Lecture Notes in Networks and Systems 1651

Simon Fong
Nilanjan Dey
Amit Joshi *Editors*

ICT Analysis and Applications

Proceedings of ICT4SD 2025, Volume 7

Springer

Lecture Notes in Networks and Systems 1651

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).

Simon Fong · Nilanjan Dey · Amit Joshi
Editors

ICT Analysis and Applications

Proceedings of ICT4SD 2025, Volume 7

Springer

Editors
Simon Fong
University of Macau
Macau, China

Amit Joshi
Global Knowledge Research Foundation
Ahmedabad, Gujarat, India

Nilanjan Dey
Computer Science and Engineering
Techno International New Town
Kolkata, West Bengal, India

ISSN 2367-3370 ISSN 2367-3389 (electronic)
Lecture Notes in Networks and Systems
ISBN 978-3-032-06687-9 ISBN 978-3-032-06688-6 (eBook)
https://doi.org/10.1007/978-3-032-06688-6

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

Preface

Tenth International Conference on ICT for Sustainable Development (ICT4SD 2025) targets Theory, Development, Applications, Experiences and Evaluation of Interaction Sciences with fellow students, researchers and practitioners.

The conference is devoted to increasing the understanding role of technology issues and how engineering has day by day evolved to prepare human-friendly technology. The conference will provide a platform for bringing forth significant research and literature across the field of ICT for Sustainable Development and provide an overview of the technologies awaiting unveiling. This interaction will be the focal point for leading experts to share their insights, provide guidance and address participant's questions and concerns.

The conference will be held on 17–19 July 2025 at Hotel Taj Cita De – Horizon, Dauna Paula, Goa. The Conference is organized by Global Knowledge Research Foundation & Managed By: G R Scholastic LLP, State Chamber Partner Goa Chamber of Commerce & Industry, National Chamber Partner Knowledge Chamber of Commerce & Industry & Eco System Partner Goa Technology Association.

Research submissions in various advanced technology areas were received and after a rigorous peer-review process with the help of program committee members and 300 external reviewers for 3100+ papers from 16 different countries out of which 618 were accepted with an acceptance ratio of 0.19. These will be presented in 64 parallel sessions in three days organised Physical at Goa and Virtual on Zoom including 1 inaugural and 2 keynote sessions.

Technology is the driving force of progress in this era of globalization. Information and Communication Technology (ICT) has become a functional requirement for the socio-economic growth and sustained development of any country. The influence of information communications technology (ICT) in shaping the process of globalization, particularly in productivity, commercial and financial spheres, is widely recognized. The ICT sector is undergoing a revolution that has momentous implications for the current and future social and economic situation of all the countries in the world. ICT plays a pivotal role in empowering people for self-efficacy and how it can facilitate this mission to reach out to grassroots level. Finally, it is concluded that ICT is a significant contributor to the success of the ongoing initiative of Start-up India.

In order to recognize and reward the extraordinary performance and achievements by ICT and allied sectors & promote universities, researchers and students through their research work adapting new scientific technologies and innovations. The two days Conference had presentations from the Researchers, Scientists, Academia, and Students on the Research work carried out by them in different sectors.

Contents

Survey on Prediction of Bipolar Disorder Using CNN and LSTM

Ankita Mehta and Shailesh Gahane(✉)

Datta Meghe Institute of Higher Education and Research, Wardha, Maharashtra 442001, India
shailesh.gahane@dmiher.edu.in

Abstract. In this research, we focus specifically on mental disorder which is bipolar disorder using machine learning techniques, utilizing a simple dataset from Kaggle for training and evaluation. The study involves applying various ML models, including R. Forest, XG-Boost, and S. Vector Machines (SVM), on the dataset. We assess the performance of these models using evaluation MSE (how far actual value to predicted value), Precision, and Fl Score to determine their effectiveness in predicting bipolar disorder. However, through extensive experimentation, we found that the combination of (CNN) and (LSTM) networks outperformed the other algorithms, achieving an overall accuracy of 95%.

Keywords: bipolar disorder · ML Models · (SVM) · random forest · gaussian NB · Boost · mean square error · F1score · precision · convolutional · Relu · sigmoid etc.

