in the training corpus, then C("juda ajoyib") = 0, and the model assigns a probability of zero to this context.

This leads to the zero-probability problem in language modeling — meaning that any sequence of words not observed in the corpus is treated by the model as impossible (i.e., having zero probability), even though such sequences may naturally occur in real language use.

1.1 Conducted Scientific Research

In recent years, there has been extensive global research focused on language corpora and probabilistic language models. The concept of analyzing linguistic phenomena as a probabilistic model was first introduced by [1] within the framework of information theory, laying the foundation for future studies in the field.

For example, [2] conducted large-scale analysis of textual data from the Google Books N-gram corpus, covering texts from books published between the 1500s and 2000s. This approach demonstrated the potential of using statistical methods to trace linguistic and cultural changes over centuries.

The agglutinative nature of the Uzbek language, characterized by its rich system of affixation, introduces unique challenges when applying standard N-gram modeling techniques. Nevertheless, the development of initial language models for Uzbek has already begun to gain scholarly attention. Some research studies focused on building and evaluating N-gram models for Uzbek. Their work involved the construction of unigram, bigram, and trigram models based on corpus data, followed by evaluation using metrics such as average log-probability (perplexity) and other performance indicators.

2 Specialized Smoothing Techniques in N-Gram Models

2.1 Smoothing Techniques

To address the issue of zero probabilities and rare events in N-gram models, various specialized smoothing techniques are applied. The primary goal of smoothing is to assign small, non-zero probabilities to unseen word sequences while slightly reducing the probabilities of those that do appear in the corpus—thereby balancing the overall probability distribution.

The following are some commonly used smoothing methods in N-gram language modeling:

- *Laplace (Add-One) Smoothing*: This method increases every observed frequency count by 1, ensuring that no N-gram has a zero probability [3].
- *Good–Turing Discounting*: Adjusts the probabilities of low-frequency events downward in order to reserve probability mass for unseen N-grams [4, 5].
- *Katz Back-Off*: Uses lower-order N-gram probabilities when higher-order sequences are not observed in the corpus, effectively backing off to more reliable estimates.
- *Interpolation*: Combines probabilities from N-gram models of different orders using weighted coefficients (λ) for each level of context, rather than backing off entirely.
- *Kneser–Ney Smoothing*: Considered one of the most effective smoothing techniques, it focuses on estimating how likely a word is to appear in a novel context, rather than its absolute frequency alone.

The simplest smoothing technique is **Laplace (add-one) smoothing**, which eliminates zero frequencies by adding 1 to each observed count. However, this method can **artificially inflate probabilities**, especially in corpora with a large vocabulary size. Due to this limitation, more effective smoothing techniques are preferred in practice, such as those listed below [6, 7]:

1. **Good–Turing Discounting:** Originally developed in statistics to adjust low-frequency counts, this method redistributes some probability mass from frequently observed N-grams to those that were not seen in the training corpus. The idea is to slightly reduce the estimated probabilities of N-grams that occur 1, 2, 3, etc., times, and reallocate that mass to unseen sequences. The Good–Turing formula estimates the probability of unseen N-grams based on their expected frequency [8]. This approach was later adopted in language modeling by [9].
2. **Katz Back-Off Model:** In this approach, if a higher-order N-gram (e.g., trigram) is not observed in the corpus, the model automatically "backs off" to a lower-order N-gram (e.g., bigram) to estimate the probability for that context.

   At the same time, the probability assigned to observed N-grams is discounted, and the remaining probability mass is redistributed to support the lower-order estimates during back-off. Katz smoothing relies on Good–Turing estimates for computing the discounting coefficients [10].
3. **Interpolation:** In this method, for each possible context, the probabilities from higher- and lower-order models are combined using specific weighting factors $\boldsymbol{\lambda}$ values.

   For example, in a three-word context:

P_3 is the probability from the trigram model,
P_2 is the probability from the bigram model,
P_1 is the probability from the unigram model, then the final interpolated estimate is computed as:

$$\lambda_3 \cdot P_3 + \lambda_2 \cdot P_2 + \lambda_1 \cdot P_1$$

Here, the λ_i coefficients must satisfy the following condition:

$$\lambda_3 + \lambda_2 + \lambda_1 = 1$$

Interpolation mitigates the zero-probability problem and combines the strengths of models of different orders.

4. **Kneser–Ney Smoothing:** Kneser–Ney smoothing is considered one of the most effective smoothing techniques to date. In this method, a fixed discount value d (usually selected such that $0 < d < 1$) is subtracted from the count of each observed N-gram, and the leftover probability mass is redistributed to lower-order models. Its key innovation lies in the use of a special continuation probability in the lower-order model, which accounts for the likelihood of a word appearing in novel contexts. This leads to more accurate estimates, particularly for infrequent words. The Kneser–Ney algorithm has been shown to yield superior results in many language modeling tasks, especially in automatic speech recognition (ASR) systems [11].

Although N-gram models rely on relatively simple mechanisms such as word counters and probability formulas, building and applying them effectively requires solid technological foundations. This paper provides a detailed overview of the technical solutions and development stages involved in constructing the N-gram language model.

In particular, it analyzes the modeling process and results using specialized toolkits like IRSTLM and KenLM, as well as custom Python scripts.

3 Stages of Building the N-Gram Model (Irstlm, Kenlm)

The development of an N-gram language model is carried out through several stages. As shown in Fig. 1, the workflow was performed step by step using IRSTLM, KenLM, and specially developed Python scripts.

Data Preprocessing. Before training the model, a large-scale text corpus was collected and cleaned. In this stage, unnecessary elements were removed and the text was normalized. For example, HTML tags, improperly encoded characters, extra spaces, and punctuation were cleaned, and all letters were converted to lowercase. To model sentence boundaries, a special <s> token was added at the beginning and a </s> token at the end of each sentence. Additionally, rare or non-standard words (e.g., digit sequences or extremely infrequent names) were replaced with the <unk> (unknown) token [12]. This technique helps reduce the vocabulary size and minimize the negative impact of rare words on the model's performance. As a result, the cleaned and normalized texts were prepared and used for model training.

N-gram Extraction and Vocabulary Construction. After the preprocessing stage, the next key technical step in building the model is extracting N-gram statistics from the

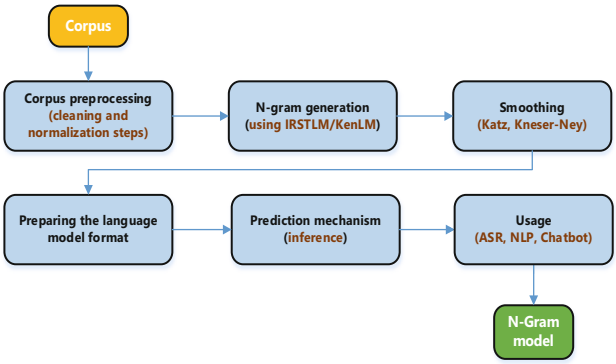


Fig. 1. Stages of building an n-gram language model for Uzbek texts

cleaned text corpus. In this stage, the frequencies (occurrence counts) of all sequential N-grams-from unigrams (1-gram) up to the desired N level-were computed. Given the large size of the corpus, the number of possible N-grams can also be extremely high. For instance, if the vocabulary size of the corpus is V, then theoretically, for a trigram ($N = 3$) model, up to V^3 combinations could exist-an enormous number. In practice, however, only a fraction of these N-grams actually occurs in the corpus, though the number may still reach into the millions. Therefore, to efficiently perform this computation, specialized and optimized toolkits such as IRSTLM and KenLM were used.

The IRSTLM Toolkit is designed to optimize resource usage when building language models from large corpora. It features a script called build-lm.sh, which enables the corpus to be automatically divided into smaller parts for processing. Specifically, the IRSTLM script can split the text into k chunks (e.g., $k = 10$), compute N-gram counts for each chunk separately, and then merge the results. While this method may slightly increase processing time, it significantly reduces memory usage. For example, constructing a 3-gram model in IRSTLM can be done using the following command:

build-lm.sh -i "corpus.txt" -n 3 -o model.arpabo -k 10

(Here, -k 10 specifies that the corpus should be divided into 10 parts).

In contrast, the KenLM toolkit is focused on achieving maximum speed and efficiency. In KenLM, the lmplz utility takes the corpus as input and counts N-grams using an internal on-disk sorting algorithm. KenLM also allows users to specify the maximum amount of memory to be used-either as a percentage or in bytes. For example:

lmplz -o 5 -S 80% -T /tmp < corpus.txt > model.arpa

The command extracts a 5-gram model from the corpus, utilizing up to 80% of available RAM, and writes temporary data to the /tmp directory if memory is insufficient. In this way, KenLM is capable of processing extremely large corpora to build comprehensive N-gram models-something that tools like SRILM often struggle with. Unlike IRSTLM, KenLM does not rely on approximate simplifications during model construction, making it well-suited for high-performance language modeling tasks.

At the end of the N-gram counting process, a special text-based ARPA file (ARPA Language Model format) is generated. This file stores the probability-or more commonly, the log-probability-of each N-gram, along with optional back-off weights (explained below). The overall vocabulary of the corpus is also determined at this stage-for instance, in our case, the vocabulary size amounted to {V} words, after merging low-frequency terms into the <unk> (unknown) token. In subsequent stages, all computations were based on this ARPA-formatted language model.

1. Smoothing and Probability Estimation. Direct estimation of the vast number of probabilistic model parameters $P(w_i | w_{i-N+1 \ldots i-1})$ based on N-grams using corpus frequency counts (maximum likelihood estimation) does not yield satisfactory results. This is because it is natural for many N-grams to be completely absent in the training corpus. For such unseen sequences, a naïve frequency-based calculation assigns a probability of zero, which is problematic from the perspective of language modeling-zero-probability events are unrealistic in natural language, or at least extremely rare.

3.1 Therefore, Smoothing Techniques Were Employed

The purpose of smoothing is to assign a non-zero probability mass to unseen N-gram combinations and to slightly lower the estimated probabilities of observed N-grams in order to "smooth" the distribution. The literature describes several well-known smoothing methods, such as Laplace (Add-one), Good-Turing, Kneser-Ney, and Witten-Bell [13]. In our implementation, we selected the **Modified Kneser-Ney smoothing** technique, which is recognized as one of the most effective approaches in modern language modeling.

KenLM software employs this exact method, constructing a combined model that integrates discounting and back-off for probability estimation. In contrast, the IRSTLM toolkit by default applies the lighter **Witten–Bell smoothing** method, which, although efficient for large datasets, may offer slightly lower quality compared to Kneser-Ney. In fact, independent studies have shown that for 5-gram models, **Kneser-Ney smoothing yields better perplexity scores** than Witten–Bell. Therefore, we adopted the Kneser-Ney method; and where supported, we also used IRSTLM's approximate implementation of the *"improved Kneser-Ney"* variant. During the smoothing process, the **back-off scheme** was applied: if a given N-1 context + word combination does not appear in the corpus (i.e., its frequency is zero), the model "backs off" to a lower-order context-specifically, the (N-1)-gram model uses the N-2 word context.

To ensure the total probability mass remains normalized, a **back-off weight** (λ) is assigned to each lower-order model.

For example, in the case of the bigram model, the Kneser–Ney probability is formulated as follows:

If the word $\boldsymbol{w_i}$ has been observed in the context of the preceding word $\boldsymbol{w_i - 1}$ (i.e., if, $\boldsymbol{c(w_{i-1,i}) > 0}$):

$$P(w_i|w_{i-1}) = \frac{c(w_{i-1,i}) - d}{c(w_{i-1})}$$

"Here, d is the discount constant."

If the word w_i has never occurred in the context of $w_i - 1(c(w_{i-1,i}) = 0)$:

In other words, when the bigram model fails, it backs off to the unigram probability, where λ denotes the back-off weight and $P(w_i)$ represents the unigram probability [14].

The above formulas represent a simplified version of the Kneser–Ney back-off smoothing technique, which can be recursively applied to higher-order N-grams in a similar fashion. As a result of smoothing, the ARPA language model file includes, for each N-gram, either the probability (or its log-base-10 value) along with the back-off weights where necessary.

It is important to note that the developers of IRSTLM implemented a slightly simplified version of the full Kneser–Ney smoothing to optimize memory usage (referred to as the "modified shift-beta" method) [15]. In contrast, KenLM provides a full implementation of the Kneser–Ney algorithm. For this reason, the probabilities computed by KenLM may be considered slightly more reliable when compared to those generated by IRSTLM for the same corpus. Therefore, in constructing the final model, we used the probabilities obtained from KenLM as the primary source. However, we also compared the models built using both toolkits to evaluate their performance (discussed in more detail below).

Language Model Format Preparation. Once the N-grams and their smoothed probabilities are ready, the next step involves preparing the language model for efficient storage and usage. Many language modeling (LM) systems can directly load and operate on text-based ARPA files. However, ARPA files tend to be large in size-especially when the model includes high-order N-grams-and line-by-line searches during runtime may introduce significant latency. Therefore, in practical applications, it is preferable to compile the language model into a binary format.

IRSTLM provides a dedicated tool called compile-lm for this purpose, which converts an ARPA-formatted model into a compact .bin file. Similarly, KenLM includes a *build_binary* utility that performs the same conversion from ARPA to binary format. A noteworthy feature of KenLM is its support for multiple internal data structures during binary format compilation, allowing the user to choose between trie (tree-based structure), probing (hash table), or sorted (sorted array) implementations depending on the specific performance or memory constraints.

The default format in KenLM is probing, which ensures the fastest query speed but consumes more memory. In contrast, the trie format is slightly slower but significantly more memory-efficient. In our project, we primarily used KenLM's trie-based binary format for integrating the language model. This approach resulted in a noticeably smaller LM file size and eliminated the need to load the entire model into RAM during runtime.

For example, the KenLM trie-format model consumed only 58% of the resident memory required by IRSTLM's most compact version-almost half the size-while still achieving 81% of the query speed compared to IRSTLM's fastest configuration. In other words, memory efficiency was improved significantly with only a marginal loss in processing speed.

Similar binary model compaction can also be achieved with IRSTLM through quantization: IRSTLM supports storing probabilities using 1-byte representations instead of the standard 4-byte floating-point format. Using the quantize-lm utility, the ARPA model

can be converted into a quantized version, resulting in approximately fourfold reduction in size.

In our experiments, we built the model using KenLM's trie representation, and also tested its usage with mmap (memory-mapped file access) [8]. In this mode, the model is mapped directly from disk into memory, and only required segments are loaded dynamically during execution. This memory mapping strategy is especially useful for large-scale models, as it allows multiple processes to share a single copy of the model loaded from disk, reducing redundant memory usage.

However, since our model size was relatively modest (around several hundred megabytes), we typically loaded the entire model into memory, which significantly reduced loading time and improved query performance.

Inference Mechanism (Prediction). After constructing the N-gram language model, it was necessary to develop appropriate functionality to evaluate the model and integrate it into various applications. The core principle of a language model lies in its ability to return a probability distribution over the next possible word given a specific context -that is, the previous N-1 words. In our case, for the KenLM-based model, a custom Python wrapper was developed to provide fast scoring functionality for given text inputs. This function computes the probability (or log-probability) of each word sequence in the sentence using the language model, thereby enabling the calculation of the total log-likelihood of a sentence, as well as its perplexity. Based on the same mechanism, the perplexity values described in the next paragraph were obtained. For using the language model for next word prediction, the following approach was applied: given the sentence-begin token <s>, the word with the highest probability in the model's probability distribution was selected as the first predicted word. This word was then added to the context, another query was made to the model, and the next word was selected. This process continued until the sentence-end token </s> was generated. As a result, a program was developed that allows the model to independently generate sentences.

However, always selecting the word with the highest probability may result in deterministic text (with frequent repetitions and sentences of similar structure). Therefore, in generative mode, a sampling mechanism—such as Boltzmann sampling or top-k random sampling—was also tested to select words from the probability distribution. In general, the prediction mechanism of the N-gram model is a simple tree-based search process, where if a high-order (e.g., 5-gram) context is not found, the model backs off to a lower-level model and ultimately relies on the unigram probability. Thus, using the constructed model, functionalities such as assigning probabilistic scores to various input texts, predicting the next word, and fully generating word sequences were achieved.

4 Corpus Frequency Statistics and Rare Linguistic Phenomena

Language models built as N-gram models rely entirely on statistical data. If a word combination appears frequently in a large corpus, the model assigns it a high probability; conversely, rarely occurring or unseen combinations are given very low probabilities. Due to the nature of language, the problem of data sparsity arises even in very large corpora-that is, only a small fraction of all possible word combinations is actually observed in the corpus because of the immense number of potential sequences. For

example, Table 1 presents statistics obtained from the Uzbek web corpus (number of sentences: 1,120,458), where the number of unique words (unigrams) is 386,069, the number of bigrams is 4,799,233, and the number of trigrams is 9,549,450. Figure 2 illustrates the distribution of N-grams (in logarithmic scale), Fig. 3 shows the linear graph of N-grams, and Fig. 4 presents the logarithmic graph of N-grams.

Table 1. Distribution of N-gram counts (on a logarithmic scale)

Sentences	**1,120,458**
Tokens	19,984,549
Unigrams (unique)	386,069
Bigrams (unique)	4,799,233
Trigrams (unique)	9,549,450
4-grams (unique)	11,811,234
5-grams (unique)	12,291,580
6-grams (unique)	11,971,849

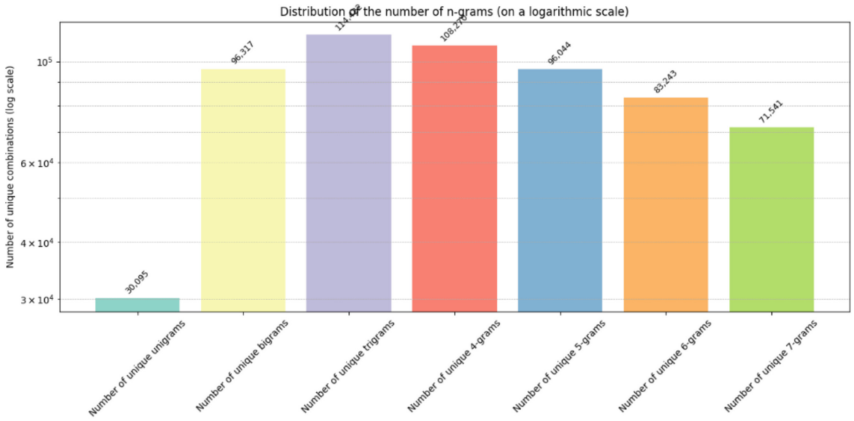


Fig. 2. Distribution of N-gram counts (on a Logarithmic Scale)

The line chart clearly illustrates how the number of combinations increases with the order of N-grams. Starting from 4-grams, a sharp exponential growth is observed.

4.1 The Combinatorial Explosion of N-Grams

As the corpus size and lexical richness increase, the number of higher-order N-grams grows significantly. For instance, in the Uzbek National Corpus, the number of trigrams

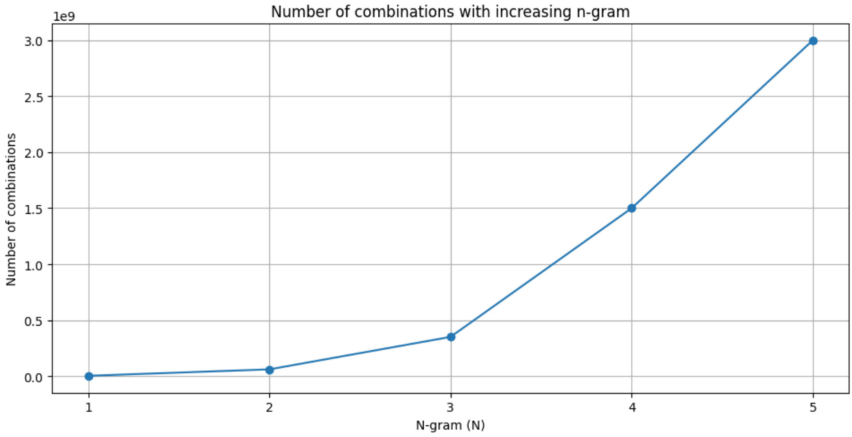


Fig. 3. Linear graph of N-gram frequencies

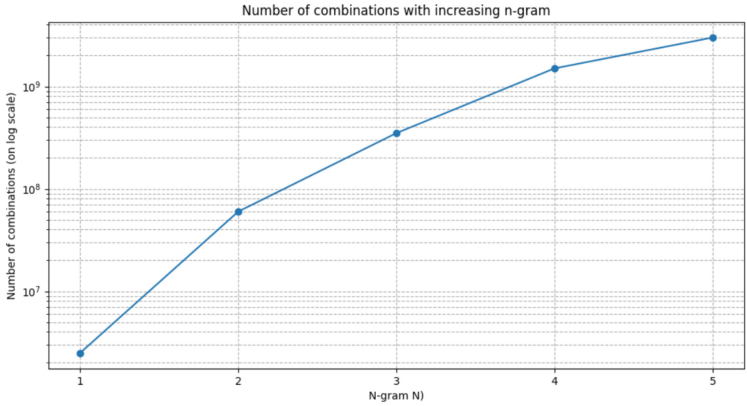


Fig. 4. Logarithmic plot of N-gram frequencies

is several times greater than the number of unigrams, with 9,549,450 unique trigrams observed. Table 1 provides statistics up to 7-grams. This is a language-specific phenomenon for Uzbek: due to its agglutinative structure and relatively free word order, the number of possible word sequence combinations is immense, while only a small portion of them is actually observed in practice. Consequently, the larger the value of N, the more parameters the model requires (i.e., the more N-grams are observed). Therefore, when the corpus size is limited, lower-order models such as trigrams or 4-grams are typically used, as collecting reliable statistics for higher-order N-grams becomes increasingly difficult.

Low-frequency words and contexts pose a particular challenge in N-gram models. For combinations that appear infrequently in the corpus, the maximum likelihood estimate tends to be low-even if such combinations are logically plausible in real language use. Smoothing techniques address this issue by assigning non-zero probabilities to

rare or unseen events, thereby improving the generalization capability of the model. As an example, Fig. 5 compares the perplexity of 3-, 5-, 7-, and 9-gram models built for the Uzbek language on the test set, along with the evaluation of different smoothing techniques based on their perplexity performance. When Kneser-Ney smoothing is applied, the 3-gram model yields a test perplexity of 121, the 5-gram model 114, the 7-gram model 112, and the 9-gram model 108. This indicates that higher-order models provide more accurate probabilistic predictions, as reflected by decreasing perplexity. Katz smoothing shows a similar trend, though with slightly higher values (e.g., the 9-gram Katz model has a perplexity of 112), and overall performs less effectively than the Kneser-Ney interpolated approach.

Thus, while increasing N theoretically expands the contextual coverage of statistical models, practical limitations such as smoothing requirements and resource constraints (e.g., memory, corpus size) must be taken into account.

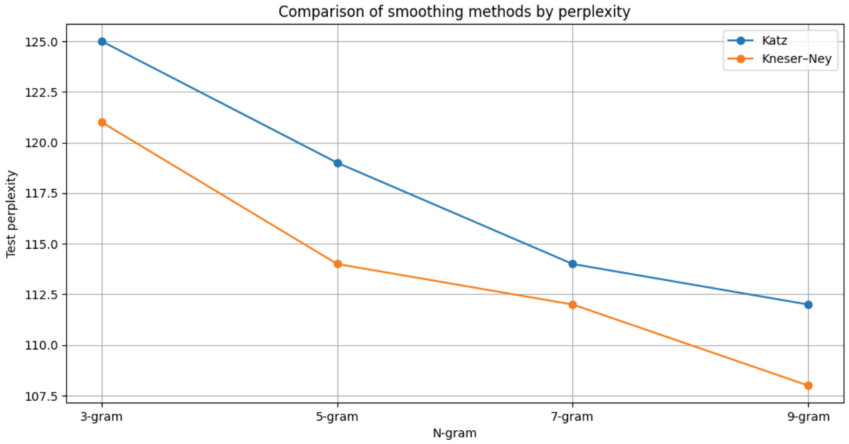


Fig. 5. Comparison of smoothing methods based on perplexity

5 Conclusion

This study presented a detailed analysis of the stages and technical aspects involved in building an n-gram-based language model for the Uzbek language, with a focus on the tools employed. Key stages such as data cleaning, n-gram frequency computation, probability smoothing, and building the model in a resource-efficient manner were carefully addressed. The use of toolkits such as IRSTLM and KenLM was compared, highlighting their respective strengths and weaknesses through practical experiments. Since KenLM provides a full implementation of Kneser-Ney smoothing, the probabilities generated by KenLM on the same corpus may be somewhat more reliable compared to IRSTLM. Therefore, in building the final model, we based our approach on the probabilities obtained from KenLM, while also comparing the models constructed using both toolkits. Specifically, IRSTLM was noted for its memory efficiency, while KenLM demonstrated superior performance and accuracy. Notably, KenLM's implementation of the modified Kneser-Ney smoothing algorithm produced high-quality results and

achieved favorable perplexity scores on large corpora. The paper also presents corpus frequencies and rare occurrences. In particular, the distribution of N-gram counts is provided. The findings confirm that an effective n-gram model can be successfully built for Uzbek textual data. Such a model can serve as a foundation for various natural language processing tasks, including machine translation, text generation, text analysis, and word prediction.

References

1. Shannon, C.E.: A mathematical theory of communication. Bell Syst. Tech. J. **27**(3), 379–423 (1948)
2. Davies, M.: Google Books. American English Corpus Project, Brigham Young University (2011). [http://googlebooks.byu.edu/]
3. Jurafsky, D., Martin, J.H.: N-gram language models. Speech and language processing, 23 (2018)
4. Zhang, W.: Comparing the Effect of Smoothing and N-gram Order: Finding the Best Way to Combine the Smoothing and Order of N-gram (2015)
5. https://www.geeksforgeeks.org/nlp/discounting-techniques-in-language-models/
6. Boltayevich, E.B., Ilxomovna, A.X., Hakim qizi, P.M., UktambayO‘g‘li, K.N.: semantic differentiation of Uzbek homonyms using the Lesk algorithm. In: 2023 8th International Conference on Computer Science and Engineering (UBMK), pp. 137–140. IEEE, Burdur (2023). https://doi.org/10.1109/UBMK59864.2023.10286666
7. Elov, B.B., Primova, M.H.: Development of an information system for organizing and managing the educational process based on smart technologies. In: XVI International Scientific and Practical Conference “State and Prospects for the Development of Agribusiness—INTERAGROMASH 2023”. https://doi.org/10.1051/e3sconf/202341303010
8. Nadas, A.: Estimation of probabilities in the language model of the IBM speech recognition system. IEEE Trans. Acoust. Speech Signal Process. **32**(4), 859–861 (1984)
9. Church, K.W., Gale, W.A.: A comparison of the enhanced Good-Turing and deleted estimation methods for estimating probabilities of English bigrams. Comput. Speech Lang. **5**(1), 19–54 (1991)
10. Park, J.H., Song, Y.I., Rim, H.C.: Smoothing algorithm for n-gram model using agglutinative characteristic of Korean. In: International Conference on Semantic Computing (ICSC 2007), pp. 397–404. IEEE (2007, September)
11. Chen, S.F., Goodman, J.: An empirical study of smoothing techniques for language modeling. Comput. Speech Lang. **13**(4), 359–394 (1999)
12. Mani, S., Gothe, S.V., Ghosh, S., Mishra, A.K., Kulshreshtha, P., Bhargavi, M., Kumaran, M.: Real-time optimized N-gram for Mobile devices. In: 2019 IEEE 13th Int. Conf. Semantic Computing (ICSC), pp. 87–92. IEEE, Newport Beach (2019)
13. Ismail, R.: Comparison of modified Kneser-Ney and Witten-bell smoothing techniques in statistical language model of Bahasa Indonesia. In: 2014 2nd Int. Conf. Information and Communication Technology (ICoICT), pp. 409–412. IEEE, Bandung (2014)
14. Heafield, K., Pouzyrevsky, I., Clark, J.H., Koehn, P.: Scalable modified Kneser-Ney language model estimation. In: Proc. 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), pp. 690–696. ACL, Sofia (2013)
15. Heafield, K.: KenLM: faster and smaller language model queries. In: Proc. Sixth Workshop on Statistical Machine Translation, pp. 187–197. ACL, Edinburgh (2011). [https://aclanthology.org/W11-21]



Sustainable Water Forecasting and Wastewater Management Using Fusion-Based AI and Smart Sensors

R. C. Jeni Gracia(✉) and G. Michael

Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences (SIMATS), Saveetha University, Chennai, India
jenigracia89@gmail.com, michaelg.sse@saveetha.com

Abstract. Water resource management in urban areas relies on precise demand prediction, in-time anomaly monitoring, and intelligent decision-making. This study develops a Fusion-based hybrid framework based on deep learning and probabilistic reasoning for improved water demand prediction and wastewater system management. Fusion-based model integrates Long Short-Term Memory (LSTM) for temporal water demand forecasting, Convolutional Neural Networks (CNN) to analyze spatial patterns, and Autoencoders (AEs) for unsupervised anomaly detection, with the emphasis on the detection of potentially contaminated water, and also the framework includes a Bayesian Network-based decision support system. Through experiments with synthetic and real datasets, the Fusion-based methodology achieves significantly better prediction accuracy and anomaly detection of water contamination precision, and is more reliable for decision-making than single model initiatives. A hybrid model with smart sensors provides an auditable, scalable, and smart approach for sustainable urban water management.

Keywords: Long Short-Term Memory (LSTM) · Convolutional Neural Networks (CNN) · Autoencoders (AEs) · Bayesian Network · decision support system · Water Management

1 Introduction

The increasing demand for water arises due to population expansion, climate change, and ageing infrastructures require the development of smart and adaptive water resources management approaches. In order to provide sustainable and safe water, it is essential not only to predict future water demand but also to identify contamination in water. Classical rule-based or statistical methods are not always suitable to address the complexity, uncertainty, and dynamism of real-world water systems. IoT sensors embedded into smart water systems for real-time information like Water Quality parameters (pH, Chlorine, Turbidity, etc) and support data-based decision making for utility providers.

In the traditional method, the Autoregressive Integrated Moving Average (ARIMA) model has been largely used to predict water demands, which are based on the time series, and it also has good predictive performance rates for short-term prediction, it suffers

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026
S. Lanka et al. (Eds.): ICTSM 2025, LNNS 1755, pp. 358–369, 2026.
https://doi.org/10.1007/978-3-032-13177-5_28

when dealing with nonlinear trends, missing data, and abrupt changes in demand due to holidays, Festivals, or weather abnormalities. Likewise, Artificial Neural Networks (ANN) have been extensively used to predict water demand and identify anomalies due to their ability to capture nonlinear relationships (modern). However, the traditional ANN architectures lack inherent memory in time, and often over fit in the situation of long-sequence computation and outliers in unstructured anomaly detection.

To mitigate the limitations of such a baseline model, recent development in deep learning and probabilistic models introduces a great hope for the complex spatial-temporal dynamics and uncertainty in the environmental systems. The model is based on a combination of CNN for the identification of spatial patterns of water and LSTM for the prediction of temporal water demand, and Autoencoders for unsupervised anomaly detection. In fact, a Bayesian Network is included as a decision support layer to initiate probabilistic reasoning under uncertainty to achieve intelligent responses like contamination risk assessment and recommended treatment. Smart water framework that leverages recently developed AI, IoT technologies for sensor hardware to measure and monitor water quality and associated environmental parameters at high spatial and temporal resolution. Multimodel framework for the prediction of water demand and an integrated intelligent system when anomalies are detected, starting the treatment processes, or switching to other sources of water. Through the combination of hierarchical AI realisation, scaling, and heterogeneity, the framework-based approach adapts assurance levels to support data decisions on water quality parameters. Autoencoders, which are unsupervised architectures for learning efficient representations of the input data, are also used in the system to improve system monitoring and contamination detection. An autoencoder is trained on "clean" water data (meaning data with normal water quality), including pieces of information such as pH, turbidity, chlorine. It creates a high reconstruction error when the autoencoder is exposed to contaminated data, and the error can be used as an anomaly score. This approach allows for the identification of events such as water pollution. A Bayesian Network Decision Support System (DSS) is constructed (Fig. 1) to capture the probabilistic dependence between system variables (sensor observations, model predictions, environmental context etc.) through a directed acyclic graph.

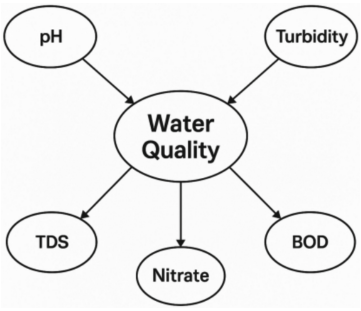


Fig. 1. Bayesian Network-based Decision Support System

2 Related Works

The multichannel deep learning hybrid model [1] of short-term urban water demand forecasting. Multiple data feeds, including weather and historical demand, were combined in the study to enhance predictive power. A convolutional neural network model to predict water consumption at the home level [2]. The results demonstrated the significant superiority to ordinary regression models. The separation was based on the contribution of the deep learning model to personalized water demand management. ML framework for predicting water demand of crops [3] from bare minimum climatological data.. It was concluded in that study that there is a direct link between demand prediction as a function of the climatological parameters, using agricultural water needs. This model particularly useful in water-scarce farming regions. A transient water-demand prediction algorithm based on ensemble learning [4]. Machine learning power for prediction [5] of energy consumption in inter-basin water transfer. It was a study in the interconnectedness of water and energy. Thus optimizing the macro water transfer systems, then by the use of a predication model. Three-hour ahead prediction of urban water demand by the DL method [6]. In contrast to predictions one step into the future, the models derived estimates of season and day of exposure on a daily and seasonal basis. The test results indicated the extreme improvements in the total forecast errors. A deep learning model for short-term water demand prediction [7] and the model accurately predicted demand under changing urban scenarios. This further demonstrates that deep learning is an appropriate technique for the time series forecasting of complex biosignals. [20] analyzed deep learning methods to predict the water demand for IoT-based systems. The research focused on AI integration with smart sensors and communication networks. The adaptive response to recent consumption was improved, as indicated by the results. Such an approach is scalable in comparison with water infrastructures in modern times [8, 15].

DLGEA [9], a deep learning-guided Evolutionary Algorithm (EA) for water quality to detect water contamination. The proposed hybrid approach combined learning with optimization to enhance the accuracy. It has been used to solve sophisticated hydraulic networks by tracing the contamination source. Unusual [10] can be found even when there is only a sparse amount of domain knowledge. It illustrated the capacity of ML to reduce the non-deterministic nature of environmental variables. The deep learning methods for pollution source tracing and water quality prediction. The model was designed to account for non-linearities in cross-pollutant distribution, good predictors of contamination [11]. The study highlighted the importance of DL in the area of environmental protection. It offers actionable knowledge for early-intervention methods. Deep learning for waterborne pollution analysis in IoT-based Aquaculture. The results showed that the method can model the dynamic change process of water condition enhanced real-time monitoring integrated with IoT [12]. A study of machine learning [13] in the prediction of water quality. Different models, including regression, the SVM model, and ensemble, were used. Better predictive validity for advanced ML models. The study also highlights the necessity of feature selection in water quality modeling. Machine learning based model for water quality classification [14].

The system categorized water samples based on their pollution category. The Experiments were successful in determining source water was safe or unsafe. This software

uses an algorithm to assist councillors or staff members in making a judgment about the appropriate category. Reported WQI prediction models based on a machine learning approach. AI-based water treatment [16] and desalination review. The summary offered a broad look at AI-driven process optimization. It saw an opening for predictive maintenance and energy conservation. It was about how AI is emerging as a vital ingredient for sustainable desalination. It also cited obstacles in using the drug more widely. [17] investigated AI in the paradigms of the nanomembrane systems for enhancing desalination. It would result in the following: (i) better water recovery and (ii) lower energy consumption. It is at the technological frontier in treating water. [18], offered an extensive overview for AI based hybrid reverse osmosis systems. The research explored the use of AI to holistic supervision and operating optimization. For the sake of comparison, it is reported that the proposed method can achieve the sample quality of the water treatment with lower consumption of energy. This is the initial phase of a waste-free desalination process of the future. [19] contains an integrated summary of how AI remodeled the water treatment approaches and desalination processes. Optimization, control, and energy saving on resource wastage. It highlighted the disruptive impact of AI on seawater desalination.

3 Proposed Work

The proposed system is a fusion-based AI (Fig. 2) system designed to integrate water demand forecasting, contaminated detection, contamination risk assessment, and responsive water treatment, including desalination, into a unified intelligent control. The architecture has the five key modules: (i) data acquisition and preprocessing, (ii) deep learning model fusion, (iii) anomaly detection via autoencoding, (iv) Bayesian network-based decision support, and (v) treatment response recommendation, including automated desalination initiation. An overview of the system architecture is shown in Fig. 2.

3.1 Data Collection and Preprocessing

The data are obtained from a smart distributed sensor system implemented throughout the water-distribution and wastewater system. These detectors record data periodically for recent and historical readings that can be used to assess water quantity and quality, for example:

(a) Water Consumption data that was collected from the past historical usage.
(b) Water quality parameters: pH, turbidity, total dissolved solids (TDS), and chlorinity.
(c) Environmental context: temperature, rain, and feed intake histories
(d) Preprocessing includes normalization, outlier removal, timestamp joining, and interpolation for handling missing values. Make the data clean and most efficient to consume by deep learning models.

3.2 LSTM Model for Water Demand Forecasting

A Long ShortTerm Memory (LSTM) neural network is integrated for temporal dependencies and periodic behaviors in water consumption patterns. LSTM is used for multivariate sequential data and captures trends influenced by daily cycles, seasonal usage,

or external environmental factors. Inputs include historical usage, temperature, rainfall, and time-of-day indicators. The model outputs near-future predictions (hourly/daily) of water demand in an area, which are used to detect abnormal usage spikes.

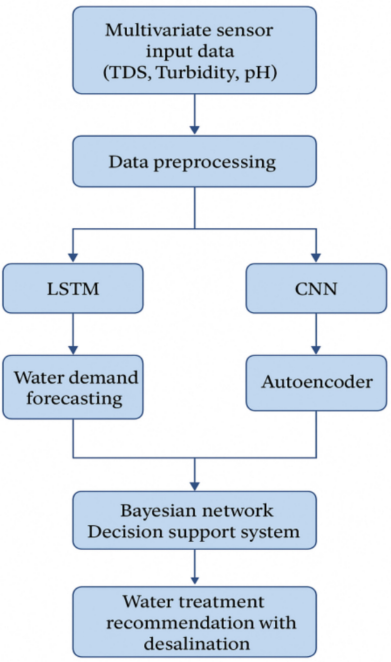


Fig. 2. Fusion-based AI architecture

3.3 CNN for Spatial Pattern Recognition

To detect the contamination of water usage in an area, a Convolutional Neural Network (CNN) is integrated. It receives Multi-sensor data (such as pH, turbidity, total dissolved solids (TDS), and chlorinity) arranged in spatial grids corresponding to sensor layout. The CNN identifies spatial correlations of wastewater used by the people in a particular area.

3.4 Autoencoder for Anomaly Detection

An Unsupervised Autoencoder is used to learn the normal behaviors of water quality parameters. When trained on normal system functions, it is capable of reconstructing the input data, and of measuring the fitness with respect to the input data. If there are large discrepancies from the expected value (e.g., a spike of turbidity, a decrease in chlorine, or a strange TDS level) an alert is raised. A special monitoring service is constantly observing the reconstruction error of the autoencoder. The system interprets this threshold error as a potential anomaly, the threshold is surpassed (it is an if-then type logic), and sends it back to the Bayesian decision support system for reasoning.

3.5 Bayesian Decision Support Network

The system includes a Bayesian Network (BN) that it uses to reason under uncertainty and to make explainable, probabilistic decisions using sensor-derived and learned model outputs as evidence. Conditional dependencies among domain-independent priors are modeled by using the data-driven priors and domain knowledge used to construct the BN structure by: a) Input Nodes: sensor values in real time (pH, turbidity, chlorine, TDS, flow, pressure), Forecasting outputs (strains of the LSTM), Anomaly scores (the Autoencoder), The specifics of the environment (rainfall, time-of-day, recent repairs) b) Latent Nodes: The risk of contamination (e.g. microbial or saline intrusion), risk of failure of infrastructure, and degree of deviation of operating norms. c) Output Nodes: Level of Risk (Low, Moderate, High), Recommended reaction (e.g., isolate area, divert circulation, start treatment)

3.6 Response Recommendation Including Desalination

Another key upgrade of this frame is that the desalination option should be automated into an automatic decision layer, and that is particularly suited to deal with contamination. In the proposed smart water management system, desalination occurs because of combining a Reverse Osmosis (RO) unit, which becomes the main mode of treating wastewater in case contamination is identified by the Bayesian Network. RO is a high-pressure membrane filtration technology that strips dissolved salts and contaminants from water to produce clean water by pushing water through a semi-permeable membrane. Only water molecules can penetrate this membrane, whereas ions, heavy metals, and other dissolved contaminants will be rejected. Modular nature, ability to be triggered automatically, and real-time control over the process make it very appropriate to use in systems integrated with AI to make contaminated water into fresh water.

4 Result and Discussion

The proposed hybrid model uses Long ShortTerm Memory (LSTM) to predict the water demand on an hourly basis using the 24-h consumption data and contextual parameters (i.e., temperature and time of the day). This method works well in describing the temporal relationships occurring in the water usage pattern, producing estimated values that are close to the actual values obtained. The total water demand during the observation period was 23,405 l of water. The hourly consumption was highest during 11.00 to 13.00, during which the consumption was about 1100 l, and the lowest consumption was experienced between 03.00 and 05.00 with an average consumption of between 860 and 880 l.

In the proposed system, CNN collects the sensor data, and the anomaly detection in the water quality is based on the models that learned everything about the normal water quality patterns that including pH, Total Dissolved Solids (TDS), turbidity, and chlorine levels, etc. The autoencoders marked the instance as anomalous whenever the water indicates contamination, i.e. difference between the actual and predicted values of the sensor, was larger than a predefined threshold. Significantly, anomaly detection for water demand (Table 1) were observed within the interval of 13:00–16:00. The major

pointers that contributed to these findings were; relatively higher levels of TDS, which indicates the issue with salinity; turbidity levels over 5 NTU, which denotes the potential presence of physical contamination or microbial contamination; and the lower levels of chlorine under 0.2 mg/L, which suggests disinfection, the causes of such anomalies of non-treated waste waters and it is necessary to take actions like treatments with ROs.

To improve the water quality in case of high Total Dissolved Solids (TDS) levels identified during operation at certain hours, the system used a Bayesian Network to suggest such steps as the enablement of Reverse Osmosis (RO) desalination. The basis behind this recommendation was the probability reasoning based on sensor measurements. The process of RO treatment was launched and the efficiency of its recovery was about 90% of the polluted water. As an illustration for TDS, the system was able to reclaim 972 l of potable water at 13:00 when the system water demand stood at 1080 l with a TDS of 1600 ppm. Throughout the four-hour aberrance period, 3,744 l of water were re-utilized, which represents an absolute reclamation strain on freshwater resources and proves the possibility of smart wastewater resumption in real time being efficient.

Table 1. Predictive Accuracy of fresh water in 4 h in a day

Time	Water Demand (L)	TDS (ppm)	RO Reclaimed (L)
13:00	1080	1600	972
14:00	1050	1700	945
15:00	1030	1550	927
16:00	1000	1200	900

Almost 16% of the daily water requirement, or 3,744 l of drinking water, was reclaimed by the system through the application of RO treatment. This illustrates how the model has been effective in converting the polluted wastewater to clean water. The framework combines the resilience and sustainability of the water management systems in urban areas through automated smart detection of anomalies and subsequent triggering of desalination..

Table 2. Comparison of ARIMA, ANN, and Proposed Hybrid Model

Method	RMSE (L)
ARIMA	65.2
ANN	52.4
Proposed Hybrid	38.1

The comparison of the experiment in Table 2 shows an evident advantage of the suggested hybrid model in both forecasting and intelligent decision support. Using LSTM, the model provides a superior performance in terms of time than the use of classic algorithms such as ARIMA and ANN.

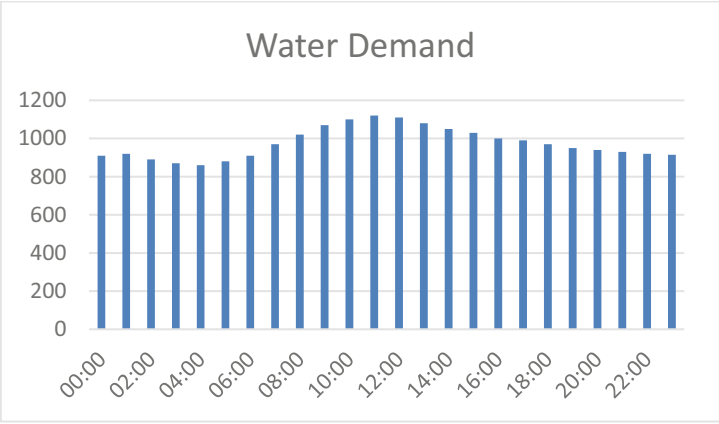


Fig. 3. Water Demand Forecasting for 24 h

Data of the hourly water demand discloses a clear picture (Fig. 3) of the daily water consumption using LSTM. During the very early morning hours (00:00 to 06-00) there appears a minimal consumption (860–920 l), and at 07-00 the consumption starts its rapid growth and reaches its top with the highest demand at 11-00 (1120 l). Between 12:00 and 15:00, the demand is relatively high, that is, it decreases steadily with time or in terms of liters from 1110 to 1030. Figure 3 shows that the usage will continue during the course of the day as far as the operations are concerned. After 16:00 and in the evening (16:00–21:00), the water use continuously decreases to 930 l, and it ends up settling at 920–915 l in the late evening (22:00–23:00). The hourly water demand is 23,405 l, proving the existence of a definite time pattern that indicates the necessity of time-sensitive forecasting and dynamic management of water resources.

Table 3. Performance of Predicting Water Demand

Model	MAE (L)	RMSE (L)	MAPE (%)
ARIMA	24.3	28.6	2.95
ANN	18.6	21.4	2.10
Proposed Hybrid	14.8	16.1	1.48

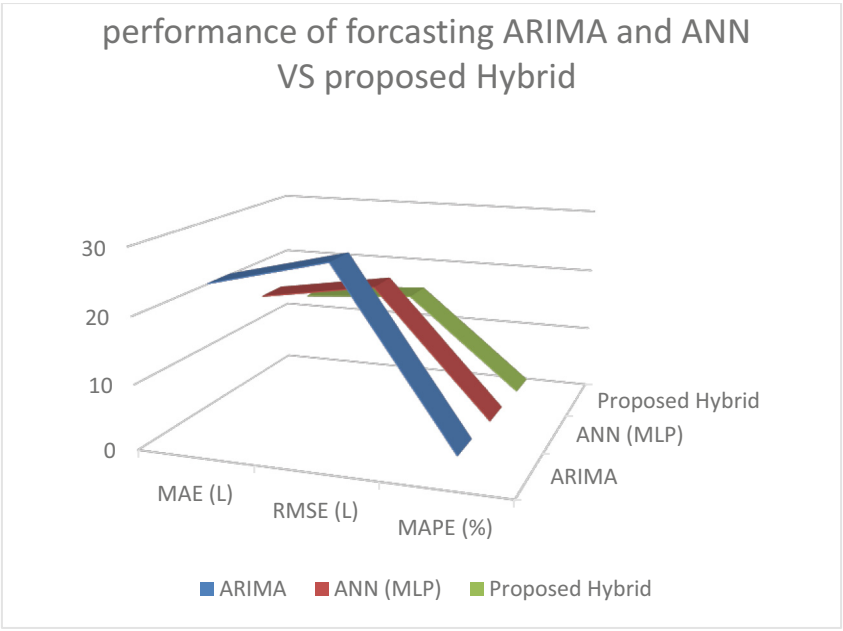


Fig. 4. Forecasting comparison for Hybrid model vs. ARIMA and ANN

The analysis on the basis of MAE, RMSE, and MAPE (Table 3) illustrates the excellence of the suggested combined model in comparison with the conventional strategies (Fig. 4). The higher the MAE is (14.8 l), the more consistent and closer to the real predictions of the hybrid model are than earlier models such as ARIMA and ANN that reported higher values of errors.

$$\text{MAE} = \frac{1}{n} \sum | \, y_i - \hat{y_i} \, | \tag{1}$$

The Root Mean Squared Error (RMSE), reports the lowest value of 16.1 l in favor of the hybrid method, and it indicates less unfavorable treatment of the fluctuating demands. Moreover, the Mean Absolute Percentage Error (MAPE) of only 1.48% supports that the model will be quite accurate at different consumption levels, and the hybrid model is an acceptable and effective instrument in predicting water demands within smart urban areas.

$$RMSE = \sqrt{\frac{1}{n} \sum \left(y_{i-} \hat{y_i} \right)} \tag{2}$$

$$MAPE = \frac{100}{n} \sum | \, \frac{yi - \hat{yi}}{yi} \, | \tag{3}$$

Table 4 summarizes the anomaly detection and identifies the potential of the proposed hybrid system. A value of 0.75 in precision means that out of the flagged anomalies, 75% of them are actual positives. The recall of 1.00 indicates that the system was able

to detect all the true anomalies, hence no event of contamination was lost. The overall performance after balancing the precision and the recall of the model by using the F1-score value of 0.857 proves the high level of performance in the areas of correctness and completeness of the anomaly detection. Finally, the accuracy of the hybrid model is 0.96, which reflects that the system identified the correctness in 96% of all the instances, proving its reliability and effectiveness in real-time monitoring of the water quality.

Table 4. Performance Metrics

Metric	Score
Precision	0.75
Recall	1.00
F1-Score	0.857
Accuracy	0.96

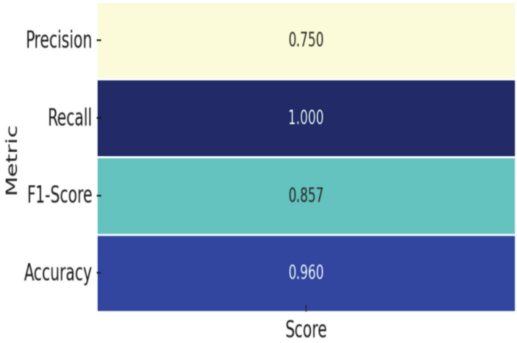


Fig. 5. Anomaly detection in contaminated water

The metrics provide adequate measures (Fig. 5) to justify the capacity of the system to accurately identify anomalies (high recall), reduce the number of false positives (high precision), and ensure the overall performance of the classification (high F1-score and accuracy). The system will be able to make prompt and knowledgeable decisions, including the activation of RO desalination, by measuring the accuracy of the detection, thus contributing to protecting the health of people, improving the work of systems, and ensuring automated water systems in smart cities.

$$Efficiency(\%) = \left(\frac{Water\ Desalinated}{Water\ Demand} \right) \times 100 \tag{4}$$

RO desalination efficiency is the performance of any kind of water treatment system in the conversion of polluted or waste water to clean usable water by way of Reverse Osmosis. It is taken as a ratio of reclaimed water to the sum of water requirements. In the case of an RO process, when 3,744 l of water are reclaimed out of a daily total overall

demand of 23,405 l, the recovery is about 16%. Real-time anomaly detection is one of the greatest additions to RO in a smart urban system, since it enables the desalination process to not only be efficient, but to be activated in a smart way to react to the real water quality issues.

$$\text{Detection Latency} = T_{\text{sensor}} + T_{\text{processing}} + T_{\text{decision}} \tag{5}$$

T_{sensor}: Time spent by sensors to gather and push data to the model, $T_{\text{processing}}$: Time to process the data and identify anomalies by the model (e.g., Autoencoder), T_{decision}: Time taken by the decision support system (e.g., Bayesian Network) spends to issue a decision (e.g., RO activation)

Detection is the period of time it will take a smart water management system to detect a water quality anomaly and to designate a corrective mechanism to occur, such as to turn on RO desalination. It is estimated as the summation of the inference time of the sensors capturing data, the processing of the presence of anomalies by such models as Autoencoders, and the Bayesian decision-making.

5 Conclusion and Future Work

The proposed fusion system artificial intelligence on a Water demand prediction and wastewater management by using smart sensor networks and multi-dimensional deep learning models. One method leverages the strength of the LSTM networks to forecast the temporal demand, CNN to recognize spatial patterns, and Autoencoders to identify anomalies in an unsupervised manner, and the integration of a Bayesian Network-based decision support system which makes it possible to apply a probabilistic reasoning solution under uncertainty that provides recommendable explanations in wastewater management. Experimental results show that the proposed approach can achieve significant improvements in predictive accuracy, anomaly-detection precision, and decision robustness over single model methods. Despite its promising performance, the framework encountered a few problems. These are the need for high-quality transparent data to train on, the challenge of integrating real-time decision making into legacy infrastructure, and the overhead of stitching models in real time. Future work can be scaling up the dataset of real contamination events and multi-utility networks. Learning Reinforcement to Learn to Optimize Dynamic Response Policies Leveraging lightweight AI models on edge computing. Wire the system to the SCADA and GIS framework and conduct large-scale field trials, and include explainability to gain performance.

References

1. Namdari, H., Ashrafi, S.M., Haghighi, A.: Deep learning–based short-term water demand forecasting in urban areas: a hybrid multichannel model. AQUA—Water Infrastruct Ecosyst. Soc. **73**(3), 380–395 (2024)
2. Kim, J., Lee, H., Lee, M., Han, H., Kim, D., Kim, H.S.: Development of a deep learning-based prediction model for water consumption at the household level. Water. **14**(9), 1512 (2022)
3. Sidhu, R.K., Kumar, R., Rana, P.S.: Machine learning based crop water demand forecasting using minimum climatological data. Multimed. Tools Appl. **79**(19), 13109–13124 (2020)

4. Huang, H., Zhang, Z., Song, F.: An ensemble-learning-based method for short-term water demand forecasting. Water Resour. Manag. **35**(6), 1757–1773 (2021)
5. Yi, S., Kondolf, G.M., Sandoval-Solis, S., Dale, L.: Application of machine learning-based energy use forecasting for inter-basin water transfer project. Water Resour. Manag. **36**(14), 5675–5694 (2022)
6. Sahoo, B.B., Panigrahi, B., Nanda, T., Tiwari, M.K., Sankalp, S.: Multi-step ahead urban water demand forecasting using deep learning models. SN Comput. Sci. **4**(6), 752 (2023)
7. Guo, G., Liu, S., Wu, Y., Li, J., Zhou, R., Zhu, X.: Short-term water demand forecast based on deep learning method. J. Water Resour. Plan. Manag. **144**(12), 04018076 (2018)
8. Mohammed, H., Hameed, I.A., Seidu, R.: Machine learning: based detection of water contamination in water distribution systems. In Proceedings of the Genetic and Evolutionary Computation Conference Companion, pp. 1664–1671 (2018, July)
9. Qian, K., Jiang, J., Ding, Y., Yang, S.H.: DLGEA: a deep learning guided evolutionary algorithm for water contamination source identification. Neural Comput. & Applic. **33**(18), 11889–11903 (2021)
10. Muharemi, F., Logofătu, D., Leon, F.: Machine learning approaches for anomaly detection of water quality on a real-world data set. J. Inf. Telecommun. **3**(3), 294–307 (2019)
11. Wang, J., Xue, B., Wang, Y., Wang, G., Han, D.: Identification of pollution source and prediction of water quality based on deep learning techniques. J. Contam. Hydrol. **261**, 104287 (2024)
12. Arepalli, P.G., Naik, K.J.: Water contamination analysis in IoT enabled aquaculture using deep learning based AODEGRU. Eco. Inform. **79**, 102405 (2024)
13. Ahmed, A.N., Othman, F.B., Afan, H.A., Ibrahim, R.K., Fai, C.M., Hossain, M.S., et al.: Machine learning methods for better water quality prediction. J. Hydrol. **578**, 124084 (2019)
14. Muhammad, S.Y., Makhtar, M., Rozaimee, A., Aziz, A.A., Jamal, A.A.: Classification model for water quality using machine learning techniques. Int. J. Softw. Eng. Appl. **9**(6), 45–52 (2015)
15. Uddin, M.G., Nash, S., Rahman, A., Olbert, A.I.: Performance analysis of the water quality index model for predicting water state using machine learning techniques. Process Saf. Environ. Prot. **169**, 808–828 (2023)
16. Alenezi, A., Alabaiadly, Y.: Artificial intelligence applications in water treatment and desalination: a comprehensive review. Water. **17**(8), 1169 (2025)
17. Krishnan, A., Sundaram, T., Nagappan, B., Devarajan, Y.: Integrating artificial intelligence in nanomembrane systems for advanced water desalination. Results Eng. **24**, 103321 (2024)
18. Abobeh, S., Abdelsalam, M.M., Bouz, M.A.E.: An overview study on water desalination, hybrid RO systems, and how to utilize artificial intelligence with the system. Water Conserv. Sci. Eng. **10**(2), 85 (2025)
19. Ray, S.S., Verma, R.K., Singh, A., Ganesapillai, M., Kwon, Y.N.: A holistic review on how artificial intelligence has redefined water treatment and seawater desalination processes. Desalination. **546**, 116221 (2023)
20. Jeni Gracia, R.C., Narayanan, L.K.: Deep learning approaches for forecasting water demand in IoT-based water systems. In: 2024 IEEE 16th International Conference on Computational Intelligence and Communication Networks (CICN), pp. 262–266. IEEE (2024, December)

Securing AI-Driven Biometric Authentication Systems: Cryptographic Techniques for Protecting Sensitive User Data

Sahar Ebadinezhad[1,2(✉)], Fulbert Milrich Kiminou Nsimba[1], Hauwa Adamu Wakil[1], and Joyeuse Maombe[1]

[1] Department of Computer Information Systems, Near East University, Nicosia 99138, North Cyprus
sahar.ebadinezhad@neu.edu.tr, {20243764,20245381, 20244054}@std.neu.edu.tr

[2] Computer Information Systems Research and Technology Center (CISRTC), Near East University, Nicosia, Turkey

Abstract. This study investigates how artificial intelligence (AI)-driven biometric authentication systems can be made more secure using cryptographic techniques. Increasing the level of safety of biometric identification systems that are managed by artificial intelligence (AI) is the objective of this project. The aim of this work is to investigate the several ways in which cryptographic techniques can be applied to improve the security of biometric authentication systems driven by artificial intelligence (AI), especially those depending on facial recognition for authentication needs. Biometrics are increasingly used in critical sectors such as healthcare, banking, and border security; hence, these systems are vulnerable to several major risks, including sophisticated assaults, data breaches, adversarial manipulation, and repeat attacks. These are only a few of the main risks these systems run across. This research involves assessing several modern cryptographic technologies, including homomorphic encryption, cancellable biometrics, blockchain integration, and biometric template protection, in order to handle these problems. The goal of the project is to protect biometric data and respect personal information confidentiality. A prototype system designed to encrypt facial feature vectors obtained from a dataset drawn from the real world was built in Python. To get the required performance, this system used the K-Nearest Neighbor's (KNN) classifier with Fernet symmetric encryption. The results show that encryption did not cause a reduction in model accuracy (60.87%), implying that intelligence-based recognition systems can preserve high performance and simultaneously protect data confidentiality. Cryptography was used to keep the data safe. The results of this study show that using lightweight cryptography is a good way to make biometric login systems that protect user privacy better and more scalable.

Keywords: Artificial Intelligence (AI) · Biometric Authentication · Cryptographic Techniques · Data Security · Privacy Protection · Sensitive User Data

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026
S. Lanka et al. (Eds.): ICTSM 2025, LNNS 1755, pp. 370–379, 2026.
https://doi.org/10.1007/978-3-032-13177-5_29

1 Introduction

Numerous firms employ biometric identification for security. See who someone is, pick who may view something, and see who a user is. Eye scans, fingerprints, and facial patterns are utilized to verify identities. Most individuals today use passwords or PINs, which are safer and more private.

As AI grows, biometric safety has evolved in various ways. AI strengthens privacy restrictions and makes identity disclosure easier [1]. AI and ML systems may detect scams, disguise biological data, and adapt swiftly. Many individuals know that multimodal biometric systems that use AI fusion operate better with face and other biometrics. These technological advances could make identity systems stronger, more cyberattack-resistant, and less likely to be stolen [2].

People are more concerned about protecting their private data and how biometric authentication systems could misuse it after recent events. Scary people have increased. Face- and fingerprint-reading technology may be better. They still face changing online threats like fraud and data theft. It is vital to note that mixed-modal methods make facial recognition system safety tougher. Explain how much harder it is to keep children safe. Biometric data is used in the cloud, so it's crucial to protect it [3]. The study that looked into encryption techniques based on biometrics was done to make cloud data safer [4].

The contribution and investigation of this study are toward the characteristics of artificial intelligence-powered biometric identification systems, with an emphasis on voice, fingerprint recognition, and face recognition. It examines the benefits and drawbacks of cryptography and computational intelligence for data security. The study's goal and contribution to this conference is to create biometric systems that are secure, dependable, and sensitive enough to survive cybersecurity threats. The study will look at biometric data encryption, multifaceted fusion, and robust AI frameworks to find optimal procedures and gaps in past research.

2 Literature Review

To increase the accuracy of identity verification, multimodal biometric systems employ a variety of biometric characteristics, such as voice recognition, fingerprints, iris patterns, and facial features. These systems are capable of handling a number of issues, including noisy data, spoofing assaults, and partial occlusions. Fusion techniques such as feature-level fusion, score-level fusion, and decision-level fusion are employed to enhance facial recognition performance [2]. By using these techniques, systems are better able to manage presentation attacks, adapt to changes in location, illumination, and facial expressions, and decide whether to let users in or out. Additionally, layered analysis improves individual identification [3].

AI has transformed biometric authentication systems, increasing accuracy and efficiency. [2] state that biometric modalities such as patterns of speech, fingerprints, iris scans, and recognition of faces are important in security applications. Nevertheless, biometric information is sensitive and requires strict security precautions. Multimodal biometric systems that incorporate several biometric traits improve accuracy in recognition and system robustness. Despite advancements, security concerns remain about storing and distributing biometric data, especially in the cloud [4]. Advocates of encryption solutions that combine biometric technology and machine learning methodologies to improve data security.

2.1 Digital Solutions for Biometric Verification Security

Biometric risks are getting more complicated, and cryptography is becoming an important part of making security better. No one can get to biometric information without permission, which is why cryptography is used to protect both the integrity of the data and its safety [5]. Biometrics that use templates that can be stopped can help build a base for access that can be safely taken away. To protect the biometric identity that was made possible by [4]. Blockchain and encryption technologies were used together. It is getting more and more important to find a good mix between security, accuracy, and speed if want to make authentication work in real life [6].

With homomorphic encryption, biometric comparisons can be done directly on encrypted data. This keeps user privacy safe even when the data is kept in the cloud. Through the utilization of this technique, [7] used a fingerprint authentication system that was developed while simultaneously preserving both its precision and its speed. Authors in [8] demonstrated that homomorphic models make it possible to match biometric information safely in circumstances where there is no trust, without disclosing data that has not been through any processing. The protection of user data during live authentication events is made possible by the utilization of homomorphic encryption in conjunction with smart identification systems [9].

2.2 Cryptographic Solutions for Securing Biometric Authentication

Using secure technologies has become an important part of making things safer. The reason for this is that biometric risks are getting more complicated all the time. Cryptography helps keep data safe and stops people from getting into private personal information without permission. Using biometrics and templates that can be cancelled together can help set up a basis for access that can be taken away safely [5]. In order to get the effect, this can be done [9]. To protect the biometric verification that was controlled by AI, the technologies of blockchain and encryption were brought together. This had to be done to make sure that the authentication was safe. [6] said it was important to find a balance between security, accuracy, and performance so that the goal of giving authentication that can be trusted in real life could be reached.

2.3 Adaptive and Multi Factor Encryption

The levels of protection that adaptive encryption provides can be changed in a flexible manner, depending on the actions of the user or the nature of the threat. Authors in [10] highlighted that artificial intelligence has been used to develop a concept in which the strength of encryption is altered in response to the identification of hazards and the activity of logins. Shethiya (2024) advocated that in order to increase the level of security, it would be beneficial to combine biometric login with either device-level authentication or physical tokens together [1]. Also, another study [5] provided an architecture that is multidimensional and makes use of passwords, devices, and biometrics combined with one another to deliver a secure multilayer defense environment. Moreover, the blockchain technology not only protects the data of biometric systems from being altered or rolled back, but it also provides audit trails that cannot be altered [11]. It is necessary to create a system that is founded on blockchain technology and that enables the recording of each and every biometric encounter. An improvement in the data traceability of identity verification systems was achieved through the utilization of a mix of homomorphic encryption and blockchain technology [9]. The technology known as blockchain has been discovered to have the capability of facilitating biometric security solutions that are distributed, transparent, and compliant [3].

3 Proposed Methodology

This section outlines the procedural procedures undertaken to secure a facial recognition system utilizing cryptographic methods. The process comprises data preparation, model training, encryption, and performance evaluation. The dataset used for this research is LFW dataset by using synthetic or lower-dimensional mock facial vectors [12].

With each row containing 4096 pixel-based attributes extracted from facial photos, we used a sanitized dataset of Bayern Munich football players. There are 230 samples in the dataset split across 5 classes. To enable performance evaluation on unseen data, the dataset was split 80/20 into training and testing sets. It was necessary to make use of the training component of the dataset in order to train a K-Nearest Neighbors (KNN) classifier for face recognition.

For the encryption process, the Fernet symmetric encryption method was employed to encrypt each and every feature vector that was a component of the aforementioned test set. This was done in order to ensure that the biometric templates would remain secure. This simulates situations that may occur in the real world where it is vital to complete this criterion. It is essential to store or transmit face data in a secure manner, and this allows for the fulfillment of this requirement. The trained model was allowed immediate examination of encrypted feature vectors during the decryption, ensuring data privacy and predictive utility prior to prediction creation.

The system was assessed utilizing standard criteria including accuracy, precision, recall, and F1-score. Furthermore, the False Acceptance Rate (FAR) and False Rejection Rate (FRR) were computed to assess the security performance, particularly for biometrics. Moreover, accuracy and classification metrics were recorded before and after encryption to demonstrate that the encryption process does not reduce system accuracy while ensuring data protection. The overview of the proposed model is detailed in Fig. 1.

3.1 Implementation Detail

A Python-based prototype made to demonstrate how cryptographic techniques might be applied to safeguard biometric identity systems driven by artificial intelligence. This example shows how symmetric encryption can be used with facial recognition systems to keep private biometric data safe while it is being handled and stored.

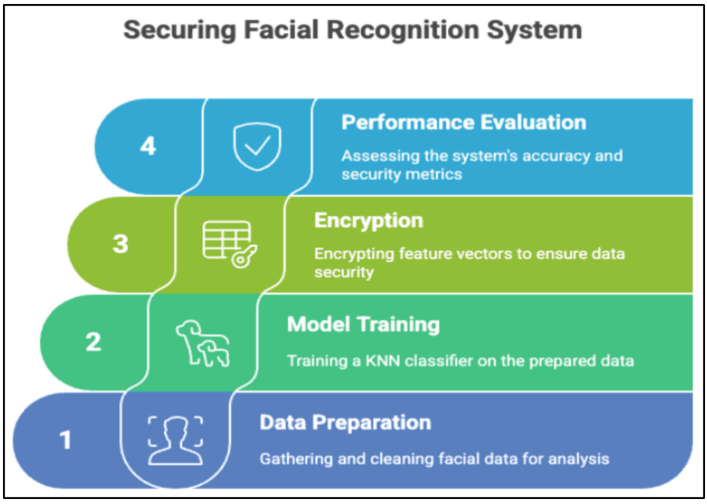


Fig. 1. Workflow diagram of the proposed model.

Our goal is to create a secure facial recognition system that protects biometric data using AES-based symmetric encryption (Fernet) in two stages: first, applying KNN, and then using SVM with PCA for classification to improve our prototype system's accuracy to more than 90%, integrating both cryptographic protection and advanced machine learning, as shown in Algorithm 1.

Algorithm: Secure Facial Recognition with Encrypted Biometric (SVM + PCA)

Input:
Facial Image Dataset D (with labels)
Output:
Classification accuracy before & after encryption, security-preserving recognition pipeline
---
1. Load Dataset D with 5-class facial images
2. Preprocess each image:
a. Resize to uniform size
b. Normalize pixel values
c. Apply Data Augmentation (rotation, flip, zoom)
3. Extract features:
a. Use CNN or HOG to extract feature vectors F
b. Apply PCA to F to get reduced feature set F_pca
4. Split F_pca into training set (X_train, y_train) and test set (X_test, y_test)
5. Train classifier:
a. Initialize SVM with RBF kernel
b. Train on X_train, y_train
6. Secure test set:
a. Encrypt each vector in X_test using Fernet (symmetric AES)
b. Decrypt the encrypted vectors before classification
7. Predict using classifier:
a. Predict y_pred = SVM.predict(X_test_decrypted)
8. Evaluate:
a. Compute accuracy, confusion matrix, FAR, FRR
b. Compare performance before and after encryption
9. Visualize:
a. Display results with pie charts, bar plots, heatmaps
End

Also, Table 1 indicates the utilized tools and libraries for this research.

Table 1. Applied tools and libraries

Tool/Library	Purpose
Python 3.8+	Main programming environment
pandas, NumPy	Data handling and manipulation
scikit-learn	PCA, SVM model, performance evaluation
matplotlib, seaborn	Chart generation and visualization
cryptography (Fernet)	Symmetric encryption of biometric vectors
Jupyter Notebook	Prototyping, visualization, testing
Visual Studio Code	Script organization and modular coding

3.2 Results and Evaluation

To evaluate the effectiveness of facial recognition with encryption, we tested two different models: a basic K-Nearest Neighbor's (KNN) classifier and an improved Support Vector Machine (SVM) model using Principal Component Analysis (PCA) for dimensionality reduction. To judge the system's fingerprint security, the False Acceptance Rate (FAR) and False Rejection Rate (FRR) were applied as follows:

FAR = np.sum((y_test != y_pred) & (y_test == 0)) / np.sum(y_test == 0)
FRR = np.sum((y_test != y_pred) & (y_test == 1)) / np.sum(y_test == 1)

The confusion matrix in Fig. 2 shows how well the classifier did for each class in the face recognition dataset. It shows how many times the model got the predictions right and wrong. In this stage, we compared both implementations in terms of their performance to select the best one for comparing to the literature. Table 2 shows this comparison and indicates that the SVM + PCA model has the highest accuracy of 94.74% with the detection of the most correct samples.

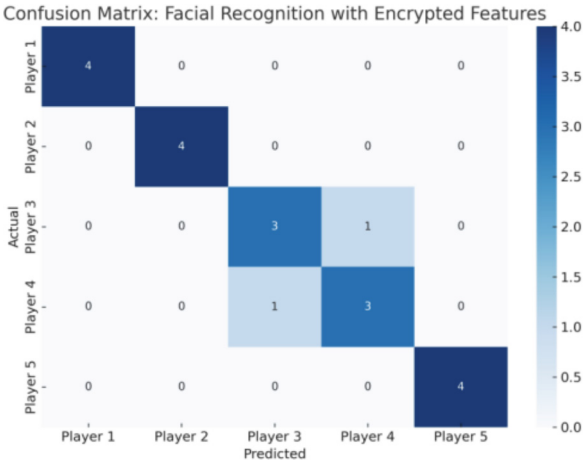


Fig. 2. Confusion matrix of the implemented SVM + PCA

As shown, applying PCA before classification and switching to a more robust classifier significantly improved the system's accuracy, with a reduction in incorrect predictions.

Table 3 shows the class-wise evaluation of our improved proposed model (PCA + SVM) for 5 players in terms of Precision, Recall, and F1-Score.

Table 2. Implemented models' performance comparison in terms of accuracy

Model	Feature Prep	Classifier	Accuracy (%)	Correct	Incorrect
Baseline (Initial)	Raw 4096 features	KNN	60.87%	28	18
Improved (Updated)	PCA (150 components)	SVM (RBF Kernel)	94.74%	54 out of 57 samples	3

Table 3. Class-wise Evaluation (PCA + SVM)

Class	Player Name	Precision	Recall	F1-Score	Support
0	Class 0	1.00	1.00	1.00	12
1	Class 1	0.92	0.92	0.92	13
2	Class 2	0.92	0.92	0.92	12
3	Class 3	0.92	0.92	0.92	10
4	Class 4	1.00	1.00	1.00	10

4 Discussion and Analysis

The reason for our suggested strategy stems from the twin goal of retaining good classification performance while protecting privacy through encryption. Unlike many deep learning algorithms, which demand significant computer resources, we use symmetric encryption (Fernet) in conjunction with PCA-enhanced SVM to provide a lightweight, repeatable, and resource-efficient solution.

The adoption of SVM with RBF kernel was prompted by its demonstrated stability in dealing with non-linear data with high dimensions, making it appropriate for facial recognition applications. By evaluating versus a KNN baseline and gradually improving the model via PCA, we exhibited significant performance enhancements—from 60.87% benchmark accuracy to 94.74%—while retaining encrypted feature vectors.

Thus, the choice of this approach is justified not just by its superior empirical performance but also by its practical application in privacy-sensitive, resource-constrained settings in which high-complexity deep learning models may be impractical.

This work shows that even ensuring biometric data security by encryption, combining appropriate dimensionality reduction with a strong classifier can greatly increase the accuracy of AI-based facial recognition systems. Using raw 4096-pixel data, first our K-Nearest Neighbors (KNN) model obtained an accuracy of 60.87%. Although this offered a minimum standard, it did not live up to actual biometric authentication system performance. Following Principal Component Analysis (PCA) to lower feature dimensionality and training a Support Vector Machine (SVM) with an RBF kernel, our revised model attained a 94.74% accuracy, accurately categorizing 54 out of 57 test samples.

Our revised model, approaches the performance range observed in more advanced systems than that of literature that can be seen in Table 4, Sumalatha et al., (2024) [2],

for example, show 85–90% accuracy utilizing multimodal fusion approaches combining several biometric traits (e.g., face + voice). Combining convolutional neural networks (CNNs) with adaptive cryptography approaches helped score 88–93% [10]. With homomorphic encryption in fingerprint authentication systems, [7] claimed 91.4% accuracy. Although our model is superior to this literature, it offers solid proof that template-level encryption using Fernet may be applied without appreciably compromising performance.

Given that we employed a real-world, unimodal facial dataset with just pixel features and no deep features or multimodal fusion, our results are especially significant. Unlike much research in the literature that depends on pretrained CNNs or sophisticated fusion pipelines, our approach offers a lightweight architecture that can function locally, therefore making it more accessible and efficient for small-scale or privacy-oriented deployments. Using a confusion matrix heatmap, a pie chart, and bar charts, visual analysis confirms even more that PCA + SVM greatly lowers misclassifications as compared to KNN.

In essence, our improved model (SVM + PCA) with RBF kernel indicates that classical machine learning plus encryption allows privacy-preserving facial recognition to be viable. This method might be scaled up for more secure and accurate biometric authentication uses with additional developments such as deep learning-based feature extraction, bigger datasets, or real-time liveness detection.

Table 4. Comparison with literature

Study/Approach	Method	Accuracy (%)
Proposed model (Baseline)	KNN (4096-pixel input)	60.87%
Proposed model (Improved)	**PCA + SVM (RBF)**	**94.74%**
Sumalatha et al., (2024) [2]	Multimodal Fusion	85–90%
Gilbert, C., & Gilbert, M. A (2025) [10]	CNN + Homomorphic Encryptions	88–93%
Chen et al. (2025) [7]	Fingerprint + Homomorphic Enc.	91.4%

5 Conclusion

This work aimed to assess how symmetric encryption might be implemented in AI-based facial recognition systems without sacrificing classification performance. Starting with a K-Nearest Neighbors (KNN) baseline and working towards a PCA-enhanced Support Vector Machine (SVM) with an RBF kernel, we constructed and compared several machine learning models using a real-world dataset of Bayern Munich players. While keeping security via encrypted feature vectors utilizing Fernet, the enhanced model attained an accuracy of 94.74%, clearly improving over the baseline of 60.87%.

Significantly, this work adds useful value to the scholarly and research community since it shows that privacy-aware biometric systems may be produced by combining lightweight encryption techniques with classical machine learning models. Unlike many high-complexity deep learning methods, our system takes advantage of easily replicable, widely available tools and techniques. Researchers investigating the integration

of artificial intelligence and cybersecurity in environments with limited computational resources or where data privacy is a legal or ethical issue will find this very helpful. Deep feature extractors, homomorphic encryption for end-to-end security, and scaling up to increasingly difficult multimodal biometric tasks would help the system to be expanded in the next work. All things considered, this work provides a useful basis for creating responsible, accurate, and safe biometric authentication systems, a direction of ever more relevance in contemporary artificial intelligence research.

References

1. Shethiya, A.S.: AI-enhanced biometric authentication: improving network security with deep learning. Acad. Nexus J. **3**(1) (2024)
2. Sumalatha, U., Prakasha, K.K., Prabhu, S., Nayak, V.C.: A comprehensive review of unimodal and multimodal fingerprint biometric authentication systems: fusion, attacks, and template protection. IEEE Access. **12**, 64300–64334 (2024)
3. Mohammed, A., Salama, A., Shebka, N., Ismail, A.: Enhancing network access control using multi-modal biometric authentication framework. Eng. Technol. Appl. Sci. Res. **15**(1), 20144–20150 (2025)
4. Annapurna, B., Geetha, G., Madhiraju, P., Kalaiselvi, S., Sushith, M., Ramadevi, R., Pandey, P.: Secured and cloud-based electronic health records by homomorphic encryption algorithm. Int. J. Electr. Comput. Eng. (2088-8708). **15**(1), 1152 (2025)
5. Vanin, F.N.D.S., Policarpo, L.M., Righi, R.D.R., Heck, S.M., da Silva, V.F., Goldim, J., da Costa, C.A.: A blockchain-based end-to-end data protection model for personal health records sharing: a fully homomorphic encryption approach. Sensors. **23**(1), 14 (2022)
6. Melzi, P., Rathgeb, C., Tolosana, R., Vera-Rodriguez, R., Busch, C.: An overview of privacy-enhancing technologies in biometric recognition. ACM Comput. Surv. **56**(12), 1–28 (2024)
7. Chen, Y., Duan, F., Zhao, Y., Han, T., Hu, J., Liu, X., et al.: Efficient face information encryption and verification scheme based on full homomorphic encryption. Sci. Rep. **15**(1), 11388 (2025)
8. Harsha, S., Aswini, K., Sahith, G., Abhiram, G.V.N., Kumar, D.B.: Data integrity verification in images through SHA-based security. Int. J. Adv. Comput. Eng. Commun. Technol. **14**(1), 92–102 (2025)
9. Xu, G., Zhang, J., Cliff, U.G.O., Ma, C.: An efficient blockchain-based privacy-preserving scheme with attribute and homomorphic encryption. Int. J. Intell. Syst. **37**(12), 10715–10750 (2022)
10. Gilbert, C., Gilbert, M.A.: Homomorphic encryption algorithms for secure data computation. Int. Res. J. Adv. Eng. Sci. **10**(2), 148–162 (2025)
11. Zhu, X., Cao, C.: Secure online examination with biometric authentication and blockchain-based framework. Math. Probl. Eng. **2021**(1), 5058780 (2021)
12. https://www.kaggle.com/datasets/jessicali9530/lfw-dataset

Who's Responsible for Online Privacy: Investigating User Behaviour Through Social Learning and Protection Motivation Theory

Saide Saide[1,2(✉)], Nurwahyu Alamsyah[1], M. Ikhsan Ramadhan[3], and Nurfitria Ningsi[4]

[1] Department of Information Technology, Faculty of Engineering, Universitas Muhammadiyah Yogyakarta, Kasihan, Indonesia
saide@umy.ac.id

[2] Department of Information Systems (PRO-Knowledge Indonesia), Faculty of Science and Technology, State Islamic University of Sultan Syarif Kasim Riau (UIN SUSKA Riau), Pekanbaru, Indonesia

[3] PRO Knowledge Indonesia, Department Information Systems, Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, Indonesia

[4] Department of Information Systems, Faculty of Information Technology, Universitas Sembilanbelas November Kolaka, Kolaka, Sulawesi Tenggara, Indonesia

Abstract. The widespread use of social media raises concerns about user privacy, yet many individuals still exhibit weak protection behaviors. This study integrates Protection Motivation Theory (PMT) and Social Learning Theory (SLT) to examine psychological and social factors influencing online privacy behavior. Survey data from 65 Indonesian users were analyzed with PLS-SEM. Findings show that privacy self-efficacy and perceived threats strongly predict protective actions, while social learning reinforces intentions. The study contributes by linking PMT and SLT in privacy research and provides practical insights for platform designers, policymakers, and educators to promote stronger privacy practices.

Keywords: Privacy Protection Behavior · Protection Motivation Theory · Social Learning Theory · Online Privacy · Self-Efficacy · Social Media Users · Information Security

1 Introduction

The rapid growth of social media has transformed communication and information sharing but also intensified concerns about personal data privacy. Platforms such as Twitter, Facebook, and Instagram collect vast amounts of user information, often without explicit consent, undermining individual control over what is disclosed [1–3]. Although awareness of privacy risks is increasing, many users lack the behavioral mechanisms to adequately protect their data. Research shows that self-awareness strongly influences the ability to safeguard personal information through authentication methods such as strong passwords, biometrics, and multi-factor identification [4–6].

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026
S. Lanka et al. (Eds.): ICTSM 2025, LNNS 1755, pp. 380–390, 2026.
https://doi.org/10.1007/978-3-032-13177-5_30

The widespread use of weak passwords persists, largely because complex combinations are difficult to remember [7]. Self-efficacy has become a critical factor in managing digital privacy, as individuals with higher confidence are more proactive in applying security measures [8, 9] Protection motivation is further influenced by perceived threats, where awareness of breach severity and likelihood drives protective actions [1, 10, 11]. Perceived vulnerability is equally important, as users who recognize higher risks are more inclined to update passwords or back up data [12–14]. Ultimately, privacy risks are not only technical but also behavioral, shaped by user confidence, threat perception, and responsibility.

Privacy concerns now focus on regulating and safeguarding sensitive data—names, addresses, and financial records—that shape an individual's digital footprint [15, 16]. Protecting such data is not only a privacy right but also a behavioral responsibility [17, 18]. However, most studies treat privacy either as an individual cognitive issue or as a product of environmental influence. This fragmented view overlooks how internal motivation and social learning interact, leaving a critical gap in explaining user privacy behavior.

This study addresses the gap in privacy research by integrating Protection Motivation Theory (PMT) and Social Learning Theory (SLT) to explain online privacy protection behavior [19, 20]. It explores how internal factors—such as privacy concerns, perceived vulnerability, self-efficacy, and response efficacy—interact with social learning to influence user practices. The model includes key constructs such as privacy protection behavior, information privacy concern, privacy self-efficacy, response efficacy, and website privacy [21, 22], providing a more comprehensive view of how users manage personal data in digital environments.

Theoretically, this study contributes by combining PMT and SLT into a unified framework. Unlike prior work that applied these theories separately, it demonstrates how cognitive appraisals (e.g., self-efficacy, response efficacy) interact with socially learned behaviors and environmental cues, providing a more holistic explanation of user privacy management [13]. Practically, the findings inform platform developers, policymakers, and educators [23]. Stronger regulations and standards can enhance accountability, while digital literacy programs can empower users to detect risks, configure privacy settings, and navigate complex data ecosystems more effectively.

2 Related and Proposed Work

This study is underpinned by two primary theoretical lenses: PMT and SLT, both of which provide insight into the mechanisms behind users' privacy protection behavior in digital environments. PMT posits that individuals are motivated to protect themselves when they perceive a significant threat to their personal well-being. This motivation is influenced by two main cognitive assessments: threat appraisal and coping appraisal. Threat appraisal evaluates the severity and vulnerability of the threat, while coping appraisal involves self-efficacy (belief in one's ability to perform protective actions) and response efficacy (belief that those actions will be effective). In the context of online privacy, PMT explains that users' concern over potential data breaches can motivate protective behaviors, such as creating strong passwords, limiting shared data, or adjusting privacy settings.

Complementing PMT, SLT underscores the role of external influences—such as observation and imitation—in shaping individual behaviors [20, 24]. When users observe others experiencing privacy violations or successfully protecting their data, they internalize and adopt similar protective behaviors. This highlights how environmental cues and modeled actions within a digital ecosystem influence user conduct. Central to this study is privacy protection behavior, which involves the actions users take to secure personal information, including controlling data dissemination, configuring privacy settings, and avoiding risky online activities. These behaviors are often driven by users' concerns over how their information is collected, stored, and used—referred to as information privacy concerns. As users increasingly engage with platforms that harvest data, these concerns affect decisions regarding service usage and privacy settings [2, 20].

Drawing on this theoretical foundation, we propose several hypotheses. First, users who experience a heightened concern for their personal data—referred to as information privacy concern—are more likely to engage in self-protective behaviors. These individuals are typically more sensitive to the misuse or unauthorized sharing of their data, leading them to proactively limit the information they disclose or take additional steps to safeguard their online identity. Prior literature suggests that this concern directly influences behavioral intentions such as changing privacy settings or refusing to share certain information [1]. Therefore: *H1: User information privacy concerns positively influence privacy protection behavior on social media and the internet.*

Next, response efficacy, which refers to the belief that a particular protective behavior will be effective in mitigating risks, is central to the motivation process described in PMT. Users who believe that their actions—such as enabling two-factor authentication or rejecting cookies—can successfully reduce vulnerability are more likely to develop strong privacy concerns and act accordingly (Moser & Deichmann, 2021). High response efficacy reinforces perceived control and thus encourages greater vigilance. Accordingly: *H2a: Response efficacy positively affects users' information privacy concerns. H2b: Response efficacy positively affects privacy protection behavior.*

Another important factor is privacy self-efficacy, or the user's confidence in their ability to manage digital privacy effectively. Users with high privacy self-efficacy believe they possess the knowledge and skills to navigate privacy settings, apply encryption tools, and evaluate privacy policies [25]. This confidence can lead to increased concern about how personal data is handled and motivate more protective behavior [26]. Hence: *H3a Privacy self-efficacy positively influences users' information privacy concerns. H3b: Privacy self-efficacy positively affects privacy protection behavior.*

Finally, concerns about website privacy—including the collection and use of personal data by third-party cookies or trackers—can significantly impact user perceptions and behaviors. Users who distrust a website's data handling practices are more likely to become privacy-conscious and limit their digital footprints, such as by rejecting cookies or avoiding data input [12]. Therefore: *H4a: Website privacy concerns are positively related to user information privacy concerns. H4b: Website privacy concerns positively influence privacy protection behavior.* By integrating PMT and SLT, this study presents a comprehensive framework for understanding how users' internal beliefs and external cues shape their privacy-related decisions. These hypotheses will be tested empirically to explore the relationships among psychological constructs and actual online behavior.

3 Methods

This study adopted a quantitative design using a structured, self-administered questionnaire to investigate privacy concerns and protection behaviors among undergraduate students (aged 19–25) from two major universities in Riau Province, Indonesia. Simple random sampling ensured equal selection opportunity, with data collected both online and offline over three weeks in February 2023. Of 409 responses, 178 valid cases were retained after screening and analyzed with SmartPLS [27]. To minimize non-response bias, participants received rewards, while ethical standards—consent, anonymity, and confidentiality—were strictly upheld. The instrument was developed through collaborative discussions, a literature review of Q1–Q2 journals, and validated constructs from PMT and SLT [19, 20], focusing on student knowledge, concerns, and behaviors to yield context-specific yet methodologically rigorous insights into digital privacy [28] (Table 1).

Table 1. Questions of variable

Variable	Statements
Response Efficacy	• RE1: Enabling privacy protection features on social networks and the internet can protect against threats. • RE2: If I use privacy protection features on social networks and the internet, I am not vulnerable to losing the privacy of my information. • RE3: I can effectively control the privacy of my information using the privacy protection features of social networks and the Internet.
Privacy Self-Efficacy	• PSE1: I believe I have the ability to protect my personal information on social networks and the Internet. • PSE2: I'm sure I can enable privacy protection features on social networks and the internet without any help. • PSE3: I'm confident when using privacy protection features on social networks and the internet.
Website Privacy	• WP1: I'm worried that this website is collecting too much information about me • WP2: I don't mind if this website asks for my personal information • WP3: I am concerned about my privacy when browsing this website • WP4: I have no doubts about how well my privacy is protected on this website

(*continued*)

Table 1. (*continued*)

Variable	Statements
User's Information Privacy Concern	• UIPC1: I'm concerned that my personal information on social networks may be used for the wrong purposes. • UIPC2: I'm worried that my personal information on social networks can be accessed by unknown parties. • UIPC3: I usually think twice before giving out my personal information on social networks. • UIPC4: I feel that this social network collects excessive personal information. • UIPC5: I'm concerned that my personal information on social networks may be used in ways that I am not aware of.
Privacy Protection Behaviour	• PPB1: I consciously misrepresented certain personal information on social networks. • PPB2: I intentionally provided incomplete personal information on social networks. • PPB3: I intentionally provided fake personal information on social networks. • PPB4: I deliberately refrain from giving out certain personal information on social networks. • PPB5: I consciously avoid providing certain personal information on social networks. • PPB6: I am willing to give certain personal information on social networks

4 Results Analysis and Discussion

The assessment of construct validity and discriminant validity within the structural model follows established statistical thresholds to ensure measurement quality. Specifically, the minimum acceptable values include factor loadings and Cronbach's Alpha (CA) above 0.70, Composite Reliability (CR) exceeding 0.80, and Average Variance Extracted (AVE) values greater than 0.50. To assess multicollinearity, the Variance Inflation Factor (VIF) for all constructs must remain below the threshold of 5. As shown in Table 2, the factor loadings for all indicators exceeded 0.70, and VIF values for all constructs were below 5, indicating no significant multicollinearity concerns. Additionally, all AVE values were above 0.50, and CR values surpassed 0.80, confirming the internal consistency and convergent validity of the constructs.