1 Introduction

A very important factor are mental health and physical well-being, our daily life can be affected by any change in the mental well-being. Very common factors for disturbing any person are depression and anxiety, and as we grow, the risk of having these conditions tends to increase. [7] Fortunately, for predicting and understanding diseases like bipolar disorder advancement in technology specifically artificial intelligence have opened up new chance. BD symptoms are from extreme mood swings, from periods of intense elation to deep depression. Managing such symptoms and enhancing the quality of life for bipolar disorder patients is very essential. Machine learning and deep Learning techniques are advancing the field of medicine by analyzing massive amount of data and it can identify [8].

Patterns which can go untouched. Researchers and scientist can very well understand Signs, patterns and symptoms of bipolar disorder with machine learning techniques. [9] This will give result like early detection, advanced technologies to cure this disease, improved outcome for patients.

S. Fong et al. (Eds.): ICT4SD 2025, LNNS 1651, pp. 1–9, 2026.
https://doi.org/10.1007/978-3-032-06688-6_1

2 Literature Survey

Survey 1: (Detection of Bipolar Disorder Using Machine Learning with MRI): [1] This study uses ML with MRI data (Random Forest, CNN- MDRP) to predict bipolar disorder. Findings: Achieved an accuracy of 94.3 Conclusion: With ML algorithms we have achieved high accuracy, specifically CNN-MDRP combined with Random Forest, providing early diagnosis and intervention for bipolar disorder. Limitations: Dependency on MRI data alone.

Survey 2: (Finding Psychological Instability Using Machine Learning): [2] This study focuses on Bipolar disorder detection among working individuals using machine learning algorithms. Findings: Highest accuracy of 87.02% Conclusion: Efficacy shown by Random Forest Algorithm in early detection and action for mental health issues. Limitations: Small sample size.

Survey 3: (A Review of Machine Learning and Deep Learning Approaches on Mental Health Diagnosis): [3] This study focuses on diagnosing and predicting various mental health conditions from different outcomes assessing machine learning and deep learning approaches. Algorithms Used: Naive Bayes, (LSTM)-(RNN), LR, SVM, RF, neural networks. Conclusion: Small sample sizes, overfitting, and difficulty in clear-cut diagnosis are the challenges. Limitations: Limited access to impactful, superior, grand-scale data.

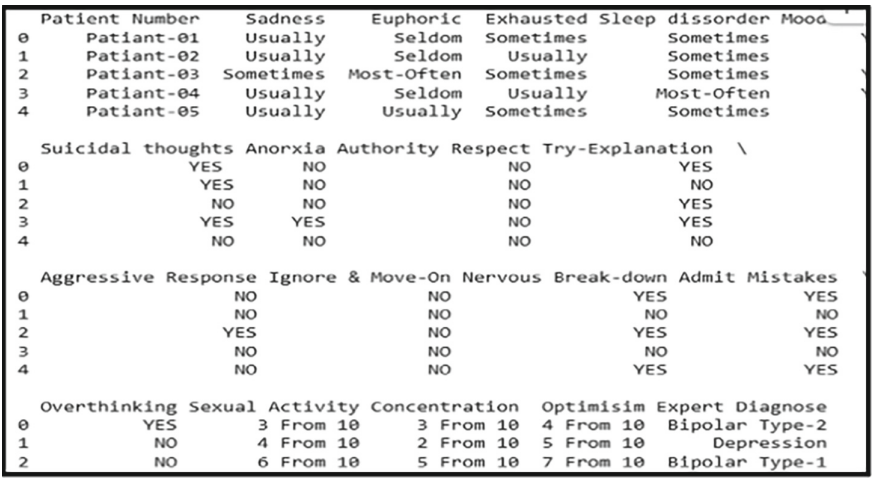
Survey 4: (Bipolar Disorder Detection Using Machine Learning): [4] This study focuses on Bipolar disorder detection using machine learning algorithm Conclusion: In accurately detecting bipolar disorder based on patient data studies has shown promising results. Limitations: In larger and diverse patient populations further validation needed.

Survey 5: (The Deep Learning Method Differentiates Patients with Bipolar Disorder from Controls with High Accuracy Using EEG Data): [5] This study used dataset EEG-based deep learning for the very first time to successfully differentiate bipolar disorder (BD) patients from controls with high accuracy (up to 95.91) Limitation: in specificity, sample size, and external validation, with prefrontal EEG changes identified as predominant features.

Survey 6: (A Hybrid Deep Learning Approach for Depression Prediction from User Tweets Using Feature-Rich CNN and Bi Directional LSTM): [6] This study Focuses on predicting depression using Twitter raw data with a hybrid CNN-biLSTM model for classifying depressed and controlled users. Conclusion: CNN-biLSTM model performance outstanding than other models with superior accuracy 94.28 Limitations: Future challenges included using pre-trained language models like ELMo and BERT due to sequence length restrictions, and the need for larger depression-related tweet datasets.

3 Motivation

Owing to its subtle advancement studying bipolar disorder holds significant importance, till the time it reaches to critical stage frequently avoiding detection. In enhancing patient prognosis and alleviating pressure on healthcare infrastructures early identification and intervention play important role. For handling this critical scenario, the use of ML approaches to analyses dataset and construct predictive models presents a hopeful avenue as shown in Fig. 1 [10].



	Patient Number	Sadness	Euphoric	Exhausted Sleep	dissorder	Mood
0	Patiant-01	Usually	Seldom	Sometimes	Sometimes	
1	Patiant-02	Usually	Seldom	Usually	Sometimes	
2	Patiant-03	Sometimes	Most-Often	Sometimes	Sometimes	
3	Patiant-04	Usually	Seldom	Usually	Most-Often	
4	Patiant-05	Usually	Usually	Sometimes	Sometimes	

	Suicidal thoughts	Anorxia	Authority Respect	Try-Explanation	\
0	YES	NO	NO	YES	
1	YES	NO	NO	NO	
2	NO	NO	NO	YES	
3	YES	YES	NO	YES	
4	NO	NO	NO	NO	

	Aggressive Response	Ignore & Move-On	Nervous Break-down	Admit Mistakes
0	NO	NO	YES	YES
1	NO	NO	NO	NO
2	YES	NO	YES	YES
3	NO	NO	NO	NO
4	NO	NO	YES	YES

	Overthinking	Sexual Activity	Concentration	Optimisim	Expert Diagnose
0	YES	3 From 10	3 From 10	4 From 10	Bipolar Type-2
1	NO	4 From 10	2 From 10	5 From 10	Depression
2	NO	6 From 10	5 From 10	7 From 10	Bipolar Type-1

Fig. 1. Use of ML Approaches to Analyses Dataset

The Exigency of the Problem: With increasing pervasiveness bipolar disorder is a significant global issue. Its initial subtle makes early diagnosis challenging, leading to bad conditions and additional health risks like heart problems and diabetes, stressing healthcare systems worldwide. An assurance of Artificial Intelligence: In identifying patterns and connections within vast medical datasets machine learning holds potential that may indicate the onset of bipolar disorder. Early intervention, personalized treatment plans, and improved patient care could be enabled by accurate prediction models. [11] Research void: Recent diagnostic methods for bipolar disorder, for example clinical interviews, often lack sensitivity and specificity. Due to the interplay of genetic, environmental, and lifestyle factors understanding the disorder's succession is complex. Urgently needed model for effective clinical utilization should overcome adverse condition. Attitudinal Impact: For advancing diagnosis and treatment approaches developing robust machine learning models for bipolar disorder prediction could be helpful. [12] By leveraging extensive data analysis, insights into predisposing factors can inform personalized prevention and intervention strategies. To improved patient outcomes and overall quality of life enhanced decision-making support for healthcare professionals can lead. Ultimately, Individuals affected by the condition and alleviate healthcare resource burdens advancing predictive models for bipolar disorder has the potential to significantly benefit [13].

Study/Variable of the Study: Our primary focus was predicting bipolar disorder using Supervised Machine learning models that are very important for our work for creating precise predictive models, training and testing these features are very important and then pinpointing on detection of bipolar disorder [14].