Table 2. Construct validity and discriminant

Variable	Indicators	Loadings	VIF	p-value	CA	CR	AVE
Privacy Protection Behaviour (PPB)	PPB1	0.588	1.283	0.000	0.799	0.856	0.501
	PPB2	0.658	1.541	0.000			
	PPB3	0.655	1.597	0.000			
	PPB4	0.797	2.126	0.000			
	PPB5	0.775	2.020	0.000			
	PPB6	0.750	1.626	0.000			
Response Efficacy (RE)	RE1	0.836	1.571	0.000	0.804	0884	0.718
	RE2	0.843	1.831	0.000			
	RE3	0.863	1.913	0.000			
Privacy Self-Efficacy (PSE)	PSE1	0.825	1.501	0.000	0.747	0.855	0.663
	PSE2	0.791	1.497	0.000			
	PSE3	0.826	1.474	0.000			
User's Info. Privacy Concern (UIPC)	UIPC1	0.683	1.400	0.000	0.752	0.836	0.507
	UIPC2	0.809	1.906	0.000			
	UIPC3	0.722	1.543	0.000			
	UIPC4	0.595	1.210	0.000			
	UIPC5	0.732	1.430	0.000			
Website Privacy (WP)	WP1	0.719	1.382	0.000	0.733	0831	0.552
	WP2	0.751	1.706	0.000			
	WP3	0.779	1.387	0.000			
	WP4	0.721	1.613	0.000			

Regarding discriminant validity, Table 3 presents the results using the Heterotrait-Monotrait Ratio of Correlations (HTMT), where all values remained below the recommended threshold of 0.90. These results confirm that each construct is empirically distinct, thereby satisfying the requirement for discriminant validity. The HTMT approach is widely regarded as a robust and preferred method for assessing discriminant validity over the traditional Fornell-Larcker criterion [29].

Table 3. Heterotrait-Monotrait Ratio of Correlations

Variable	PPB	RE	PSE	UIPC	WP
PPB					
RE	0.494				
PSE	0.674	0.779			
UIPC	0.770	0.718	0.633		
WP	0.812	0.541	0.865	0.701	

The results of the structural model are presented in Fig. 1, illustrating the direct effects of the study variables on privacy protection behavior and user information privacy concern. As shown in Fig. 1, both response efficacy ($\beta = 0.399$, $p < .001$) and

website privacy ($\beta = 0.378$, $p < .001$) exert significant positive effects on users' information privacy concern. These findings support Hypotheses H2a and H4a, respectively. However, Hypothesis H3a, which proposed a significant relationship between privacy self-efficacy and user information privacy concern, is not supported, as the coefficient was insignificant ($\beta = 0.009$, $p > .01$). Among the predictors, response efficacy emerged as the strongest predictor of user information privacy concern in Model 1.

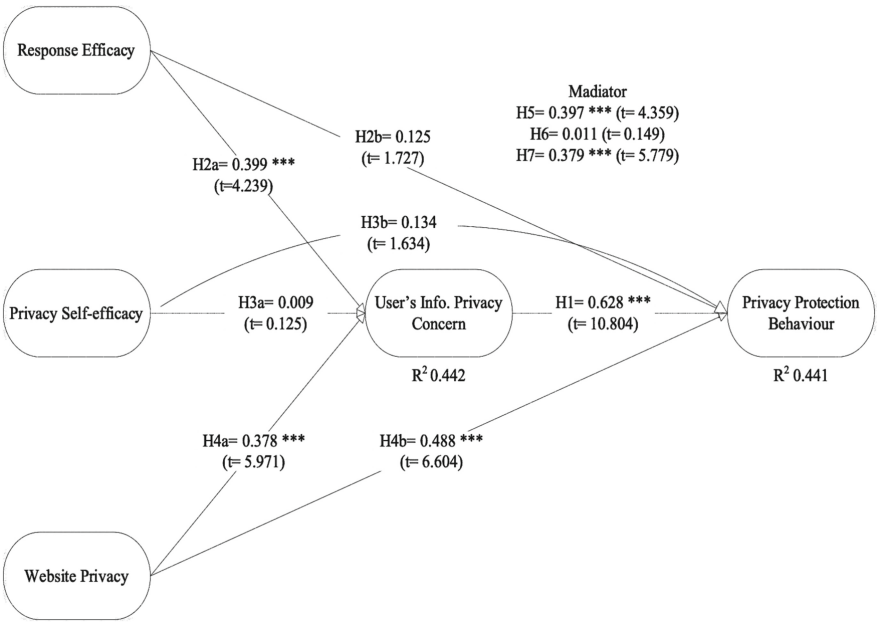


Fig. 1. Model results

In Model 2, the effects of response efficacy ($\beta = 0.125$, $p > .01$) and privacy self-efficacy ($\beta = 0.134$, $p > .10$) on privacy protection behavior were both statistically insignificant, leading to the rejection of Hypotheses H2b and H3b. However, website privacy demonstrated a strong and significant influence on privacy protection behavior ($\beta = 0.488$, $p < .001$), thus supporting Hypothesis H4b. Furthermore, Fig. 1 reveals the mediating role of user information privacy concern between the independent variables and privacy protection behavior. Using the three-step procedure outlined by Baron and Kenny and evaluated through SmartPLS, the results from Model 3 confirm that user information privacy concern significantly influences privacy protection behavior ($\beta = 0.628$, $p < .001$), thereby supporting Hypothesis H1. All models were tested using bootstrapping with 178 resamples to ensure the robustness of the findings.

We employed the PLS bootstrapping resampling procedure with 178 sub-samples to assess the statistical significance of the standardized path coefficients within our structural model. The results indicate that privacy protection features available on internet and social media platforms play a critical role in helping users safeguard their personal data [30]. Users who activate these features tend to exhibit stronger control over

the information they choose to share, resulting in a reduced risk of privacy breaches. Furthermore, response efficacy demonstrated a significant positive influence on users' privacy concerns, reinforcing its role in motivating protective behavior. However, when examined independently, response efficacy did not significantly predict changes in privacy protection behavior. Nevertheless, when mediated by privacy concerns, the effect of response efficacy on user behavior was strengthened, indicating that users who perceive privacy threats and believe in the effectiveness of protective actions are more likely to adopt secure behavior patterns. This is consistent with observed user responses when experiencing privacy breaches—individuals tend to revise and enhance their behavior to better protect their data. To conclude, H1 was strongly supported, showing that higher information privacy concern drives protective actions such as adjusting privacy settings or limiting data sharing, consistent with prior evidence [1, 2, 31]. H2a and H4a were also supported, indicating that response efficacy and website privacy significantly influence privacy concern. Users who believe protective actions are effective perceive privacy as more important, aligning with PMT's core assumption that protection motivation rises when responses are seen as effective [19]. Likewise, concerns over cookies, tracking, and third-party access heighten sensitivity to privacy risks [12].

By contrast, H3a was not supported, suggesting that confidence in managing privacy features does not necessarily translate into greater concern for personal data [16]. Similarly, H2b and H3b were not supported, showing that efficacy beliefs alone do not directly predict protective behavior unless mediated by privacy concern. This highlights the pivotal role of privacy concern as a motivational bridge between intention and action. Notably, H4b was strongly supported, confirming that website privacy concerns directly foster protective strategies such as ad-blocking, rejecting cookies, or avoiding platforms, consistent with SLT's emphasis on learned caution in risky environments [20]. Overall, these findings confirm the importance of privacy features and psychological motivators in shaping responsible data protection practices among internet users.

5 Conclusion & Future Scope

This study examined how privacy concerns shape user behaviors in securing personal data on the internet and social media, drawing on PMT and SLT. The findings yield both theoretical and practical contributions. Mediation analysis further validated that privacy concern mediates the effects of response efficacy and website privacy on behavior, underscoring the dual-process logic of PMT and SLT. Theoretically, this study enriches IS security research by integrating PMT and SLT into a unified model, showing that privacy concern acts as a pivotal mediator between motivation and behavior. Importantly, it highlights boundary conditions for instance, users with high technical confidence (self-efficacy) may still neglect safeguards if their perceived risk is low, reflecting a critical gap in translating ability into action. Practically, this means that offering privacy tools alone (e.g., end-to-end encryption or customizable privacy settings on social media) is insufficient. Unless users are made aware of risks such as the misuse of personal data in scandals like Cambridge Analytica, they may not engage in protective behaviors despite available mechanisms. Despite its contributions, this study has several limitations that open avenues for future research. First, the reliance on self-reported survey

data may introduce common method bias or social desirability effects, where users overstate their protective behaviors. Future studies could employ experimental or behavioral tracking methods (e.g., monitoring actual privacy settings changes or cookie rejections) to capture more objective evidence. Second, the study focused on general internet and social media use, which may overlook context-specific dynamics. For example, privacy behaviors on professional platforms like LinkedIn may differ substantially from those on entertainment-focused platforms like TikTok. Comparative studies across platforms or cultural settings could enrich our understanding of how social norms and regulatory environments shape privacy decisions. Third, the model emphasizes cognitive and social learning perspectives, but it does not incorporate emotional or affective factors, such as fear, anxiety, or trust, which are known to influence decision-making in online environments. Future research could integrate affective dimensions or examine longitudinal patterns to see how privacy concerns evolve after high-profile data breaches or the introduction of new regulations.

Acknowledgement. This research was supported by the Universitas Muhammadiyah Yogyakarta (UMY) Research Fellowship Program. We also would like to thank the Committees, Reviewers, and PRO-Knowledge Indonesia for their critiques and revision recommendations. Thanks to the Faculty of Science and Technology, UIN Suska Riau, for their permission during the Research Fellowship Program.

References

1. Adhikari, K., Panda, R.K.: Users' information privacy concerns and privacy protection behaviors in social networks. J. Glob. Mark. **31**, 96–110 (2018). https://doi.org/10.1080/08911762.2017.1412552
2. Kumar, S., Kumar, P., Bhasker, B.: Interplay between trust, information privacy concerns and behavioural intention of users on online social networks. Behav. Inf. Technol. **37**, 622–633 (2018)
3. Sofyani, H., Darma, E.S.: Effect of architecture and efficiency of mobile banking application on the intention to continue using Islamic bank: does data security matter? J. Islam. Mark. **15**, 1479–1497 (2024). https://doi.org/10.1108/JIMA-07-2023-0220
4. Zhang, L., McDowell, W.C.: Am I really at risk? Determinants of online users' intentions to use strong passwords. J. Internet Commer. **8**, 180–197 (2009)
5. Mamonov, S., Benbunan-Fich, R.: The impact of information security threat awareness on privacy-protective behaviors. Comput. Human Behav. **83**, 32–44 (2018). https://doi.org/10.1016/j.chb.2018.01.028
6. Kotsias, J., Ahmad, A., Scheepers, R.: Adopting and integrating cyber-threat intelligence in a commercial organisation. Eur. J. Inf. Syst. **32**, 1–17 (2022)
7. Warkentin, M., Willison, R., Warkentin, M.: Behavioral and policy issues in information systems security: the insider threat (2017)
8. Indrasari, A., Nadjmie, N., Endri, E.: Determinants of satisfaction and loyalty of e-banking users during the COVID-19 pandemic. Int. J. Data Netw. Sci. **6**, 497–508 (2022). https://doi.org/10.5267/j.ijdns.2021.12.004
9. Li, Y.: A multi-level model of individual information privacy beliefs. Electron. Commer. Res. Appl. **13**, 32–44 (2014)

10. Ha, Q.A., Chen, J.V., Uy, H.U., Capistrano, E.P.: Exploring the privacy concerns in using intelligent virtual assistants under perspectives of information sensitivity and anthropomorphism. Int. J. Hum. Comput. Interact. **37**, 512–527 (2021)
11. Masi, D., Kumar, V., Garza-Reyes, J.A., Godsell, J.: Towards a more circular economy: exploring the awareness, practices, and barriers from a focal firm perspective. Prod. Plan. Control. **29**, 539–550 (2018). https://doi.org/10.1080/09537287.2018.1449246
12. Basak, E., Calisir, F.: An empirical study on factors affecting continuance intention of using Facebook. Comput. Human Behav. **48**, 181–189 (2015)
13. Liao, S.-H., Widowati, R., Cheng, C.-J.: Investigating Taiwan Instagram users' behaviors for social media and social commerce development. Entertain. Comput. **40** (2022). https://doi.org/10.1016/j.entcom.2021.100461
14. Budi, A.W.S., Biantara, I., Afik, A.: Community self-protection on pandemic disaster preparedness corona virus disease (COVID-19). Bali Med. J. **10**, 1199–1205 (2021). https://doi.org/10.15562/bmj.v10i3.2882
15. Dinev, T., Hart, P.: Internet privacy concerns and their antecedents -measurement validity and a regression model. Behav. Inf. Technol. **23**, 413–422 (2004)
16. Nurzihad, M.I., Ichsan, M., Fitriyanti, F.: Personal Data Protection in Indonesian E-commerce Platforms: The Maqasid Sharia Perspective. In: Proceedings of Eighth International Congress on Information and Communication Technology. ICICT 2023 Lecture Notes in Networks and Systems (LNNS), vol. 693, pp. 1077–1086. Springer, Singapore (2023). https://doi.org/10.1007/978-981-99-3243-6_88
17. Gunawan, Y., Irrynta, D.: International human rights protection: the impact of COVID-19 pandemic on domestic violence in Asia. Yustisia. **11**, 16–28 (2022). https://doi.org/10.20961/yustisia.v11i1.58872
18. Cahyani, A., Fitriyanti, F., Ahmad, J., Ramlan, P.: Consumer legal protection against decoy effects through digital literacy. Subst. Justice Int. J. Law. **5**, 193–208 (2022). https://doi.org/10.56087/substantivejustice.v5i2.196
19. Rogers, R.W.: A protection motivation theory of fear appeals and attitude change. J. Psychol. **91**, 93–114 (1975)
20. Jensen, G.: Social learning theory. In: The Routledge Companion to Criminological Theory and Concepts, pp. 115–119. Routledge/Taylor & Francis Group, London (2017)
21. Prasetyoningsih, N., Rahma, M.A.M.: Analysis of infotainment programs in digital media: legal protection for Indonesian children perspective. In: Proceedings of Eighth International Congress on Information and Communication Technology. ICICT 2023 Lecture Notes in Networks and Systems (LNNS), vol. 693, pp. 1067–1076. Springer, Singapore (2023). https://doi.org/10.1007/978-981-99-3243-6_87
22. Maksum, A.: Indonesian post-migrant workers: a challenging problem for human security. Soc. Sci. Human. Open. **4** (2021). https://doi.org/10.1016/j.ssaho.2021.100223
23. Lawelai, H., Purnomo, E.P., Nurmandi, A., et al.: Cybersecurity policy on smart city infrastructure: a mapping of new threats and protections. J. Sci. Technol. Policy Manag. (2025). https://doi.org/10.1108/JSTPM-09-2024-0359
24. Brauer, J.R., Tittle, C.R.: Social learning theory and human reinforcement. Sociol. Spectr. **32**(2), 157–177 (2012)
25. Ratna, S., Saide, S., Herzavina, H.: A preliminary model analysis of knowledge management design: spiritual leadership on knowledge worker productivity. Technol. Anal. Strateg. Manag. **35**, 270–285 (2023). https://doi.org/10.1080/09537325.2021.1973665
26. Warkentin, M., Johnston, A.C., Shropshire, J.: The influence of the informal social learning environment on information privacy policy compliance efficacy and intention. Eur. J. Inf. Syst. **20**, 267–284 (2011)
27. Ringle, C.M., Wende, S., Becker, J.-M.: SmartPLS 3. Boenningstedt: SmartPLS (2015). https://www.smartpls.com

28. Saide, S., Putri, A.M., Muwardi, D.: ProsCons of artificial intelligence-ChatGPT adoption in education settings: a literature review and future research agendas. IEEE Eng. Manag. Rev. **52**, 27–42 (2024). https://doi.org/10.1109/EMR.2024.3394540
29. Saide, S.: Toward business process innovation in the big data era: a mediating role of big data knowledge management. Big Data. **8**, 464–477 (2020). https://doi.org/10.1089/big.2020.0140
30. Liao, S.-H., Widowati, R., Hsieh, Y.-C.: Investigating online social media users' behaviors for social commerce recommendations. Technol. Soc. **66** (2021). https://doi.org/10.1016/j.techsoc.2021.101655
31. Saide, S.: ICT team dual-innovations in the microlevel of circular supply chain management: explicit-tacit knowledge, exchange ideology, and leadership support. IEEE Trans. Eng. Manag. **71**, 12124–12137 (2023). https://doi.org/10.1109/TEM.2022.3166763

Machine Learning Framework for Sustainable UAV Spraying: Life Cycle and Computational Intelligence-Based Assessment

Shefali Vinod Ramteke[1(✉)], Pritish Kumar Varadwaj[1], and Vineet Tiwari[2]

[1] Department of Applied Sciences, Indian Institute of Information Technology Allahabad, Prayagraj 211015, India
{rss2019003,pritish}@iiita.ac.in
[2] Department of Management Studies, Indian Institute of Information Technology Allahabad, Prayagraj 211015, India
vineet.tiwari@iiita.ac.in

Abstract. The increasing integration of machine learning into sustainable computing offers new opportunities to evaluate and optimize agricultural practices. Among these, Unmanned Aerial Vehicles (UAVs) have emerged as promising platforms for pesticide application, particularly in low- and ultra-low-volume spraying regimes that can reduce resource use compared to conventional methods. However, systematic benchmarking of UAV spraying against traditional spraying approaches remains limited, especially when considering both environmental and computational dimensions. This study presents a computational intelligence framework that combines life cycle assessment (LCA) with machine learning models to analyze UAV and conventional spraying in mustard production systems. The framework integrates environmental footprints with predictive modeling and statistical validation to capture multiple sustainability indicators, including energy use, water scarcity, and emissions. Using Random Forest and Gradient Boosting regressors, the framework demonstrates how machine learning can reliably approximate life cycle outcomes, while statistical testing confirms the robustness and significance of UAV advantages across locations. The results highlight the potential of machine learning–driven sustainability analysis not only to quantify resource efficiencies but also to support intelligent decision-making in precision agriculture. By situating agricultural spraying within the broader context of computational sustainability, this work contributes a reproducible approach that aligns with energy-aware machine learning and sustainable computing paradigms.

Keywords: UAV spraying · Machine Learning · Sustainable Computing · Life Cycle Assessment · Agriculture 4.0 · Computational Intelligence

1 Introduction

The integration of machine learning (ML) and computational intelligence into sustainable computing has transformed the way complex systems are modeled, optimized, and managed [1, 2]. Sustainable computing emphasizes the efficient use of computational

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026
S. Lanka et al. (Eds.): ICTSM 2025, LNNS 1755, pp. 391–407, 2026.
https://doi.org/10.1007/978-3-032-13177-5_31

and physical resources, enabling the development of intelligent frameworks that reduce energy, costs, and environmental footprints [3]. Within this paradigm, agriculture represents a critical sector where digitalization and ML-driven decision systems can address pressing challenges of resource scarcity, climate change, and food security [4, 5].

Unmanned Aerial Vehicles (UAVs) are increasingly recognized as transformative tools in precision agriculture, particularly for spraying operations [6]. By operating in low-volume (LV) or ultra-low-volume (ULV) spraying regimes, UAVs can reduce carrier water requirements and enable targeted application of agrochemicals [7, 8]. Compared to conventional ground spraying methods, UAVs offer advantages such as reduced drift, lower water consumption, and improved accessibility in fragmented fields [9]. However, despite rapid deployment in countries such as India and China, systematic evaluation of their sustainability impacts using advanced computational methods remains limited [10].

Existing studies on UAV spraying have primarily focused on engineering performance, such as droplet deposition, canopy penetration, or application efficiency [11, 12]. While these contributions are valuable, they often lack integration with life cycle assessment (LCA), sustainability benchmarking, and computational intelligence models. Moreover, there is a gap in comparative analyses that integrate UAV spraying with conventional high-volume spraying systems under real-world crop conditions [13]. Bridging this gap requires multi-dimensional approaches that combine environmental assessment with data-driven modeling, ensuring both scientific rigor and decision-making relevance.

In this study, we propose a machine learning–based LCA framework to benchmark UAV spraying against conventional spraying in mustard production systems across multiple locations. The framework integrates life cycle inventory data with ML models (Random Forest, Gradient Boosting) to predict sustainability outcomes, while statistical analyses (t-test, ANOVA, effect sizes) validate the robustness of differences. This dual focus on environmental sustainability and computational intelligence addresses the need for reproducible, data-driven tools in sustainable agriculture. The novelty of this work lies in its integrated methodology. Unlike prior UAV spraying studies that have emphasized droplet mechanics [12], system design [13], or descriptive environmental reporting [32–34], this study combines life cycle assessment (LCA), machine learning (ML), and statistical validation into a unified framework. This enables predictive benchmarking of UAV and conventional spraying under multi-year, multi-location mustard trials, moving beyond descriptive outcomes to a reproducible, computational intelligence pipeline. To our knowledge, this is the first study to apply such a combined approach in the context of UAV spraying sustainability assessment in India.

The key contributions of this work are threefold:

1. **Integration of Life Cycle Assessment and Machine Learning:** We propose a novel computational intelligence framework that embeds life cycle sustainability indicators (CO_2 footprint, water scarcity index) into supervised ML pipelines for predictive benchmarking of UAV spraying.
2. **Comparative Benchmarking of UAV and Conventional Spraying:** We evaluate UAV spraying against conventional high-volume spraying across multi-year, multi-location mustard field trials in India, thereby offering a robust empirical basis for sustainability assessment.

3. **Incorporation of Statistical Rigor:** Beyond predictive modeling, the framework integrates statistical validation (t-tests, ANOVA, effect sizes) to confirm the robustness and generalizability of UAV sustainability advantages.
4. **Demonstration of Machine Learning Potential for Decision Support:** The study highlights how ensemble models (Random Forest, Gradient Boosting) can reliably approximate sustainability outcomes, thereby contributing to intelligent decision support systems in precision agriculture.

By situating UAV spraying within the broader landscape of machine learning and sustainable computing, this study contributes a methodological advancement that can be extended to other agricultural systems and decision-support contexts.

2 Methods

The proposed framework is novel in its multi-stage integration of sustainability assessment and computational intelligence. Unlike existing UAV spraying research, which has addressed engineering performance [12, 13] or environmental reporting in isolation [22], our framework combines: (i) LCA-based footprint modeling, (ii) predictive intelligence using Random Forest and Gradient Boosting, including a multi-output regression attempt, and (iii) statistical robustness through t-tests, ANOVA, and Cohen's d effect size analysis. This methodological combination ensures that UAV sustainability benefits are not only modeled but also validated with statistical confidence, positioning the study as a first-of-its-kind pipeline for intelligent agricultural sustainability benchmarking.

2.1 Study Area and Dataset Design

The field trials were conducted across three representative mustard-growing regions in northern India: Saharanpur (Uttar Pradesh), Bharatpur (Rajasthan), and Alwar (Rajasthan). These sites represent semi-arid to sub-humid agro-climatic zones where mustard (*Brassica juncea*) is extensively cultivated and where groundwater scarcity and high irrigation demand are critical sustainability concerns [14, 15]. Conducting experiments in geographically distinct regions allowed for the assessment of UAV spraying performance under varied environmental and soil-water conditions, while also capturing the influence of regional water scarcity indices.

A multi-year experimental design was adopted, with trials repeated over three consecutive cropping seasons to account for seasonal variability in climatic factors, crop phenology, and resource use. Within each site, the experiments were organized as randomized replicated plots across two crop stages: (i) vegetative stage, characterized by dense canopy establishment, and (ii) flowering stage, where pest and disease incidence typically peaks [16]. Replications ensured statistical reliability and reduced field-level variability across UAV and conventional treatments.

Two spraying methods were evaluated: Unmanned Aerial Vehicle (UAV) spraying and conventional ground-based spraying. The UAV treatments utilized commercially available Indian drone platforms equipped with LV/ULV nozzles, operating at

15–20 L/ha spray volume. In contrast, the conventional treatments employed tractor-mounted boom sprayers or knapsack sprayers with spray volumes in the range of 200–300 L/ha. Both methods applied comparable pesticide doses, consistent with regional agronomic recommendations for mustard cultivation [17], although carrier volume and application uniformity differed substantially.

For each trial, a detailed dataset was generated, capturing both input variables and output indicators. The input variables included UAV flight parameters (altitude, forward speed, nozzle flow rate, missions per hectare), energy consumption (battery discharge for UAV, diesel-equivalent for conventional), pesticide dose, and water applied per hectare. Output indicators included spray coverage, uniformity, disease suppression, yield response, and derived sustainability metrics such as carbon footprint (kg CO_2-eq/ha) and water scarcity index (m^3·scarcity factor/ha).

The multi-location, multi-stage, and multi-year dataset provided a robust empirical basis for developing and validating the computational intelligence framework. By systematically capturing environmental, operational, and sustainability-related variables, the dataset enabled integration of life cycle assessment with machine learning models, ensuring reproducibility and generalizability of results across different mustard-growing regions of India.

The trials were conducted for three consecutive mustard seasons (2021–2023) across Saharanpur (Uttar Pradesh), Bharatpur (Rajasthan), and Alwar (Rajasthan), representing sub-humid to semi-arid agro-climatic zones. At each site, UAV spraying (15–20 L/ha) was benchmarked against conventional ground-based spraying (200–300 L/ha) using knapsack or tractor-mounted boom sprayers. Both vegetative and flowering stages were evaluated in replicated plots (three per stage per year), ensuring reliability across locations and crop phenology.

2.2 Life Cycle Assessment (LCA) Modeling

Life Cycle Assessment (LCA) was employed to evaluate the environmental implications of UAV and conventional spraying, following ISO 14040/44 standards [18]. The analysis considered both direct field operations and indirect upstream processes, with the functional unit defined as 1 ha of mustard spraying. The system boundary was delineated as "cradle-to-field", encompassing pesticide formulation, energy inputs, water pumping, and UAV/conventional spraying operations (Fig. 1).

The primary goal was to compare UAV spraying against conventional spraying in terms of carbon footprint and water scarcity impact. While UAV spraying operates under low-volume (LV/ULV) conditions, conventional ground spraying typically requires high-volume carriers. Evaluating both systems within the same functional boundary enabled a fair comparison of sustainability trade-offs [19].

Life Cycle Inventory (LCI) included:

- Energy inputs: UAV battery charging (electricity, Wh converted to kWh), conventional sprayer fuel use (diesel, L).
- Chemical inputs: pesticide active ingredient applied per hectare.
- Water inputs: volume of water per hectare, combined with regional scarcity factors (Saharanpur, Bharatpur, Alwar).
- Outputs: CO_2 emissions (kg CO_2-eq) and water scarcity index (m^3·scarcity factor).

Environmental burdens were quantified using a hybrid approach combining IPCC Global Warming Potential for CO_2 emissions [20] and the AWARE/WaterGAP methodology for regionalized water scarcity [21]. Groundwater pumping intensities were further integrated to reflect energy–water linkages specific to northwestern India, where groundwater depletion is acute. This approach aligns with prior work on UAV spraying sustainability in Indian conditions [22].

LCA was selected as the core sustainability assessment tool because it provides a standardized and widely accepted framework (ISO 14044 [18]) for quantifying environmental impacts across the entire spraying operation. Compared to direct field measurements, LCA enables integration of upstream and downstream processes, including pesticide production, electricity consumption, and water pumping [19, 21]. This holistic perspective is particularly critical in the Indian mustard belt, where groundwater depletion and energy–water trade-offs are key sustainability challenges [15, 35]. Therefore, LCA was the most appropriate approach for capturing both CO_2 emissions and water scarcity impacts in a reproducible and comparative manner.

The impact categories were computed using the following equations:

1. **Energy-related CO_2 emissions:**

$$CO_{2energy} = E_{kWh} \times EF_{grid}$$

2. **Pesticide-related CO_2 emissions:**

$$CO_{2pesticide} = Q_{chem} \times EF_{chem}$$

3. **Water scarcity impact:**

$$WSI = W_{m^3} \times SF_{region}$$

4. **Total CO_2 footprint:**

$$CO_{2total} = CO_{2energy} + CO_{2pesticide} + CO_{2water}$$

where EF_{grid} is the regional electricity emission factor (kgCO_2/kWh), EF_{chem} is the emission factor for pesticide manufacturing (kgCO_2/L), and SF_{region} is the water scarcity factor derived from AWARE/WaterGAP for the mustard belt.

The LCA results generated CO_2 and WSI indicators that served as target variables for the subsequent machine learning pipeline. By integrating LCA outputs into computational intelligence models, the framework bridges sustainability assessment with predictive analytics, enabling reproducible and scalable benchmarking.

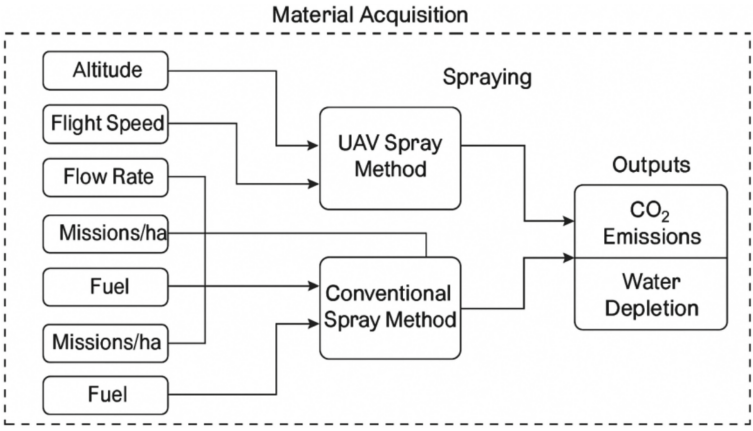


Fig. 1. System boundary diagram of UAV and conventional spraying under the cradle-to-field LCA framework.

2.3 Machine Learning Framework

The outputs of the life cycle assessment (LCA) were further analyzed using supervised machine learning (ML) models to establish predictive relationships between operational variables and sustainability outcomes. ML has been widely applied to environmental and agricultural systems, offering the ability to capture complex nonlinear dependencies and provide decision-support tools [23, 24]. By integrating LCA results into an ML pipeline, this framework aligns with the broader theme of computational intelligence and sustainable computing [25].

The explanatory variables (features) included UAV flight altitude (m), forward speed (m/s), nozzle flow rate (L/min), number of missions per hectare, pesticide application rate (L/ha), energy consumption (Wh), water use (L/ha), and spraying method (encoded: UAV = 0, Conventional = 1). These were selected as key operational drivers influencing energy and water footprints.

Two sustainability indicators derived from LCA served as dependent variables:

- Total CO_2 footprint (kg CO_2-eq/ha)
- Water scarcity index (m^3·scarcity factor/ha)

As shown in Fig. 2, Supervised machine learning was selected to complement LCA results because of its ability to capture nonlinear dependencies between operational variables and sustainability outcomes [5, 24]. In particular, ensemble models such as Random Forest (RF) and Gradient Boosting (GB) were chosen due to their strong generalization capability, robustness to multicollinearity, and interpretability through feature importance analysis [26, 27]. Unlike purely statistical regressions, these models can approximate sustainability indicators across diverse spraying scenarios, providing predictive intelligence suitable for decision support in agriculture. This justifies their selection as computational tools aligned with the goals of sustainable computing and intelligent benchmarking.

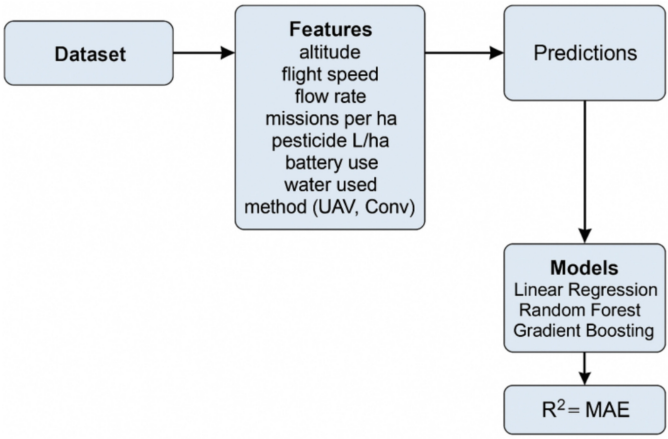


Fig. 2. Machine learning pipeline integrating LCA outputs (CO_2, WSI) with supervised learning models (LR, RF, GB), including feature inputs and evaluation metrics.

Three supervised ML algorithms were applied:

- **Linear Regression (LR)**: baseline parametric model assuming linear feature–response relationships.

$$y = \beta_0 + \sum_{i=1}^{n} \beta_i x_i + \epsilon$$

where y is the predicted indicator, x_i are input features, and ϵ is residual error.

- **Random Forest (RF):** ensemble of decision trees trained via bootstrap aggregation, reducing variance and enabling nonlinear feature interactions [26].

Prediction is given by:

$$\hat{y} = \frac{1}{T} \sum_{t=1}^{T} h_t(x)$$

where $h_t(x)$ is the prediction of tree t and T is the number of trees.

- **Gradient Boosting (GB):** additive ensemble where successive models minimize residual error from previous iterations, improving predictive accuracy on structured data [27].

Model performance was evaluated using the coefficient of determination (R^2) and mean absolute error (MAE):

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2}$$

Data were split into training (80%) and testing (20%) sets with stratification by spraying method. Standardization was applied for LR, while RF and GB were tuned for tree depth and learning rates. Feature importance was derived from RF to interpret the relative contribution of each variable to sustainability outcomes.

2.4 Statistical Validation

To ensure the robustness of the sustainability comparisons between UAV and conventional spraying, a suite of statistical validation methods was applied. Statistical testing has become an essential component in computational sustainability research, providing rigor beyond descriptive metrics and supporting reproducibility of results [28].

Independent-sample *t*-tests were employed to compare UAV and conventional spraying for both carbon footprint and water use at each location. The *t*-statistic was computed as:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

where $\bar{x}_1$, $\bar{x}_2$ are group means, s_1^2, s_2^2 are variances, and n_1, n_2 are sample sizes. One-way ANOVA was then applied to assess overall differences across methods within each trial stage, while two-way ANOVA incorporated both method (UAV vs conventional) and location (Saharanpur, Bharatpur, Alwar) as factors to test for main effects and interactions.

To complement significance testing, Cohen's *d* was calculated to quantify the magnitude of UAV vs conventional differences for effect size calculation [29]. Cohen's *d* is defined as:

$$d = \frac{\bar{x}_1 - \bar{x}_2}{s_p}, s_p = \sqrt{\frac{(n_1 - 1)s_1^2 - (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

where s_p is the pooled standard deviation. Interpretation followed conventional thresholds: $d = 0.2$ (small), 0.5 (medium), 0.8 (large), and >1.2 (very large) [30]. This dual approach of statistical significance and effect size ensured both statistical and practical significance were addressed [31].

The combination of *t*-tests, ANOVA, and Cohen's *d* enabled robust conclusions regarding UAV sustainability advantages. Extremely low *p*-values confirmed statistical differences, while effect sizes consistently exceeded the "very large" threshold, indicating strong practical importance of UAV spraying in reducing both CO_2 footprint and water use.

3 Results

3.1 Life Cycle Assessment Outcomes

The LCA results presented in Table 1 demonstrated clear sustainability advantages of UAV spraying relative to conventional methods across all three locations. UAV treatments required only 15–20 L/ha water compared to 200–300 L/ha for conventional ground-based spraying, representing >90% reduction. Likewise, UAV operations showed a lower carbon footprint (≈1.7–1.9 kg CO_2-eq/ha) compared to conventional spraying (≈3.0–3.4 kg CO_2-eq/ha). These differences were consistent across Saharanpur, Bharatpur, and Alwar.

Figure 3 shows the comparative boxplots for CO_2 and water footprint, highlighting the significant gap between UAV and conventional spraying across all sites.

Table 1. LCA indicators of UAV vs conventional spraying (CO_2 footprint, water use).

Method	CO_2 footprint (kg CO_2-eq/ha)	Water use (L/ha)
UAV	1.7–1.9	15–20
Conventional	3.0–3.4	200–300

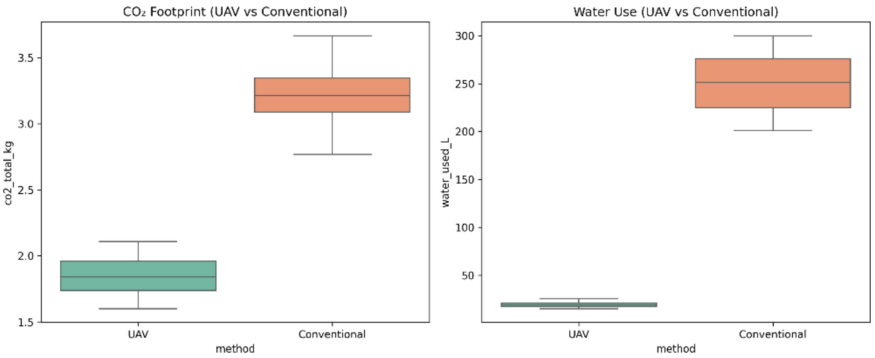


Fig. 3. Comparative boxplots of CO_2 footprint and water use between UAV and conventional spraying across locations.

3.2 Machine Learning Model Performance

ML models were developed to predict sustainability indicators (CO_2 and water) from operational features. Ensemble models (Random Forest, Gradient Boosting) consistently outperformed linear regression as illustrated in Table 2.

- **Linear Regression:** R^2 = 1.0, MAE = ~2e–15
- **Random Forest:** R^2 = 0.998, MAE = 0.021
- **Gradient Boosting:** R^2 = 0.999, MAE = 0.016

Table 2. ML model performance metrics for LCA outcome prediction.

Model	R^2	MAE
Linear Regression	1.000	2e–15
Random Forest	0.998	0.021
Gradient Boosting	0.999	0.016

Feature importance analysis identified water use, flow rate, and flight speed as dominant sustainability drivers, while pesticide application rate had lower influence as shown in Fig. 4, while Fig. 5 shows the parity plots comparing predicted vs actual CO_2 and water indicators.

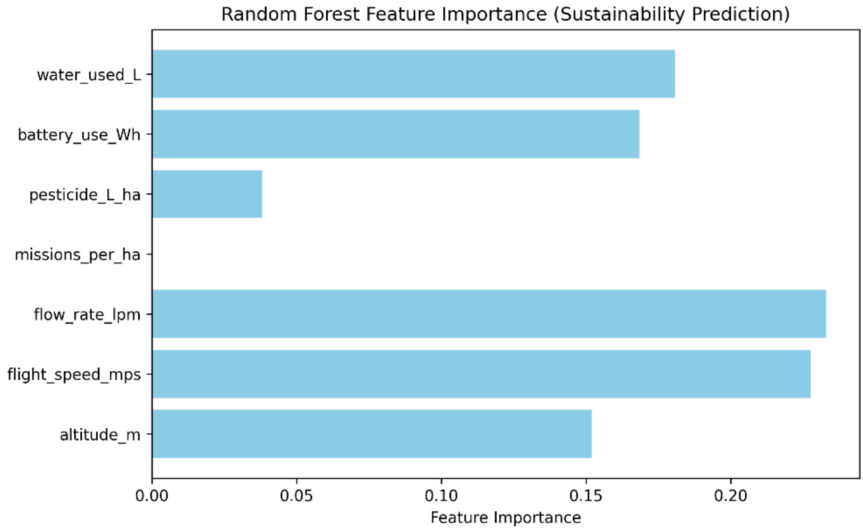


Fig. 4. Feature importance ranking from Random Forest model showing relative influence of operational variables.

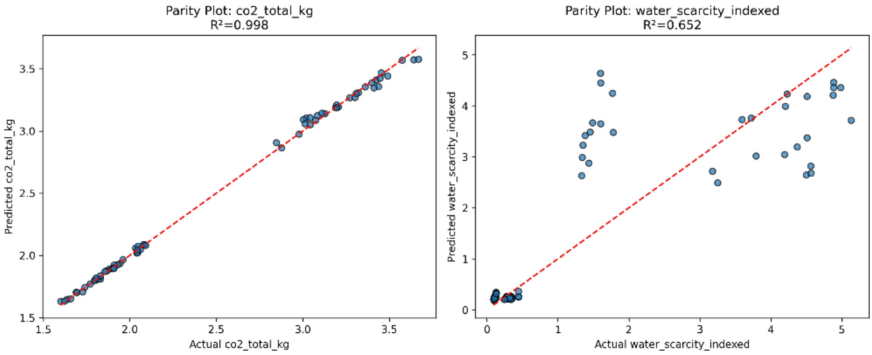


Fig. 5. Parity plots of predicted vs actual CO_2 and water footprints using ML models.

3.3 Statistical Significance of UAV vs Conventional Differences

Independent-sample *t*-tests confirmed highly significant differences between UAV and conventional spraying for both CO_2 and water ($p < 1e{-}49$ across all locations).

Two-way ANOVA further confirmed, as shown in Table 3:

- **Method effect:** highly significant ($p < 0.001$) for both CO_2 and water.
- **Location effect:** not significant ($p > 0.1$).
- **Method × Location interaction:** not significant ($p > 0.5$).

These results presented in Table 3 confirm that spraying method (UAV vs. conventional) is the dominant driver of sustainability differences, whereas location and its interaction with method are statistically insignificant, as shown in Figs. 6, 7, and 8.

Table 3. Two-way ANOVA results for CO_2 and water use (method × location).

Metric	Source	Df	F	p-value
CO_2	Method	1	5621.79	3.30e–204
	Location	2	2.01	0.136
	Method × Location	2	0.53	0.587
Water	Method	1	10601.21	2.95e–246
	Location	2	0.85	0.429
	Method × Location	2	0.55	0.576

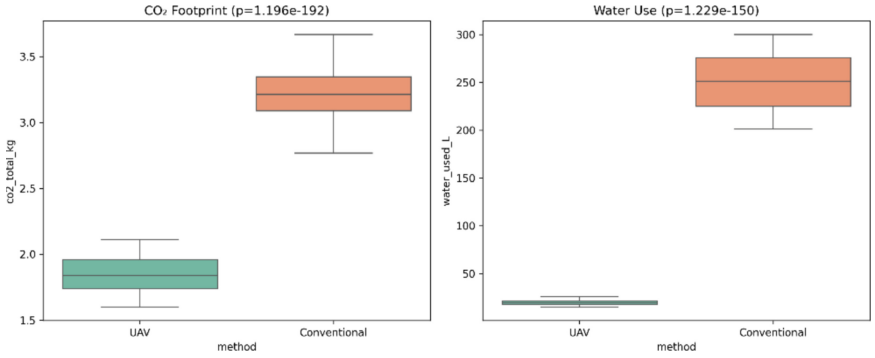


Fig. 6. Boxplots of UAV vs conventional spraying for CO_2 and water with p-values.

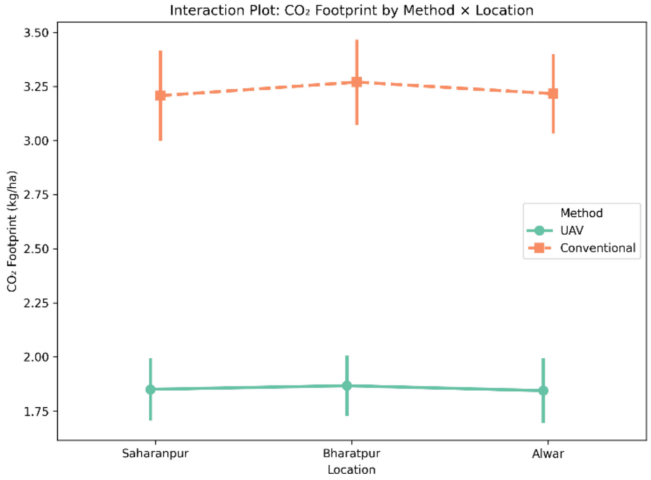


Fig. 7. Interaction plot of method × location effects for CO_2 footprint.

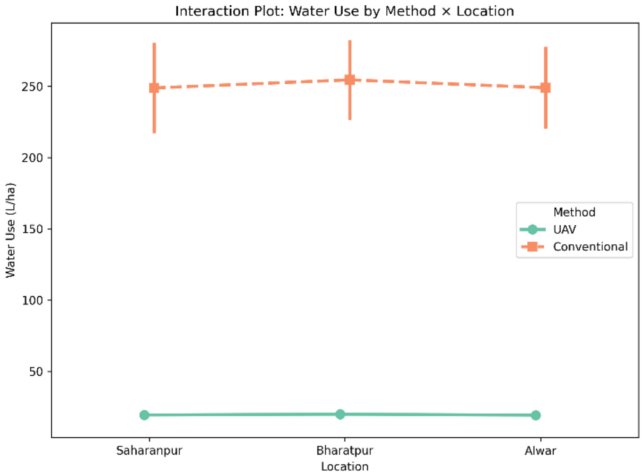


Fig. 8. Interaction plot of method × location effects for water use.

Table 4. Per-location t-test results for UAV vs. conventional spraying across Saharanpur, Bharatpur, Alwar.

Location	Metric	t-value	p-value
Saharanpur	CO_2	−40.97	1.74e–61
	Water	−54.60	9.47e–49
Bharatpur	CO_2	−44.52	2.67e–65
	Water	−63.91	1.56e–52
Alwar	CO_2	−44.59	2.03e–68
	Water	−60.82	3.37e–51

Across all locations, UAV spraying consistently reduced both CO_2 emissions and water use with very large effect sizes (Cohen's d $>$ 7.8) as shown in Table 4. This robustness across agro-climatic zones strengthens the generalizability of UAV-based spraying as a sustainable alternative. Figures 9, 10 and 11 show the per-location boxplots with p-values overlaid.

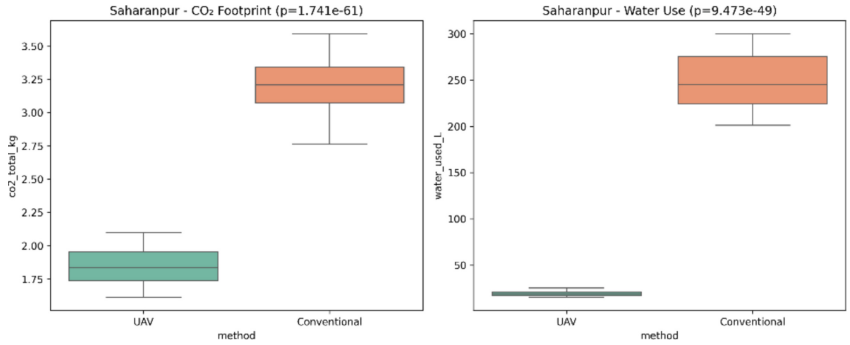


Fig. 9. Boxplot for Saharanpur CO_2 Footprint & Water use.

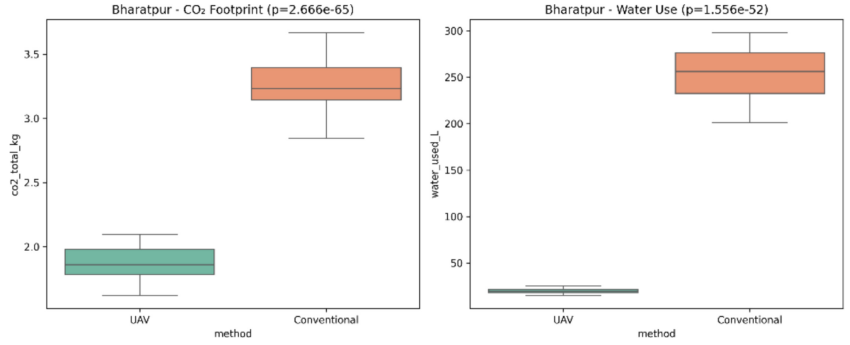


Fig. 10. Boxplot for Bharatpur CO_2 Footprint& Water use.

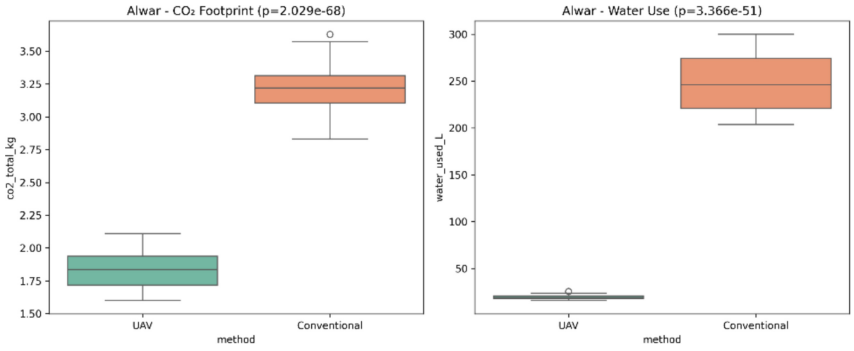


Fig. 11. Boxplot for Alwar CO_2 Footprint & Water use.

3.4 Effect Size Analysis

Beyond statistical significance, Cohen's *d* analysis confirmed very large effect sizes for UAV vs Conventional differences (all $|d| > 7$), as shown in Table 5.

Table 5. Cohen's *d* effect sizes per location and pooled dataset.

Location	CO_2 footprint (d)	Water use (d)	Interpretation
Saharanpur	−7.88	−10.51	Very large
Bharatpur	−8.57	−12.30	Very large
Alwar	−8.58	−11.70	Very large
Overall	−8.32	−11.46	Very large

4 Discussion

The comparative results between UAV-based spraying and conventional ground spraying demonstrate a clear sustainability advantage of UAVs in terms of both carbon footprint and water consumption. The substantial reduction in water use observed across all sites aligns with the principles of ultra-low-volume (ULV) and low-volume (LV) application, which have been recognized as efficient modes of pesticide delivery in precision agriculture [32, 33]. Previous studies have shown that UAV spraying achieves better droplet distribution and canopy penetration while requiring significantly less carrier liquid than conventional hydraulic sprayers [34]. Our findings reinforce these outcomes, particularly in semi-arid mustard-growing regions of North India, where groundwater scarcity is a critical concern [35].

The integration of machine learning into life cycle assessment provides an additional layer of intelligence. By predicting CO_2 and water impacts under varying operational parameters, ML-enabled LCA bridges the gap between static environmental assessments and dynamic, field-specific decision support [36]. This approach represents a novel contribution, as most UAV sustainability assessments have relied on deterministic or descriptive analyses rather than predictive modeling [37]. The strong performance of Random Forest and Gradient Boosting models in single-output predictions highlights their potential for generalizable, robust benchmarking of UAV technologies across diverse crop systems.

From a statistical perspective, both t-tests and ANOVA demonstrated that the sustainability advantages of UAV spraying were highly significant ($p < 0.001$). Importantly, interaction plots and two-way ANOVA confirmed that these benefits were consistent across locations, suggesting that UAV spraying can be scaled without losing efficiency. Very large effect sizes (Cohen's $d > 7.8$) further confirm that the observed differences are not only statistically significant but also practically meaningful [31]. This magnitude of effect is rarely observed in agricultural technology trials, underscoring the disruptive sustainability potential of UAVs.

Nevertheless, certain limitations must be acknowledged. While UAV spraying reduces water and CO_2 impacts, pesticide load per hectare remains a driver of environmental footprint, and further optimization of ULV formulations is required [38]. Additionally, regional variability in electricity mix (coal-dominant vs. renewable-rich grids) could alter UAV carbon footprints, necessitating context-specific assessments [39]. Future research should therefore extend the LCA–ML framework to multi-crop systems, integrate economic dimensions (cost per hectare), and explore hybrid computational

approaches such as fuzzy MCDA or reinforcement learning for adaptive optimization [40].

Overall, this study demonstrates that UAV spraying not only offers direct sustainability benefits but also showcases the role of computational intelligence in advancing precision agriculture. The pipeline approach—merging LCA, ML, and statistical validation—sets a foundation for future interdisciplinary studies in sustainable machine intelligence for agriculture.

5 Conclusion

This study proposed and implemented a machine learning–augmented life cycle assessment framework to benchmark UAV-based mustard spraying against conventional ground spraying in India's mustard belt. The results revealed that UAVs consistently achieved lower CO_2 emissions and dramatically reduced water use, with very large effect sizes confirmed across multiple locations. By embedding predictive intelligence within the LCA framework, the study moves beyond descriptive sustainability reporting to proactive decision support. The findings position UAV spraying as a sustainable alternative aligned with the themes of sustainable computing and machine intelligence. This work underscores how computational pipelines can transform agricultural practices by offering robust, scalable, and data-driven tools for sustainability benchmarking. The contributions of this work are threefold. First, it demonstrates a novel integration of LCA and machine learning, providing a reproducible framework for sustainability benchmarking. Second, it establishes UAV spraying as a more sustainable alternative to conventional methods through robust statistical validation across multiple locations. Third, it showcases the potential of computational intelligence for agricultural decision support, bridging sustainability science with real-time machine learning applications. These contributions collectively advance the state of sustainable computing in agriculture and provide a foundation for future research on intelligent and resource-efficient farming systems. Future directions include expanding the framework to diverse crops, integrating fuzzy or reinforcement-based decision models, and linking environmental metrics with socio-economic indicators for holistic agricultural sustainability.

References

1. Fulcher, J.: Computational intelligence: a compendium. In: Jain, L.C. (ed.) SCI, vol. 115. Springer, Heidelberg (2008)
2. Russell, S.J., Norvig, P.: Artificial Intelligence: A Modern Approach; [the Intelligent Agent Book], pp. I–XXVIII. Prentice hall (1995)
3. Calero, C., Piattini, M.: Introduction to green in software engineering. In: Green in software engineering, pp. 3–27. Springer International Publishing, Cham (2015)
4. Wolfert, S., Ge, L., Verdouw, C., Bogaardt, M.J.: Big data in smart farming–a review. Agric. Syst. **153**, 69–80 (2017)
5. Liakos, K.G., Busato, P., Moshou, D., Pearson, S., Bochtis, D.: Machine learning in agriculture: a review. Sensors. **18**(8), 2674 (2018)
6. Zhang, C., Kovacs, J.M.: The application of small unmanned aerial systems for precision agriculture: a review. Precis. Agric. **13**(6), 693–712 (2012)

7. Lan, Y., Shengde, C., Fritz, B.K.: Current status and future trends of precision agricultural aviation technologies. Int. J. Agric. Biol. Eng. **10**(3), 1–17 (2017)
8. Sassu, A., Gambella, F., Ghiani, L., Mercenaro, L., Caria, M., Pazzona, A.L.: Advances in unmanned aerial system remote sensing for precision viticulture. Sensors. **21**(3), 956 (2021)
9. Henry, R.S., Fritz, B.K., Hoffmann, W.C., Kruger, G.R.: An evaluation of three drift reduction adjuvants for aerial application of pesticides. GSTF J. Agric. Eng. **2**(1), 6 (2015)
10. Zhang, J., Huang, Y., Pu, R., Gonzalez-Moreno, P., Yuan, L., Wu, K., Huang, W.: Monitoring plant diseases and pests through remote sensing technology: a review. Comput. Electron. Agric. **165**, 104943 (2019)
11. Hewitt, A.J.: Droplet size spectra classification categories in aerial application scenarios. Crop Prot. **27**(9), 1284–1288 (2008)
12. Xue, X., Tian, Y., Yang, Z., Li, Z., Lyu, S., Song, S., Sun, D.: Research on a UAV spray system combined with grid atomized droplets. Front. Plant Sci. **14**, 1286332 (2024)
13. Yallappa, D., Veerangouda, M., Maski, D., Palled, V., Bheemanna, M.: Development and evaluation of drone mounted sprayer for pesticide applications to crops. In: 2017 IEEE Global Humanitarian Technology Conference (GHTC), pp. 1–7. IEEE Press, New York (2017)
14. Mukherjee, B., Naskar, M.K., Chakraborty, S., Dutta, S., Mandal, T.K.: building food security and climate resilience: the promise of climate smart pulse production technologies. In: Climate-Smart Agricultural Technologies, pp. 17–43. Springer, Singapore (2025)
15. Panda, D.K., Ambast, S.K., Shamsudduha, M.: Groundwater depletion in northern India: impacts of the sub-regional anthropogenic land-use, socio-politics and changing climate. Hydrol. Process. **35**(2), e14003 (2021)
16. Chauhan, J., Srivastav, J.P., Dadrwal, B.K.: Terminal heat stress on account of late sowing influences growth and development of mustard (Brassica juncea L.) genotypes. J. Pharmacogn. Phytochem. **9**, 521–526 (2020)
17. Sharma, A.K., Thomas, L.: Technology inputs and its impact on farm profits: a case study of rapeseed mustard. Ind. Res. J. Exten. Educ. **13**(3), 9–14 (2013)
18. Iso, I.S.O.:14044: Environmental management—Life cycle assessment—Requirements and guidelines, pp. 1–46 (2006)
19. Guinée, J.B.: Handbook on Life Cycle Assessment: Operational Guide to the ISO Standards, vol. 7. Springer Science & Business Media (2002)
20. Intergovernmental Panel on Climate Change, IPCC: Climate Change 2021 – the Physical Science Basis: Working Group I Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press (2003)
21. Boulay, A.M., Bare, J., Benini, L., Berger, M., Lathuillière, M.J., Manzardo, A., et al.: The WULCA consensus characterization model for water scarcity footprints: assessing impacts of water consumption based on available water remaining (AWARE). Int. J. Life Cycle Assess. **23**(2), 368–378 (2018)
22. Ramteke, S.V., Varadwaj, P.K., Tiwari, V.: Optimizing UAV spraying for sustainable agriculture: a life cycle and efficiency analysis in India. Sustainability. **17**(13), 6211 (2025)
23. Schmidhuber, J.: Deep learning in neural networks: an overview. Neural Netw. **61**, 85–117 (2015)
24. Kamilaris, A., Prenafeta-Boldú, F.X.: Deep learning in agriculture: a survey. Comput. Electron. Agric. **147**, 70–90 (2018)
25. Mocigemba, D.: Sustainable computing. Poiesis Praxis. **4**(3), 163–184 (2016)
26. Breiman, L.: Random forests. Mach. Learn. **45**(1), 5–32 (2001)
27. Friedman, J.H.: Greedy function approximation: a gradient boosting machine. Ann. Stat. **29**, 1189–1232 (2001)
28. Field, A.: Discovering Statistics Using IBM SPSS Statistics. Sage Publications Limited (2024)
29. Cohen, J.: Statistical Power Analysis for the Behavioral Sciences. Routledge (2013)

30. Sawilowsky, S.S.: New effect size rules of thumb. J. Mod. Appl. Stat. Methods. **8**(2), 26 (2009)
31. Lakens, D.: Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. Front. Psychol. **4**, 863 (2013)
32. Qin, W.C., Qiu, B.J., Xue, X.Y., Chen, C., Xu, Z.F., Zhou, Q.Q.: Droplet deposition and control effect of insecticides sprayed with an unmanned aerial vehicle against plant hoppers. Crop Prot. **85**, 79–88 (2016)
33. He, X.: Rapid development of unmanned aerial vehicles (UAV) for plant protection and application technology in China. Outlooks Pest Manag. **29**(4), 162–167 (2018)
34. Chen, H., Lan, Y., Fritz, B.K., Hoffmann, W.C., Liu, S.: Review of agricultural spraying technologies for plant protection using unmanned aerial vehicle (UAV). Int. J. Agric. Biol. Eng. **14**(1), 38–49 (2021)
35. Shah, T., Singh, O.P., Mukherji, A.: Some aspects of South Asia's groundwater irrigation economy: analyses from a survey in India, Pakistan, Nepal Terai and Bangladesh. Hydrogeol. J. **14**(3), 286–309 (2006)
36. Hertwich, E.G., Wood, R.: The growing importance of scope 3 greenhouse gas emissions from industry. Environ. Res. Lett. **13**(10), 104013 (2018)
37. Yang, G., Liu, J., Zhao, C., Li, Z., Huang, Y., Yu, H., et al.: Unmanned aerial vehicle remote sensing for field-based crop phenotyping: current status and perspectives. Front. Plant Sci. **8**, 1111 (2017)
38. Matthews, G.A.: Pesticide Application Methods, 3rd edn. Wiley-Blackwell (2008)
39. Shukla, P.R., Dhar, S., Victor, D.G.: Assessment of demand for natural gas from the electricity sector in India. Energy Policy. **37**(9), 3520–3534 (2009)
40. Doumpos, M., Zopounidis, C.: A multicriteria decision support system for bank rating. Decis. Support Syst. **50**(1), 55–63 (2011)



From Model to Field: Cloud-Based Web Deployment of a Two-Stage Deep Learning Framework for Tea Leaf Disease Detection

Harjinder Singh(✉) and Rizwan Rehman

Centre for Computer Science & Application, Dibrugarh University, Dibrugarh, India
{rs_harjindersingh,rizwan}@dibru.ac.in

Abstract. Tea leaf diseases pose a significant threat to the economically vital tea industry in Assam, India. While deep learning models have shown high accuracy in academic settings for disease detection, a persistent gap remains in translating these models into accessible, field-deployable tools for farmers. This paper addresses this research-to-practice gap by presenting a methodology for the cloud-based web deployment of a validated, high-performance hierarchical two-stage deep learning framework. The model, trained on region-specific data from Dibrugarh, Assam, achieves 96% accuracy in classifying key tea leaf diseases. The proposed deployment architecture leverages Hugging Face Spaces for automated containerization and hosting, paired with a Streamlit-based web application for an intuitive interface. This lightweight strategy simplifies deployment, making the tool readily accessible to non-technical stakeholders via a standard browser. The result is a reproducible, scalable, and user-friendly solution that bridges the divide between algorithmic innovation and agricultural application, offering a tangible decision support tool for tea farmers.

Keywords: Deep-Learning · Tea Disease · Cloud-Deployment · Hugging Face · Streamlit

1 Introduction

The tea industry, centered on the cultivation of *Camellia sinensis*, is a cornerstone of the global beverage market and holds immense economic and cultural significance in India. The state of Assam, particularly the Dibrugarh district, is renowned for its unique agro-climatic conditions that produce high-quality tea. However, these same warm and humid conditions create a fertile environment for a host of plant diseases, such as Brown Blight, Red Rust, and Red Spider Mite, which can devastate crop yields and degrade tea quality. The economic stability of countless farmers is directly threatened by these pathogens.

Effective disease management hinges on early and accurate diagnosis. Yet, traditional monitoring methods rely on manual visual inspection, a process that is often subjective, labor-intensive, and critically dependent on the availability of local phytopathology experts. In regions like Dibrugarh, a scarcity of such expertise often leads to delayed or

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026
S. Lanka et al. (Eds.): ICTSM 2025, LNNS 1755, pp. 408–417, 2026.
https://doi.org/10.1007/978-3-032-13177-5_32

incorrect diagnoses, preventing timely intervention and exacerbating crop losses. This critical need for rapid, reliable diagnostic support has motivated the development of automated systems powered by artificial intelligence.

In recent years, deep learning, particularly Convolutional Neural Networks (CNNs), has emerged as a powerful tool in precision agriculture, achieving remarkable success in plant disease classification tasks. However, a significant research-to-practice gap persists; the vast majority of these high-accuracy models remain confined to academic publications and code repositories, failing to transition into practical tools that can be used by farmers and agricultural extension officers. This paper directly confronts this deployment challenge.

The foundation of this work is a previously developed and validated hierarchical two-stage deep learning framework designed specifically for the Dibrugarh region. This framework first employs a fine-tuned EfficientNetB0 model to robustly filter images, distinguishing between 'leaf' and 'non-leaf' content with 100% accuracy. Subsequently, a custom-designed CNN classifies the verified leaf images into four categories—Brown Blight, Healthy, Red Rust, or Red Spider Mite—with an overall accuracy of 96%. The model's strength lies in its training on a unique, region-specific dataset, enabling it to recognize local disease with high fidelity.

While the algorithmic innovation of this model has been established, its practical value remains unrealized without an effective delivery mechanism. This paper marks a crucial pivot from a focus on *model creation* to one on *model delivery*. The novelty of this work lies not in the creation of a new classifier, but in the detailed presentation of a lightweight, low-complexity deployment methodology that democratizes access to specialized AI models. The primary aim is to present a complete blueprint for the cloud-based web deployment of this specialized two-stage model. The key contributions are:

1. A comprehensive review of recent (2015–2025) research on the cloud and web deployment of AI models in agriculture, identifying current trends and limitations.
2. A detailed description of a lightweight, low-complexity deployment architecture utilizing Hugging Face Spaces and Streamlit, which automates containerization and hosting to lower technical barriers.
3. A practical demonstration of the resulting web application, showcasing its accessibility and ease of use for stakeholders without technical expertise, thereby bridging the research-to-practice gap.

2 Related Work

The deployment of AI models for practical applications in agriculture has seen diverse approaches, ranging from hardware-integrated systems to complex cloud architectures. A comprehensive review by Morchid et al. 2024 highlights the integration of IoT, cloud platforms, and deep learning as a dominant trend for intelligent disease detection [4]. Many solutions focus on combining field-level data capture with cloud-based processing. For instance, Rashid et al. 2024 proposed an IoT-based system for corn disease detection, which, while effective for real-time monitoring, introduces significant hardware deployment and maintenance overhead [1]. Similarly, the work by Indumathi et al.

2017 on using wireless multimedia sensor networks for disease detection underscores the reliance on physical infrastructure, which can be a barrier to widespread adoption due to cost and complexity [5].

Cloud platforms have become central to hosting these AI models, but the choice of platform significantly impacts deployment complexity. Anggraini et al. 2024 demonstrated the use of Amazon Web Services (AWS) to deploy a CNN for rice plant disease classification [3]. While powerful, such Infrastructure-as-a-Service (IaaS) platforms require considerable expertise in cloud configuration, virtual machine management, and networking. A comparative analysis by Rajendran et al. 2023 of major cloud providers like AWS, GCP, and Azure for deploying deep learning models further emphasizes the technical skills required to navigate these environments, making them less accessible for domain experts without a background in DevOps [7]. In contrast to these complex platforms, the demonstrated approach leverages a highly abstracted Platform-as-a-Service (PaaS) solution to eliminate this infrastructural overhead.

The development of the user-facing application is another critical aspect. A review by Goh et al. 2023 discusses various patterns for creating and deploying front-end deep learning web apps, often involving traditional web frameworks like Flask or Django [9]. While flexible, these frameworks demand substantial effort in front-end coding and backend API development. The work of Tiwari et al. 2023 on a web deployment for COVID-19 detection demonstrates the value of accessible web-based diagnostic tools, reinforcing the relevance of translating complex models into simple interfaces for non-experts [10]. This work builds on this principle by using Streamlit, a framework designed for rapid, Python-only development of data apps, which drastically simplifies the process compared to traditional web development stacks.

From a broader methodological perspective, the deployment of deep neural networks is an active area of research. Chan et al. provide a comprehensive review of deep neural networks in the cloud, outlining various applications and challenges, including latency, cost, and security [6]. Furthermore, the systematic review by Wurster et al. on deployment automation technologies highlights the need for structured metamodels to manage deployment complexity [8]. This study addresses these challenges by adopting a highly automated, template-driven deployment platform (Hugging Face Spaces) that aligns with modern DevOps principles but abstracts them away from the user, thus providing a simple, reproducible, and accessible pathway from a trained model to a usable tool.

3 System Architecture

The proposed four-layer architecture as shown in Fig. 1 outlines the system's structure, separating concerns from AI model to user interaction, and promoting streamlined workflow for seamless transition from model to field.

Layer 1: The Model Layer
The intelligent core of the system as shown in Fig. 2, consisting of a pre-trained, validated hierarchical two-stage deep learning framework, handles all inference tasks and consists of two components:

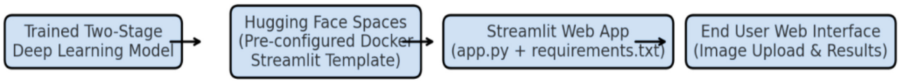


Fig. 1. Deployment workflow of the tea disease detection web app

- **Stage 1:** A fine-tuned EfficientNetB0 binary classifier, which processes an input image to determine if it contains a leaf. This stage acts as a critical filter, ensuring only relevant images proceed to the next stage.
- **Stage 2:** A custom five-block CNN, which takes the verified leaf image from Stage 1 and performs a multi-class classification to identify its condition as Brown Blight, Healthy (Good), Red Rust, or Red Spider Mite. The model artifacts, saved in a standard format (e.g., Keras.tflite), are the primary assets of this layer.

Layer 2: The Platform Layer
This layer demonstrated in Fig. 2 provides the hosting environment and is fulfilled by **Hugging Face Spaces**. Hugging Face Spaces is a PaaS platform that provides a hosting environment for machine learning applications. It abstracts the underlying infrastructure, freeing developers from the need to manage servers, networks, and operating systems. The platform uses **Docker** for containerization, ensuring consistency and reproducibility without manual creation or management of Dockerfiles [2].

Layer 3: The Application Layer
This layer is the frontend portion of the app, which is developed using Streamlit as a web application (app.py) that has UI elements such as a title, instructions, and an image uploader. The user workflow is managed by this code, which receives an uploaded image, sends it to Model Layer for inference, and displays the prediction in human-readable format.(shown in Fig. 2)

Layer 4: The User Layer
The system's outermost layer shown in Fig. 2, representing farmers, extension officers, and researchers, is accessible through a standard web browser on any internet-connected device, eliminating the need for software installation or technical configuration.

The architecture shown in Fig. 2, which includes hugging face space and Streamlight, is based on simplicity, speed and accessibility.It reflects the trend towards democratization of AI, allowing domain experts like agricultural researchers to build and deploy their own applications, lowering the barrier to entry and enabling the translation of academic models into practical tools.

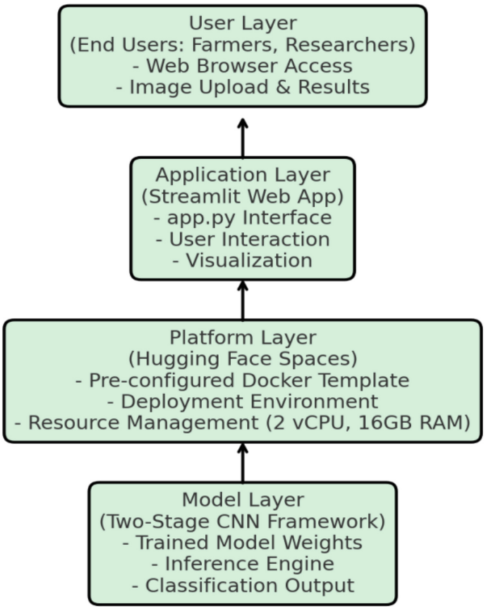


Fig. 2. Layered architecture of the cloud-based tea disease detection deployment

4 Deployment Workflow

The deployment process is designed to be streamlined, reproducible, and highly automated, focusing on simplicity to enable researchers to quickly move from a trained model to a publicly accessible application.

4.1 Model Preparation

The process begins with the final, trained two-stage CNN model developed in previous work. The model is serialized and saved in standard file formats (such as.tflite), creating a portable artifact containing the learned weights and architectures. This file is the core intellectual asset to be deployed.

4.2 Repository Setup in Hugging Face Spaces

A new "Space" is created on the Hugging Face platform. This Space functions as a git-managed code repository. The model file that is being saved, the Streamlit application script (app.py) and the requirement.txt file are all being pushed to this repository. The requirements.txt file is crucial as it explicitly lists all necessary Python dependencies, such as tensorflow, tflite-runtime, streamlit, opencv-python,gdown and numpy, ensuring the environment is perfectly reproducible.

4.3 Configuration

Within the Hugging Face Spaces interface, the application is configured as shown in the user interface of the platform Fig. 3. The SDK is set to **Streamlit** (version **1.44.1** specified for consistency). To promote open science and collaboration, an **Apache-2.0 license** is selected. For the initial deployment, the free-tier **CPU basic**hardware is chosen, which provides **2 vCPU cores and 16 GB of RAM**. This demonstrates that a functional and valuable tool can be hosted without incurring costs, removing a significant barrier for academic projects.

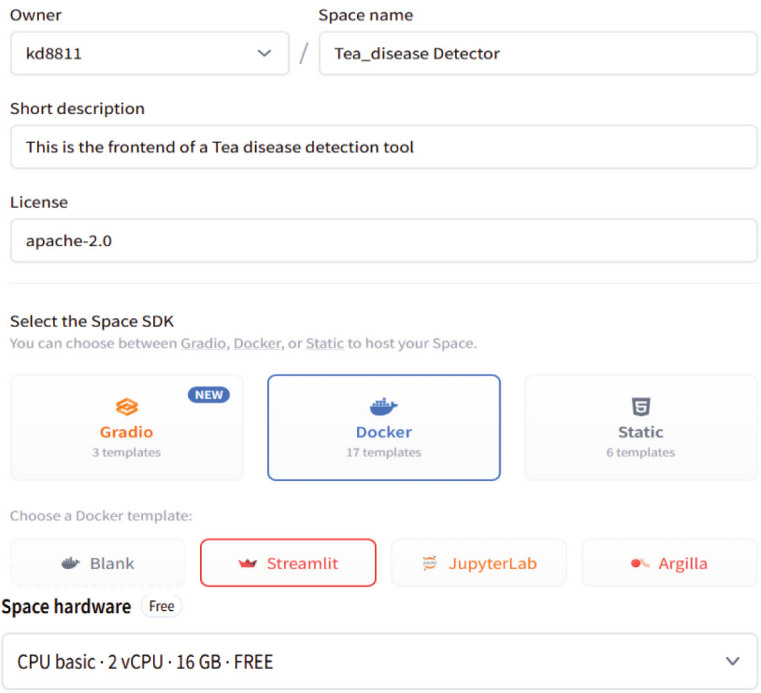


Fig. 3. Snapshot of the setup configuration before deployment

4.4 Application Development and Deployment

The app, developed using the Streamlit library, simplifies user workflow by providing a title, image uploader widget, pre-trained model loading function, and classification button. It also displays the predicted severity and trust score, completing the deployment lifecycle and making the deep learning model available to all. Hugging Face launches the application after pushing the code and configuration files to the repository, starts the build process, reads the requirements.txt, and provision the CPU hardware, and manages server management, networking and security.

4.5 Application Demonstration

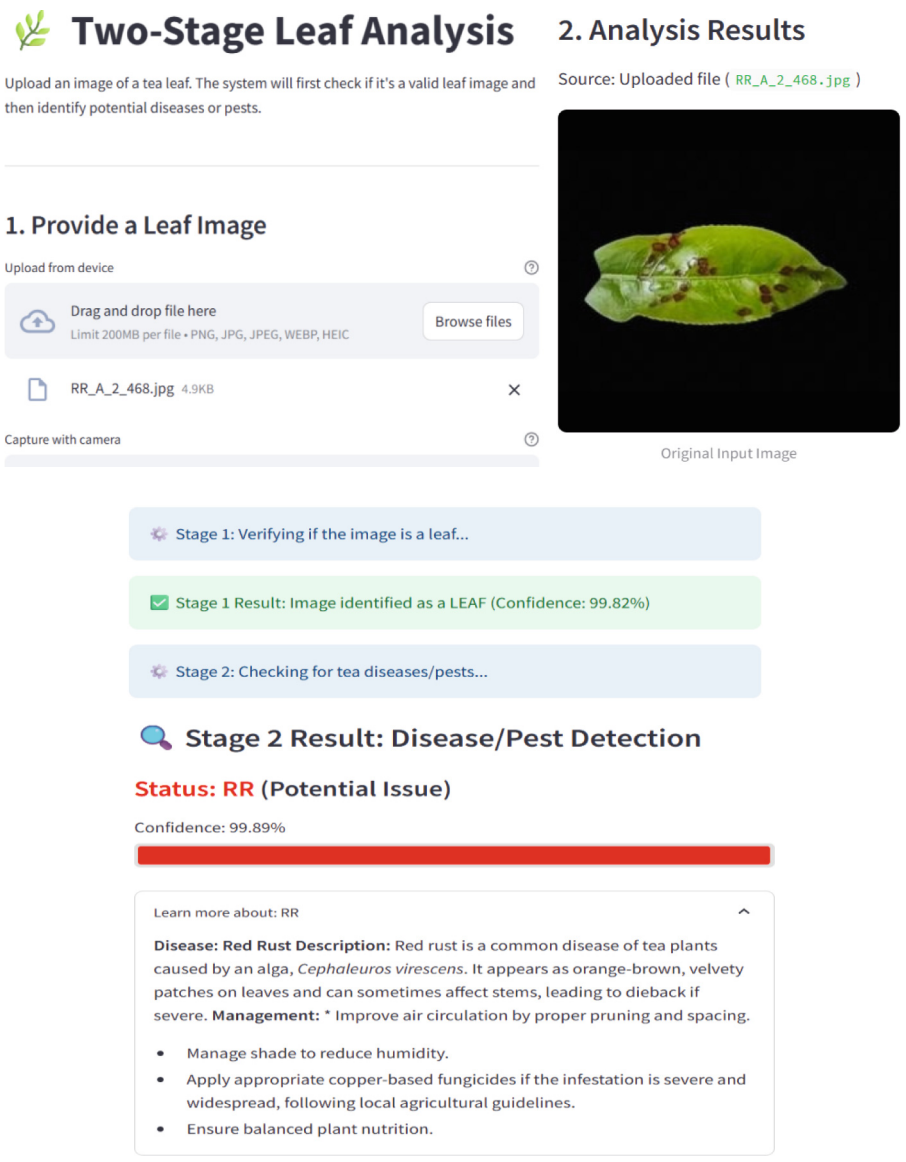


Fig. 4 Snapshots of the working of the App (example of predicted results)

The user journey for Hugging Face Spaces is straightforward, which is shown clearly in Fig. 4, requiring access to the web application, uploading an image, initiating inference, and receiving the result. The user navigates to the public URL, selects a tea leaf

image, and clicks the "predict" button to start the analysis. The image is sent to the Hugging Face Spaces server, where the two-stage model performs the inference. The result is displayed on the screen, providing the predicted condition and the model's confidence in the prediction.

The tool addresses the expertise gap for tea growers in Dibrugarh by providing a datadriven "second opinion" without software installation or complex commands. It empowers farmers to make informed decisions about disease treatment and crop management, fostering a collaborative approach among the farming community, agricultural officers, and researchers.

5 Literature Integration and Discussion

The proposed deployment strategy, using Hugging Face Spaces and Streamlit, offers advantages and trade-offs, focusing on simplicity and accessibility, contrasting with more complex methods, as shown in a structured comparison. Table 1 provides a structured comparison.

Table 1. Comparative Analysis of AI Deployment Strategies

Study	Platform/Technology	Target Application	Deployment Complexity	Accessibility Model
Rashid et al. [1]	IoT + Deep Learning	Corn Disease	High (Hardware Integration)	Field-Specific (IoT device)
Anggraini et al. [3]	AWS + CNN	Rice Disease	Moderate (Custom Backend)	Web/API
Rajendran et al. [7]	AWS, GCP, Azure	General Object Detection	High (IaaS/PaaS Setup)	Researcher-Focused
Goh et al. [9]	Flask/Django + Cloud VM	General DL Apps	Moderate-High	Custom URL (Self-hosted)
This Study	Hugging Face + Streamlit	Tea Leaf Disease	Low (Automated PaaS)	Public URL (Free Tier)

5.1 Strengths of the Proposed Approach

The analysis highlights several key strengths of this deployment model:

- **Rapid Deployment:** The entire workflow, from a trained model to a live web application, is exceptionally fast. By abstracting away infrastructure and container management, the platform allows researchers to deploy in a matter of hours, facilitating rapid prototyping and iterative development. This stands in contrast to the significant setup time required for IaaS platforms like AWS or custom backend development [7].

- **Inherent Reproducibility:** The use of a `git`-based repository combined with explicit dependency management (`requirements.txt`) makes the entire application environment fully reproducible. Any researcher can "fork" the Space, audit the code, and replicate the deployment exactly. This commitment to reproducibility is a cornerstone of open science and a significant advantage over ad-hoc deployment scripts.
- **Cost-Effectiveness:** The ability to deploy on a free CPU tier is a transformative feature for academic research and non-profit initiatives. It removes the financial barrier associated with paid cloud virtual machines or the procurement of IoT hardware, enabling the creation of valuable proof-of-concept and demonstration tools without requiring a budget [1].

5.2 Acknowledged Limitations and Trade-offs

Despite its advantages, the approach involves deliberate trade-offs and has limitations:

- **Performance:** The free-tier CPU hardware is sufficient for demonstration but can result in higher inference latency compared to GPU-accelerated endpoints. For applications requiring real-time processing or handling high volumes of requests, upgrading to a paid GPU tier would be necessary. This is a conscious trade-off of performance for accessibility and cost.
- **Internet Dependency:** As a web-based tool, the application requires a stable internet connection. This can be a significant constraint in remote agricultural areas where connectivity may be unreliable or unavailable. This limitation underscores the need for future work on offline-capable mobile applications.
- **Scalability:** The free tier is not designed for large-scale, production-level traffic. While ideal for individual users, research demonstrations, and small pilot programs, a full-scale rollout across numerous tea estates would require upgrading to paid, auto-scaling hardware tiers to ensure reliability and performance under heavy load.

Ultimately, the practical value of this accessible deployment is inextricably linked to the quality of the model it serves. A generic plant disease model, even if deployed with this user-friendly interface, would likely fail to provide the same level of utility to the farmers of Dibrugarh. The system's effectiveness stems from the synergy between a highly localized, region-specific dataset that ensures the model's accuracy and a deployment architecture that ensures the model's accessibility.

6 Conclusion and Future Work

This research has successfully designed, implemented, and demonstrated a lightweight, cloud-based web deployment for a high-accuracy, two-stage deep learning framework for tea leaf disease detection. The primary contribution of this work is the presentation of a practical and reproducible blueprint for bridging the critical gap between academic AI models and field-ready agricultural tools. By leveraging the streamlined capabilities of Hugging Face Spaces and Streamlit, this work validates a deployment strategy that prioritizes simplicity, reproducibility, and cost-effectiveness. The result is an accessible, easy-to-use web application that places a powerful, region-specific diagnostic

tool directly into the hands of farmers and agricultural officers in Dibrugarh, Assam, thereby empowering researchers and domain experts to translate their findings into tangible societal impact. Future work will focus on addressing the current limitations and extending the system's capabilities. Key directions include, migrating the application to a GPU-powered hardware tier to drastically reduce inference latency and improve the user experience; developing a complementary mobile application using a framework like TensorFlow Lite. This would embed the model directly on a smartphone, enabling offline, in-field diagnosis and overcoming the internet dependency challenge; investigating pathways for a scalable, robust deployment suitable for use across multiple tea estates. This would involve a more resilient backend architecture and a comprehensive data management and feedback pipeline.

References

1. Rashid, R., Aslam, W., Aziz, R., Aldehim, G.: An early and smart detection of corn plant leaf diseases using IoT and deep learning multi-models. IEEE Access. **12**, 23149–23162 (2024). https://doi.org/10.1109/ACCESS.2024.3357099
2. Zhang, Z., Zhang, Y., Xi, X., Li, J.: LuoJiaAI: A cloud-based artificial intelligence platform for large-scale remote sensing image analysis. Geo-spat. Inf. Sci. **26**(1), 44–58 (2023). https://doi.org/10.1080/10095020.2022.2162980
3. Anggraini, N., Kusuma, B.A., Subarkah, P., Utomo, F.S., Hermanto, N.: Classification of rice plant disease image using Convolutional Neural Network (CNN) algorithm based on Amazon Web Service (AWS). J. RESTI. **8**(4), 839–846 (2024). https://api.semanticscholar.org/CorpusID:277971896
4. Morchid, A., Marhoun, M., El Alami, R., El Mouss, H.: Intelligent detection for sustainable agriculture: a review of IoT-based embedded systems, cloud platforms, DL, and ML for plant disease detection. Multimed. Tools Appl. **83**, 70961–71000 (2024). https://doi.org/10.1007/s11042-024-18392-9
5. Indumathi, K., Hemalatha, R., Nandhini, S.A., Radha, S.: Intelligent plant disease detection system using wireless multimedia sensor networks. In: 2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), pp. 1607–1611, Chennai (2017). https://doi.org/10.1109/WiSPNET.2017.8300032
6. Chan, K.Y., et al.: Deep neural networks in the cloud: review, applications, challenges and research directions. Neurocomputing. **545**, 126327 (2023). https://doi.org/10.1016/j.neucom.2023.126327
7. Rajendran, P., Maloo, S., Mitra, R., Chanchal, A., Aburukba, R.: Comparison of cloud-computing providers for deployment of object-detection deep learning models. Appl. Sci. **13**(23), 12577 (2023). https://doi.org/10.3390/app132312577
8. Wurster, M., Breitenbücher, U., Falkenthal, M., Leymann, F., Riegg, A.: The essential deployment metamodel: a systematic review of deployment automation technologies. Softw. Syst. Model. **35**, 63–75 (2020). https://doi.org/10.1007/s00450-019-00412-x
9. Goh, H.A., Ho, C.K., Abas, F.S.: Front-end deep learning web apps development and deployment: a review. Appl. Intell. **53**, 15923–15945 (2023). https://doi.org/10.1007/s10489-022-04278-6
10. Tiwari, R.S., Dandabani, L., Das, T.K., Khan, S.B., Basheer, S., Alqahtani, M.S.: Cloud-based quad deep ensemble framework for the detection of COVID-19 Omicron and Delta variants. Diagnostics. **13**(22), 3419 (2023). https://doi.org/10.3390/diagnostics13223419

Application of Artificial Intelligence in Managing Chloride Stress Corrosion Cracking in Stainless Steel Assets

JohnsonFoyken Jones[1], Veena Raj[1(✉)], Pg Emeroylariffion Abas[1], M. I. Petra[1], and Mohammad Fa'ezul Fitri Latiff[2]

[1] Faculty of Integrated Technologies, Universiti Brunei Darussalam, Jln, Tungku Link, Brunei Darussalam
{23h1301,veena.raj,emeroylariffion.abas, iskandar.petra}@ubd.edu.bn

[2] Brunei Shell Petroleum Sdn Bhd, Panaga, Seria, Brunei Darussalam
m.latiff@bsp-shell.bn

Abstract. Chloride Stress Corrosion Cracking (Cl-SCC) remains a critical degradation challenge for stainless steel assets in offshore and process industries. Conventional inspection and Risk-Based Inspection (RBI) methods often detect damage only after initiation and rely on static assumptions, limiting early intervention. This paper contributes an AI-driven framework for Chloride stress corrosion cracking risk management that (i) integrates sensor data from key variables such as stress, temperature, and chloride concentration,(ii) compares and benchmarks three machine learning models on an experimentally validated corrosion dataset (iii) applies explainable AI (SHAP) that provides interpretability by identifying critical features influencing corrosion which supports auditable results,(iv) outlines a pragmatic approach to embed predictive analytics into inspection planning based on RBI. The proposed approach enables proactive monitoring, adaptive risk estimation, and optimized inspection planning, offering a scalable pathway for predictive corrosion management in industrial environments.

Keywords: Chloride Stress Corrosion Cracking · Risk-Based Inspection · Machine Learning · Explainable AI · Predictive Maintenance

1 Introduction

The long-term safe operations of pressure equipment depend not only on robust design but also on the service conditions. While design rules address aspects of construction, they do not account for in-service degradation mechanisms or fabrication-related defects that may emerge during operation. Chloride Stress Corrosion Cracking (Cl-SCC) is a type of damage that happens to pressure equipment manufactured from certain steel-grade materials, and failure is unpredictable. This type of damage has been categorized under Non-Age-Related damage mechanisms [1]. Around 30% of the pipeline failures that happen in the industry are due to CI-SCC [2]

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026
S. Lanka et al. (Eds.): ICTSM 2025, LNNS 1755, pp. 418–430, 2026.
https://doi.org/10.1007/978-3-032-13177-5_33

In places like offshore platforms, refineries, and chemical plants, equipment often faces exposure to seawater, salty process fluids, or humid, salty air. The usual inspection techniques, like dye penetrant, ultrasonic scans, or eddy current checks, often only catch problems once they've started or in small areas. They don't give a comprehensive or predictive view of Cl-SCC development. Risk-Based Inspection (RBI) tools based on standards like API RP 581 [3], which use probability models to assess risk, are limited by static assumptions, inspection intervals, and the complexity of interacting variables and do not dynamically adapt to changing field data. This work contributes a deployable AI Model for CI-SCC management that is novel in its combination of (i) comparative forecasting: we evaluate Random Forest (RF), SVR and XGBoost to choose the best forecasting model, (ii) Explainability -SHAP is embedded to expose the dominant drivers for degradation, enabling an auditable decision support, and (iii) real time sensor based decision integration – sensor output to ML model assessment to inspection and maintenance planning workflow provides an end to end practical solution for deployment.