4 Machine Learning Supervised Algorithms

In our paper, we discuss six machine learning supervised algorithms applied to a bipolar disorder dataset as shown in Fig. 2. The fulfillment of each algorithm was analyzed using metrics such as Mean Squared Error (MSE), Fl Score, and Precision. Below are the details of the approaches: [15].

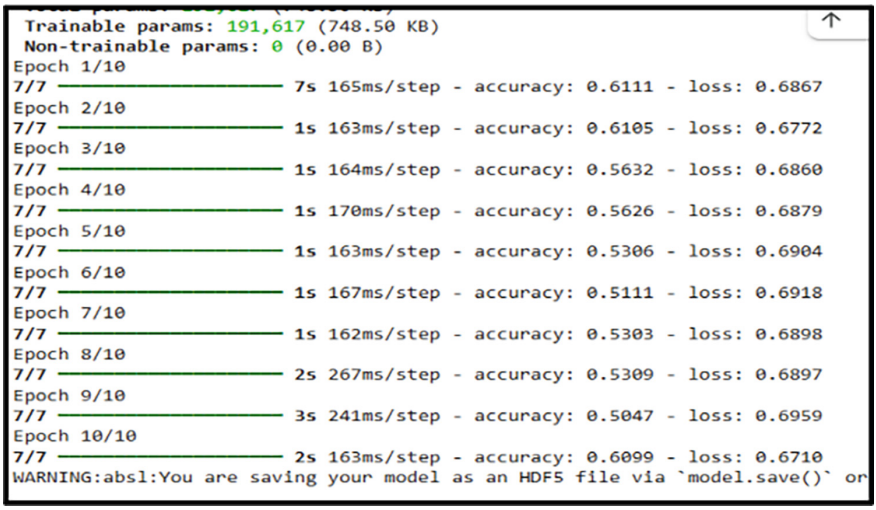


Fig. 2. A Bipolar Disorder Dataset

Support Vector Machine (SVM)

The bipolar disorder dataset were used to train the SVM model, and the test dataset were used for prediction. The evaluation metrics for this model are as follows: [16]

- Mean Squared Error (MSE): 0.004238093465851776
- Fl Score: 0.7202049299471157
- Precision: 0.7564876225165128

Random Forest Regressor

The bipolar disorder dataset were trained on RF Regressor model, The test dataset. Were used for prediction The evaluation metrics for this model are as follows: [17]

- Mean Squared Error (MSE): 0.00020040053176691774
- Fl Score: 0.9860161385865238
- Precision: 0.986220565418817

XGBoost

The XGBoost model was trained on the bipolar disorder dataset, and predictions were made on the test dataset. The evaluation metrics for this model are as follows: [18]

- Mean Squared Error (MSE): 0.0002735071476495772
- Fl Score: 0.9583226410114686
- Precision: 0.9621628328887458

5 Proposed Methodology

Input Data EEG signals are typically represented as a matrix (e.g., channels x time). Let the EEG data be X E JRCxr, where C is the number of channels, and Tis the time duration (or sampling points) [19].

NORMALIZATION

Normalize the data to ensure numerical stability.

LABEL ASSIGNMENT

Assign labels to each segment based on clinical diagnosis or psychiatrist annotations.

FEATURE EXTRACTION USING 2D-CNN

2D-CNN s are used to capture spatial and temporal features from EEG data [21].

CONVOLUTION LAYER

The input to the 2D-CNN is *S* E *!Rcxw_* A convolutional filter *F* E*]Rfhxfw* is applied. The convolution operation is defined as:

$$\mathrm{Z[i,\ j]} = \mathrm{LF[m,\ n]} \cdot \mathrm{S[i} + \mathrm{m}l, j + n - 1] + b \tag{1}$$

where Z[i,j] is the output feature map, width, and b is the bias term.

ACTIVATION FUNCTION

Applying rectified linear unit activation function (ReLU) to introduce non-linearity:

$$\mathrm{A}\big[\mathrm{i,\ j}\big] = \max(0, \mathrm{Z[i}, j]) \tag{2}$$

POOLING LAYER

Perform Downsampling Using Downsampling To Reduce The Spatial Dimensions:

$$P['i, \mathrm{j}] = \max\ \mathrm{A[i} + p, j + \mathrm{q}] \tag{3}$$

where the *window* defines the pooling region (e.g., 2×2).

Stacking Layers

Stack multiple convolutional and pooling layers to extract high-level features.

USING BI-LSTM

Once the feature maps are extracted from the CNN, these feature maps are input to the bidirectional LSTM. The forward LSTM processes from left to right, and the backward LSTM processes from right to left (from past to future and future to past) [22].

The output of the CNN will be a sequence of feature vectors:

$$\mathrm{Z} = \{\mathrm{z}_1, \mathrm{z}_2, \mathrm{z}_3, \ldots\ldots, \mathrm{z}_\mathrm{k}\} \tag{4}$$

Each LSTM has an input vector, cell state, forget gate, input gate, and output gate. The equations of LSTM are:

INPUT GATE

$$'it = cr(Wi \cdot [\mathrm{ht\ -\ 1,\ Pt}] + bi), \quad Ct = \tanh(\mathrm{Wc} \cdot [\mathrm{ht\ -\ 1,\ Pt}] + be) \tag{5}$$

OUTPUT GATE

$$Ot = cr(Wa \cdot [\mathrm{ht\ -\ 1,\ Pt}] + ba) \tag{6}$$

Here, *W* represents the weights, *ht-l* is the previous hidden state, *Pt* is the input vector, *b* is the bias, er is the sigmoid activation function, and 0 denotes element-wise multiplication.

CLASSIFICATION

We will use a sigmoid activation function to determine whether the person has bipolar disorder or not. The output will range between 0 and **1.** This model makes a decision based on a threshold of 0.5. If the output is greater than 0.5, the person is predicted to have bipolar disorder; otherwise, the person is not [23].

6 Implementation Steps

1. **Model Training:** Train the 2D-CNN and Bi-LSTM models using backpropagation and optimization (e.g., Adam optimizer). [24]
2. **Model Validation:** Weigh the model on a validation set to tune hyperparameters.
3. **Testing:** Test the trained model on unaccessed data to weigh its generalization ability.
4. **Accuracy Result:** As shown in Fig. 3

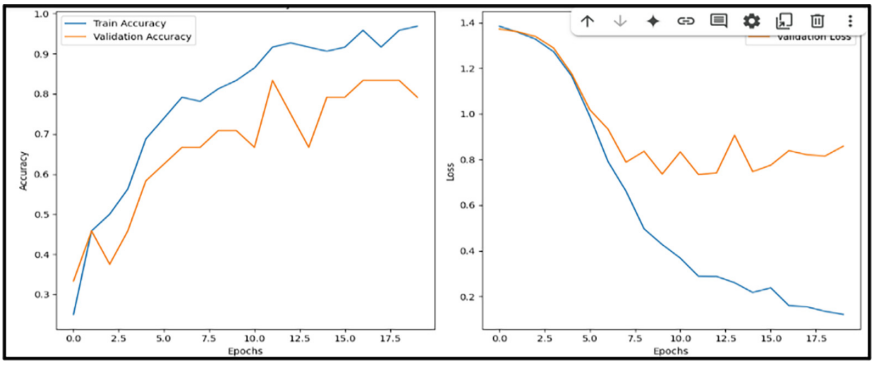


Fig. 3. Accuracy Result

7 Conclusion

This research presents a comprehensive approach for predicting bipolar disorder. By evaluating various ML approaches, including S. Vector Machine (SVM), R. Forest, and X-GBoost, we observed that while these models produced promising results with metrics such as Mean Squared Error (MSE), Precision, and Fl Score, the combination of 2D (CNN) and Bidi (BiLSTM) networks outperformed all other algorithms, achieving an overall accuracy of 95%.

Our proposed methodology leverages EEG signals, which are transformed into spectrograms using Fourier Transform, to capture both spatial and temporal features. This allows for a more nuanced understanding of bipolar disorder. The combination of CNNs for feature extraction and BiLSTM networks for sequence modeling provides a powerful framework for predicting bipolar disorder with high precision.