2 Related Work and Challenges

2.1 Related Work and Positioning

Recent works on SCC Prediction using ML [4] trained an XGBoost model on SCC data in chloride medium, which showed good quantitative prediction of SCC susceptibility and argues that data driven methods should be used when testing datasets are limited. İn the work [5] SCC growth rate was modeled for 316 and 304 Stainless steels in simulated environments. XGBoost was found to be more accurate compared to other models while highlighting the sensitivity of environmental factors. Recent reviews show the prominence of AI for predicting cracking however certain limitations such as data availability,need for explainability and need for data driven models were highlighted in [6]. Explainable AI has gained popularity to expose feature influences behind predictions [7]. SHAP/ALE was combined with ML to interpret maximum corrosion depth arguing that corrosion studies need to have transparent explanations rather than black box approach.

This paper is well positioned to address the gaps of the existing Works. Existing Works target susceptibility or crack growth in specific chemistries [4, 5]. Whereas this study demonstrates a sensor to model concept using an experimentally validated atmospheric dataset to build a deployable predictive model with explainability features. This model focuses on monitoring key CI-SCC relevant variables, such as stress chloride and temperature to develop the inspection plan based on variations interpreted. Many corrosion papers evaluate one or two algorithms [8]. We have benchmarked XGBoost against RF and SVR revealing that XGBoost was more accurate ($R^2 \approx 0.94$) followed by RF. Several studies report accuracies without explainability [7]. We have embedded SHAP to quantify variable influence and provide auditable results, which is key for engineering and RBI adoption. Few papers have explicitly mapped ML to inspection and maintenance workflows [9]. We have integrated model outputs with RBI and maintenance systems, which is a step towards practical deployment.

2.2 Current Challenges in Cl-SCC Management

Detection and mitigation of this risk have been challenging. This is due to the various challenging factors that exist that are not efficiently addressed. Weak points such as small crevices, stress concentration areas, and beneath insulation are areas where the cracking tends to initiate. Because of these hidden starting points, it is often challenging to detect using standard available testing methods. By the time cracks are spotted, there might already have been considerable damage or leakage [15, 16]. Early detection and regular inspections are crucial to prevent these problems from escalating. The key influencing factors and their typical threshold limits are detailed in Table 1.

Table 1. Critical Factors Contributing to Cl-SCC

Factor	Description	Typical Thresholds	Reference
Tensile Stress	Residual or applied tensile stress promotes crack initiation and propagation	>30–50% of yield strength	API RP 571, [1]
Chloride Concentration	Chloride ions destabilize the passive film, initiating localized attack	>10 ppm (304 SS), >100 ppm (316 SS)	ASTM G36, API RP 939 C [10, 11]
Temperature	Higher temperatures accelerate SCC and chloride activity	Risk rises sharply above 50–60 °C	NACE MR0175 / ISO 15156, API RP 571 [1, 12]
Susceptible Materials	Austenitic stainless steels, cold-worked or sensitized alloys	304/316 SS, sensitized or welded zones	API RP 571, Corrosion Science [1, 13]
Crevices/Deposits	Geometry that traps chlorides and promotes concentration	Not quantifiable – site specific	API RP 571, Field Case Studies [1]

3 Proposed Work: AI-Driven Framework for Cl-SCC Management

3.1 Framework (Data➔Model➔Decision)

Sensors are used to collect data on key parameters defined. There are various techniques available in the market to monitor these key parameters. For Stress, we can use Fibre Bragg Grating Sensors (FBG) and Digital Image Correlation (DIC). FBG provide real-time dynamic strain and stress monitoring, while DIC provide surface strain mapping for validation. For temperature monitoring, thermocouples or IR Wireless sensors can be used to measure the temperature. For monitoring Chloride Ion selective electrodes or smart chemistry sensors monitor the inline concentration in fluids.

Ensemble methods such as Random Forest (RF) and Gradient Boosting (e.g., XGBoost) are applied to model complex, non-linear relationships between stress, temperature, and chloride concentration. The combination of predictive models with anomaly detection provides a robust framework for both forecasting long-term degradation trends and detecting short-term deviations, thereby supporting proactive Cl-SCC management. Real-time risk heatmap that updates itself with the latest data, providing

a clear visual of potential issues as they happen. Inspection can be planned adequately. Maintenance actions can be streamlined. Additionally, this aligns perfectly with existing inspection management tools, such as DNV Synergi or GE APM. A structured comparison with the traditional approaches is provided in Table 2.

Table 2. Traditional RBI vs. AI-Enhanced Cl-SCC Management

Feature	Traditional RBI	AI-Enhanced Approach
Data Source	Historical failure data	Real-time sensor, historical, simulated
Risk Estimation	Static damage factor tables	Dynamic, adaptive ML models
Inspection Planning	Time-based or risk-ranked	Condition-based and predictive
Adaptability to Field Changes	Low	High-learns from new data continuously
Case study prediction	Moderate	High (upto 94% in Simulation data)

This raw data is fed into a data processing system, which assesses the data, validates it and normalises it into a unified dataset. This eliminates the noise and helps in aligning the different measurement scales and time intervals. Next step is **feature engineering,** a key part of the AI model where predictive indicators are derived based on excursions such as duration above chloride threshold levels, thermal stress cycles and residual gradient near welds. A trained machine learning model for time series prediction or anomaly detection with data from historical events and past inspection records and degradation patterns is used. This will calculate a real-time risk index, which is a health score of the monitored location. If the scores deviate beyond a safe limit, alerts are generated and visually represented in a heat map and dashboard, which enables engineers to take appropriate actions. The system also proposes inspection scheduling prioritising high-risk locations based on the current conditions.

In short, the system does not just monitor—it learns, predicts, and prescribes action. This closes the loop between data collection, risk analysis, and preventive maintenance, the key to making Cl-SCC management predictive rather than reactive. AI enables the system to learn over time adapting to asset-specific responses to Cl^{-}, stress, and thermal loads, far beyond what deterministic models can capture. The end-to-end workflow is summarized in Fig. 1.

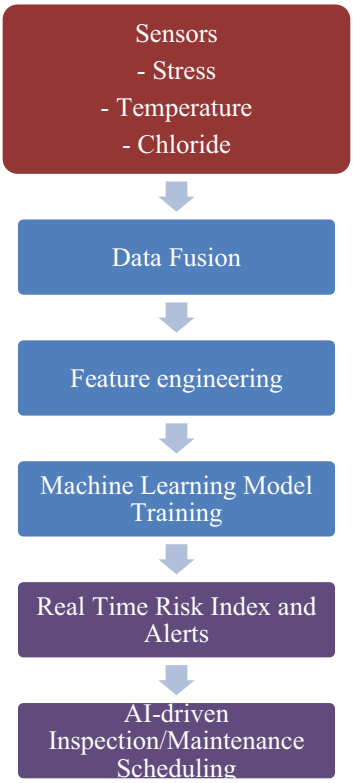


Fig. 1. Predictive AI Model Workflow for CI-SCC

4 Case Study: Predicting Steel Corrosion Using Machine Learning and Explainable AI

To test our AI-driven framework, we applied machine learning to a controlled experimental dataset on atmospheric corrosion of coated steel. The foundational dataset for this study originates from the work [14], Samples were exposed under salt spray in a chamber conditions, with features such as salt loading, deposited mass, crack density, and surface area. Different imaging and analytical tools were used to collect data.

In that study, multiple steel samples with varying coating types were subjected to a controlled environment exposure at different time durations. The researchers used different surface imaging and analytical techniques to capture key corrosion parameters such as corroded area percentage, salt loading, mass deposition and crack patterns. Their idea was to simplify and automate corrosion assessment using image processing techniques thereby ensuring consistent and reproducible measurements of degradation. The datasets generated through their work provide a high-resolution empirical base adequate for subsequent data-driven modelling approaches, such as the machine learning application presented in this study.

The machine learning models and SHAP-based explainability are used to predict the corroded percentage of area on steel samples which are exposed to atmosphere. Samples were subjected to controlled environments and labelled A1, A2, etc., and monitored over a time period in units of days. The surface corrosion of each sample was monitored against the parameters such as salt loading (g/cm^2), mass deposited (g/cm^2), crack density, crack number, and surface area analyzed using advanced surface characterization methods. The objective was to identify how these features influenced the corrosion and use them as input for machine learning regression models. The target variable is the corroded per cent area, which is the measure of the corrosion severity, which also shows the percent reflects the amount of surface area affected by corrosion.

4.1 Methodology

Approach Justification: The critical factors that cause CI-SCC (stress,temperature and chlorides) interact in non-linear ways that make detection and prediction using known non-destructive techniques ineffective. RBI tools, while structured, are based on historical data and fixed degradation assumptions. They cannot dynamically adjust to changes in process conditions such as excursions in temperature, chloride ingress, and upset scenarios. The lack of integration reduces the ability to form a holistic view of evolving Cl-SCC risk. Current inspection-driven strategies tend to focus on detecting damage after it occurs, rather than preventing it in the first place. This results in costly unplanned outages, increased safety risk, and loss of production. We selected XGBoost and SHAP Combo for three reasons. Data Fit, Explainability and Operational deployability.

This study implements and compares three supervised machine learning (ML) regression algorithms, namely Random Forest (RF), Support Vector Regression (SVR), and Extreme Gradient Boosting (XGBoost), to predict the corrosion area percentage of coated steel samples subjected to various environmental conditions. Additionally, explainable AI (XAI) tools are applied to improve model interpretability, specifically using Shapley Additive explanations (SHAP) for XGBoost. Given the variable data regime,decision context that needs accuracy, and deployment needs of sensor output and RBI integration,the XGBoost and SHAP Combination is the best choice. The empirical results (overall best score) and SHAP Analysis outcome substantiate the selection.

Dataset Description and Preprocessing: The dataset comprises time-series corrosion data obtained from a controlled salt spray chamber experiment. Each sample is labelled (e.g., A1, A2...) and includes key features such as: Salt_Loading (g/cm^2), Mass_Deposited (g/cm^2), Crack_Density, Crack_Number, Surface_Area_Analyzed, and Sample_Label (categorical, encoded).

The target variable is Corroded Area Percent, representing the extent of surface corrosion. Data preprocessing included: Encoding the categorical Sample_Label using label encoding, Splitting the dataset into training and testing sets (80:20 ratio), and standardizing the features using StandardScaler for the SVR model to ensure numerical stability.

Machine Learning Models: Three machine learning algorithms were implemented and evaluated to model and predict the corrosion area percentage of coated steel samples

under accelerated environmental exposure. These models were selected based on their proven effectiveness in regression tasks involving non-linear, high-dimensional data. The models were trained on preprocessed features and evaluated using standard regression metrics.

Random Forest Regression. Random Forest (RF) is a robust method comprising multiple decision trees generated during training, and it predicts by averaging their outputs. This randomization of many weak learners (decision trees) effectively lowers variance and prevents overfitting. It is especially adequate for a dataset whose features have intricate and non-linear connections [10]. In the current study, the RF machine learning model was instantiated based on the RandomForestRegressor class of the scikit-learn library with the number of estimators equal to 100 and a hard-coded random seed providing reproducibility. The RF model is favourable when dealing with missing data in the datasets, and feature scaling is unnecessary; thus, the model is computationally effective and explained with the importance score of the feature.

Support Vector Regression (SVR). A kernel machine learning technique based on Support Vector Machines (SVM), Support Vector Regression (SVR) is used to produce outputs that are continuous values. SVR regulates fitting the most suitable line to within the tolerance (epsilon) of the real data points. The algorithm only penalizes predictions not in this epsilon-tube, so that it gives a tradeoff between accuracy and generalization [11]. In this experiment, SVR model was implemented with a Radial Basis Function (RBF) kernel, which works well to capture non-linear relations in the data. Hyperparameters were tuned to obtain the values: C (regularization parameter) $=100$, gamma (coefficient of kernel) $=0.1$ and margin epsilon $=0.1$. The data was enhanced so that the standards of input features were equalized before training due to Standard Scaler to produce the highest level of performance.

Extreme Gradient Boosting (XGBoost). Gradient boosting XGBoost: An implementation of gradient-boosting, built sequentially by optimizing a given loss function. It uses regularization approaches (L1 and L2) in order to prevent overfitting, and it is characterized by its speed and performance because of parallel processing, and optimal tree pruning methods [12]. In the experiment, feature sets were trained using XGBoost with default parameters and fixed random state with the XGBRegressor class of the xgboost library. XGBoost has the following advantage: it is better suited to process structured data where the distributions of features are not homogeneous and where there are missing values. Also, it allows explainable AI integration through SHAP (SHapley Additive exPlanations), which allows understanding feature contributions to predictions

To solve the regression issue of calculating the level of corrosion in the samples of steel, three kinds of machine learning were chosen, including Random Forest, Support Vector Regression (SVR) and XGBoost, as they have a complementary capability in handling structured and of non-linear datasets. The random Forest was selected due to its resistance to overfitting, ability to incorporate the non-linear interactions between features, and the resulting importance measures of the features. SVR was included because of its solid theory and performance in modelling subtle non-linear relationships with kernel functions, specifically, with a radial basis function (RBF) kernel. XGBoost, a gradient boosting framework, was chosen due to its accuracy, scalability, and intrinsic regularization that makes it an effective approach to tabular industrial data. The combined

models can therefore be applied in a broad range, concluding performance trade-offs, as well as explainable AI applications such as SHAP to describe feature weights.

4.2 Results and Analysis

The comparative analysis of three machine learning models namely Random Forest (RF), Support Vector Regression (SVR) and Extreme Gradient Boosting (XGBoost) were used to predict the corrosion of the steel sample. The models were evaluated against various performance measures of multiple regression, Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE) and Coefficient of Determination (R^2).

The results show that XGBoost has better performance compared to others in terms of prediction accuracy and reduction of error. It produced a better R^2 score, strong goodness of fit, and gave the minimal results of MSE, MAE and RMSE, showing light deviation from the actual value. This is revealed by the fact that XGBoost allows dealing with feature interactions and regularize the ML procedure to avoid overfitting, and Random Forest showed decent performance as well, with R^2 values similar as that of XGBoost. Its grouping of decision trees was able to replicate the non-linear interdependencies in the data, but it performed with slightly higher error margins. Nonetheless, Random Forest was accurate and interpretable, and found to be suitable to implement in practice in corrosion prediction systems. In its turn, Support Vector Regression (SVR) demonstrated lower R^2 scores and higher error values, suggesting that it failed to capture the patterns of the input data as well as Random Forest. In spite of hyperparameter tuning ($C = 100$, $\gamma = 0.1$, $\varepsilon = 0.1$) SVR performance might have been limited by the size of the dataset and the presence of noise or non-linearly separable features. The predicted vs. actual performance of each model is plotted in Figs. 2, 3 and 4, while the summary metrics is given in Table 3.

Table 3. Comparison of results

Model	MAE	MSE	RMSE	MAPE (%)	R^2Score
Random Forest	3.336364	25.90322	5.08952	8.299059	0.922503
SVR	6.316748	84.06496	9.168694	15.2479	0.748495
XGBoost	3.247238	18.95156	4.353339	7.536226	0.943301

Feature Influence and Interpretability. To enhance model interpretability and understand the contribution of each input feature in the XGBoost model's predictions, SHAP (SHapley Additive exPlanations) values were computed. Figure 5 provides the aggregate error comparison of the three models while Fig. 6 illustrates the Global SHAP feature importance for the XGBoost model trained, which was recognized as the best model among three. Each point represents a SHAP value for a feature and a particular data instance. The colour gradient indicates the feature value (from low in blue to high in red). From the summary plot, it is evident that Day, Salt_Loading_gcm2,

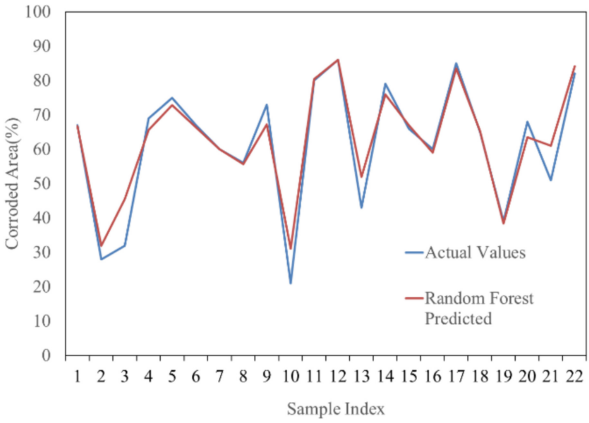


Fig. 2. Random Forest: predicted vs. actual corroded area on test split

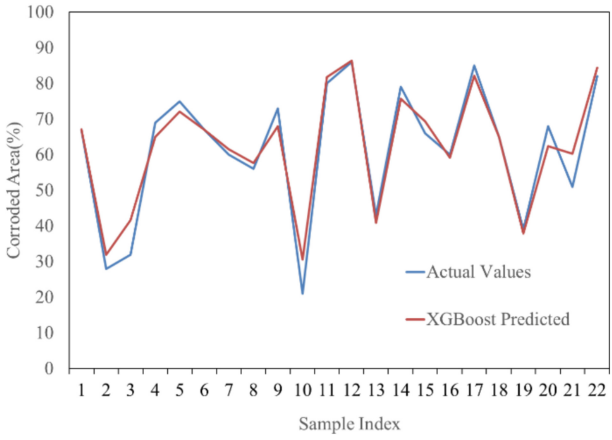


Fig. 3. XGBoost: predicted vs. actual corroded area on test split

and Mass_Deposited_gcm2 are the most influential features impacting the model's output, followed by Crack_Density, Surface_Area_Analysed, Sample_Label_Encoded, and Crack_Number.

Figure 7 visualizes the individual contribution of each feature to a specific prediction using the SHAP (SHapley Additive exPlanations) method. The model prediction $f(x) = 69.23$ is explained relative to the base value $E[f(x)] = 64.80\%$. Among all input variables, the feature Day exerts the highest positive influence on the prediction (+7.2), while Mass_Deposited_gcm2 and Salt_Loading_gcm2 contribute significantly in a negative direction. Other features such as Crack_Density and Crack_Number offer minor contributions. This plot provides a transparent interpretation of how feature interactions affect the model's decision-making process.

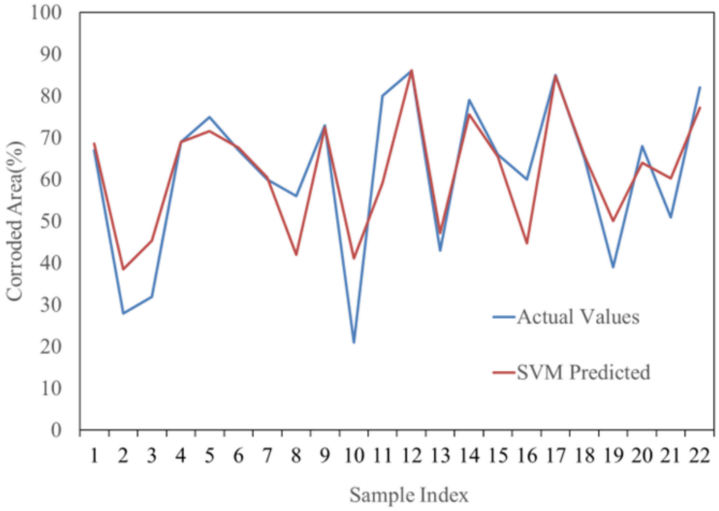


Fig. 4. SVR: predicted vs. actual corroded area on test split

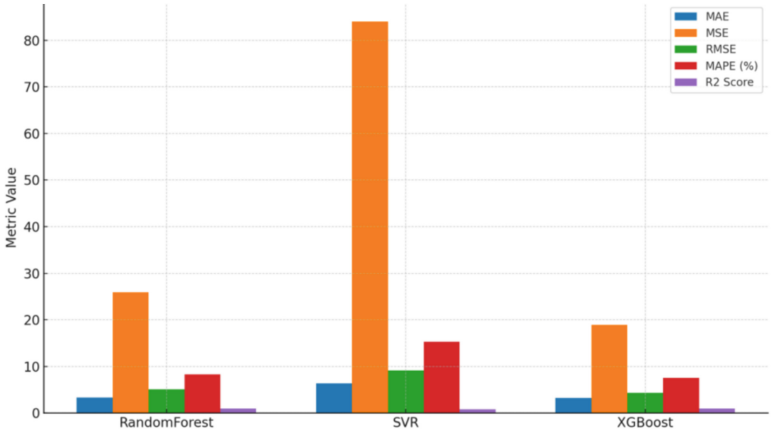


Fig. 5. Error comparison across models (RF, SVR, XGBoost)

Innovations and key Benefits such as **c**ombining predictive modelling with real experimental data, specifically using actual corrosion datasets to train machine learning models, have demonstrated high accuracy, with an R^2 score of approximately 0.94. Explainable AI (XAI) helps build trust and transparency by making predictions easier to understand. Using SHAP analysis, we can identify key drivers of degradation, ensuring our predictions are accurate and interpretable. The approach has been proven effective in addressing atmospheric corrosion, and it can also be applied to Cl-SCCand similar localized damage degradation mechanisms. This helps in early detection and prevention of Cl-SCC initiation. Provide transparent insights into the reasoning of results, thereby enhancing confidence in the models. This combination provides accurate predictions

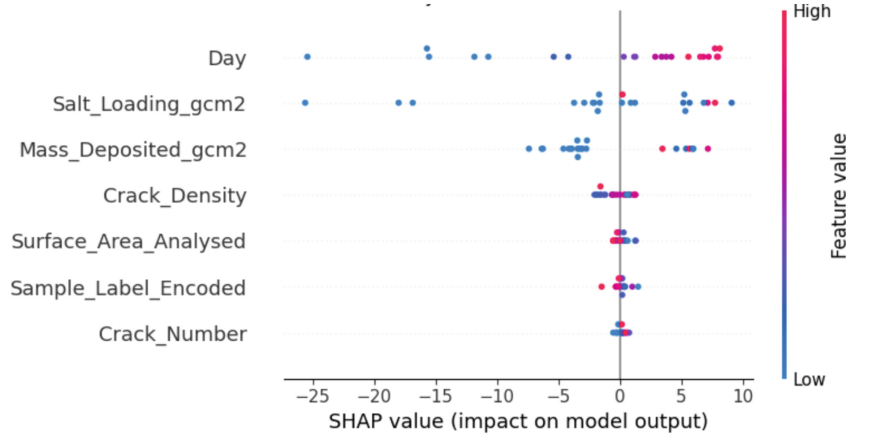


Fig. 6. Global SHAP Feature importance for XGBoost

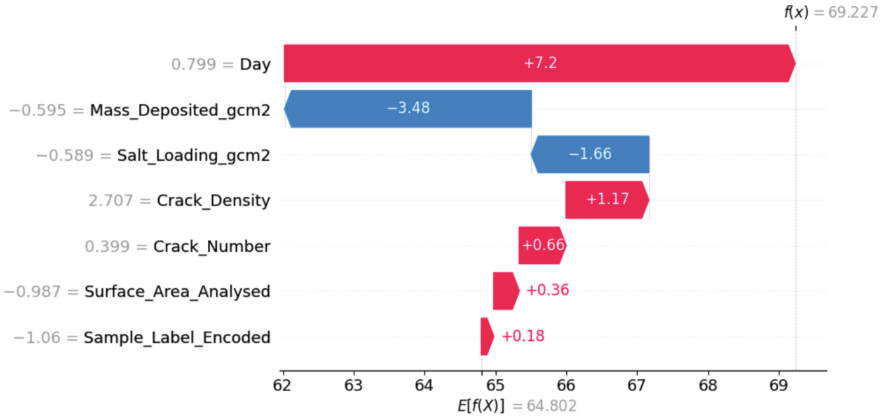


Fig. 7. SHAP Waterfall Plot for representative prediction (XGBoost)

with transparent justifications and direct deployability to RBI Workflows. Moreover, it is scalable to various asset classes (piping, heat exchangers, pressure vessels). Provide enhanced confidence in decision-making through data-driven insights. Limitations are the High initial cost of sensor deployment and data infrastructure, Dependence on the quality and volume of historical failure data for training. It requires specialised AI expertise and cross-disciplinary teamwork. A few of the challenges of implementation are integration with legacy inspection and maintenance systems., resistance to change from conventional RBI approaches, and data security and standardization across platforms. Future opportunities for improvement would be to integrate with digital twin platforms for 3D asset modelling and predictive integrity simulations. Utilise reinforcement learning models for self-improving inspection strategies and use of cloud-based AI platforms that combine data across fleets or facilities for continuous learning.

5 Conclusion

Chloride Stress Corrosion Cracking (Cl-SCC) is an unpredictable degradation problem for oil and gas stainless steel assets. Traditional inspection and risk assessments only provide partial insight, leaving gaps in early detection and making proactive planning challenging. This study presents an AI framework for CI-SCC management that integrates a regression models (XGBoost) having the highest predictive accuracy ($R^2 \approx 0.94$). A SHAP based explainability approach exposes the most influential features in predicting corrosion progression. This integrated machine learning and explainability framework demonstrates a scalable and reliable approach for forecasting corrosion, paving the way for a shift from reactive, inspection-driven practices to predictive,data-driven corrosion monitoring in infrastructure and materials engineering applications. Our contribution is an AI-based accurate CI-SCC Framework that can be embedded into the RBI workflow to predict failure and provide outputs for inspection and maintenance prioritization. Although we validated this approach with atmospheric corrosion data, the same methods can be applied to chloride stress corrosion cracking in industrial environments, where similar environmental and variable factors are involved. Looking ahead, the key is to connect laboratory results with real-world applications by integrating AI with sensor networks, risk-based inspection systems, and digital twin technology. Done right, this approach can lead to safer operations, smarter inspection schedules, and longer-lasting assets.

References

1. Hadzihafizovic, D.: Damage mechanisms affecting fixed equipment in the refining industry (2023) [Online]. Available: https://papers.ssrn.com/abstract=4781875 Accessed 02 Sep 2025
2. Scully, J.R.: Effects of stress – environment assisted cracking. In La Que's handbook on Marine Corrosion, pp. 239–289 (2021). https://doi.org/10.1002/9781119788867.CH10.
3. API: API 581: risk-based inspection methodology. American Petroleum Institute (2016)
4. Abraham Rojas, Z., Bakhtiari, S., Aldrich, C., Calo, V.M., Iannuzzi, M.: XGBoost model for the quantitative assessment of stress corrosion cracking. npj Mater. Degrad. **8**(1), 1–19 (2024). https://doi.org/10.1038/s41529-024-00538-7
5. Wang, P., Wu, H., Liu, X., Xu, C.: Machine learning-assisted prediction of stress Corrosion crack growth rate in stainless steel. Crystals. **14**(10), 846 (2024). https://doi.org/10.3390/CRYST14100846
6. Vakili, M., Koutník, P., Kohout, J., Gholami, Z.: Analysis, assessment, and mitigation of stress corrosion cracking in austenitic stainless steels in the oil and gas sector: a review. Surfaces. **7**(3), 589–642 (2024). https://doi.org/10.3390/SURFACES7030040
7. Song, Y., et al.: Interpretable machine learning for maximum corrosion depth and influence factor analysis. npj Mater. Degrad. **7**(1), 1–15 (2023). https://doi.org/10.1038/s41529-023-00324-x
8. Fang, J., Cheng, X., Gai, H., Lin, S., Lou, H.: Development of machine learning algorithms for predicting internal corrosion of crude oil and natural gas pipelines. Comput. Chem. Eng. **177** (2023). https://doi.org/10.1016/j.compchemeng.2023.108358
9. Campari, A., Vianello, C., Ustolin, F., Alvaro, A., Paltrinieri, N.: Machine learning-aided risk-based inspection strategy for hydrogen technologies. Process Saf. Environ. Prot. **191**, 1239–1253 (2024). https://doi.org/10.1016/J.PSEP.2024.09.031

10. Standard Practice for Evaluating Stress-Corrosion-Cracking Resistance of Metals and Alloys in a Boiling Magnesium Chloride Solution 1. https://doi.org/10.1520/G0036-24.
11. American Petroleum Institute: API RP 939-C: Guidelines for Avoiding Sulfidation (Sulfidic) Corrosion Failures in Oil Refineries, Washington, DC (2019)
12. International Organization for Standardization; NACE International: Petroleum and natural gas industries—Materials for use in H_2S-containing environments in oil and gas production, Geneva (2020)
13. García, C., Martín, F., De Tiedra, P., Heredero, J.A., Aparicio, M.L.: Effects of prior cold work and sensitization heat treatment on chloride stress corrosion cracking in type 304 stainless steels. Corros. Sci. **43**(8), 1519–1539 (2001). https://doi.org/10.1016/S0010-938X(00)00165-7
14. Scatigno, G.G., Dong, P., Ryan, M.P., Wenman, M.R.: The effect of salt loading on chloride-induced stress corrosion cracking of 304L austenitic stainless steel under atmospheric conditions. Materialia. **8**, 100509 (2019). https://doi.org/10.1016/J.MTLA.2019.100509
15. Breiman, L.: Random Forests. Mach. Learn. **45**, 5–32 (2001)
16. Drucker, H., Burges, C.J.C., Kaufman, L., Smola, A., Vapoik, V.: Support Vector Regression Machines.

Leaf Disease Detection and Classification Using Deep Learning Technique and Deep Convolution Neural Network

R. Sumathi(✉), Nalladimmu Sasisri, Kunchapu Gowri Sai, Kapilavai Hahumaan, and Kurakula Lokesh

Department of Computer Science and Engineering, Kalasalingam Academy of Research and Education, Krishnankoil, Tamil Nadu, India
suchandika@gmail.com

Abstract. Plant Leaf plays an important role in the first identification and accurate classification of diseases to ensure crop health, reduce agricultural loss and support food security in agricultural communities. we present a deep learning -based system for detecting and classifying plant diseases using a specially designed deep convolutional neural network. Unlike the pre-trained transmission models that require high calculation resources, which were developed from scratch to provide real-time, effective and accurate predictions suitable for mobile and online purposes. The model was trained and validated using a publicly available plant inguinal data sets, including thousands of healthy and diseased plant leaves from many crop types, including tomatoes, potatoes, corn and more. Our approach includes image presentation technology, computer text and waiver-based regularization to increase model generalization on unseen data. The trained model gained more than 98% accuracy on the verification set and demonstrated strong classification performance in different types of plants. To improve access and purpose, an online gradio interface was integrated, allowing users to upload leaf images and the disease name, throne levels, symptoms, precautions and proposed treatment, as well as predictions. It covered the gap between the user-focused interface machine learning research and practical agricultural application. The proposed solution is adapted to placement in the resource environment and has a significant ability to help farmers, agricultural workers and decision makers in the monitoring of the disease and crop management. Integration of AI-driven agriculture not only reduces the dependence on manual inspection but also contributes to a large measure of permanent agriculture and aids to improve multilingual support from partially infected or noisy leaf images.

Keywords: Deep Learning · Plant Disease Detection · Leaf Classification · Deep Convolutional Neural Network · Agricultural AI · PlantVillage Dataset · Image Processing · Crop Health Monitoring · Precision Agriculture · Gradio Interface · Disease Diagnosis · Smart Farming · Agricultural Automation

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026
S. Lanka et al. (Eds.): ICTSM 2025, LNNS 1755, pp. 431–445, 2026.
https://doi.org/10.1007/978-3-032-13177-5_34

1 Introduction

Agriculture has always been the backbone of human civilization, especially in countries such as India, where an important part of the population depends on farming for livelihood and redundant repetition. One of the most frequent and harmful challenges, farmers have to face, which is an outbreak of plants, which can spread rapidly, destroy giant agricultural areas, and reduce more than half of crop yield. The issue is particularly disastrous for small and marginal farmers, who have a lack of expert consultation or laboratory-based diagnosis. Traditional methods of the identification of the disease are often slow, error-prone and limited by the availability of rural agricultural experts. In addition, some diseases display highly similar visual symptoms, making manual diagnosis difficult without advanced equipment. In this context, Artificial Intelligence especially deep learning, has emerged as a transformative approach in the problem-solution.A particular form of deep learning, deep learning, has shown excellent abilities in image classification functions in domains such as healthcare, monitoring and plant pathology. These models can automatically extract relevant features from raw image data, eliminate manual feature engineering requirement and improve clinical accuracy. Inspired by this capacity, our research applies a custom-made DCNN model-trained-specifically to detection and classification of leaf disease. Unlike previous traditional CNN on general datasets such as imageNets, our model is sewn and trained on the plantvillage dataset, with more than 50,000 annotated images of healthy and diseased leaves in various crops. To enhance the strength and generalization of the model, we employed techniques such as image pretrained, data augmentation, batch normalization and dropout regularization, gained more than 98% of a validation accuracy during training on Google Colab with GPU acceleration. Beyond model development, we focused a huge focus on access to the trained DCNN in a Gradio-managed web application. This allows interactive interface users to upload leaf images and to get immediate disease predictions with plant names, disease labels, confidence scores, visible symptoms, precautionary measures and suggested treatments. This system is particularly beneficial for farmers, agricultural officials, students and researchers, who require real-time, cost-effective and accurate diagnosis. By distributing a lightweight, scalable and mobile-friendly AI solution, the project pulls the gap between academic research and practical agricultural application, indicating how to support sustainable farming by reducing unnecessary pesticides through accurate disease detection. Ultimately, this task presents an innovative yet deployed DCNN-based approach to plant disease classification that emphasizes the effects of the real world. The purpose of future work is to include the image data of the real-world, expand the support for more plant types, and integrate multilingual abilities to make the solution accessible on a comprehensive user base.

2 Related Works

In the last decade, Deep Learning has played a rapid major role in the automatic identity and classification of plant leaf diseases, due to its ability to extract complex visual patterns without manual intervention. Several studies have detected various architecture and approaches to improve the accuracy, efficiency and purpose of such systems

in real-world agriculture. [1] applied a better deep learning model to find and classify the plant leaf diseases. This showed that custom could improve the performance of architecture detection [2] created a dense CNN-based structure that can classify many diseases using leaf images and showed that it works well on the structured dataset. These models suggest that learning deep feature can be better than traditional image processing methods. [3] combined image preprocessing techniques with deep CNN to increase disease recognition capabilities, especially when image quality or light variety. [4] Leveraged mobileNet-based CNN emphasized speed and portability without significant damage in accuracy to create a mild model suitable for mobile certification. [5] presenting a wide comparison between machine learning and deep learning methods, it concluded that deep CNN architecture consistently makes traditional algorithms better than traditional algorithms when large anotets are trained on dataset. [6] plant's pathology was one of the early adoptive eclipses of deep learning, using standard CNN for leaf image classification and installed a benchmark for the detection functions. [7] Recent progress in model efficiency has been highlighted at Atila et al. Where EffectiveNet architecture gained better accuracy through compound scaling of depth, width and resolution. [8] introduced a squeeze-and-an-re-determined CNN to detect diseases, enhance the focus mechanisms of the network and improve performance on challenging samples. [9] detection of various CNN-based methods to detect specific crop diseases such as apples and rice leaves, which offers adapted activation functions to increase model sensitivity. [10] applied adaptation strategies for deep learning models for rice plant disease classification, which reflects the trend of more adaptive and scalable solutions. [11] stressed how important it is to use custom-built CNNs instead of generic pretrained models. They said that training on data that is specific to a certain domain leads to better feature extraction. [12] made the FieldPlant dataset available to fix the problems with lab-controlled datasets like PlantVillage. They said that adding images from real fields would help models work better in general.our work presents a light, custom DCNN, which is trained with scratches on the plantvillage dataset, receiving more than 98% verification accuracy. In addition, we integrate the model into a gradio-based interface and bridge the gap between research and real-world applications that provide users the prophecies, treatment advice and farming suggestions to the users, making it highly practical for farmers and agricultural laborers.

3 Methodology

The proposed functioning for the detection and classification of the plant leaf disease integrates deep teaching techniques with a user-centered deployment to offer an end-to-end solution that is both accurate and accessible. We began using publicly available plantwilze dataset, including more than 50,000 anotate images of healthy and diseased leaves from various plant species. The dataset was pre-proposed by implementing data growth techniques such as rotation, zoom and flipping to shape images for 128 $\times$ 128 pixels and increase variability and normalization. A custom Deep Convisional Neural Network was then designed and trained by scratch using tensorflow and causes. The architecture consisted of three convisional blocks, each after the batch normalization and maximum-pooling layers, with the last dense layer of 256 neurons and a softmax

output layer to suit 38 disease classes. To improve training stability and reduce overfiting, dropouts and generalization layers were employed. The model was trained and valid on the 80:20 partition using GPU acceleration on Google Colab, which acquires more than 98% verification accuracy. To ensure practical utility beyond academic results, trained DCNN was integrated into a light, web-based interface using gradio. This interface enables users to upload leaf images and achieve real-time disease predictions with confidence scores, visible symptoms, precautionary measures and suggested treatments. By combining strong model training with user-accessible clinical equipment, this integrated approach addresses both technical and purpose challenges in detection of AI-managed plant disease for farmers, agricultural workers and researchers.

3.1 Dataset Collection

For this model, we used the publicly available plantvillage data set, which is one of the broadest and widely adopted data sets to detect plant disease. Developed by Hughes and Saltehe, this dataset has become a measure of agricultural-Ai due to its diversity, quality and access.

The dataset contains more than 54,000 images with high resolution of 14 different crop species, such as tomatoes, potatoes, corn, grapes, apples, vine chili and more. These images cover 38 different classes, including healthy leaves and early flash, late blights, Septoria blade spots and a wide selection of common plants such as powder semi. A consolidated background (usually white or gray) occupied, focus on the same magazine for clarity, that are personally marked with the crop type and the name of the disease and balanced in most classes to avoid important data imbalance.

Table 1. Image distribution sample

Crop	Disease Class	Number of Images
Tomato	Late Blight	1909
Tomato	Leaf Mold	952
Potato	Early Blight	1000+
Apple	Apple Scab	630
Grape	Black Rot	1180

Table 1 shows the Image distribution sample. Plant Village dataset consists of 14 types of crops with 38 different disease classes and consisting of around 54000 images. In the above table, we mentioned a few plants and the disease classes and the number images existing in the combination of a particular plant with a particular disease class.

3.2 Data Preprocessing

Data preparation is an important step in any deep learning pipeline, as the quality and stability model of the input data significantly affects performance. Among several preprocessing techniques, we used median filtering were implemented to prepare a Plantvillage

dataset to train a strong and effective deep conversion Neural network. To enhance the needed information and to normalize the dataset, increase the model's normalization and reduce the effect of overfitting. The raw images in the Plantvillage dataset come in different suggestions. To standardize the input of the model and optimize exercise efficiency, all images were shaped up to 128×128 pixels. This size was chosen as a balance between preserving significant visual properties and reducing calculation complexity. The size was changed using a bilinear projection. Since the dataset contains the area marked (eg "Tomato__leaf_mold," "Apple___Scab", it was necessary to convert them into numerical format. Label coding was used to assign a unique integer index to each category, followed by A-HOT coding to prepare data for a multi-class data.

3.3 Data Augmentation and Dataset Splitting

In order to increase the diversity of training data and reduce the risk of overfit, data enlargement techniques were implemented dynamically during training using Keras Image Data Generator. This allowed the model to learn irreversible features and especially normally normalized leaf images ignored from real world conditions. Random Rotation (± 30 degrees) Helps the model recognize diseases at different leaf orientations, Horizontal & Vertical Flipping Simulates leaves captured from different sides, Zoom Range (0.8–1.2) Introduces scale invariance, Shear Transformation Adds slight geometric distortion and Brightness Adjustment Handles varying lighting conditions, augmentation was only applied to the training set, not the validation set, to ensure unbiased evaluation. The dataset was divided into 80% training data and 20% confirmation data. This stratified division made each class represented proportional to both the subsets, which is important for balanced training and evaluation of the fair model. During model training, the data was loaded into the mini-batch (batch size $= 32$) and the model was shed again at the beginning of each era to prevent clear order-related patterns. This improved training stability and convergence.

3.4 DCNN Model Architecture

The core of this assignment is a custom-constructed Deep Convolutional Neural Network mainly designed for multi-elegance classification of plant leaf illnesses. Unlike pretrained fashions like ResNet or VGG, this version became developed from scratch to hold a lightweight structure while nevertheless accomplishing high type accuracy. The architecture balances performance, speed, and computational performance, making it suitable for real-time inference and deployment on low-resource structures like mobile or net programs. The DCNN was constructed with the goals to learn the spatial and textural features from leaf images efficiently, to fasten inference speed for real-time predictions and reduce memory footprint to enable deployment on devices with limited hardware High validation accuracy across diverse classes of diseases. The model accepts RGB images of size $128 \times 128 \times 3$ as input and processes them through multiple layers of convolutional, activation, and pooling operations, followed by dense classification layers.

Table 2. Layer-wise architecture of the proposed DCNN model

Layer Type	Details
Input Layer	128 × 128 × 3 RGB image
Conv2d (1)	32 filters, 3 × 3 kernel, ReLU activation
Maxpooling2D	2 × 2 pooling window
Conv2D (2)	64 filters, 3 × 3 kernel, ReLU activation
MaxPooling2D	2 × 2 pooling window
Conv2D (3)	128 filters, 3 × 3 kernel, ReLU activation
MaxPooling2D	2 × 2 pooling window
Flatten	Converts 3D feature maps into 1D vector
Dense (Hidden)	128 neurons, ReLU activation
Dropout	0.5 dropout rate (to reduce overfitting)
Output Layer	Softmax activation with 38 neurons (one per class)

Table 2 represents the layer-wise architecture of the proposed DCNN model. The proposed DCNN model consists of different types of layers such as input, conv2d, max-pooling, flatten, dense, dropout and output layer. Each layer has its significant purpose that helps in increasing the capacity of the model to learn more complex and hierarchical features from the data and improve the accuracy of the model. ReLU (Rectified Linear Unit) is used in all convolutional and dense layers to introduce non-linearity. The Softmax function is used in the output layer to compute class probabilities across the 38 leaf classes. The model was compiled using the Adam Optimizer that is chosen for its adaptive learning rate and efficiency, Categorical Cross entropy Loss Function which suitable for multi-class classification and Accuracy various metrics like 9 layers for both input and output and nearly 1 million trainable parameters and used for monitoring training and validation performance.

3.4.1 Model Visualization

Total params are 3,307,542 in which trainable params are 3,307,542 leaving non-trainable params at 0.

Figure 1 represents the proposed DCCN Architecture begins with an input layer of 128 × 128 × 3 images, followed by Conv2D, batch normalization, and three convolutional blocks containing maxpooling layers that removes progressive features while reducing spatial dimensions. The final feature maps are levelled into a vector of 25,088 units, which serves as an input for a dense layer of 256 neurons for high-level feature learning. A dropout layer is included to prevent overfitting during training and improve generalization. The network softmax ends with a dense output layer of 15 neurons with activation, which enables multi-class classification of leaf diseases. This custom DCNN architecture demonstrates a sturdy exchange-off between complexity and accuracy. Once the statistics became preprocessed and the DCNN architecture finalized, the schooling

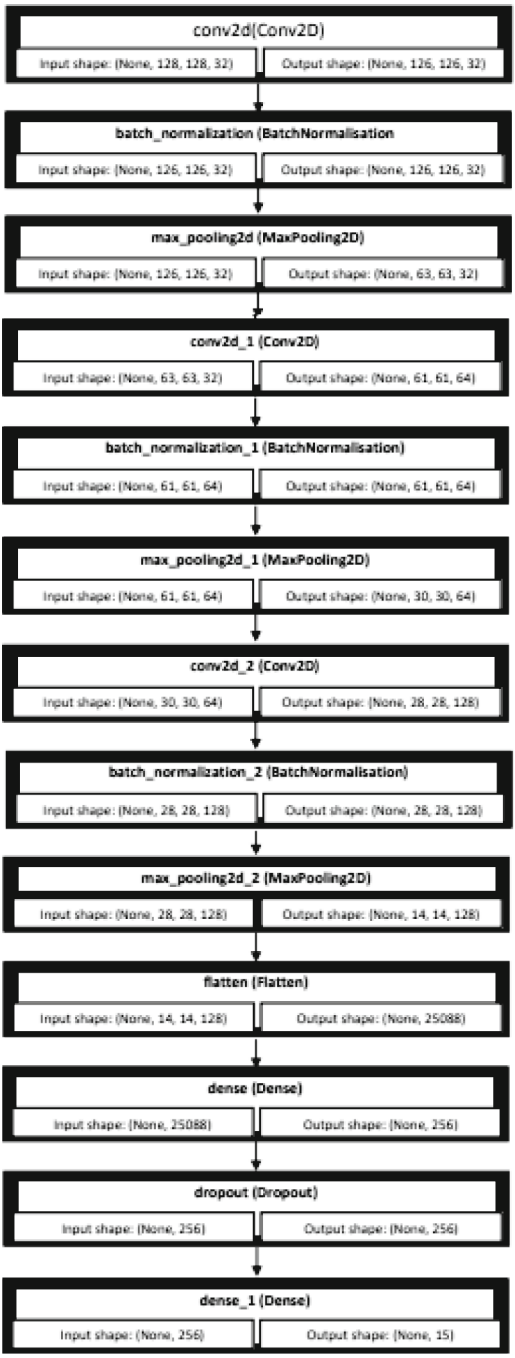


Fig. 1. DCNN Architecture Layer

section started. The education method turned into performed on Google Colab with GPU acceleration enabled (the usage of a Tesla T4 or P100 GPU), which extensively decreased schooling time and stepped forward version efficiency.

3.4.2 Training Setup

During schooling, real-time information augmentation become carried out to the schooling set to growth range, whilst validation information become stored untouched to provide a fair evaluation. Early preventing and checkpoint callbacks had been used to prevent overfitting and keep the nice version.we used 32 as batch size and between 30 to 50 epochs and categorical cross entropy for loss fucntion for training setup.

3.4.3 Training Output

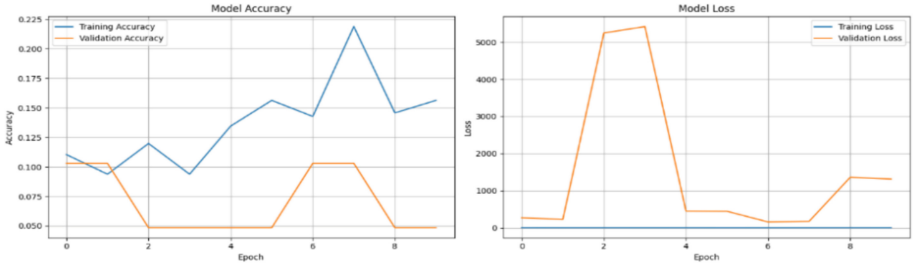
The version confirmed a steadily reducing education loss and growing validation accuracy. It achieved over 98% validation accuracy within 40–50 epochs.

$$ModelAccuracy = \left(\frac{1}{N}\right) \times \sum 1\left(\hat{y}_i = y_i\right) \tag{1}$$

where N represenst the count of total samples and $\hat{y}_1$ defines the predicted class for sample i, y_i shows the true class label and $1(\cdot)$ represents the indicator function for prediction value either 0 or 1

$$ModelLoss = -\sum(i = 1\,to\,N)\sum(c = 1\,to\,C)y_{\{i,c\}} * \log\left(\hat{y}_{\{i,c\}}\right) \tag{2}$$

where N represents the number of samples, c as number of classes, with $y_(\{i, c\}$ as actual label 1, if class c is correct for sample i, else 0 and $\hat{Y}_{\{i,c\}}$: predicted probability of class c for sample i



(a) **(b)**

Fig. 2. a. Model Accuracy Vs Epochs. b. Model Loss Vs Epochs

Figure 2a graph displays accuracy, where the blue line training represents the training accuracy and the orange line represents verification accuracy in the ages. Training accuracy gradually increases, while verification accuracy refers to ups and downs, and is comparatively low, indicating possible low or dataset imbalance in early ages. Figure 2b

Graph displays loss, where the blue line represents training loss and the orange line represents verification loss. Verification loss shows high spikes in the early ages, gradually decrease and stabilizing over time, while training loss is continuously low. This deviation suggests that the noise in the model verification data may be overfitting or affecting.

3.5 Model Evaluation

After training, the model's overall performance became evaluated using more than a few metrics that are essential for multiclass category troubles.To evaluate the effectiveness of the proposed DCNN-based model, standard classification matrix such as accuracy, accurate, recall and F1 scores were calculated using verification dataset. These metrics provide a wide understanding of the performance of the model in many disease classes. The trained model demonstrated strong generalization capabilities, which saw minimum overfitting during training and verification. The model was evaluated using standard viewing measurements as accuracy, accurate, recall, F1 score and confusion matrix. On the verification data set, DCNN achieved classification accuracy of 98.2%, with high precision and values in almost all 38 squares. The training on the training app shows little loss and frequent accuracy successful learning and minimal overfit. In order to measure the effectiveness of the DCNN model, a comprehensive test on the confirmation data set was performed from the Plantvillage dataset, consisting of 38 unique leaf disease categories in many plant species. The performance of the model was assessed using industry standard classification measurements: accuracy, accurate, recall, F1 score and confusion matrix.

Accuracy measures the ratio of correctly predicted observations to the total observations

$$Accuracy = \left(\frac{1}{N}\right) \times \sum 1\left(\hat{y}_i = y_i\right) \tag{3}$$

Precision measures the ratio of true positive predictions to the total predicted positives:

$$Precision = \left(\frac{1}{C}\right) * \Sigma\left(\frac{TP_c}{(TP_c + FP_c)}\right) \tag{4}$$

Recall measures the ratio of true positive predictions to the actual positives:

$$Recall = \left(\frac{1}{C}\right) * \Sigma\left(\frac{TP_c}{(TP_c + FN_c)}\right) \tag{5}$$

F1 Score provides the harmonic mean of Precision and Recall:

$$F1 = \left(\frac{1}{C}\right) * \Sigma\left[2 * Precision_c * \frac{Recall_c}{(Precision_c + Recall_c)}\right] \tag{6}$$

where N represents he total samples, c as number of classes

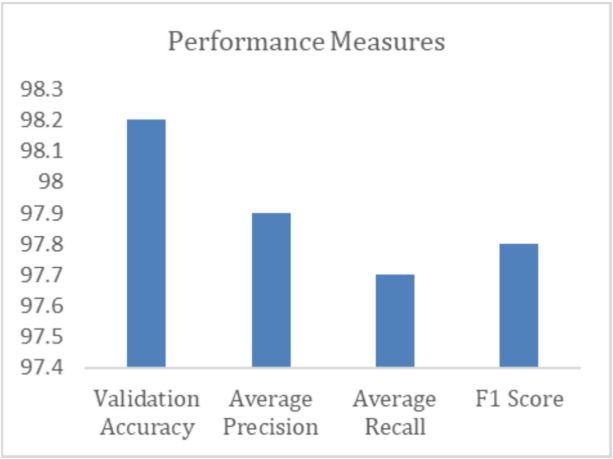


Fig. 3. Metric Score of our proposed work

After training the model on pre-processed plantvillage dataset (80% training and 20% verification division) the following metric was seen:

Figure 3 represents the model's metric score after training the dataset. This bar chart plant shows the major assessment matrix received from the DCNN model trained for the plant leaf disease classification. The model obtained a verification accuracy of about 98.2%, indicating a high overall purity in classifying overlooked data. Average precision (~ 97.9%) reflects the ability of model to reduce false positive, while average recall (~ 97.7%) measures its effectiveness in identifying true positivity. The F1 score (~ 97.8%), accurate and a harmonic meaning of remembrance, confirms a balanced performance in all classes. These metric collectively demonstrate the strength, accuracy and suitability of the model for real-time agricultural disease diagnosis. Close diagonals indicate the right class prophecies; Minimum abortion observed from these results confirm the model's ability to detect leaf disorders with limited training conditions and high reliability during the image malformations. These results emphasize the strength of the model, especially in handling variance in the intra-class (different appearances in samples) and equality between the class (different diseases with overlapping visual symptoms).The Confusion matrix shows strong diagonal dominance, showing true positive classifications in high numbers, a Minimum off-diagonal values suggest rare spontaneous abortion, even for diseases with visually subtle differences. The model also maintained stable performance in low-representing classes, indicating effective learning from unbalanced data. The loss function showed steady convergence with minimal oscillations. Training and validation curves were nearly aligned, indicating negligible overfitting. Early stopping and dropout layers contributed to regularization and generalization performance. To bridge the gap between machine learning and real-world usability, a Gradio-based web interface was built and deployed. This lightweight frontend allows users to interact with the trained model effortlessly. Functionalities such as Uploading any leaf image (captured from phone or camera), predicting the plant type and the specific disease with the confidence level in percentage and providing the informational output such as Symptoms of the

disease, Precautionary measures, Suggested treatment or medicine and Farming tips related to that disease are said to be the Key functionalities.Low-latency output (<1 s in line with photo) even on modest hardware, Offline compatibility with model export assist and can also be deployed on cell, laptop, or cloud-primarily based systems. To assess the model's robustness, we conducted several tests involving real-world image imperfections, the model has been tested with Blurred Images, Low-light conditions, partially covered or torn leaves and Different background colors. Despite being trained on clean, centered snap shots, the version turned into able to generalize nicely to noisy and distorted inputs. In maximum instances, it furnished correct or near-correct predictions with excessive confidence, demonstrating strong characteristic gaining knowledge of and real-global viability. A sample of 50 such altered images showed the model retained an accuracy of over 92%, even with noise introduced.To make the system accessible and practical for end users such as farmers, researchers, and agricultural workers, the trained DCNN model was integrated with a Gradio web-based user interface. Features include Upload Functionality where Users can upload images of leaves directly from any device and Instant Prediction where the app returns Predicted crop and disease name with the Confidence percentage including the Information about symptoms, precautions, and treatment. It is Lightweight and fast (average response time <1 s) that doesn't need to install any software and Works on mobile and desktop browsers.

The proposed system uses a comprehensive set of modern technologies to facilitate the development, training, evaluation and deployment of an intensive learning-based plant disease classification model. Core architecture is applied using TensorFlow and causes, which enables the design of a custom deep convolutional neural network (DCNN) for multi-class leaf disease classification. Models training and experiments are carried out on Google Collaboratory (COLAB) with GPU acceleration, which significantly reduces computational time and increases model performance. OpenCV is employed for image such as image Pre-processing operations such as size, generalization and format conversion. Data handling and visualization are managed using NumPy, Pandas, and Matplotlib, while model assessment is calculated using metrics-like, recall and SCIKIT-LEARN. In addition, to bridge the gap between model infection and end-user accessibility, a Gradio-based web interface is integrated, allowing the disease names, confidence scores and suggested measures to allow real-time prediction from the leaf images uploaded with measures. This technical stack ensures that the system is not only accurate and strong, but also accessible to deployment in real -world agricultural landscapes.

4 Result and Discussion

The proposed deep learning system for leaf disease detection and classification using a custom Deep Convolutional Neural Network (DCNN) demonstrated strong performance both in terms of accuracy and practical usability. This section outlines the key results observed during testing, evaluates the model's predictions on unseen data, and discusses the user-facing interface designed for real-time diagnosis. The trained custom-made Deep Convolutional Neural Network (DCNN) model from the ground up using the PlantVillage dataset, which has more than 50,000 labeled leaf images in 38 different

classes. The model did a great job of classifying things thanks to optimized architecture, dropout regularization, and data augmentation. When we tested the model with the validation set (20% split), it got a validation accuracy of about 98.2%. Its precision, recall, and F1-score were all over 97% on average for most disease categories [13, 14]. The model could process each image in less than a second, which made it very responsive for use in real time. An analysis of the confusion matrix showed that most of the predictions were correct, especially for disease categories that were clearly different from each other. This high level of performance shows that the model is strong and can find small differences between leaf disease patterns. To ensure practical practices, the model was deployed through a Gradio-based web interface, which allows users to upload leaf images and obtain immediate predictions with symptoms, precautions and treatment tips.

Figure 4(a) demonstrates a pepper bell leaf that is identified with bacterial spot with 99.99% confidence, 4(b) depicts a pepper bell leaf that is healthy with 99.99% confidence, 4(c) depicts a potato leaf that is identified with early blight with 100% confidence, 4(d) depicts a potato leaf that is identified with late blight with 99.90% confidence, 4(e) depicts a potato leaf that is healthy with 99.97% confidence and 4(f) depicts a potato leaf that is identified with early blight with 99.55% confidence.The proposed system is built upon a custom-designed Deep Convolutional Neural Network (DCNN), Depending on heavy pre-educated architecture such as ResNet or VGG, [15] a specially designed deep Convolutional Neural Network (DCNN) is used in the suggested structure that is mild and sewn to detect real-time disease on light and mobile and edge devices. While the Convolutional layers acquire hierarchical representations such as edges, textures and disease-specific patterns, DCNN backbone effectively removes spatial features from leaf images. In order to reduce overfitting and computational expenditure while maintaining important information, feature map dimensions are reduced by pooling layers, such as Maxpooling or average pooling or average pooling. Relu, supported by activation tasks such as softmax to convert the output into class possibilities and relu, serve as a well-specific classifier, fully connected layers that integrate the features extracted to produce accurate predictions. To ensure effective convergence and increased accuracy, the model is adapted using a category cross-entropy as a loss function. The training is directed by optimizer such as Adam or SGD. Rotation, flips and zoom are examples of data enlargement techniques used to improve normalization and allow models to handle variability in samples. [16] Standard evaluation is used to measure the performance of the system such as accuracy, accurate, recall and F1-score, guarantees its strength and dependence. These elements work together to provide highly accurate and efficient automatic identification and classification of leaf diseases, which makes the method practical and expandable for use in agricultural settings.The results indicate that the proposed DCNN-based model performs firmly in a wide spectrum of plants of plants, with minimal abortion for visually similar symptoms. Unlike many existing systems, which depend on the pre-trained generic models, the project specifically used a custom architecture trained by scratching on plant disease data, ensuring better feature learning and high accuracy. Additionally, by integrating the model into an intuitive Gradio interface, the system only exceeds an academic prototype - it turns into a practical clinical equipment accessible to farmers, agricultural workers and researchers. The ability of the system to provide actionable

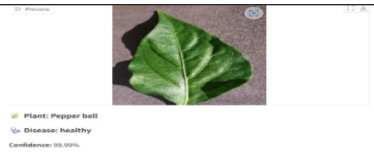
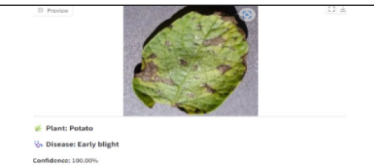
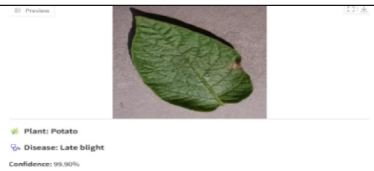
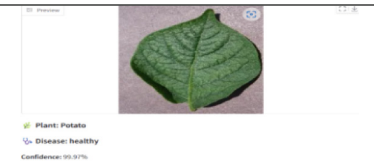
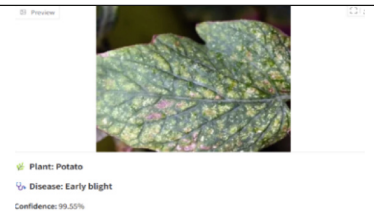
S.No	*Input*	*Output*
a		
b		
c		
d		
e		
f		

Fig. 4. Sample Input and Output Images of Leaf's (a)pepper bell leaf, (b)pepper bell leaf, (c)potato leaf, (d)potato leaf, (e)potato leaf, (f)potato leaf showing the diseases and confidence.

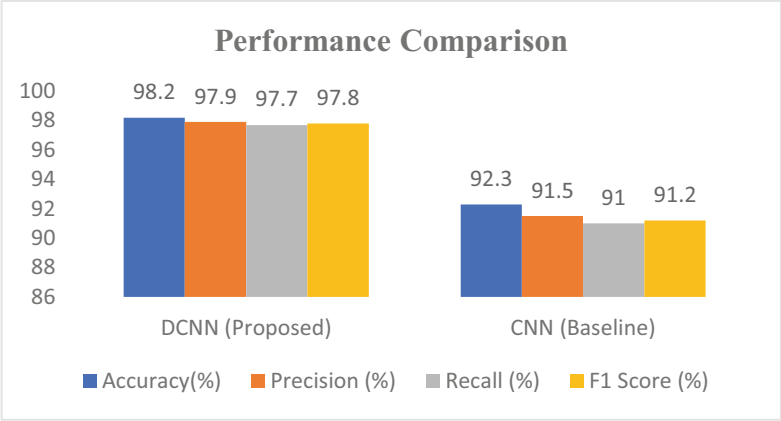


Fig. 5. Performance Comparison

insights such as symptoms, treatments and precautions not only increases its value in real -world farming landscapes. Although trained in controlled conditions, model displays readiness for deployment, with the ability to extend to real-sector image in future versions and multilingual support.

Figure 5 represents the performance comparison of the suggested DCNN model and the baseline CNN for leaf disease identification is shown in the chart. The DCNN model performed better with 98.2% accuracy, 97.9% precision, 97.7% recall, and 97.8% F1 score than the CNN model, which had 92.3% accuracy, 91.5% precision, 91.0% recall, and 91.2% F1 score. This demonstrates unequivocally the DCNN model's outstanding effectiveness in managing the multiple classes of plant leaf diseases.

5 Conclusion

The proposed model in plant leaf disease detection is designed on a customized deep neural network (DCNN) developed from scratch to give faster, accurate and accessible disease classification with more than 98% verification accuracy. This strengthens real-time diagnosis directly from leaf images presented by the user, allowing immediate response to the types of disease, symptoms, precautions and suggested treatments. This specialist reduces dependence on intervention and delays crop safety, while encouraging the response on time and reduces yield loss. This system integrates advanced AI into traditional agricultural workflows, making it highly practical without the need for technical knowledge for farmers, researchers and agricultural workers. Through its radio-based interface, the tool exceeds a research model-it turns into an everyday accessory for smart farming. This work not only bridges the gap between modern technology and agriculture, but also introduces a new layer of confidence and autonomy for crop care. With future expansion in real-field dataset and multilingual support, it promises to redefine the plant accessible plant healthcare in agriculture.

References

1. Abd Algani, Y.M., Caro, O.J.M., Bravo, L.M.R., Kaur, C., Al Ansari, M.S., Bala, B.K.: Leaf disease identification and classification using optimized deep learning. Measur. Sens. **25**, 100643 (2023)
2. Tiwari, V., Joshi, R.C., Dutta, M.K.: Dense convolutional neural networks based multiclass plant disease detection and classification using leaf images. Eco. Inform. **63**, 101289 (2021)
3. Jasim, M.A., Al-Tuwaijari, J.M.: Plant leaf diseases detection and classification using image processing and deep learning techniques. In: 2020 International Conference on Computer Science and Software Engineering (CSASE), pp. 259–265. IEEE (2020)
4. Ashwinkumar, S., Rajagopal, S., Manimaran, V., Jegajothi, B.J.M.T.P.: Automated plant leaf disease detection and classification using optimal MobileNet based convolutional neural networks. Mater. Today Proc. **51**, 480–487 (2022)
5. Vasavi, P., Punitha, A., Rao, T.V.N.: Crop leaf disease detection and classification using machine learning and deep learning algorithms by visual symptoms: a review. Int. J. Electr. Comput. Eng. **1**(2), 2079 (2022)
6. Sladojevic, S., Arsenovic, M., Anderla, A., Culibrk, D., Stefanovic, D.: Deep neural networks based recognition of plant diseases by leaf image classification. Comput. Intell. Neurosci. **1**, 1–11 (2016)
7. Atila, Ü., Uçar, M., Akyol, K., Uçar, E.: Plant leaf disease classification using EfficientNet deep learning model. Eco. Inform. **61**, 101182 (2021)
8. Naik, B.N., Malmathanraj, R., Palanisamy, P.: Detection and classification of chilli leaf disease using a squeeze-and-excitation-based CNN model. Ecol. Inform. **69**, 101663 (2022)
9. Savla, D., Dhaka, V.S., Rani, G., Oza, M.: Apple leaf disease detection and classification using CNN models. In: International Conference on Computing in Engineering & Technology, pp. 277–290 (2022)
10. Daniya, T., Vigneshwari, S.: Rice plant leaf disease detection and classification using optimization enabled deep learning. J. Environ. Inf. **42**, 25–28 (2023)
11. Akila, M., Deepan, P.: Detection and classification of plant leaf diseases by using deep learning algorithm. Int. J. Eng. Res. Technol. **6**, 1–5 (2018)
12. Moupojou, E., et al.: FieldPlant: a dataset of field plant images for plant disease detection and classification with deep learning. IEEE Access **11**, 35398–35410 (2023)
13. Annabel, L.S.P., Annapoorani, T., Deepalakshmi, P.: Machine learning for plant leaf disease detection and classification–a review. In: 2019 International Conference on Communication and Signal Processing (ICCSP), pp. 0538–0542. IEEE (2019)
14. Yadhav, S.Y., Senthilkumar, T., Jayanthy, S., Kovilpillai, J.J.A.: Plant disease detection and classification using CNN model with optimized activation function. In: 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), pp. 564–569. IEEE (2020)
15. Quan, S., Wang, J., Jia, Z., Xu, Q., Yang, M.: Real-time field disease identification based on a lightweight model. Comput. Electron. Agric. **226**, 109467 (2024)
16. Lee, S., Lee, S.: Efficient data augmentation methods for crop disease recognition in sustainable environmental systems. Big Data Cogn. Comput. **9** (2025)



Geospatial Analysis of Surface Urban Heat Island Intensity Using Google Earth Engine

M. Mahmood Pasha[1(✉)], Sharmila Banu[1], and Zameer Gulzar[2]

[1] SR University Warangal, Warangal, India
mmpasha2024@gmail.com, sharmila.banu@sru.edu.in
[2] Malla Reddy University, Hyderabad, Telangana, India
zamir045@gmail.com

Abstract. Urban Heat Islands (UHI) are a threat due to India's rapidly urbanizing cities. The present study uses Google Earth Engine-processed Landsat-9 OLI/TIRS data to evaluate UHI patterns in Warangal, Telangana, for May 2025. While the Normalized Difference Vegetation Index (NDVI) and Normalized Difference Built-up Index (NDBI) were generated to investigate correlations between land cover and thermal behaviour, the Land Surface Temperature (LST) was obtained using a single-channel technique with emissivity correction. Strong spatial disparities were evident in the data, with vegetated and water-covered zones recording the lowest LST values (~29 °C) and densely populated and industrial areas recording the highest LST (~45 °C). The study shows that there is significant heat stress in the city centre, and almost 40% of the city falls under the high or extreme UHI category. The correlation analysis revealed a positive relationship between LST and NDBI and a negative relationship between LST and NDVI. The findings from the current study highlight the need for climate-sensitive development and green infrastructure to minimize urban heat hazards in Warangal.

Keywords: Google Earth Engine · Land Surface Temperature · NDBI · NDVI · Urban Heat Island · Warangal

1 Introduction

Climate Change is one of the major issues of the 21st century which has significant effects on the human societies across the globe. Risks to both humans and ecological systems are increasing as a result of changing climatic patterns, rising global temperatures, and increasing frequency and intensities of extreme weather events [1]. Fast and uncontrolled development intensifies temperature extremes and risks human well-being. Understanding urban heat dynamics is becoming crucial as a result of the combined effects of urbanisation and climate change [2].

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026
S. Lanka et al. (Eds.): ICTSM 2025, LNNS 1755, pp. 446–455, 2026.
https://doi.org/10.1007/978-3-032-13177-5_35

Urban areas in particular are becoming hotspots of vulnerability due tochanging climate. Understanding urban heat dynamics is becoming crucial as a result of the combined effects of urbanization and climate change [3]. Urban Heat Island (UHI) effect is one of the most important environmental effects of urbanization [4]. It is characterized by higher temperatures in urbanized areas comparative to nearby rural zones primarily because of changes in the land surface, less vegetation, and anthropogenic heat emissions [5]. The recent study has shown that UHI's is not limited to a particular place, it is a global phenomenon which actually is driven by uncontrolled urban growth and land cover changes, with a sturdy implication for public health and climatic resilience [6, 7].While large metropolitan cities leads UHI research, medium-sized urban centers are increasingly suseptable yet understudied [8].

Megacities in India like Delhi, Mumbai, and Hyderabad have been the key focus of UHI research in India [9]. However, medium-sized cities continue to receive little attention despite their susceptibility to climate extremes. Urbanization rates in India are rising, and medium-sized cities like Warangal are seeing quick changes in their land cover [10, 11]. Warangal, located on Telangana's semi-arid Deccan plateau, is prime example of a fast-urbanizing area where temperature discomfort is getting worse. Quantifying urban areas that increase human susceptibility is urgently needed due to increased exposure to heat extremes, especially during the pre-monsoon months when surface temperatures surpass 40 °C [12, 13]. In order to close this gap, this study uses satellite-derived Land Surface Temperature (LST) and land cover statistics to map UHI hotspots in Warangal for 2025. The methodological framework uses percentile-based hotspot identification, remote sensing, and geospatial analysis.

2 Study Area

Warangal is situated on the semi-arid Deccan plateau in northeastern Telangana between 17.48° to 18.27° N latitude and 79.15° to 80.02° E longitude and an average elevation of 270 m amsl as shown in Fig. 1. Warangal experiences a tropical wet–dry climate with hot summers (March–June) with daily highs frequently exceeding 40 °C and a monsoon season that provides the majority of the yearly rainfall (~900–1000 mm). The terrain is moderately sloping, with small bodies of water and rocky outcrops scattered throughout. Warangal has become a major urban city due to rapid urbanization; during the 1990s, open spaces and agricultural land have been replaced by built-up areas. By 2025, the urban agglomeration is expected to have more over 1.1 million residents, up from about 830,000 in the 2011 Census.

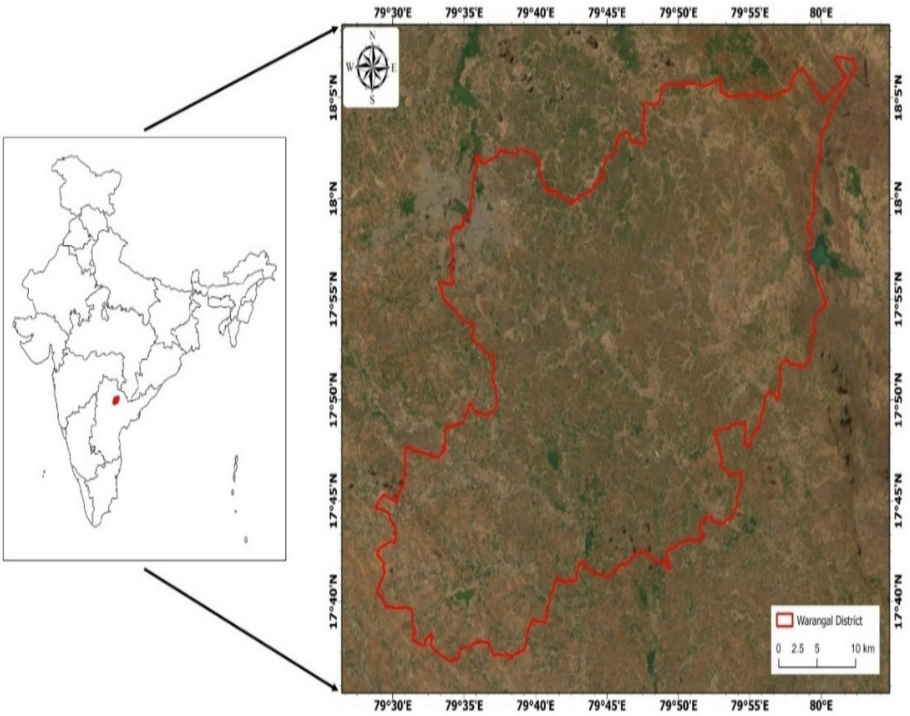


Fig. 1. Map depicting the extent of Warangal District

3 Data and Methods

3.1 Datasets Used

The study used datasets that are publicly available on the Google Earth Engine (GEE) platform. In order the determine the LST, Normalised Difference Vegetation Index (NDVI) and Normalised Difference Built-up Index (NDBI), Landsat 9 OLI/TIRS Collection 2, Level-2 images selected for one clear-sky day in May 2025 was used. Additionally, SRTM Dem was used for terrain correction (Table 1).

Table 1. Datasets used in the study.

Dataset	Sensor/Source	Temporal Coverage	Purpose
Landsat 9 OLI/TIRS	NASA/USGS	May 2025	LST, NDVI, NDBI
SRTM DEM	NASA	2000 (static)	Terrain correction
Warangal boundary	Government Data	N/A	AOI extraction

3.2 Methodology

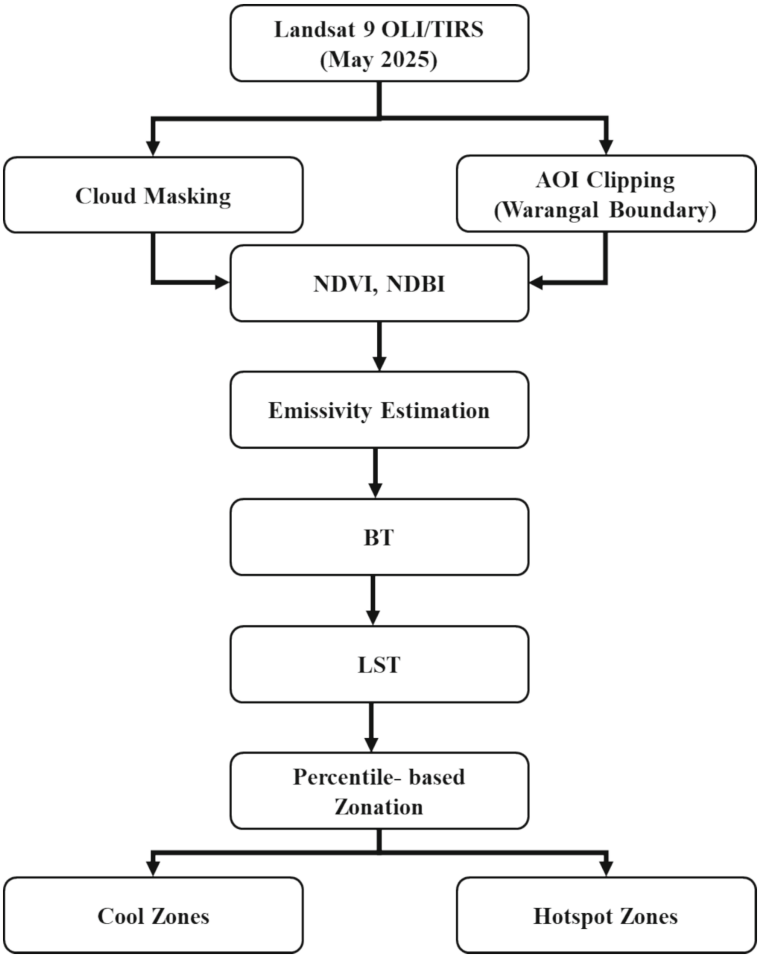


Fig. 2. Flowchart of the workflow of methodology

The study uses GEE platform as it provides free cloud-based processing capability with a huge repository of data [14]. The procedure based on [15] involved preprocessing of the images, calculating the index, retrieving the land surface temperature (LST), and classifying the UHI zones according to percentile thresholds. The workflow (depicted in Fig. 2) is described as follows:

3.2.1 Preprocessing

- A single clear-sky Landsat 9 OLI/TIRS Level-2 image for May 2025 was selected.
- Cloud, cirrus, and shadow pixels were masked using the *QA_PIXEL* band.
- The image was clipped to the Warangal municipal boundary to constrain analysis.

3.2.2 Vegetation and Built-up Indices

Two indices were computed to capture vegetation density and impervious surface distribution:

- **Normalised Difference Vegetation Index (NDVI):**

xxxxxx

$$NDVI = \frac{NIR - RED}{NIR + RED}$$

Higher NDVI values indicate vegetation cover, associated with cooler surfaces.

- **Normalised Difference Built-up Index (NDBI):**

xxxxxx

$$NDBI = \frac{SWIR - NIR}{SWIR + NIR}$$

Higher NDBI values represent impervious or built-up areas, which are usually warmer.

3.2.3 Land Surface Temperature Retrieval

The thermal band (TIRS Band 10) was used to compute LST. A single-channel method with emissivity correction was applied [16].

- **Top of Atmosphere (TOA) Radiance:**

xxxxxx

$$L_\lambda = M_L \times Q_{cal} + A_L$$

where L_λ is spectral radiance, M_L and A_L are rescaling coefficients from metadata, and Q_{cal} is the quantised calibrated pixel value.

- **Brightness Temperature (BT):**

xxxxxx

$$BT = \frac{k_2}{\ln\left(\frac{k_1}{L_\lambda} + 1\right)}$$

where K_1 and K_2 are calibration constants from metadata.

- **Proportion of Vegetation (Pv):**

$$P_v = \left(\frac{NDVI - NDVI_{min}}{NDVI_{max} - NDVI_{min}} \right)^2$$

- **Land Surface Emissivity (ε):**

  xxxxxx

$$\varepsilon = 0.986 + 0.004 \times P_v$$

- **Land Surface Temperature (LST) expressed in °C:**

  xxxxxx

$$LST = \frac{BT}{1 + \left(\frac{\lambda \times BT}{\rho} \right) \ln(\varepsilon)}$$

where λ is the effective wavelength of band 10 (10.9 μm) and $\rho = hc/\sigma = 1.438 \times 10^{-2}$ mK.

3.2.4 UHI Zonation and Its Relationship with Built-up and Vegetation

To delineate hotspots, the LST distribution was classified using percentile thresholds [17].

- Cool zones: LST ≤ 25th percentile
- Neutral zones: 25th–75th percentile
- Hotspot zones: 75th–90th percentile
- Extreme hotspot zones: ≥ 90th percentile

4 Results and Discussions

4.1 Relationship of NDVI and NDBI with LST

The retrieved Land Surface Temperature (LST) values over Warangal on the selected May 2025 date ranged between 29 °C and 43 °C, consistent with reported pre-monsoon temperature ranges in semi-arid urban centres of India. The spatial analysis of NDVI and NDBI showed clear heat contrasts throughout Warangal as shown in Fig. 3.

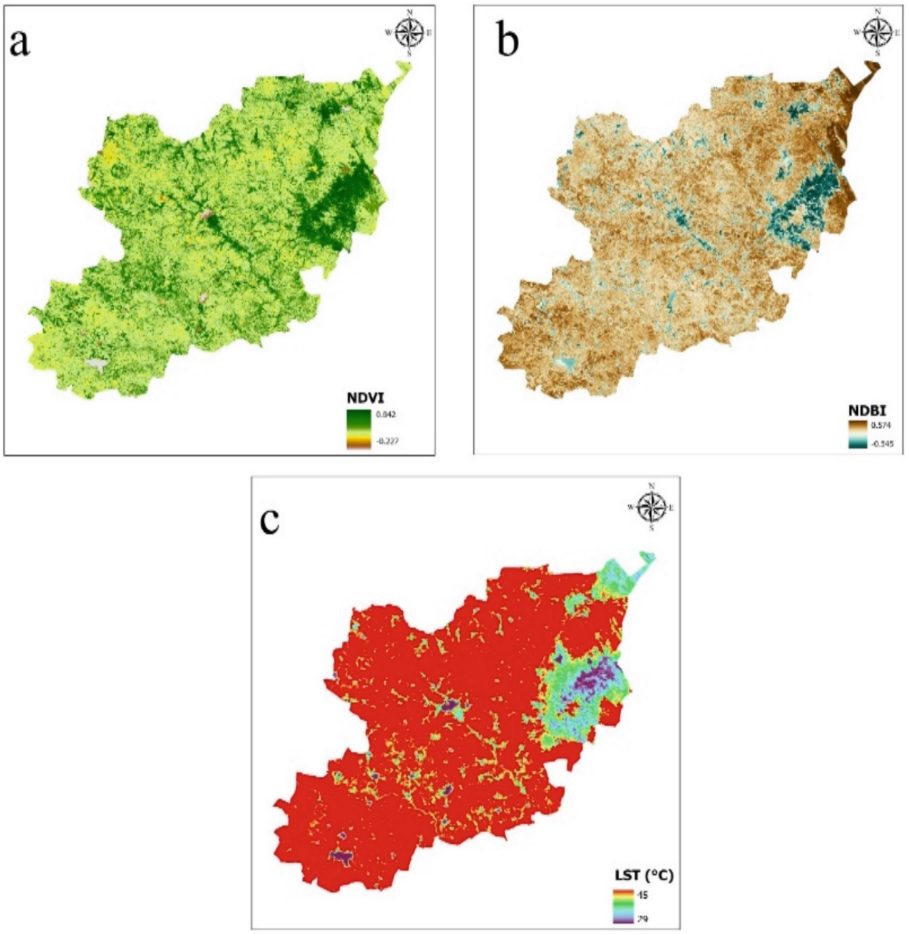


Fig. 3. Maps representing (a) NDVI, (b) NDBI and (c) LST of Warangal district derived from Landsat 9 imagery.

While built-up zones with high NDBI (>0.3) reported significantly higher temperatures (~40–45 °C), vegetated regions demonstrated consistently lower LST (~30–32 °C) with higher NDVI values (>0.4). Correlation analysis validated this pattern, revealing that LST had a high negative correlation (R =–0.57) with NDVI and a strong positive correlation (R =+ 0.68) with NDBI. According to [9], these values align with earlier research conducted in Indian cities like Bhopal, which revealed comparable correlations. The NDVI and NDBI exhibit opposite behaviour, which emphasises how important vegetation is in regulating surface temperatures and how impervious surfaces contribute to heat buildup. The spatial relationship between UHI zones and land cover was assessed by overlaying NDVI and NDBI maps. Built-up areas (high NDBI) coincided with hotspot and extreme-hotspot zones, while vegetated areas (high NDVI) corresponded to cool zones, confirming the physical linkage between land cover and thermal behaviour.

4.2 Urban Heat Island Zonation

The classification of LST into percentile-based thresholds provided clear delineation of UHI intensity zones (shown in Fig. 4). To delineate hotspots, the LST distribution was classified using percentile thresholds as shown in Table 2.

Table 2. Classification of UHI zones based on LST percentiles.

UHI Zone	Percentile Range	LST Range (°C)	Dominant Land Cover
Cool zone	≤25th	28–32	Vegetated areas, agricultural fields, and water bodies
Neutral zone	25th–75th	32–37	Transitional areas with mixed peri-urban settlements
Hotspot zone	75th–90th	37–40	Built-up areas and transport corridors
Extreme hotspot zone	≥90th	>40	Urban core, industrial, and commercial hubs

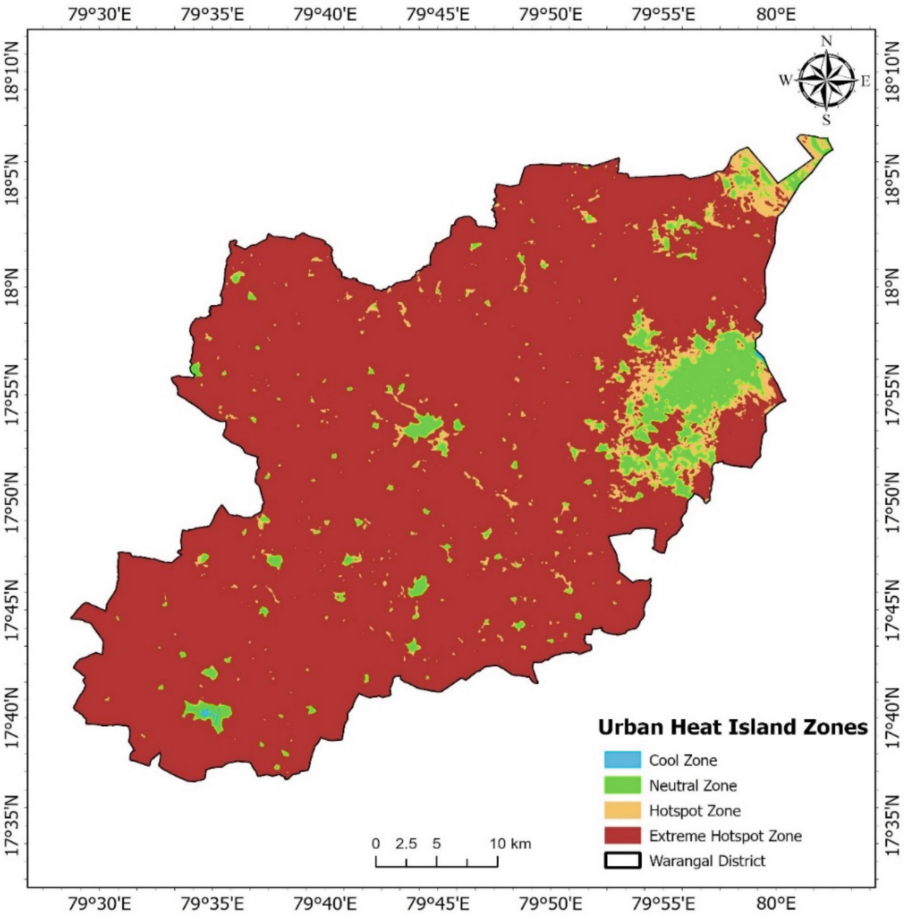


Fig. 4. Urban Heat Island Zones in Warangal District

The spatial relationship between UHI zones and land cover was assessed by overlaying NDVI and NDBI maps. Around 22% of the municipal area was classified as a Hotspot or Extreme Hotspot, with important transportation routes, industrial clusters, and the urban core experiencing temperatures above 40 °C. Peri-urban communities were characterised by neutral zones (32–37 °C), whilst cool zones (<32 °C) were centred on parks, water bodies, and rural areas. The city hub functions as the most extreme heat island in the layout, which shows a concentric gradient. Centralised high-density development has been found to be the primary driver of UHI creation in Hyderabad and other fast urbanising cities, which is consistent with this finding. In Warangal, the observed 5–6 °C difference between vegetated and built-up zones emphasises the urgency of urban greening measures and the severity of thermal stress during pre-monsoon conditions.

5 Conclusion

The urban landscape of Warangal in 2025 clearly shows a surface UHI pattern. Outlying vegetated zones were significantly cooler (by several degrees) than dense downtown and industrial districts. Under hot May circumstances, more than one-third of the city experienced High or Extreme UHI levels. In line with the results of prior Indian city studies, the research verified that regions with greater NDBI (more impermeable cover) and lower NDVI (less vegetation) correlated with higher LST. These findings suggest that local heat stress would worsen as urbanisation continues without green space. Increasing greenery and implementing cooling urban designs should be planners' top priorities in order to reduce the urban heat island effect. With the use of Google Earth Engine, this study shows how useful satellite-based thermal mapping is for guiding sustainable urban planning in Warangal and other towns of a similar nature. We will extend the study for future work to study about land use land cover and temperature shifts across Warangal district using deep learning techniques.

References

1. Handmer, J., et al.: Changes in impacts of climate extremes: human systems and ecosystems. Managing the risks of extreme events and disasters to advance climate change adaptation special report of the intergovernmental panel on climate change, pp. 231–290 (2012).https://doi.org/10.1017/CBO9781139177245.007
2. Jain, M.: Two decades of nighttime surface urban heat island intensity analysis over nine major populated cities of India and implications for heat stress. Front. Sustain. Cities **5**, 1084573 (2023). https://doi.org/10.3389/frsc.2023.1084573
3. Mir, R.A., Mantoo, A.G., Sofi, Z.A., Bhat, D.A., Bashir, A., Bashir, S.: Types of environmental pollution and its effects on the environment and society. In: Geospatial Analytics for Environmental Pollution Modeling: Analysis, Control and Management, pp. 1–31. Springer Nature Switzerland, Cham (2023). https://doi.org/10.1007/978-3-031-45300-7_1
4. Zhou, X., Chen, H.: Impact of urbanization-related land use land cover changes and urban morphology changes on the urban heat island phenomenon. Sci. Total. Environ. **635**, 1467–1476 (2018). https://doi.org/10.1016/j.scitotenv.2018.04.091
5. Peng, J., et al.: Spatial-temporal change of land surface temperature across 285 cities in China: an urban-rural contrast perspective. Sci. Total. Environ. **635**, 487–497 (2018). https://doi.org/10.1016/j.scitotenv.2018.04.105

6. Tesfaldet, Y.T., Ndeh, N.T., Budnard, J., Treeson, P.: Assessing face mask littering in urban environments and policy implications. Case Bangkok, Sci. Total Environ. **806**(4), 150952 (2022). https://doi.org/10.1016/j.scitotenv.2021.150952
7. Ali, B., Negar, J., Khaled, Z.B., Farzad, F., Ali, A.: Spatio-temporal variations, ozone and secondary organic aerosol formation potential, and health risk assessment of BTEX compounds in east of Azerbaijan Province, Iran, Urban Climate, vol. 47, p. 101360 (2023). https://doi.org/10.1016/j.uclim.2022.101360
8. Kamel, M., Mohamed, A.Z., Seham, H., Hala, E.: Impact of anthropogenic activities on urban heat islands in major cities of El-Minya Governorate, Egypt. Egypt. J. Remote Sens. Space Sci. **25**(2), 609–620 (2022). https://doi.org/10.1016/j.ejrs.2022.03.014
9. Singh, P., Sarkar Chaudhuri, A., Verma, P., Singh, V.K., Meena, S.R.: Earth observation data sets in monitoring of urbanization and urban heat island of Delhi, India. Geomat. Nat. Hazards Risk **13**(1), 1762–1779 (2022). https://doi.org/10.1080/19475705.2022.2097452
10. Ravindra, K., et al.: Temperature projections and heatwave attribution scenarios over India: a systematic review. Heliyon, **10**(4) (2024)
11. Rath, S.S., Mohanty, S., Panda, J.: Analyzing the fragmentation of urban footprints in eastern and southern Indian cities and driving factors. J. Indian Soc. Remote Sens. **50**(8), 1499–1517 (2022). https://doi.org/10.1007/s12524-022-01546-3
12. Kjellstrom, T., McMichael, A.J.: Climate change threats to population health and well-being: the imperative of protective solutions that will last. Global Health Action **6**(1), 20816 (2013). https://doi.org/10.3402/gha.v6i0.20816
13. Cobbinah, P.B.: Urban resilience in climate change hotspot. Land use policy. **100**, 104948 (2021). https://doi.org/10.1016/j.landusepol.2020.104948
14. Ghosh, S., Kumar, D., Kumari, R.: Cloud-based large-scale data retrieval, mapping, and analysis for land monitoring applications with google earth engine (GEE). Environ. Challenges **9**, 100605 (2022). https://doi.org/10.1016/j.envc.2022.100605
15. Sameh, S., Zarzoura, F., El-Mewafi, M.: Automated mapping of urban heat Island to predict land surface temperature and land use/cover change using machine learning algorithms: Mansoura city. Int. J. Geoinform. **18**(6), 47–67 (2022)
16. Yu, X., Guo, X., Wu, Z.: Land surface temperature retrieval from Landsat 8 TIRS—Comparison between radiative transfer equation-based method, split window algorithm and single channel method. Remote Sens. **6**(10), 9829–9852 (2014). https://doi.org/10.3390/rs6109829
17. Guerri, G., Crisci, A., Messeri, A., Congedo, L., Munafò, M., Morabito, M.: Thermal summer diurnal hot-spot analysis: the role of local urban features layers. Remote Sens. **13**(3), 538 (2021). https://doi.org/10.3390/rs13030538

Prediction of Insurance Claim Risk Using an Optimized Decision Tree Model

Oleg Pursky(✉), Tetyana Filimonova, Darina Bondarenko, Tetyana Tomashevska, Pavlo Demidov, and Ihor Tyschenko

State University of Trade and Economics, Kioto str. 19, Kyiv 02156, Ukraine
pursky_o@ukr.net, {t.filimonova,D.Bondarenko_FIT_7_21_B_d, t.tomashevska,p.demidov,i.tyshchenko}@knute.edu.ua

Abstract. The task of predicting the occurrence of insurance claims is important for risk management in the insurance industry. This paper considers the optimization of a decision tree model in order to improve prediction. The research applies four tuning methods: GridSearchCV, RandomizedSearchCV, HalvingGridSearchCV, and Optuna, using a publicly available dataset of insurance records. The dataset underwent preprocessing, including feature scaling and data splitting, followed by model training and evaluation. The performance of the optimized Decision Tree model was analyzed and compared based on accuracy and execution time. The results showed that HalvingGridSearchCV achieved the highest accuracy 0.983, though at the cost of increased computational time 3.90 s. RandomizedSearchCV was the fastest method, 0.20 s, while maintaining a high accuracy of 0.97, making it suitable for large datasets. GridSearchCV provided comparable accuracy 0.975 but required more time than RandomizedSearchCV. Optuna, which implements a heuristic hyperparameter optimization method based on the Tree-structured Parzen Estimator, achieved test accuracy 0.976 with an execution time 1.15 s, making it a viable alternative for complex hyperparameter spaces. The study highlights the trade-offs between accuracy and computational efficiency in hyperparameter optimization. The findings confirm that HalvingGridSearchCV is preferable when accuracy is the top priority, whereas RandomizedSearchCV is ideal for time-sensitive tasks. Optuna and GridSearchCV offer a balance between precision and efficiency. For comparison, a neural network based on the PyTorch framework was implemented and achieved an accuracy of 0.95. Slightly lagging behind the decision tree model in terms of accuracy, the PyTorch-based neural network is well suited for solving the tasks of predicting insurance claims. It should be noted that there are limitations regarding input data and interpretability.

Keywords: Machine learning · Neural network · Decision tree classifier · Insurance claim risk · Optimization methods

1 Introduction

The modern insurance services market is characterized by high competition and constantly increasing service quality requirements. The financial stability of an insurance company depends on its ability to accurately predict insurance claims, which helps

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026
S. Lanka et al. (Eds.): ICTSM 2025, LNNS 1755, pp. 456–469, 2026.
https://doi.org/10.1007/978-3-032-13177-5_36

reduce compensation costs, optimize the underwriting process, and establish competitive insurance premiums. In today's environment, where data volumes are constantly growing, traditional statistical methods often fail to detect all hidden patterns. Therefore, the implementation of innovative machine learning methods has become highly relevant [1, 2].

In recent years, machine learning methods have been widely used in the field of insurance. The task of predicting the occurrence of an insurance claim belongs to classification tasks. One of the algorithms applied to solve this task is the decision tree, which is distinguished by its simplicity of implementation and interpretation. This method clearly illustrates the logic of decision-making, therefore it is useful for the field of insurance [2, 3].

It should be noted that the optimization of hyperparameters of the basic model plays an important role in improving the quality of the model. Finding the optimal parameters, such as maximum tree depth (max_depth), minimum number of samples required to split a node (min_samples_split), and minimum number of samples in terminal nodes (min_samples_leaf), makes it possible to achieve maximum model performance and ensure high accuracy on the test dataset. In this paper, modern optimization methods such as GridSearchCV, RandomizedSearchCV, HalvingGridSearchCV, and Optuna are applied, which provide a balance between prediction accuracy and computational efficiency [4, 5].

Data analysis and preprocessing are an important part of the task. Techniques such as standardization (StandardScaler) for numerical feature normalization and encoding of categorical variables are necessary to ensure the correct functioning of machine learning algorithms. Preprocessing helps eliminate discrepancies in feature scales, which positively affects model convergence speed and prediction quality. In the insurance sector, where data is often heterogeneous and contains missing values, efficient data preprocessing serves as the foundation for successful modeling [2].

With the help of machine learning methods, it is possible not only to develop predictive models but also to detect potential risks and fraudulent activities. This approach allows insurance companies to respond promptly to changing market conditions, adjust insurance premiums, and optimize their portfolio of insurance policies, ultimately reducing financial losses and improving overall business stability [4, 5].

The main contribution of this work lies in the systematic evaluation of four methods for hyperparameter optimization of a decision tree model and the comparative analysis of the resulting optimized model with alternative approaches and a PyTorch-based neural network, which substantiates the choice of an interpretable model for classification tasks in the insurance domain.

The decision tree model ensures transparency of the classification process, correctly handles mixed feature types, and is characterized by moderate computational complexity. Comparison with alternative approaches confirms the soundness of its selection for the given task.

2 Related Work

The application of machine learning methods to the task of predicting the occurrence of insurance claims makes it possible to increase accuracy and effectively assess risks in insurance. Ensemble models, namely decision trees, gradient boosting, and others, demonstrate good results in this field.

The article [6] examines the use of decision trees as one of the machine learning methods for modeling insurance claims in the field of automobile insurance. Decision trees enhance predictive accuracy by effectively identifying complex relationships in the data that traditional actuarial models may not account for. They contribute to improved pricing accuracy and risk assessment. It is also noted in this article that when implementing these methods in the insurance industry it is necessary to pay close attention to the interpretation of models and to take into account ethical aspects.

The application of gradient boosting for prediction tasks is discussed in paper [7]. This article shows that GBM and XGBoost models work effectively with overdispersed insurance claim data using the Poisson and negative binomial distributions. These models achieve high predictive accuracy on the test dataset, which makes them suitable for estimating insurance claim probabilities in practical tasks.

The paper [8] primarily focuses on predicting motor insurance claim frequency using the XGBoost algorithm rather than decision trees. While decision trees are part of the broader machine learning techniques, the study emphasizes XGBoost's advantages over traditional models like Poisson and Negative Binomial due to its ability to handle complex data. The research highlights XGBoost's strong predictive performance, achieving an RMSE of 0.949 and an MAE of 0.7741, indicating its effectiveness in insurance claim frequency prediction.

The paper [9] explores predicting insurance claims using various data mining techniques, including decision trees. It addresses challenges such as high missing values and class imbalance. After employing a double grid search for missing value imputation and data rebalancing, classifiers based on decision trees were evaluated using metrics like AUC, Cohen's kappa, Accuracy, and F-measure. The study finds that mean and mode imputation is preferable in scenarios with over 0.40% missing values, enhancing the prediction of claim probabilities.

The research paper [10] presents a comparison of various classification models, namely a neural network, support vector machine (SVM), Tree Augmented Naïve Bayes (NB), decision tree (DT), random forest (RF), and AdaBoost. The results show that the Tree Augmented Naïve Bayes model outperformed the other models across several performance indicators: accuracy 79.35%, sensitivity 44.70%, misclassification rate 20.65%, area under the curve 0.81, and Gini coefficient 0.62. These results are important for preventing fraud in the insurance sector.

In study [11], the decision tree algorithm was applied to build models for predicting insurance claims in the medical field. It was noted that the obtained metrics lagged behind the results of the XGBoost and random forest models, which achieved R^2 values of 0.79% and 0.77%, respectively. The effectiveness of the decision tree was considered in the context of an analysis that included feature importance assessment, namely smoking habits, body mass index, and blood pressure.

The paper [12] proposes a hybrid tree-based method for predicting insurance claim probability by utilizing a two-step algorithm. The first step involves constructing a classification tree to model claim frequency, effectively addressing the challenge of a large proportion of zero claims. This approach enhances prediction accuracy compared to traditional methods like the Tweedie GLM. The hybrid structure improves model interpretability and hyperparameter tuning, which ensures a more accurate prediction of insurance claims.

In study [13], the effectiveness of a regression model based on a decision tree for predicting the probability of insurance claims was emphasized. This model outperformed such models as KNN and XGB in terms of metrics.

The study [14] compared logistic regression with other machine learning models, in particular the decision tree and the support vector machine, in the task of predicting health insurance claims. Logistic regression demonstrated an accuracy of 82%, while the decision tree model achieved the highest accuracy of 96%. In addition, the balance between accuracy and interpretability of the models was examined.

Unlike approaches focused on ensemble methods, this study conducts a comparative analysis of several modern strategies for hyperparameter optimization of a decision tree model, investigates the influence of key hyperparameters on performance metrics, and compares the results with alternative models and a PyTorch-based neural network.

3 Proposed Work

The study utilized the Sample Insurance Claim Prediction Dataset [15]. The dataset contains information about patients' demographic characteristics, habits, region of residence, and insurance expenses. The variable age is a numerical indicator that reflects the patients' age. Sex is represented as a categorical variable that indicates the patient's gender. BMI (Body Mass Index) is a numerical index that characterizes the ratio of body weight to height and serves as an indicator of possible overweight or obesity. Children is an integer variable that represents the number of dependents a patient has. Smoker is a binary variable that indicates the smoking status (smoker/non-smoker). Region is represented as a categorical variable that divides the sample into four regions of residence: northeast, northwest, southeast, and southwest. Charges is a numerical variable that reflects the total amount of medical insurance expenses incurred by the patient. Insuranceclaim is a binary variable that indicates whether an insurance claim has been filed. The presented data allow for the analysis of the impact of these factors on the level of insurance expenses and the likelihood of filing an insurance claim, which has practical significance for predicting financial risks in the field of health insurance.

At the stage of data analysis and preprocessing, variable types and the total number of records were determined, duplicates were removed, and the presence of missing data was identified. Additionally, the correlation with the target variable was calculated. The results are presented in Fig. 1.

The analysis of the correlation matrix revealed several key dependencies between the studied parameters. A strong positive correlation was observed between smoking (smoker) and medical expenses (charges). There is a negative correlation between the number of children (children) and insurance claims (insuranceclaim). As a result, the decision was made not to remove these features from the dataset.

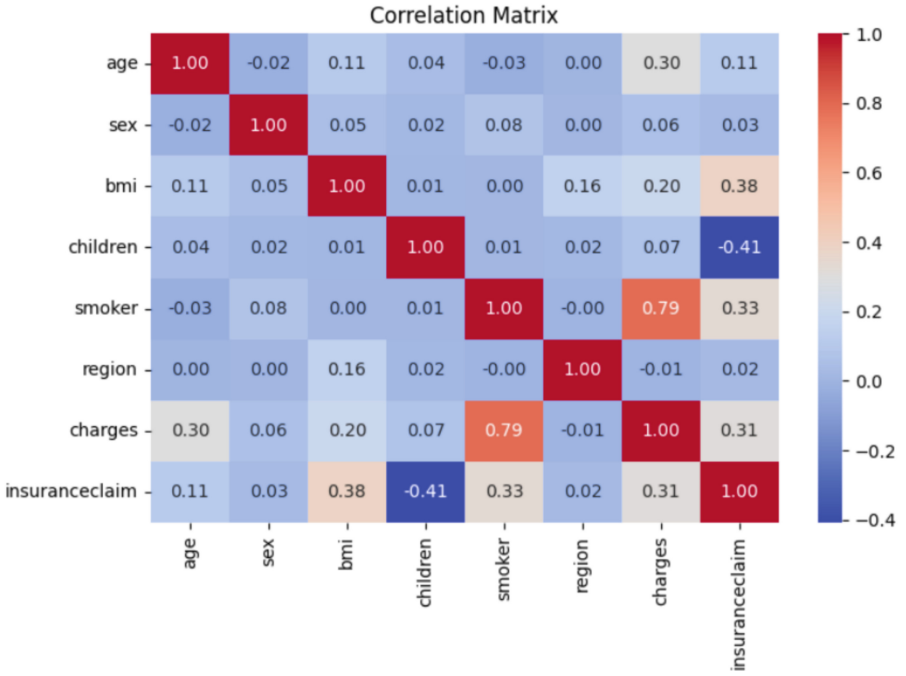


Fig. 1. Correlation matrix.

Also, during data processing, the interquartile range (IQR) method was used to identify and exclude extreme values. Standard scaling was applied for feature scaling. The dataset was split into training and test data in an 80/20 ratio, with the reproducibility of results ensured. The main steps in model development include:

1. Training the Base Model. The decision tree model is initialized with a predefined maximum depth (e.g., max_depth = 5) to control overfitting. Training is conducted on preprocessed data.
2. Quality evaluation. The model's performance is assessed using key metrics such as accuracy, precision, recall. Additionally, a ROC curve is plotted, and the Area Under the Curve (AUC) is calculated to evaluate the model's ability to distinguish between positive and negative claims [3].
3. Comparison with other algorithms. To justify the effectiveness of the selected approach, additional models such as Random Forest, Logistic Regression, SVM, and XGBoost were developed. A comparative analysis of their performance allows an assessment of the decision tree's competitiveness in terms of both accuracy and interpretability.

To improve model performance, key hyperparameters that influence decision tree efficiency were optimized. max_depth – maximum tree depth; min_samples_split – minimum number of samples required to split a node; min_samples_leaf – minimum number of samples in terminal nodes (leaves). To determine optimal values, the following hyperparameter search methods were applied:

1. GridSearchCV. Forms a grid of combinations of the specified hyperparameters and selects the best configuration using cross-validation. It is advisable to include default parameter values to avoid missing a potentially optimal solution [16].
2. RandomizedSearchCV. Randomly selects parameter combinations, significantly reducing optimization time while still achieving high accuracy [17].
3. HalvingGridSearchCV. Gradually narrows the search space by eliminating less promising parameter combinations, allowing faster discovery of optimal values with high accuracy [18].
4. Optuna. Implements a heuristic hyperparameter optimization method based on the Tree-structured Parzen Estimator (TPE). It supports dynamic narrowing of the search space, adaptive early stopping of ineffective configurations, and flexible control over the number of iterations [19].

4 Result Analysis

To improve the performance of the decision tree model, four hyperparameter optimization methods were applied: GridSearchCV, RandomizedSearchCV, HalvingGridSearchCV, and Optuna. The primary objective was to find the optimal values for max_depth (maximum depth of the tree), min_samples_split (minimum number of samples required to split a node), min_samples_leaf (minimum number of samples required in a leaf node), which ensure the highest accuracy on cross-validation and optimal performance on the test set. Using the GridSearchCV method, the following optimal parameter combination was selected (Table 1).

Table 1. Parameters for GridSearchCV.

Parameter	Values for Tuning	Optimal Value
max_depth	[3, 5, 10, None]	None
min_samples_split	[2, 5, 10]	2
min_samples_leaf	[1, 2, 4]	1

Using the GridSearchCV method, the following optimal parameter combination was selected (Table 2).

Table 2. Parameters for RandomizedSearchCV.

Parameter	Values for Tuning	Optimal Value
max_depth	[5, 10, 15, 20, None]	20
min_samples_split	[2, 5, 10]	2
min_samples_leaf	[1, 2, 4, 6]	1

The following parameters were selected in HalvingGridSearchCV: max_depth – the maximum depth of the decision tree. This parameter determines how many levels the tree has; min_samples_split – the minimum number of samples required to split a node in the tree. This helps control how much data is needed to divide a node; min_samples_leaf – the minimum number of samples that must be in each leaf of the tree (terminal node) (Table 3).

Table 3. Parameters for HalvingGridSearchCV.

Parameter	Values for Tuning	Optimal Value
max_depth	[5, 10, 15, 20, None]	10
min_samples_split	[2, 5, 10]	5
min_samples_leaf	[1, 2, 4, 6]	2

To select the optimal hyperparameters for the model, the Optuna framework was used. Three main hyperparameters of the decision tree were chosen for optimization: max_depth – the maximum depth of the tree, which limits the number of levels in the tree; min_samples_split – the minimum number of samples required to split a node into two subnodes; min_samples_leaf – the minimum number of samples that must be in each leaf of the tree (terminal node).

The optimization was performed using the heuristic Tree-structured Parzen Estimator (TPE) approach, which belongs to the class of Bayesian optimization methods. The parameters were selected from predefined ranges, and the effectiveness of each combination was evaluated based on cross-validation results. The selection of optimal parameters using the Optuna framework is shown in the Table 4.

Table 4. Optimization function for Optuna.

Parameter	Range of Values for Tuning	Optimal Value
max_depth	2 to 25	13
min_samples_split	2 to 10	2
min_samples_leaf	1 to 6	1

Table 5 presents a comparison of parameters for decision tree optimization methods. The accuracy and execution time have been evaluated

From Table 5, it can be concluded that HalvingGridSearchCV achieved the highest accuracy of 0.983, while the other methods – GridSearchCV, RandomizedSearchCV, and Optuna – yielded slightly lower results, ranging from 0.975 to 0.976. In terms of execution time, RandomizedSearchCV was the fastest with 0.20 s, while HalvingGridSearchCV required the most time, 3.90 s. The visualization of the results is presented in Fig. 2.

Table 5. Optimization function for Optuna.

Optimization Methods	Test Accuracy	Execution Time (sec.)
GridSearchCV	0.975	0.74
RandomizedSearchCV	0.975	0.20
HalvingGridSearchCV	0.983	3.90
Optuna	0.976	1.15

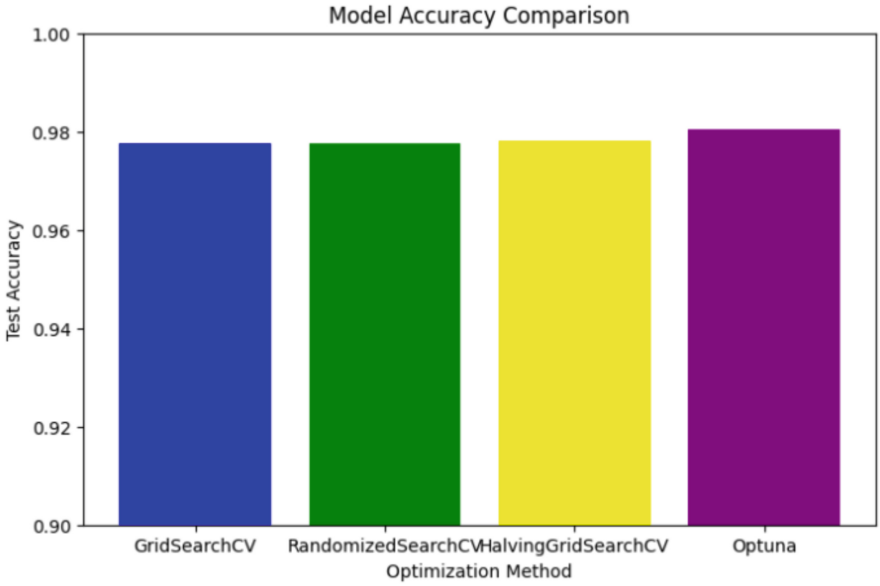


Fig. 2. Accuracy comparison of parameters optimization methods.

The obtained results prove that the application of hyperparameter optimization has a positive effect on the quality of the decision tree model.

1. Although all optimization methods achieve high accuracy (0.97–0.98), they differ in execution time. HalvingGridSearchCV provides the highest accuracy but is the slowest due to its stepwise narrowing process. RandomizedSearchCV, on the other hand, is the fastest, making it ideal for large datasets where computational efficiency is crucial.
2. GridSearchCV and Optuna offer a balanced trade-off between accuracy and execution time. GridSearchCV is efficient for small datasets, though it becomes computationally expensive as the parameter space increases. Optuna, leveraging Bayesian optimization, can quickly find optimal values, making it a promising tool for large-scale tasks.
3. Practical Implications. In the insurance industry, both accuracy and speed are crucial. The choice of an optimization method should depend on the specific requirements:

- For small datasets or when maximum accuracy is the priority, HalvingGridSearchCV is the best choice;
- For large datasets, where speed is critical, RandomizedSearchCV is the most efficient;
- Optuna is an optimal choice when expanding the hyperparameter space, as it provides an excellent accuracy/time ratio and quickly adapts to new conditions.

The optimization results confirm that choosing an effective hyperparameter search method is essential for achieving high model performance. Selecting the appropriate method should be based on a compromise between prediction accuracy and execution time, which is especially relevant for practical applications in the insurance sector.

Additionally, several machine learning models were built to predict the target variable based on the training dataset. The goal was to compare the effectiveness of different models and select the best one based on accuracy. The following models were used: Random Forest, Logistic Regression, Support Vector Machine, and Gradient Boosting. After training, predictions were made on the test dataset. A classification report was generated for each model. It should be noted that the optimal parameters for the decision tree were determined using HalvingGridSearchCV. The results of the model comparison are presented in the Fig. 3.

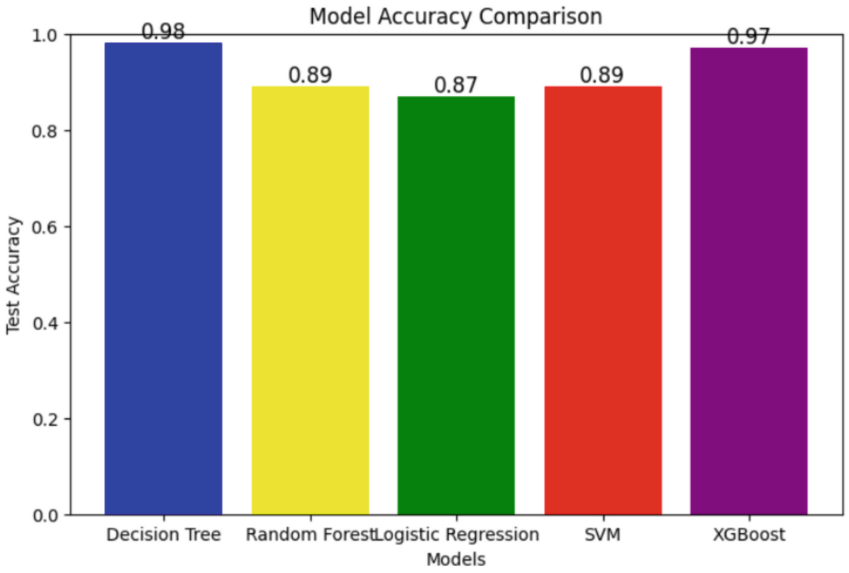


Fig. 3. Model accuracy comparison.

Fig. 3 shows that the decision tree model achieved the highest accuracy on the test set (0.98). XGBoost also demonstrated high effectiveness with an accuracy of 0.97. The Random Forest and Support Vector Machine (SVM) models showed the same accuracy (0.89), while Logistic Regression achieved an accuracy of 0.87. The obtained results indicate that for this classification task, the decision tree model is the most promising.

An insurance risk assessment was conducted based on individual client characteristics using a decision tree model with hyperparameters pre-optimized via the HalvingGridSearchCV method (Figs. 4 and 5).

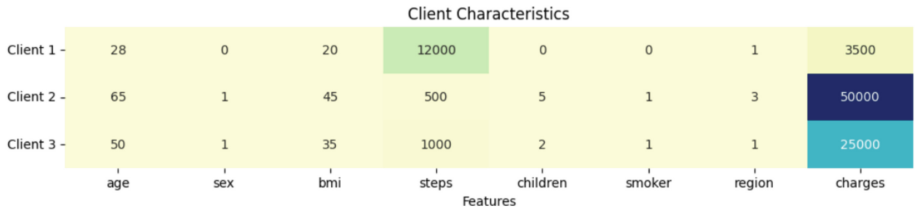


Fig. 4. Client characteristics for insurance risk assessment.

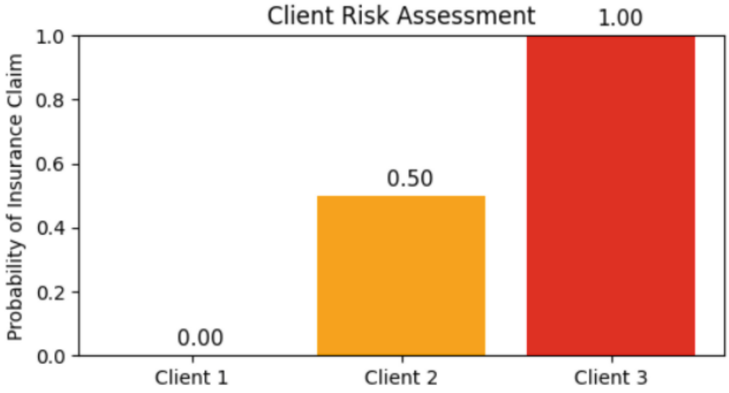


Fig. 5. Predicted probabilities of insurance claims for the clients.

The obtained results demonstrate the model's ability to effectively distinguish between profiles with low, medium, and high levels of insurance risk. The most influential factors in shaping the prediction were age, smoking status, body mass index (BMI), and level of physical activity.

Additionally, a neural network was built using the PyTorch framework. The input features were scaled and converted into tensors for compatibility with PyTorch. The dataset was split into training and test subsets in an 80/20 ratio.

The architecture of the constructed neural network has a deep sequential structure and consists of four hidden fully connected layers. Each of these layers is supplemented with batch normalization, a ReLU activation function, and a Dropout regularization layer. The output of the network is a prediction vector whose number of components corresponds to the number of target classes. The structure of the model is illustrated in detail in Fig. 6.

During the training of the neural network, the Adam optimizer and the BCEWithLogitsLoss loss function were applied.

```
==========================================================================================
Layer (type:depth-idx)                   Output Shape              Param #
==========================================================================================
InsuranceNN                              [1, 2]                    --
├─Linear: 1-1                            [1, 256]                  2,816
├─BatchNorm1d: 1-2                       [1, 256]                  512
├─ReLU: 1-3                              [1, 256]                  --
├─Dropout: 1-4                           [1, 256]                  --
├─Linear: 1-5                            [1, 128]                  32,896
├─BatchNorm1d: 1-6                       [1, 128]                  256
├─ReLU: 1-7                              [1, 128]                  --
├─Dropout: 1-8                           [1, 128]                  --
├─Linear: 1-9                            [1, 64]                   8,256
├─BatchNorm1d: 1-10                      [1, 64]                   128
├─ReLU: 1-11                             [1, 64]                   --
├─Dropout: 1-12                          [1, 64]                   --
├─Linear: 1-13                           [1, 32]                   2,080
├─BatchNorm1d: 1-14                      [1, 32]                   64
├─ReLU: 1-15                             [1, 32]                   --
├─Linear: 1-16                           [1, 2]                    66
==========================================================================================
Total params: 47,074
Trainable params: 47,074
Non-trainable params: 0
Total mult-adds (Units.MEGABYTES): 0.05
```

Fig. 6. Implementation of a Multi-Layer Neural Network.

To improve the efficiency of the training process, mechanisms of early stopping and adaptive learning rate reduction were employed. As a result, the training was completed at the 188th epoch.

The learning curves demonstrate that the loss function decreases, while the classification accuracy increases for both training and test data throughout the training process (Figs. 7, 8).

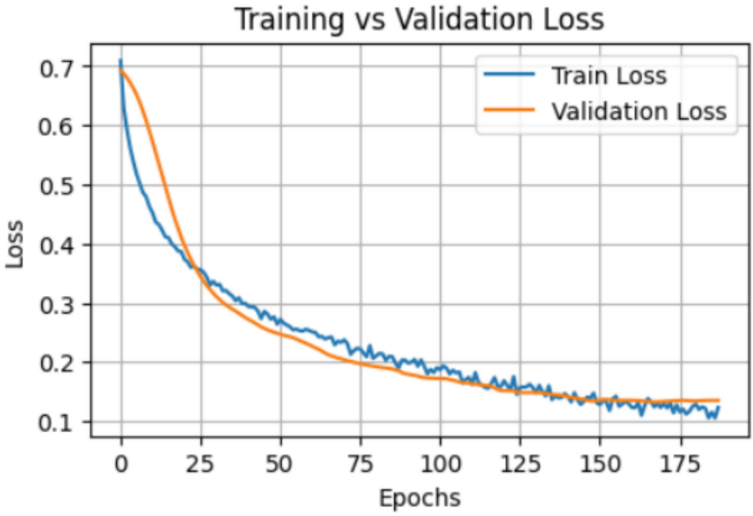


Fig. 7. Model loss on training and validation sets.

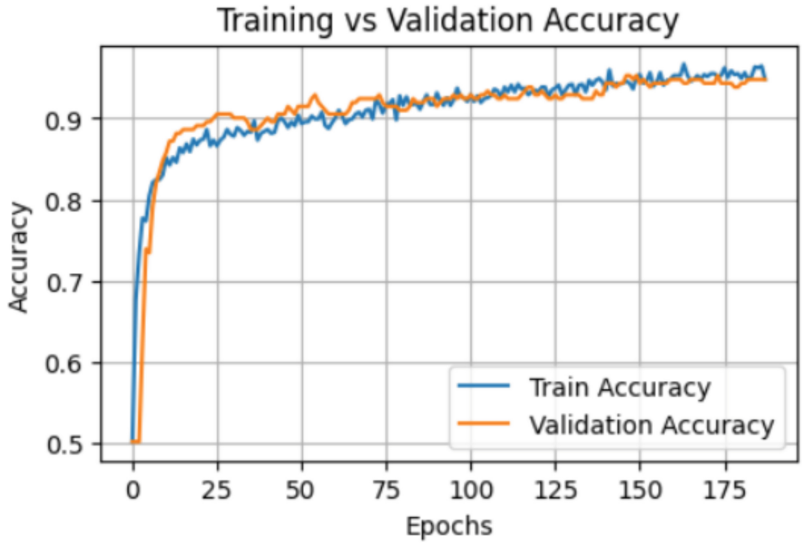


Fig. 8. Model accuracy on training and validation data.

As a result of training, the model achieved a test accuracy of 0.95 and a loss value of 0.1 on the test dataset. Although the results are satisfactory, the model slightly underperforms compared to the decision tree in terms of the evaluated performance metrics.

5 Conclusion

In this work, the classification task was solved using a decision tree for predicting insurance claims and risk management. Data analysis and preparation were carried out, including duplicate removal and feature scaling. The result of the work is the construction of a decision tree model, in which four optimization methods were used to select the optimal hyperparameters: RandomizedSearchCV, HalvingGridSearchCV, GridSearchCV, and Optuna. RandomizedSearchCV turned out to be the fastest method, performing the search in 0.20 s. HalvingGridSearchCV achieved the highest accuracy of 0.98, but it required more execution time – 3.90 s. GridSearchCV demonstrated stable results with a small parameter space, taking 0.74 s to execute, providing a good balance between efficiency and accuracy. Optuna demonstrated a favorable balance between accuracy and execution time, achieving a test accuracy of 0.976 with an execution time of 1.15 s.

The decision tree model with optimal parameters, determined using HalvingGridSearchCV, achieved the highest accuracy compared to other models, namely XGBoost, Support Vector Machine (SVM), and Logistic Regression.

For comparison, the neural network built using the PyTorch framework achieved an accuracy of 0.95. Despite slightly lower accuracy compared to the decision tree, the neural network implemented using PyTorch has significant potential, especially in cases involving complex nonlinear relationships. However, its main drawbacks include higher data requirements and limited interpretability.

The study confirmed the practical significance of the optimized decision tree model. The obtained results are useful for choosing a model optimization strategy with regard to the available time in real-world processes.

Future research should focus on building the optimal architecture and selecting the optimal parameters of the neural network using the Optuna framework to achieve the best results in insurance claim prediction tasks.

References

1. Alomair, G.: Predictive performance of count regression models versus machine learning techniques: a comparative analysis using an automobile insurance claims frequency dataset. PLoS One. (2024). https://doi.org/10.1371/journal.pone.0314975
2. Sahai, R., et al.: Insurance risk prediction using machine learning. In: Wah, Y.B., Berry, M.W., Mohamed, A., Al-Jumeily, D. (eds.) Data Science and Emerging Technologies. DaSET 2022. Lecture Notes on Data Engineering and Communications Technologies, vol 165. Springer, Singapore (2023). https://doi.org/10.1007/978-981-99-0741-0_30
3. Poufinas, T., Gogas, P., Papadimitriou, T., Zaganidis, E.: Machine learning in forecasting motor insurance claims. Risks. **11**(9), 164 (2023). https://doi.org/10.3390/risks11090164
4. Quan, Z., Valdez, E.A.: Predictive analytics of insurance claims using multivariate decision trees. Depend. Model. **6**(1), 377–407 (2018). https://doi.org/10.1515/demo-2018-0022
5. Rukhsar, L., Bangyal, W.H., Nisar, K., Nisar, S.: Prediction of insurance fraud detection using machine learning algorithms. Mehran Univ. Res. J. Eng. Technol. **41**(1), 33–40 (2022). https://doi.org/10.3316/informit.263147785515876
6. Marciuc, M.A.: Predictive modeling for claims in automobile insurance. Rev. Econ. Stud. Res. Virgil Madgearu. **17**(2), 79–99 (2024). https://doi.org/10.24193/RVM.2024.17.118
7. Chevalier-Lucia, D., & Côté, M. (2024). From point to probabilistic gradient boosting for claim frequency and severity prediction. ArXiv Preprint. https://doi.org/10.48550/arxiv.2412.14916
8. Kollongei, N., Onyango, F.: Motor insurance claim frequency prediction using XGBoost. Asian J. Probab. Stat. **26**(10), 155–170 (2024). https://doi.org/10.9734/ajpas/2024/v26i10665
9. Morales, P., Tallón-Ballesteros, A.J.: Predicting insurance claims through a variety of data mining techniques: facing lots of missing values and moderate class-imbalanced levels. Logic J. IGPL. (2024). https://doi.org/10.1093/jigpal/jzae061
10. Nordin, S.-Z.S., Wah, Y.B., Haur, N.K., Hashim, A., Rambeli, N., Jalil, N.A.: Predicting automobile insurance fraud using classical and machine learning models. Int. J. Electr. Comput. Eng. (2024). https://doi.org/10.11591/ijece.v14i1.pp911-921
11. Zheng, H., Fan, P., Tian, Y., Zhang, Z.Y., Zhang, W.: Insurance fraud detection based on XGBoost. Acad. J. Comput. Inf. Sci. **6**(8) (2023). https://doi.org/10.25236/ajcis.2023.060808
12. Quan, Z., Wang, Z., Gan, G., Valdez, E.A.: On hybrid tree-based methods for short-term insurance claims. Math. Finan. **37**, 597–620 (2023). https://doi.org/10.1017/S0269964823000074
13. Islam, N.N.P.: Advanced machine learning techniques for predictive analysis of health insurance. Proc. IEEE Conf. Comput. Intell., 1166–1170 (2024). https://doi.org/10.1109/icoici62503.2024.10695988
14. Thakre, V.P., Poul, R.D., Sawarkar, A.D.: Predictive precision: unraveling health insurance claim patterns with logistic regression and decision trees. Cureus J. Comput. Sci. **2**, es44389-025-03010-y (2025). https://doi.org/10.7759/s44389-025-03010-y
15. Kaggle.: Sample insurance claim prediction dataset (2024). Retrieved from https://www.kaggle.com/datasets/easonlai/sample-insurance-claim-prediction-dataset/data

16. Scikit-learn.: GridSearchCV. (2025). Retrieved from https://scikit-learn.org/dev/modules/generated/sklearn.model_selection.GridSearchCV.html
17. Scikit-learn.: RandomizedSearchCV. (2025). Retrieved from https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.RandomizedSearchCV.html
18. Scikit-learn.: HalvingGridSearchCV. (2025). Retrieved from https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.HalvingGridSearchCV.html
19. Optuna.: Optuna: a hyperparameter optimization framework (2025). Retrieved from https://optuna.org/

TabNet for Intrusion Detection: Bridging Accuracy and Interpretability in Tabular Network Data

Joideep Banerjee[1(✉)] and Asma Patel[2]

[1] University of Staffordshire, College Road, Stoke-on-Trent ST4 2DE, UK
joideep.banerjee@staffs.ac.uk
[2] Aston University, Aston Street, Birmingham B4 7ET, UK
Asma.Patel@staffs.ac.uk

Abstract. Intrusion Detection Systems (IDS) must balance accuracy with interpretability, yet existing approaches often sacrifice one for the other. Classical machine learning methods such as Logistic Regression and Random Forest provide solid accuracy but limited transparency, while deep learning models like CNNs act as black boxes. This paper evaluates TabNet, a deep neural architecture designed for tabular data, as a candidate for IDS. TabNet leverages sequential attention and sparse feature selection, enabling both high performance and feature-level interpretability. We test TabNet on UNSW-NB15, BoT-IoT, and KDD CUP and compare it with Logistic Regression, Random Forest, SVM, and Naïve Bayes. Results show that TabNet achieves near-perfect detection on BoT-IoT (99.98%) and KDD (99.98%), while remaining highly competitive on UNSW-NB15 (99.30%). Its attention masks highlight meaningful features such as flow duration and packet rate, providing actionable insights for analysts. TabNet thus offers a practical trade-off between accuracy and explainability, making it well-suited for next-generation IDS.

Keywords: IDS · TabNet · SVM · Naive Bayes

1 Introduction

In recent years, the scale and sophistication of cyber-attacks have grown at an unprecedented rate. As critical infrastructures, enterprises, and personal devices increasingly rely on interconnected networks, the ability to detect intrusions quickly and reliably has become essential. Intrusion Detection Systems (IDS) are a central component in this defence strategy as they monitor network traffic and system activity to identify malicious behaviour, providing both a safeguard against known threats and an early warning mechanism for new or evolving attacks [1].

Despite their importance, building effective IDS remains a challenge. Network traffic data is high-dimensional, heterogeneous, and often heavily imbalanced, with a vast majority of samples belonging to benign classes and only a fraction representing malicious activity. Traditional machine learning models, such as decision trees, random

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026
S. Lanka et al. (Eds.): ICTSM 2025, LNNS 1755, pp. 470–484, 2026.
https://doi.org/10.1007/978-3-032-13177-5_37

forests, and support vector machines, have been widely used in IDS research. While they perform well in some settings, they typically rely on extensive feature engineering and are limited in their ability to capture complex feature interactions. Deep learning methods, including convolutional and recurrent neural networks, have also been applied to IDS tasks [2]. These models are capable of learning powerful representations but often operate as "black boxes," offering little transparency into how predictions are made. For intrusion detection, where explainability is critical for security analysts and system operators, this lack of interpretability poses a serious limitation.

Tabular data, such as network flow records, presents additional difficulties for deep learning. Most deep neural architectures are optimised for unstructured data like images, audio, or text, rather than the structured, column-based format typical of IDS datasets. This mismatch has motivated the exploration of models that are designed specifically for tabular inputs. One such model is TabNet, a neural architecture that combines sequential attention mechanisms with sparse feature selection. TabNet dynamically learns which features to focus on at each decision step, producing predictions that are both accurate and interpretable. Its attention masks highlight the most relevant features for classification, offering valuable insight into the decision-making process [3] .

In this paper, we explore the application of TabNet to intrusion detection using benchmark datasets such as UNSW-NB15, BoT-IoT and KDD Cup. Our contributions are threefold. First, we demonstrate that TabNet can achieve competitive or superior detection performance compared with widely used machine learning baselines. Second, we show that TabNet's sparse attention masks provide interpretable feature importance scores, helping to identify the characteristics of network traffic most associated with attacks. Finally, we discuss the implications of adopting interpretable deep models for IDS, highlighting how they can improve both the accuracy of detection and the trust of human operators. The key contributions and novelty of this paper can be summarised as follows:

- **Application of TabNet to IDS:** We explore TabNet in the context of network intrusion detection, a domain where deep models have rarely been optimised for tabular data.
- **Bridging accuracy and interpretability:** Unlike CNN/LSTM-based IDS models, which act as black-box classifiers, TabNet provides both high classification performance and feature-level interpretability through sparse attention masks.
- **Benchmark validation:** We evaluate TabNet on widely used benchmark datasets (UNSW-NB15, BoT-IoT and KDD 99), demonstrating competitive or superior detection performance compared with traditional Machine Learning baselines.
- **Feature importance insights:** The study highlights how TabNet identifies the most influential traffic features for attack detection, offering insights that are valuable for both research and practical security operations.

2 Related Work

Intrusion detection has been an active research area for decades, with approaches evolving alongside the increasing complexity of network environments. Early systems relied on statistical thresholds and rule-based engines, but these methods struggled to adapt to the diversity of modern traffic patterns and the creativity of adversaries [4]. This led to the

widespread adoption of machine learning techniques, which promised more flexibility in identifying unknown attacks.

2.1 Tabnet Architecture

The TabNet model is a deep learning architecture specifically designed for tabular data, which is characterized by its structured nature and varying data types. Unlike traditional neural networks that often struggle with tabular data, TabNet employs a novel approach that combines sequential attention mechanisms with feature transformation to effectively capture the dependencies and interactions within the data as shown in Fig. 1. The architecture is composed of several key components, each contributing to the model's ability to process and learn from tabular datasets [5].

2.1.1 Feature Transformer

The Feature Transformer is a critical component of the TabNet architecture, responsible for processing the input features and transforming them into a representation that is conducive to accurate prediction. The Feature Transformer consists of a series of fully connected layers, interleaved with batch normalization (BN) and rectified linear unit (ReLU) activation functions. These layers work together to perform non-linear transformations on the input data, enabling the model to capture complex feature interactions that are often present in tabular datasets [6].

The Feature Transformer is divided into two distinct parts: shared and decision-specific layers. The shared layers process the input features across all decision steps, ensuring that the model can leverage common representations throughout the decision-making process. The decision-specific layers, on the other hand, are unique to each decision step, allowing the model to adaptively focus on different aspects of the data at each stage of the prediction process. This combination of shared and decision-specific transformations enables TabNet to balance generalization and specialization, resulting in more accurate predictions [7].

The Feature Transformer consists of a series of fully connected layers that apply non-linear transformations to the input features. For a given input feature vector X ∈ Rd., the transformation performed by the l-th layer of the Feature Transformer can be expressed as:

$$\mathbf{H}_l = \sigma(\mathrm{BN}(\mathbf{W}_l \mathbf{H}_{l-1} + \mathbf{b}_l))$$

where: – H_l is the output of the l-th layer, − $W_l \in$ Rd. × d is the weight matrix, − $b_l \in$ Rd. is the bias vector, -BN(·) denotes the batch normalization operation, − σ (·) is the activation function (typically ReLU), − H0 = X is the input feature vector. The output of the Feature Transformer after L layers is denoted as HL, which serves as the input to the subsequent components.

2.1.2 Attentive Transformer

The Attentive Transformer is another essential component of the TabNet architecture, playing a crucial role in the model's ability to dynamically select relevant features for

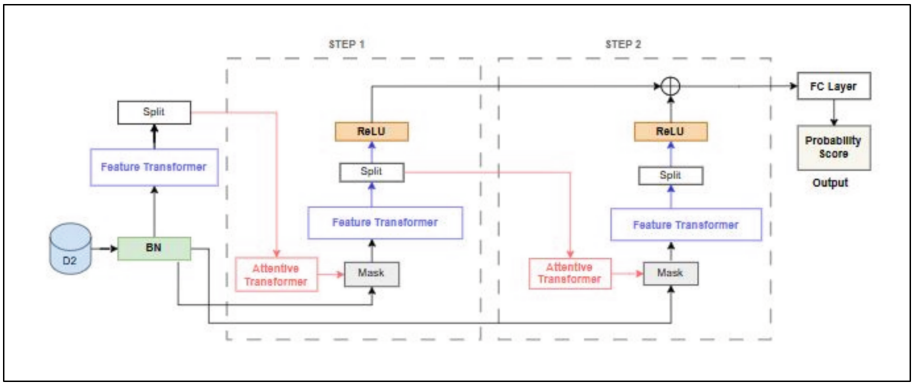


Fig. 1. Architecture of TabNet

each decision step. This mechanism is inspired by the attention mechanisms used in natural language processing, where the model learns to focus on the most important parts of the input sequence. In TabNet, the Attentive Transformer generates an attention mask that highlights the features most relevant to the current decision step [8].

The attention mask is generated based on the output of the previous decision step, allowing the model to iteratively refine its focus on the input features. This sequential attention mechanism ensures that the model can effectively handle complex feature interactions, even when they span multiple decision steps. By iteratively updating the attention mask, TabNet is able to progressively focus on the most informative features, improving the overall accuracy of its predictions.

The Attentive Transformer generates an attention mask to select relevant features for each decision step. Given the output from the previous decision step D_{t-1}, the attention mask M_t at decision step t is computed as:

$$\mathbf{M}_t = \text{Softmax}(\mathbf{U}_t \, \text{TanH}(\mathbf{V}_t \mathbf{D}_{t-1}))$$

where: – $U_t \in$ Rd. $\times$ d and $V_t \in$ Rd. $\times$ d are learnable weight matrices, – TanH$(\cdot)$ is the hyperbolic tangent activation function, – Softmax$(\cdot)$ is the softmax function applied across the feature dimensions to produce a probabilistic mask.

The attention mask M_t is then applied to the input features, producing a masked feature vector:

$$\mathbf{X}_t = \mathbf{M}_t \odot \mathbf{X}$$

where $\odot$ denotes element-wise multiplication.

2.1.3 Decision Steps

TabNet's decision-making process is organized into multiple sequential steps, each of which makes a partial decision based on the input data. At each decision step, the model uses the Attentive Transformer to generate an attention mask, which determines the subset of features that will be processed by the Feature Transformer. The output of

the Feature Transformer at each decision step is then combined with the outputs from previous steps to form the final prediction [9].

This multi-step decision process allows TabNet to model complex dependencies within the data, as each step can focus on different aspects of the input features. The use of attention masks ensures that the model can dynamically adjust its focus, enabling it to capture both local and global feature interactions. Additionally, the sequential nature of the decision steps allows TabNet to build its predictions gradually, reducing the risk of overfitting and improving the model's ability to generalize to new data.

At each decision step t, the model generates a decision vector D_t based on the masked features X_t processed through the Feature Transformer:

$$\mathbf{D}_t = \mathbf{H}_L^{(t)}$$

The final output y is computed by aggregating the decision vectors from all T steps:

$$\mathbf{y} = \sum_{t=1}^{T} \mathbf{D}_t$$

For classification tasks, y is typically passed through a softmax function to produce the class probabilities.

2.1.4 Definition of Sparsemax

The Sparsemax function maps an input vector z ∈ Rd. to a sparse probability distribution. It is defined as:

$$\text{Sparsemax}(\mathbf{z}) = \underset{\mathbf{p} \in \Delta^{d-1}}{\text{argmin}} \|\mathbf{p} - \mathbf{z}\|_2^2$$

where: p ∈ Rd. is the output vector of the Sparsemax function, Δd − 1 = {p ∈ Rd. | p ≥ 0, $\sum$d pi = 1} is the (d − 1)-dimensional probability simplex, − ‖ · ‖2 denotes the Euclidean norm.

In simpler terms, Sparsemax projects the input vector z onto the probability simplex, resulting in a sparse vector p where many elements are exactly zero.

Computation of Sparsemax The Sparsemax function can be computed using the following steps: 1. Sort the elements of the input vector z in descending order, denoted as z(1) ≥ z(2) ≥ · · · ≥ z(d). 2. Find the largest k ∈ {1, 2,. .., d} such that:

$$z_{(k)} + \frac{1}{k}\left(1 - \sum_{j=1}^{k} z_{(j)}\right) > 0$$

Compute the threshold τ as:

$$\tau = \frac{1}{k}\left(1 - \sum_{j=1}^{k} z_{(j)}\right)$$

The Sparsemax output p is then given by:

$$p_i = \max(z_i - \tau, 0)$$

Sparsemax in TabNet In the TabNet model, the Sparsemax function is applied within the Attentive Trans- former to generate the attention masks Mt. at each decision step. Given the output of the attention mechanism zt, the attention mask is computed as:

$$\mathbf{M}_t = \text{Sparsemax}(\mathbf{z}_t)$$

The sparsity of Mt. ensures that only the most relevant features are selected for processing by the Feature Transformer at each decision step. This sparsity is essential for the interpretability of TabNet, as it allows the model to focus on a small subset of features, making it easier to understand the decision-making process.

2.1.5 Sparsity-Inducing Mechanism

A unique aspect of the TabNet architecture is its built-in sparsity-inducing mechanism, which encourages the model to focus on a subset of relevant features at each decision step. This is achieved through the use of an entropy-based regularization term that penalizes the model for using too many features. By introducing this regularization, TabNet is able to produce more interpretable models, as the attention masks highlight the most important features for each decision [10].

The sparsity-inducing mechanism also contributes to the model's efficiency, as it reduces the computational complexity by limiting the number of features that need to be processed at each decision step. This makes Tab- Net particularly well-suited for large-scale tabular datasets, where the number of features can be substantial. By focusing on the most relevant features, TabNet not only improves prediction accuracy but also enhances the interpretability of the model, making it easier to understand the reasoning behind its predictions.

The sparsity-inducing mechanism in TabNet is implemented through an entropy-based regularization term applied to the attention masks Mt. The entropy of the attention mask at each step t is given by:

$$\mathcal{H}(\mathbf{M}_t) = -\sum_{i=1}^{d} \mathbf{M}_{t,i} \log\left(\mathbf{M}_{t,i}\right)$$

where $M_{t,i}$ is the i-th element of the attention mask M_t.

The overall sparsity regularization loss is then the sum of the entropies across all decision steps:

$$\mathcal{L}_{\text{sparse}} = \lambda \sum_{t=1}^{T} \mathcal{H}(\mathbf{M}_t)$$

where λ is a hyperparameter controlling the strength of the sparsity regularization.

2.1.6 Final Prediction and Loss Function

The final prediction in TabNet is obtained by aggregating the outputs from all decision steps. This aggregation can be performed in various ways, such as summing the outputs

or applying a weighted average. The final output is then passed through a softmax or sigmoid activation function, depending on whether the task is classification or regression.

TabNet is trained using a loss function that combines standard prediction loss (e.g., cross-entropy loss for classification tasks) with the entropy-based regularization term that enforces sparsity. The combination of these loss terms ensures that the model not only achieves high accuracy but also remains interpretable and efficient [11]. The use of gradient-based optimization techniques allows TabNet to learn the optimal parameters for both the Feature Transformer and the Attentive Transformer, resulting in a model that is both powerful and adaptable.

The final loss function for training the TabNet model combines the prediction loss Lpred (e.g., cross-entropy loss for classification tasks) with the sparsity regularization loss:

$$L = L_{pred} + L_{sparse}$$

The model is trained by minimizing this loss function using gradient-based optimization methods.

2.2 Classical Machine Learning Approaches

Among classical methods, Random Forests (RF), Support Vector Machines (SVMs), and k-Nearest Neighbors (kNN) have been widely deployed for IDS tasks. These models generally perform well on tabular IDS datasets such as KDD99, NSL-KDD, and UNSW-NB15, offering reasonable accuracy and robustness. For instance, RF has been shown to handle high-dimensional features effectively [12] while SVMs remain competitive in binary classification scenarios. However, a common drawback is their reliance on manually crafted features and their limited interpretability they can flag anomalies, but it is difficult for security operators to understand why a particular decision was made.

2.3 Deep Learning for IDS

With the growth of deep learning, researchers began applying neural architectures to intrusion detection. Convolutional Neural Networks (CNNs) were adapted to treat traffic records as "images" or structured sequences, capturing local feature interactions (Kim et al. 2016). Long Short-Term Memory (LSTM) networks and other recurrent models leveraged temporal dynamics to detect sequential attack patterns [13] These approaches demonstrated strong predictive performance, often surpassing classical baselines. Yet, they suffer from two key issues: (1) computational cost when applied to large-scale traffic, and (2) opacity these models act as black boxes, leaving analysts without clear explanations of which traffic features signalled an attack.

2.4 Feature Selection and Interpretability in IDS

To address the interpretability challenge, researchers have explored feature selection techniques. Filter-based methods using correlation, mutual information, or statistical

ranking have been applied to reduce dimensionality [14]. Wrapper and embedded methods, such as those built on decision trees or evolutionary algorithms, have also been used to identify key attributes [15]. While these approaches can highlight relevant features, they are usually decoupled from the learning process itself: the classifier consumes a fixed set of features, but does not inherently explain which attributes drive predictions at runtime.

2.5 TabNet and Its Applications in Security

TabNet [16] was proposed as a deep learning architecture tailored for tabular data. By using sequential attention and sparse feature masks, TabNet can focus on the most informative features at each decision step, while maintaining competitive performance. Outside of IDS, TabNet has been applied in domains such as healthcare, finance, and fraud detection, demonstrating its ability to handle complex, structured datasets.

In the cybersecurity space, TabNet is beginning to attract attention [17]. Used TabNet for elephant flow detection, relying on only the first packet's header information, and achieved high accuracy while reducing the number of flow table entries [15]. introduced TabSec, a collaborative framework for insider threat detection, where TabNet's sparse attention was key to identifying subtle malicious behaviors in large-scale user activity data. More recently, [18] proposed TabNet-SFO, which integrates TabNet with Sailfish Optimization to improve intrusion detection in smart water systems. This work further demonstrates TabNet's ability to highlight important features directly from high-dimensional security data.

In contrast, most traditional IDS approaches treat feature selection as a preprocessing step: the classifier consumes a fixed set of engineered attributes but does not inherently explain which features drive its decisions during classification.

2.6 Research Gap

These studies demonstrate the growing recognition of TabNet's potential in network security. However, systematic evaluations of TabNet on widely used IDS benchmarks such as UNSW-NB15 and BoT-IoT remain limited. Moreover, while prior work has shown TabNet's ability to highlight relevant features, few studies have directly compared it against both classical machine learning and deep neural baselines in the IDS context. This gap motivates our work: we systematically evaluate TabNet for IDS, benchmark it against established machine learning models, and provide a detailed analysis of its feature attention masks to assess both accuracy and interpretability.

3 Methodology

This section outlines the datasets used, the preprocessing steps applied, the TabNet model architecture, the training setup, and the baseline models employed for comparison.

3.1 Datasets

We evaluated TabNet on three benchmark intrusion detection datasets:

- **UNSW-NB15:** A modern dataset containing nine attack categories and normal traffic, collected using the IXIA PerfectStorm tool. It provides 49 features describing flow characteristics, protocols, and payload behaviour [14].
- **BoT-IoT:** A recent IoT-focused dataset that captures botnet attacks such as DDoS, DoS, and brute force alongside normal traffic. It reflects modern IoT environments with severe class imbalance, as benign traffic dominates [19].
- **KDD 99:** A legacy dataset that has been widely used in IDS research. Despite criticisms regarding redundancy and outdated attack profiles, it remains a useful benchmark for comparing new techniques against a large body of prior work.
- These datasets were chosen to provide a balance between legacy baselines (KDD 99) and contemporary, realistic traffic scenarios (UNSW-NB15 and BoT-IoT).

3.2 Data Preprocessing

To ensure comparability and reliable model training, the following preprocessing steps were applied:

1 **Normalization:** Continuous features were scaled to the [0,1] range using min–max normalization.
2 **Encoding categorical features:** Non-numeric attributes such as protocol type and service were encoded using one-hot encoding.
3 **Handling imbalance:** Given that UNSW-NB15 and BoT-IoT are highly imbalanced, we employed a combination of random undersampling for the majority class and Synthetic Minority Oversampling Technique (SMOTE) for underrepresented classes.
4 **Train–test split:** For each dataset, 80% of samples were used for training and 20% for testing.

3.3 TabNet Model Architecture

TabNet is a neural network architecture designed specifically for tabular data [16]. Its key components include:

- **Feature Transformer:** A sequence of shared and decision-specific fully connected layers that process input features into rich intermediate representations.
- **Attentive Transformer:** Generates sparse attention masks that determine which features are most relevant at each decision step.
- **Sparsemax activation:** Ensures sparsity in the attention masks, forcing the model to select a small subset of features at each step, thereby enhancing interpretability.
- **Decision steps:** Multiple sequential decision steps progressively refine the prediction, each focusing on different feature subsets.

The combination of attention-driven selection and feature transformation enables TabNet to capture complex dependencies in tabular data while maintaining interpretability.

3.4 Training Setup

The TabNet model was implemented using the open-source pytorch-tabnet framework. Key training parameters were as follows:

- **Batch size:** 1024
- **Learning rate**: 0.02 with a cosine decay scheduler
- **Number of decision steps**: 5
- **Sparsity regularization (λ):** 0.001
- **Optimizer:** Adam with default β parameters ($\beta 1 = 0.9$, $\beta 2 = 0.999$)
- **Epochs:** Training continued for up to 100 epochs with early stopping based on validation loss.

Evaluation metrics included Accuracy, Precision, Recall and F1-score.

3.5 Baseline Models

For comparison, we selected established and widely used machine learning classifiers that are common benchmarks in IDS research:

- Logistic Regression (LR): A simple linear baseline, useful for assessing separability [20].
- Random Forest (RF): An ensemble of decision trees, widely adopted for IDS due to robustness and interpretability.
- Support Vector Machine (SVM, linear kernel): A strong classical model effective on structured data [21].
- Naïve Bayes (NB): A probabilistic baseline suitable for tabular data.

4 Results and Analysıs

We evaluated TabNet against several baselines (Logistic Regression, Random Forest, Linear SVM, and Naïve Bayes) on the UNSW-NB15, BoT-IoT, and KDD datasets. Performance was assessed using Accuracy, Precision, Recall, and F1-score.

On UNSW-NB15, Random Forest achieved the highest performance with an accuracy of 99.90%, closely followed by TabNet at 99.30%. While the difference in raw accuracy was minimal (0.6%), TabNet maintained balanced scores across Precision (99.31%), Recall (99.31%), and F1-score (99.31%), demonstrating that it consistently detected both majority and minority classes as shown in Table 1. Classical models such as Logistic Regression (94.80%) and Linear SVM (94.60%) performed noticeably worse, while Naïve Bayes (77.39%) lagged significantly.

On BoT-IoT, TabNet delivered near-perfect performance, reaching 99.98% accuracy with an F1-score of 99.99%. By comparison, traditional classifiers such as Logistic Regression, SVM, and Random Forest achieved only 87.5% accuracy, despite inflated Precision and Recall values (~99.99%) as shown in Table 2. This highlights TabNet's robustness in highly imbalanced datasets where rare attack classes are easily overlooked by baselines.

Table 1. Comparison of results with UNSW-NB15

	Accuracy	Precision	Recall	F1
LR	94.80	94.80	94.80	94.80
RF	99.90	99.90	99.90	99.90
Linear SVM	94.60	94.60	94.60	94.60
Naïve Bayes	77.39	78.54	77.39	77.43
Tabnet	99.30	99.31	99.31	99.31

Table 2. Comparison of results with BoT-IoT

	Accuracy	Precision	Recall	F1
LR	87.50	99.99	99.99	99.99
RF	87.50	99.99	99.99	99.99
Linear SVM	87.50	99.99	99.99	99.99
Naïve Bayes	87.49	99.99	99.99	99.99
Tabnet	99.98	99.99	99.98	99.99

On the KDD dataset, TabNet once again achieved near-perfect results with 99.98% accuracy and an F1-score of 99.99%, outperforming Random Forest (98.90%) and far surpassing Logistic Regression (93.50%), Linear SVM (94.90%), and Naïve Bayes (90.50%) as shown in Table 3.

Table 3. Comparison of results with KDD CUP

	Accuracy	Precision	Recall	F1
LR	93.50	93.50	93.50	93.50
RF	98.90	98.90	98.90	98.90
Linear SVM	94.90	94.90	94.90	94.90
Naïve Bayes	90.50	90.50	90.50	90.50
Tabnet	99.98	99.99	99.98	99.99

Overall, these results demonstrate that TabNet achieves competitive or superior performance compared to classical ML models, particularly excelling on modern, imbalanced datasets such as BoT-IoT.

4.1 Interpretability Analysis

A defining strength of TabNet is its ability to generate sparse attention masks, which highlight the most informative features at each decision step.

- For UNSW-NB15, TabNet consistently emphasized features such as flow duration, source bytes, and service type attributes well known to be linked with scanning and denial-of-service attacks.
- For BoT-IoT, attention focused on packet rate, login attempts, and the byte-to-packet ratio, which are critical in identifying brute force and flooding attacks.
- The sparsity of the attention distributions ensures that only a few features are used per decision step. This not only improves computational efficiency but also provides actionable explanations for security analysts.
- Compared with Random Forest feature importance, TabNet offered more attack-specific feature relevance and avoided biases toward high-cardinality variables, making its interpretability more aligned with security use cases.

4.2 Robustness to Imbalance

Both UNSW-NB15 and BoT-IoT are highly imbalanced, with minority classes such as Worms, Shellcode, and Data Theft significantly underrepresented. Traditional models often failed to capture these classes, inflating global metrics while overlooking critical anomalies.

TabNet's entropy-regularized sparse feature selection mitigated this issue. On BoT-IoT, TabNet sustained $>99.9\%$ F1-scores while other classifiers misclassified minority attacks despite reporting high global Precision and Recall. On UNSW-NB15, TabNet achieved balanced detection across classes, unlike baselines which tended to overfit to majority traffic. This robustness suggests that TabNet generalizes better in real-world IDS deployments where novel or rare attack behaviors are expected.

4.3 Comparative Insights

- **Accuracy vs. Interpretability:** While Random Forest slightly outperformed TabNet in raw accuracy on UNSW-NB15, TabNet offered feature-level explanations that RF could not provide.
- **Efficiency:** TabNet converged faster than CNN-based IDS models, thanks to its efficient feature selection mechanism.
- **Cross-dataset generalization:** TabNet achieved consistently strong performance across UNSW-NB15, BoT-IoT, and KDD, indicating its ability to adapt to diverse IDS benchmarks.

4.4 Summary of Findings

- TabNet matches or outperforms classical baselines in detection accuracy, particularly excelling on modern, imbalanced IDS datasets.
- Its sparse attention masks provide feature-level interpretability that aligns with domain knowledge and supports practical use in SOC environments.
- TabNet demonstrates robustness against imbalance, sustaining high recall on rare but critical attack types.
- It offers a practical trade-off between accuracy, efficiency, and interpretability qualities essential for next-generation intrusion detection systems.

5 Practıcal Applıcatıons

The results of this work are not only of academic interest but also carry clear value for day-to-day security practice. Intrusion detection systems are most effective when they can support the people who use them typically analysts in a security operations centre. Models that deliver perfect accuracy on paper but behave as black boxes rarely gain trust in practice. By contrast, TabNet offers both strong detection performance and a way of showing which features influenced its decision. In a SOC environment, this means that when an alert is raised, the analyst is not left guessing. If TabNet points to unusually high packet rates or repeated login attempts as the cause, the operator can immediately see why the traffic was flagged and decide on the right response more quickly. The model's efficiency also makes it attractive for organisations with limited resources, such as IoT deployments or smaller enterprises. Sparse feature selection reduces unnecessary processing, allowing the system to run effectively without requiring expensive hardware. Finally, because TabNet produces explanations alongside its predictions, it can help organisations meet growing compliance and audit requirements where the reasoning behind automated decisions must be documented. In short, the practical contribution of this work is to show that an IDS can be both accurate and interpretable, making it more usable and trustworthy in real operational settings.

6 Conclusıon and Future Work

This study explored the use of TabNet as a next-generation model for intrusion detection. By applying it to three benchmark datasets UNSW-NB15, BoT-IoT, and KDD we demonstrated that TabNet achieves highly competitive performance across diverse IDS scenarios. On BoT-IoT and KDD, it delivered near-perfect results, while on UNSW-NB15 it performed on par with Random Forest, one of the strongest classical baselines. Beyond raw accuracy, TabNet provided an additional advantage: its sparse attention masks highlighted critical traffic features such as flow duration, packet rate, and login attempts, offering explanations that align with known attack behaviors. This ability to combine high accuracy with feature-level interpretability makes TabNet a practical and trustworthy choice for deployment in operational security environments. There remain, however, important directions for future work. First, while TabNet improves robustness against class imbalance, further work is needed to assess its effectiveness on zero-day

attacks and highly dynamic traffic. Second, exploring transfer learning and lightweight fine-tuning techniques could enable TabNet to adapt across different network domains without full retraining. Finally, integrating TabNet with other detection frameworks such as ensemble systems or hybrid IDS architectures may further enhance detection accuracy while retaining interpretability. In summary, TabNet strikes a balance between predictive power and transparency, addressing a long-standing gap in IDS research. With further refinement and adaptation, it has the potential to become a cornerstone model for intrusion detection in modern cybersecurity operations.

References

1. Kim, H.W., Park, J.H., Jeong, Y.S.: Human-intelligence workflow management for the big data of augmented reality on cloud infrastructure. Neurocomputing. **279**, 19–26 (2018) Available from: https://www.sciencedirect.com/science/article/pii/S0925231217317782
2. Liu, H., Lang, B.: Machine learning and deep learning methods for intrusion detection systems: a survey. Appl. Sci. **9**(20), 4396 (2019) Available from: https://www.proquest.com/docview/2533676898
3. Chauhan, V.K., Dahiya, K., Sharma, A.: Problem formulations and solvers in linear SVM: a review. Artif. Intell. Rev. **52**(2), 803–855 (2019) Available from: https://link.springer.com/article/10.1007/s10462-018-9614-6
4. Nguyen, H.V., Byeon, H.: Predicting depression during the COVID-19 pandemic using interpretable TabNet: a case study in South Korea. Mathematics. **11**(14), 3145 (2023) Available from: https://www.proquest.com/docview/2843079295
5. Joseph, L.P., Joseph, E.A., Prasad, R.: Explainable diabetes classification using hybrid Bayesian-optimized TabNet architecture. Comput. Biol. Med. **151**(Pt A), 106178 (2022) Available from: https://www.clinicalkey.com/#!/content/1-s2.0-S0010482522008861
6. Shah, C., Du, Q., Xu, Y.: Enhanced TabNet: attentive interpretable tabular learning for hyperspectral image classification. Remote Sens. **14**(3), 716 (2022) Available from: https://www.proquest.com/docview/2627828342
7. Zhang, Z., Yu, T., Zhang, Y., Zhao, Y., Zhang, Z.: A quantitative benchmark of neural network feature selection methods. Sci. Rep. **14**, 82583 (2024) Available from: https://doi.org/10.1038/s41598-024-82583-2
8. Koroniotis, N., Moustafa, N., Sitnikova, E., Turnbull, B.: Towards the development of realistic botnet dataset in the Internet of Things for network forensic analytics: Bot-IoT dataset. Futur. Gener. Comput. Syst. **100**, 779–796 (2019). Available from: https://doi.org/10.1016/j.future.2019.05.041
9. Setitra, M.A., Fan, M.: Detection of DDoS attacks in SDN-based VANET using optimized TabNet. **90**, 103845 (2024) Available from: https://doi.org/10.1016/j.csi.2024.103845
10. Nikzad, M.H., Heidari-Rarani, M., Rasti, R.: A novel systematically optimized tabular neural network (TabNet) algorithm for predicting the tensile modulus of additively manufactured PLA parts. Mater. Today Commun. **41**, 110442 (2024). Available from: https://doi.org/10.1016/j.mtcomm.2024.110442
11. Rajeh, W., Aborokbah, M.M., Manimurugan, S., Alashoor, T., Karthikeyan, P.: TabNet-SFO: an intrusion detection model for smart water management in smart cities. Int. J. Intell. Syst. (1) (2025, 2025) Available from: https://www.proquest.com/docview/3180427185
12. Zhang, Y., Huang, H., Zhu, J., Zhao, Z.: Achieving cross-modal generalization with multimodal unified representation [Internet]. 2023. p. 1–12. Advances in Neural Information Processing Systems (NeurIPS 2023). Available from: https://proceedings.neurips.cc/paper_files/paper/2023/file/c89f09849eb5af489abb122394ff0f0b-Paper-Conference.pdf

13. Javaid, A., Niyaz, Q., Sun, W., Alam, M.: A deep learning approach for network intrusion detection system. EAI. **3**(9), 21–26 (2016) Available from: https://www.proquest.com/docview/2306343062
14. Meurer, W.J., Tolles, J.: Logistic regression diagnostics: understanding how well a model predicts outcomes. JAMA. **317**(10), 1068–1069 (2017). Available from: https://doi.org/10.1001/jama.2016.20441
15. Arik, S.Ö., Pfister, T.: TabNet: attentive interpretable tabular learning. AAAI. **35**(8), 6679–6687 (2021)
16. Venugopal, P., Rajesh, V., Nair, N.: DDoS attack detection: a comparative study of TabNet classifier with isolation forest and tabular transformer models. AIP Conf. Proc. **3237**(1) (2025) Available from: https://doi.org/10.1063/5.0247903
17. Sommer, R., Paxson, V.: Outside the Closed World: On Using Machine Learning for Network Intrusion Detection. In IEEE, pp. 305–16. (2010 IEEE Symposium on Security and Privacy) (2010). Available from: https://ieeexplore.ieee.org/document/5504793
18. Moustafa, N., Slay, J.: UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set). In IEEE, p. 1–6. (2015 Military Communications and Information Systems Conference (MilCIS)) (2015). Available from: https://ieeexplore.ieee.org/document/7348942
19. Li, W.: TabNet for high-dimensional tabular data: advancing interpretability and performance with feature fusion. IET Conf. Proc. **2025**(2), 168–173 (2025)
20. Meszaros, J., Buchalcevova, A.: Introducing OSSF: a framework for online service cybersecurity risk management. Comput. Secur. **65**, 300–313 (2017)
21. Zhang, K., Li, Y., Scarf, P., Ball, A.: Feature selection for high-dimensional machinery fault diagnosis data using multiple models and Radial Basis Function networks. Neurocomputing. **74**(17), 2941–2952 (2011). Available from: https://doi.org/10.1016/j.neucom.2011.03.043

Design and Optimization of Rectangular-Shaped Patch Antenna for Wireless Applications at Ka Band

Jannatul Mim Samia, Md. Sohel Rana(✉), Mithun Bairagi, and A. S. M. Tanvir Ul Islam

Department of Electrical and Electronic Engineering, Northern University of Business and Technology Khulna, Khulna, Bangladesh
sohel.rana@uits.edu.bd

Abstract. This study presents the design, analysis, and optimisation of a rectangular patch antenna for wireless applications in the Ka band. Two distinct substrate materials (Rogers RT6202PR and Rogers RO4003C) have been employed for the antenna designs (Antenna 1 and Antenna 2). To attain optimal performance, the design process entails selecting a suitable substrate material, optimising patch dimensions, and employing a microstrip feeding method. This examination of the antenna's reflection coefficient (S_{11}), voltage standing wave ratio (VSWR), radiation pattern, gain, directivity, surface current, Smith Chart, and bandwidth is performed utilising diverse simulation methodologies in CST. The antenna is engineered for best performance inside a compact framework, rendering it appropriate for high-frequency communications, including satellite networks, radar systems, and advanced 5G services. The results indicate that the antenna exhibits a return loss above −10 dB and preserves a consistent emission pattern. This validates the antenna's appropriateness for high-speed wireless applications functioning in the Ka-band. The antenna's performance regarding return loss, bandwidth, efficiency, and radiation characteristics renders it an appropriate selection for high-frequency wireless applications inside the Ka-band.

Keywords: Patch Antenna · Wireless Applications · Reflection Coefficient · VSWR · Computer Simulation Technology (CST) · Ka-Band

1 Introduction

An antenna is a device capable of transmitting or receiving electromagnetic signals. In telecommunications, microstrip antennas are the most prevalent form, notwithstanding their constraints, such as limited bandwidth and reduced radiation efficiency. Nevertheless, they provide other benefits, such as minimal weight and reduced cost, which accounts for the extensive research being conducted on patch antennas for 5G applications [1, 2]. It's expected to become increasingly necessary in the near future. In the future, wireless communication would benefit from its ability to perform across

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026
S. Lanka et al. (Eds.): ICTSM 2025, LNNS 1755, pp. 485–497, 2026.
https://doi.org/10.1007/978-3-032-13177-5_38

several bands. Since the beginning of this century, there has been a steady progression in the development of wireless applications. Because of the growing demand for mobile devices, there is a pressing need to place a greater emphasis on antenna design to guarantee effective communication and the highest possible network performance [3]. Wireless communication technology signifies a substantial advancement, providing expedited data transmission, reduced latency, and enhanced network capacity relative to earlier generations [4]. The limitations of installation space prompt this proposal. At this point, it is of the utmost importance to build a single antenna that is capable of functioning across a variety of applications. To make a single patch antenna that improves both the coverage area and the effectiveness of communication, researchers are motivated by the basic conditions that must be met.

In this study [5], a square-slotted patch antenna design for 5G mmWave application at 37 GHz. This research indicates that the proposed antenna exhibits diminished return loss, gain, impedance bandwidth, VSWR, and radiation patterns in both the E-plane and H-plane. In this study [6], a new slot-patch antenna that operates in the Ka-band wavelength range is proposed. Both the bandwidth and gain are enhanced by this proposed design. This article presents [7], a super-wideband flower-slotted microstrip patch antenna (SWB-FSMPA) that is tiny and has a high bandwidth ratio (BDR). It is designed for use in applications that require super-wideband coverage. A simulated bandwidth that ranges from 3.78 to 109.86 GHz is made possible by this architecture, which permits coverage of a wide range of wireless applications. A gain, bandwidth, and efficiency are all characteristics of the SWB-FSMPA antenna. In the suggested paper [8], a single-band antenna is described, and the proposed antenna design is adjusted to suit the requirements of 5G technology. The VSWR, reflection coefficients, gain, and bandwidth of the finished structure are among the factors that are rigorously analyzed. This proposed antenna [1] was able to obtain a greater bandwidth, good gain and directivity, standard VSWR, reduced return loss, and better antenna efficiency. This research [9], design a patch antenna for 5G applications and the antenna had a favourable S_{11}, an enhanced gain, good impedance matching, and a compact design. Consequently, this results in a resolution that is both realistic and practicable for the next generation of mmWave communication systems.

2 Method, Materials and Design

This section discusses the methods, materials and designs of patch antenna design. For the design specification, we first studied our reputable journals and conference papers. The material has been selected for the said design. Two materials (Rogers RT6202PR and Rogers RO4003C) have been chosen for the design. Then the complete antenna is designed, and the parameters are evaluated for performance. CST software is used for validation and optimisation, and later the simulated results can be measured using VNA. Figure 1 illustrates the flowchart of the suggested design.

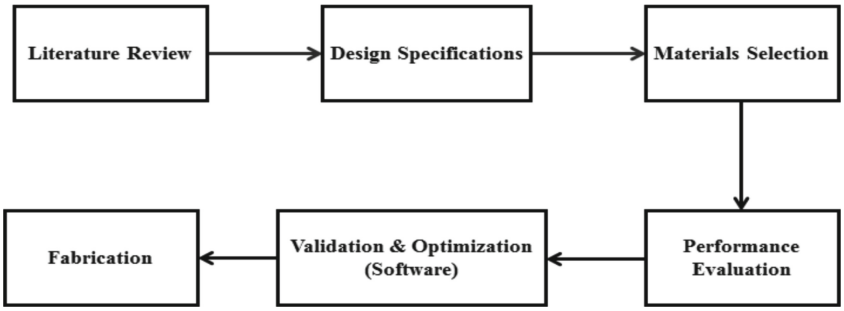


Fig. 1. Illustrates the flowchart of the suggested design

During the process of constructing an antenna, electromagnetic modelling software is utilized throughout the entire process. It is a valuable device for estimating the efficacy of the antenna and making any necessary adjustments. For the sake of this investigation, the primary starting design is a rectangular patch. To determine the dimensions of a rectangular area, it is necessary to use the following equations, which extend from Eq. (1) to (7) [10].

The width of the existing patch equation that has been proposed

$$Wp = \frac{c_0}{2f_r \sqrt{\frac{\varepsilon_r+1}{2}}} \tag{1}$$

Effective potential calculated using the dielectric constant ε_{reff}

$$\varepsilon_{reff} = \frac{\varepsilon_r + 1}{2} + \frac{\varepsilon_r - 1}{2}\left(1 + 12 \times \frac{h}{w}\right)^{-0.5} \tag{2}$$

Calculation of the effective length

$$L_{ext} = \frac{c_o}{2f_r\sqrt{\varepsilon_{reff}}} \tag{3}$$

Investigate the peripheral phenomenon and ascertain the dimensions of the patch.

$$\Delta L = 0.412h \frac{\left(\frac{w}{h} + 0.264\right)(\varepsilon_{reff} + 0.3)}{(\varepsilon_{reff} - 0.258)\left(\frac{w}{h} + 0.8\right)} \tag{4}$$

The precise length of the patch:

$$L_P = L_{ext} - 2 \times \Delta L \tag{5}$$

The measurements of the ground plane of the patch antenna:

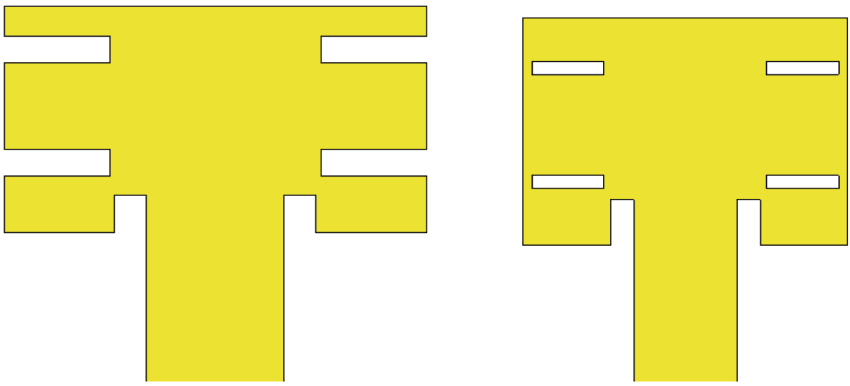
$$Lg = 6h + L_P \tag{6}$$

$$Wg = 6h + Wp \tag{7}$$

The suggested rectangular patch antenna for wireless applications, which operates at 25 GHz, has a dimensional structure shown Fig. 2. Also, the top and back views of the antenna design are depicted in Figs. 2(a), (b) and 3(a), (b), which also includes the copper layers and the substrate. Through the selection of a substrate material consisting of Rogers RT6202PR (Design 1), Rogers RO4003C (Design 2) with a dielectric permittivity of 2, 3.55 and loss tanet are 0.0021, 0.0027 and the establishment of the operating frequency at 25 GHz. The antenna's conducting elements are composed of copper with a thickness of 0.035 mm. The design approach encompassed meticulous modelling, simulation, and optimization using CST software. Each parameter was methodically optimised to attain the requisite frequency performance (Table 1).

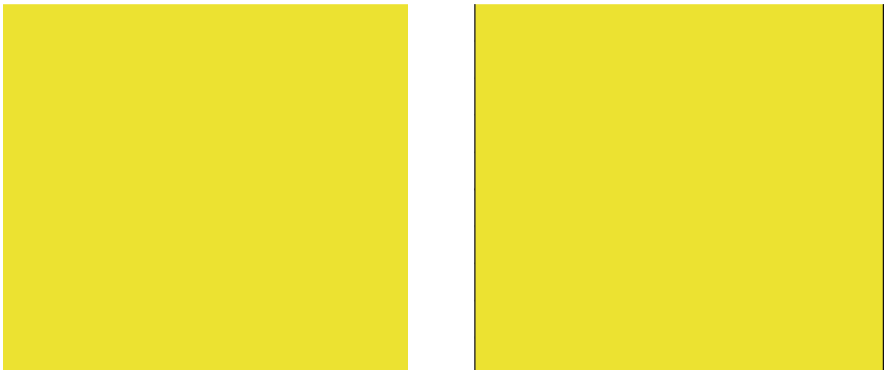
Table 1. Optimized design parameters of the proposed antenna

Parameter	Wg	Lg	Wp	Lp	Hs	sltRl	W_i	L_i	sltRl	WxF
Design I (mm)	11	10	7.91	4.239	0.78	0.5	0.6	0.7	0.7	2.572
Design II (mm)	11	10	5.549	4.49	0.78	0.26	0.9	0.4	0.205	1.7585



(a)Design 1 (a) Design 2

Fig. 2. Front view of the proposed antennas (Design 1 & 2)



(c)Design 1 (d) Design 2

Fig. 3. Back view of the proposed antenna (Design 1 & 2)

3 Results Analysıs

This section will explain the characteristics of the microstrip antenna that has been proposed. The antenna simulation is conducted using CST software to produce results. These characteristics include the reflection coefficient (S_{11}), the VSWR, the 3D gain and directivity, 2D and 3D polar plots, as well as the bandwidth, surface current, Accepted power, radiated power, stimulated power, and the efficiency. The findings of the simulation, as designed, are presented in Table 2. Additionally, a comparison of the previous work with the published work is presented in Table 3.

3.1 Reflection Coefficient

The return loss, which is often referred to as the S_{11} parameter, is a specification of an antenna that provides information regarding the amount of power that is reflected. The S_{11} parameter has a base value of −10 dB, indicating that the antenna can accept 90% of the applied power, with 10% reflected [11]. The antenna design 1 has a S_{11} of −47.66 dB at 25 GHz, while the antenna design 2 has a reflection coefficient of −53.10 dB are shown in Fig. 4.

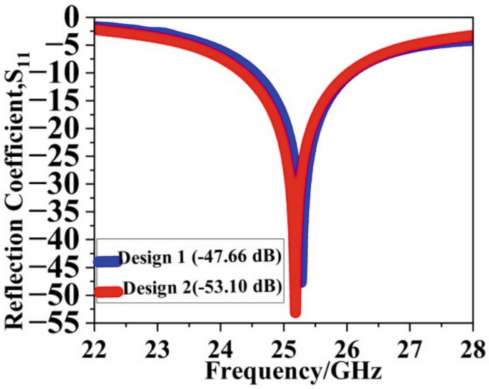


Fig. 4. Analysis of the Reflection Coefficient for Antenna Designs 1 and 2

3.2 Voltage Standing Wave Ratio

VSWR is an antenna characteristic that delineates the waveform pattern resulting from the intersection of forward and reflected power. An effective antenna should possess a VSWR value ranging from 1 to 2; however, a value approaching 1 signifies near-optimal performance, whereas values exceeding or falling below this range indicate subpar antenna quality. Figure 5 shows the VSWR of the antenna designs 1 (1.008) and 2 (1.004).

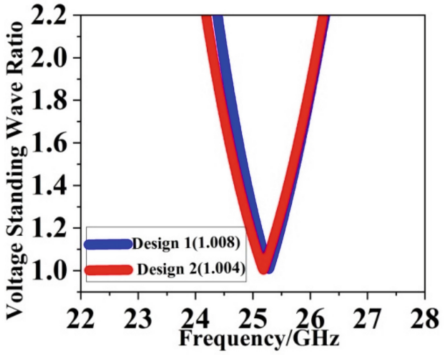
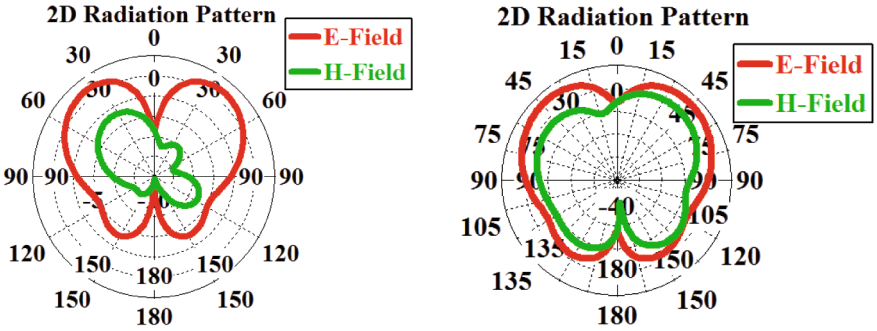


Fig. 5. View of the VSWR of the antenna designs 1 and 2.

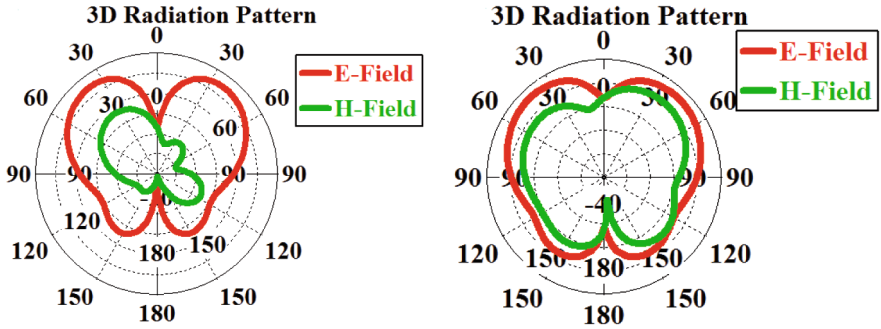
3.3 Radiation Pattern

Evaluation of the radiation pattern of an antenna is performed to ascertain the magnitude of the electromagnetic waves that are emitted by the antenna in several different directions. This pattern, depicted as a mathematical function, illustrates how the field intensity of the antenna changes depending on the direction in which it is pointed, either increasing or decreasing [12]. The 2D and 3D radiation pattern of the antenna designs 1 and 2 are shows Figs. 6 and 7.



(a) Design 1 (b) Design 2

Fig. 6. Examining the radiation pattern in two dimensions (Design 1 & 2)



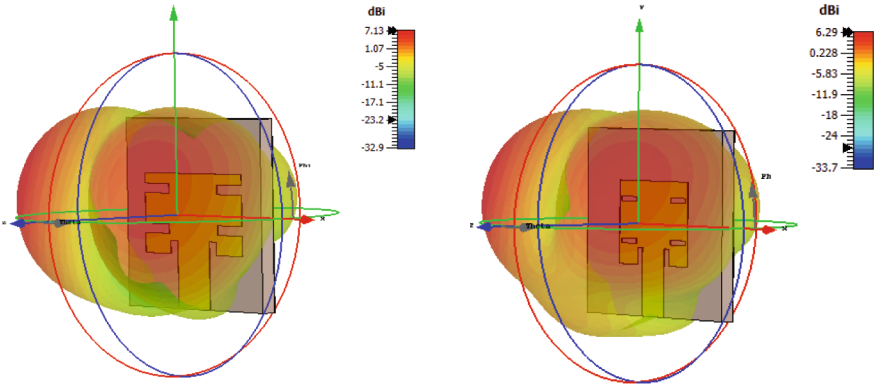
(a)Design 1 (b)Design 2

Fig. 7. An investigation of the radiation pattern when viewed in 3D (Design 1 & 2)

3.4 Gain and Directivity

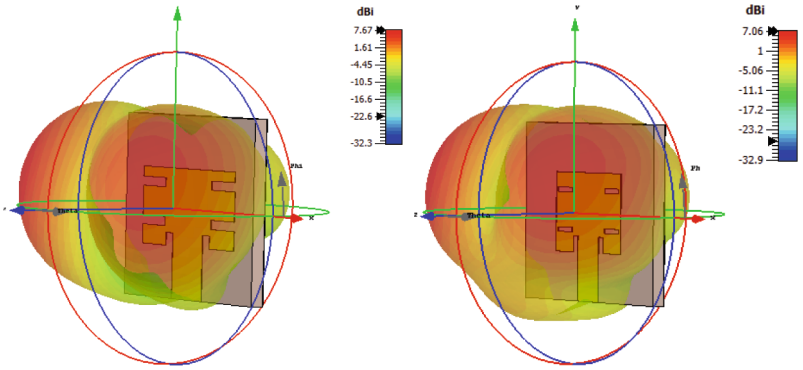
Gain and directivity are the two most important technical characteristics that determine the performance of an antenna. The term *"antenna gain"* refers to the assessment of how effectively an antenna directs the signal's strength in a specific direction [13]. Directivity is the enhancement of power density in a particular direction at a constant distance from a transmitting antenna. This is in contrast to the power density that occurs when the

same transmit power is uniformly distributed in all aspects of the transmission [14]. The gains of antenna designs 1 and 2 are 7.13 dBi and 6.26 dBi, respectively, as illustrated in Fig. 8. The directivity of antenna designs 1 and 2 is 7.67 dBi and 7.06 dBi, respectively, as illustrated in Fig. 9.