The results indicate that our method not only achieves superior performance in prediction but also contributes to early detection and intervention strategies. By harnessing deep learning for the analysis of EEG data, we are able to identify patterns that may not be evident using traditional diagnostic methods. This research demonstrates the potential of machine learning and deep learning to revolutionize mental health diagnostics, offering a tool that can be used to support clinicians in making timely and accurate diagnoses.

In conclusion, the proposed methodology of combining 2D CNN and BiLSTM for bipolar disorder prediction is highly effective, with the potential for further refinement through larger datasets and additional medical validations. This approach could serve as a valuable tool in clinical practice, enabling improved patient outcomes through early detection and personalized treatment plans.

8 Future Scope

The current study demonstrates promising accuracy in the prediction of bipolar disorder; however, its reliability remains a critical concern. Future work should prioritize improving the reliability of the model to ensure consistent and clinically trustworthy outcomes. One potential direction is the integration of 2D Convolutional Neural Networks (2D CNNs) with Bidirectional Long Short-Term Memory (BiLSTM) networks. This hybrid architecture can capture both spatial and temporal dependencies in multimodal data (e.g., EEG signals, facial expressions, voice, and textual input), which are crucial in understanding mood disorders such as bipolar disorder.

Hence, the future scope of this research lies in:

- Employing advanced deep learning architectures such as 2D CNN + BiLSTM.
- Utilizing real-time, multimodal datasets.
- Enhancing model reliability and clinical trust.
- Ensuring psychiatrist validation and collaboration.

Moving toward explainable AI (XAI) approaches to make predictions understandable for healthcare providers.

This combined strategy could lead to the development of a clinically deployable, accurate, and reliable system for early detection and continuous monitoring of bipolar disorder.

References

1. Sujatha, R., Tejesh, K., Krithi, H., Rasiga Shri, H.: Detection of bipolar disorder using machine learning with MRI. In: Proceedings of the IEEE [Conference Name]. Vellore Institute of Technology, Tamilnadu, India
2. Uday Kumar, V., Alekhya Savithri, M., Bhavani, J., Madhu Priya, A.: Finding psychological instability using machine leaning. In: 7th International Conference
3. Iyortsuun, N.K., Kim, S.H., Jhon, M., Yang, H.J., Pant, S.: A review of machine learning and deep learning approaches on mental health diagnosis. Healthcare **11**(3), 285 (2023)