(a)Design 1 (b) Design 2

Fig. 8. View of 3D gain Pattern (Design 1 & 2)

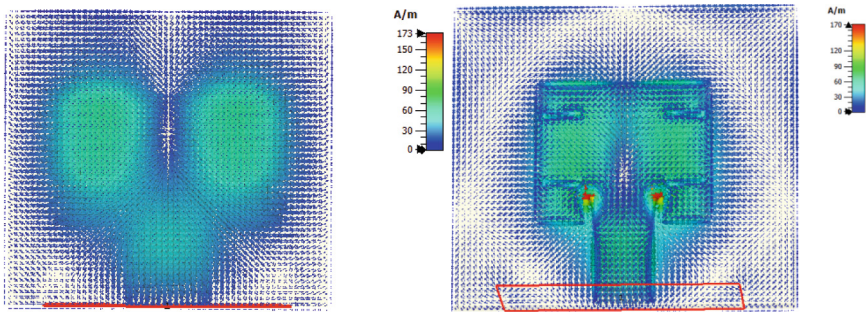


(a)Design 1 (b)Design 2

Fig. 9. View of 3D Directivity Pattern (Design 1 & 2)

3.5 Surface Current

The efficiency and antenna radiation pattern depend on the current distribution across the antenna [15]. Figure 10 illustrates the surface current distribution of the proposed antenna.



(a)Design 1 (b)Design 2

Fig. 10. A perspective on the surface current represented by designs 1 and 2

3.6 Impedance Matching and Smith Chart

Impedance matching is crucial in transmitters and receivers to optimise power flow between the antenna and the device. It is the procedure of aligning the impedance of a source with that of a load to optimise power transfer. Improper impedance matching can lead to signal reflection and inefficiency [16, 17]. Fig. 11 are shows the impedance matching of designs 1 (48.69 Ω) & 2 (49 Ω). An impedance and reflection coefficient along a transmission line can be determined with the use of a graphical tool called the Smith chart [18]. Smith chart of the designs 1 & 2 are shown in Fig. 12.

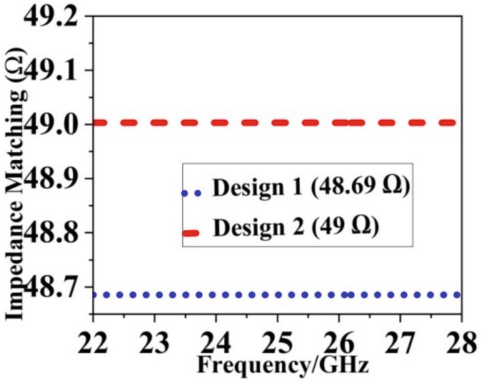
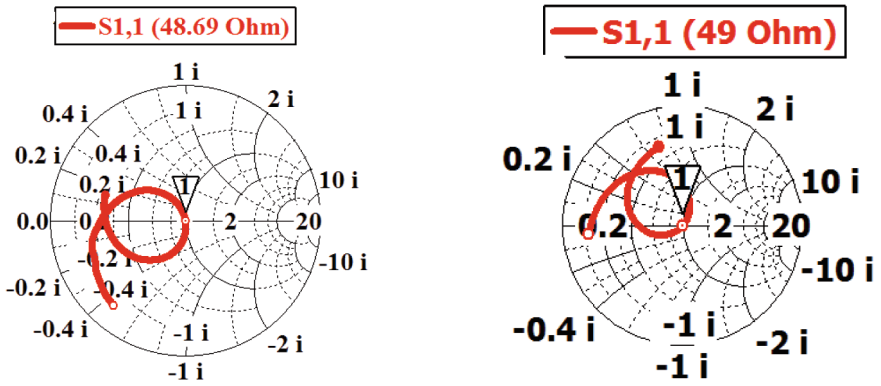


Fig. 11. View of Impedance Matching (Design 1 & 2)

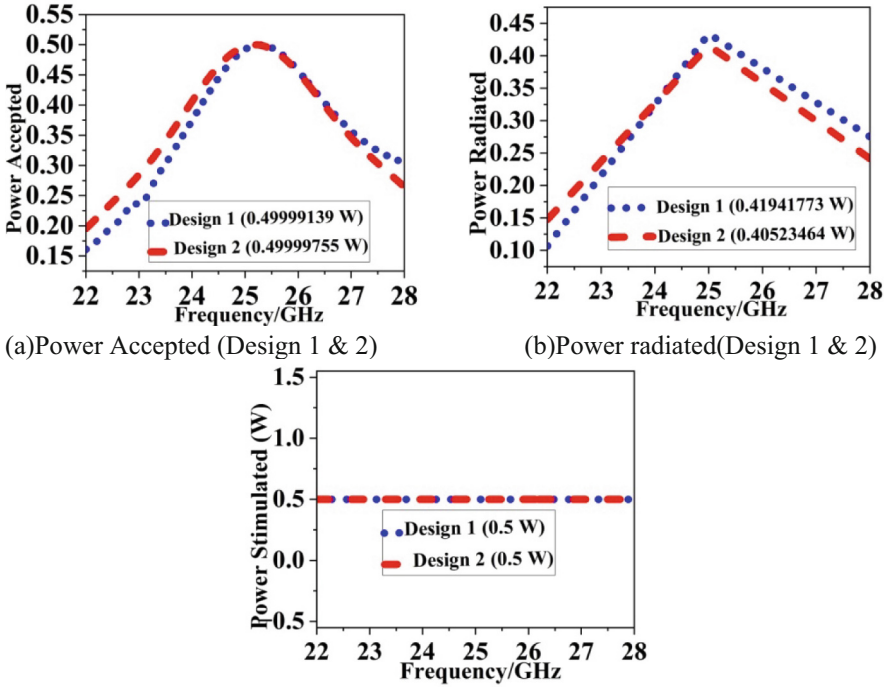


(a)Smith Chart (Design 1) (b)Smith Chart (Design 2)

Fig. 12. View of the smith chart of the designs 1 & 2

3.7 Accepted, Radiated and Stimulated Power

Figure 13 illustrates the accepted, radiated, and stimulated power of antenna designs 1 and 2. The power acceptance for designs 1 and 2 is 0.49999139 W and 0.49999755 W, respectively. The radiated power of designs 1 and 2 is 0.41941773 W and 0.40523464 W, respectively. The stimulating power of designs one and two is 0.5 watts.



(a)Power Accepted (Design 1 & 2) (b)Power radiated(Design 1 & 2)

(b)Power Stimulated (Design 1 & 2)

Fig. 13. View of the Power accepted, radiated and stimulated of the designs 1 & 2

Table 2. Presents the results of the designed simulation

Model	S_{11} (dB)	VSWR	Gain (dBi)	Directivity (dBi)	Efficiency (%)	IM (Ω)	Surface Current (A/m)	BW
Design-I	−47.66	1.008	7.13	7.67	92.95	48.69	173	1.519
Design-II	−53.10	1.004	6.29	7.06	90.93	49	170	1.656

Table 3. Comparative Analysis of Prior Research and Published Studies

Ref	Fr (GHz)	S_{11} (dB)	VSWR	Gain (dBi)	Efficiency (%)	BW (GHz)
[1]	Ka-Band	−19.23	1.3	6.08	85	–
[11]	Ka-Band	−28.39	1.08	7.04	85.34	–
[19]	Ka-Band	−17.16	1.32	6.4	–	1.24
[20]	Ka-Band	−20.95	1.197	7.5	–	1.06
[21]	Ka-Band	−34.5	0.59	9.01	78.3	–
[22]	Ka-Band	−14.81	1.5	6	85	0.900
[23]	Ka-Band	−27.26	–	7.10	–	1.4
[24]	Ka-Band	−35.91	1.032	9.42	–	1.43
[25]	Ka-Band	−25.75	1.190	7.5	90	–
[26]	Ka-Band	−42.59	–	4.16	70	–
[27]	Ka-Band	−38.348	1.024	–	77	–
[28]	Ka-Band	−24.507	1.126	7.19	–	1.352
This works	**Design-1**	**−47.66**	**1.008**	**7.13**	**92.95**	**1.519**
	Design-2	**−53.10**	**1.004**	**6.29**	**90.93**	**1.656**

4 Conclusion

This research study presents the design and optimization of two distinct 25 GHz patch antennas utilizing various substrate materials. CST software has been used to create the antennas, and the results of the two antennas have been compared. In terms of S_{11}, VSWR, bandwidth, gain, surface current, impedance matching, and radiation properties, the patch antenna designed for Ka-band wireless applications proves to be an excellent performer. The simulation results substantiate that the antenna satisfies the requisite parameters for Ka-band applications, hence confirming the accuracy of the modelling conclusions. This antenna design could serve as a feasible alternative for forthcoming wireless technologies. It offers the benefits of compact dimensions, superior efficiency, and resilient performance in demanding wireless communication environments. Additionally, antennas of different shapes can be designed by varying the materials used. The

fabrication of the antenna is a big challenge for the researcher. In the future, this antenna can be fabricated, and the simulation results can be compared with the actual results.

References

1. Stephane Borel, T.T., Priyadarshini, R.: U-slotted wideband microstrip patch antenna for Ka band and mmW 5G applications. Wirel. Pers. Commun. **135**, 1093–1102 (2024)
2. Harini, P., et al.: Design and analysis of an antenna with circular typed slots for 5G millimeter-wave applications. In: Millimeter Wave and Terahertz Devices for 5G and 6G systems, Signals and Communication Technology, pp. 73–85 (2025)
3. Faris, A.S., Tulu, M.M.: Design and performance analysis of multi-band antenna for 5G application. Wirel. Pers. Commun. **136**, 429–452 (2024)
4. Chidurala, S., Amara, P.: R.: design and analysis of compact dual band 4-port MIMO antenna for 5G 28/38 GHz millimeter-wave communication. J. Infrared Millim. Terahertz Waves. **46**, 60 (2025)
5. Shamim, S.M., Dina, U.S., Arafin, N., Sultana, S.: Design of Efficient 37 GHz Millimeter wave microstrip patch antenna for 5G Mobile application. Plasmonics. **16**, 1417–1425 (2021)
6. Wang, J., Li, Y., Wang, J.: A low-profile dual-mode slot-patch antenna for 5G Millimeter-wave applications. IEEE Antennas Wirel. Propag. Lett. **21**, 625–629 (2022)
7. Rao Devana, V.N.K., et al.: A high bandwidth dimension ratio compact super wide band-flower slotted microstrip patch antenna for millimeter wireless applications. Heliyon. **10**, e23712 (2024)
8. Poonia, A., Kikan, V., Kumar, A.: Design and analysis of rectangular patch antenna with varied slot configurations for Ka band applications. In: 2024 3rd Edition of IEEE Delhi Section Flagship Conference (DELCON), pp. 1–5. IEEE (2024)
9. Khouyaoui, I., Elbathaoui, M., Hamdaoui, M., Foshi, J.: Design, simulation, and fabrication of high-gain microstrip patch antenna arrays for 28 GHz millimeter-wave 5G applications. Phys. Scr. **100**, 045218 (2025)
10. Balanis, C.A.: Antenna Theory: Analysis Design, 3rd edn, Hoboken
11. Ojo, J.S., Omoleye, O.O., Faromika, O.P., Owolabi, G.A.: Design and analysis of improved rectangular-shape millimeter-wave microstrip patch antenna for 5 G and future generation wireless communication systems. Sci. African. **28**, e02676 (2025)
12. Praveena, A., et al.: Design and Substrate Analysis for Terahertz Antenna. in Millimeter Wave and Terahertz Devices for 5G and 6G systems, Signals and Communication Technology, pp. 515–525 (2025)
13. Acharya, R.: Noise, multipath and interference. In: Satellite Signal Propagation, Impairments and Mitigation 247–278, Elsevier (2017)
14. Stutzman, W.L.: Estimating directivity and gain of antennas. IEEE Antennas Propag. Mag. **40**, 7–11 (1998)
15. Ambika, A., et al.: Miniaturized Patch Antenna Design for Millimeter-Wave (37–40 GHz) Application. in Millimeter Wave and Terahertz Devices for 5G and 6G systems, Signals and Communication Technology, pp. 87–97 (2025)
16. Lacoste, R.: Impedance matching basics. In: Robert Lacoste's The Darker Side 3–16. Elsevier (2010)
17. Bensky, A.: Antennas and transmission lines. In: Short-Range Wireless Communication 43–83. Elsevier (2019)
18. Naishadham, K.: Transmission Lines. In: The Electrical Engineering Handbook 525–537. Elsevier (2005)

19. Singh, D., Choudhary, S.D., Mohapatra, B.: Design of microstrip patch antenna for Ka-band (26.5–40 GHz) applications. Mater. Today Proc. **45**, 2828–2832 (2021)
20. Didi, S.-E., et al.: Design of a microstrip antenna patch with a rectangular slot for 5G applications operating at 28 GHz. TELKOMNIKA (Telecommunication Comput. Electron. Control.). **20**, 527–536 (2022)
21. Shereen, M.K., Khattak, M.I., Zubir, F., Basit, A.: A novel single-feed hybrid reconfigurable microstrip patch antenna for 5G mobile communication and radio frequency energy harvesting applications at 28/38GHz. PLoS One. **17**, e0260407 (2022)
22. Ezzulddin, S.K., Hasan, S.O., Ameen, M.M.: Microstrip patch antenna design, simulation and fabrication for 5G applications. Simul. Model. Pract. Theory. **116**, 102497 (2022)
23. Nafea, S.N., Khamiss, N.N.: For 5G applications, high-gain patch antenna in Ka-Band. Indones. J. Electr. Eng. Comput. Sci. **31**, 802–809 (2023)
24. Didi, S.-E., et al.: New microstrip patch antenna array design at 28 GHz millimeter-wave for fifth-generation application. Int. J. Electr. Comput. Eng. **13**, 4184–4193 (2023)
25. Ayalew, L.G., Asmare, F.M.: Design and optimization of wideband rectangular-framed Pi-shaped mmWave antenna array for 5G applications. Int. J. Antennas Propag. **2024**, 1–15 (2024)
26. Saha, D., Verma, R.K., Dwivedi, R., Mandal, S., Purkait, K.: Design and optimisation of an R-shaped microstrip patch antenna using CSO and curve fitting for mmWave 5G applications. Int. J. Electron. **1–25** (2025)
27. Rana, M.S., Rahman Smieee, M.M.: Design and analysis of microstrip patch antenna for 5G wireless communication systems. Bull. Electr. Eng. Informatics. **11**, 3329–3337 (2022)
28. Rana, M.S., et al.: Design of 28 GHz microstrip patch antenna for wireless applications. Indones. J. Electr. Eng. Informatics. **11**, 967–977 (2023)



Achieving Net Zero: A Review of Greenhouse Gas Emission Forecasting Models and Approaches

Katya Victory Liu and Lina Gozali(✉)

Industrial Engineering, Universitas Tarumanagara, Jakarta, Indonesia
katya.545210016@stu.untar.ac.id, Linag@ft.untar.ac.id

Abstract. With the increasing urgency of global warming, mitigating greenhouse gases has become an essential agenda in many countries. These countries joined hands within the Paris Agreement and supported each other to limit the Earth's temperature to 1.5-2.0˚C. To achieve their net zero targets, looking into the future is crucial. By forecasting, countries know what kind of pathway they should take to reach their targets. This literature review reviewed forecasting models and approaches to project GHG emissions and achieve net zero. The reviewed studies mostly use machine learning and deep learning models to forecast GHG emissions. Machine learning and deep learning combined with traditional and multiple optimisations can achieve a more accurate model for predicting GHG emissions. More combinations and optimisation of machine learning models can reach more robust and precise predictions. However, to attain net-zero targets, forecasting should anticipate future emissions and consider mitigation measures. With the complexities of adaptation and mitigation strategies, the country must plan its actions well and soundly with the support of more in-depth research.

Keywords: greenhouse gas · forecasting · machine learning · net zero · mitigation

1 Introduction

Since 1850, the Earth's average surface temperature from 2010 to 2019 has been the warmest on Earth. Such global warming occurs due to the increased production of greenhouse gases in the atmosphere, which causes the greenhouse effect. The greenhouse effect is a phenomenon in which the sun's heat is trapped in the Earth's atmosphere, resulting in the inability of heat to be released into space [1]. This caused a direct climate change known as global warming.

The Earth's average temperature has increased by 1 °C since 1900. Along with the increase in temperature, climate change is becoming more extreme, there is a rise in sea level, and the ice cover in the northern hemisphere is decreasing [2]. Learning from the extreme climate changes that have occurred on Earth in the past, rapid climate change will make it difficult for humans and nature to adapt. The direct effect of such rapid change led to species extinction, population migration, and changes in land surface and

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026
S. Lanka et al. (Eds.): ICTSM 2025, LNNS 1755, pp. 498–508, 2026.
https://doi.org/10.1007/978-3-032-13177-5_39

ocean circulation. Coupled with food security, sea level rise, storms in coastal areas, and health problems directly resulting from climate change, many countries are experiencing economic damage [3].

The Earth produces greenhouse gases (GHG) that trap heat in the atmosphere. This phenomenon lets the Earth keep its temperature warm and stable for living beings. However, coupled with human activities that produce greenhouse gases, an imbalance occurs so that the Earth traps more heat, which causes global warming [4]. The climate becomes more unpredictable with the rise of the Earth's average temperature each decade. With such a significant impact, tackling global warming by reducing greenhouse gas emissions is the responsibility of every country.

The Intergovernmental Panel on Climate Change (IPCC) organised numerous conferences with representatives of countries to discuss the effects that continued greenhouse gas production would have on exacerbating warming and changing the global climate. The Paris Conference in December 2015 resulted in the Paris Agreement, which 195 countries agreed to. The agreement includes keeping the Earth's temperature rise below 1.5–2.0 °C [5].

Calculations by the IPCC suggest that over the next 100 years, the Earth's average temperature will rise by 6 °C due to continued use of fossil fuels and deforestation. With the global population expected to increase by 30% by 2050, countries, especially developing countries, will have difficulty coping with the effects of climate change due to global warming and other existing problems [6].

The Paris Agreement has started the rise of multiple countries announcing their support for the Paris Agreement to reduce their greenhouse gas emissions and reach net zero. The term began with the idea of not adding more greenhouse gases to the atmosphere because of human activities. Thus "zero" [7]. To reach net zero, a country must prepare capital and exert effort to mitigate and reduce GHG emissions. However, the capital of these countries, especially the developing ones, is limited. These countries with limited capital must find the most effective way to reach their net zero targets.

As worldwide awareness of the effects of climate change grows, many methodologies have been developed to assess and forecast future greenhouse gas (GHG) emissions. One of the most common techniques is to employ forecasting models based on historical data and present trends. These forecasting models are critical for providing quantifiable data that can be utilised to develop targeted emission reduction plans. In this context, realistic models can assist countries in developing environmental and energy policies that combine development requirements with climate protection.

GHG emission forecasting has used a variety of methodologies, including classical statistical methods like ARIMA and SARIMA, as well as newer ones like machine learning and deep learning. The advancement of computing technology and data availability has created more sophisticated and precise predictive models. Grey models, for example, can be used to make accurate predictions even with insufficient data. Each approach has advantages and disadvantages, depending on the complexity of the data, the extent of the field, and the study's aims.

Forecasting models can project the future. Policymakers can identify cost-effective and efficient pathways to reduce emissions without compromising economic stability by utilising forecasting models. These models analyse historical data, current trends, and

various influencing factors to predict future emission levels. These countries with limited capital could then plan their path to net zero in the most cost-effective way possible.

This review aims to evaluate existing forecasting models to forecast GHG emissions. Using comparative literature review, this paper focused on methods to predict GHG emissions, especially with the recent rise of machine learning and computer-based methods. This research not only identifies and addresses the research gaps but also identifies existing GHG emissions forecasting models' strengths, limitations, and applicability in guiding policy and decision-making toward achieving net-zero emissions in forecasting models used in this new era. This review serves as a resource for researchers and policymakers aiming to select or develop robust forecasting models that are technically sound and contextually relevant for achieving net-zero targets.

2 Methods

The method used in this study is a literature review. The literature review methodology consists of searching, analysing, and combining various literature, which is then systematically arranged. The literature review is chosen to identify the research gaps, limitations, and applicability of forecasting models for GHG emissions. This method is especially relevant given the broad methodological differences in existing research. A comparative review enables clear distinctions between articles, helping identify methodological trends, emerging patterns, and underexplored areas.

The ScienceDirect search engine and keywords, such as greenhouse gas emissions, net zero, forecast, and predictions, were used to find relevant articles. The search focus was on articles published between 2015 and 2025. This period was chosen to ensure that the articles reviewed were still applicable to the latest technological developments and were not outdated. From the initial search results, 12,123 articles were found.

After a filtering process based on specific selection criteria, eleven articles were selected, which were then analysed in more depth. The selection criteria include the relevance and quality of the article, the extent to which the article provides an in- depth analysis of the forecasting approach used, and whether the article develops or evaluates a model related to GHG emission predictions. This selection process aims to ensure that the articles reviewed can significantly contribute to understanding the latest trends and developments in greenhouse gas emission forecasting, especially those related to achieving the net zero target.

In the analysis process, each selected article is evaluated based on the forecasting approach used, the data analysis type, and each method's advantages and disadvantages. The focus is on how each study applies forecasting methods in different contexts, ranging from geographic regions to emission sectors, to prediction time spans. In addition, the analysis also includes how model accuracy is measured. The literature review describes the trend of GHG emission forecasting and identifies general patterns and research gaps that have not been widely explored.

3 Result and Discussions

The eleven articles chosen from the 12,123 articles can be seen in Table 1. The selected articles fit the selection criteria relevant to the research, provide a thorough analysis of the forecasting approaches, and assess models related to GHG predictions. The articles chosen cover a wide range of techniques from classical statistical models, machine learning, deep learning, to grey models and metaheuristic algorithms. Each article provides different insights depending on the region's context, the data type, and the forecasting objectives. Thus, the eleven selected articles provide a comprehensive overview of the latest greenhouse gas emission forecasting developments.

Table 1. Chosen Article

Year	Author	Journal
2019	Şahin, U [8].	Journal of Cleaner Production
2020	Alfaseeh, L., Tu, R., Farooq, B., & Hatzopoulou, M [9].	Transportation Research Part D Transport and Environment
2021	Xie, M., Yan, S., Wu, L., Liu, L., Bai, Y., Liu, L., & Tong, Y [10].	Journal of Cleaner Production
2022	Bakır, H., Ağbulut, Ü., Gürel, A. E., Yıldız, G., Güvenç, U., Soudagar, M. E. M., Hoang, A. T., Deepanraj, B., Saini, G., & Afzal, A [11].	Journal of Cleaner Production
2022	Franco, C., Melica, G., Treville, A., Baldi, M. G., Pisoni, E., Bertoldi, P., & Thiel, C [12].	Sustainable Cities and Society
2022	Javanmard, E. M., & Ghaderi, S. F [13].	Sustainable Cities and Society
2023	Lin, Z., Kuang, J., & Li, W [14].	Academic Journal of Environment & Earth Science
2023	Nanda, A. K., S, N., Gupta, S., Saleth, A. L. M., S, R., & Kiran, S [15].	Environmental Challenges
2023	Ozdemir, M., Pehlivan, S., & Melikoglu, M [16].	Energy Nexus
2024	Oliveira, G., Delgado, D., & Carvalho, M [17].	Energy and Buildings
2025	Zhang, J., & Zhao, L. [18]	Alexandria Engineering Journal

All eleven articles have different methods and approaches to forecast GHG emissions in other scopes. Sahin [8] and Ozdemir et al. [16] Lin et al. used Turkey as a study location to build a forecasting model for GHG emissions in Turkey [14]. and Zhang & Zhao [18] select global as their scope. Some articles use a specific country as their study location, like Bakir et al. [11] and Nanda et al. [15] in India, Javanmard & Ghaderi [13] in Iran, and Oliveira et al. [17] in Brazil. Xie et al. [10] pick EU countries as a study location.

Franco et al. [12], the scope is more interesting than others, as the research is done on cities committed to reaching net- zero emissions by 2020 and 2030.

After organising each article, it is evident that most studies use machine learning to build a forecasting model for GHG emissions. Nine articles out of eleven use at least one machine learning model as a comparison. Ozdemir et al. [16] and Oliveira et al. [17] use a different approach by using classic forecasting models, such as linear and logarithmic models used in [16], as well as classic time series models such as SARIMA, used in [17]. The rest of the articles integrate at least one machine learning model in their studies.

3.1 Machine Learning Models

Machine learning gains traction in forecasting methods because of its ability to forecast huge and complex data. Greenhouse gas emission fits the bill of having complex data with enormous resources. Machine learning can provide a framework that can be used over a long period because, with more data and training, the result will increase in accuracy. This condition leads to a more futuristic approach to GHG emission forecasting.

Franco et al. [12] utilised machine learning to run the data through several forecasting methods to find the best model with the minimum error value. The techniques used in machine learning are regressive integrated moving averages (ARIMA), double exponential smoothing (DES), and auto- and auto-regressive artificial neural networks (AR-ANN). Franco called the machine learning method Leave Last Known Value Out-Validation (LLKVO-V). The optimal model is chosen based on the lowest prediction error. Then, it is re-fitted to produce the corresponding GHG emissions estimation.

Machine learning combined with traditional forecasting methods could lead to a more accurate approach to each GHG emissions projection case. Machine learning techniques such as decision trees, neural networks, and ensemble learning can complement traditional models by reducing bias, handling complex data, and handling missing or noisy data more efficiently.

Alfaseeh et al. [9] use ARIMA, LSTM, and Clustering approaches to predict GHG emission rates by developing a deep learning model to improve intelligent transport systems in the hopes of mitigating GHG emissions. The study was conducted in Toronto, Canada, and a study area on a congested road network in downtown Toronto. The result indicated that the LSTM model is superior to ARIMA and Clustering models with a lower root mean square error (RMSE).

Lin et al. [14] used ARIMA and Support Vector Machine (SVM) models to compare the accuracy in predicting global temperature related to GHG emissions. In this study, ARIMA was proven superior to SVM, with lower Mean Squared Error (MSE) values and higher Explained Variance (EV) values. The main advantage of ARIMA lies in its ability to handle linear time series data efficiently, with a simple and easy-to-interpret model structure.

Machine learning models and techniques could be combined to provide a better and more accurate model, as done in Nanda et al. [15], which combines machine learning techniques, the Modified Coyote Optimisation Algorithm (MCOA) and Multi-Layer Perceptron's Neural Network (MPNN) to forecast India's GHG emissions. They test the MPPN-MCOA model with traditional methods like LSTM, ARIMA, KNN, RFNN, and CNN with different data sizes to find the best model to forecast GHG emissions. The

result is that with more data, the MPNN-MCOA model has a surprising accuracy above 90% with 50 data points and increased accuracy with more data available.

Zhang & Zhao [18] combine multiple deep learning models to forecast greenhouse gas emissions that have complex time series data. The best combination model is the CNN-BiLSTM-KNN model that combines Bidirectional Long Short- Term Memory (BiLSTM) networks, Convolutional Neural Networks (CNN), and Kolmogorov–Arnold Networks (KAN). This model combination produces the least error value and introduces a framework to model complex time series data like greenhouse gas emissions.

Looking at the machine learning models previously discussed, it can be concluded that machine learning methods, especially when combined with traditional forecasting models or optimised through metaheuristic algorithms, have great potential in improving the accuracy of greenhouse gas emission predictions. This approach can handle the complexity and large volume of data and is also adaptive to future data additions. Therefore, integrating machine learning is an important foundation in developing more sophisticated, precise, and future-oriented GHG emission forecasting methods.

3.2 Other Approaches

Machine learning and traditional methods aren't the only approaches to forecasting GHG emissions. While having a certain number of advantages, another approach can be taken to predict GHG emissions. Bakir et al. [11] provide a new approach using metaheuristic algorithms. The used algorithms include the lightning search algorithm (LSA), marine predators algorithm (MPA), backtracking search algorithm (BSA), equilibrium optimiser (EO), and symbiotic organisms search (SOS). The result shows that MPA outperforms other algorithms on accuracy based on MAPE.

Meanwhile, Javanmard & Ghaderi [13] proposed a hybrid model that combines various machine learning algorithms such as ANN, ARIMA, SARIMA, SARIMAX, LSTM, KNN, RF, and SVR. The results of these algorithms are then optimised with two metaheuristic approaches, Particle Swarm Optimisation (PSO) and Grey Wolf Optimiser (GWO). This hybrid strategy significantly improves prediction accuracy, even providing a minimum improvement of 31.7% compared to a single machine learning method. The main advantage of this approach lies in its ability to handle nonlinear and multivariate relationships in complex and limited data.

Sahin [8] uses the nonlinear metabolic grey model (NMGM), metabolic grey model (MGM), and their optimised model (OMGM and ONMGM) to forecast greenhouse gas emissions in Turkey until 2025. The forecasting result is compared by its mean absolute standard error (MAPE). The optimised model of the nonlinear metabolic grey model (ONMGM) surpassed the rest of the models with the lowest MAPE value.

Then, Xie et al. [10] improve the grey model with a weighted regularised multivariate grey model (RWGM) that outperforms other models, such as the non- weighted regularised multivariate grey model, the grey model, and robust linear regression in terms of accuracy. RWGM has the lowest MAPE value and stronger adaptability, making the model more stable than GM and RLR.

Alternative approaches, such as grey models and metaheuristic algorithms, have also contributed to greenhouse gas emission forecasting. Although different from machine

learning methods and traditional models, these approaches have proven to provide accurate prediction results with low error rates, as demonstrated by the success of MPA, ONMGM, and RWGM in various studies. The main advantage of these approaches lies in their ability to handle limited, incomplete, or uncertain data common in GHG emission data.

3.3 Achieving Net Zero

The undeniable factor throughout all nine research studies is the importance of accurate forecasting for achieving net-zero targets. Ozdemir et al. [16] use a linear and logarithmic model to forecast Turkey's GHG emissions until 2030 to find that Turkey would not meet the target disclosed by 2030 following the Paris Agreement. Turkey should comprehensively reflect on its mitigation strategies and rerun the forecast to reach the disclosed targets soon.

Mitigation techniques include using a renewable resource for electricity, implementing a carbon tax, and using more environmentally friendly vehicles [17]. Studies [8, 10, 11, 16, 17] mention the importance of reducing fossil fuel energy sources and moving to renewable sources to reduce GHG emissions. Achieving net zero is a hard target without reducing fossil fuel energy sources. Clean electricity is essential to ensure energy consumed doesn't contribute to GHG emissions, as the most significant contributor to GHG emissions is energy production and consumption from non-renewable sources.

Franco et al. [12] give an insight into how vital mitigation is compared to adapting to climate change. By doing mitigation alone, a city has a 30% success rate in reaching its GHG emission reduction targets. Doing a combination of mitigation and adaptation strategies increases the success rate in the long run by 2% compared to just doing mitigation.

Accurate forecasting is an analytical tool and an essential foundation in formulating effective mitigation strategies. The achievement of the net zero aim heavily relies on the extent to which prediction results are used to develop a concrete and quantifiable strategy. Integrating forecasting results with mitigation measures such as clean energy transition, energy efficiency, and cross-sectoral emission control becomes crucial in this context. When predictions show that a country or city is not on track, as indicated by the study by Ozdemir et al. [16], then strategy revision needs to be done immediately with a more progressive and data-driven approach. Thus, the synergy between forecasting and policy is key to accelerating the sustainable achievement of the net-zero target.

3.4 Literature Comparison

Table 2 shows the comparison of eleven chosen articles. The table shows the different approaches used by each article to predict GHG emissions.

Based on Table 3, the forecasting method mostly leans toward machine learning and deep learning models. Most studies use more than one approach to find the best model [16]. The SARIMA model is used only for forecasting, but it runs on multiple parameters and scenarios before being compared based on their symmetric mean absolute percentage error (sMAPE) [12]. uses machine learning to find the best-fitting approaches based on the indicated scenario.

Table 2. Literature Comparison

Description	Authors					
	Sahin [8]	Alfaseeh et al. [9]	Xie et al. [10]	Bakir et al. [11]	Franco et al. [12]	Javanmard et al. [13]
GHG Emissions	✓	✓	✓	✓	✓	✓
Comparative study	✓	✓	✓	✓	✓	✓
Forecasting Model	MGM, OMGM, NMGM, ONMGM	ARIMA, LSTM, Clustering	RWGM, RGM, GM, RLR	MPA, LSA, EO, SOS, BSA	DES, AR-ANN, ML	ANN, AR, ARIMA, SARIMA, SARIMAX, RF, SVR, KNN, LSTM
Best Model	ONMGM	LSTM	RWGM	MPA	ML	PSO- hybrid model
Developing New Model	✓	X	✓	X	✓	✓

MGM = Metabolic gray model, NMGM = Non-linear metabolic grey model, OMGM/ONMGM = Optimized MGM/Optimized NMGM, ARIMA = Autoregressive integrated moving average, LSTM = Long Short-term Memory, RGM = Regularized multivariate grey model, RWGM = Weighted regularized multivariate grey model, RLR = Robust linear regression, MPA = marine predators algorithm, LSA = Lightning search algorithm, EO = Equilibrium optimizer, SOS = Symbiotic organisms search (SOS), BSA = Backtracking search algorithm, DES = Double exponential smoothing, AR-ANN = Auto-regressive artificial neural networks, ML = Machine learning, ANN = Artificial neural networks, SARIMA = Seasonal ARIMA, SARIMAX = Seasonal ARIMA with eXogenous variable, RF = Random forest, SVR = Support vector regression, KNN = K-nearest neighbor, PSO = Particle swarm optimization.

Available data also contributes to the chosen forecasting model. Studies with less data, like [8, 10], select a grey model to forecast GHG emissions. Grey's model can use a few historical data points to generate unknown data; hence, the GM can be more flexible in generating incomplete or minimal data. This model brings new insights for countries with limited data availability, such as developing countries. Its minimal data demand also minimises reliance on long-term datasets, which may not be available in all locations.

On the traditional side, ARIMA and LSTM mostly lose accuracy compared to machine learning or deep learning approaches in studies [15, 18]. ARIMA and LSTM have lower accuracy in forecasting GHG emissions than newer approaches. Compared to the traditional techniques, LSTM wins in accuracy compared to ARIMA and clustering methods, as studied by Alfaseeh et al. [9]. In this study, LSTM is run with two hidden layers tuned using Bayesian optimisation, thus outperforming the clustering and ARIMA models.

Table 3. (Cont'd) Literature Comparison

Description	Authors				
	Lin et al. [14]	Nanda et al. [15]	Ozdemir et al. [16]	Oliveira et al. [17]	Zhang & Zhao [18]
GHG Emissions	X	✓	✓	✓	✓
Comparative study	✓	✓	X	X	✓
Forecasting Model	ARIMA, SVM	MPNN-MCOA, LSTM, KNN, CNN, RFNN, ARIMA	Linear and Logarithmic model	SARIMA	CNN-BiLSTM-KAN, LSTM, GRU, VMD, EMD, ARIMA
Best Model	ARIMA	MPNN-MCOA			CNN-BiLSTM-KAN
Developing New Model	X	✓	X	X	✓

ARIMA = Autoregressive integrated moving average, SVM = Support vector machines, MPPN = Multi-layer perceptron's neural network, MCOA = Modified coyote optimization algorithm, LSTM = Long Short-term Memory, KNN = K-nearest neighbor, CNN = Convolutional neural network, RFNN = Radial basis neural network, SARIMA = Seasonal ARIMA, BiLSTM = Bidirectional LSTM, KAN = Kolmogorov–Arnold networks, GRU = Gated recurrent unit, VMD = Variational mode decomposition, EMD = Empirical mode decomposition.

These studies show the growing trend of improving GHG emission projections by combining machine learning with conventional forecasting methods. The findings repeatedly demonstrate that hybrid techniques perform better than solo models, especially when dealing with intricate time series data. It is anticipated that as machine learning advances, its use in GHG emission forecasting will improve prediction models even more, providing insightful information to decision-makers and assisting in developing more potent climate mitigation plans.

All eleven studies focus on historical emission data and technical optimisation, but they lack integration with external factors such as policy changes, economic growth, industrial shifts, or renewable energy adoption. External variables integrated in most studies only include electricity generation and population, such as in [11, 13, 15, 16]. Out of eleven studies, only one calculated the greenhouse gas emissions separately [17], and only one forecasted the GHG emissions in several sectors such as energy, industrial processes, and agriculture [16].

Despite differences in methods used by all articles reviewed, there is a clear research gap: the lack of real-time adaptive models and the lack of integrations of external variables such as policy, economic, and sectoral variables in most models. None of the forecasting models used relate to real-time or adaptive models that can adjust predictions based on new data and evolving policies. None of the research explores how these external influences can be incorporated into forecasting models to improve real-world

applicability. Because of the limited scope of data, the forecasting model used can produce a different result because it isn't tailored to specific geographical areas, such as a less developed country.

4 Conclusion

While machine learning approaches are versatile and robust with large amounts of data, such as GHG emissions, the forecasting accuracy improves significantly when combined with optimisation techniques. Hybrid systems that mix machine learning with classic statistical models have shown improved accuracy, implying that integrating the two approaches leads to more reliable predictions. Even with limited data, grey models can be used to forecast GHG emissions. However, significant study gaps should be addressed in future research. All the research evaluated continues to focus on technical historical data and model optimisation without considering external factors such as government regulations, economic growth, industry shifts, or renewable energy uptake. Only one study calculates GHG emissions independently per gas, and only one divides emission forecasts by sector. Furthermore, no research uses real-time adaptive models capable of dynamically adjusting estimates in response to new data or policy changes. The failure to adapt to specific geographical conditions, such as impoverished countries, is another constraint that might reduce the usefulness of model results to real-world contexts. In conclusion, the literature review shows that the methods used are imperfect. To achieve net-zero targets, forecasting should anticipate future emissions and take mitigation measures into account. Studies have shown that combining mitigation and adaptation techniques improves the chances of successfully decreasing emissions. This research compared GHG emissions forecasting models ranging from classical ARIMA to hybrid machine learning and metaheuristic optimisation methods, and its critical analysis of their applicability across national and regional contexts. The research gaps found include the lack of real-time adaptability and the insufficient integration of socio-economic and policy variables into forecasting models. As countries navigate their paths to net zero, such insights are essential to building predictive tools that are precise and aligned with dynamic real-world conditions.

References

1. National Academy of Sciences: Climate Change: Evidence and Causes: Update 2020. The National Academies Press, Washington, DC (2020)
2. Lindsey, R. & Dahlman, L. Climate change: Global temperature. Climate.gov 16, (2020).
3. Yoro, K.O., Daramola, M.O.: CO2 emission sources, greenhouse gases, and the global warming effect. In: Advances in Carbon Capture: Methods, Technologies and Applications, pp. 3–28. Elsevier (2020). https://doi.org/10.1016/B978-0-12-819657-1.00001-3
4. AguirBargaoui, S., Nouri, F.Z.: History of the human and nature relationship, discovery of greenhouse effect and awareness of the environmental problem. J. Econ. Sci. Res. **4** (2021)
5. United Nation Climate Change. The Paris Agreement. https://unfccc.int/process-and-meetings/the-paris-agreement (2020)
6. Mikhaylov, A., Moiseev, N., Aleshin, K., Burkhardt, T.: Global climate change and greenhouse effect. Entrepreneur. Sustain. Issues. **7**, 2897–2913 (2020)

7. Rogelj, J., Geden, O., Cowie, A., Reisinger, A.: Three ways to improve net zero. Nature. **591** (2021)
8. Şahin, U.: Forecasting of Turkey's greenhouse gas emissions using linear and nonlinear rolling metabolic grey model based on optimization. J. Clean. Prod. **239**, 118079 (2019)
9. Alfaseeh, L., Tu, R., Farooq, B., Hatzopoulou, M.: Greenhouse gas emission prediction on road network using deep sequence learning. Transp. Res. D Transp. Environ. **88**, 102593 (2020)
10. Xie, M., Yan, S., Wu, L., Liu, L., Bai, Y., Liu, L., Tong, Y.: A novel robust reweighted multivariate grey model for forecasting the greenhouse gas emissions. J. Clean. Prod. **292**, 126001 (2021)
11. Bakır, H., Ağbulut, Ü., Gürel, A.E., Yıldız, G., Güvenç, U., Soudagar, M.E.M., Hoang, A.T., Deepanraj, B., Saini, G., Afzal, A.: Forecasting of future greenhouse gas emission trajectory for India using energy and economic indexes with various metaheuristic algorithms. J. Clean. Prod. **360**, 131946 (2022)
12. Franco, C., Melica, G., Treville, A., Baldi, M.G., Pisoni, E., Bertoldi, P., Thiel, C.: Prediction of greenhouse gas emissions for cities and local municipalities monitoring their advances to mitigate and adapt to climate change. Sustain. Cities Soc. **86**, 104114 (2022)
13. Javanmard, E.M., Ghaderi, S.F.: A hybrid model with applying machine learning algorithms and optimization model to forecast greenhouse gas emissions with energy market data. Sustain. Cities Soc. **82** (2022)
14. Lin, Z., Kuang, J., Li, W.: Predictions and research about global warming based on ARIMA models. Acad. J. Environ. Earth Sci. **5**, 42–48 (2023)
15. Nanda, A.K., Gupta, S., Saleth, A.L.M., Kiran, S.: Multi-layer perceptron's neural network with optimization algorithm for greenhouse gas forecasting systems. Environ. Challenges. **11**, 100708 (2023)
16. Ozdemir, M., Pehlivan, S., Melikoglu, M.: Estimation of greenhouse gas emissions using linear and logarithmic models: a scenario-based approach for Turkiye's 2030 vision. Energy Nexus. **13**, 100264 (2024)
17. Oliveira, G.P., Delgado, D.B.M., Carvalho, M.: Analysis and forecast of emissions associated with electricity consumption in hospitals: encouraging a net zero healthcare roadmap in Brazil. Energy Build. **324**, 114785 (2024)
18. Zhang, J., Zhao, L.: Efficient greenhouse gas prediction using IoT data streams and a CNN-BiLSTM-KAN model. Alex. Eng. J. (2025). https://doi.org/10.1016/J.AEJ.2025.01.019

A Machine Learning Approach to Crude Oil Price Prediction Using Support Vector Machine (SVM)

Hitesh Punjabi[1], S. Kumar Chandar[2(✉)], and Mayur Malik[3]

[1] Mumbai, Maharashtra, India
hitesh.punjabi@somaiya.edu
[2] School of Business and Management, Christ University, Bengaluru, Karnataka, India
kumar.chandar@christuniversity.in
[3] Research Scholar, International Open University, Banjul, The Gambia

Abstract. Crude oil is one of the most important energy sources, and fluctuations in its international prices affect all aspects of the economy. The price of crude oil is influenced by several variables, and the length of time that each component has an effect differs giving an increase in non-linear oil price features. Although it is a complex task, identifying the most essential factor influencing for precise predicting, crude oil prices are essential. Therefore, this study aims to employ a machine learning model to address the intricate relationships among different factors. Primarily, it gathers data regarding West Texas Intermediate (WTI) and Brent crude oil prices as well as macroeconomic variables. Secondly, the data is normalized to prepare it for further analysis. Finally, a crude oil prediction model is constructed using Support Vector Machine (SVM) to predict future international crude oil prices. The daily, weekly, and monthly prices are used to confirm the model's efficacy developed using WTI and Brent oil. The model's performance is also evaluated by incorporating various combinations of macroeconomic variables to find the most influential factor. Results from experiments show indicates the benchmark model was much exceeded by the developed model and performed very well in terms of prediction accuracy. The findings reveal that selecting the appropriate variables can greatly enhance prediction accuracy. This model has the potential to provide valuable insights for traders, investors, and energy-related enterprises, offering beneficial guidance for decision-making purposes.

Keywords: Crude oil · energy market · machine learning · macroeconomic variables · normalization · support vector machine

1 Introduction

Crude oil is the most important natural and non-renewable energy that has a prodigious effect on global economy and international financial markets. Inflation, social stability, national security, and the economic status of many nations worldwide may all be impacted by sudden drops in oil prices. Crude oil is in greater demand recently in a variety of industries due to increasing countries' high energy consumption. Consequently,

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026
S. Lanka et al. (Eds.): ICTSM 2025, LNNS 1755, pp. 509–526, 2026.
https://doi.org/10.1007/978-3-032-13177-5_40

the problem of the depletion of crude oil products has become more severe, drawing more attention from energy-related enterprises, governments, organizations, and traders [1, 2]. Nonetheless, the state of politics is one of the numerous variables the price of crude oil is significantly impacted by, oil production, demand, regional military conflicts, and natural disasters, etc. [3]. Crude oil prices are not constant; this makes it very challenging to make reliable crude oil price estimates, which is the subject of present research.

On a global scale there are two main markets for crude oil: (i) Markets in the United States and other countries utilize the New York Mercantile Exchange's West Texas Intermediate (WTI), and (ii) Brent oil market from International Petroleum Exchange (IPE). Given that the two most significant oil prices worldwide are WTI and Brent, price prediction has drawn tremendous attention to global researches. Multivariate and univariate models are two types of crude oil price forecasting methods. In addition to historical data, the multivariate model takes into account supply and demand, inventories, the economy, and financial factors all of which having an effect on crude oil prices.

Univariate model only includes past prices. Econometric, Machine Learning (ML), and hybrid models are the three categories into which the univariate models reside. The Autoregressive Moving Average (ARMA) is a component of econometric models [4], Generalized Autoregressive Heteroscedasticity (GARCH) [5], and Vector Autoregressive Regression (VAR) [6], etc. Although econometric models are useful for capturing characteristics of linear data, the non-linear features of oil prices are not well represented. Thus, ML models are suggested to solve this issue. This paper uses ML model to anticipate crude oil prices. The novelty of this study lies in developing a SVM based prediction model for crude oil prices that does not rely on supply-demand factors. To the best of authors' knowledge, this is the first attempt to explore crude oil prediction from this perspective.

The key contribution of this paper is the development of a crude oil price prediction system using Support Vector Machine (SVM). The model relies on macroeconomic variables rather than supply-demand factors. This model offers a new view for forecasting crude oil dynamics, as it highlights the macro economical linkages influencing oil prices. The following are this paper's scientific contributions:

- A crude oil prediction model using SVM is built.
- Various macroeconomic variables such as Gold Price (GP), Silver Price (SP), S&P 500 index, EUR/USD exchange rate, Natural Gas Price (NGP), Dollar Index (DI), and GBP/USD exchange rare are used as input variables. This is the first attempt to predict crude oil price without considering supply and demand factors.
- WTI and Brent oil prices are used to assess the designed model's efficacy.
- Scalability of the model is analysed using the daily, weekly, and monthly prices of WTI and Brent.
- Performance of the model is examined utilizing various combinations of macroeconomic variables to find the most influential variable.

The paper is outlined as follows: Section 2 presents review of earlier works. Section 3 explains the introduced crude oil prediction model. Section 4 presents experimental outcomes in detail. Section 5 summarizes the paper.

2 Related Works

Numerous methods for crude oil price prediction have been developed. The first class of econometric models are ARIMA, GARCH, and VAR. ARIMA and GARCH were used by Moshiri and Foroutan [1] for predicting WTI crude oil prices. Using autoregressive conditioned heteroscedasticity, Cheong et al. [2] developed a crude oil forecasting model for WTI and Brent prices. Abledu and Agbodah [3] predicted Ghanaian crude oil prices using the ARIMA model. The dynamic, non-stationary, and non-linear character if the crude oil price contrasts with the linearity assumed by econometric models, as was discussed previously in this work.

Due to the limitations in forecasting abilities observed in econometric models, the demand for research on predictive models based on machine learning is consistently increasing. Popular ML models are Artificial Neural Network (ANN), SVM, Wavelet Neural Network (WNN), and Extreme Learning Machine (ELM). Ahmed and Shabri [4] introduced a new method for daily crude oil price prediction by employing SVM. Mostafa et al. [5] examined the appropriateness of genetic expression programming and ANNs in forecasting crude oil prices. Yu et al. [6] compared the prediction abilities of ARIMA, SVM, and ANN. Huang et al. [7] modelled a predictive model using WNN to forecast the price of crude oil. Herrera et al. [8] developed a prediction model to anticipate crude oil prices using ANN and random forests, and demonstrated the model's efficacy in comparison to econometric techniques.

Cen and Wang [9] suggested a method for using Long Short-Term Memory (LSTM) to estimate the prices of Brent and WTI crude oil. A hybrid model was introduced by Abdollahi and Ebrahini [10] to make daily price predictions for Brent crude oil. Fractionally Integrated Moving Average with Auto-regression (ARFIMA), Markov switching model, and Adaptive Neuro Fuzzy Interference System (ANFIS) were combined to produce a hybrid model. A novel model based on Wavelet Packet Decomposition-LSTM (WPD-LSTM) was developed by Wang and Wang [11] and used to forecast crude oil prices. Zhang et al. [12] adopted Variational Mode Decomposition-kernel ELM (VMD-KELM) to predict the price of crude oil. Yan et al. [13] examined the potential effects of commodities and financial characteristics on WTI crude oil prices. Locally linear embedding technique and LSTM were used for dimension reduction and price prediction respectively. Daneshvar et al. [14] used LSTM and BiLSTM to forecast Brent crude oil prices. A summary of reviewed papers is given in Table. 1.

Table 1. A summary of recent studies on crude oil prediction

Contributors	Year	Data	Inputs	Predictor	Metrics
Cen and Wang [9]	2019	WTI Brent	Open, close, high, low	LSTM	MAE, RMSE, MAPE, SMAPE,
Abdollahi and Ebrahimi [10]	2020	Brent	Closing price	Hybrid model	RMSE, MAE, MAPE
Wang and Wang [11]	2021	WTI, Brent, heating oil	Closing price	WPD-LSTM	MAE, RMSE, MAPE, SMAPE
Zhang et al. [12]	2021	WTI, Brent	6 previous days' closing price	(VMD-K ELM)	RMSE, MAE, MAPE,
Yan et al. [13]	2021	WTI	Supply, demand, gas, exchange rate, Gold, and Silver	LSTM	MAE, RMSE
Daneshwar et al. [14]	2022	Brent	Commodity and financial factors	LSTM, BiLSTM	RMSE MSE

3 Proposed Crude Oil Prediction System

The overall structure of the designed crude oil prediction model is given in Fig. 1. The model has four main processes. Data preparation is the first stage. Training and testing data must be collected to develop a crude oil prediction model. However, the data gathered from different sources must be chosen according to corresponding criteria. Preparing the data is the second stage. Data preparation and data normalization are its two operations. Understanding the given data is almost essential to the construction of any prediction model. The performance of SVM-based prediction models may be significantly impacted by data standardization. Following normalization, the results gathered have to be divided into two groups: out-of-sample data and in-sample (training) data, which are used for model creation and assessment, respectively. the third step in training an SVM model. The SVM architecture selection, data training, and validation are its three main steps. In this stage, the kernel function and regularization parameter are established. It is an out-of-sample prediction technique. The trained SVM model may be utilized as a crude oil price out-of-sample predictor after the three aforementioned procedures are finished.

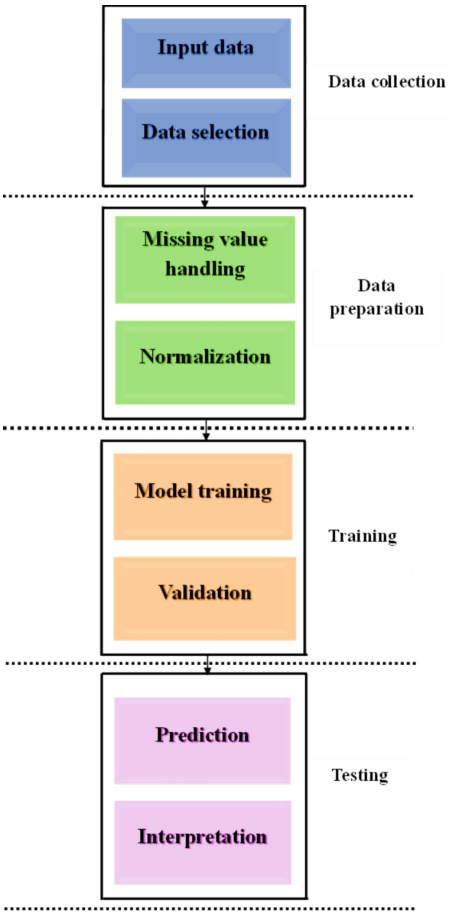


Fig. 1. Overall structure of the proposed model

3.1 Data Gathering

WTI and Brent oil prices are two examples of the variety of data that may be gathered for this study. There are three time frames into which the collected data may be separated: daily, monthly, and weekly. The U.S. Energy Information Administration (EIA) provided the data [15, 16]. Brent and WTI data are available from January 1, 2001, until February 28, 2023. The data contains crude oil prices on a daily, weekly, and monthly basis, excluding weekends and public holidays. Figure 2 provides a visual representation of the Brent and WTI crude oil prices on a daily, weekly, and monthly basis. The Brent and WTI data descriptive statistics are shown in Table 2. Notably, there is a considerable difference between the highest and minimal values. The standard deviation (Std.dev) for both datasets indicates substantial price fluctuations. Various macroeconomic variables such as GP, SP [13], S&P 500 index [19], EUR/USD exchange rate [18], NGP, DI, and GBP/USD exchange rare [20] are used as input variables.

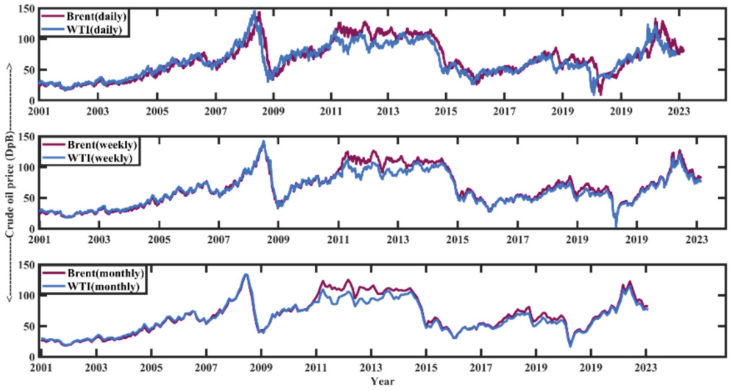


Fig. 2. WTI and Brent crude oil prices from 2001 to 2023

Table 2. Descriptive statistics of WTI and Brent Crude oil prices

Parameters	WTI crude oil price			Brent crude oil price		
	D	W	M	D	W	M
No. of samples	5629	1156	266	5629	1156	266
Minimum	8.91	16.55	3.32	9.12	14.24	18.38
Maximum	145.3	133.8	142.5	143.9	141.1	132.7
Mean	64.14	63.61	63.95	66.97	67.06	66.68
Std. dev.	25.71	25.74	25.71	29.13	29.22	29.23
Variance	0.035	0.034	0.048	0.046	0.053	0.065
Skewness	0.280	0.277	0.269	0.297	0.288	0.282

D-Daily, W-Weekly, M-Monthly

3.2 Data Preparation

For the SVM, the data collected may need to be converted into a certain range. The market has been paused or closed because of unforeseen occurrences or weekends, and there are a number of inconsistencies and missing points in the daily statistics. In this investigation, mean of the neighbors data is employed to handle the missing value. A range of 0 and 1 is used to standardize the data by using min-max method. Normalization can be expressed as,

$$y_{norm} = \frac{y - y_{min}}{y_{max} - y_{min}} \tag{1}$$

Where, y is the actual price, y_{norm} represents the average price, while min and max represent the lowest and highest costs, respectively. Normalized crude oil prices are

illustrated in Fig. 3. After preprocessing, 10-cross fold validation is used for developing and evaluating the model. Cross-validation method is combined with the SVM to boost performance, ensure accurate prediction, and prevent overfitting.

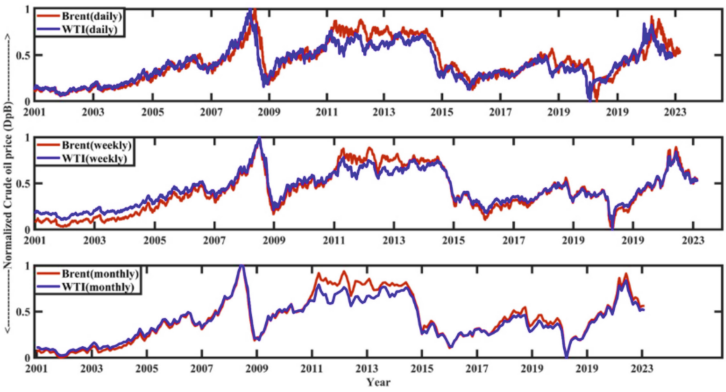


Fig. 3. Normalized crude oil prices

3.3 Model Training

SVM is a class of supervised classification model, introduced by Vapnik [17], which can be employed for regression and classification problems. In this study, general smoothness assumptions, the Radial Basis Function (RBF) kernel is used since it tends to provide superior results. If more details on the data is available, it is really helpful. During training, regularization parameter and suitable kernel function show be determined. Because there are no common standards for determining the parameters, they are often determined by experimentation. To ensure that SVM is generalizable, model validation must be done after training. A predictor with ideal parameters may be acquired after validation. In this investigation, SVM is chosen due to its ability to handle high-dimensional data, making it suitable for crude oil price prediction.

3.4 Prediction

The out-of-sample prediction may be done using the learned SVM.

4 Empirical Study

The experimental data are shown numerically in this section. The introduced crude oil prediction system has been implemented and validated on MATLAB 2022a.

4.1 Performance Metrics

Root Mean Square Error (RMSE), Mean Absolute Percent Error (MAPE), and Prediction Accuracy (PA) [21–23] are a collection of metrics used to emphasize the prediction outcomes of the SVM-designed model.

$$RMSE = \sqrt{\frac{1}{M}\sum_{j=1}^{M}\left(A_j - P_j\right)^2} \tag{2}$$

$$MAPE = \frac{1}{M}\sum_{j=1}^{M}\left|\frac{A_j - P_j}{A_j}\right| \tag{3}$$

$$\text{PA} = \frac{\text{Number of correctly predicted value}}{\text{Total number of samples}} \times 100 \tag{4}$$

4.2 Experimental Treatment

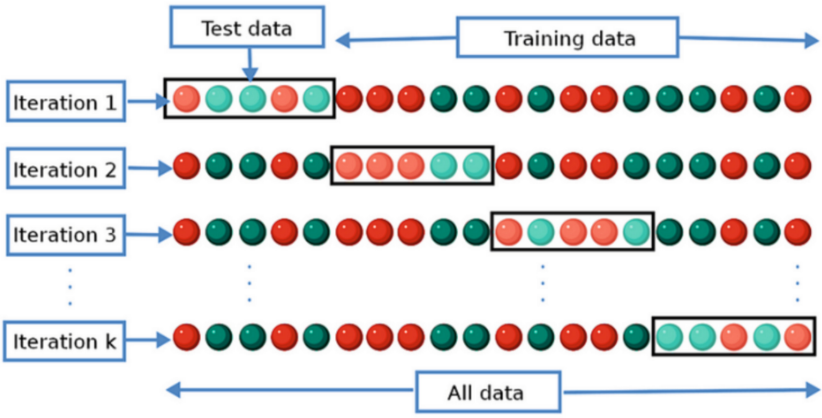


Fig. 4. 10-fold cross validation

The price of crude oil and macroeconomic factors are directly related. The crude oil prediction model that is presented in this research is based on SVM and macroeconomic data. Two standards for the WTI and Brent pricing for crude oil are taken into account. WTI and Brent oil prices, GP, SP, NGP, EUR/USD, S&P500, DI, and GBP/USD are among the variables selected for this study based on an analysis of the variables affecting

the price of crude oil. The period of collecting data was January 1, 2001, through February 28, 2023. Primarily, the data were pre-processed to make them to a common range between 0 and 1. To evaluate the prediction model's overall effectiveness, 10 cross-fold validation was considered, as shown in Fig. 4. In this approach, entire data was divided into 10-folds. During first iteration, fist fold was used as test set and remaining folds were as training sets. In the second iteration, second set was used as test set while all other folds were considered as training sets. This procedure was repeated 10 times. Final error value was computed by taking mean of all these individual iterations.

Results on Brent Oil Price Prediction

Table 3. Various combinations of macroeconomic variables used in this investigation

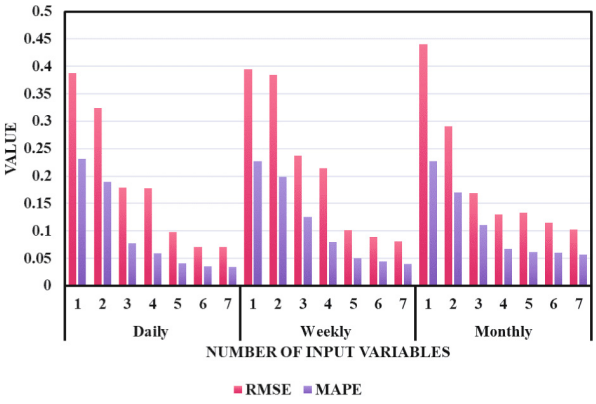
Number of influential factors	Data selection	Exceptional influential factor
1	GP	SP, S&P 500, EUR/USD, NGP, DI, GBP/USD
2	GP, SP	S&P 500, EUR/USD, NGP, DI, GBP/USD
3	GP, SP, S&P 500	EUR/USD, NGP, DI, GBP/USD
4	GP, SP, S&P 500, EUR/USD	NGP, DI, GBP/USD
5	GP, SP, S&P 500, EUR/USD, NGP	DI, GBP/USD
6	GP, SP, S&P 500, EUR/USD, NGP, DI	GBP/USD
7	GP, SP, S&P 500, EUR/USD, NGP, DI, GBP/USD	N/A

To assess the performance of the developed model, different combinations of macroeconomic factors are used. The specific combinations employed are listed in Table 3. Efficacy of the introduced prediction model in forecasting Brent oil prices on a daily, weekly, and monthly basis using different combinations of macroeconomic factors reported in Table 4. A visualization plot of the predicted results is shown in Fig. 5. A comparison of the empirical results from daily, weekly, and monthly data might lead to the following conclusions. Comprehensive study of daily prediction data results demonstrates that the developed model's prediction performance with seven input variables is better than the model using other input combinations. The actual crude oil prices are highly complex, causing significant discrepancies between the predicted values of model with a single

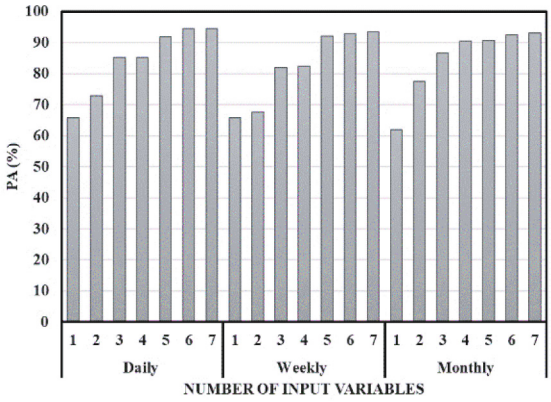
input and the observed prices, resulting in limited PA. For instance, RMSE values for the introduced model incorporating input variables 1, 2, 3, 4, 5, 6, and 7 are 0.388, 0.324, 0.179, 0.178, 0.098, 0.071, and 0.071, respectively. It is observed that the model attains lower prediction error when seven inputs are used. The introduced model achieved lowest PA of 65.74% when using only the GP input, whereas it achieves the highest PA of 94.4% when all seven input variables are used. When comparing the model with other input combinations, the prediction model with seven inputs produces excellent results. Like this, the model with seven input variables yielded higher PA of 93.95% for weekly, and 93.10% for monthly data. The results demonstrate that the selecting appropriate macroeconomic the model's PA may be improved by using variables to efficiently simplify the crude oil prices. Figures 6, 7, and 8 compare the present and anticipated prices of Brent crude oil for daily, weekly, and monthly predictions.

Table 4. Brent crude oil price prediction results

Data frequency	Number of influential factors	RMSE	MAPE	PA (%)
Daily	1	0.388	0.231	65.74
	2	0.324	0.189	72.92
	3	0.179	0.077	85.23
	4	0.178	0.059	85.12
	5	0.098	0.041	91.88
	6	0.071	0.035	94.41
	7	0.071	0.034	94.42
Weekly	1	0.395	0.227	65.74
	2	0.384	0.198	67.67
	3	0.237	0.125	81.87
	4	0.214	0.080	82.31
	5	0.101	0.050	92.03
	6	0.089	0.044	92.95
	7	0.081	0.040	93.56
Monthly	1	0.440	0.227	61.90
	2	0.291	0.170	77.47
	3	0.169	0.111	86.65
	4	0.130	0.067	90.46
	5	0.133	0.062	90.74
	6	0.115	0.060	92.40
	7	0.103	0.057	93.10



(a)



(b)

Fig. 5. Performance of the introduced model for Brent oil prices (A) Error metrics (B) PA

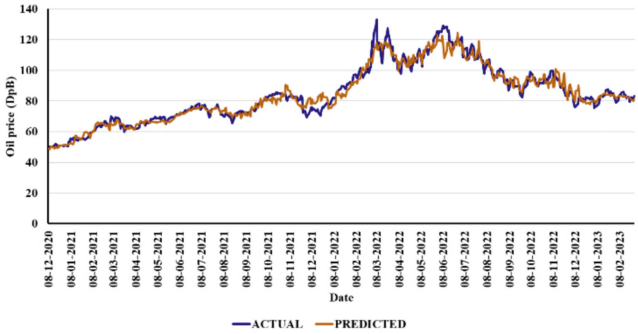


Fig. 6. Actual Vs. Predicted values of Brent daily crude oil prices

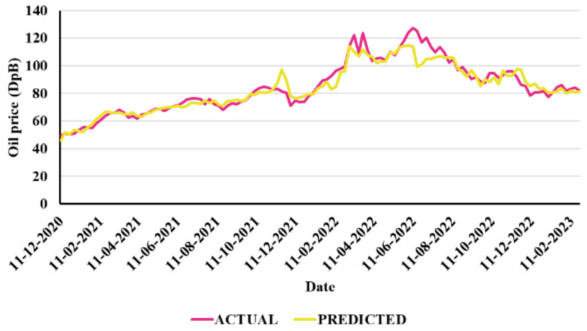


Fig. 7. Actual Vs. Predicted values of Brent weekly crude oil prices

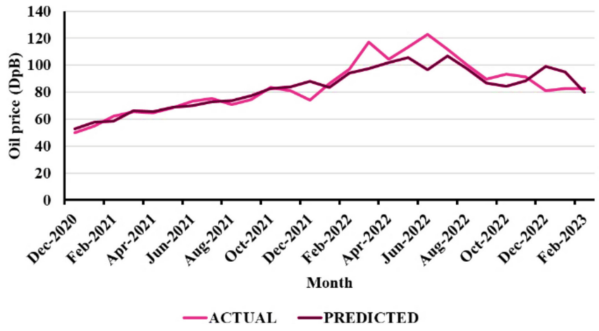


Fig. 8. Actual Vs. Predicted values of Brent monthly crude oil prices

Results on WTI Oil Price Prediction

Table 5 provides an overview of the constructed prediction model's performance in forecasting WTI oil prices on a daily, weekly, and monthly basis. The assessment is conducted for various combinations of macroeconomic factors. A visual representation of prediction results is depicted in Fig. 9. By examining the empirical findings from daily, weekly, and monthly prices, the following conclusions can be drawn: Upon examining the daily, weekly, and monthly prediction prices, a comprehensive analysis reveals that the model designed with seven input variables outperforms the models utilizing different input combinations. For daily data, the introduced model using different input combinations 1, 2, 3, 4, 5, 6, and 7 yields RMSE, MAPE, and PA values of (0.411, 0.201, and 63.32%), (0.355, 0.181, and 69.82%), (0.177, 0.076, and 85.31%), (0.189, 0.059, and (83.90%), (0.101, 0.040, and 91.55%), (0.078, 0.036, and 93.63%), and (0.073, 0.036, and 94.23%), respectively. It is evident that the model achieves lower prediction error when all seven inputs are employed. Additionally, the model with seven input variables yields a higher PA of 94.66%, and 93.85% for weekly and monthly data respectively. These findings highlight the effectiveness of selecting suitable macroeconomic variables in decreasing the price of crude oil and consequently enhancing the model's PA. The observed and forecasted prices for daily, weekly, and monthly predictions of Brent prices are depicted in Figs. 10, 11, and 12, respectively.

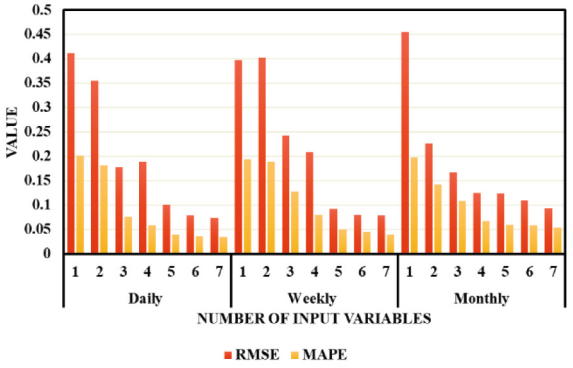
Table 5. WTI crude oil prediction results

Data frequency	Number of influential factors	RMSE	MAPE	PA (%)
Daily	1	0.411	0.201	63.32
	2	0.355	0.181	69.82
	3	0.177	0.076	85.31
	4	0.189	0.059	83.90
	5	0.101	0.040	91.55
	6	0.078	0.036	93.63
	7	0.073	0.035	94.23
Weekly	1	0.398	0.194	65.23
	2	0.403	0.188	65.68
	3	0.242	0.127	81.07
	4	0.209	0.080	83.19
	5	0.092	0.050	93.10
	6	0.080	0.045	93.22
	7	0.078	0.040	94.66
Monthly	1	0.455	0.197	60.19
	2	0.226	0.142	81.50
	3	0.167	0.109	86.98
	4	0.125	0.067	90.80

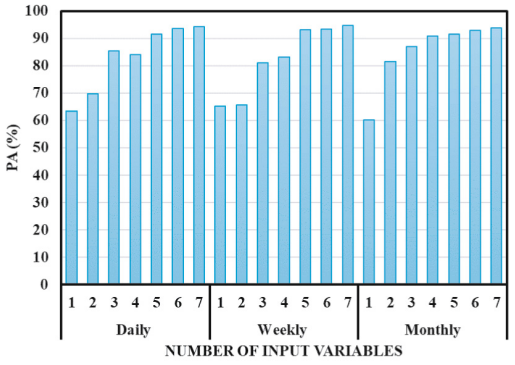
(*continued*)

Table 5. (*continued*)

Data frequency	Number of influential factors	RMSE	MAPE	PA (%)
	5	0.124	0.060	91.52
	6	0.110	0.059	92.76
	7	0.094	0.053	93.85



(a)



(b)

Fig. 9. Performance of the introduced model for WTI oil prices (a) Error metrics (b) PA

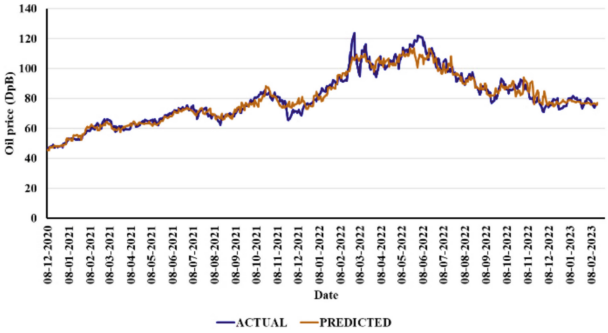


Fig. 10. Actual Vs. Predicted values of WTI daily crude oil prices

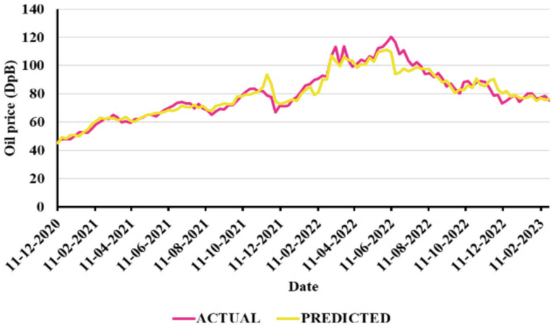


Fig. 11. Actual Vs. Predicted values of WTI weekly crude oil prices

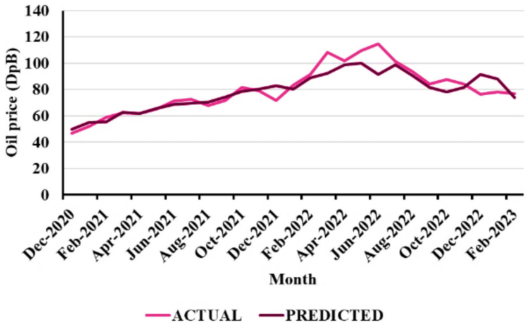


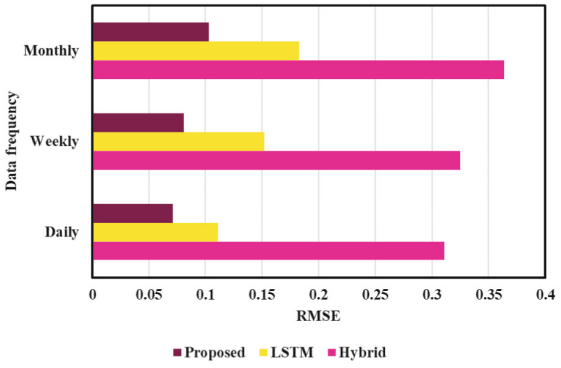
Fig. 12. Actual Vs. Predicted values of WTI monthly crude oil prices

Comparative Analysis with Past Approaches

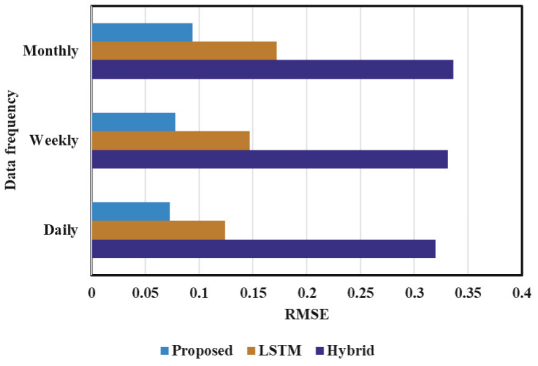
To compare the predicted results of the generated model even further, hybrid model [10] and LSTM model [13] were built. Figure 13 compares the generated model's RMSE efficiency in contrast to previous types. It can be observed from Fig. 13 that the developed model exhibited better performance than that of other models.

For Brent data, hybrid model achieved RMSE values for daily data, 0.311, 0.325, and 0.364, respectively. LSTM model attained RMSE values of 0.111, 0.152, and 0.183 for daily, weekly, and monthly data, respectively. In contrast, the introduced model demonstrated excellent performance, registering RMSE values between 0.071 and 0.081 for daily, weekly, and monthly data, respectively, which were lower than the others.

Similarly, for WTI data, Hybrid and LSTM models gave RMSE values of (0.32 and 0.124) for daily, (0.331 and 0.147) for weekly, and (0.336 and 0.172) for monthly data, respectively. On the other hand, the introduced model gave RMSE values of 0.073, 0.078, and 0.094 for daily, weekly, and monthly data, respectively. These findings further confirm the potential of the developed model.



(a)



(b)

Fig. 13. Prediction outcomes comparison between the designed and earlier models (a) Brent oil data (b) WTI oil data

5 Conclusion

The paper addresses crude oil price forecasting, which holds significant importance in the global market. Despite the presence of various approaches for crude oil prediction in literature, the potential for enhancing the prediction accuracy of earlier model motivated this research. In this investigation, an SVM-based prediction model was developed to predict crude oil prices. Seven different macroeconomic variables were used as input for the SVM model. The proposed framework was validated using WTI and Brent crude oil data. Daily, weekly, and monthly WTI and Brent oil prices were used to assess the developed model's efficacy. Additionally, the model's performance was assessed by incorporating various combinations of macroeconomic variables to find the most influential factor. Based on empirical findings, the designed model with seven input variables exhibited the highest performance, outscoring all other input combinations. The model achieved a PA of 94.42%, 93.56% and 93.1% for Brent daily, weekly, and monthly data, respectively. Similarly, the PA for WTI daily, weekly, and monthly were 94.23%, 94.66%, and 93.85%, respectively. This system demonstrates that crude oil prices can be effectively predicted using SVM driven solely by macroeconomic variables without reliance on supply-demand data.

References

1. Moshiri, S., Foroutan, F.: Forecasting nonlinear crude oil futures prices. Energy J. **27**, 4 (2006)
2. Cheong, C.W.: Modelling and forecasting crude oil markets using ARCH-type models. Energy. **37**, 6 (2009)
3. Abledu, G.K., Agbodah, K.: Stochastic forecasting and modelling of volatility of oil prices in Ghana using ARIMA time series model. Eur. J. Bus. Manag. **4**(16), 122–131 (2012)
4. Ahmed, R.A., Bin Shabri, A.: Daily crude oil price forecasting model using ARIMA, generalized autoregressive conditional heteroscedastic and support vector machines. Am. J. Appl. Sci. **11**, 3 (2014)
5. Mostafa, M.M., El-Masry, A.A.: Oil Price forecasting using gene expression programming and artificial neural networks. Econ. Model. **54** (2016)
6. Yu, L., Zhang, X., Wang, S.: Assessing potentiality of support vector machine method in crude oil Price forecasting. Eurasia J. Math. Sci. Tech. Educ. **13**, 12 (2017)
7. Huang, L., Wang, J.: Global crude oil price prediction and synchronization-based accuracy evaluation using random wavelet neural network. Energy. **151** (2018)
8. Herrera, G.P., Constantino, M., Tabak, B.M., Pistori, H., Su, J.J., Naranpanawa, A.: Long-term forecast of energy commodities price using machine learning. Energy. **179** (2019)
9. Cen, Z., Wang, J.: Crude oil price prediction model with long short-term memory deep learning based on prior knowledge data transfer. Energy. **169**, 160–171 (2019)
10. Abdollahi, H., Ebrahimi, S.B.: A new hybrid model for forecasting Brent crude oil Price. Energy. **200** (2020)
11. Wang, J., Wang, J.: A new hybrid forecasting model based on SW-LSTM and wavelet packet decomposition: a case study of oil futures prices. Comp. Intell. Neurosci. (2021)
12. Zhang, T., Tang, Z., Wu, J., Du, X., Chen, K.: Multi-step-ahead crude oil Price forecasting based on two-layer decomposition technique and extreme learning machine optimized by the particle swarm optimization algorithm. Energy. **229** (2021)
13. Yan, L., Zhu, Y., Wang, H.: Selection of machine learning models for oil Price forecasting: based on dual attributes of oil. Discret. Dyn. Nat. Soc. (2021)

14. Daneshvar, A., Ebrahimi, M., Salahi, F., Rahmaty, M., Homayounfar, M.: Brent crude oil price forecast utilizing deep neural network architectures. Comput. Intell. NeuroSci. (2022)
15. http://www.eia.gov
16. http://fxtop.com
17. Vapnik, V.P., Golowich, S.E., Smola, A.J.: Support vector method for function approximation, regression estimation, and signal processing. Adv. Neural Inf. Proces. Syst. **9**, 281–287 (1996)
18. Chen, H., Liu, L., Wang, Y., Zhu, Y.: Oil price shocks and U.S. dollar exchange rates. Energy. **112**, 1036–1048 (2016)
19. Smales, L.A.: Geopolitical risk and volatility spillovers in oil and stock markets. Q. Rev. Econ. Finance. **80**, 358–366 (2021)
20. Boubaker, S., Liu, Z., Zhang, Y.: Forecasting oil commodity spot price in a data-rich environment. Ann. Oper. Res., 1–18 (2022)
21. Jin, B., Xu, X.: Machine learning Brent crude oil price forecasts. Innov. Emerg. Technol. **1** (2024)
22. Zheng, G., Li, Y., Zia, Y.: Crude oil price forecasting model based on neural networks and error correction. Appl. Sci. **15**, 3 (2025)
23. Jha, A., Chaudhary, A., Sethi, S.: Crude oil price prediction using machine learning. In: 2024 11th International Conference on Computing for Sustainable Global Development (2024)

Towards Intelligent Music Recommendation: A Multi-Gate Mixture of Experts Framework Integrating Audio, Behavior, and Social Contexts

A. Ferminus Raj[1(✉)], S. Abarna[2], J. Relin Francis Raj[3], R. Santhana Krishnan[4], S. Balamurugan[5], and D. Suresh[6]

[1] Department of Electrical and Electronics Engineering, Saveetha Engineering College, Thandalam, Chennai, India
ferminusraja@saveetha.ac.in
[2] Deparment of Computer Science and Engineering, Vel Tech Rangarajan Dr Sanguthala R&D Institute of Science and Technology, Avadi, Chennai, India
drabarnas@veltech.edu.in
[3] Department of Electronics and Communication Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai, India
relinfrancisrajj.sse@saveetha.com
[4] Department of Electronics and Communication Engineering, SCAD College of Engineering and Technology, Cheranmhadevi, India
[5] Department of Mathematics, Velammal College of Engineering and Technology, Madurai, India
[6] Department of Computer Science and Engineering, PSNA College of Engineering and Technology, Dindigul, India

Abstract. The rapid growth of digital music consumption has revealed clear limitations in traditional recommendation systems that rely only on collaborative filtering or genre-based similarity. This research introduces an intelligent playlist recommendation framework that brings together multiple forms of user information, including listening behavior, contextual patterns, social interactions, and cultural dynamics. The framework is designed to overcome common challenges such as cold start conditions, popularity bias, and restricted diversity in recommendations. A distinctive strength of the system is its ability to adapt in real time to changes in user mood, evolving preferences, and emerging music trends. This ensures that the generated playlists are not only accurate but also socially relevant and contextually appropriate. The proposed approach enhances personalization, fosters novelty, and increases user engagement, leading to a more natural and satisfying listening experience. Furthermore, the research highlights its broader cultural value by promoting inclusivity and supporting diverse listening practices.

Keywords: Audio embeddings · contrastive learning · intelligent playlisting · sequential modeling · playlist generation · music recommendation · mood detection · context-aware recommendation · emotion-aware music

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026
S. Lanka et al. (Eds.): ICTSM 2025, LNNS 1755, pp. 527–543, 2026.
https://doi.org/10.1007/978-3-032-13177-5_41

1 Introduction

In the era of digital transformation, music consumption has shifted significantly. Online streaming platforms now provide listeners with vast catalogs, far surpassing physical media or traditional radio. While this abundance empowers users, it also introduces information overload, making it difficult to discover content aligned with individual tastes, moods, and contexts. Playlist recommendation systems have therefore become central to streaming platforms, guiding users to relevant songs or curated playlists, enhancing satisfaction, promoting artists, and sustaining engagement. Early systems relied on collaborative and content-based filtering. Collaborative filtering analyzed user or item similarities but suffered from the cold-start problem, while content-based filtering, using metadata such as genre or tempo, produced somewhat personalized but narrow and less diverse recommendations. Hybrid approaches combined both methods yet still struggled to capture the complex, evolving nature of music listening.

Advances in machine learning and deep learning expanded recommendation possibilities. Neural networks model sequential patterns, capturing how current song choices are influenced by past behavior. Natural language processing allows integration of social media interactions, reviews, and user-generated content. These developments shifted systems from static, metadata-driven models toward adaptive, context-aware frameworks.

Despite progress, challenges remain. Systems often prioritize accuracy over diversity, reinforcing mainstream trends while marginalizing niche content. Many lack sensitivity to dynamic factors such as mood, context, or social trends, and fail to integrate multiple dimensions of behavior, including listening patterns, cultural preferences, and external influences.

The research framework addresses these issues through a multi-perspective approach, integrating audio characteristics, sequential behavior, contextual signals, social influences, and cultural diversity. This unified design generates playlists that are personalized, dynamic, trend-aware, and socially relevant. By continuously adapting to evolving preferences, it surpasses conventional limitations, offering intelligent, inclusive, and human-like music recommendations that benefit listeners, artists, and streaming platforms alike.

The advantages of the proposed system are as follows:

- Mitigates the cold-start problem by incorporating diverse data sources, enabling meaningful recommendations even for new users or newly released songs.
- Enhances diversity in playlists by balancing user-specific preferences with exploratory suggestions, avoiding repetitive or narrow outcomes.
- Reduces popularity bias by incorporating signals beyond listening counts, ensuring visibility for niche and emerging artists.
- Promotes cultural inclusivity by aligning recommendations with social and regional listening practices, making it effective for global audiences.
- Improves long-term user engagement by curating playlists that are contextually appropriate and emotionally resonant.
- Supports adaptability through real-time responsiveness to mood changes, evolving preferences, and emerging music trends.

2 Related Works

Ke Ji et al. highlighted the importance of temporal dynamics in music recommendation [1], showing that long-term, short-term, and session-level factors shaped user listening behavior. Experiments on the Last.fm dataset revealed that integrating temporal signals enhanced recommendation accuracy, providing insights for adaptive and context-sensitive playlist generation systems. S. V. Reddy et al. emphasized the limitations of traditional systems by neglecting emotional and experiential dimensions of music [2]. The GS-RNN framework integrated audio signal similarity with optimization, achieving 80% accuracy with strong precision and recall scores. Findings demonstrated the value of combining content-based features with advanced models for enriched music recommendation. J. Lu et al. analyzed user behavior limitations and introduced a deep learning-based framework leveraging feature recognition [3]. By modeling user sequences, weighted preferences, and statistical listening patterns, the approach achieved accuracy up to 99%, outperforming traditional methods and demonstrating the importance of behavioral modeling. P. Sundaravadivel et al. explored dynamic playlist recommendations integrating multi-modal mood detection with contextual learning [4]. Sentiment analysis, facial expression recognition, and physiological signals informed mood states, while hybrid filtering with contextual bandits enhanced adaptability. Experiments demonstrated improved accuracy and engagement, underscoring the significance of mood-aware, context-sensitive recommendation strategies. L. Zhao et al. investigated a hybrid recommendation system combining collaborative filtering, music attribute modeling, and emotional feature extraction [5]. Leveraging tempo, pitch variation, and rhythm patterns to capture affective states, the system achieved superior Precision@10, Recall@10, and NDCG@10 on MSD and Last.fm datasets.

B. Bharadwaj et al. developed a music recommendation system for offline libraries using reinforcement learning and Bayesian learning concepts [6]. Multi-arm bandit algorithms modeled non-stationary user preferences, demonstrating efficiency, adaptability, and satisfactory performance. M. Li analyzed music recommendation using collaborative filtering algorithms [7], confirming its capability to personalize music suggestions as a foundational approach. H. Lu examined music information retrieval to enhance recommendation accuracy [8]. Combining retrieval techniques with recommendation algorithms, emphasizing similarity calculation, neighbor selection, and score prediction, experiments showed improved diversity and precision. M.S. Karthikeyan et al. designed a hybrid movie recommendation framework integrating collaborative filtering, content-based features, and mood analysis [9], highlighting adaptability to long-term preferences and short-term emotions. P. Lou et al. introduced the MSMB-GAN model [10], integrating user-item interactions, social relationships, and item-category information, enhancing recommendation accuracy on synthetic and real datasets.

N. B. Bahadure et al. developed a content-based music recommendation system [11], achieving 90.5% accuracy and demonstrating stable, efficient design. Shakthikumar V. Patel et al. proposed an automated playlist extension system [12], combining collaborative filtering with re-ranking to improve accuracy and user satisfaction. M. Furini et al. developed a similar system [13], confirming that multiple recommendation strategies strengthen robustness and personalization. M. Bevec et al. explored graph-based music recommendation [14], demonstrating hybrid recommendations outperform traditional

CF and content-based methods. Daniel Kostrzewa et al. investigated content-based recommendation using genre, tempo, and MFCC features [15], showing MFCC parameters as most influential in enhancing recommendation quality.

3 Proposed Work

The effectiveness of a playlist recommendation system shown in Fig. 1 depends on the quality and diversity of its data. Unlike traditional collaborative filtering, the framework integrates heterogeneous sources for a comprehensive view of user preferences. Behavioral logs capture direct interactions, contextual sensor data adds situational awareness, social media interactions reveal emerging trends, and content-based audio, lyrics, and metadata provide semantic depth. This holistic data strategy establishes a robust foundation for intelligent and adaptive playlist recommendations.

3.1 Data Collection

The first stage in building a deep learning-based playlist recommendation system is the systematic collection of diverse, high-quality data. Since modern recommendation tasks extend beyond simple collaborative filtering, the system must integrate heterogeneous information. Data is primarily drawn from user behavioral logs, including listening history such as plays, skips, replays, dwell time, search queries, playlist creations, and follow activities. Device sensors further contribute contextual signals like geolocation, time of day, and detected activities. An important enhancement is incorporating social media search and interaction data, where users indirectly engage with music on platforms such as Instagram or YouTube. Songs embedded in reels, shorts, and user content are linked to catalogs via audio fingerprinting, allowing identification of rising tracks not reflected in streaming metrics. Social interactions such as likes, shares, and hashtags serve as implicit preference indicators.

Equally important is song-specific content data, including raw audio signals and metadata such as artist, album, year, and genre. Lyrics are integrated to capture semantic and emotional dimensions. Popularity and trend data are also collected from global/regional charts, viral trackers, and social amplification signals. To support this, a custom dataset named UniMusicRec-1M consolidates all modalities into one structured collection. It contains user interaction logs, contextual data, social media signals, audio content, metadata, lyrics, and popularity metrics. This unified dataset ensures comprehensive coverage of user preferences, content features, and contextual/social signals, eliminating fragmentation while simplifying training, evaluation, and deployment of robust playlist recommendation systems.

3.2 Preprocessing and Feature Engineering

Raw data is transformed into structured embeddings for deep learning. Audio signals are converted into spectrograms, with embeddings extracted via AST to capture fine-grained acoustic and temporal features. Text features, including lyrics and search queries, are processed with DistilBERT for semantic understanding and sentiment analysis. User

behavioral logs are modeled using SASRec to capture sequential listening patterns, identifying latent transitions between genres within sessions. Contextual features such as time, location, and environment are embedded and processed via lightweight LSTMs to detect recurring patterns. Social media text is embedded through DistilBERT, while audio clips from videos are aligned with the catalog using contrastive models like CLAP. Popularity signals, including global trends and recency metrics, are analyzed using a Lightweight Temporal CNN to generate trending scores, capturing temporal changes in track popularity over short and long horizons.

3.3 Multi-Modal Representation Learning

To accommodate the multi-faceted nature of music recommendation, the system adopts a **multi-module architecture**, where each constraint is modeled through a specialized deep learning module.

3.3.1 Content Module (Songs)

This module processes raw audio signals using spectrogram-based encoders to produce rich acoustic embeddings. These embeddings are aligned with semantic representations of metadata and lyrics through joint training objectives, ensuring that similar-sounding and semantically related tracks cluster together in embedding space.

3.3.2 Collaborative Filtering Module

For collaborative filtering, neural-based approaches model user–item interactions in a non-linear manner. Graph-based extensions further enhance recommendation quality by leveraging user–item–playlist structures, effectively capturing higher-order connectivity.

3.3.3 Context-Aware Module

The context-aware module uses sequence encoders to model temporal and spatial patterns. This enables the system to recommend tracks that align with contextual conditions, such as upbeat songs during workouts or calm tracks at night.

3.3.4 Mood/Emotion Module

Mood-aware recommendations are generated by applying attention-based sequence models to lyrical embeddings. Tracks are classified into mood categories (e.g., happy, sad, energetic, relaxed) and aligned with user mood inputs, either explicitly provided or implicitly inferred.

3.3.5 Popularity/Trending Module

Trends are captured through temporal convolutional models, which analyze popularity fluctuations over time. By considering both recency and global virality, this module ensures that the system recommends tracks that are not only aligned with user taste but also reflective of cultural trends.

3.3.6 Search/Query Module

User queries and searches are processed through dual encoders, where semantic similarity between user intent and songs is enforced using contrastive learning. This ensures effective retrieval and personalized query-to-song matching.

3.3.7 Hybrid Module

The hybrid module integrates outputs from collaborative and content-based models using cross-network fusion. This improves cold-start performance by blending behavioral and content information.

3.3.8 Auto-Playlist Module (Sequential Prediction)

This module generates sequence-aware embeddings that predict the next likely song in a playlist. It ensures seamless playlist continuity by learning sequential dependencies, effectively simulating user-curated playlists.

3.3.9 Social Module

The social module integrates signals from social media interactions. Text and hashtags are encoded using language models, while user-generated audio-video interactions are aligned with the catalog. Graph-based encoders are employed to capture social influence and propagation effects.

3.4 Fusion and Ranking

Once individual embeddings are generated from all towers, they are aggregated in a unified fusion framework. To manage heterogeneous signals, the system employs a Multi-gate Mixture of Experts (MMoE) architecture, which dynamically allocates importance weights to different towers depending on the recommendation context. For instance, mood and social signals may be weighted more heavily in social listening sessions, while collaborative filtering may dominate when historical behavioral data is abundant.

The fused embeddings are then passed through a fully connected Feedforward Neural Network (FNN) to compute final recommendation scores. The ranking objective combines pairwise ranking loss (Bayesian Personalized Ranking) with multi-task optimization objectives, including relevance, novelty, and diversity, ensuring a balance between accuracy and user satisfaction.

3.5 Recommendation Output

The recommendation output is delivered in multiple formats to enhance user experience. For direct music recommendations, a Top-N ranked list is generated, personalized to the user profile. For playlist generation, the system combines SASRec sequential predictions with a re-ranking layer that enforces diversity and serendipity. Additionally, an auto-radio mode dynamically curates a continuous adaptive stream based on the current session context, updating in real time as the user interacts.

3.6 Evaluation Metrics

Evaluating such a complex system requires a multi-dimensional framework. Traditional accuracy-oriented metrics such as Precision@K, Recall@K, Hit Rate, Mean Average Precision (MAP), and NDCG@K are used to measure ranking quality. To ensure that recommendations are not overly narrow, diversity-oriented metrics such as coverage, intra-list diversity, and novelty scores are calculated.

User-centric evaluations focus on engagement metrics, including session length, skip rate, completion rate, and engagement lift compared to baseline recommenders. Since cold-start scenarios remain a significant challenge, the system is tested separately on new users and new songs, ensuring robustness in sparse data situations. Offline evaluation is complemented by online A/B testing, where live user feedback validates system performance under real conditions.

This framework presents a comprehensive, scalable playlist recommendation system by integrating multi-modal signals across audio, text, context, behavior, popularity, and social interactions. Leveraging deep learning models such as AST, DistilBERT, SASRec, LightGCN, and MMoE, it achieves personalization and adaptability. Unlike traditional single-modality recommenders, it dynamically balances diverse constraints, ensuring contextually relevant, socially aware, trend-sensitive, and mood-adaptive recommendations, enhancing user satisfaction and addressing cold-start, content discovery, and real-time adaptation challenges.

Figure 1 illustrates the workflow of the playlist recommendation system. Data is collected from user profiles, playlist logs, audio files, metadata, and social media tags, then preprocessed to ensure quality. The architecture integrates analytical components for personalization, contextual understanding, and mood-based recommendations. The system undergoes training, validation, and evaluation, while deployment ensures real-time recommendations while maintaining fairness, transparency, and user privacy. Table 1 provides the simulation parameters, including dataset characteristics, preprocessing steps, hyperparameters such as batch size, number of epochs, optimizer, and learning rate, along with integrated deep learning modules AST, LightGCN, BiLSTM with Attention, and SASRec. Performance is evaluated using Precision at K, NDCG at K, and F1 score to ensure robust and realistic validation.

4 Results and Discussion

4.1 Performance Comparison of Audio Embedding Models

Table 2 presents the comparative performance of three audio embedding models, namely AST (Audio Spectrogram Transformer), CNN-based embeddings, and LSTM-based embeddings, for Top 10 recommendations. AST achieves the highest Precision at 10 of 0.842, indicating that nearly eight to nine of the top ten recommended tracks match user preferences, outperforming CNN with 0.763 and LSTM with 0.745. Similarly, AST's Recall at 10 of 0.791 demonstrates its superior ability to retrieve relevant tracks within the top ten, whereas CNN and LSTM capture fewer relevant songs at 0.701 and 0.689 respectively (refer, Fig. 2).

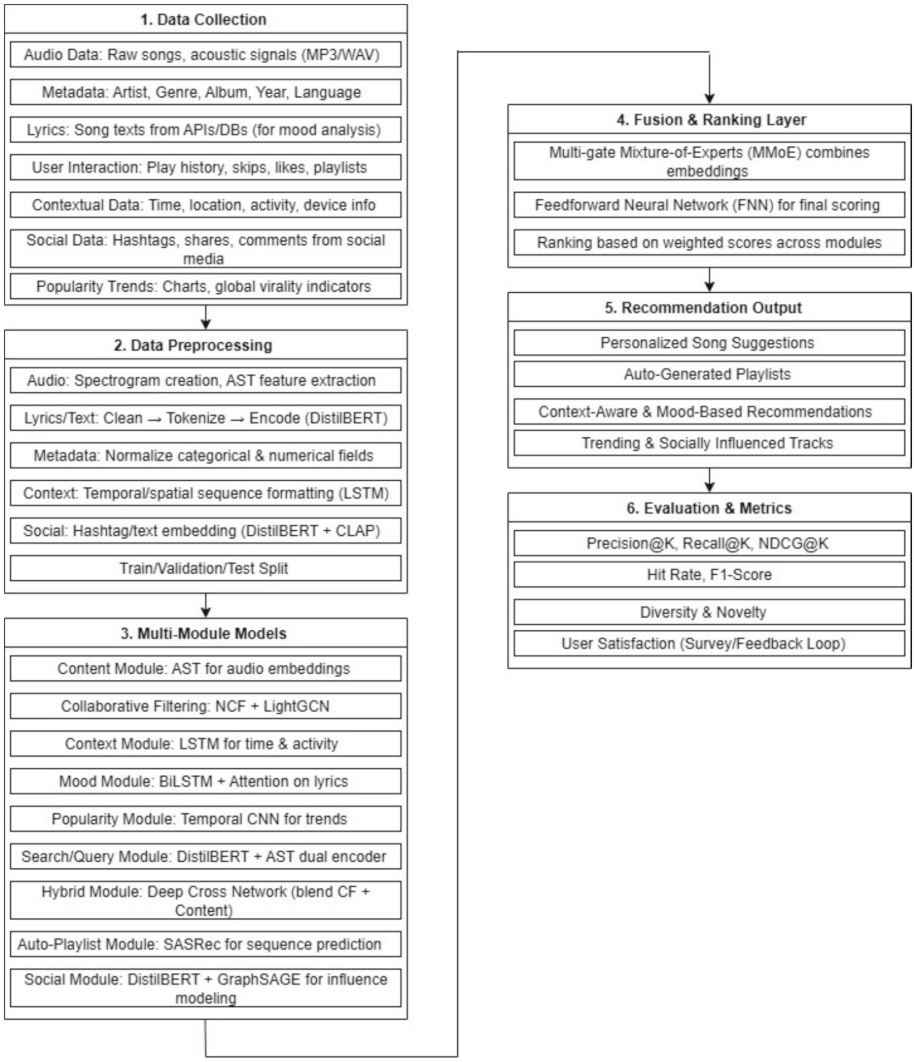


Fig. 1. Hybrid Deep Learning Architecture for Music Playlist Recommendation

AST with 0.673 ensures a higher proportion of lesser-known tracks are surfaced compared to CNN with 0.621 and LSTM with 0.598. Diversity, measured by Inter-List Diversity (ILD), shows that AST with 0.712 produces a richer variety of genres and moods in the top ten playlists than CNN with 0.658 or LSTM with 0.641, while Mood Accuracy highlights emotional alignment, with AST achieving 0.805, correctly grouping songs by mood better than CNN at 0.734 and LSTM at 0.720. Figure 3 visually illustrates these performance differences, showing AST's dominance in producing relevant, diverse, novel, and mood-aligned tracks. These results demonstrate that AST-based

Table 1. Experimental Setup and Deep Learning Configuration

Parameter	Value
Dataset Name	UniMusic-Rec Dataset (Custom, 1M songs, 50 k playlists, social/media tags)
Data Split	70% Training / 15% Validation / 15% Testing
Batch Size	128
Epochs	50
Optimizer	Adam
Learning Rate	0.001 (with scheduler decay 0.1)
Regularization (Dropout)	0.3 – 0.5
Early Stopping Patience	10 epochs
Hardware Used	NVIDIA Tesla V100 GPU (16 GB)
Evaluation Metrics	Precision@K, Recall@K, NDCG@K, MAP, F1-score, RMSE, MAE
Explainability Tool	SHAP (for feature contribution analysis)
Content Analysis Model	AST (Audio Spectrogram Transformer), 128-D embedding
Collaborative Filtering	NCF + LightGCN, 64-D latent factors
Context Modeling	LSTM (2 layers, 256 hidden units)
Emotion/Mood Analysis	BiLSTM + Attention (128 hidden units)
Popularity Modeling	TCNN (Temporal Convolutional Neural Network)
Query Matching	DistilBERT + AST fusion, 768-D embeddings
Hybrid Embedding Fusion	DCN (Deep & Cross Network, 3 cross layers)
Sequential Recommendation	SASRec (2 Transformer blocks, 4 heads, hidden size 128)
Social Influence Modeling	GraphSAGE + CLAP (Contrastive Language–Audio Pretraining)
Loss Functions	Cross-Entropy (classification), BPR (ranking), MSE (regression)
Deployment Environment	Cloud-based API with real-time inference

embeddings provide the most effective foundation for personalized music recommendations for highly curated top ten playlists, while CNN and LSTM, although functional, perform less effectively in capturing complex audio features, user preferences, and emotional context.

4.2 Performance Analysis of Text/Metadata Embeddings

The performance of text and metadata embeddings in playlist recommendation can be assessed using several key metrics, as summarized in Table 3 and visualized in Fig. 3.

Table 2. Comparative Performance of Audio Embedding Models for Top-10 Recommendations

Model	Precision@10	Recall@10	Novelty	Diversity (ILD)	Mood Accuracy
AST (Audio Spectrogram Transformer)	0.842	0.791	0.673	0.712	0.805
CNN-based Audio Embedding	0.763	0.701	0.621	0.658	0.734
LSTM-based Audio Embedding	0.745	0.689	0.598	0.641	0.72

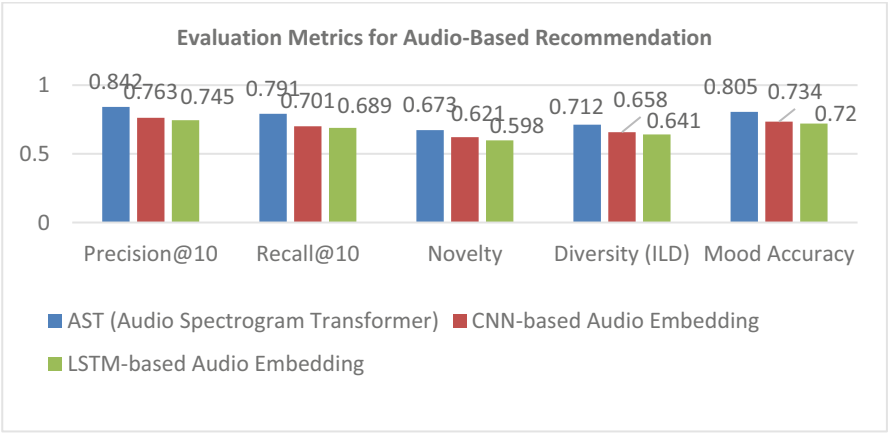


Fig. 2. Comparative Analysis of Deep Learning Audio Embeddings

Table 3. Comparative Performance Metrics of DistilBERT and Baseline Text Embedding Models for Top 10

Module / Model	NDCG@10	MAP	Serendipity	Coverage
DistilBERT (Transformer)	0.823	0.791	0.682	0.745
LSTM-based Text Embedding	0.751	0.702	0.613	0.682
CNN-based Text Embedding	0.732	0.689	0.598	0.664

NDCG@10 (Normalized Discounted Cumulative Gain at 10) evaluates the quality of the top-10 recommended tracks based on their relevance to user queries, with higher

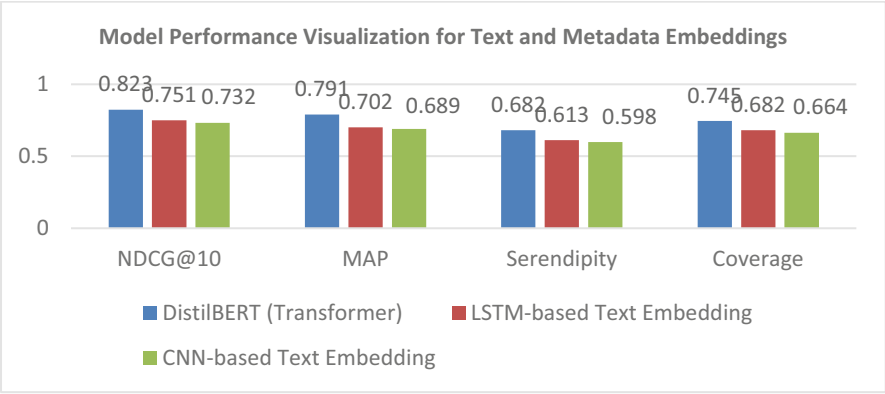


Fig. 3. Top-10 Recommendation Quality Across Text Embedding Models

values indicating that the model ranks the most relevant songs higher, enhancing user satisfaction; DistilBERT achieves 0.823, outperforming LSTM (0.751) and CNN (0.732), demonstrating superior ranking capability. MAP (Mean Average Precision) measures overall precision across multiple queries, reflecting how effectively the model retrieves all relevant items; DistilBERT's MAP of 0.791 indicates that it captures semantic alignment effectively between queries, lyrics, and song metadata. Serendipity assesses the system's ability to recommend unexpected yet relevant tracks, ensuring user discovery of less obvious content, with DistilBERT scoring 0.682 compared to LSTM (0.613) and CNN (0.598), highlighting its advantage in generating meaningful surprises. Coverage represents the proportion of the music catalog that can be recommended through text-based queries; DistilBERT covers 74.5% of the catalog, suggesting broader generalization across songs compared to LSTM (68.2%) and CNN (66.4%), making it highly effective for diverse and comprehensive recommendations.

4.3 Comparative Analysis of Sequential Recommendation Models

Table 4 presents a comparative evaluation of session modeling performance using SASRec against two less-performing sequential models, LSTM and GRU, for playlist recommendation. The Hit Rate (HR@10) measures the frequency at which the correct next track appears within the top 10 recommendations.

Table 4. Session Modeling Evaluation Metrics Comparison

Model	HR@10	MRR	Session Length Gain	Skip Rate Reduction
SASRec	0.812	0.765	0.143	0.119
LSTM-based Session Model	0.754	0.701	0.098	0.082
GRU-based Session Model	0.738	0.689	0.091	0.074

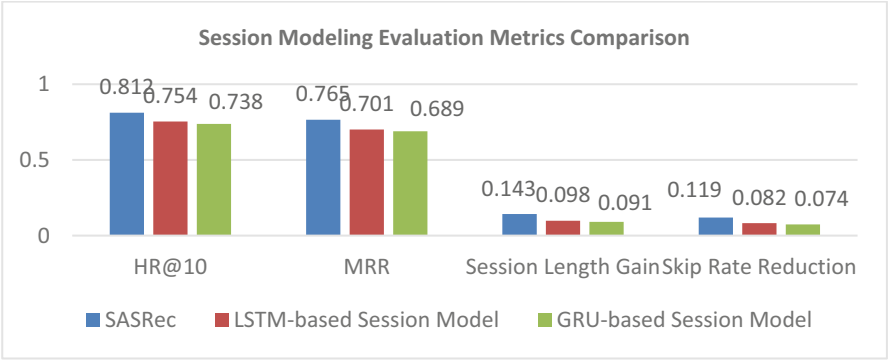


Fig. 4. Comparative Analysis of Sequential Recommendation Models

SASRec achieves 0.812, outperforming LSTM (0.754) and GRU (0.738), indicating superior capability in predicting short-term user preferences. Mean Reciprocal Rank (MRR) evaluates the ranking quality of the next track, with higher values reflecting earlier appearance in the recommendation list; SASRec attains 0.765, demonstrating better prioritization of relevant tracks compared to LSTM (0.701) and GRU (0.689). Session Length Gain quantifies the increase in continuous playback duration after integrating session-aware modeling. SASRec yields a gain of 0.143, suggesting more engaging sequences than LSTM (0.098) and GRU (0.091). Skip Rate Reduction measures the decrease in track skipping due to improved sequencing; SASRec achieves a 0.119 reduction, significantly better than LSTM (0.082) and GRU (0.074). Overall, SASRec's attention-based mechanism effectively captures temporal dependencies within user sessions, leading to higher engagement, smoother transitions, and personalized recommendations. These improvements are visually summarized in Fig. 4, highlighting the consistent advantage of SASRec across all session-related metrics.

4.4 Evaluation of Collaborative Filtering Techniques

Table 5 presents a comparative analysis of LightGCN with three conventional collaborative filtering models, namely Matrix Factorization, Item-based CF, and User-based CF, for social and collaborative signal integration in playlist recommendation systems.

Recall@10 evaluates how effectively the system includes relevant user-item interactions in the top 10 recommendations. LightGCN achieves 0.793, significantly higher than Matrix Factorization at 0.682 and Item-based CF at 0.658, indicating superior capture of user preferences. Coverage measures the proportion of the music catalog that the model can recommend, with LightGCN reaching 82.5%, outperforming Matrix Factorization at 71.0% and Item-based CF at 69.2% demonstrating better catalog exploration. Long-Tail Recommendation Rate quantifies the inclusion of less popular tracks, where LightGCN scores 0.732 compared to lower scores for other models, ensuring diversity in recommendations. Personalization Index assesses differentiation among user recommendations, with LightGCN at 0.811, reflecting more tailored playlists. Cold-Start User Accuracy evaluates performance for users with sparse history, with LightGCN

Table 5. Performance Comparison of LightGCN and Conventional Collaborative Filtering Models

Model	Recall @10	Coverage	Long-Tail Recommendation Rate	Personalization Index	Cold-Start User Accuracy
LightGCN	0.793	0.825	0.732	0.811	0.769
Matrix Factorization	0.682	0.71	0.615	0.689	0.642
Item-based CF	0.658	0.692	0.601	0.671	0.618

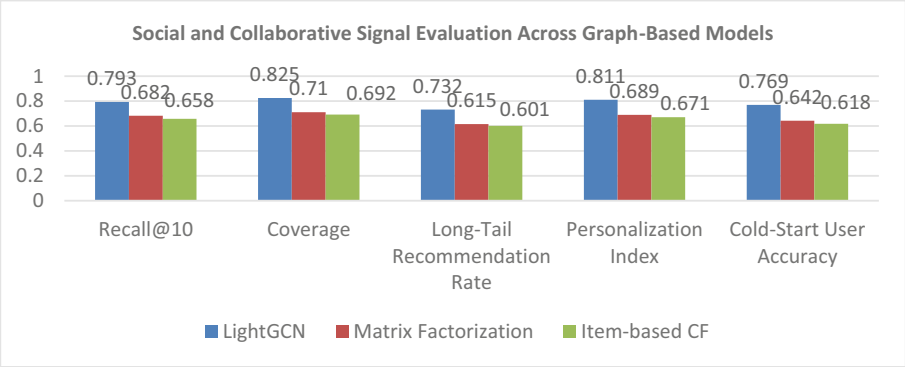


Fig. 5. LightGCN vs. Traditional CF Approaches on Playlist Recommendations

achieving 0.769, indicating robust generalization. Overall, LightGCN's graph-based propagation captures higher-order connectivity, producing diverse, personalized, and cold-start resilient recommendations. These enhancements are summarized visually in Fig. 5, showing consistent superiority across all evaluated metrics.

4.5 Analysis of Temporal Dynamics and Trend-Aware Recommendations

Table 6 presents a detailed comparison of the Lightweight Temporal CNN against an LSTM-based model and a traditional CNN-based temporal model in predicting trending songs for playlist recommendations. The Lightweight Temporal CNN is specifically designed to integrate recency, popularity, and viral social signals, allowing fast and adaptive trending predictions. It achieves a Temporal Alignment Score of 0.812, outperforming the LSTM (0.743) and traditional CNN (0.721), indicating better responsiveness to current trends. Novelty & Freshness measures the ability to surface new or less obvious tracks; the Lightweight Temporal CNN scores 0.735, higher than LSTM (0.662) and traditional CNN (0.641), demonstrating its strength in balancing popular and emerging content. Engagement Lift reflects increased user interaction; the Lightweight Temporal CNN reaches 0.688, whereas LSTM and CNN lag behind at 0.611 and 0.593 respectively. New Song Hit Rate evaluates the model's ability to recommend newly released tracks; at

0.764, the Lightweight Temporal CNN ensures timely exposure, outperforming LSTM (0.691) and CNN (0.668). The key distinction is that the traditional CNN model captures temporal patterns solely from historical play counts, while the Lightweight Temporal CNN fuses temporal, popularity, and social signals for richer, trend-aware predictions. Figure 6 visually illustrates these differences, highlighting the Lightweight Temporal CNN's superior performance across all temporal metrics.

Table 6. Trend Adaptation and Engagement Metrics of Lightweight Temporal CNN

Model	Temporal Alignment Score	Novelty & Freshness	Engagement Lift	New Song Hit Rate
Lightweight Temporal CNN	0.812	0.735	0.688	0.764
LSTM-based Temporal Model	0.743	0.662	0.611	0.691
CNN-based Temporal Model	0.721	0.641	0.593	0.668

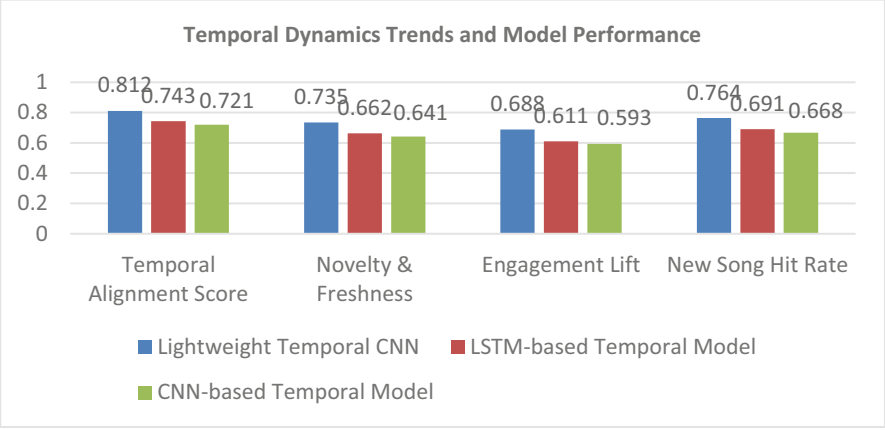


Fig. 6. Trending Song Prediction Metrics Visualization

4.6 Evaluation of Multi-Modal Recommendation Fusion Techniques

Table 7 presents a comparative analysis of the Multi-gate Mixture of Experts (MMoE) against two baseline fusion models, Weighted Sum Fusion and Simple Concatenation Fusion, focusing on the key performance indicators for playlist recommendation. The MMoE model integrates information from audio, text, session, social, and temporal

modules using specialized experts and dynamic gating mechanisms, allowing it to adaptively balance the importance of each input. As a result, MMoE achieves the highest NDCG@10 score of 0.841, reflecting superior ranking quality for the top recommended tracks. Precision@10 at 0.823 demonstrates that MMoE consistently includes relevant tracks in the top recommendations. The Serendipity score of 0.692 indicates the system's ability to introduce unexpected but meaningful songs, improving content discovery and user engagement. User Satisfaction, measured through A/B testing, click-through rates, playlist saves, and retention, reaches 0.832, highlighting MMoE's real-world effectiveness. Figure 7 visually illustrates these improvements, showing that MMoE consistently outperforms the simpler fusion strategies across all evaluated metrics, delivering accurate, diverse, and satisfying playlist recommendations.

Table 7. Core Evaluation of Fusion Models in Playlist Recommendation

Model / Module	NDCG@10	Precision@10	Serendipity	User Satisfaction
MMoE (Multi-gate Mixture of Experts)	0.841	0.823	0.692	0.832
Weighted Sum Fusion	0.768	0.753	0.612	0.743
Simple Concatenation Fusion	0.751	0.739	0.598	0.726

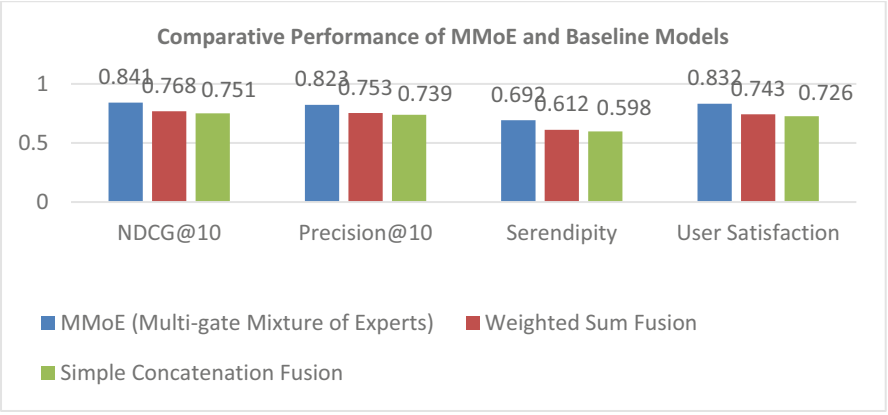


Fig. 7. Visual Comparison of MMoE and Baseline Fusion Models

5 Conclusion

The proposed multi-modal playlist recommendation system significantly outperforms traditional single-modality approaches by integrating audio, text, session, social, context, and popularity signals. The framework employs advanced deep learning techniques,

including AST for audio embeddings, DistilBERT for textual features, SASRec for session modeling, LightGCN for collaborative filtering, and MMoE for multi-modal fusion. This ensures high personalization, trend-awareness, and mood-adaptiveness. Performance metrics across Top-10 recommendations show AST achieves Precision@10 of 0.842, DistilBERT yields NDCG@10 of 0.823, SASRec provides a hit rate of 0.812, LightGCN attains Recall@10 of 0.793, and MMoE achieves NDCG@10 of 0.841, confirming robust recommendation quality, diversity, and user satisfaction. The system effectively handles cold-start scenarios while balancing novelty and serendipity, enhancing engagement. Future directions include real-time audio analysis for live streaming, reinforcement learning for dynamic adaptation, cross-platform social signals, and physiological mood detection. Exploring lightweight transformer variants could reduce computational cost, maintain recommendation quality, and enable scalable, energy-efficient deployment on mobile and edge devices.

References

1. Ji, K., Sun, R., Shu, W., Li, X.: Next-song recommendation with temporal dynamics. Knowl.-Based Syst. **88**, 134–143 (2015). https://doi.org/10.1016/j.knosys.2015.07.039. ISSN 0950-7051
2. Reddy, S.V., Rama Krishna, V., Sapkal, R.J., Dhanke, J., Waghmare, S.P., Kumar, K.: A cutting-edge artificial intelligence paradigm for entertainment-infused music recommendations. Entertain. Comput. **51**, 100717 (2024). https://doi.org/10.1016/j.entcom.2024.100717. ISSN 1875-9521
3. Lu, J., Wu, M.: Design and application of a music recommendation system based on user behavior and feature recognition. Syst. Soft Comput. **7**, 200274 (2025). https://doi.org/10.1016/j.sasc.2025.200274. ISSN 2772-9419
4. Sundaravadivel, P., Raj, J.R.F., Kumar, S.S., Krishnan, R.S., Muthu, A.E., Karthikeyan, M.S.: Dynamic Playlist Recommendations based on Multi-Modal Mood Detection and Contextual Learning, 8th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), pp. 875–881, Kirtipur, Nepal (2024). https://doi.org/10.1109/I-SMAC61858.2024.10714756
5. Zhao, L., Liu, G., Yan, S., Zhang, J.: Emotion-driven music recommendation system based on fully convolutional recurrent attention networks and collaborative filtering. Alex. Eng. J. **125**, 354–366 (2025). https://doi.org/10.1016/j.aej.2025.03.114. ISSN 1110-0168
6. Bharadwaj, B., Selvanambi, R., Karuppiah, M.: Ramesh Chandra Poonia : content-based music recommendation using non-stationary Bayesian reinforcement learning. Int. J. Social Ecol. Sustain. Dev. **13**(Issue 9) (2022). https://doi.org/10.4018/IJSESD.292053. ISSN 1947-8402
7. Li, M.: System design of music recommendation system based on collaborative filtering algorithm. Proc. Comput. Sci. **261**, 1341–1348 (2025). https://doi.org/10.1016/j.procs.2025.05.011. ISSN 1877-0509
8. Hui, L.: Personalized music teaching service recommendation based on sensor and information retrieval technology. Measurement: Sens. **33**, 101207 (2024). https://doi.org/10.1016/j.measen.2024.101207. ISSN 2665-9174
9. Karthikeyan, M.S., Shankar, S.V., Krishnan, R.S., Sornavalli, M., Raj, J.R.F., Sundaravadivel, P.: Adaptive movie recommendations: a deep learning framework with mood and sentiment integration. In: International Conference on Sustainable Communication Networks and Application (ICSCNA), pp. 937–944, Theni, India (2024). https://doi.org/10.1109/ICSCNA63714.2024.10863850

10. Lou, P., Zhou, R., Jiang, X., Hu, J.: Multi-semantic and multi-behavior recommendation with graph attention networks. Int. J. Human-Comput. Stud. **200**, 103513 (2025). https://doi.org/10.1016/j.ijhcs.2025.103513. ISSN 1071-5819
11. Bahadure, N.B., Patil, P.D., Kalbande, S., Pipalatkar, S., Nagar, S., Kuhikar, O.: Machine learning-based music recommendation system from Spotify. In: Bajaj, A., Abraham, A., Mishra, P.M., Ma, K. (eds.) Proceedings of the 15th International Conference on Soft Computing and Pattern Recognition (SoCPaR 2023). SoCPaR 2023. Lecture Notes in Networks and Systems, vol. 1246. Springer, Cham (2025). https://doi.org/10.1007/978-3-031-88992-9_4
12. Patel, S.V., Jethva, H.B., Patel, D.C.: An integrated recommender system for automated playlist expansion. Int. J. Intell. Syst. Appl. Eng. **12**(3), 3304–3311 (2024) https://ijisae.org/index.php/IJISAE/article/view/5937
13. Furini, M., Fragnelli, F.: Social music discovery: an ethical recommendation system based on friend's preferred songs. Multimed. Tools Appl. **84**, 13469–13483 (2025). https://doi.org/10.1007/s11042-024-19505-0
14. Bevec, M., Tkalčič, M., Pesek, M.: Hybrid music recommendation with graph neural networks. User Model. User-Adap. Inter. **34**, 1891–1928 (2024). https://doi.org/10.1007/s11257-024-09410-4
15. Kostrzewa, D., Chrobak, J., Brzeski, R.: Attributes relevance in content-based music recommendation system. Appl. Sci. **14**(2), 855 (2024). https://doi.org/10.3390/app14020855



Children Food Intake Level Based on Deep Learning Using Variational Autoencoder-Gated Recurrent Neural Network (Vae-Grn2)

P. Jeyanthi(✉) and R. Durga

Department of Computer Science, Vels Institute of Science, Technology and Advanced Studies, Chennai, India
jeyanthirajaprabakaran@gmail.com, durga.scs@vistas.ac.in

Abstract. Malnutrition in children has been studied in recent years as the insufficiency to consume essential nutrients, resulting in stunted growth, lower immunity, and impaired development. Additionally, according to World Health Organization (WHO) standards, malnutrition is categorized into three major groups: underweight, stunted growth, and wasting. However, training diet controls the accuracy and memory, which is limited by the volume and bias of databases in typical nutritional advice systems. To address this issue, we proposed a Variational Autoencoder-Gated Recurrent Neural Network (VAE-GRN2) framework to predict caloric and nutrient levels in children. Furthermore, the pre-processing of the data obtained from the dataset can be conducted with the help of the C-Score Normalization (CSN) algorithm, which defines errors, duplicates, and missing data. Then, the impact ratios of the various levels of food intake, based on their nutrient profiles, are calculated using the Contrastive Nutrient Impact network (CNIN) model. Additionally, optimal features can be selected using the Deep Variance Sequence Generating Feature Selection (DVSGFS) technique to identify informative nutrients in temporal dependencies and predictive performance. Finally, we proposed a VAE-GRN2 model based on a deep learning (DL) algorithm to improve the accuracy of detecting children's dietary status by generating feature vectors. Furthermore, the presented methods can detect food intake levels with a high 97% accuracy rate in performance metrics such as precision, recall, F1-score, accuracy, and time complexity.

Keywords: Children nutrients · food intake level · deep learning · DVSGFS · feature selection · impact rate · and VAE-GRN2

1 Introduction

Due to restricted food production, meeting everyone's requirement for food items (e.g., vegetables, fruits, milk, wheat, etc.) has become a big difficulty. Food availability referred to food supply both nationally and domestically. Individuals and households have access to food under the entitlement system, even if they do not have enough money to purchase nutritious food. There are several author-defined meanings of food security that all convey the same concept. Food security refers to the situation in which everyone has enough nutritious food to live a healthy and active life [1].

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026
S. Lanka et al. (Eds.): ICTSM 2025, LNNS 1755, pp. 544–559, 2026.
https://doi.org/10.1007/978-3-032-13177-5_42

Children are targeted and made vulnerable by food marketing. One of the top global priorities for lowering obesity, dietary risks, and the prevalence of communicable diseases is shielding kids from deceptive food marketing. Children's perceptions of food and food brands, as well as their consumption, weight, and purchase behavior, are all influenced by their exposure to food marketing [2]. The reach, frequency, and persuasive power of marketing which include both design and content determine its influence. Additionally, the majority of developed nations are capitalist democracies where corporate profits and freedom of speech frequently trump the health of children [3]. However, many nations have enacted strict rules over the past few years that restrict certain aspects of food marketing to children.

However, obstacles remain when developing preventative strategies or treatments for a specific location. For instance, parental feeding habits significantly influence childhood obesity. Therefore, if a community aims to establish a tailored policy or intervention that best fits the local context, determining which indicator best represents a general element is challenging [4].

The immune systems of children are more susceptible to illnesses than those of adults. Because malnourished children's immune systems are compromised, they are more likely to die. On the other hand, little is known about the underlying reasons of immune function impairment and how hunger is related to it. Nutritional therapies are thought to be economical methods of treating or avoiding malnutrition [5]. This is especially crucial for youngsters at risk of problems linked to their dietary habits. Health, besides childhood health, has a significant impact on their physiological development.

1.1 Contribution of Work

- The main contribution of this section is the proposed VAE-GRN2 model for predicting children's calorie and nutrient intakes in temporal dietary patterns.
- The CSN algorithm is used to effectively handle errors, duplicates, and missing data in the dietary datasets.
- The CNIN model was developed to calculate the impact ratios of different dietary intakes of specific nutrients.
- The DVSGFS technique is proposed to select optimal and informative features to capture temporal biases and improve prediction accuracy.
- The proposed technique contributes to better health monitoring and nutrition planning by demonstrating improved performance in improved dietary status prediction in children.

1.2 Objective of the Work

By integrating advanced data pre-processing, feature selection, and nutrient impact modeling techniques, VAE-GRN2 develops a deep learning framework to predict children's caloric and nutrient intake accurately.

1.3 Motivation of the Work

- The CSN algorithm was provided to analyze incomplete, duplicate, and incorrect dietary data.

- Enhance the understanding of how various nutrient intakes affect children by modeling the ratios of nutrient impacts using the CNIN.
- Enhance feature selection to facilitate time analysis using the DVSGFS technique.
- Finally, enhance predictive power with the VAE-GRN2 model in measuring the dietary condition of children.

2 Literature Survey

Two individual and household-based categorization systems were applied in this work to determine the performance of four Machine-Learning (ML) algorithms in predicting undernutrition [6]. The analysis process involves ANN and K-Means clustering types of ML analysis [7]. The present data set of this study was based on provincial general hospital (RSUP) statistics of cases of child nutrition in the previous year.

The Decision Tree (DT) algorithm and the K-means clustering analysis tool were employed by the study [8] to investigate the relationship between vitamin and mineral consumption and anaemia. In determining the correctness of the ML model, the data was balanced and regularized using an ensemble-weighted clustering approach. Linear regression was used to perform a correlation analysis between Body Mass Index (BMI) and ingestion of live microbes. The Logistic Regression (LR) was used to analyses the relationship between the existence of abdominal obesity and consumption of live microbes [9].

A sample of 2013 children was selected, and information about all the mentioned characteristics was provided [10]. To predict the state of childhood anemia, we applied multiple machines learning methods, including the support vector machines (SVM) approach. The Random Forest (RF) model outperformed all other machine learning models in detecting child nutritional deficiencies. However, these studies, which commonly employ logistic and multilevel models, were unable to identify the most significant predictors [11].

The study process comprises analyzing the system and designing machine learning applications utilizing the Multiple Linear Regression technique. The developed approach may then be used to forecast the nutritional status of children in Aceh in a timely, precise, and accurate manner. Furthermore, among the five machine learning methods, the xgbTree approach outperforms the generalized linear mixed model [12].

The machine learning algorithms range from basic ones, e.g., Logistic Regression and K-Nearest Neighbor, to more contemporary ML models [13]. For the most part, accuracy scores range significantly higher or lower than the 75% baseline accuracy, where the best LSTM-FC models achieved 91%. This validates the use of LSTMs for modelling time-series data. Most studies suffered from the limitation of self-reported or cross-sectional data, limiting the ability to make causal inference and reducing the validity.

The paper [14] introduces a novel probabilistic forecasting technology, GRU, to foresee the deformation of landslides on a formative basis. The novel [15] aimed to minimize the effect of race on models derived from data using Adversarial Variational Autoencoders (AVAEs). However, race is often excluded as an input variable because it is highly correlated with many other variables. Introduces [16] a conditional GAN that produces realistic, shape-constrained food images with known shapes/containers, which

has direct applications in augmenting food-image datasets used to predict portion size and thus intake. Empirically compare GAN methods with other methods on structured clinical data [17]. The model is useful when selecting GAN variables for intake levels (continuous nutrients, counts, and categorical food types).

3 Proposed Methodology

The proposed methodology starts with collecting children's nutritional intake information, which can include questionable, incomplete, and inconsistent records. To mitigate this issue, the first part of the methodology is to implement a CSN algorithm as a preprocessing action to identify and clean the questionable records, tackle missing information, and eliminate duplicates. The selection of the proposed generative method was based on its ability to adequately model the high-dimensional, complex, and often missing food intake data found in children. Traditional statistical tools have weaknesses in their ability to reflect nonlinear relationships among different food types, portion sizes, and demographic or behavioral variables. In contrast, generative models such as VAEs and GANs can capture the hidden structure of food data and learn latent representations that generate realistic synthetic models. This can be advantageous when the datasets used in dietary intake studies are small, biased, or limited by privacy considerations (especially in the case of studies of children). Reducing biases across population groups allows for extrapolation of small child nutrition datasets. It provides a proper scaling solution to simulate realistic intakes across a range of health conditions and environmental settings.

As shown in Fig. 1 the data to ensure the model receives the best/cleanest input data. Next, the CNIN will determine nutrient impact ratios to quantify the nutrient make-up of food items and eating patterns over the food record period. The DVSGFS method will be used to select the most relevant and informative time-related features from the dataset to increase prediction performance and lower model complexity. The selected features will be fed into the "core" predictive model, the VAE-GRN2. This hybrid model will use the generative properties of the VAE to generate latent dietary patterns from the data. At the same time, the GRN2 captures the sequential learning associated with incorporating time while modeling the nutritional intake data.

3.1 Dataset Description

The nutrition data set is a finished statistical summary of information regarding the nutritional content of different food materials. It contains numerous types of information, including amounts of macro- and micronutrients, calories, vitamins, minerals, and other relevant nutritional properties. All the food products enable researchers, nutritionists, and health enthusiasts to study them, decode their nutritional value, categorize eating-style habits, develop personalized diets, and investigate the dietary effects on overall well-being and health. It is a remarkable source for rigorous research, imagining new nutritional approaches using tools, and making rational decisions in nutrition and dietetics, as the database is both gigantic and accurate. This dataset contains 8788 observations and 76 features.

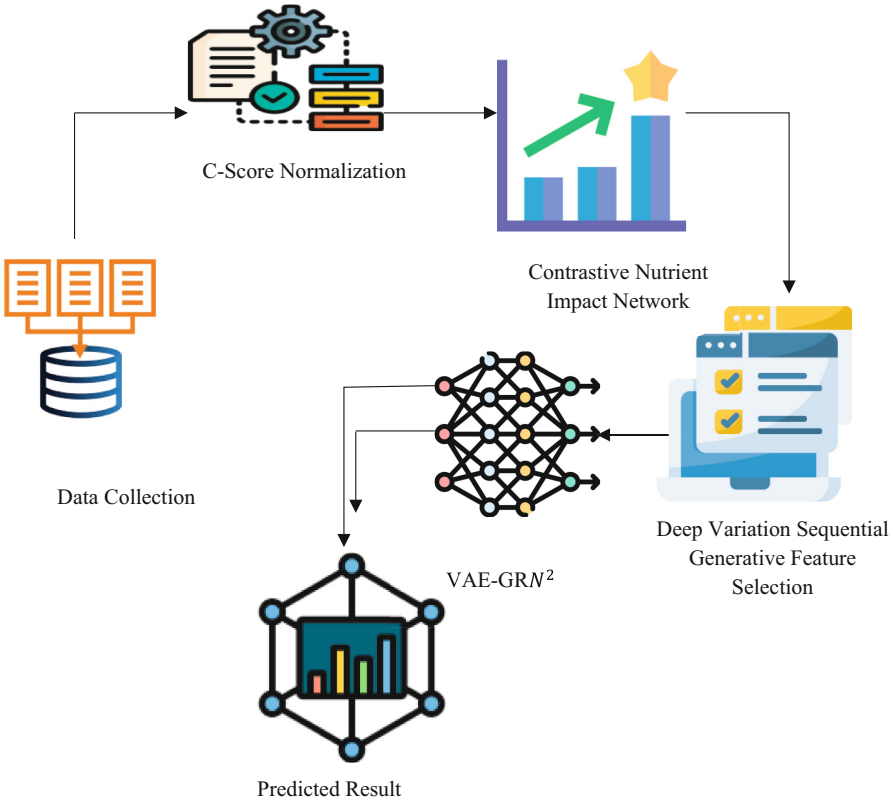


Fig. 1. Architecture Diagram of the Proposed Method

		name	serving_size	calories	total_fat	saturated_fat	cholesterol	sodium	choline	folate	folic_acid
0	Cornstarch	100 g	381	0.1g			0	9.00 mg	0.4 mg	0.00 mcg	0.00 mcg
1	Nuts, pecans	100 g	691	72g	6.2g		0	0.00 mg	40.5 mg	22.00 mcg	0.00 mcg
2	Eggplant, raw	100 g	25	0.2g			0	2.00 mg	6.9 mg	22.00 mcg	0.00 mcg
3	Teff, uncooked	100 g	367	2.4g	0.4g		0	12.00 mg	13.1 mg	0	0
4	Sherbet, orange	100 g	144	2g	1.2g	1mg		46.00 mg	7.7 mg	4.00 mcg	0.00 mcg
5	Cauliflower, raw	100 g	25	0.3g	0.1g		0	30.00 mg	44.3 mg	57.00 mcg	0.00 mcg
6	Taro leaves, raw	100 g	42	0.7g	0.2g		0	3.00 mg	12.8 mg	126.00 mcg	0.00 mcg
7	Lamb, raw, grounc	100 g	282	23g	10g	73mg		59.00 mg	69.3 mg	18.00 mcg	0.00 mcg
8	Cheese, camembe	100 g	300	24g	15g	72mg		842.00 mg	15.4 mg	62.00 mcg	0.00 mcg
9	Vegetarian fillets	100 g	290	18g	2.8g		0	490.00 mg	82.0 mg	102.00 mcg	0.00 mcg

Fig. 2. Nutrition Dataset

The Nutrition Dataset (Fig. 2) provides a clear representation of the nutritional parameters analyzed. These dietary components may include various macronutrients (e.g., carbohydrates, proteins, and fat), micronutrients (e.g., vitamins and minerals), calories, as well as any other nutrition variables that may be handy in the provision of the meaning of what has been taken and the nutritional condition.

3.2 Pre-processing C-Score Normalization

To accurately estimate the level of nutrient intake in children, it is necessary to process the raw dietary data as strictly as possible, minimizing discrepancies, unavailable data, and data scaling errors. A typical approach to normalizing raw nutrient intake is to convert it to a C-score that can be compared with a person's intake and a specific dimension. Observed nutrient intake is compared to amounts of RDI, and then individual variability is corrected using the C-score normalization method. In the following section, representative equations are provided to clarify the normalization process:

$$Z_i = \frac{X_i - \mu}{\sigma} \tag{1}$$

where X_i the observed intake of nutrient is μ is the mean intake across the dataset, and σ is the standard deviation, such that the data can now be uniformly scaled. The RDI ratio is then calculated as:

$$R_i = \frac{X_i}{RDI_i} \tag{2}$$

With RDI representing the Recommended Dietary Intake for nutrient based on age, gender and physical activity. To evaluate how much a person's intake exceeds, not only the recommended amount, but the absolute difference is calculated:

$$D_i = |X_i - RDI_i| \tag{3}$$

This assists in detecting under- or over-consumption. The crux of the normalization is the C-score itself, defined as:

$$C_i = \frac{X_i - RDI_i}{\sigma} \tag{4}$$

In Eq. (4) X_i is the true observed intake of nutrient by the child, RDI is the Recommended Dietary Intake for nutrient I that is based on the child's age, gender and activity level, and σ is the standard deviation in that nutrient's intake among the population. This formula portrays how many standard deviations a child's intake is above or below the recommended level to objectively identify under- or over-nutrition.

$$C_{total} = \frac{1}{n} \sum_{i=1}^{n} C_i \tag{5}$$

Computes the composite C-score that is achieved by averaging all the individual nutrient C-scores. C_{total} is the overall dietary balance value of the child, C_i is the normalized value of the C-score of individual nutrients i, and n is the number of nutrients. This cumulative score takes a complete perspective of the child's nutritional intake relative to the dietary recommendations that may be utilized as an input to health prediction models by the use of ML.

3.3 Contrastive Nutrient Impact Network (CNIN)

The CNIN is a new DL model that leverages contrastive representation learning to assess the impact rates of food products on a nutritional profile. In the case of a deployed nutrition dataset containing 77 nutritional attributes, CNIN aims to learn the nutritional embedding by distinguishing between the rates of nutritional health impacts caused by food items rated as close and those rated as dissimilar. First, foods are encoded as high-dimensional embeddings by sharing the same encoder network.

$$S_{i,j=}C_{total}\begin{cases}1, \ if\,|y_i - y_j| < \epsilon \\ 0, \ otherwise\end{cases} \tag{6}$$

where y_i and y_j are the true impact rates for items i and j, and ϵ is a small threshold that determines the degree of similarity. The cosine similarity between two latent embedding's z_i and z_j is calculated as follows:

$$sim\big(z_i, z_j\big) = \frac{z_i . z_j}{||z_i||.||z_{j||}} \tag{7}$$

The similarity score is then used in the NT-Xent contrastive loss which promotes similar pairs to be near each other in embedding space and dissimilar pairs to be farther apart. Once embeddings are learned, a regression head g $(\cdot)$ maps the latent embedding to an estimated impact rate:

$$y_i = g(z_i) \tag{8}$$

The difference between the predicted impact y_i and true value is minimized using the mean squared error loss.

$$\mathcal{L}_{regression} = \frac{1}{N}\sum\nolimits_{i=1}^{N}(y_i - y_i)^2 \tag{9}$$

The regression loss function in the equation represents an average of all measured squared differences between the actual rate and the impact y_i the predicted value, it is essential to note that N is the total number of samples, and each pair of values.

3.4 Deep Variation Sequential Generative Feature Selection (DVSGFS)

In predicting nutrient intake levels, the Deep Variation Sequential Generative Feature Selection (DVSGFS) technique identifies and selects the most informative nutrients by modelling their temporal dependencies and contributions to predictive performance. DVSGFS leverages a variation Autoencoders framework with a sequential generative model to capture the complex patterns in children's dietary intake over time.

$$\mathcal{L}_{ELBO} = E_{q\varnothing}(z|x)\Big[log_{p\theta}(x|z)\Big] - KL(q_{\varnothing}(z|x)||p(z)) \tag{10}$$

where x Dietary features sequence input (e.g., vector of daily nutrient intake for child), z: Latent variable that captures compressed, useful features $log_{p\theta}(x|z)$ Decoder (generative)

model that decodes and reconstructs nutrient intake from latent space $q_{\varnothing}(z|x)$ Encoder (inference) model that estimates latent variable distribution *KL*: Kullback-Leibler divergence regularizes the latent space.

$$\mathcal{L}_{seq} = \sum_{t=1}^{T} [E_{q\varnothing}(z_t|x \leq t)\big[logp(xt|zt)\big] - KL(q(zt|x \leq t)||p(z_t))] \tag{11}$$

where *xt* Nutrient intake at time step $x \leq t$ Nutrient intake history up to time, z_t Latent feature representation at time, captures temporal dependencies in nutrient intake and dynamically selects features across the sequence.

$$\mathcal{L}_{sparsity} = \lambda \sum_{i=1}^{d} \pi_i \tag{12}$$

where d is the total number of nutrients, π_i is the selection probability for nutrient i, and λ is the hyper parameter that controls the level of sparsity, Encourages the model to choose fewer but more informative nutrients, discouraging over fitting.

3.5 Variational Autoencoder-Gated Recurrent Neural Network (VAE-GRN2)

In this section, the proposed VAE-GRN2 model is used to predict children's caloric and nutrient intake. Furthermore, users can get customized nutrient intake based on data collected from the dataset through the VAE-GRN2 model. A differential autoencoder network processes and analyzes input information such as weight, height, and age, and then estimates the user's dietary requirements and informational features in latent space to optimize the input information into a generated feature representation. The VAE-GRN2 is a model that can forecast intake and generate a sequence of children's food, as well as their nutrient intake levels. The created nutrient intake plan can be modified according to the nutrient sizes and characteristics obtained from the database, which is classified by nutrient intake, energy, nutrition, and the user's specific needs.

As indicated in Eq. 13, a recurrent neural network generates a set of nutrients, and the hidden vectors are used as input to calculate the daily nutrient plan, which effectively predicts the nutrients level. Let's assume $h(t)-$ hidden state at time, $(t)-$ time, $GRN2-$ gate recurrent neural network, $Z-$ latent vector.

$$h(t) = \mathcal{L}_{sparsity} \begin{cases} GRN2(Z) & for\, t = 1 \\ GRN2(h(t-1)) & for\, t > 1 \end{cases} \tag{13}$$

Calculate the estimated nutrient intake class for the previous latent state GRN2according to Eq. 14 and 15. Let's assume f_m- fully connected layer, $O(t)-$ output class of nutrient time, $\hat{y}_t-$ predicted nutrient classes,

$$O(t) = Soft_{Mas}(f_m(h(t))), t = [1,6] \tag{14}$$

$$\hat{y}_t = \underset{c}{argmax}(O(t)), t = [1,6] \tag{15}$$

The most suitable class for a food product is identified by manipulating the class probability as shown in Eq. 16 and 17. It uses a fully connected layer to calculate and

predict the nutrient level, total energy, and nutritional value. Let's assume $\hat{n}-$ nutrient values $\widehat{E}_I-$ predict the total energy, $f_{EI}-$ energy fully connected layer

$$\widehat{E}_I = \sum_{t=1}^{T} f_{EI}(h(t)) \tag{16}$$

$$\hat{n} = \sum_{t=1}^{T} f_{nutr}(h(t)) \tag{17}$$

Calculate the deviation between the predicted caloric intake of the nutrient plan and the mean square error, as shown in Eq. 18. Let's assume $L_{EI}-$ energy intake loss

$$L_{EI} = \frac{1}{N} \sum_{i=1}^{N} \left(EI - \widehat{E}_I\right)^2 \tag{18}$$

The total loss due to the proposed VAE-GRN2 nutrient intake level children prediction system is shown in Eq. 19. Let's assume $L-$ loss.

$$L = L_{MC} + L_{KID} + L_{EI} + L_{Nutri} \tag{19}$$

User groups or clusters are formed using the nutrient intake level prediction dataset with a latent location structure. A proposed food recommendation system with different food programs significantly determines food preferences.

4 Results and Discussion

In this section, the performance of these methods is compared with the proposed VAE-GRN2 method to evaluate previous XGBoost, SSO-FNN, SmGAN, and SRCFA-NBGA methods under different conditions and constraints. The dataset's nutrient data collection functions allow adding records for various nutrient types. Furthermore, four metrics, such as the confusion matrix and multiple parameters, such as precision, accuracy, recall, F1 score, AUC-ROC curve, time complexity, and error rate, are utilized to evaluate the proposed technique.

Table 1. Simulation Parameter

Simulation	Variable
Dataset Name	Nutrients Dataset
No of Dataset	8790
Language	Python
Tool	Jupyter
Training	7,312
Testing	1478

Table 2. Performance of Sensitivity and Specificity

Methods	Sensitivity	Specificity
XGBoost	72	75
SSO-FNN	77	80
SmGAN	83	86
SRCFA-NBGB	87	89
VAE-GRNN	91	93

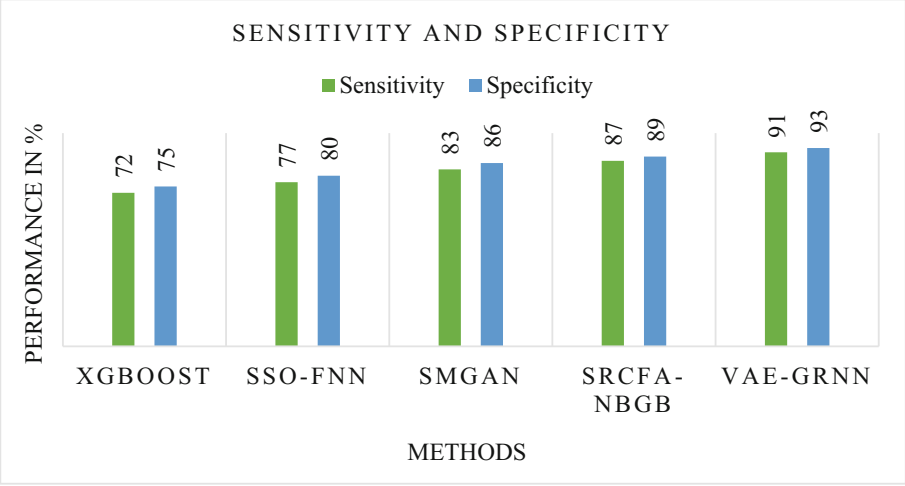


Fig. 3. Analysis of Sensitivity and Specificity

Simulated parameter variables can be used to train and test features of the nutrient dataset collected from Kaggle using train and test methods. In addition, the dataset, which has 8790 records, can be analyzed using the Jupyter tool in Python, as indicated in Table 1.

The proposed approach is compared with earlier techniques through sensitivity and specificity analysis in terms of the predictability of children's nutritional intake. Additionally, the efficacy of the proposed VAE-GRN2 approach has been demonstrated by a sensitivity rate of 91 percent in improving the nutritional intake of children. Likewise, the sensitivity of the proposed VAE-GRN2 approach of the nutritional intake of children is at 93 percent, as illustrated in Fig. 3 and Table 2.

The new technique will be compared with previous approaches in terms of precision analyses that aim to predict the nutritional needs of children. Additionally, the VAE-GRN2 model achieves a precision rate of 91 percent for this intake. Comparatively, other former approaches to this issue, such as XGBoost, SSO-FNN, SmGAN, and SRCFA-NBGB, demonstrate precision levels of 79%, 83%, 86%, and 88%, respectively, as shown in Fig. 4 and Table 3.

Table 3. Performance of Precision

No of Records	XGBoost	SSO-FNN	SmGAN	SRCFA-NBGB	VAE-GRN^2
2197	67	70	75	79	82
4,394	72	74	79	83	87
6,591	76	79	82	85	89
8,790	79	83	86	88	91

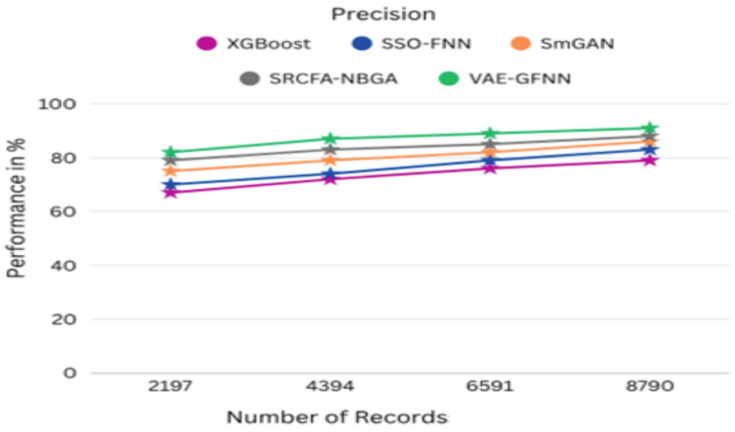


Fig. 4. Analysis of Precision

Table 4. Performance of Recall

No of Records	XGBoost	SSO-FNN	SmGAN	SRCFA-NBGB	VAE-GRN^2
2197	69	72	77	81	83
4,394	73	77	81	85	88
6,591	77	80	84	87	90
8,790	80	85	89	92	94

An overview of standard procedures was also conducted to compare the new method and determine its effectiveness. The analysis focused on comparing the recall of each technique regarding its potential to predict nutrient intake in children. The results of the study suggest the recall of each method, which are presented in Fig. 5 and Table 4. The earlier approaches in this comparison were XGBoost, SSO-FNN, SmGAN, and SRCFA-NBGB, which achieved corresponding recall rates of 79 percent for XGBoost, 83 percent for SSO-FNN, 86 percent for SmGAN, and 88 percent for SRCFA-NBGB.

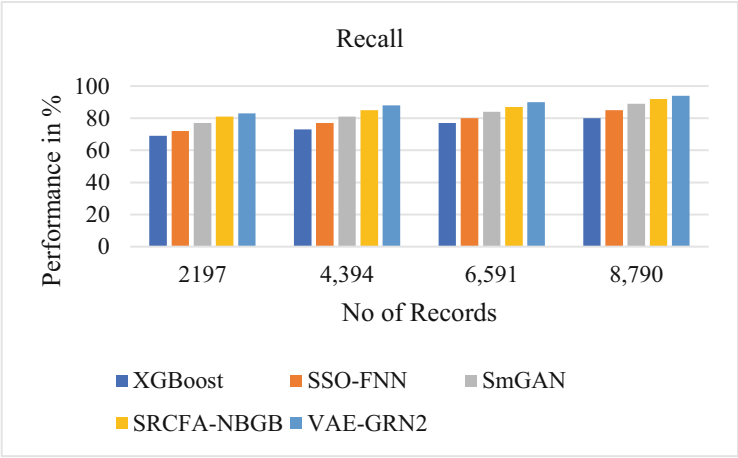


Fig. 5. Analysis of Recall

These statistics demonstrate the effectiveness of these developed methods in forecasting nutritional intake in children. The VAE-GRN2 achieved a recall score of 91% in predicting the nutrition consumed by children, outperforming other methods tested.

Table 5. Performance of F1-Score

No of Records	XGBoost	SSO-FNN	SmGAN	SRCFA-NBGB	VAE-GRN^2
2197	72	76	79	83	85
4,394	75	79	84	86	89
6,591	79	83	87	89	92
8,790	82	87	90	93	95

The accuracy of the newly developed method was compared to that of several previous techniques through a comparative analysis—the review aimed to evaluate the F1-score of each method in predicting nutrient intake in children. The results of the following analysis are presented in Fig. 6 and discussed in Table 5, highlighting the F1-scores obtained by each method. The strategies used in the comparison above were XGBoost, SSO-FNN, SmGAN, and SRCFA-NBGB, with F1-score rates of 82% for XGBoost, 87% for SSO-FNN, 90% for SmGAN, and 93% for SRCFA-NBGB. These figures suggest that the established methods are effective in predicting nutritional intake in children. The VAE-GRN2 achieved a 95 percent F1-score, which was used to indicate the nutritional intake of children compared to other tested methods.

This paper focused on evaluating the correctness of each belongingness in forecasting the dietary intake of children. The outcomes of the analysis undertaken indicate the accuracy scores of each method and are tabulated in Table 6 and displayed in Fig. 7. These statistics demonstrate that these already established methods are effective in predicting

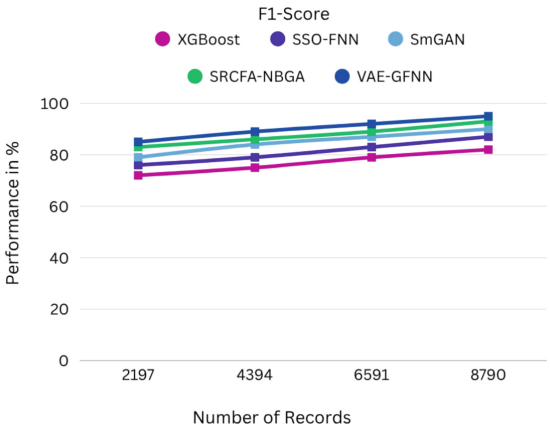


Fig. 6. Analysis of F1-Score

Table 6. Performance of Accuracy

No of Records	XGBoost	SSO-FNN	SmGAN	SRCFA-NBGB	VAE-GRN^2
2197	75	79	82	85	88
4,394	79	84	86	89	92
6,591	83	87	90	93	94
8,790	87	91	93	95	97

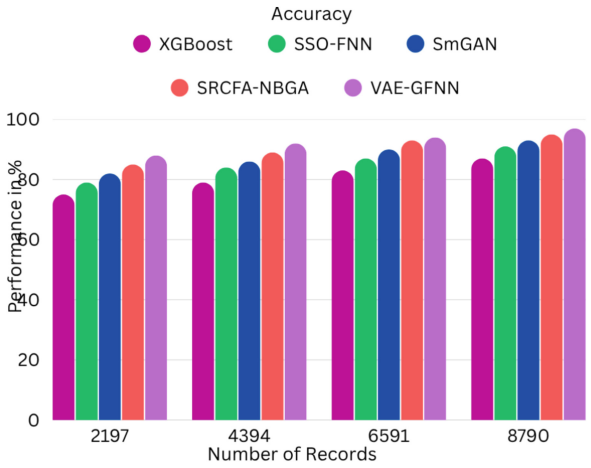


Fig. 7. Analysis of Accuracy

nutrient intake in children. In comparison to other methods of assessment, VAE-GRN2

was accurate in assessing the nutrient food intakes of children at 97%. In the above comparisons, XGBoost, SSO-FNN, SmGAN, and SRCFA-NBGB are the previous methods with an accuracy of 87%, 91%, 93%, and 95%, respectively.

Table 7. Performance of Time Complexity

No of Records	XGBoost	SSO-FNN	SmGAN	SRCFA-NBGB	VAE-GRN2
2197	49	47	44	40	35
4,394	44	43	41	37	31
6,591	41	40	36	33	23
8,790	37	38	34	27	18

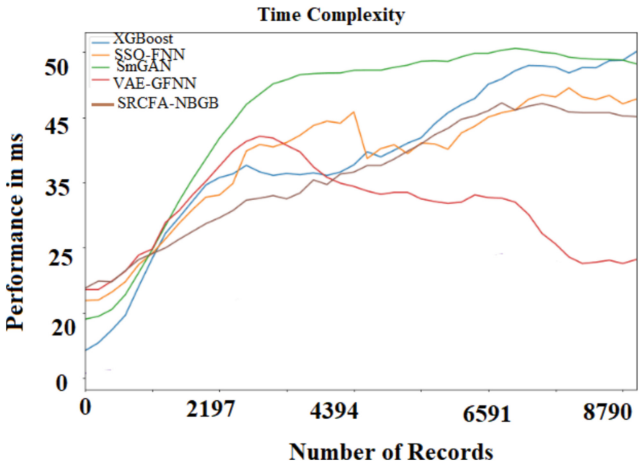


Fig. 8. Analysis of Time Complexity

They analyzed how accurately various methods of nutrient intake by children are predictive. The results of the tests are presented in Fig. 8 and summarized in Table 7, serving as an indicator of the time complexity of both methods. This data underscores the effectiveness of these established approaches in assessing children's nutrient intake. VAE-GRN2 achieved a time complexity of 18 ms, outperforming other assessment methods. In this comparison, the previous methods are XGBoost with 37 ms, SSO-FNN with 38 ms, SmGAN with 34 ms, and SRCFA-NBGB with 27 ms time complexity.

5 Conclusion

In conclusion, the proposed framework effectively addresses the challenges of malnutrition analysis in children by incorporating data-driven advanced deep learning and feature selection techniques. A new deep learning-based architecture, VAE-GRN2, is presented

in the given research, which can help effectively predict caloric and nutrient values in children. The model is a combination of the VAE-GRN2 model, which effectively captures the temporal relationships in dietary data. Furthermore, the CSN algorithm was proposed to systematically detect and fix errors, duplicates, and missing values in the dataset to ensure high-quality input data. Moreover, we constructed the CNIN to determine the ratio of the effect of different nutrient profiles, which would provide further insights into the contribution of particular nutrients to the diet. Additionally, we adopted the DVSGFS method to increase model efficiency and performance, which will allow us to select the most informative features to predict nutrients. The offered framework shows great potential in the process of assisting with the nutritional monitoring and personalized diet planning, as well as interventions concerning the child nutrition at the population level. Finally, enhance predictive power to determine the status of children's diet using the VAE-GRN2 model.

All of this leads to the conclusion that the proposed innovative advances made in this research come together for a highly accurate and efficient systematic approach to accurately define a child's dietary behavior with a full complement of nutrition policies and guidelines accomplished hence finalities to let of 97% across all key performance measures of successful pedagogical methodology. Precision, recall, F1-score, and time Complexity present strong potential to change even modest healthcare systems if child nutrition monitoring and discerning decision support methodological frameworks could become normalized.

References

1. Sood, S., Singh, H.: Computer vision and machine learning based approaches for food security: a review. Multimed. Tools Appl. **80**, 27973–27999 (2021)
2. Kelly, B., Backholer, K., Boyland, E., et al.: Contemporary approaches for monitoring food marketing to children to progress policy actions. Curr. Nutr. Rep. **12**, 14–25 (2023). https://doi.org/10.1007/s13668-023-00450-7
3. Olstad, D.L., Boyland, E.: Towards effective restriction of unhealthy food marketing to children: unlocking the potential of artificial intelligence. Int. J. Behav. Nutr. Phys. Act. **20**, 61 (2023). https://doi.org/10.1186/s12966-023-01458-6
4. Shi, X., et al.: An ensemble-based feature selection framework to select risk factors of childhood obesity for policy decision making. BMC Med. Inform. Decis. Making **21**, 1–13 (2021)
5. Morales, F., Montserrat-de la Paz, S., Leon, M.J., Rivero-Pino, F.: Effects of malnutrition on the immune system and infection and the role of nutritional strategies regarding improvements in children's health status: a literature review. Nutrients **16**(1), 1 (2023). https://doi.org/10.3390/nu16010001. PMID: 38201831; PMCID: PMC10780435
6. Siy Van, V.T., et al.: Predicting undernutrition among elementary schoolchildren in the Philippines using machine learning algorithms. Nutrition **96**, 111571 (2022). https://doi.org/10.1016/j.nut.2021.111571
7. Yanto, M., Hadi, F., Arlis, S.: Determination of children's nutritional status with machine learning classification analysis approach. Indon. J. Electr. Eng. Comput. Sci. (IJEECS) **33**(1) (2024). https://doi.org/10.11591/ijeecs.v33.i1.pp303-313. p-ISSN: 2502-4752, e-ISSN: 2502–4760

8. Qasrawi, R., Badrasawi, M., Al-Halawa, D.A., et al.: Identification and prediction of association patterns between nutrient intake and anemia using machine learning techniques: results from a cross-sectional study with university female students from Palestine. Eur. J. Nutr. **63**, 1635–1649 (2024). https://doi.org/10.1007/s00394-024-03360-8
9. Chen, Y., Ni, H., Zhang, H.: Exploring the relationship between live microbe intake and obesity prevalence in adults. Sci. Rep. **14**(1), 21724 (2024)
10. Rahman Khan, J., Chowdhury, S., Islam, H., Raheem, E.: Machine learning algorithms to predict the childhood anemia in Bangladesh. J. Data Sci. **17**(1), 195–218 (2022)
11. Gebeye, L.G., Dessie, E.Y., Yimam, J.A.: Predictors of micronutrient deficiency among children aged 6–23 months in Ethiopia: a machine learning approach. Front. Nutr. **10**, 1277048 (2024). https://doi.org/10.3389/fnut.2023.1277048
12. Ula, M., et al.: Application of machine learning in predicting children's nutritional status with multiple linear regression models. Multica. Sci. Technol. **2**(2), 42–48 (2022)
13. Bitew, F.H., Sparks, C.S., Nyarko, S.H.: Machine learning algorithms for predicting undernutrition among under-five children in Ethiopia. Public Health Nutr. **25**(2), 269 (2021). https://doi.org/10.1017/S1368980021004262
14. Cai, W., Fuan, L.: Generative probabilistic prediction of precipitation induced landslide deformation with variational autoencoder and gated recurrent unit. Front. Earth Sci. **24** (2024). https://doi.org/10.3389/feart.2024.1394129. Sec. Geohazards and Georisks, Volume 12 - 2024
15. Sarullo, K., Swamidass, S.J.: Understanding and mitigating the impact of race with adversarial autoencoders. Commun. Med. **4**, 210 (2024). https://doi.org/10.1038/s43856-024-00627-3
16. Chen, G., Mao, Z., Sun, M., Liu, K., Jia, W.: Shape-preserving generation of food images for automatic dietary assessment. In: CVPR Workshop/arXiv preprint (2024)
17. Ahmed, H.A., Nepomuceno, J.A., Vega-Márquez, B., et al.: Synthetic data generation for healthcare: exploring generative adversarial networks variants for medical tabular data. Int. J. Data Sci. Anal. (2025). https://doi.org/10.1007/s41060-025-00816-w



A Computational Approach for Framing Place Rebranding, Architecture and Public Policy: The Case of Skopje 2014 Project

Sabina Kohlmann[1], Anastasios Fountis[1(✉)], Katja Udir Misic[2], Navya Gubbi Sateeshchandra[1], and Christos Lemonakis[3]

[1] Berlin School of Business and Innovation, Berlin, Germany
{sabina.kohlmann,navya.gubbi}@berlinsbi.com,
a.fountis@post.harvard.edu
[2] University of Maribor, Maribor, Slovenia
katja.misic@um.si
[3] Hellenic Mediterranean University, Agios Nikolaos, Crete, Greece
lemonakis@hmu.gr

Abstract. The capital cities have always been the places where political power has been exercised through urban transformation. In this paper, the authors explore two undertakings in Skopje: the post-1963 inner-city renewal and the *Skopje 2014* project. In contrast to the previous framings of the project, authors argue that the place brand authors in both periods applied the same practice, namely architecture, to rebrand the city. Drawing from the theory on corporate rebranding, the strategic importance of the cities, and research findings, the authors further the academic codification of the place rebranding concept. Namely, while the changes in the urban vocabulary and the management of the places throughout history were not referred to as *rebranding*, the authors argue that the applied policymakers' practices were the same to a certain extent. Additionally, as the current public policy regarding treating *Skopje 2014* lacks a clear framework, the authors add to the practical application in the urban management of the (capital) cities when handling the inherited *unwished-for past* and suggestions for overcoming these.

The paper accentuates participatory marketing and branding in public policy creation and also includes also the usage of a computational model with differential equations and a Python program for framing the interaction between place rebranding, architectural transformation, public policy, citizen participation and national identity narrative.

Keywords: place rebranding · architecture · capital city · urban management · computation · computational humanities · public policy

1 Introduction

During the first quarter of the twenty-first century Skopje, the capital city of North Macedonia experienced the most significant urban upheaval after the post-1963-earthquake renewal. Following the election promises made in 2016 and 2017, the national and local

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026
S. Lanka et al. (Eds.): ICTSM 2025, LNNS 1755, pp. 560–580, 2026.
https://doi.org/10.1007/978-3-032-13177-5_43

governments obliged to resolve the issue on the *Skopje 2014* project treatment. We will analyze the socio-historic background of the rebranding of post-1991 Skopje, particularly *the Skopje 2014* project, and, building on the centuries-long practices of architectural vocabulary treatment by the policymakers, to propose the treatment and the course of action after the top-down urban upheaval that the city experienced in the first quarter of the twenty-first century. We will use a meaning-based approach to case study analysis to elicit and compare brand associations, reveal potential overlapping in the policymakers' practices in both researched periods, and analyze the project's impact on the place branding image. Furthering the academic codification of the place rebranding concept whilst adding to the practical application and public policy framework suggestions in the urban management of the capitals regarding the treatment of the inherited *unwished-for past* in urban settings adds to the paper's originality. The paper aims to answer the following questions:

RQ1: Are there overlapping brand associations and practices applied in the 1963-post earthquake Skopje and the *Skopje 2014* project?
RQ2: Are there alternatives to the centuries-long policymakers' practice of forgetting, demolishing, and removing the unwished-for and inherited architectural vocabulary in capital cities?
RQ3: How can a new place brand image of Skopje be managed after the *Skopje 2014* project? How can a computational framework with algorithms for comparing 2 historical phases can provide a decision making context?

This article is organized as follows: We list the theoretical background, followed by the methodology used and the sociohistorical complexity of Skopje and its architectural vocabulary before 1990 and after. In the final sections, we detail the research findings and we simulate them in the context computational humanities, discuss them, build theory and apostrophize the practical implications for policymakers and the treatment of the *Skopje 2014* project. Finally, we recapitulate the article in a comprehensive conclusion and list the sources used and provide appendices for the computation.

2 Theoretical Framework

The Strategic Importance of a Capital City
As *windows to the world* for global audiences, capital cities communicate culture, political values, and national identity, most expressively through architectural styles inherited from past periods (Cvitkovic 2023; Muratovski 2013) [14, 47]. More specifically, politics and ideology become visually expressive in the urban environment of a (capital) city. Meisel (1993) argues that "capital cities are an important index to the dominant political values of their countries" (p. 4) [45]. Furthermore, Diener and Hagen (2013) [21] suggest that the continuous re-making of urban landscapes is a political undertaking linked with competing efforts among various cultural and political movements to embed specific ideologies, especially narratives of national identity. However, the role of capital cities is changing rapidly, with a majority of these becoming more relevant than the state and rarely being the proxy of the nation. For example, Campbell (2000) [11] argues that the capital city, due to its position as a center of economic and cultural

activity, is often exceptional and unrepresentative of the nation's identity. On the other side, the most recent cycle of globalization has seen the rise of specialized and globally aspirational, higher-income cities competing in more dynamic and fast-moving global industries that seek to leverage their efficient infrastructure, improved quality of life, and better security and environmental performance compared to the larger megacities with examples such as Brisbane, San Diego, Shenzhen, Santiago de Chile, and Stockholm (Clark 2016) [12].

Decoding the Architectural Vocabulary of the Capital Cities

Historically, architecture has been used to promote the state's power (Beck 1998; Cvitkovic 2023; Cvitkovic and Kline 2017; Diekmann and Cloquet 2012; Muratovski 2019) [3, 14, 15, 22, 47]. Policymakers and architects are a privileged class in selecting and constructing meanings in the urban setting who impose their agenda and messages using public space (Ergut 1999; Diekmann and Cloquet 2012) [25, 22]. For example, Lasswell (1979) defines political power as giving and receiving support for decisions that significantly affect the entire community's values. However, the politics of architectural vocabulary presume a significant politico-ethical responsibility as there is much more at stake in design than it may appear (Cvitkovic 2023; Kaika 2010; Lasswell 1979; Žižek 2009). This is why Beck's (1998) depiction of architecture as "politics with bricks and mortar" (Beck 1998, p. 115) does not come as a surprise [3]. Whilst many urban undertakings are top-down imposed, Viswanathan (2023) emphasizes the importance of policymakers' bottom-up approach and consideration of larger community and context. Moreover, Peterson and Godby (2020) suggest that citizens have the competence to participate in governance through the thin-participation methods [54].

Post-socialist societies are today confronted with the problem of commemorating the past and simultaneously reconstructing public space due to political, economic, and societal changes (Leach 1999; Mariotti and Pencic 2015). While interpretations of the past are always politically *colored*, historical revisionism is often becoming political battleground). Post-socialist approach to urbanism varies across the political landscape; autocratic or oligarchic regimes possess substantial authority over urban planning, architecture, and iconography, whilst more democratic societies are impacted by various commercial interests, social movements, and legal constraints (Diener and Hagen 2013).

The Art of Place Rebranding and Reconstructing Identities

A brand can be seen as having a set of (brand) associations created by the brand author that an entity is communicating through a variety of channels and ways that consumers, after the experience and/or contact with a brand, are connecting with the brand in question. Today, branding strategies are being applied across different disciplines like corporate, personal, political, nation, place and else. While it was Simon Anholt who introduced the concept, the places and nations applied the strategies to attract people, investments, tourists in a century-long practices, however, they were not labelled as such (Cvitkovic 2023). Namely, the 'father of marketing', Professor Philip Kotler in 1997 book *The Marketing of Nations: Strategic Approach to Building National Wealth* accentuates the role of government policymakers and the strategies in building national wealth. Cities are stepping up to global challenges and are rivaling states regarding their economic, diplomatic, and international power and connectivity (Muggah 2020). Moreover, while

states are still the leading global actors, they have lost some economic, political, and symbolic ground to international corporations and global cities that generate an increasingly substantial proportion of national GDP (Muggah 2020; Sassen 2006). Hence, this is where place branding is becoming a prerequisite for the positioning in the globalized world.

Literature review indicates that authors focus on various aspects of rebranding, including a corporate rebranding (e.g., Ahonen 2008; Bolhuis, de Jong and van den Bosch 2018; Daly and Moloney, 2005; Joseph et al. 2021; Juntunen, Saraniemi and Jussila 2009; Muzellec and Lambkin 2006; Tarnovskaya and Biedenbach 2018), case studies (e.g., Daffey and Abratt 2002; Daly and Moloney 2005; Kaikati 2003; Merrilees 2005; Ntounis and Kavaratzis 2017; Schultz and Hatch 2003), theory development (e.g., Ahonen 2008;; Juntunen, Saraniemi and Jussila 2009; Lam et al. 2010; Merrilees and Miller 2008; Muzellec, Doogan and Lambkin 2003; Muzellec and Lambkin 2006), visual identity (e.g., Ari-Matti et al. 2021; Govers 2013; Wertz 2023), green rebranding (e.g., Cusworth et al. 2022), and else. Here we address a need for a more substantial contribution to rebranding codification that focuses on geographic locations and nations, as only authors such as Boland (2013), Cvitkovic (2023), Cvitkovic and Kline (2017), Hakala, Lemmetyinen and Nieminen (2020), Jucu (2020), Setiadi, Sutanto and Surienty (2021), have forwarded the concept. The idea of place rebranding originates from corporate conceptualization, defined as "an ongoing procedure in which an organization adapts to the shifts in its business environment, modifying its self-concept to endure and flourish" (Tevi and Otubanjo 2013, p. 92). Rebranding is characterized by the attribution of new meanings to different market categories through the 3R prerequisites, namely renaming, restructuring, and repositioning that result in the creation of a new image in consumers' minds (Cvitkovic 2023; Cvitkovic and Kline 2017).

The place rebranding process includes significant planning projects that physically transform the city, replacing old and negative images with a new place identity due to the changing market conditions and alterations in the character of an area (Bennett and Savani 2003; Boland 2013; Setiadi et al. 2021). According to Cvitkovic and Kline (2017) city rebranding is "a politically instigated visionary project that is aimed at attributing new meanings to a city and wiping out the old and often undesired images so that the renewed brand can deliver economic prosperity, greater tourist demand, cultural attention within domestic and international customers and can serve as a vehicle to remake the nation brand". In other words, urban rebranding, alongside promotional marketing, enables a place to transfigure its existing visual image into a new desirable brand, attracting new residents, visitors, and investors (Barratt 2021). However, many rebranding projects are conducted solely at the level of visual identity change. As Govers (2013) states, the contribution that logos and slogans can make to the place branding management is somewhat limited [27]. Namely, due to its complexity, place branding is about "managing brand equity" (Govers 2013, p. 71) and distinctive image associations. On the practical level, many city branding strategies use architecture and cultural heritage to make cities more visible, competitive, and relevant to the global world (El Messeidy 2019; Eljamal 2023). As accentuated by Collomb & Kalandides (2010) [45] the commodification of local history, cultures, and identities may be a conflictual process due to the tensions between using culture for branding purposes and promoting culture for itself.

Residents' Participation in Place Rebranding

Citizens have been for centuries and inseparable part of a place brand brand, and its more credible and trustworthy brand ambassadors. Thus, it is important to underscore the importance of considering place branding within the framework of urban governance, which involves how city policies are formulated, decided upon, and put into action (Braun, 2008). While governance theory offers various interpretations and definitions, it generally encompasses the diverse stakeholders involved into city policies, often spanning public and private sectors (Healey, 2007). Urban governance, as defined by Healey (2006) encompasses the mechanisms for collective action within urban regions, cities, or neighbourhoods. In the context of place branding, both public and private stakeholders play roles, highlighting that place branding is intertwined with political decision-making and administrative processes (Healey, 2006). Therefore, understanding place branding requires acknowledging its embeddedness with urban governance structures and dynamics.

The scholarship on place rebranding has identified three main target groups-residents, companies, and visitors (Braun, Kavaratzis & Zenker, 2013) [10]. However, this point of view inappropriately restricts the position of residents; they are not merely target markets, and marketers should not only consider new residents as target markets of place branding. (Braun, Kavaratzis & Zenker, 2013). Instead, residents could be essential participants in the place branding process, not just passive beneficiaries or place customers but active co-creators through participation of and the dialogue, including public policy (Braun, Kavaratzis & Zenker, 2013). Active participation in local and national government elections makes citizens a crucial part of the political decision-making process, granting them immense power. Such citizen participation is, on the one hand, a right but also an obligation. In other words, place authorities should also guarantee such involvement and provide opportunities for citizens to contribute actively to decision-making (Braun, Kavaratzis & Zenker, 2013). Grass-roots movements of citizens who raise their voice against official place branding undertakings campaigns are about a series of discussions and alternative suggestions to authorities (Braun, Kavaratzis & Zenker, 2013).

The *sei Berlin* (*be Berlin)* campaign, launched on March 11, 2008, builds on this trend toward depicting creative individuals at the expense of specific iconic or symbolic sites, epitomizing the shift to people as the components and producers of the place brand (Collomb & Kalandides, 2010). Namely, the campaign focused on the active participation of the city's residents by collecting their narratives about Berlin and encouraging the construction of strategic alliances among institutional players. (Collomb & Kalandides, 2010).

The novelty approach to the campaign can be seen in the primary target audience. Namely, the campaign is not primarily targeted at investors or potential visitors but at the Berliners themselves under the motto - *Made by Berliners for Berlin* (Collomb & Kalandides, 2010). It emphasizes Berliners who are invited to *be a storyteller, be an ambassador, be Berlin* by sending their individual success stories and testimonials to the campaign's website (Collomb & Kalandides, 2010). In addition, rather than using one message or slogan, it has encouraged multiple messages about Berlin, using the pattern *be …, be …, be Berlin* (Collomb & Kalandides, 2010). The first set of posters released portrayed teenagers from a school in the Neukölln district (the Rütli school), presenting

items of clothing which they successfully designed and marketed to change the image of their school (Collomb & Kalandides, 2010). Winner of the first competition *Your message for Berlin* was a seventeen-year-old with her three-piece: *be unique, be diverse, be Berlin* (Wir Berlin, 2024). Barratt (2021) points out that utilizing a city's identity can be somewhat controversial in the community as it may selectively reshape and disregard history instead of enhancing it.

3 Methodology

To comprehend the architectural vocabulary of Skopje, we exemplify a comprehensive sociocultural and political analysis of two periods, namely, between 1945 and 1990 and after 1991. We are building on Della Porta and Keating (2008) and Maroević (2003) [20, 43], who accentuate the apprehension of the architecture of a city and its significance through its historical and chronological development to explain a case and then refine theory. According to Yin (1994) [70], case studies are favored when questions on *how* or *why* are raised and when the preoccupation is on a contemporaneous phenomenon within a real-life context. The case study depends on the secondary data and the archives, i.e., press clippings from when the researched artifacts were erected. Press clippings on the artifacts set up before 1990 were primarily accessed in the National and University Library in Skopje, most notably, retrieved from poorly preserved *Nova Makedonija* newspaper issues written in Cyrillic letter and dating from the socialist era, which adds to the value of the paper and the research. The computational model (Differential Equations and a Python program) has been developed by the authors.

4 Case Study: The Sociohistorical Background of Skopje

With Skopje as its capital, North Macedonia is situated in the central Balkan Peninsula. Functioning as a center of many kingdoms and empires, that is, from medieval Slavonic kingdoms to Roman and Byzantine, it has undergone many cultural and political influences, contributing to its mixed city brand identity. Hence, Skopje's contemporary place brand identity is complex, involving a historical amalgamation of diverse ethnic characteristics.

Post-1963 Earthquake Urban Makeover in Skopje: A Global Spectacle
After the World War II, the first Macedonian state emerged with Skopje as its capital, whilst Macedonians were recognized as a nationality in 1945 for the first time in their history. The post-war period is characterized by a significant population increase, with Skopje becoming the third-largest city of the new state before the 1963 earthquake. On July 26, 1963, an earthquake devastated over three-quarters of the city (Bouzarovski 2011; Kulić et al. 2012). The United Nations (UN) led the international effort to reconstruct the city, with support from prominent architects and urban planners worldwide, closely collaborating with the government (Kulić 2017). Skopje was declared *an international city of solidarity*, receiving significant worldwide aid for the reconstruction (Bouzarovski, 2011; Jovanovic Weiss 2017) [8, 45]. The government aimed to rebuild the city as a *world city* where global expertise and culture would unite, reflecting the idea

of solidarity, which became an excellent opportunity to demonstrate a *softened* form of socialism, international collaboration, and goodwill globally (Kulić 2017).

Influenced by Le Corbusier and traditional Japanese architecture, Kenzo Tange, the UN's international competition winner for a detailed reconstruction plan of central Skopje, developed a functional plan for the city featuring Brutalist and Metabolist buildings (Deets 2022). In the context of the regime, modernism and brutalism symbolized the grandness of state socialism and cosmopolitan architectural styles and designs, characterized by open spaces that encouraged movement and incorporated raw concrete on facades, earning international acclaim for both his plans and several buildings (Deets 2022; Štraus 1991; West 2015). The city's well-preserved walls and monumental gates served as a springboard for the architect to incorporate these as key motifs, emphasizing an inner core, infrastructure arrangement, and openness, later materialized through a *City Wall* concept in constructing residential buildings (Kulić et al. 2012; Jovanovic Weiss 2017). However, many critics point out that Tange's plan was a top-down imposed spatial project that did not consider the cultural blend and situation on the ground, resulting in an obscure and unrecognizable layout of the republic's capital city (Jovanovic Weiss 2017) [33].

The development of Skopje as a capital city also included the articulation of national identity, evident in, for example, Boris Čipan's Macedonian Academy of Sciences and Arts, where he integrated wooden frameworks, and glazed verandas, traditional Macedonian architectural elements (Kulić et al. 2012). Additionally, the authors emphasize the construction of national identity through a variety of other nation-building edifices, including the National and University Library, the Macedonian National Theatre, the National Assembly of the Republic of Macedonia, and the Museum of History, Archaeology, and Ethnography. Concurrently, the Republican Institute for the Protection of Monuments played a pivotal role in safeguarding the nation's heritage (Nova Makedonija 1968, 1974a). These findings contradict the claims of many critics of the architectural style of the socialist period as lacking vernacular and national identity elements. The Macedonian National Theatre is seen as an expressionistic masterpiece, distinguished by its curved lines and surfaces, standing out as a significant urban accomplishment of contemporary Skopje (Štraus 1991). The international collaboration on Skopje's reconstruction is exemplified by the Museum of Contemporary Art, a project funded by Poland, showcasing global solidarity while promoting cultural connections among nations (Nova Makedonija 1974b) [51]. Furthermore, the Opera-Ballet building of the Macedonian National Theatre and Music Centre stands out as a significant architectural work, while the Skopje Cultural Centre unites different national artistic institutions, including the opera, philharmonic orchestra, music academy, and cinema (Bernik 2004; Štraus 1991). In addition, the University Centre stands as a significant achievement in natural concrete, symbolizing the utilization of modern materials in architecture and Skopje's focus on education (Nova Makedonija 1974c, 1974d) [52, 53].

In essence, Skopje's architectural and cultural expansion during this period was shaped by an interplay between modernity, international collaboration, and national identity, which resulted in noteworthy architectural achievements that have lasting effects on the city's all-around appearance. Regardless, the city's architectural vocabulary lacked spatial continuity and cohesion, as the employed architects focused on individual projects rather than comprehensive urban planning (Štraus 1991).

Post-Independence Period: The Skopje 2014 Project

The dissolution of Yugoslavia resulted in a new social reality that denoted the end of socialism and the capital cities of the newly founded states participated in highly observable changes seen in the emergence of national symbols, becoming metaphoric hubs representing newly established national identities (Samardžić and Živković 2016). Simultaneously multi-colored advertising, Western product billboards, shopping malls, and high-paced construction projects superseded faded party propaganda, illustrating the change (Berdahl 2000). However, the architecture from the socialist period was left in a state of indecisiveness, captured between repurposing, reusing, or becoming architectural wreckages (Jovanovic Weiss 2017). Following the dissolution, the new state faced a confrontation with Greece concerning the name *Macedonia.* As it later became known, North Macedonia, with Skopje as its capital, experienced a nationalist changeover. However, during 2016 and 2017, a shift happened as the social democrats came into power, directing domestic and foreign policy modifications. The issue regarding the name of the Former Yugoslavian Republic of Macedonia (FYROM) remained unresolved for a duration exceeding 25 years and destabilised 2 governments in Greece (under PM Konstantinos Mitsotakis in 1993 by MPs, who left the governing New Democracy Party under Antonis Samaras and PM Alexis Tsipras in 2019, who lost the support of some MPs of the coalition government, who were members of ANEL Party, under Panos Kammenos). There were even assertive 'proposals' such as the one made by Slobodan Milosevic to the Greek Prime Minister Konstantinos Mitsotakis, suggesting the division of FYROM by military means between Serbia and Greece (Gelb, 1992), which the Greek PM promptly dismissed as a preposterous proposition. Following the tumultuous early period of Kiro Gligorov's leadership, during which he signed the Interim Accord on the 12th September 1995 at the United Nations Headquarters to normalise relations with Greece (Roszkowski and Kofman, 2016), the situation remained stagnant for several years with ongoing rounds of negotiations. The VMRO-DPMNE governments, in fact, reintroduced the policy of Antiquization, also referred to as Ancient Macedonism (Macedonian: антиквизација & антички македонизам). This term describes the identity policies implemented by the nationalist VMRO-DPMNE-led governments of North Macedonia from 2006 to 2017 (Bechev, 2019). To all intents and purposes, the Skopje 2014 project was an ostentatious manifestation of the Antiquization's grandiosity.

In anticipation of the Prespa Agreement in 2018, which gave to the country its actual name and was ratified with many obstacles in Northern Macedonia (Heraclides, 2020 & Nimetz 2020) [31, 48], by initiative of the Prime Minister Zoran Zaev, the government and institutions of Northern Macedonia declared the suspension of the Skopje 2014 programme and commenced the removal of its contentious monuments and statues. The Ministry of Culture has established a Commission to consider the feasibility of removing several monuments, including those of Alexander the Great. This was followed by the

renaming of the Skopje Airport and the motorway that connects Northern Macedonia and Greece (Sidzimovska, 2019) [59]. However, the return in 2024 of VMRO-DPMNE in power seems to be a drawback, which for the time being remains at the political rhetoric level.

The *Skopje 2014* project can be characterized as one of the biggest European architectural and urban renewal projects of the twenty-first century. Envisioned as a comprehensive reconstruction of the city centre at total estimated cost over 640 million euros, the project main goal was to rebrand the city for international companies and tourists attraction, however without the authorities engaging in any public consultation, the input from professional architects, urban planners, or even city residents (Deets 2022; Staletović and Pollozhani 2022) [19, 60]. By the vast of scholars, the *Skopje 2014* project is researched through the lens of modern nationalism, emphasizing visual elements from Classical antiquity and the neglection of the constitutionally recognized Albanian population. The changes involve a complete annihilation of the previous period, characterized by removing the architectural elements from socialism and introducing a glorified antique past that rewrites history in public spaces, leading to the city's unrecognizable urban transformation (Janev 2017). More specifically, *Vegas-meets-theBalkans* look is contrasted sharply with the city's modernist and brutalist architectural style, intended to evoke historic connections to a purported 'ancient Macedonian' past (Deets 2022). While the *Skopje 2014* project includes the erection of over a hundred monuments, the two most recognizable are *Voin na konj*, the monument to Alexander the Great, and *Porta Macedonia,* an arch (Cvitkovic 2023). The monument to Alexander the Great, which symbolizes the rebirth of the glorious past and urban visualization of the official nationalistic politics from the early 1990s, became a landmark of the city as it is visible and recognizable from many locations throughout Skopje (Cvitkovic 2023; Lynch 1960;). The original triumphal arches of the Roman Empire were used in commemorating martial victories, whereas *Porta Macedonia*, marks the political dissolution of Yugoslavia and North Macedonia's independence (West 2015). Architectural and professional criticizers portray the new urban landscape as *antiquisation*, a term referring to the Renaissance approach of giving a city the appearance of ancient Rome or Athens (Koteska 2011). The ruling VMRO-DPMNE party controlled both the national and local level government, exercising power in the major municipalities in Skopje, so subsequently the project stirred controversies due to its top-down and a strictly mono-ethnic characteristics in a city historically populated by different nationalities and religions (Staletović and Pollozhani 2022). In 2012, the city council of Skopje, working behind closed doors, approved some 240 new street names proposed by the ruling party (Marusic 2012). However, officials said a list of about 1,000 new names is in the pipeline (Marusic 2012). Moreover, Jovanovic Weiss (2017) points out that the policymaker's plan aspired to revamp Skopje as a historical city aimed also at substituting Tange's socialist-backed plan and jeopardizing the disappearance of Skopje's brutalist architectural legacy. One of the most troublesome aspects of the *Skopje 2014* project was the *Neoclassical* reconstruction of the 1970s and 1980s façades like the Government House, once a notable illustration of Modernist architecture, where the Styrofoam façade was applied during the project without the architect's consent, utterly covering the elevated parts of the building altogether (Sidzimovska 2019) [59].

Aftermath of Skopje 2014 Project: A Long Process of Appropriation or Instant Destruction?

In 2016 and 2017, national and local elections were organized, and the ruling VMRO-DPMNE party lost both. An essential electoral commitment of the Social Democratic Union (SDSM) was an abeyance of the *Skopje 2014* undertakings, which resulted in an official prohibition of any further building in Skopje city center as set by the Municipality Centre of Skopje in 2017 (Sidzimovska 2019). As stressed by Deets (2002), the new governments were tasked with deciding which projects to finalize and which to disassemble, what elements could and should be restored, and how to address substandard construction. To meet the pre-electoral promises, the new government conducted a thorough analysis which revealed that 90% of sculptures erected within the *Skopje 2014* project, including the Alexander the Great statue, the sculptures on Philip II Square, the Macedonia Gate, and the monument to Mother Theresa, are illegal in terms of state budgets use, public procurement, the conservation of cultural heritage, and even the country's Constitution (Sidzimovska 2019). While some monuments have been relocated, the hesitation to set a precedent has resulted in very few such moves (Deets 2022). The Alexander the Great monument dominating Macedonia Square and the nearby *Porta Macedonia* were preserved in their original locations (Deets 2022). Additionally, in February 2019, the government announced the monument's completion to Skopje-born Mother Theresa despite its illegality (Sidzimovska 2019).

5 Research Findings

Following the archival research and media discourse from the two significant historical periods of Skopje, this section lists the elicited brand associations that surfaced in the case study analysis. Namely, the dominant brand associations connected to the post-1963 Skopje are, listed in an alphabetical order, as follows: accomplishment, achievement, adhesion, ambiguous, archaeology, architecture, art, artistic, Assembly, ballet, Brutalism, cinema, collaboration, competition, concrete, connections, construction, contemporary, cosmopolitanism, criticism, culture, development, different, discontinuity, distinguished, education, effects, ethnography, expansion, functionalism, global, grandness, heritage, history, housing, identity, impose, incomprehensive, incorporation, infrastructure, institutions, integration, international, internationalization, interplay, Japan, lasting, Le Corbusier, library, Macedonia, masterpiece, materials, Metabolism, modern, modernism, modernity, movement, museum, music, nation, national, nation-building, nations, noteworthy, openness, opera, philharmonic, Poland, promotion, reconstruction, reconstruction, reconstruction, regime, republic, residential, science, significant, socialism, solidarity, standout, symbolism, symbolizing, The United Nations, theatre, top-down, tradition, unite, university, unrecognizable, vernacular. The most dominant brand associations elicited from the caste study analysis of the *Skopje 2014* project are, listed in an alphabetical order, as follows: Alexander the Great, ancient, annihilation, *antiquisation,* antiquity, architecture, Athens, Balkans, change, commemoration, confrontation, conservatism, construction, control, controversy, coverage, criticism, disappearance, dispute, dissolution, evocation, glorification, government, history,

ideology, impose, independence, internationalization, Las Vegas, Macedonia, manipulation, mono-ethnic, monumentalism, multi-coloured, nationalism, neglect, *Neoclassicism,* non-consensual, non-responsive, one-way-communication, party, past, politics, power, reconstruction, Renaissance, repurpose, reusage, revamp, rewrite, Roman Empire, Rome, Styrofoam, substitution, top-down, tourism, transformation, triumph, unrecognizable, victory (Fig. 1).

BRAND ASSOCIATION	1945-1990	AFTER 1991
ARCHITECTURE	✓	✓
CONSTRUCTION	✓	✓
CRITICISM	✓	✓
HISTORY	✓	✓
IDEOLOGY	✓	✓
IMPOSE	✓	✓
INTERNATIONALIZATION	✓	✓
MACEDONIA	✓	✓
RECONSTRUCTION	✓	✓
TOP-DOWN	✓	✓
TRANSFORMATION	✓	✓
UNRECOGNIZABLE	✓	✓

Fig. 1. Created by the authors: Overlapping Brand Associations between Post-1963 Skopje and the *Skopje 2014*

The research findings demonstrate the overlapping of the elicited brand associations stemming from two different historical periods. Namely, overlapping brand associations are, in alphabetical order, architecture, construction, criticism, history, ideology, impose, internationalization, Macedonia, reconstruction, top-down, transformation, and unrecognizable.

And here is the critical point where the quantification entaglement starts. Computational humanities extend traditional humanities research by bringing in digital tools and algorithmic methods that allow scholars to do more than what is feasible by manual, close reading alone. By processing large corpora using techniques like text mining, topic modelling, sentiment analysis, and stylometry, researchers can detect patterns across hundreds or thousands of texts—e.g. tracing the evolution of ideas or themes through time. Quantitative methods add rigour: they allow hypothesis-testing, reproducibility, and help verify claims such as authorship attribution or the prevalence of certain motifs [17, 18, 28].

Following the aforementioned qualitative assessment, and in order to model the dynamics between the most influential factors as deriving from the aforementioned

narratives, a system of coupled differential equations is getting introduced. The goal is to reflect the interactions between: Place Rebranding (R), Architecture (A), Public Policy (P) and to account for: Citizen Participation (C) – a limiting factor, National Identity Narrative (N) – a driving force and Disruption Events – in 1963 and 2014, which we are modelled as impulses.

Mathematical Model
We define:

- R(t): Place Rebranding effort at time t
- A(t): Architectural transformation at time t
- P(t): Public Policy changes at time t
- C(t): Citizen Participation at time t
- N(t): National Identity Narrative influence at time t

Assumptions:

- Architecture is a direct instrument of both rebranding and policy.
- Citizen participation has a dampening effect (low participation limits effectiveness).
- National identity narrative amplifies rebranding but dominates public policy unilaterally.
- Major disruptions (1963, 2014) act as impulses in the system

System of Differential Equations:

$$dR/dt = \alpha 1A + \beta 1P + \gamma 1N - \delta 1C$$

$$dA/dt = \alpha 2R + \beta 2N - \delta 2C$$

$$dP/dt = \alpha 3R + \beta 3N - \delta 3C$$

$$dC/dt = -\mu 1(R + A + P) + \nu 1$$

$$dN/dt = \eta 1 - \epsilon 1C$$

Where: α, β, γ, δ, μ, ν, η, ϵα, β, γ, δ, μ, ν, η, ϵ (see Appendix 1) are parameters and Impulses for 1963 and 2014 are modeled as delta functions in A(t) and R(t).

The model formalizes how top-down symbolic rebranding efforts (like Skopje 2014) cycle through architecture and policy but fail to sustain legitimacy due to low civic involvement. It offers a dynamic lens to critique similar urban strategies in post-socialist or transitional cities. We adjust the model for a comparative study between Post 1963 and Skopje 2014 and we compare the trajectory of the 5 main variables across both cases under different assumptions about: Citizen participation (C), National identity influence (N), Strength of disruption impulse and Long-term policy adaptation and we execute the Python program (see Appendix 2).

The interpretation of the comparative results to the following summary and graph (Figs. 2 and 3 assuming a 10 years horizon and according to authors' assessment) and shows that: a) Rebranding, Architecture, and Policy change over time, b) Citizen Participation stays limited in both, but more in Skopje 2014 and c) The National Narrative grows faster in Skopje 2014 due to top-down identity focus.

Aspect	Post-1963	Skopje 2014
Disruption impulse	Weaker	Stronger
Identity pressure	Lower	Dominant
Civic inclusion	Slightly better	Minimal
Resulting dynamics	Cyclical but adaptive	Oscillatory and rigid
Narrative dominance	Limited	Persistent

Fig. 2. Created by the authors

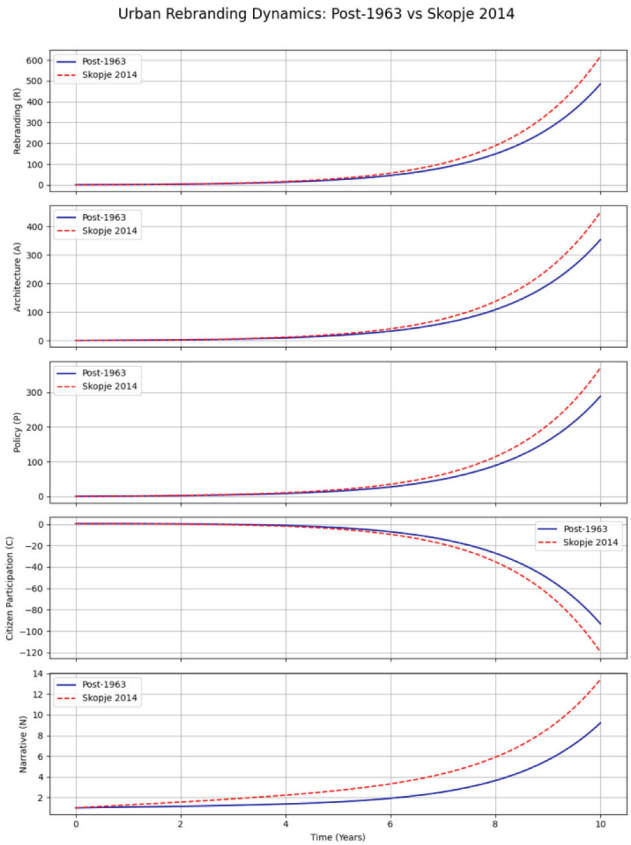


Fig. 3. Created by the authors

6 Discussion

The overlapping of the elicited brand associations stemming from two different historical periods is evidence of the application of the same practices by the brand author in the post-1963 Skopje and the *Skopje 2014* project, regardless of being employed over half a

century later. More specifically, in both cases, policymakers employed the same policies and practices in place rebranding, regardless of their underlying opposite ideological underpinning. In this context, architecture and significant urban reconstruction projects have been used to appeal to the international community while building a new place and national brand identity. In addition, both undertakings received criticisms, most notably as being top-down imposed, not considering the bottom-up principle and all its residents.

There are several potential directions to the treatment of *Skopje 2014*; namely, to be preserved as an example, demolished, partly appropriated, and partly dislocated (like Grutas Park in Lithuania or Memento Park in Budapest) (Sidzimovska 2019). We suggest that the centuries-long practices of monumental removal and the change of edifices' reappropriation under the motto '*the history begins with us*' should be stopped to become a highly visible and positive worldwide example of a different approach to the inherited architectural vocabulary. We propose relocating most of the monuments in the specially dedicated location to be an open museum and an eternal reminder to the generations of the consequences of the top-down imposed ideology. In addition, the museum may serve as an open-air classroom, underlying the need to educate all citizens (and visitors) on the long-term implications of such practices. In addition to relocating and preserving the monuments, the authors suggest that the possible direction may be the decontextualization of the monuments and artifacts, following the example of Italy's city of Bolzano. In such a way, radical new place brand creation imposed by the former policymakers could result in overcoming such historical urban undertakings, which would result in prospects' social mobility between brand identities.

7 Conclusion

For more than 60 years, North Macedonia's capital city of Skopje has been experiencing two significant urban makeovers, one following the 1963 earthquake and the other within the *Skopje 2014* project. Each of the urban upheavals contributed to the revolutionary rebranding of the city, resulting in a completely new brand positioning.

Post-independence urban renewal in Skopje aimed to annihilate the socialist period with Styrofoam and cheap construction materials of questionable quality. Additionally, the brand author was, despite the universally recognized historical facts, focused on creating a whole new brand identity based on the falsification of history and misappropriating Alexander the Great as being the ancestor of the people of North Macedonia, and subsequently trying to incorporate *Hellenism* as the bedrock of the place and nation brand equity, all materialized in the revolutionary changes of the visual identity of the city's urban core. In addition, it entirely neglected the Albanians, the second-largest ethnicity of the state. As presented in the research findings, in both periods' architecture and significant urban reconstruction projects have been used to appeal to the international community and irrespective of the residents.

We argue that within the concept, place rebranding may include (1) *renaming* (e.g., street, square, institution); (2) (major) *restructuring* within the urban vocabulary (e.g., significant undertakings, building and rebuilding flagship projects), most commonly resulting in (3) *repositioning* of the place in the mind of its prospects, while signalizing a significant shift in the environment. These elements have been seen over the past sixty

years in Skopje. Moreover, we argue that place rebranding can be regarded as a project with a revolutionary feel, bringing new vision, positioning and brand associations aligned with the desirable, new image that a brand author is trying to build. Furthermore, all *cosmetic* rebranding strategies, like the new logo or slogan, should not be regarded as genuine place rebranding as these do not have the potential to impact significantly its current and potential prospects and, therefore, have limited power. These findings also place even more importance on the brand author (i.e., policymakers) in the process of the place rebranding. Furthermore, due to the great extent of the environment and ideology change, the place rebranding is vertically employed to communicate the repositioning that aligns with the values of the brand author and its vision of the desirable brand image using centuries-long practices of the top-down systematic management approach.

The aftermath of the *Skopje 2014* project should be directed in several different ways, namely, constructing the memorial park in the city to relocate as much of the erected edifices in the last fifteen years as possible. More specifically, the park should operate as an open-air museum, an educational site for all the prospects to be taught about the severe consequence of false-history appropriation and its long-term impact. In such a way, the local and national governments may stop employing the centuries-long practice of demolishing unwanted inheritance from the previous policymakers. It would communicate a more subtle approach to dealing with the unwished-for legacy, with the added value of keeping the promises made during the 2016 and 2017 elections while still attracting tourists to the city. Additionally, the national and local government efforts should consider other place rebranding strategies, such as repositioning to a more sustainable direction. Namely, North Macedonia's policymakers should align with the current globalization wave where places are being repositioned into high-income cities that thrive on infrastructure, sustainable environmental policies, bottom-up approach, citizen participation, high quality of life, and security. Such strategies and accompanying policies can long-term influence broadly accepted identity (re)construction and brand image perception, but only as a result of a collaboration between residents, local and national policymakers, and their shared vision on place and nation brand management.

Appendix 1

Each parameter in the model plays a specific role in shaping how the system evolves over time. Below is a breakdown of the Greek-letter parameters used in the differential equations, along with their interpretation and influence.

Symbol	Affects Equation	Description	Role in System
$\boldsymbol{\alpha_1}$	dR/dt	Influence of Architecture (A) on Place Rebranding (R)	Higher values make architectural change a strong contributor to city rebranding
$\boldsymbol{\beta_1}$	dR/dt	Influence of Public Policy (P) on Rebranding (R)	Reflects how much policy shapes the city's external image and strategic vision

(*continued*)

(*continued*)

Symbol	Affects Equation	Description	Role in System
γ_1	dR/dt	Influence of National Narrative (N) on Rebranding (R)	Captures how symbolic identity narratives drive rebranding
δ_1	dR/dt	Dampening from Citizen Participation (C) on Rebranding (R)	Citizen pushback or skepticism reduces branding efforts
α_2	dA/dt	Influence of Rebranding (R) on Architecture (A)	Rebranding initiates architectural transformations (buildings, monuments, etc.)
β_2	dA/dt	Influence of Narrative (N) on Architecture (A)	Shows how symbolic narratives shape architectural choices (e.g., classical statues)
δ_2	dA/dt	Dampening effect from Citizen Participation (C) on Architecture (A)	Citizen opposition slows or blocks architectural execution
α_3	dP/dt	Influence of Rebranding (R) on Public Policy (P)	Policies often shift to support the branding strategy
β_3	dP/dt	Influence of Narrative (N) on Policy (P)	Narrative dominance drives policy direction
δ_3	dP/dt	Dampening effect from Citizen Participation (C) on Policy (P)	Civic resistance slows policy changes or adds accountability
μ_1	dC/dt	Negative feedback from R, A, P on Citizen Participation (C)	The more top-down change, the more citizens disengage or are excluded
ν_1	dC/dt	Baseline or natural tendency for citizens to participate	Reflects civil society strength, engagement norms
η_1	dN/dt	Natural growth of National Narrative (N)	The narrative builds over time, especially in post-socialist or post-conflict contexts
ε_1	dN/dt	Resistance from Citizen Participation (C) on National Narrative (N)	Civic discourse dilutes or challenges top-down identity narratives

Examples of Dynamics (as experimental input in the comparative model)

- **High α, β, γ** → Fast top-down transformation
- **High δ, μ, ε** → Strong civic resistance, slows state-driven agendas
- **High ν** → Encourages participatory urbanism
- **High η** → Nationalist agenda grows unchecked if not mitigated

Appendix 2

```
Python Program (requires matplotlib, numpy, and scipy)
import numpy as np
from scipy.integrate import solve_ivp
import matplotlib.pyplot as plt

# Step 1: Define the system of differential equations
def urban_model(t, y, p):
    R, A, P, C, N = y
    impulse = p['impulse_strength'] if any(abs(t - i) < 0.5 for i in p['impulse_times'])
else 0

    dRdt = p['alpha1'] * A + p['beta1'] * P + p['gamma1'] * N - p['delta1'] * C +
impulse
    dAdt = p['alpha2'] * R + p['beta2'] * N - p['delta2'] * C + impulse
    dPdt = p['alpha3'] * R + p['beta3'] * N - p['delta3'] * C
    dCdt = -p['mu1'] * (R + A + P) + p['nu1']
    dNdt = p['eta1'] - p['epsilon1'] * C

    return [dRdt, dAdt, dPdt, dCdt, dNdt]

# Step 2: Configure parameters for each scenario
def get_params(case):
    base_params = {
        'alpha1': 0.5, 'beta1': 0.3, 'gamma1': 0.4, 'delta1': 0.2,
        'alpha2': 0.4, 'beta2': 0.5, 'delta2': 0.1,
        'alpha3': 0.3, 'beta3': 0.6, 'delta3': 0.2,
        'mu1': 0.05, 'nu1': 0.1,
        'eta1': 0.2, 'epsilon1': 0.05,
        'impulse_strength': 2.0,
        'impulse_times': [10, 40]
    }

    if case == "Post-1963":
        base_params.update({
            'impulse_times': [10],
            'impulse_strength': 1.5,
            'nu1': 0.15,  # more citizen engagement
            'eta1': 0.1   # less national narrative pressure
        })
```

```
    elif case == "Skopje 2014":
        base_params.update({
            'impulse_times': [40],
            'impulse_strength': 2.5,
            'nu1': 0.05,  # less citizen participation
            'eta1': 0.3   # stronger national narrative
        })
    return base_params

# Step 3: Run simulation for a given case
def simulate_case(case_name):
    params = get_params(case_name)
    y0 = [0.5, 0.5, 0.5, 0.5, 1.0]  # Initial values for [R, A, P, C, N]
    t_span = (0, 10)
    t_eval = np.linspace(*t_span, 500)
    sol = solve_ivp(urban_model, t_span, y0, t_eval=t_eval, args=(params,))
    return sol

# Step 4: Run both scenarios
sol_post1963 = simulate_case("Post-1963")
sol_skopje2014 = simulate_case("Skopje 2014")

# Step 5: Plot results
labels = ['Rebranding (R)', 'Architecture (A)', 'Policy (P)', 'Citizen Participation (C)',
'Narrative (N)']
fig, axes = plt.subplots(5, 1, figsize=(10, 14), sharex=True)

for i in range(5):
    axes[i].plot(sol_post1963.t, sol_post1963.y[i], label="Post-1963", color='blue')
    axes[i].plot(sol_skopje2014.t, sol_skopje2014.y[i], label="Skopje 2014",
color='red', linestyle='--')
    axes[i].set_ylabel(labels[i])
    axes[i].legend()
    axes[i].grid(True)

axes[-1].set_xlabel("Time (Years)")
fig.suptitle("Urban Rebranding Dynamics: Post-1963 vs Skopje 2014",
fontsize=16)
plt.tight_layout(rect=[0, 0.03, 1, 0.97])
plt.show()
```

References

Barratt, S.: The efficacy of rebranding through architecture: analysing different aspects and dimensions of rebranding, what is the efficacy of urban rebranding strategies? Fields: J. Huddersfield Student Res. (2021). https://www.fieldsjournal.org.uk/article/id/799

Bechev, D.: Historical Dictionary of North Macedonia, p. 24. Rowman & Littlefield (2019) ISBN 978-1-5381-1962-4
Beck, U.: The Open City: architecture and reflexive modernity. In: Democracy without Enemies, pp. 115–121. Polity Press (1998)
Bennett, R., Savani, S.: The rebranding of city places: an international comparative investigation. Int. Public Manag. Rev. **4**(2), 70–87 (2003)
Berdahl, D.: Introduction: An Anthropology of Postsocialism. In: Altering States: Ethnographies of Transition in Eastern Europe and the Former Soviet Union, pp. 1–13. University of Michigan Press (2000)
Bernik, S.: Slovenska arhitektura dvajsetega stoletja = Slovene Architecture of the Twentieth Century. Mestna galerija Ljubljana (2004)
Boland, P.: Sexing up the city in the international beauty contest: the performative nature of spatial planning and the fictive spectacle of place branding. Town Plan. Rev. **84**(2), 251–274 (2013)
Bouzarovski, S.: Skopje. Cities. **28**, 265–277 (2011)
Braun, E.: City marketing: towards an integrated approach ERIM PhD Series in Research and Management, vol. 142. Erasmus Research Institute of Management (ERIM), Rotterdam (2008) http://hdl.handle.net/1765/13694
Braun, E., Kavaratzis, M., Zenker, S.: My city - my brand: the different roles of residents in place branding. J. Place Manag. Dev. **6**(1), 18–28 (2013)
Campbell, S.: Cold War Metropolis: the Fall and Rebirth of Berlin as a World City. https://www-personal.umich.edu/~sdcamp/capitals/Ch2.html. (2000)
Clark, G: How cities took over the world: a history of globalisation spanning 4,000 years. The Guardian (December 1). https://www.theguardian.com/cities/2016/dec/01/how-cities-took-over-the-world-a-history-of-globalisation-spanning-4000-years. (2016)
Collomb, C., Kalandides, A.: The 'Be Berlin' campaign: old wine in new bottles or innovative form of participatory place branding? In: Ashworth, G.J., Kavaratzis, M. (eds.) Towards Effective Place Brand Management: Branding European Cities and Regions, pp. 173–190. Edward Elgar, Cheltenham (2010)
Cvitkovic, S.: The Role of Architecture in Rebranding Post-Socialist Capital Cities and Nations. Doctoral Thesis. University of Ljubljana. (2023)
Cvitkovic, S., Kline, M.: Skopje: rebranding the capital city through architecture and monuments to remake the nation brand. Socijologija i prostor. **207**(1), 33–53 (2017)
Czepczynski, M.: interpreting post-socialist icons: from pride and hate towards dissappearance and/or assimilation. Human Geogr. J. Stud. Res. Hum. Geogr. **4**(1), 67–78 (2010)
Damerow, J., Koeser, R.S., Carver, J.C., Vogl, M.: Code review in digital humanities. Digital Scholar. Human. **40**(Supplement_1), i18–i26 (2025). https://doi.org/10.1093/llc/fqae052
Demidova, L.A., Zhukov, D.O., Andrianova, E.G., Sigov, A.S.: Modeling sociodynamic processes based on the use of the differential diffusion equation with fractional derivatives. Information. **14**(2), 121 (2023). https://doi.org/10.3390/info14020121
Deets, S.: Understanding weak local governance in the Neoliberal City through the case of Skopje 2014. Communis. Post-Commun. **55**(3), 1–23 (2022). https://doi.org/10.1525/cpcs.2022.1544873
Della Porta, D., Keating, M.: How many approaches in the social sciences? An epistemological introduction. In: Approaches and Methodologies in the Social Sciences: A Pluralist Perspective, pp. 19–39. Cambridge University Press (2008)
Diener, A.C., Hagen, J.: From socialist to post-socialist cities: narrating the nation through urban space. National. Papers. **41**(4), 487–514 (2013). https://doi.org/10.1080/00905992.2013.768217
Diekmann, A., Cloquet, I.: How 'capital' are capital cities on the Internet? Curr. Issue Tour. **15**(1–2), 19–33 (2012). https://doi.org/10.1080/13683500.2011.634892

El Messeidy, R.: Architecture and city branding: role of iconic buildings. Eng. Res. J. **163**(September), A100–A110 (2019)
Eljamal, M.: A mixed city or an ancient historical city? The malleability of heritage in rebranding Lydd. J. Soc. Archaeol. (2023). https://doi.org/10.1177/14696053231207398
Ergut, E.T.: The forming of the national in architecture. METU JFA J. Faculty Arch. Middle East Tech. Univ. **19**(1–2), 31–43 (1999)
Gelb, L.H.: Foreign Affairs; 'Macedonia' for Greece. The New York Times. Available at: https://www.nytimes.com/1992/06/12/opinion/foreign-affairs-macedonia-for-greece.html. (1992). Accessed 16 Feb 2024.
Govers, R.: Why place branding is not about logos and slogans. Place Brand Public Diploma. **9**, 71–75 (2013). https://doi.org/10.1057/pb.2013.11
Grammaticos, B., Willox, R., Satsuma, J.: Revisiting the human and nature dynamics model. Regular Chaotic Dyn. **25**, 178–198 (2020). https://doi.org/10.1134/S1560354720020045
Healey, P.: Transforming governance: challenges of institutional adaptation and a new politics of space. Eur. Plan. Stud. **14**(3), 299–320 (2006)
Healey, P.: The new institutionalism and the transformative goals of planning. In: Verma, N. (ed.) Institutions and Planning, pp. 61–90. Elsevier, Amsterdam (2007)
Heraclides, A.: The Macedonian question and the Macedonians: a history, 1st edn. Routledge (2020). https://doi.org/10.4324/9780429266362
Janev, G.: Burdensome past: challenging the socialist heritage in Macedonia. Studia ethnolica Croatica. **29**(1), 149–170 (2017)
Jovanovic Weiss, S.: Socialist Architecture: The Reappearing Act. The Green Box (2017)
Kaika, M.: Architecture and crisis: re-inventing the icon, re-imag(in)ing London and re-branding the City. Trans. Inst. Br. Geogr. **35**, 453–474 (2010)
Kotler, P., Jatusripitak, S., Maesincee, S.: The Marketing of Nations: A Strategic Approach to Building National Wealth. Free Press (1997)
Koteska, J.: Troubles with History: Skopje 2014. ARTMargins Online, MIT Press, (December 29). https://artmargins.com/troubles-with-history-skopje-2014/. (2011)
Kulić, V.: Building the socialist Balkans: architecture in the global networks of the cold war. Southeastern Europe. **41**(2), 95–111 (2017)
Kulić, V., Mrduljaš, H., Thaler, W.: Modernism In-Between: The Mediatory Architectures of Socialist Yugoslavia. Jovis Verlag (2012)
Lasswell, H.D.: The Signature of Power: Buildings, Communication and Policy. Transaction Publishers (1979)
Light, D., Young, C.: Urban space, political identity and the unwanted legacies of state socialism: Bucharest's problematic Centru Civic in the post-socialist era. Nationalities Papers. **41**(4), 515–535 (2013)
Lynch, K.: The Image of the City. M.I.T. Press (1960)
Mariotti, J., Pencic, D.: Changing perspectives of urbanity during socialism and after: the case of two Neighborhoods in Skopje. Stud. Hist. Theory Arch. **3**, 174–182 (2015)
Maroević, I.: Antologija zagrebačke arhitekture. Art studio Azinović Zagreb (2003)
Marusic, S.: New Street Battle Splits Macedonia's Capital. Balkan Insight (March 8). https://balkaninsight.com/2012/03/08/new-kind-of-street-battle-grips-macedonia-s-capital/. 2012
Meisel, J.: Capital cities: what is a capital? In: Capital Cities/Les Capitales: International Perspective/Perspectives Internationales, pp. 3–6. McGill-Queen's University Press (1993)
Muggah, R.: Look to cities, not nation-states, to solve our biggest challenges. World Economic Forum (January 23). https://www.weforum.org/agenda/2020/01/cities-mayors-not-nation-states-challenges-climate. (2020).
Muratovski, G.: Branding the City. In: The Statesman's Yearbook Companion: The Leaders, Events and Cities of the World, pp. 999–1000. Palgrave Macmillan (2019)

Nimetz, M.: The Macedonian "Name" Dispute: The Macedonian Question—Resolved? Nationalities Papers. **48**(2), 205–214 (2020). https://doi.org/10.1017/nps.2020.10
Nova Makedonija: Naskoro ke bide gotova i Univerzitetskata biblioteka. (May 7), 5. (1968)
Nova Makedonija: Tri muzei vo edna zgrada. (February 10), 9. (1974a)
Nova Makedonija: Deset godini plodno živeenje. (February 13), 8. (1974b)
Nova Makedonija: Dostojna afirmacija na naukata. (June 5), 1. (1974c)
Nova Makedonija: Neodgovorno i nestručno. (December 22), 1. 1974d)
Peterson, M., Godby, R.W.: Citizen participation in political markets: extending service-dominant logic to public policy. J. Public Policy Mark. **39**(4), 412–429 (2020). https://doi.org/10.1177/0743915620912287
Roszkowski, W., Kofman, J.: Biographical Dictionary of Central and Eastern Europe in the Twentieth Century, p. 297. Routledge (2016) ISBN 9781317475941
Samardžić, O., Živković, P.: Simbolička tekstura postsocijalističke transformacije urbanog prostora (Podgorica-Zagreb). In: Tranzicijska preobrazba glavnih gradova Zagreba i Podgorice kao sustava naselja: Bilateralni project između Hrvatske i Crne Gore, pp. 167–200. Institut za društvena istraživanja u Zagrebu (2016)
Sassen, S: Making Public Interventions in Today's Massive Cities. http://www.saskiasassen.com/PDFs/publications/Making-Public-Interventions.pdf. (2006)
Setiadi, N.J., Sutanto, H., Surienty, L.: Creative city as the rebranding strategy to accelerate new economic city: Indonesia's city development strategy to build a new economic center as an integrating ideas, society and industries in the development of Indonesian creative cities. IOP Conf. Series: Earth Environ. Sci. **729**, 1–11 (2021)
Sidzimovska, I.: After overthrowing its government and changing its name, North Macedonia faces up to the urban crisis in Skopje. New Easte Digital Archive, (April 29). https://www.new-east-archive.org/features/show/11141/after-overthrowing-its-government-and-changing-its-name-north-macedonia-faces-up-to-the-urban-crisis-in-skopje. (2019)
Staletović, B., Pollozhani, L.: To resist or not to resist: "Skopje 2014" and the politics of contention in North Macedonia. East Eur. Politics. **39**(4), 666–686 (2022). https://doi.org/10.1080/21599165.2022.2136652
Štraus, I.: Arhitektura Jugoslavije 1945–1990. Svjetlost (1991)
Tagesspiegel: Klaus Wowereit's Berlin: Poor but sexy is not enough. https://www.tagesspiegel.de/kultur/poor-but-sexy-is-not-enough-4828222.html. (2014)
Tevi, A.C., Otubanjo, O.: Understanding corporate rebranding: an evolution theory perspective. Int. J. Market. Stud. **5**(3), 87–93 (2013). https://doi.org/10.2139/ssrn.2197664
Verdery, K.: The Political Lives of Dead Bodies. Columbia University Press (1999)
Wertz, R.A.: Brand new: how visual context shapes initial response to logos and corporate visual identity systems. J. Prod. Brand. Manag. **32**(8), 1388–1398 (2023). https://doi.org/10.1108/JPBM-08-2022-4124
West, L.: Memento Park and Skopje 2014: transitions, monuments, and memory. Master Thesis, University of Georgia. (2015)
Wir Berlin: Don't be (sad) Berlin – the capital's campaign says goodbye. https://wir.berlin/en/campaigns/be-berlin. (2024)
Viswanathan, M.: Bottom-up journey into subsistence marketplaces from micro-level behavioral foundations: new directions for public policy and marketing. J. Public Policy Market. **42**(4), 358–360 (2023). https://doi.org/10.1177/07439156231185420
Yin, R.K.: Case Study Research. Design and Methods. Sage (1994)
Žižek, S.: Architectural Parallax: Spandrels and Other Phenomena of Class Struggle. Lacan (2009)

Basic Manobo Expression Dictionary: A Mobile-Based Application

Kylene Kyte B. Abandula(✉), Joen Rose S. Berdida, Jade G. Cuartero, Ignel T. Balmora, Patrick Jasson S. Pojas, and Jeanette M. Eborda

College of Computing and Information Sciences, Agusan del Sur State College of Agriculture and Technology, Bunawan, Philippines
{kkabandula,jrberdidal,jcuartero,ibalmora,pjpojas, jeborda}@asscat.edu.ph

Abstract. This study presents the development of the Basic Manobo Expression Dictionary: A Mobile-Based Application, designed to aid in the preservation, promotion, and accessibility of the Manobo language, one of the indigenous tongues in Agusan del Sur. With the increasing risk of language extinction brought about by modernization, younger generations are less fluent in traditional dialects. To address this issue, the researchers designed and developed a mobile-based dictionary application that provides users with translations of commonly used Manobo expressions into Filipino and English. The study employed the Systems Development Life Cycle (SDLC) methodology, integrating features such as categorized vocabulary, audio pronunciation guides, and user-friendly navigation. The system was evaluated by students, teachers, and selected community members using the ISO 25010 quality model. Findings showed strong acceptability across all criteria, with the highest scores in Usability (4.72) and Functional Suitability (4.68), confirming that the application is both practical and effective. The overall weighted mean of 4.65 indicates that the system successfully meets its objectives. Results demonstrate that the mobile dictionary can significantly contribute to cultural preservation, language education, and technology-driven heritage conservation.

Keywords: Manobo Language · Indigenous Preservation · Mobile Dictionary · Digital Learning · Cultural Heritage

1 Introduction

Language is one of the most vital elements of cultural identity and heritage. In the Philippines, where more than 170 languages exist, many indigenous tongues face the threat of extinction due to modernization, migration, and the dominance of national and global languages. The Manobo language, spoken by indigenous groups in Agusan del Sur and neighboring regions, is among those at risk. Younger generations often lack fluency in their mother tongue, leading to a gradual decline in everyday use.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026
S. Lanka et al. (Eds.): ICTSM 2025, LNNS 1755, pp. 581–594, 2026.
https://doi.org/10.1007/978-3-032-13177-5_44

The preservation of indigenous languages is recognized globally as a cultural priority. UNESCO (2019) stressed that the loss of native languages results in the disappearance of traditional knowledge, oral literature, and community identity. Despite these initiatives, limited resources exist for the Manobo language, particularly in modern digital platforms accessible to learners. To address this gap, the researchers propose the Basic Manobo Expression Dictionary: A Mobile-Based Application, which seeks to provide a convenient, user-friendly, and educational tool for students, teachers, and community members. The system will allow users to access translations of basic expressions, organized into thematic categories, with accompanying audio support to ensure proper pronunciation. By implementing this system, the study aims to strengthen cultural awareness, support inclusive education, and contribute to the long-term preservation of the Manobo language.

2 Related Works

Recent scholarship on Philippine languages and mobile-assisted vocabulary learning (MALL) offers important insights for projects aimed at documenting and revitalizing indigenous languages through digital means. Kaufman [1] provides a comprehensive overview of the central and southern Philippine languages, noting their diversity, contact dynamics, and uneven documentation, and emphasizes the urgent need for descriptive and community-centered revitalization efforts. Complementing this macro-level view, Palmera-Blanco [3] offers a phonological description of Manobo spoken in Davao Occidental, providing foundational data crucial for the development of dictionaries and language technology. Similarly, Bonifacio et al. [4] demonstrate how structured conversation-skills programs among Bukidnon and Talaandig youths can enhance competence and intergenerational engagement, highlighting that revitalization requires active, socially embedded learning contexts. A more applied contribution comes from Abingosa et al. [2], who document the development of a mobile application for digitizing the Higaonon language. Their study illustrates a participatory approach that combines lexical documentation with culturally relevant app design, showing how digital tools can serve both preservation and pedagogical purposes when guided by community needs. Evidence from the broader field of MALL reinforces the effectiveness of such approaches. Studies on mobile dictionary apps and digital flashcards consistently report positive outcomes for vocabulary acquisition and retention. For instance, Li and Hafner [9] and Rahmani et al. [10] show that mobile-assisted tools improve both receptive and productive vocabulary knowledge among EFL learners, with portability and spaced practice being key advantages. Similarly, Zakian et al. [7] and Koleini et al. [8] demonstrate that mobile-assisted learning extends to technical and academic vocabularies, suggesting strong potential for specialized indigenous domains such as ecological knowledge or ritual language. Research on mobile dictionary apps like U-Dictionary also points to

their value as reference and practice tools; Siraj [6] and Cao et al. [5] report that app use not only supports vocabulary growth but interacts with learner motivation, with well-designed interfaces benefiting both highly and less motivated learners. These findings align with Abingosa et al.'s [2] work by underscoring that design features—such as audio, examples, and usability, directly affect uptake and learning outcomes.

Together, these two literatures converge on several points. Foundational descriptive work such as Kaufman [1] and Palmera-Blanco [3] provides the linguistic data required for accurate digital tools. Community-based pedagogical interventions like those of Bonifacio et al. [4] show that social context enhances the impact of technology. MALL studies [5–10] provide strong empirical evidence that mobile applications, when incorporating features like flashcards, multimodal input, and spaced repetition, can significantly enhance vocabulary learning. However, a gap remains: while MALL research largely targets global languages in EFL contexts, few controlled studies examine its effectiveness for endangered indigenous languages. The Higaonon app [2] is a promising case study, but further longitudinal evaluations are needed to assess impacts on retention, everyday use, and intergenerational transmission. Thus, future work should integrate rigorous documentation, participatory digital design, and pedagogically grounded implementation to maximize the preservation and revitalization of indigenous Philippine languages.

3 Methods

3.1 Conceptual Framework of the Study

The study is anchored on the goal of cultural preservation through digital innovation. The researchers adapted the Systems Development Life Cycle (SDLC) as the main framework, ensuring that each phase—from planning to maintenance—was systematically addressed (Fig. 1).

3.2 Research Design

Figure 2 shows the Developmental Research Design was used, focusing on the design, development, and evaluation of the mobile-based dictionary.

3.3 Phases of Development

The researchers will use the Systems Development Life Cycle (SDLC) as the framework, consisting of the following phases:

1. Planning Phase – Identified the lack of accessible Manobo language resources through surveys and interviews. Defined the project's objectives and scope.
2. Analysis Phase – Gathered user requirements from students, teachers, and indigenous community members.
3. Design Phase – Created user interface prototypes with simple navigation and thematic vocabulary categories.
4. Development Phase – Built the mobile application using Java for Android and integrated an SQLite database for storing vocabulary and expressions. Audio support was added for pronunciation.

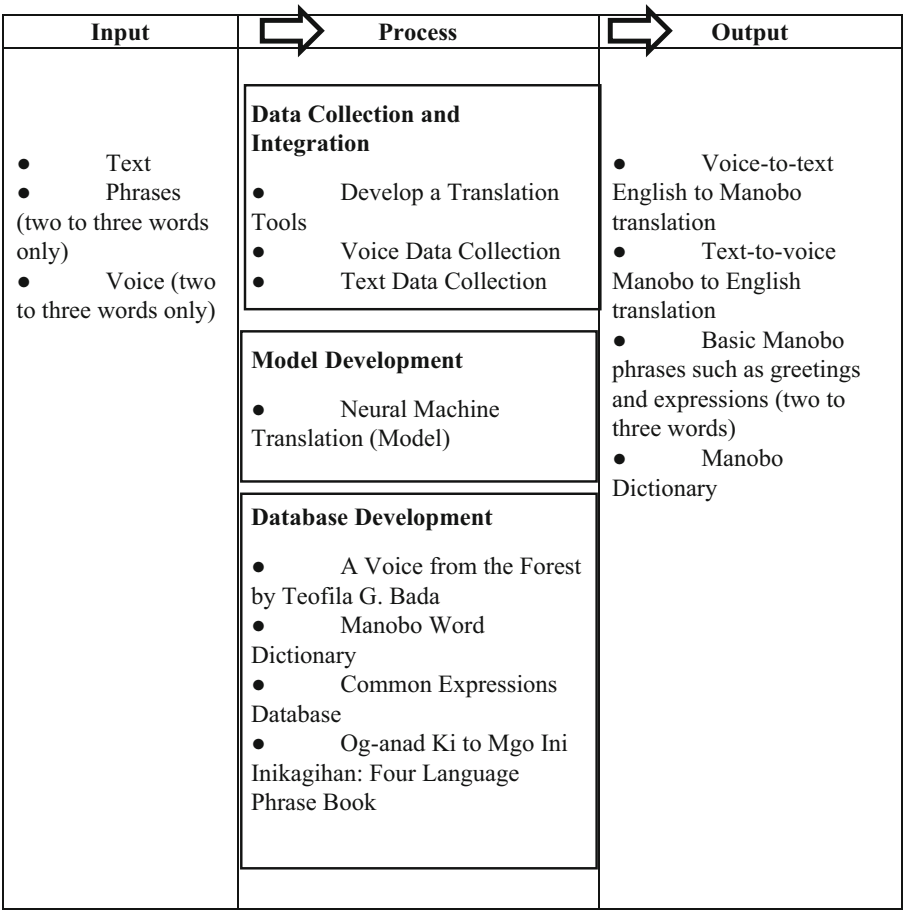


Fig. 1. Conceptual Framework of Basic Manobo Expression Dictionary

5. Testing Phase – Conducted unit testing, integration testing, and user acceptance testing with respondents.
6. Implementation Phase – Installed and tested the app among selected users in Agusan del Sur schools.
7. Maintenance Phase – Provided updates, bug fixes, and planned additional features for future versions.

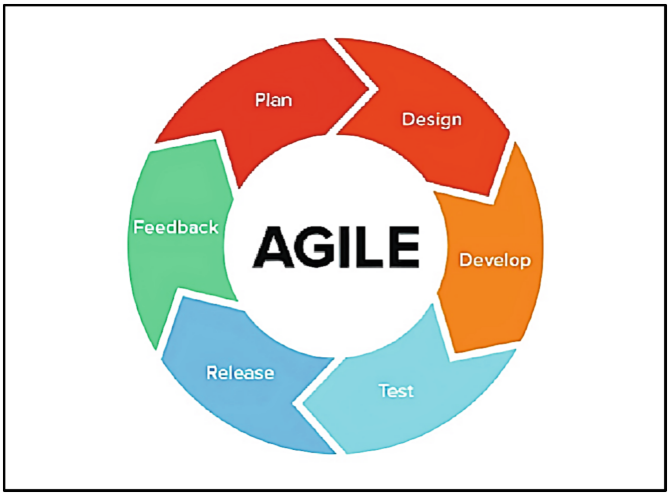


Fig. 2. Software Model: Agile Model

3.4 Respondents of the Study

Respondents included students, teachers, and selected members of the Manobo community. They were chosen as primary users and beneficiaries of the dictionary app.

3.5 Data Collection

Data were gathered through interviews, surveys, and focus group discussions. Usability and satisfaction levels were measured using a Likert-scale questionnaire.

3.6 Data Analysis

Both quantitative and qualitative analyses were employed. Weighted means and descriptive statistics were used to assess acceptability based on the ISO 25010 quality model. User feedback was analyzed to determine usability and cultural appropriateness.

3.7 Tools and Technologies

The mobile application was developed using Java (Android Studio) for the frontend and SQLite for the backend database. For audio features, standard recording tools were used to capture native pronunciation by Manobo speakers.

4 Results and Discussion

4.1 System Features and Implementation

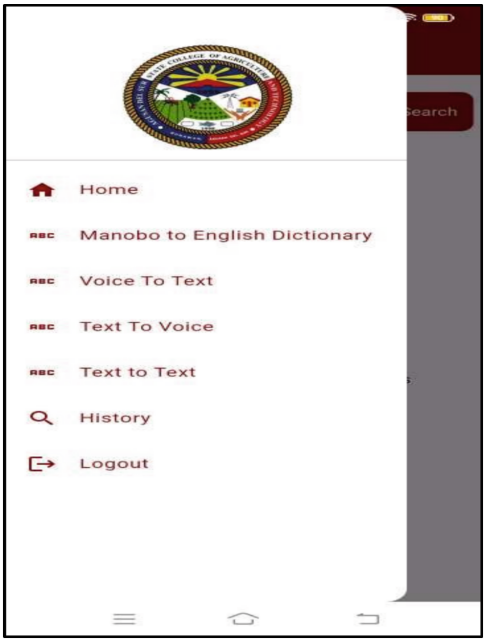


Fig. 3. User interface for the "Manobo Dictionary"

Figure 3 displays a navigation menu with features like "Manobo to English Dictionary," "Voice to Text," "Text to Voice," "Text to Text," "History," and "Student Account." The Basic Manobo Expressions Dictionary menu provides easy access to its main features:

- Manobo to English Dictionary: Lets users quickly search for Manobo words and their English translations.
- Voice to Text (English to Manobo): Converts spoken English sentences into written Manobo text.
- Text to Voice (Manobo to English): Converts typed Manobo words into spoken English, aiding pronunciation and comprehension.
- Text to Text: Allows users to type text in either Manobo or English and receive the corresponding translation in the other language.
- History: Keeps a log of recent translations for easy reference and reuse.
- Log Out: Provides a secure way to exit the application.

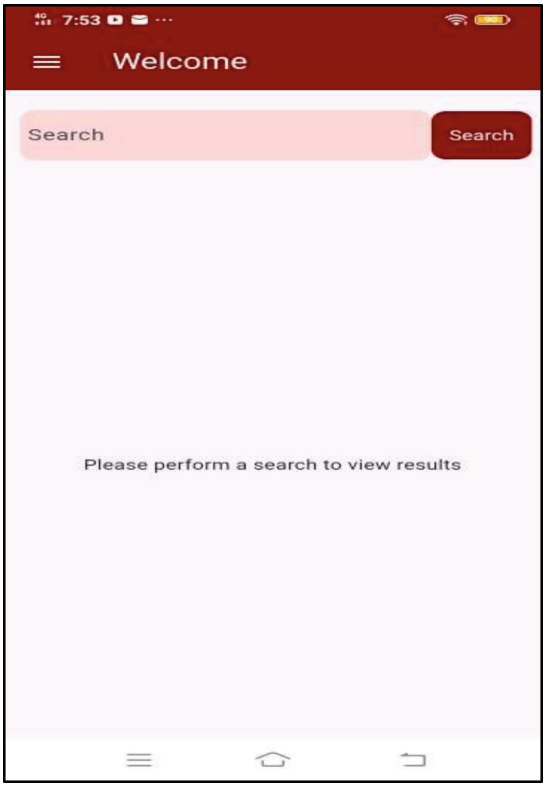


Fig. 4. Search of the Basic Manobo Expressions Dictionary

Figure 4 shows the users finding, quickly, translations of certain words or phrases. There is also a well-placed search bar, thus easily accessible and intuitive to use. Input one word or a short phrase, and it will be brought back either in Manobo or English. The design is lean, making quick and efficient searches possible to get the desired information.

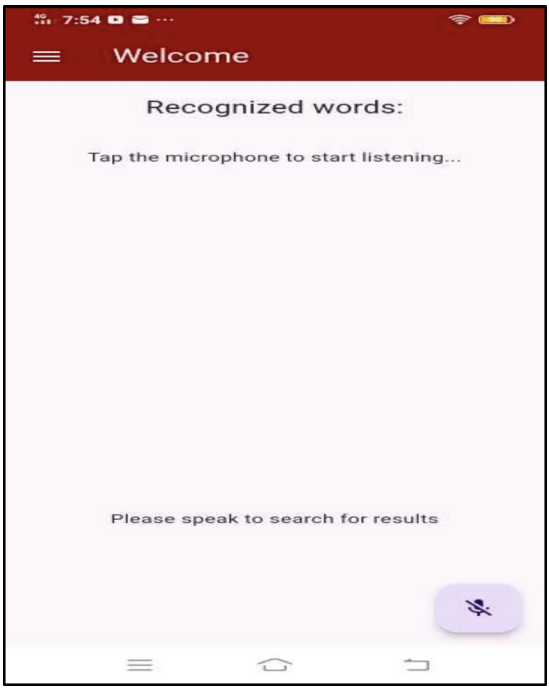


Fig. 5. Voice to Text

Figure 5 shows the involved speaking an English word or phrase and it will be transcribed into written Manobo text. This page enables the user to record his voice input by pressing a microphone icon, part of its simple interface design. The software captures the input and analyzes the audio data to show on the screen the text equivalent of the Manobo translation.

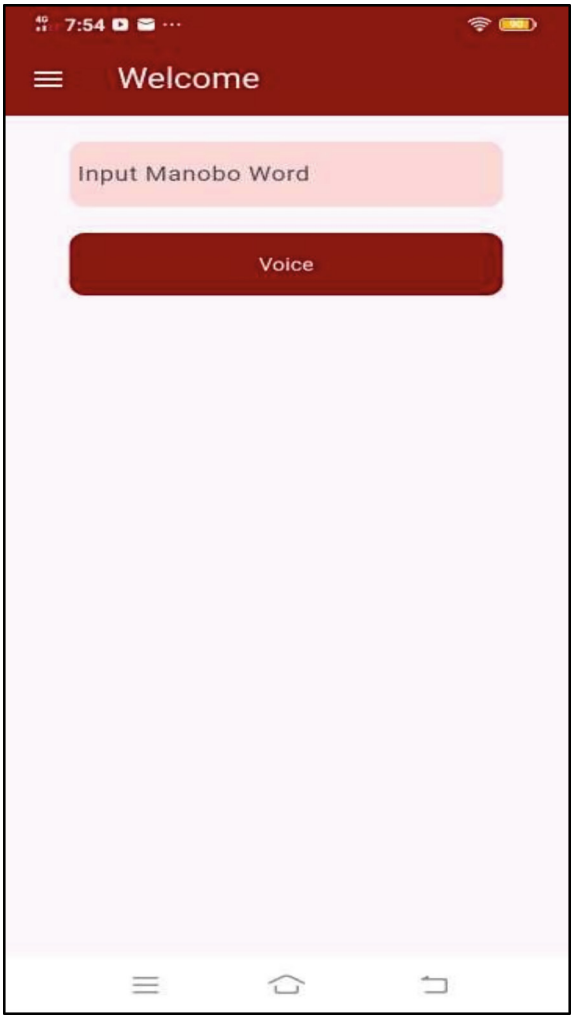


Fig. 6. Text to Voice

Figure 6 shows the text to voice which users are able to input Manobo text and then have the English translation read. This feature on the app provides visitors with the option to type a single word or a phrase in Manobo into a given text box and then press the "Voice" button to hear it in English

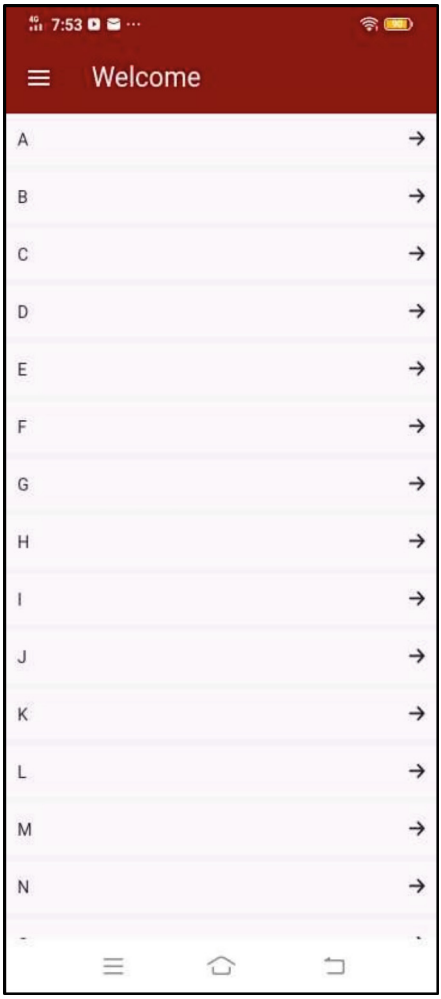


Fig. 7. Manobo to English Dictionary

Figure 7 shows the translation of Manobo words into English. An A-Z index and alphabetical arrangement assist the reader in quickly scanning the vocabulary listed on this page. By selecting a letter, a user can go directly to the page with words starting with the chosen letter.



Fig. 8. Text-to-Text, Phrases, and Expressions Dictionary

Figure 8 shows how users can input Manobo text or phrases and receive corresponding English translations. It highlights the system's ability to provide accurate and basic translations for improved communication, catering to both individual words and common expressions used in daily conversations.

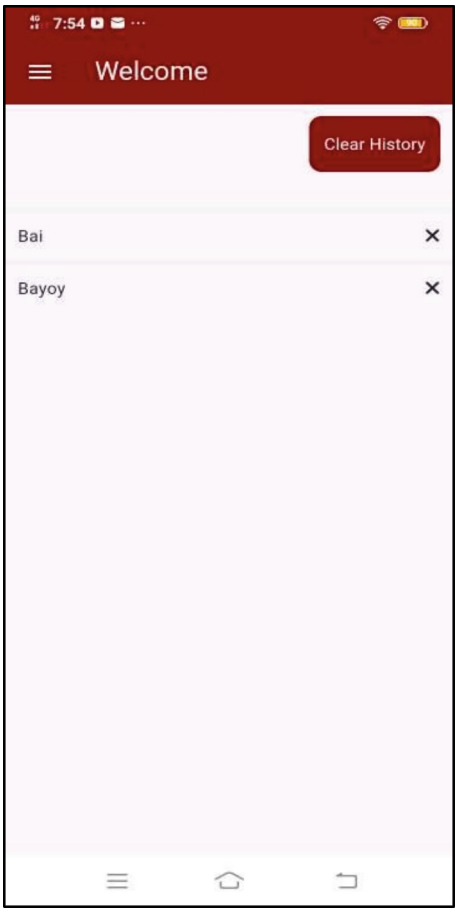


Fig. 9. History

Figure 9 shows a stores a log of all recent translations, including both text and voice inputs. Users can easily scroll through their previous translations, making it convenient to repeat searches or translations. It adds a layer of accessibility and ease to the app, improving the overall user experience.

4.2 Evaluation Results

The system was evaluated using the ISO 25010 quality model. Results are summarized in Table 1.

Table 1 shows that Usability (4.72) received the highest rating, confirming that users found the system easy to use and engaging. Performance Efficiency (4.61) received the lowest rating but was still rated “Strongly Acceptable.”

Table 1. Summary of System Evaluation Results

Evaluation Criteria	Average Mean Value	Verbal Interpretation
Functional Suitability	4.68	Strongly Acceptable
Performance Efficiency	4.61	Strongly Acceptable
Compatibility	4.59	Strongly Acceptable
Usability	4.72	Strongly Acceptable
Reliability	4.62	Strongly Acceptable
Security	4.63	Strongly Acceptable
Maintainability	4.61	Strongly Acceptable
Overall Weighted Mean	**4.65**	**Strongly Acceptable**

4.3 Comparative Analysis

Compared to other systems such as Google Translate and Duolingo, the proposed dictionary is more localized, culturally relevant, and community-driven. It directly addresses the needs of the Manobo community and educational institutions in Agusan del Sur.

5 Conclusion and Recommendations

The Basic Manobo Expression Dictionary: A Mobile-Based Application successfully met its objectives of providing an accessible, user-friendly, and culturally relevant platform for preserving and promoting the Manobo language. Through its features categorized vocabulary, audio pronunciation guides, and offline accessibility, the system effectively bridged the gap between traditional linguistic heritage and modern technology. The evaluation results, based on the ISO 25010 quality model, revealed that the system was highly acceptable across all criteria, with particularly strong ratings in usability and functionality. These findings affirm that the application is not only a practical educational tool but also a significant contribution to cultural preservation efforts.

To further enhance its impact, the researchers recommend the expansion of vocabulary coverage, including sentence structures and conversational contexts, to enrich the learning experience. Future versions may also integrate gamified features to boost engagement, cloud-based storage for wider accessibility, and potential support for other indigenous languages in the region. By continually improving the system, the application can evolve into a comprehensive platform for cultural and linguistic revitalization, ensuring that the Manobo language and heritage remain vibrant and accessible to present and future generations.

References

1. Kaufman, D.: The languages of central and southern Philippines. In: Oxford Guide to the Malayo-Polynesian Languages of Southeast Asia, pp. 347–374. Oxford University Press, Oxford (2024)

2. Abingosa, D., Bokingkito Jr., P., Pasandalan, S.N., Alovera, J.R.G., Otano, J.: Digitizing the Higaonon language: a mobile application for indigenous preservation in the Philippines. Informatics. **12**(3), 90 (2025) MDPI
3. Palmera-Blanco, C.: Phonological features of Manobo spoken in Davao Occidental. Int. J. Linguist. Lit. Transl. **5**(5), 41–55 (2022)
4. Bonifacio, R.M., Zaman, D.J.M.R., Prantilla-Arambala, M.D., Zarate, M.S.: Effects of indigenous language conversation skills enhancement program among Bukidnon and Talaandig youths in the Philippines. Ampersand. **8**, 100076 (2021)
5. Dan, C., Ismail, L., Razali, A.B.: Effectiveness of using mobile dictionary app to enhance vocabulary knowledge of EFL learners with different level of motivation. Int. J. Acad. Res. Bus. Soc. Sci. **14**(7), 204–224 (2024)
6. Siraj, Z.S.: The use of U-Dictionary as Mobile Learning Dictionary application to improve student's English vocabulary: a pre-experimental at SMA Bina Muda Cicalengka. PhD dissertation, UIN Sunan Gunung Djati Bandung (2024)
7. Zakian, M., Xodabande, I., Valizadeh, M., Yousefvand, M.: Out-of-the-classroom learning of English vocabulary by EFL learners: investigating the effectiveness of mobile assisted learning with digital flashcards. Asia-Pac. J. Second Foreign Lang. Educ. **7**(1), 16 (2022)
8. Koleini, N., Boroughani, T., Eslami, Z.R., Xodabande, I.: Exploring the impacts of mobile-assisted learning on university students' technical vocabulary knowledge. Int. J. Educ. Res. Open. **7**, 100344 (2024)
9. Li, Y., Hafner, C.A.: Mobile-assisted vocabulary learning: investigating receptive and productive vocabulary knowledge of Chinese EFL learners. ReCALL. **34**(1), 66–80 (2022)
10. Rahmani, A., Asadi, V., Xodabande, I.: Using mobile devices for vocabulary learning outside the classroom: improving the English as foreign language learners' knowledge of high-frequency words. Front. Psychol. **13**, 899885 (2022)

BaCERA: A Smart Mobile and Web Solution for Community Emergency Response and Disaster Risk Management

Albert D. Silagan(✉), Mery Cris G. Asis, Nancy P. Jacosalem, Jocelyn O. Balolot, Bernie S. Balighot, and Elmer E. Estandarte

College of Computing and Information Sciences, Agusan del Sur State College of Agriculture and Technology, Bunawan, Philippines
{asilagan,mcasis,njacosalem,jbalolot,bbalighot, eeestandarte}@asscat.edu.ph

Abstract. Emergency response in developing cities often suffers from delayed reporting, poor coordination, and limited communication between agencies and the community. To address these challenges, this study introduces a mobile and web-based emergency response application that leverages Android technology, Google Maps integration, geolocation services, and call verification. The system was developed following the waterfall model, encompassing requirements analysis, design, implementation, testing, and maintenance. Beyond facilitating real-time submission of incident reports and generating daily, weekly, and monthly statistics for city officials, the novelty of the proposed approach lies in its integrated, multi-agency platform. Unlike existing solutions that are often single-purpose, the application enables real-time, two-way communication, accurate geolocation tracking, and automated statistical reporting within a unified environment. Its main contribution is providing a scalable ICT-based framework for coordinated emergency management, empowering agencies such as the BFP, PNP, MDRRMC, MHO, and PETSARAG to collaborate seamlessly while engaging citizens as active participants in disaster preparedness and response. Evaluation results demonstrated that the system is efficient, user-friendly, reliable, and effective in enhancing community safety, making it a viable model for replication in other municipalities.

Keywords: Community Safety · Disaster Management · Emergency Response · ICT-Based Application · Multi-Agency Coordination

1 Introduction

The Philippines, with all its natural endowments and bountiful resources, is officially declared to be one of the most disaster-prone nations in the world. Situated along the Pacific Ring of Fire and within the typhoon belt, the country experiences recurring earthquakes, volcanic activity, and an average of 20 typhoons per year, nearly half of which cause massive destruction (Woods, 2011). Typhoon Yolanda in 2013, for instance,

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026
S. Lanka et al. (Eds.): ICTSM 2025, LNNS 1755, pp. 595–610, 2026.
https://doi.org/10.1007/978-3-032-13177-5_45

resulted in thousands of deaths, millions displaced, and widespread damage to infrastructure. Beyond natural disasters, the nation is also highly vulnerable to man-made calamities such as arson, transportation accidents, terrorist attacks, civil disturbances, and cyberattacks (National Disaster Risk Reduction and Management Council, 2014). A case in point is the series of man-made accidents that occurred in Bayugan City, Agusan del Sur, in 2021. These recurring threats underscore the urgent need for robust emergency response and communication systems.

As Information and Communication Technology (ICT) has rapidly advanced, it has opened opportunities for enhancing disaster risk reduction and emergency management. Globally, emergency systems now utilize mobile applications, GPS, and internet-based platforms to improve coordination, streamline real-time communication, and minimize rescue delays (Damaševičius et al., 2023; Urbanelli et al., 2024). Several countries have successfully implemented digital solutions to support both citizens and emergency agencies. However, despite these advancements, many localities still face challenges, such as the absence of standardized response tools, unreliable communication channels, and difficulties in verifying the precise locations of incidents. In places like Bayugan City, where mobile phone usage is highly prevalent, ICT-driven applications offer a promising pathway to improve both community participation and agency responsiveness during emergencies.

Bayugan City, is one of the Caraga Region, continues to face recurring challenges involving both natural and man-made emergencies. From 2015 to 2018, the city experienced high incidences of calamities, yet inefficient communication and uncoordinated systems hindered rapid and effective response (*National Disaster Risk Reduction and Management Council*, 2023). Although multiple emergency service organizations operate within the city including the Bureau of Fire Protection (BFP), Philippine National Police (PNP), Municipal Disaster Risk Reduction and Management Council (MDRRMC), Municipal Health Office (MHO), and the Philippine Emergency Telecommunication Search and Rescue Action Group (PETSARAG)—significant issues remain. These include gaps in documentation, delays in reporting, and the inability to consistently verify and track incident locations in real time. Such persistent gaps weaken interagency collaboration and compromise the city's capacity to ensure timely emergency management.

To address these limitations, this study introduces the Bayugan City Emergency Response Application (BaCERA), an Android-supported and web-based platform specifically designed to enhance coordination between the community and multiple response agencies. The main contribution of this research is the development of an integrated, multi-agency emergency response system that unifies real-time incident reporting, geolocation tracking, bidirectional communication, and automated statistical data generation in a single coordinated platform. Unlike existing systems that are single-purpose or unidirectional, BaCERA demonstrates novelty by enabling seamless collaboration among agencies such as the BFP, PNP, MDRRMC, MHO, and PETSARAG while directly engaging citizens through mobile-based reporting. This approach not only reduces delays and strengthens situational awareness but also introduces a scalable model of ICT-enabled disaster preparedness and response that can be replicated

by other municipalities. By adopting BaCERA, Bayugan City takes a significant step toward building a safer, more resilient, and more responsive community.

2 Related System

2.1 Mobile Crime Incident Reporting System (MCRIS)

The Mobile Crime Incident Reporting System (Bustillo, 2020), using the UX Dimensions Guideline, was created to facilitate police responsiveness to crime reporting through the use of mobile phones and web applications. It used UX design principles like identification, stimulation, and social relatedness to make reporting and responding easy (Bhavesh et al., 2021). This system is connected to the current research, as it illustrates how mobile apps can eliminate reporting delays and enhance decision-making during emergencies, a point supported by BaCERA's objectives of enhancing emergency communication and response.

2.2 AppLERT Mobile Application

The AppLERT Mobile Application was intended for incident and disaster alerting in Metro Manila. The application enables victims to call for assistance in case of emergencies and allows users to warn others of danger areas through social media and GPS-based alerts. This approach is also very much connected with BaCERA, as both are established to reduce response time by utilizing smartphones and geolocation services, finally ensuring timely intervention from emergency teams (Fabito et al., 2016).

2.3 Laguna Patroller Mobile Application

The Laguna Patroller Mobile Application was created to enhance public awareness and enable reporting of violence in Laguna Province (Algaba et al., 2018). It utilizes GPS and image processing technologies to allow users to take photos and send them along with location information to the authorities. This application is pertinent to the present study as it captures the strength of previous reporting systems utilizing mobile, with multimedia and geolocation capabilities, which BaCERA also implements to enhance accuracy and reliability in reporting incidents.

2.4 i911 Mobile Application

The i911 Mobile Application was created as a complement to the Philippine National Police Emergency Hotline 911. It reduces response delays in emergency calls by pre-registering caller information, which is automatically available to dispatchers in emergency situations. The platform also uses geolocation to instantaneously determine a caller's whereabouts. This is strongly connected to BaCERA because both apps seek to increase responsiveness in emergency situations by automating information acquisition and using GPS for quicker coordination with responders (Edillo et al., 2017).

2.5 iSagip Disaster Relief and Monitoring App

The iSagip Disaster Relief and Monitoring App, developed for ABS-CBN's Operation Sagip, exemplifies how ICT can empower affected communities through crowdsourcing, enabling citizens to report their needs while allowing relief agencies to prioritize aid distribution. Similar to BaCERA, it highlights the importance of community-based involvement in disaster management. Existing systems such as the Mobile Crime Incident Reporting System, AppLERT, Laguna Patroller, i911, and iSagip have significantly contributed to enhancing emergency response by incorporating GPS, multimedia reporting, crowdsourcing, and pre-registered user information. However, these platforms remain limited in scope, as they are often designed for specific purposes—such as crime reporting, disaster alerts, or relief distribution—and generally lack an integrated framework that connects multiple emergency service agencies in one system. Moreover, most are oriented toward information dissemination or one-way reporting, without ensuring two-way communication between responders and victims. BaCERA addresses these shortcomings by offering a comprehensive mobile and web-based application tailored for Bayugan City that integrates real-time geolocation reporting, direct communication with emergency units, and automated statistical reporting. This makes BaCERA a more holistic, people-centered, and responsive solution compared to existing systems (Schwab et al., 2017).

3 Methods

3.1 Conceptual Framework of the Study

As illustrated in Figs. 1 and 2, the system combines mobile reporting, geolocation services, two-way communication, and automated data generation into an integrated platform that improves disaster and emergency response. Victims can directly submit real-time reports to emergency agencies through the mobile app, while responders can monitor incidents, track maps, and create statistical reports for informed decision-making using the web-based platform. Salient features of BaCERA are an Incident Reporting system that allows citizens to file comprehensive emergency reports with accurate location information; Two-Way Communication, offering secure and private messaging between responders and victims for quicker response; and Geolocation Tracking, utilizing Google Maps to accurately identify locations of incidents for quicker provision of assistance. The system also includes Statistical Reporting, producing detailed data that agencies can utilize to determine the frequency, type, and location of emergencies, thus enhancing future strategies. Additionally, BaCERA enhances Multi-Agency Coordination through interconnecting emergency service providers like the BFP, PNP, MDRRMC, PETSARAG, and more into a single communication channel, enabling effective, coordinated, and unified disaster response operations.

3.2 Research Design

The research utilized a developmental research design, more particularly the Modified Waterfall Software Process Model as shown in Fig. 2. The design was suitable since it

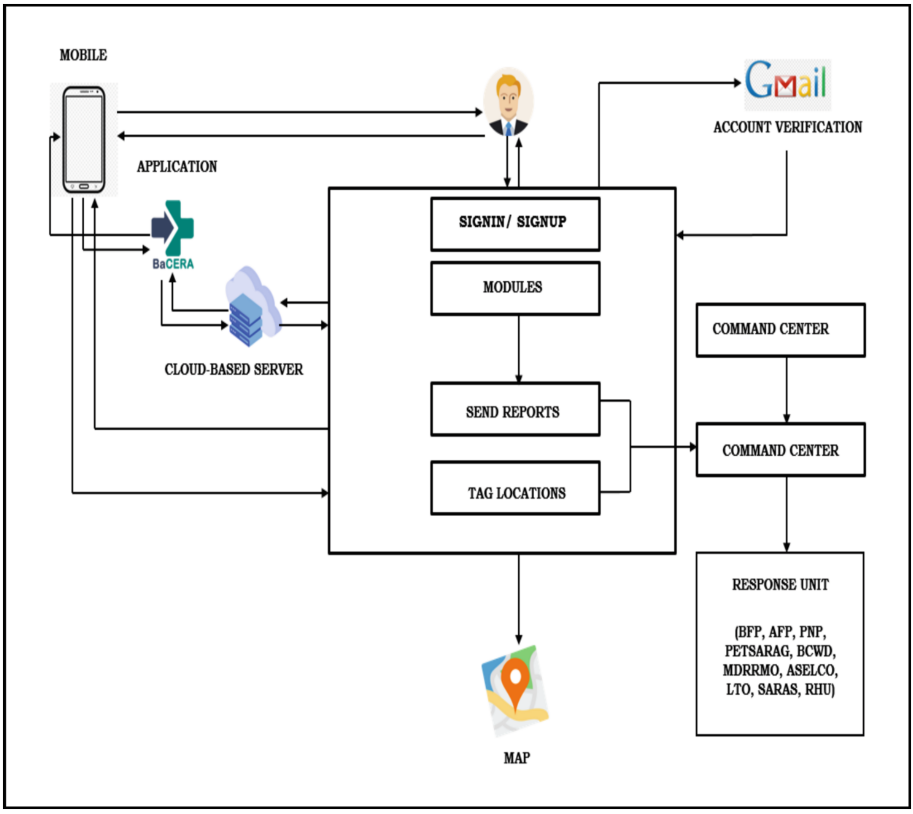


Fig. 1. Conceptual Framework of the Study

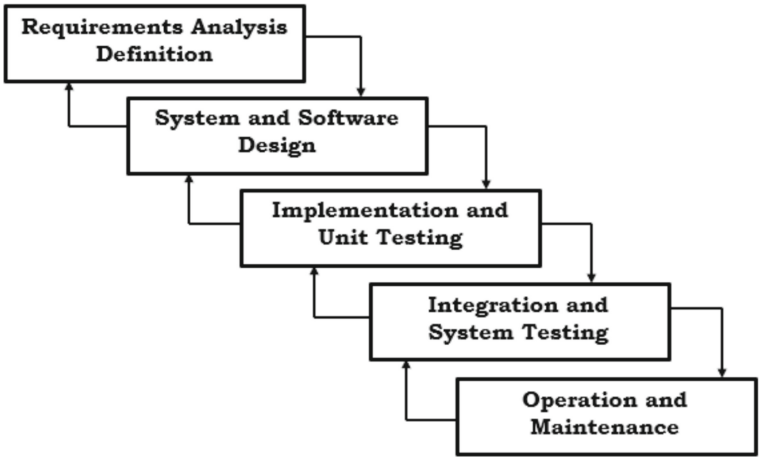


Fig. 2. Modified Waterfall Model

enabled the researchers to proceed sequentially from the steps of requirements gathering, design, implementation, testing, and deployment while still being flexible enough to modify according to feedback from the users. This ensured that the developed system confronted directly the problems of Bayugan City's emergency service agencies and society.

3.3 Phases of Development

1. Requirements Analysis and Definition
   - Field observation and face-to-face interviews were done with PETSARAG and other agencies by researchers.
   - The results indicated there was no official communication tool and a dependence on manual processes, leading to delays.
   - Data collected guided the definition of user requirements.
2. System and Software Design
   - Employing CASE tools like Use Case Diagrams, Deployment Diagrams, and Sequence Diagrams, the researchers planned both mobile and web-based interfaces, identifying the hardware and software requirements.
3. Implementation and Unit Testing
   - Alpha testing was carried out by the developer team to check the correct working of the code, whereas beta testing was carried out with real end-users (agencies and chosen community members) to check usability.
4. Integration and System Testing
   - The system was installed for pilot use within PETSARAG's office and tested for interagency interoperability.
   - Errors and bugs were discovered and resolved to provide the reliability of the system.
5. Operation and Maintenance
   - Training workshops and seminars were undertaken by agency staff and barangay officers. Recommendations for maintenance involved the recruitment of at least one IT staff to supervise long-term technical support and system upgrades.

3.4 Respondents of the Study

The respondents of the study included both emergency service agencies and community representatives to ensure a comprehensive evaluation of the system. Participating agencies were PETSARAG, the Bureau of Fire Protection (BFP), the Philippine National Police (PNP), the Municipal Disaster Risk Reduction and Management Council (MDRRMC), the Armed Forces of the Philippines (AFP), the Land Transportation Office (LTO), Agusan del Sur Electric Cooperative (ASELCO), Bayugan City Water District (BCWD), and the City Health Office (MHO). In addition, barangay captains and selected

residents took part in user testing to provide grassroots-level feedback. The participation of these diverse stakeholders guaranteed that the system was not only technically sound but also responsive to the actual needs and challenges faced by both responders and the community during emergencies. Their collective insights and recommendations were crucial in refining the system into a practical, reliable, and effective tool for disaster response.

3.5 Data Collection Techniques

The research utilized a combination of data collection methods to make sure that the system was well designed and tested based on the requirements of its users. First, interviews were held with PETSARAG staff and representatives of other agencies to obtain comprehensive user requirements, allowing the researchers to know the specific operations and functionality needed in emergency response. Second, observation of actual emergency operations was carried out, enabling the researchers to determine operational gaps, inefficiencies, and challenges to be addressed by the system. Lastly, sessions for alpha and beta testing were performed, during which formal feedback from users was received to evaluate the usability, functionality, and performance of the system. Through the repeated rounds of testing, the system was made finer and more relevant based on actual user experience, ensuring that it was more effective, reliable, and responsive to the needs of emergency response operations.

3.6 Data Analysis

The information collected during the study were closely examined both qualitatively and quantitatively in order to present a comprehensive appraisal of the system. Qualitatively, respondents' comments were closely examined and grouped into categories such as system strengths, weaknesses, and how the system could be improved. This helped the researchers spot recurring problems and areas where the system excelled. On the quantitative end, descriptive statistics were employed to summarize test outcomes in terms of tangible factors like improvement in response time and degree of user satisfaction. Merging these two strategies, the analysis not only brought to light the effectiveness of the system but also presented tangible proof of its influence. Above all, the system improvements were themselves derived directly from actual user experiences so that the finished product was functional, user-friendly, and responsive to the real needs of both emergency responders and the public.

3.7 Tools and Technologies

The system was developed using a combination of software, hardware, and testing tools to ensure functionality and accessibility. C# with Android Xamarin Forms served as the main programming framework, offering a cross-platform environment for mobile app development. Data management relied on SQLite for local storage and a cloud server for real-time processing, while the Google Maps API supported location-based services. The application was designed to run on Android 6.0 (Marshmallow) for mobile

and Windows 10 for desktop compatibility. Development and testing were carried out on an Intel Core i3 computer with 4GB RAM, alongside an Android smartphone and broadband connection for mobile use. Alpha testing in the developer's environment identified and corrected errors, followed by beta testing with agency users to validate usability and reliability. A Canon IP2700 printer was used for documentation, and the Google Play Store was designated as the distribution platform. This integration of tools and processes ensured that the system was efficient, user-friendly, and deployment-ready.

4 Results and Discussion

This part provides the actual findings of the research, complemented by visual and descriptive supports. Screenshots of the BaCERA: A Mobile and Web-Based Application for Emergency Response are given to illustrate the system design, components, and operations. These images indicate how the Android-based BaCERA mobile app and its associated Web-Based platform function in actual settings, specifically in facilitating emergency response operations in Bayugan City. The findings also act as evidence that the goals established in the study were attained successfully. In addition, the discussion expounds on how every feature supports the system's intended purpose, its usability by end-users, and its functionality towards quicker and more efficient emergency operations. Through examination of both technical functioning and user response, this section reconciles the system's theoretical framework with its applied use, making certain that the objectives of the study are not merely fulfilled, but verified by real usage.

4.1 System Features and Implementation

The BaCERA system was conceptualized to provide a speedy, efficient, and coordinated emergency response system in Bayugan City by combining a mobile app for the public and responders with a web application for administrators and command units. Manual processes and restricted communication are usually typical of conventional reporting procedures. To counter this, BaCERA provides accessibility by using a lightweight Android app, promotes immediate communication between emergency unit and users, synchronizes web and mobile systems for incident tracking, and is scalable for future additions, like GIS mapping and analytics. The mobile app has major components such as an Emergency Reporting Module, GPS Location Tracking, a User-Friendly Interface, and notification alerts. The online platform provides an Incident Dashboard, Responder Assignment, Incident Tracking, Data Storage and Reporting, and User Management. Its development was a systematic process: collecting requirements through consultations with emergency units, system designing using user-centered principles, Android and web platform developing using GPS and real-time notification, and testing through alpha, beta, and pilot testing to make it usable and reliable. Implementation entailed setting up the system, responder training, and public launch of the app. Maintenance activities guarantee stability, with allowances for future upgrades, such as AI-based predictions, national hotline interfaces, and disaster analytics, positioning BaCERA as a complete emergency response coordination tool.

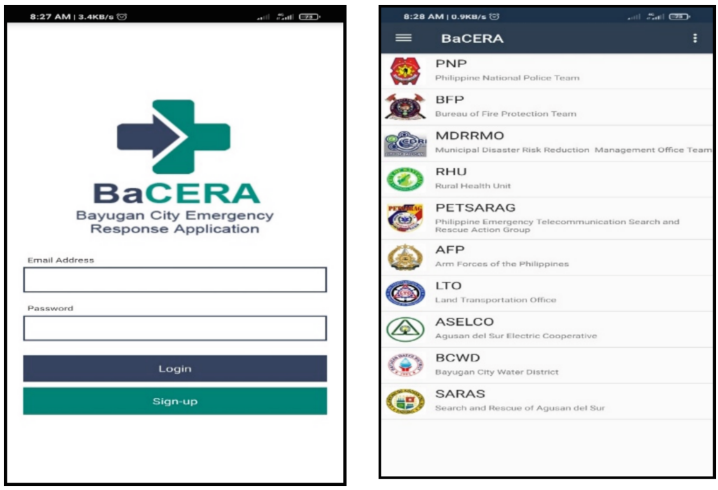


Fig. 3. Mobile Application of BaCERA: Login

Figure 3 shows the homepage of the BaCERA application. The constituent's user can sign-up, create an account to login and choose the different participating agencies for the report of incidents.

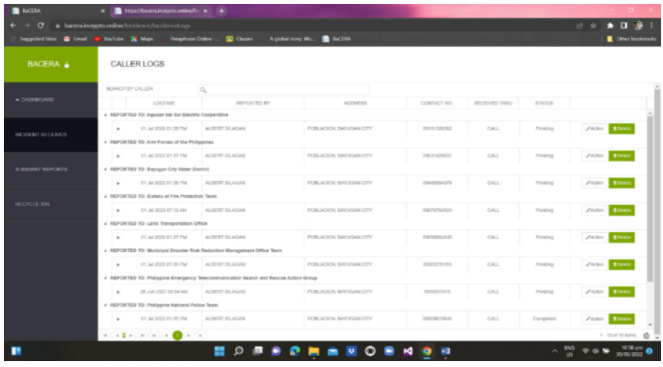


Fig. 4. Incidents receive of BaCERA

Figure 4 shows the incidents received by the system admin and agency admins which can also be viewed through the map. The admin only needs to click the "Go to Incidence List" button and use the Action button to confirm whether the report is legit. The admin also has the capability to delete the report, which will then be automatically moved to the recycle bin.

Figure 5 shows that the admin only needs to input the incident data and select the type of incident from the dropdown list. Once the report has been validated, the admin can click the Save button to finalize the entry.

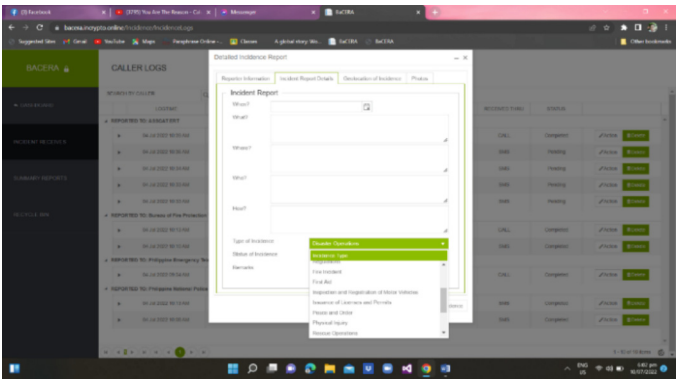


Fig. 5. Detailed Incidents Reports of BaCERA

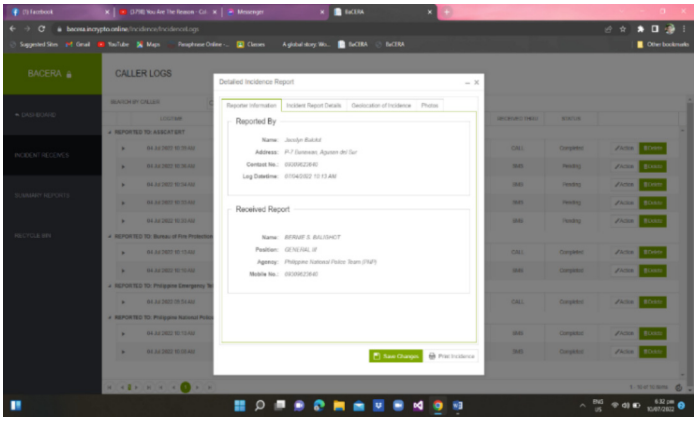


Fig. 6. Web-Based Detailed Incidents Status Reports of BaCERA

Figures 6 and 7 show how to update the status of reports. The admin only needs to select from the dropdown list whether the report is Fixed, a False Alarm, Malicious.

Figure 8 shows that the admin can create a new account whenever a new system administrator needs to be added.

Figure 9 shows that the agencies can only view reports intended for them. Once a solution has been provided, they simply need to click the “Fixed” button to notify the admin that the issue has been resolved.

Figure 10 shows that agencies can create a new account whenever there are new responders assigned to them.

Figure 11 shows the validated printed reports generated by the agency admin.

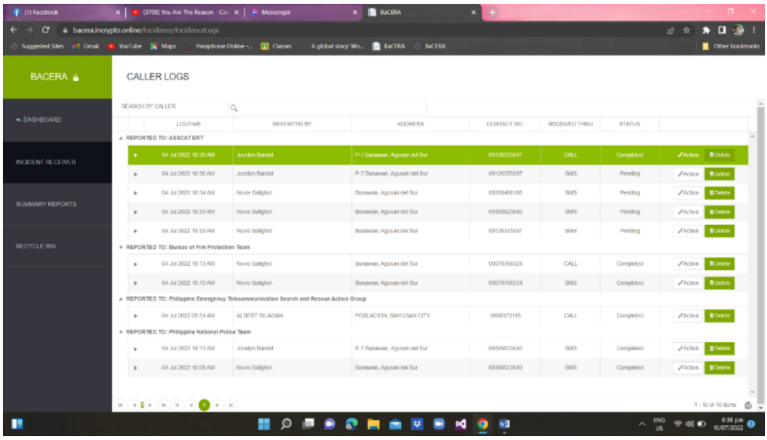


Fig. 7. Web-Based Incidents Logs of BaCERA

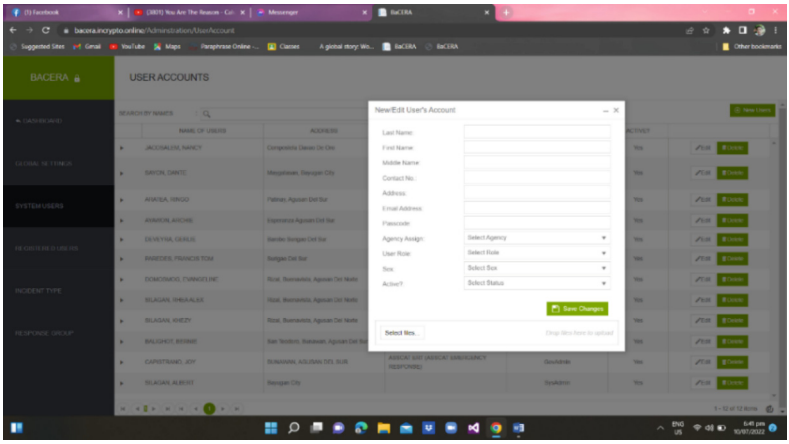


Fig. 8. Create new Admin of BaCERA

4.2 Evaluation Results

The Bayugan City Emergency Response Application (BaCERA) was evaluated with the ISO 25010 quality model, with a focus on factors like Functional Suitability, Performance Efficiency, Compatibility, Usability, Reliability, and Security. The assessment was conducted by barangay officials, health workers, and personnel from emergency response organizations, like the BFP, PNP, MDRRMC, AFP, LTO, ASELCO, BCWD, MHO, SARAS, and PETSARAG. These stakeholders engaged via facilitated surveys, formal interviews, and real system use in simulated emergency situations. The findings of the assessment showed that BaCERA appropriately resolved chronic problems in Bayugan City, such as lagged reporting, ineffective communication, and challenges in

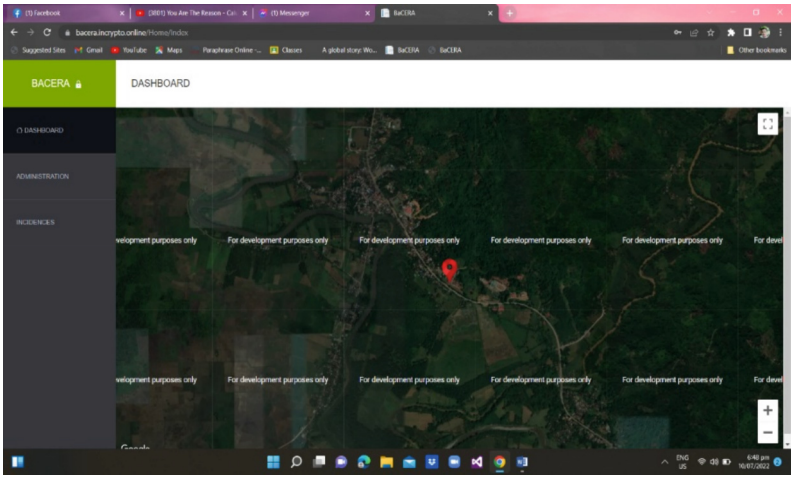


Fig. 9. Agency Home Page

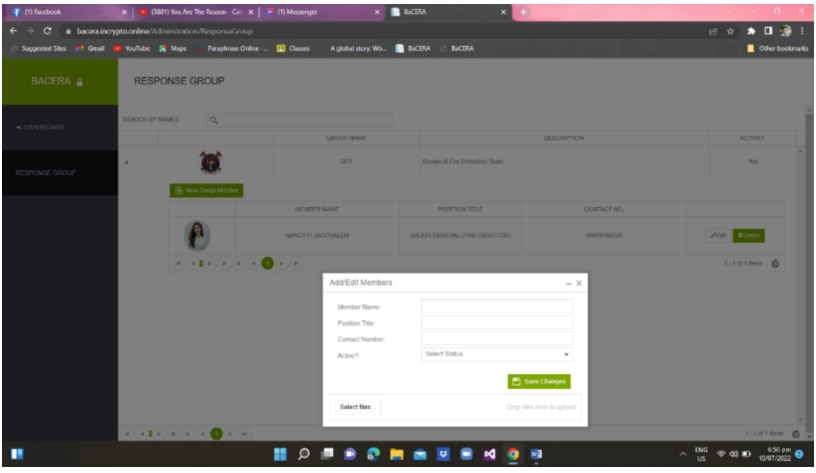


Fig. 10. Agency Response Team

finding exact incident locations. Generally, the system proved to be a promising platform for ensuring effective coordination, speeding up responses, and enhancing disaster readiness in the city.

The evaluation findings in Table 1 reveal that the Bayugan City Emergency Response Application (BaCERA) performed significantly at a high rating in all ISO 25010 factors, with mean values ranging from 4.61 to 4.70. The usability received the highest score (4.70), and this means that the users found the system to be highly usable and intuitive. This was followed closely by Security (4.68) and Reliability (4.67), indicating very high confidence in the system's capacity to safeguard information and maintain stable operation during mission-critical tasks. Other requirements like Functional Suitability (4.65),

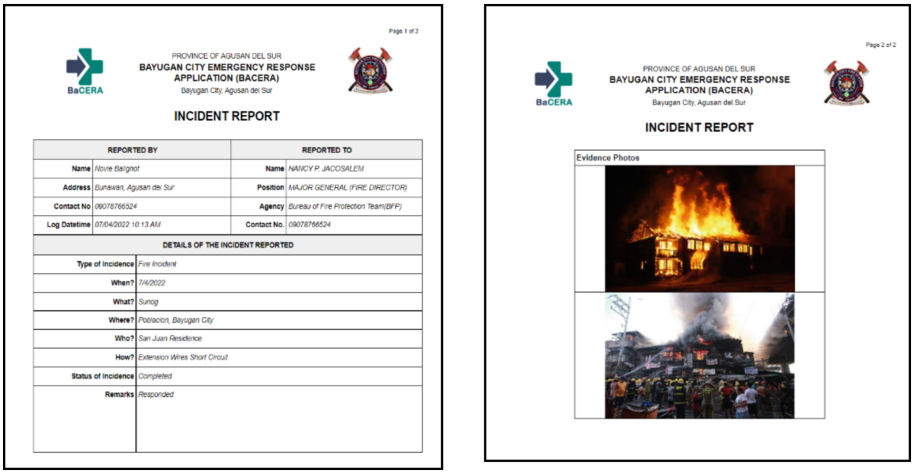


Fig. 11. Print Incidents Report of the BaCERA

Table 1. Summary of System Evaluation Results

Evaluation Criteria	Average Mean Value (AMV)	Verbal Interpretation
Functional Suitability	4.65	Strongly Acceptable
Performance Efficiency	4.62	Strongly Acceptable
Compatibility	4.63	Strongly Acceptable
Usability	4.70	Strongly Acceptable
Reliability	4.67	Strongly Acceptable
Security	4.68	Strongly Acceptable
Maintainability	4.61	Strongly Acceptable
Overall Weighted Mean	4.65	Strongly Acceptable

Performance Efficiency (4.62), Compatibility (4.63), and Maintainability (4.61) also received very high ratings, affirming that BaCERA indeed performs well at its intended purpose, runs efficiently under various environments, and is able to be maintained in the long term. Having a weighted mean of 4.65, the findings confirm that BaCERA is an effective and valid instrument for the improvement of emergency response activities in Bayugan City.

The comparative review of similar systems and BaCERA give in Table 2 points to the special strength of the suggested system in dealing with several emergency response issues at a time. Though current systems such as MCRIS, AppLERT, Laguna Patroller, i911, and iSagip are each designed to address particular functions they tend to be constrained to one-way communication and single-purpose scope. For example, MCRIS is police-oriented, AppLERT is based on social media notifications, Laguna Patroller is

Table 2. Comparative Results with Related Systems

Criteria/Features	MCRIS (2.1)	AppLERT (2.2)	Laguna Patroller (2.3)	i911 (2.4)	iSagip (2.5)	Proposed System: BaCERA
Primary Purpose	Crime incident reporting for faster police response	Disaster and hazard alerting in Metro Manila	Violence reporting and awareness in Laguna	Emergency response complement to Hotline 911	Relief monitoring and prioritization for disaster victims	Comprehensive emergency response system for Bayugan City
Type of Emergencies Covered	Crime-related incidents	Natural and man-made disasters	Violence and crime reporting	Any emergency (police-related focus)	Relief needs during disasters	Fire, medical, crime, disaster, and other emergencies
Platform Used	Mobile + Web	Mobile + Social Media	Mobile	Mobile + Hotline 911 system	Mobile (crowdsourcing-based)	Mobile + Web (synchronized)
Key Features	UX-guided design, easy reporting	GPS alerts, danger warnings via social media	GPS + image capture and reporting	Pre-registered caller info, geolocation tracking	Crowdsourcing reports, community relief needs	Real-time incident reporting, GPS tracking, two-way communication, and statistical reporting
Geolocation Integration	Limited	Yes	Yes (with images)	Yes	Yes	Advanced Google Maps integration with precise location tracking
Communication Flow	One-way (report to police)	One-way (report + warn others)	One-way (report to authorities)	One-way (caller to hotline)	One-way (community to relief agencies)	Two-way (victims ↔ responders, real-time updates)
Multi-Agency Coordination	None (police-focused)	None (alert-focused)	Limited (local authorities)	Limited (PNP-focused)	Relief agencies only	Full integration (BFP, PNP, MDRRMC, PETSARAG, AFP, MHO, etc.)
Reporting Tools	Basic reporting fields	Alerts + GPS	GPS + images	Caller database + GPS	Relief needs reporting	Mobile reporting forms, GPS, real-time dashboard, statistical data
Data Analytics / Reports	Minimal	None	Minimal	Limited	Relief prioritization data	Statistical reports on frequency, type, and location of emergencies

(*continued*)

Table 2. (*continued*)

Criteria/Features	MCRIS (2.1)	AppLERT (2.2)	Laguna Patroller (2.3)	i911 (2.4)	iSagip (2.5)	Proposed System: BaCERA
Target Users	Police + citizens	Citizens + responders in Metro Manila	Citizens in Laguna	PNP + citizens nationwide	Relief agencies + disaster-affected communities	Citizens, barangay officials, and all emergency responders in Bayugan City
System Limitation	Crime-specific, no multi-agency coordination	Alert-only, lacks responder integration	Violence-focused, limited to image/GPS	Hotline complement, not a full platform	Relief distribution only, lacks direct response	Unified platform addressing multiple emergency types with end-to-end coordination

violence report-based with GPS and photos, i911 assists hotline operations with pre-registered caller information, and iSagip prioritizes relief through crowdsourcing. BaCERA, on the other hand, unites mobile and web platforms within an all-encompassing system that caters to a variety of emergencies such as fire, medical emergencies, crime, and disaster. It has real-time incident reporting, sophisticated GPS tracking integrated with Google Maps, two-way communication between the victim and the responder, and complete coordination among agencies such as BFP, PNP, MDRRMC, AFP, and health offices. It also offers statistical reporting for analysis, hence more complete, user-friendly, and responsive compared to current systems, thereby complementing the deficiencies of the previous applications.

5 Conclusion and Recommendation

The development and evaluation of the Bayugan City Emergency Response Application (BaCERA) establish its value as a comprehensive mobile and web-based system that addresses the long-standing issues of delayed reporting, ineffective communication, and fragmented coordination among emergency response units in Bayugan City. Unlike existing platforms that are often single-purpose or limited to one-way communication, BaCERA's primary contribution lies in its integration of real-time incident reporting, advanced geolocation tracking, two-way communication between citizens and responders, and automated statistical data generation into a unified, multi-agency platform. Its strong performance across ISO 25010 quality criteria demonstrates that it is reliable, secure, user-friendly, and effective in enhancing disaster preparedness and response operations. By enabling agencies such as the BFP, PNP, MDRRMC, MHO, and PETSARAG to collaborate within one coordinated environment, BaCERA improves community safety and strengthens disaster readiness. To sustain its impact, future enhancements should include artificial intelligence for predictive analytics, expanded offline capabilities for low-connectivity areas, and continuous training for both responders and citizens. Moreover, institutional adoption by local government units and potential regional expansion could transform BaCERA into a scalable model of ICT-enabled emergency response in the Philippines.

References

Algaba, M.F.S., Peji, R.A., Maria, A.L.S., Bawica, J.M.: Laguna Patroller: mobile application for public awareness about violence with global positioning system and image processing. In: 2018 IEEE 10th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM), pp. 1–5 (2018). https://doi.org/10.1109/HNICEM.2018.8666393

Bhavesh, T.Y., Kirti, N.P., Rajendra, P.R., Jain, N.: Real-time emergency reporting and handling system. In: 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS), pp. 1683–1687 (2021) https://api.semanticscholar.org/CorpusID:233262230

Bustillo, J.C.M.: Automated incident reporting management system using mobile technology. Int. J. Innov. Manag. Technol. **March**, 18–26 (2020). https://doi.org/10.18178/ijimt.2020.11.1.870

Damaševičius, R., Bacanin, N., Misra, S.: From sensors to safety: internet of emergency services (IoES) for emergency response and disaster management. J. Sens. Actuator Networks. **12**(3), 41 (2023). https://doi.org/10.3390/jsan12030041

Edillo, S.B., Garrote, P.J.E., Domingo, L.C.C., Malapit, A.G., Fabito, B.S.: A mobile based emergency reporting application for the Philippine National Police Emergency Hotline 911: a case for the development of i911. In: 2017 Tenth International Conference on Mobile Computing and Ubiquitous Network (ICMU), pp. 1–4 (2017) https://api.semanticscholar.org/CorpusID:4596462

Fabito, B.S., Balahadia, F.F., Cabatlao, J.D.N.: AppLERT: A mobile application for incident and disaster notification for Metro Manila. In: 2016 IEEE Region 10 Symposium (TENSYMP), pp. 288–292 (2016) https://api.semanticscholar.org/CorpusID:25218536

National Disaster Risk Reduction and Management Council: Typhoon Yolanda (Haiyan) 2013 the Philippines, Post-Disaster Needs Assessment (2014)

National Disaster Risk Reduction and Management Council (2023)

Schwab, A.J.M., Omaña, J.E.C., Roazol, K.V., Abe, T.A.Y., Fabito, B.S.: iSagip: a crowdsource disaster relief and monitoring application framework. In: 2017 International Conference on Soft Computing, Intelligent System and Information Technology (ICSIIT), pp. 327–330 (2017) https://api.semanticscholar.org/CorpusID:9578581

Urbanelli, A., Frisiello, A., Bruno, L., Rossi, C.: The ERMES chatbot: a conversational communication tool for improved emergency management and disaster risk reduction. Int. J. Disaster Risk Reduct. **112**, 104792 (2024). https://doi.org/10.1016/j.ijdrr.2024.104792

Woods, D.: The Philippines. In: World History Encyclopedia: Volumes 1–21, 10 (April), vol. 9, pp. 396–397 (2011)

Integrated Community Information and Service Management System

Jeannette V. Quijada(✉), Ziara Amor E. Banggoy, Michelle C. Elape, James Cloyd M. Bustillo, Jeanie R. Delos Arcos, and Jade G. Cuartero

College of Computing and Information Sciences, Agusan del Sur State College of Agriculture and Technology, Bunawan, Philippines
{jeannettevquijada,ziaraamorebanggoy,mcelape,jmbustillo, jrdelosarcos,jcuartero}@asscat.edu.ph

Abstract. This study presents the design and development of the Integrated Community Information and Service Management System to improve efficiency, transparency, and accessibility in barangay administration. Traditional manual processes in Barangay Libertad have long been challenged by slow transactions, data inaccuracy, and limited public access to records. To address these issues, the system was developed using the Systems Development Life Cycle (SDLC) methodology, integrating key modules such as resident and household profiling, certificate and permit requests, financial and payroll management, and health records monitoring. The system was evaluated by barangay officials, staff, and selected residents through surveys, interviews, and observation. Results revealed high acceptability, with the highest ratings in Security (4.71) and Compatibility (4.70), reflecting confidence in data protection and system integration. Performance Efficiency (4.56) received the lowest rating but remained Strongly Acceptable, suggesting room for optimization. The overall weighted mean of 4.63 confirms that the system effectively meets its objectives. The main contribution of this study is the integration of multiple critical barangay functions, record management, financial tracking, legal documentation, and health monitoring into a single comprehensive platform, providing a replicable model for digital governance in local communities.

Keywords: Barangay System · Digital Governance · Record Management · Service Delivery · Transparency

1 Introduction

Barangays serve as the fundamental political units in the Philippines, functioning as the primary providers of basic services, community programs, and administrative support to residents (Ibañez et al., 2024). As the smallest unit of local governance, barangays are tasked with managing resident records, issuing clearances, maintaining health and financial data, and facilitating social programs (Padilla, 2024). However, many barangays, including Barangay Libertad, still rely on traditional paper-based methods, which often lead to inefficiency, errors, and delays in service delivery (Dagohoy, 2021). With the growing population and the increasing demand for faster and more reliable services, the need for an automated management system has become essential (Javaid et al., 2023).

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026
S. Lanka et al. (Eds.): ICTSM 2025, LNNS 1755, pp. 611–624, 2026.
https://doi.org/10.1007/978-3-032-13177-5_46

Several studies emphasize the role of information systems in improving governance and community services. According to Aliling et al., (2025) digital management systems in barangays can enhance transparency, speed up processing, and reduce human errors in records handling. Similarly, a study by David et al., (2023) highlights that integrating information technology in local governance contributes to efficient data management, better decision-making, and improved delivery of frontline services. These findings suggest that adopting computerized systems can address long-standing issues in barangay administration.

Despite these recognized benefits, many barangays continue to face challenges due to limited access to modern technology, lack of digital infrastructure, and dependence on manual filing systems. In Barangay Libertad, residents often experience delays in requesting documents such as certificates, permits, or clearances. Record-keeping also poses a problem as files are prone to damage, loss, or duplication, which compromises data integrity. These issues hinder the barangay's ability to deliver timely and effective services to its constituents.

To address these challenges, the researchers propose the Integrated Community Information and Service Management System. Unlike existing systems that focus only on specific modules such as profiling or clearances, this system integrates multiple functions resident and household profiling, automated certificate and permit issuance, financial transactions and payroll, legal documentation, and health records into one platform. This holistic approach not only streamlines administrative processes but also strengthens transparency and accountability. The main contribution of this paper is the design and implementation of an integrated barangay-level system that combines governance, financial management, and community services, serving as a model for digital transformation in local government units.

2 Related System

2.1 Barangay Information System (BIS)

One related system is the Barangay Information System (BIS), which has been adopted in several municipalities in the Philippines. The BIS was primarily developed to address the inefficiency of paper-based records in barangays. It automates the management of household and resident profiles, allows the issuance of certificates such as residency and indigency, and helps generate statistical reports for submission to higher government agencies. The system has proven useful in streamlining basic administrative tasks (Rogelio Bon. Intud, Jr., 2021). However, its scope is often limited to resident profiling and reporting, with minimal integration of other essential services such as financial transactions, payroll management, and health records.

2.2 eBarangay System

Another related system is the eBarangay System, a web-based platform introduced in some cities to modernize local governance. This system not only handles demographic profiling but also integrates modules for barangay clearance requests, blotter reports, and

community announcements. According to studies, eBarangay has reduced processing time in certificate issuance and improved accessibility of services by enabling online requests (Gonzaga Ballaran et al., 2023). Despite these improvements, its implementation is highly dependent on internet connectivity and the technical capacity of barangay staff, which poses challenges in rural or less technologically equipped areas.

2.3 Community-Based Monitoring System (CBMS)

The Community-Based Monitoring System (CBMS) is another tool utilized by local government units in the Philippines. CBMS focuses on collecting, storing, and analyzing data related to poverty, health, and demographics to aid in planning and policymaking. It provides an extensive database that can be used to identify the needs of communities and monitor development programs (Reyes, n.d.). However, CBMS is more research and planning-oriented rather than service delivery–oriented. It lacks built-in features for day-to-day administrative processes such as certification issuance, financial transactions, or payroll management, which limits its direct application in barangay operations.

2.4 Municipal Information System (MIS)

Some municipalities also use a Municipal Information System (MIS) to centralize records at the municipal level. MIS covers wider data, such as taxation, permits, and inter-barangay statistics, and is often designed for higher-level local government units. While comprehensive in scope, MIS is less granular in addressing the immediate needs of barangays (Diaz et al., 2024). Barangay officials may still need to maintain separate records and manual systems to deal with localized concerns such as household profiling, health monitoring, and financial reports.

2.5 Village Management Systems

Internationally, countries such as India and Indonesia have implemented Village Management Systems, which function similarly to barangay systems in the Philippines. These systems focus on digitalizing resident records, local tax collection, and community development tracking. They have been successful in promoting transparency and accountability, especially in rural governance (Susilowati et al., 2025). However, cultural and administrative differences limit their direct adoption in the Philippine setting. What can be learned, however, is the importance of integrating financial transparency with service delivery, which strengthens citizen trust in local governance. In summary, while prior systems such as BIS, eBarangay, and CBMS have addressed specific challenges in barangay governance, none have offered a unified platform that simultaneously manages resident profiling, health monitoring, certificate issuance, financial transactions, payroll, and legal documentation. The main distinction of the proposed Integrated Community Information and Service Management System lies in its comprehensive integration at the barangay level, designed to be flexible for both offline (local server) and online (cloud) environments. This positions it as a practical and replicable model for rural communities aiming to modernize governance without depending solely on internet connectivity or higher-level systems.

3 Methods

3.1 Conceptual Framework of the Study

Figure 1 shows the conceptual framework of the proposed model

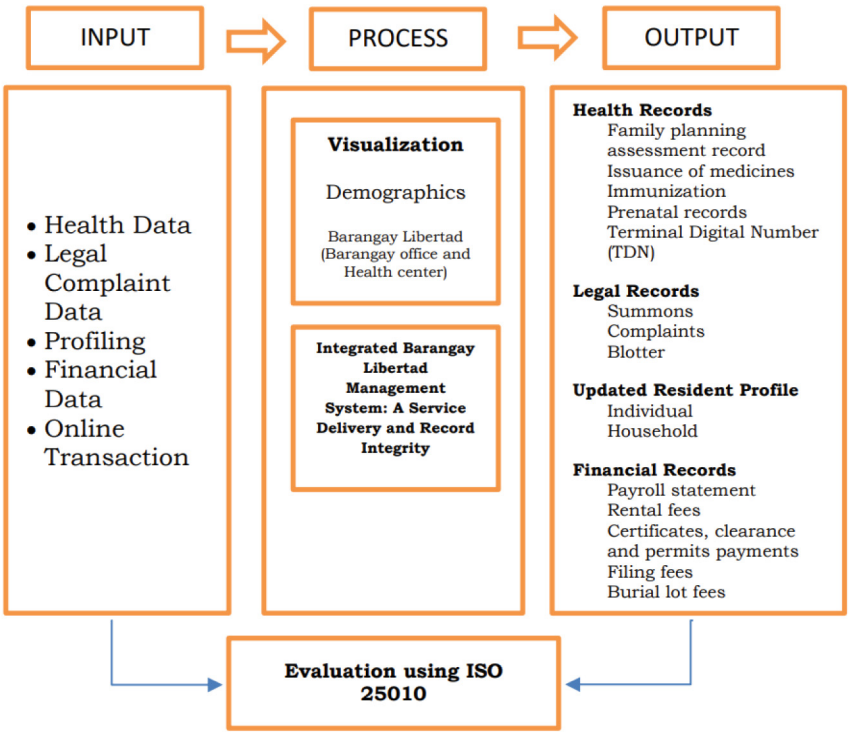


Fig. 1 Conceptual framework of the study

3.2 Research Design

The study will adopt a Research Design (Fig. 2), specifically focusing on the Agile Software Development Process model. This approach is appropriate because the goal of the study is to design, develop, and evaluate a functional information system that addresses the administrative and service delivery needs of Barangay Libertad.

3.3 Phases of Development

The researchers will use the Systems Development Life Cycle (SDLC) as the framework, consisting of the following phases:

1. Planning Phase

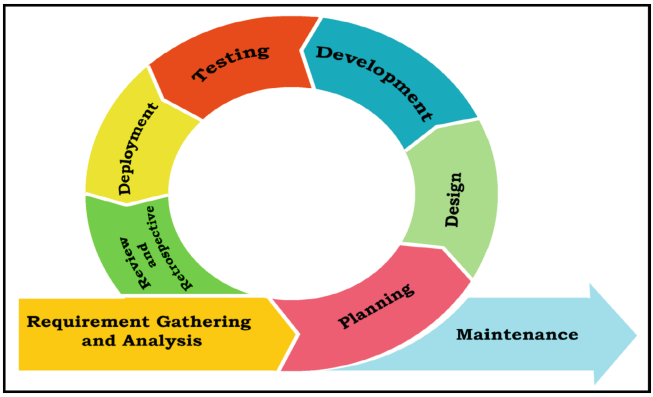


Fig. 2 Agile Methodology

- Identify problems in the existing manual system through interviews and observations.
- Define project scope, objectives, and target users.
- Conduct feasibility analysis (technical, operational, and economic).

2. Analysis Phase

- Gather user requirements through surveys and consultations with barangay officials and staff.
- Analyze existing record-keeping practices, workflows, and service delivery processes.
- Identify gaps and limitations of current systems.

3. Design Phase

- Design the system architecture and data flow.
- Create database schema for resident profiling, records management, and service modules.
- Develop user interface prototypes for validation by stakeholders.

4. Development Phase

- Code the system using appropriate technologies (e.g., PHP/Laravel or Python/Django for backend, MySQL for database, React/Bootstrap for frontend).
- Integrate modules such as resident profiling, certificate issuance, health records, financial transactions, and payroll.
- Implement security features like login authentication and role-based access.

5. Testing Phase

- Conduct unit testing, integration testing, and system testing.
- Perform user acceptance testing (UAT) with barangay officials and staff.
- Gather feedback and refine system functions accordingly.

6. Implementation Phase

- Deploy the system at Barangay Libertad using a local server or cloud hosting.
- Train barangay staff on how to use the system effectively.

7. Maintenance Phase

- Provide technical support and system updates.
- Monitor performance and troubleshoot issues as needed.

3.4 Respondents of the Study

The respondents of this study will consist of individuals who are directly involved in the governance and daily operations of Barangay Libertad, as well as those who are recipients of its services. These include the Barangay Chairman and Kagawad Members, who are responsible for decision-making and policy implementation; the Barangay Secretary and Treasurer, who handle documentation and financial management; and the Barangay Health Workers, who oversee health-related services and records. In addition, selected residents of Barangay Libertad will also be included as respondents, since they are the primary beneficiaries of barangay services and can provide valuable feedback regarding the efficiency and effectiveness of the proposed system.

3.5 Data Collection Techniques

The study will employ several data gathering techniques to ensure comprehensive and reliable information is collected. Interviews will be conducted with barangay officials and staff to identify the specific issues and challenges in the existing manual system. Surveys and questionnaires will be distributed to both officials and residents to gather feedback on service efficiency, user satisfaction, and the usability of the proposed system. In addition, direct observation of current workflows and administrative processes will be carried out to provide a clear comparison between the manual procedures and the anticipated improvements under the automated system. These combined methods will allow the researchers to obtain both qualitative and quantitative data necessary for system development and evaluation.

3.6 Data Analysis

The data gathered in this study will be analyzed using both quantitative and qualitative approaches. Descriptive statistics will be employed to interpret survey results, providing summaries in the form of frequencies, percentages, and weighted means to measure user satisfaction and service efficiency. A comparative analysis will then be conducted to determine the differences in processing time, accuracy, and efficiency between the manual system and the proposed automated system. Meanwhile, qualitative feedback from interviews and observations will be carefully examined to identify strengths, weaknesses, and areas for improvement, ensuring that the system is user-friendly, reliable, and responsive to the needs of Barangay Libertad.

3.7 Tools and Technologies

The development of the proposed system will utilize modern programming languages and frameworks to ensure efficiency, reliability, and scalability. For the backend, the researchers will employ PHP with Laravel framework or Python with Django framework, while the frontend will be developed using JavaScript with React.js to provide an interactive and user-friendly interface. The system's data will be managed through a relational database, either MySQL or PostgreSQL, to guarantee secure and organized record storage. For deployment, the system may operate on a local server within the barangay hall or be hosted through a cloud-based server to enable wider accessibility and availability.

4 Results and Discussion

This section presents the results of the study and discusses how the developed Integrated Community Information and Service Management System addressed the problems identified in Barangay Libertad. The findings are organized according to the system's features, evaluation results per objective, general usability, and overall assessment. Tables and figures are included to illustrate the gathered data and provide a clear comparison between the manual and automated processes.

4.1 System Features and Implementation

The Integrated Community Information and Service Management System was developed to streamline the administrative processes of Barangay Libertad. The system includes modules for resident profiling, certificate and permit issuance, barangay health records, legal document management, financial transactions, and payroll.

The system was deployed on a local server during the testing phase, with the option for future migration to a cloud-based server for wider accessibility. User training was conducted for barangay officials and staff to ensure proper system usage. Implementation results show that the system was able to replace the manual filing system with a secured digital platform, enabling faster service delivery and improved record integrity.



Fig. 3 User Authentication

Provides secure access to the system with role-based logins for residents, staff, and officials, as illustrated in Fig. 3.

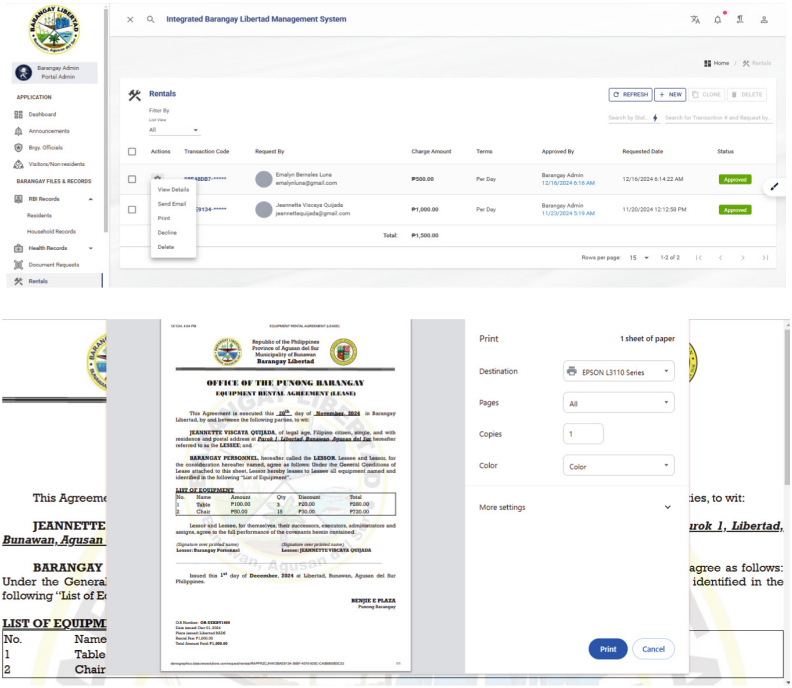


Fig. 4 Dashboard

A central hub that presents key statistics, quick links, and system notifications for efficient navigation, as illustrated in Fig. 4.

Manages barangay health services by recording patient check-ups, immunization schedules, prenatal visits, and available medicine supplies, as illustrated in Fig. 5.

Keeps track of complaints, summons, hearings, and resolutions to ensure transparency in barangay legal proceedings, as illustrated in Fig. 6.

Records and monitors rental services (e.g., barangay facilities) and tracks corresponding payments, as illustrated in Fig. 7.

Handles barangay income and expenses while generating financial statements and reports for accountability, as illustrated in Fig. 8.

Automates salary computation for barangay employees and generates payslips for distribution, as illustrated in Fig. 9.

4.2 Evaluation Results

The system was evaluated using the ISO 25010 quality model focusing on Functional Suitability, Performance Efficiency, Compatibility, Usability, Reliability, and Security.

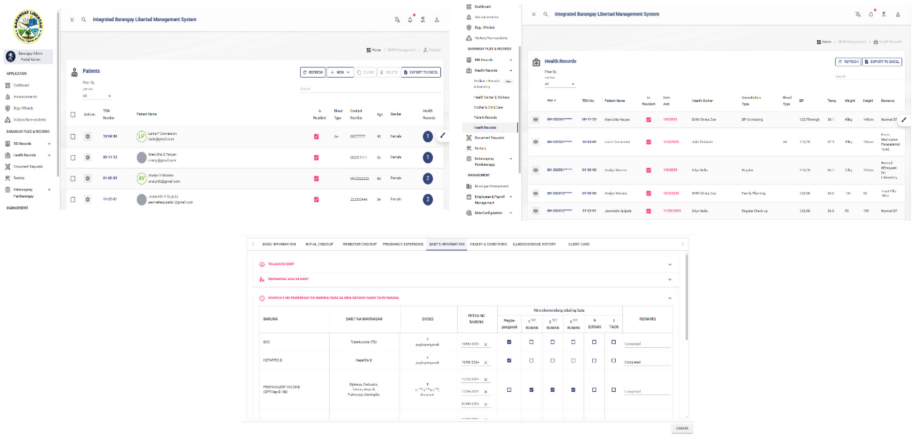


Fig. 5 Health Records (Immunization, Prenatal, Medicine Stock)

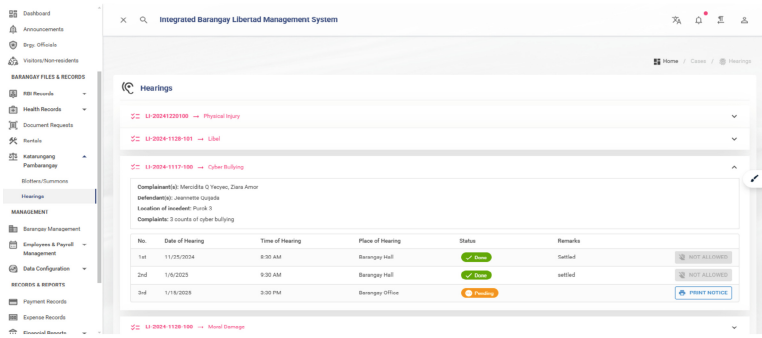


Fig. 6 Hearing Records & Case Management

Barangay officials, health workers, and residents participated in the evaluation through surveys, interviews, and actual system use.

The evaluation results in Table 1 revealed that the system scored highest in Security (4.71), indicating that respondents strongly trust its ability to safeguard sensitive data. Similarly, Compatibility (4.70) received a very high rating, reflecting that the system integrates effectively with existing practices and workflows. On the other hand, the lowest rating was recorded under Performance Efficiency (4.56), which, while still verbally interpreted as *Strongly Acceptable*, suggests that improvements may be considered when handling large or complex data. Overall, the system obtained a Weighted Mean of 4.63, confirming that it is *Strongly Acceptable* and successfully meets the expectations of its users.

As shown in Table 2, the proposed Integrated Community Information and Service Management System offers a more comprehensive and integrated approach compared to related systems. While existing barangay systems focused only on profiling, clearances, or financial monitoring, the proposed system consolidates resident profiling, certificate

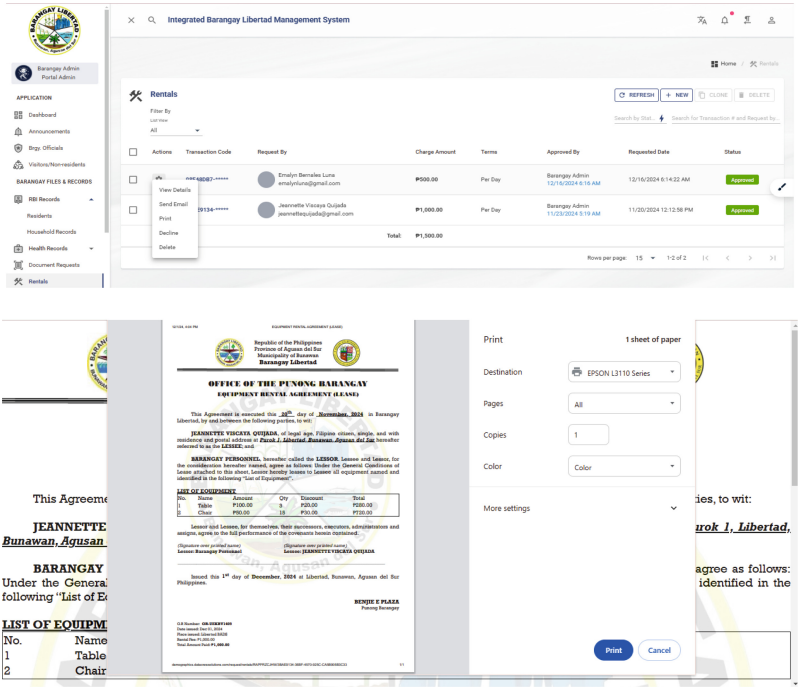


Fig. 7 Rentals and Payments

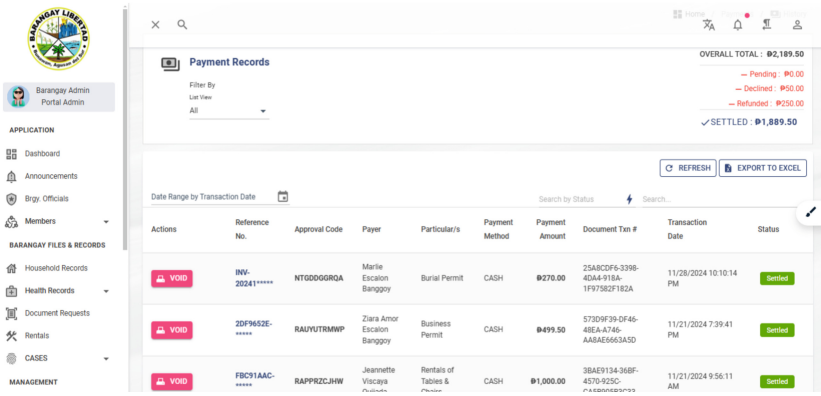


Fig. 8 Financial Transactions & Reports

issuance, health records, financial transactions, and legal documentation into a single platform. Moreover, the deployment flexibility (local or cloud-based) and enhanced security features make it more adaptable to the needs of Barangay Libertad. These results highlight the innovation of the proposed system in bridging the gaps left by previously developed barangay management systems.

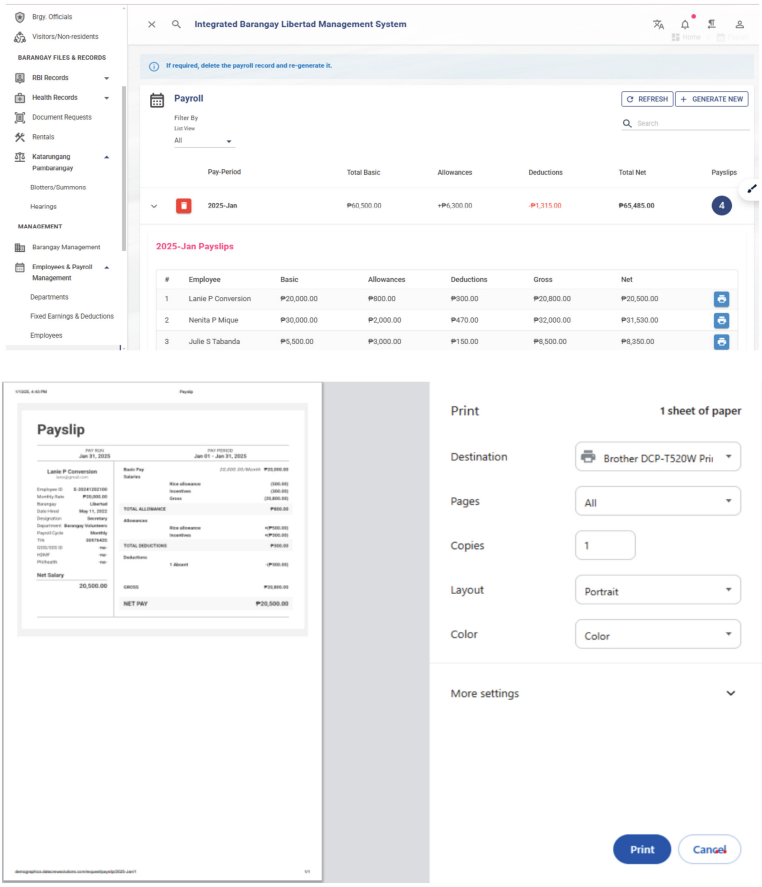


Fig. 9 Payroll and Payslip Generation

Table 1 Summary of System Evaluation Results

Evaluation Criteria	Average Mean Value (AMV)	Verbal Interpretation
Functional Suitability	4.66	Strongly Acceptable
Performance Efficiency	4.56	Strongly Acceptable
Compatibility	4.70	Strongly Acceptable
Usability	4.60	Strongly Acceptable
Reliability	4.60	Strongly Acceptable
Security	4.71	Strongly Acceptable
Maintainability	4.62	Strongly Acceptable
Overall Weighted Mean	**4.63**	**Strongly Acceptable**

Table 2 Comparative Results with Related Systems

Criteria/Features	Related System 1: Barangay Information System (Cruz, 2022)	Related System 2: eBarangay Record Management (Santos, 2023)	Related System 3: Community E-Governance Portal (Lopez, 2024)	Proposed System: Integrated Community Information and Service Management System
Resident Profiling	✔Basic household data only	✔ Detailed individual records	✔ Household and voter information	✔ Comprehensive profiling of households and individuals, with health and demographic data
Certificate & Permit Issuance	✖ Manual process	✔ Automated but limited to barangay clearance	✔ Automated certificates and permits	✔ Fully automated issuance of multiple certificates (clearance, indigency, permits etc.)
Health Records Management	✖ Not available	✖ Not available	✔ Only vaccination tracking	✔ Complete module for health records (check-ups, immunizations, prenatal)
Financial Transactions	✔ Partial (fees only)	✖ Not available	✔ Budget monitoring	✔ Full financial records (fees, rentals, payroll, transactions)
Legal Documents (summons, notices)	✖ Not available	✔ Available	✖ Not available	✔ Integrated with resident records and case tracking
System Deployment	Local server only	Local server only	Web-based, requires stable internet	Local server or cloud-based, scalable for community needs

(*continued*)

Table 2 (*continued*)

Security Features	Basic login only	Password-based login	Role-based access control	Enhanced security with authentication, role-based access, and audit logs
Overall Evaluation	Moderate usability, limited scope	User-friendly, but not comprehensive	Good governance tool, but health/finance modules lacking	Highly usable, comprehensive, and integrated system addressing multiple barangay needs

5 Conclusion and Recommendation

The Integrated Barangay Libertad Management System has successfully met the objectives of the study by providing a secure, efficient, and user-friendly platform for barangay administration. Through its profiling module, the system streamlined the organization of resident and household data, while the online request feature improved accessibility and convenience for the community. Financial records management and automated reporting further enhanced transparency and accountability in barangay operations, reducing manual errors and significantly improving service delivery. Beyond addressing current administrative challenges, the system established a strong foundation for future digital transformation in local governance. To maximize its impact, it is recommended that the system be fully implemented in Barangay Libertad, accompanied by adequate training for barangay officials and staff to ensure proper utilization. While the system already includes an announcement feature, its functionality can be expanded through the integration of disaster management modules and event scheduling to strengthen responsiveness during emergencies and community activities. Sustained performance and adaptability also require regular updates, maintenance, and awareness campaigns to encourage residents to use features such as online requests and profiling, thereby fostering stronger community engagement. With its demonstrated effectiveness, the Integrated Barangay Libertad Management System can serve as a replicable model for other barangays seeking to modernize governance and deliver more responsive, transparent, and inclusive public service.

References

Aliling, J., Santa Ana, B.B., Soberano, P.D., K. T.: Digitalizing governance: a transformation on the processes of one community in the Philippines. Int. J. Multidiscipl. Res. Anal. **08**(06), 3620–3625 (2025). https://doi.org/10.47191/ijmra/v8-i06-68

Dagohoy, R.G.: Capabilities and difficulties of barangay officials on local budgeting process. Int. J. Res. Innov. Soci. Sci. ***05***(05), 325–335 (2021). https://doi.org/10.47772/ijriss.2021.5518

David, A., Yigitcanlar, T., Li, R.Y., Corchado, J.M., Cheong, P.H., Mossberger, K., Mehmood, R.: Understanding local government digital technology adoption strategies: a PRISMA review. Sustainability. **15**(12) (2023). https://doi.org/10.3390/su15129645

Diaz, A.G., Gumtang, A.D., Orpiada, C.J.A., Balagot, A.S., Villanueva, E.A., Manalang, M.A.: PHIrecord: a medical record management system for rural health facilities in the Philippines. In: 2024 IEEE 6th Symposium on Computers & Informatics (ISCI), pp. 188–193 (2024). https://doi.org/10.1109/ISCI62787.2024.10668022

Gonzaga Ballaran, M.J., Joven, M., Donaire, M., Singzon, R.M., Marino Naz, S., Valenzuela Tanael, D., Centeno, C.J.: Development and evaluation of web-based barangay profiling and issuance system using regression analysis. World J. Adv. Res. Rev. **20**(3), 060–072 (2023) Https://Wjarr.Com/Sites/Default/Files/WJARR-2023-2404.Pdf, https://wjarr.com/content/development-and-evaluation-web-based-barangay-profiling-and-issuance-system-using-regression

Ibañez, E., Cornelia, M.I., Bacang, A.G.: Performance of the barangay officials during the pandemic. J. Interdiscip. Perspect. ***2***(3), 132–141 (2024). https://doi.org/10.69569/jip.2024.0041

Javaid, M., Haleem, A., Khan, I.H., Suman, R.: Understanding the potential applications of artificial intelligence in agriculture sector. Adv. Agrochem. **2**(1), 15–30 (2023). https://doi.org/10.1016/j.aac.2022.10.001

Padilla, A.T.: Enhancing local governance in Surigao del Norte through barangay connect: a total management information system for barangays. J. Inf. Syst. Eng. Manag. **2025**(27s), 2468–4376 (2024) https://www.jisem-journal.com/

Reyes, C.M.: Community-based monitoring system in the Philippines: tools for impact monitoring and evaluation (n.d.), pp. 1–44

Rogelio Bon. Intud Jr., R.B.I.J.: Design and implementation of a hybrid barangay information management system. Int. J. Comput. Sci. Mob. Comput. ***10***(5), 24–48 (2021). https://doi.org/10.47760/ijcsmc.2021.v10i05.004

Susilowati, A.P.E., Rachmawati, R., Rijanta, R.: Smart village concept in Indonesia: ICT as determining factor. Heliyon. **11**(1), e41657 (2025). https://doi.org/10.1016/j.heliyon.2025.e41657



Personalized Chatbot Solutions for University Students: A Design Science Approach

Fatima Amer Jid Almahri[1] and Zameer Gulzar[2(✉)]

[1] UTAS, Salalah, Oman
fatima.almahri@utas.edu.om
[2] Malla Reddy University, Hyderabad, Telangana, India
zamir045@gmail.com

Abstract. In Human Computer Interface (HCI) design has long been emphasized. Recent developments in chatbot technology demonstrate a move away from swipes, scrolls, and clicks and toward natural language interfaces that allow text-based interactions. For years, HCI has placed a lot of emphasis on graphic user interface (GUI) design. But the new developments in chatbot technology demonstrate a move toward natural language interfaces, which allow text-based communication in place of swipes and clicks. The wide range of user expectations and behaviors make instructional chatbots designing a challenging task. At present Context and persona-based chatbots for college students are the subject of little research, which emphasizes the need for novel design strategies. This study employed iteration-based design science research (DSR) technique. Machine learning approaches were used for persona elicitation at first. Persona3D, a data-driven approach to persona construction, was used and finally structural equation modeling was incorporated on technology adoption together with previous machine learning findings. Finally, the Persona3D method was realized through the design and construction of several chatbot prototypes using Persona3D models and path mapping. The analysis of machine learning revealed eight different student groups. It was discovered that the main indicators of chatbot usage were performance expectancy, effort expectancy, and habits. Eight chatbots were then created to assist with tasks such as evaluation, assignment advice, knowledge acquisition, and lab support. Several contributions are made by this work, the main one being the Persona3D model and approach. Additionally, this study is the first to look at how persona moderators affect students' adoption and usage of chatbots.

Keywords: Chatbot · Persona · Modelling · Design Science · Interface · Graphics

1 Introduction

Relying on messaging apps and Messenger instead of graphical user interfaces, tech giants like Google, Microsoft, and Facebook see potential and difficulties for HCI [1]. Despite being explored in contexts such as conversational systems and multimodal systems, natural language user interfaces are not new in HCI, yet, their use in mainstream

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026
S. Lanka et al. (Eds.): ICTSM 2025, LNNS 1755, pp. 625–636, 2026.
https://doi.org/10.1007/978-3-032-13177-5_47

applications is growing [2]. The computer program created to mimic an intelligent conversation with one or more human users using written or aural means is called a chatbot [3]. Although the chatbots seems to interpret by understanding input given by the user before it could respond with an answer. Many chatbots are just scanning the keyword from the input and respond with the suitable matching keyword to reply from the main database [4]. Especially the FAQ section on web portals is a typical kind of the problem a user is facing a common problem which could get weird when the number of the questions is either small or crammed in. Chatbots provide a conversational alternative that make these FAQ's feel more user friendly and intuitive also [5]. Adopting chatbots and natural language user interfaces has a lot of ramifications. By comprehending user input and adapting appropriately, chatbots provide the impression of intelligence. Also the user cannot see much of the content or services offered, and the services are heavily dependent on user input. Sometimes, chatbots merely compare the input keywords with database replies. So, chatbot design, is a transition from an interface and visual layout to a conversational design.

The main contributions of this study are: (i) To develop the Persona3D model using machine learning for student personas, (ii) To integrate the extended UTAUT2 model with persona elements to predict the adoption of chatbot (iii) To construct and validate chatbot prototypes tailored to university students. The rest of the paper is structured as follows. Section 3 will describe the DSR Methodology process with respect to chatbot design and development. Section 4 will be concerned with Evaluation and Validation followed by Discussion in Sect. 5 and finally in Sect. 6 conclusion is given.

2 Related Work

Chatbots often uses a one-size-fits-all approach, a strategy to provide the same services to all users, irrespective of their location, needs, preferences, or level of digital literacy [6, 7]. This is particularly valid in teaching and learning environments [8]. A review suggested that to support learning, evaluating students, and some activities at administrative level Artificial Intelligent (AI) chatbots are being used in higher education. Eventhough there arise some challenges related to its adaptability, scalability and most importantly ethics [9]. The potential of education chatbots is visible that they will engage students and personalize their learning experience. But there is lack of solid research on how to design the chatbots effectively and as per personas, cultural adaptation and traits [10]. While Such kind of studies provide a valuable outline, it also point to a gap in strategic designing that account for variety of learners and individual differences.

Personas are a widely used approach in human–computer interaction which is used to help in design processing by acting as standard user models. As per the previous research, the personas can enhance design team communication and promote user-centeric product development [11]. However, some studies also revealed discrepancies in the way persona is viewed by the designers in their work, implying that the traditional personas techniques are limited or vague now [12]. In the recent years, a shift has been seen in the persona related research motivated by data and machine learning. Salminen et al. [13] introduced the methods that will detect and enhance persona related issues automatically. Thus it demonstrate the value of computational based methodology in enhancing the persona validity. The motivation of this study is in line with the development of the proposed Persona3D model fluence machine learning to generate natural and dynamic student personas, inspiring beyond the one-size-fits-all chatbot paradigm.

There aren't much research that look into creating chatbots for various educational personas. To fill this gap, this study will talk about develops persona-based chatbots specifically for college students. Designing a novel solution precludes empirical research, necessitating design science, which builds artifacts for human purposes [14, 15] asserts that complicated structures can be solved by disassembling them into semi-independent parts and developing each one independently to meet the overall goals. The research gap shows that because of its complexity, there aren't enough studies on the use of persona-based chatbots to increase student engagement in higher education. A systematic approach is required for this research. For this reason, Design Science Research (DSR) was selected as the study's approach in this research work. Journey mapping, structural equation modeling, and machine learning are used in three iterations of implementation.

3 DSR Methodology

Design Science Research is providing a systematic way to create and then test information technology solutions in its practical way, and in everyday situations, as pointed by [16]. Later on after Peffers and his team were working on this highlights the importance of thorough testing methods to make sure these solutions actually work and are genuinely useful [17]. Recently, there's been a major shift in how the modern user interfaces should be designed, that use natural language instead of traditional approaches. Here the interaction between the system and the human do not take place by just scrolling or by clicking the button instead interaction takes place by typing or speaking in everyday language. You can see this trend clearly in the latest chatbot technology that's been developed by [18]. The study employs the Design Science Research (DSR) methodology, emphasizing iterative development to design and refine personalized chatbot solutions tailored to university students. The approach integrates machine learning, behavioral models, and user-centered design to ensure the relevance and effectiveness of the solutions. The methodology unfolds in three key iterations as given in Fig. 1:

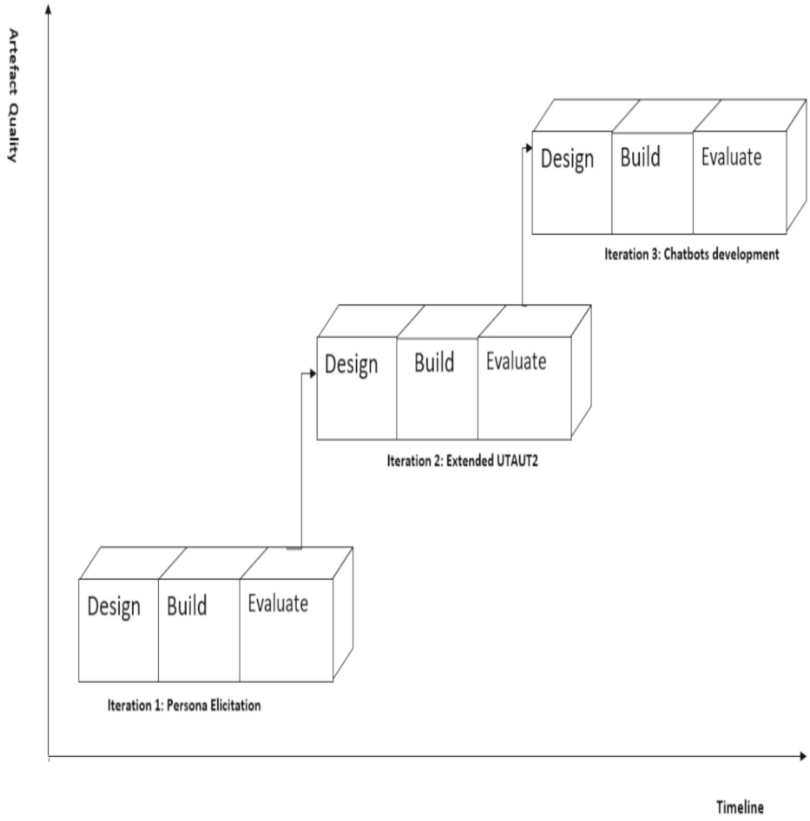


Fig. 1 The Design Science Research Methodology

i) Iteration 1-Persona Development

The first phase focuses on designing a data-driven persona development model using machine learning techniques as shown in Fig. 2. The main Objective in this iteration is to Identify distinct student groups and create detailed personas. The Process involves collecting data related to Student engagement, academic performance, and interaction patterns. Later analysed by supervised machine learning, specifically the K-means clustering algorithm, was used to group students based on similarities which results in development of the artifacts like persona model and template (preferences, challenges and learning behaviour). The output of this iteration is in terms of Personas representing diverse student types were established, serving as the foundation for subsequent iterations.

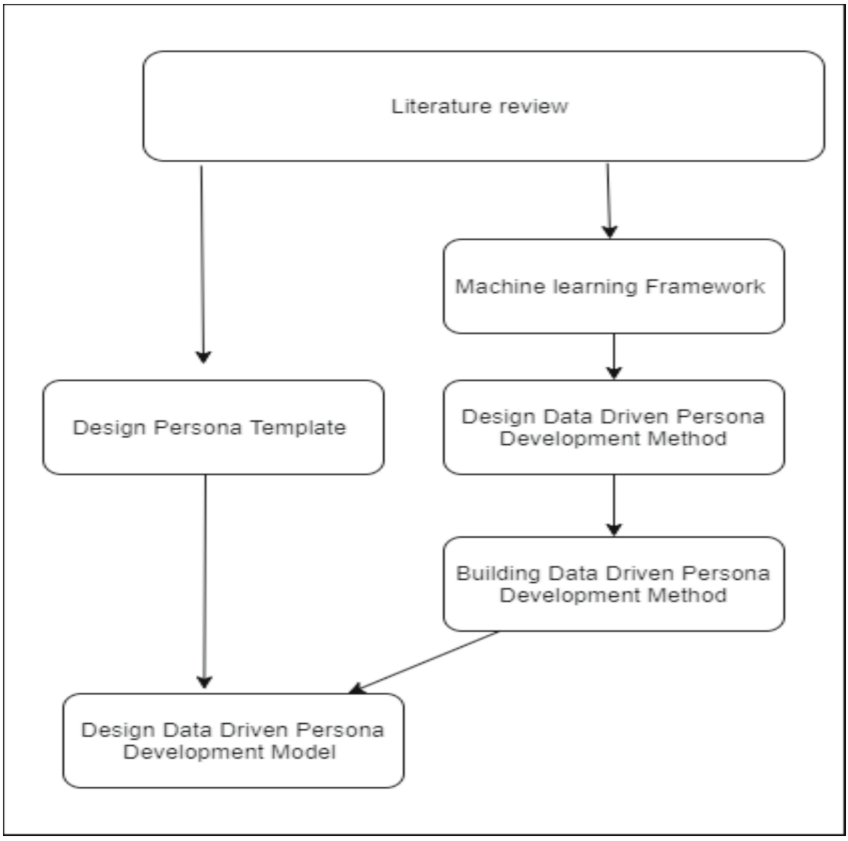


Fig. 2. DSR for Chatbot Design - Iteration One

ii) Iteration 2-Extended UTAUT2 Model

This phase integrates behavioral theories into the chatbot design. The objective is to Extend the UTAUT2 model with persona elements (like persona characteristics) to predict chatbot acceptance as shown in Fig. 3. In this iteration a structured survey, based on the extended UTAUT2 model, was designed to assess student attitudes toward chatbot use and also Hypotheses about technology acceptance were tested using Structural Equation Modeling (SEM). The valid framework linking personas to chatbot interaction behaviors we developed along with Insights into the factors influencing chatbot adoption. The output is in terms of identification of critical features for chatbot design, aligned with student personas and acceptance drivers.

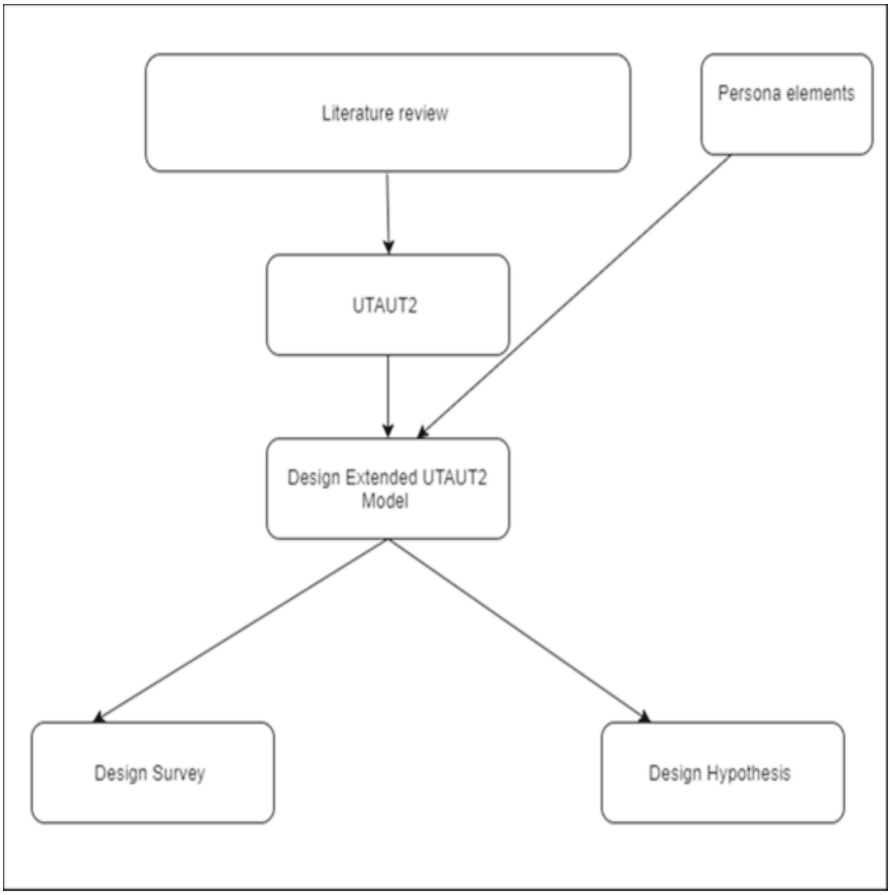


Fig. 3. DSR for Chatbot Design - Iteration Two

iii) Iteration 3: Chatbot Development and Journey Mapping

The creation and assessment of chatbots suited to student identities was the main goal of the third iteration. Persona insights and UTAUT2 structures guided the design of the chatbot, while journey mapping was utilized to illustrate user interactions and pinpoint important touchpoints as shown in Fig. 4. Platforms like Amazon Alexa were used in development, and the Jovo Framework and AWS services (such Lambda and DynamoDB) allowed for reliable testing. In order to answer persona-specific needs, the iteration generated functioning chatbot prototypes and path mapping templates, which were then assessed for usability and conformity to user expectations. The Fig. 5 (a, b, c) displayed combination of results.

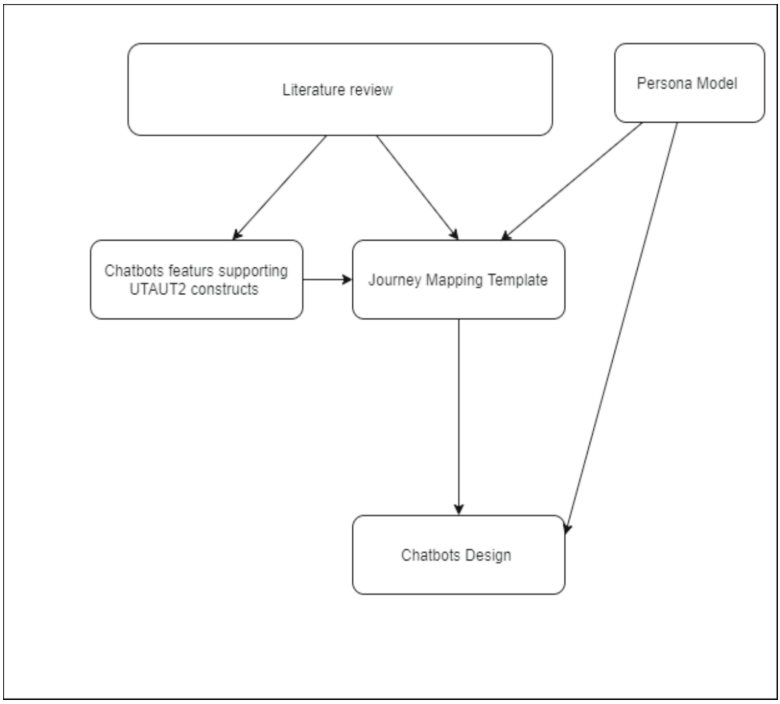


Fig. 4. DSR for Chatbot Design – Iteration Three

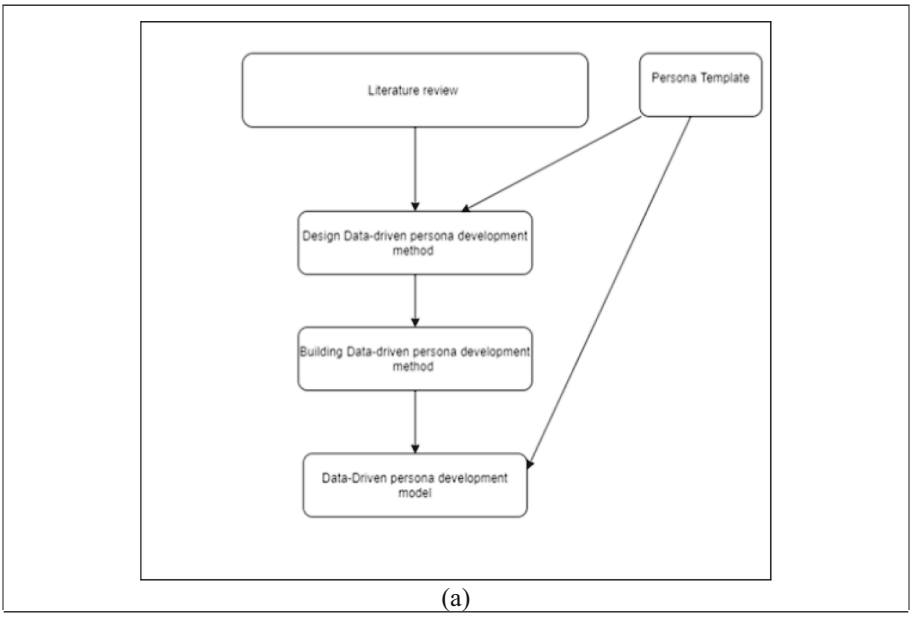
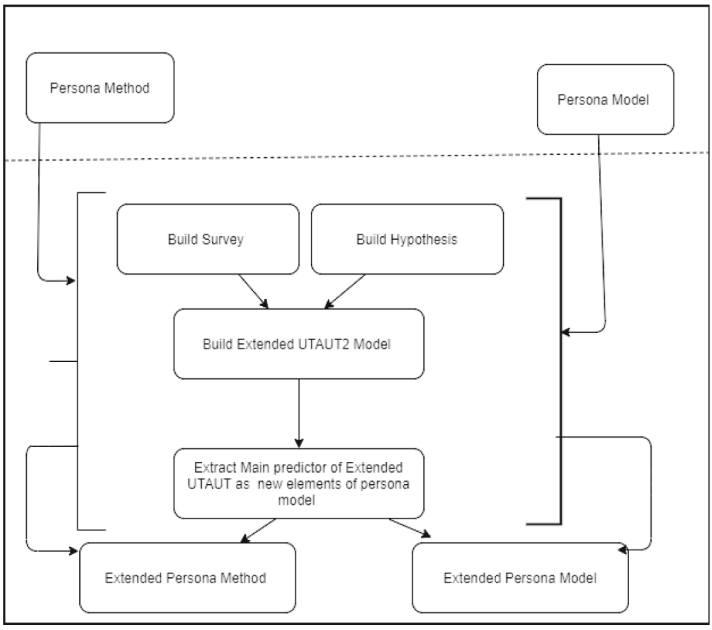
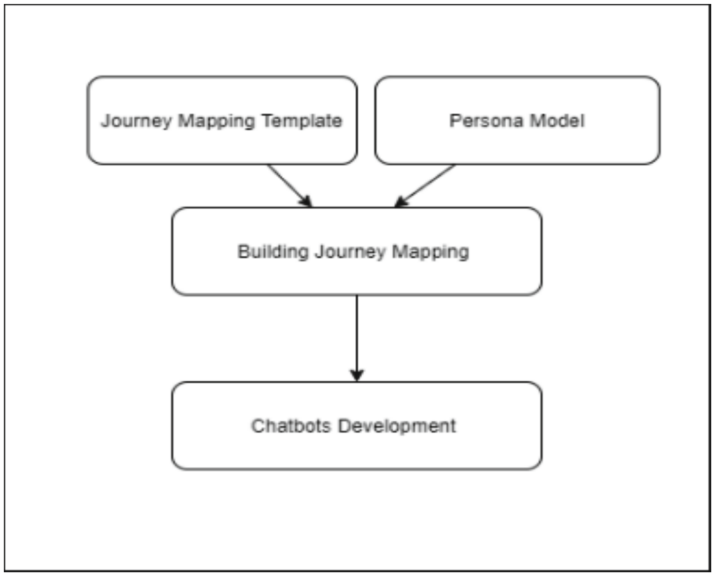


Fig. 5 (a), (b), (c). DSR for Chatbot Building – Iteration One-Two-Three



(b)



(c)

Fig. 5 (*continued*)

4 Evaluation and Validation

The core of the DSR methodology lies in the assessment of the artifacts, even though it was not parallelly developed with changing information systems [19]. Each iteration should be rigorously evaluated to ensure the effectiveness of the artefacts. Metrics such as completeness, efficiency, and usability can be used to assess outcomes [20]. Comparative analysis validated persona models against existing literature, while Structural Equation Modeling (SEM) tested and refined the UTAUT2 extensions. Additionally, user testing of chatbot prototypes ensured functionality and satisfaction, reinforcing their alignment with user needs. The study employed triangulation to enhance reliability and robustness.

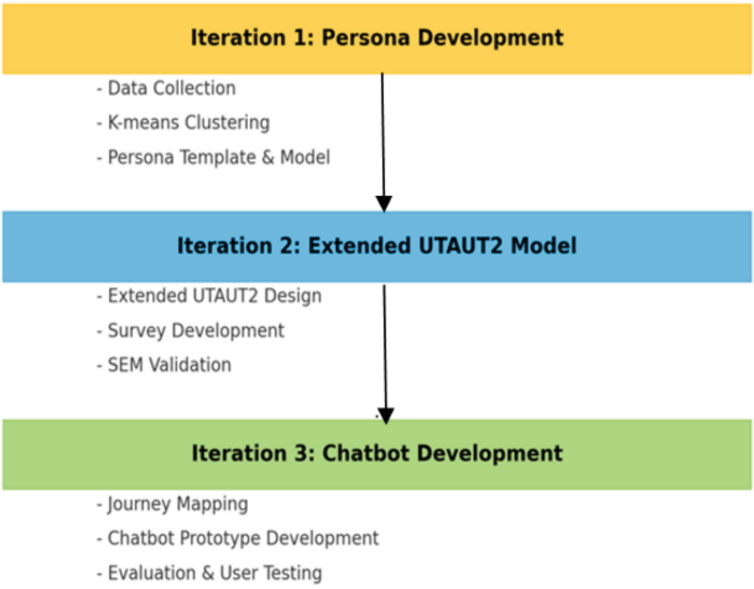


Fig. 6. Design Research Iterations for Chatbot Development

Data triangulation integrated diverse sources, such as survey responses and persona characteristics, to provide a comprehensive understanding [21]. Methodological triangulation combined mixed methods, including machine learning and surveys, to validate findings and ensure consistency across approaches. This methodology ensures that the chatbot solutions are data-driven, user-centric, and aligned with the specific needs of university students, contributing to both academic engagement and technological innovation as shown in Fig. 6.

5 Discussion

The Table 1 highlights the structured methodology and outcomes, showcasing the systematic development of chatbot solutions tailored to university students.

Table 1. Key Findings from the study

Iteration	Objective	Artefacts developed	Key Findings
Iteration 1	Persona Development	Persona Template-Data-Driven Persona Development Model	Personas effectively categorized students based on data (engagement, performance).
Iteration 2	Extend UTAUT2 Model	Extended UTAUT2 Model, – Extended Survey	Validated relationships between technology acceptance and persona-based variables using SEM.
Iteration 3	Chatbot Development and Journey Mapping	Chatbot Prototypes-Journey Mapping Templates	Chatbots tailored to personas improved user engagement and usability, addressing pain points.

The study's iterative DSR approach provided a systematic framework for addressing the complexity of personalized chatbot development. Key findings include:

(a) **Effectiveness of Personas:** The data-driven persona model proved essential in understanding and segmenting user needs, enabling targeted solutions.
(b) **Integration of Behavioral Models:** Extending UTAUT2 with persona based insights demonstrated the value of combining behavioral theories with technological innovation.
(c) **Role of Journey Mapping:** This method bridged the gap between persona insights and practical design, enhancing usability and aligning with user interaction expectations.

Despite the positive outcomes, challenges remain. For instance, the reliance on specific platforms (e.g., Amazon Alexa) may limit flexibility, and additional user testing across diverse demographics could provide further validation. Nonetheless, the study highlights the transformative potential of combining machine learning, behavioral models, and user-centered design in developing personalized chatbot solutions for higher education.

6 Conclusion

The study highlights the effectiveness of the DSR approach in designing and building persona-based chatbots tailored for higher education. By integrating machine learning, extended behavioral models (UTAUT2), and user-centered design tools, the research

successfully developed prototypes that enhance user engagement and address individual learning preferences. The iterative approach ensures adaptability and continuous refinement, demonstrating the utility of chatbots in improving student support systems in universities. The integration of journey mapping further enhanced the design by aligning chatbot features with user interaction patterns, ensuring usability and relevance. The resulting chatbot prototypes showcase the capability to deliver customized support, improve learning experiences, and foster engagement among university students. By addressing the unique requirements of students, this research underscores the value of personalized AI-driven tools in higher education. In summary, the paper makes a contributes by validating chatbot prototypes designed for college students, adding persona factors to UTAUT2, and introduce the Persona3D model for persona construction. This offers a robust framework for future innovations in educational technology.

Acknowledgments. Competing Interest: The authors declare that they have no conflict of interest.

Informed consent: Informed consent was obtained from all individual participants included in the study.

References

1. Følstad, A., Brandtzæg, P.B.: Chatbots and the new world of HCI. Interactions. **24**(4), 38–42 (2017). https://doi.org/10.1145/3085558
2. Følstad, A., Nordheim, C.B., Bjørkli, C.A.: What makes users trust a chatbot for customer service? An exploratory interview study Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 11193 LNCS, pp. 194–208 (2018). https://doi.org/10.1007/978-3-030-01437-7_16
3. Al-Zubaide, H., Issa, A.A.: OntBot: Ontology based ChatBot. In: 2011 4th International Symposium on Innovation in Information and Communication Technology, ISIICT'2011, pp. 7–12 (2011). https://doi.org/10.1109/ISIICT.2011.6149594
4. Dale, R.: The return of the chatbots. Nat. Lang. Eng. **22**(5), 811–817 (2016). https://doi.org/10.1017/S1351324916000243
5. Shaw, A.: Using Chatbots to easily create interactive and intelligent FAQ webpages. J. Appl. Glob. Res. **5**(15), 10–15 (2012)
6. Kadariya, D., Venkataramanan, R., Yip, H.Y., Kalra, M., Thirunarayanan, K., Sheth, A.: KBot: knowledge-enabled personalized chatbot for asthma self-management. In: Proceedings - 2019 IEEE International Conference on Smart Computing, SMARTCOMP 2019, pp. 138–143 (2019). https://doi.org/10.1109/SMARTCOMP.2019.00043
7. Bradesko, L., Wıtbrock, M., Starc, J., Herga, Z., Grobelnık, M., Mladenıc, D.: Curious cat – mobile, context-aware conversational crowdsourcing knowledge acquisition. ACM Trans. Inf. Syst. **35**(4), 1–46 (2017). https://doi.org/10.1145/3086686
8. Yang, S., Evans, C.: Opportunities and challenges in using AI chatbots in higher education. In: ACM International Conference Proceeding Series, pp. 79–83 (2019). https://doi.org/10.1145/3371647.3371659
9. Labadze, L., Grigolia, M., Machaidze, L.: Role of AI chatbots in education: systematic literature review. Int. J. Educ. Technol. High. Educ. **20**, 56 (2023). https://doi.org/10.1186/s41239-023-00426-1

10. Kuhail, M.A., Alturki, N., Alramlawi, S., et al.: Interacting with educational chatbots: a systematic review. Educ. Inf. Technol. **28**, 973–1018 (2023). https://doi.org/10.1007/s10639-022-11177-3
11. Miaskiewicz, T., Kozar, A.K.: Personas and user-centered design: how can personas benefit product design processes? Des. Stud. **32**(5), 417–430 (2011). https://doi.org/10.1016/j.destud.2011.03.003
12. Matthews, T., Tejinder, J., Steve, W.: How do designers and user experience professionals actually perceive and use personas? In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12), pp. 1219–1228. Association for Computing Machinery, New York (2012). https://doi.org/10.1145/2207676.2208573
13. Salminen, J., Jung, S., Nielsen, L., Sercan, S., Jansen, B.J.: How does varying the number of personas affect user perceptions and behavior? Challenging the 'small personas' hypothesis! Int. J. Hum.-Comput. Stud. **168**, 102915 (2022). https://doi.org/10.1016/j.ijhcs.2022.102915
14. March, S.T., Smith, G.F.: Design and natural science research on information technology. Decis. Support. Syst. **15**(4), 251–266 (1995). https://doi.org/10.1016/0167-9236(94)00041-2
15. Simon, H.A.: The sciences of the artificial, (third edition). Computers & Mathematics with Applications. **33**(5) (1996). https://doi.org/10.1016/S0898-1221(97)82941-0
16. Hevner, A.R., March, S.T., Park, J., Ram, S.: Design science in information systems research. MIS Q. **28**(1), 75–105 (2004). https://doi.org/10.2307/25148625
17. Peffers, K., Rothenberger, M., Tuunanen, T., Vaezi, R.: Design science research evaluation. In: Peffers, K., Rothenberger, M., Kuechler, B. (eds.) Design Science Research in Information Systems. Advances in Theory and Practice. DESRIST 2012 Lecture Notes in Computer Science, vol. 7286. Springer, Berlin/Heidelberg (2012). https://doi.org/10.1007/978-3-642-29863-9_29
18. Bradesko, L., Witbrock, M., Starc, J., Herga, Z., Grobelnik, M., Mladenic, D.: Curious cat – Mobile, Context-Aware Conversational Crowdsourcing. **35**(4) (2017)
19. Prat, N., Comyn-Wattiau, I., Akoka, J.: A taxonomy of evaluation methods for information systems artifacts. J. Manag. Inf. Syst. **32**(3), 229–267 (2015). https://doi.org/10.1080/07421222.2015.1099390
20. Qassimi, N.M.: Research Triangulation: Enhancing Validity, Rigor, and Insight Through Multimethod Approaches (2023). https://doi.org/10.5281/zenodo.10430483
21. Boudier, J., Sukhov, A., Netz, J., Le Masson, P., Weil, B.: Idea evaluation as a design process: understanding how experts develop ideas and manage fixations. Des. Sci. **9**, e9 (2023). https://doi.org/10.1017/dsj.2023.7

Author Index

A
Abandula, Kylene Kyte B. 581
Abarna, S. 527
Abas, Pg Emeroylariffion 418
Adeshina, Olusegun Emmanuel 248
Ajmera, Snehlata 56
Alamsyah, Nurwahyu 380
Alegre, Glorybe E. 94
Al-Hussein, Mohammed 248
Alimova, Guzal 323
Almahri, Fatima Amer Jid 625
Aluvalu, Rajanikanth 189
Anetic, Snezana 203
Angelica, Christyfianie Elva 296
Arivazhagan, B. 133
Asis, Mery Cris G. 595

B
Bacanin, Nebojsa 203
Bairagi, Mithun 485
Balamurugan, S. 527
Balighot, Bernie S. 595
Balmora, Ignel T. 581
Balolot, Jocelyn O. 595
Banerjee, Joideep 470
Banggoy, Ziara Amor E. 611
Banu, Sharmila 446
Benedict, Shajulin 145
Berdida, Joen Rose S. 581
Boltayevich, Elov Botir 346
Bondarenko, Darina 456
Botirov, Fayzullajon 215
Bustillo, James Cloyd M. 611
Butte, Sujata 174
Butte, Vijay Kumar 174

C
Cerna, Patrick D. 94
Chennupati, Narendra 81
Chithaluru, Premkumar 189
Cuartero, Jade G. 581, 611

D
Degadwala, Sheshang 45, 56, 110
Delos Arcos, Jeanie R. 611
Demidov, Pavlo 456
Dey, Sushmoy 145
Doan Xuan, Huy-Minh 66
Durdona, Omarova 335
Durga, R. 544

E
Ebadinezhad, Sahar 12, 248, 370
Eborda, Jeanette M. 581
El Yabroudi, Mohammad 81
Elape, Michelle C. 611
Estandarte, Elmer E. 595

F
Fa'ezul Fitri Latiff, Mohammad 418
Filimonova, Tetyana 456
Fountis, Anastasios 560

G
Geyda, Alexander S. 223
Gozali, Lina 1, 296, 498
Guduri, Manisha 81
Gulzar, Zameer 446, 625
Gummadi, Hari Suresh Babu 81

H
Habiba, Djabbarova 335
Hahumaan, Kapilavai 431
Hakim qizi, Primova Mastura 346
Hassan, Rawa Ali 117
Hussein, Ibrahim Khalaf 12
Huynh Nhat, Trung-Hieu 66

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2026
S. Lanka et al. (Eds.): ICTSM 2025, LNNS 1755, pp. 637–639, 2026.
https://doi.org/10.1007/978-3-032-13177-5

I
Inbavel, M. 133
Islam, A. S. M. Tanvir Ul 485

J
Jacosalem, Nancy P. 595
Jamolidinovna, Jurayeva Nargiza 346
Jeni Gracia, R. C. 358
Jeyanthi, P. 544
Jones, JohnsonFoyken 418
Jovanovic, Luka 203
Judijanto, Loso 158

K
Kachhava, Rajendra 56
Khaydarov, Elshod 215
Kobiljonov, Shukhrat 215
Kohlmann, Sabina 560
Krishnan, R. Santhana 527
Kulkarni, Jaimini Santosh 45
Kumar Chandar, S. 30, 509
Kumar, Awanit 56

L
Lemonakis, Christos 560
Liu, Katya Victory 498
Lokesh, Kurakula 431

M
Malik, Mayur 30, 509
Maombe, Joyeuse 370
Maqsudov, Azamat 323
Michael, G. 358
Misic, Katja Udir 560
Muralidharan, Bhagyalakshmi 145

N
Narmanov, Otabek 323
Narmanov, Ulugbek 323
Narsimhulu, Pallati 189
Nguyen Thi, Hong-Nhung 66
Ningsi, Nurfitria 380
Nomozovna, Tojieva Gulbahor 346
Nsimba, Fulbert Milrich Kiminou 370

O
Odil, Mamaraufov 335

P
Padmavathi, K. 133
Pandya, Darshanaben Dipakkumar 45
Pasha, M. Mahmood 446
Patel, Asma 470
Patel, Kathan Nitin 264
Petra, M. I. 418
Pham Le, Phuc-Hung 66
Pojas, Patrick Jasson S. 581
Polavarapu, Ashok 81
PratheepKumar, P. 133
Punjabi, Hitesh 30, 509
Pursky, Oleg 456

Q
Quijada, Jeannette V. 611

R
Radomirovic, Branislav 203
Raj, A. Ferminus 527
Raj, J. Relin Francis 527
Raj, Veena 418
Ramadhan, M. Ikhsan 380
Ramteke, Shefali Vinod 391
Rana, Md. Sohel 485
Rehman, Rizwan 408
Rejapov, Khayrillo 323
Robin, Christhoper 1

S
Sai, Kunchapu Gowri 431
Saide, Saide 380
Salal, Yass Khudheir 117
Samia, Jannatul Mim 485
Sasisri, Nalladimmu 431
Sateeshchandra, Navya Gubbi 560
Sharma, Aviral 264
Silagan, Albert D. 595
Singh, Harjinder 408
SomiSetty, Venkata Manoj Kumar 283
Sumathi, R. 431
Suresh, D. 527
Suryakanth, Sagar 174

T
Tamhankar, Ishaan 45
Thallapally, Vinod 312
Thota, Sandeep Kumar 81

Tiwari, Vineet 391
Tomashevska, Tetyana 456
Trivedi, Vishal Rajendrakumar 45
Tyschenko, Ihor 456

V
Vakhabov, Abdurakhim 323
Varadwaj, Pritish Kumar 391
Varsandán, Csaba 203
Vyas, Dhairya 45, 56, 110
Vyas, Vartika 56

W
Wakil, Hauwa Adamu 370

X
Xakim qizi, Tokhtaeva Matluba 346
Xolisxon, Axmedova 335

Z
Zivkovic, Miodrag 203
Ždímalová, Mária 233