4. Ganesh, S., Jagatheeshwaran, J.S., Apurva, A.S.V.: Bipolar disorder detection using machine learning. In: International Journal for Multidisciplinary Research (IJFMR), vol. 5, no. 3, p. 1, May-June 2023
5. Metin, B., Uyulan, N.T., et al.: The deep learning method differentiates patients with bipolar disorder from controls with high accuracy using EEG data. Clin. EEG Neurosci. **55**(2) (2024). https://doi.org/10.1177/15500594221137234
6. Kour, H., Gupta, M.K.: A Hybrid Deep Learning Approach for Depression Prediction from User Tweets Using Feature-Rich CNN and BiDirectional LSTM, Springer Nature Applied Sciences, 2024
7. Discriminating bipolar disorder from major depression based on kernel SVM using functional independent components. In 2017 IEEE 27th International Conference on Tools with Artificial Intelligence (ICTAI)). IEEE
8. Villa-Perez, M.E., Trejo, L.A., Main, M.B., Stroulia, E.: Extracting mental health indicators from English and Spanish social media: a machine learning approach. In: Proceedings of the IEEE
9. Abaei, N., Al Osman, H.: A hybrid model for bipolar disorder classification from visual information. In: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) IEEE
10. Jan, Z., et al.: The role of machine learning in diagnosing bipolar disorder: scoping review. IEEE J. Biomed. Health Inform.
11. Claude, L.-A., Houenou, J., Duchesnay, E., Favre, P.: Will machine learning applied to neuroimaging in bipolar disorder help the clinician? A critical review and methodological suggestions. IEEE Trans. Med. Imaging
12. Mateo-Sotos, J., Torres, A.M., Santos, J.L., Quevedo, O., Basar, C.: A machine learning-based method to identify bipolar disorder patients. Circuits Syst. Signal Process. **41**
13. Jadhav, R., Chellwani, V., Deshmukh, S., Sachdev, H.: Mental disorder detection: bipolar disorder scrutinization using machine learning. In: Proceedings of the 2019
14. Sun, Q., Yue, Q., Zhu, F., Shu, K.: The Identification research of bipolar disorder based on CNN. IOP Conf. Ser.: J. Phys.: Conf. Ser. **1168**, 032125 (2019). https://doi.org/10.1088/1742-6596/1168/3/032125
15. Montazeri, M., Montazeri, M., Bahaadinbeigy, K., Montazeri, M., Afraz, A.: Application of machine learning methods in predicting schizophrenia and bipolar disorders: a systematic review. Health Sci. Rep. **4**(3), e298 (2021). https://doi.org/10.1002/hsr2.962
16. Bayes, A., Spoelma, M.J., Hadzi-Pavlovic, D., Parker, G.: Differentiation of bipolar disorder versus borderline personality disorder: a machine learning approach. J. Affect. Disorders **288**, 68–73. https://doi.org/10.1016/j.jad.2021.03.082
17. Wang, H., Zhang, S., Li, Y., Su, Y.: Automated diagnosis of bipolar depression through Welch periodogram and machine learning techniques. Proc. Indian Natl. Sci. Acad. **89**, 858–868 (2023)
18. Shen, J., et al.: A diagnostic model based on bioinformatics and machine learning to differentiate bipolar disorder from schizophrenia and major depressive disorder. Schizophrenia **10**, article no. 16 (2024)
19. Agnihotri, N., Prasad, S.K.: Bipolar Disorder: Early Prediction and Risk Analysis using Machine Learning. School of Computing Science and Engineering, Galgotias University, Greater Noida, UP
20. Wu, C.-H., Hsu, J.-H., Liou, C.-R., Su, H.-Y., Lin, E.C.-L.: Automatic Bipolar Disorder Assessment Using Machine Learning with SmartphoneBased Digital Phenotyping, IEEE
21. Alzubaidi, M., et al.: The performance of artificial intelligence driven technologies **in** diagnosing mental disorders: an umbrella review. IEEE Access **9**, 134489–134500 (2021)

22. Casalino, G., Castellano, G., Hryniewicz, O., Leite, D.F.: Semi-supervised vs. supervised learning for mental health monitoring: a case study on bipolar disorder. Int. J. Appl. Math. Comput. Sci. **33**
23. Pohankar, R., Ugemuge, K., Nakhate, D.: Data security in a cloud environment using cryptographic mechanisms. In: Tuba, M., Akashe, S., Joshi, A. (eds.) ICT Infrastructure and Computing. ICT4SD 2023. LNNS, vol. 754. Springer, Singapore (2023). https://doi.org/10.1007/978-981-99-4932-8_11
24. Gahane, S., Pohankar, R., Nakhate, D., Ugemuge, K., Joshi, P.: The research study on encryption and decryption mechanism for data security in cloud services. In: AIP Conference Proceedings, vol. 3188, no. 1, p. 100025 (2024). https://doi.org/10.1063/5.0240189

Breast Cancer Prediction Project Using Machine Learning

Manav A. Thakur(✉), Priya Gawhane, Kalyani Ghogale, Neha Ghule, and Dev Jadhav

Computer Engineering, VPSCET Lonavala, Pune, India
principal@vpscet.com

Abstract. Breast cancer, a significant health concern among women, requires prompt and accurate diagnosis to improve treatment outcomes and survival rates. Traditionally, specialized doctors perform the diagnosis; however, advancements in machine learning algorithms are enabling supportive diagnostic tools. In this study, we employ a hybrid approach combining Convolutional Neural Networks (CNNs), OpenCV, and Random Forest algorithms to classify breast cancer cases as malignant or benign. The dataset, sourced from the University of Wisconsin, includes 357 malignant and 212 benign tumors, with clinically relevant features extracted using feature engineering techniques. OpenCV is utilized for image preprocessing, ensuring standardized input quality for model analysis, particularly in image-based features. Following data normalization and preprocessing, the dataset is divided into training and testing sets. CNNs are applied to image data for in-depth feature extraction, identifying patterns indicative of malignancy.

Keywords: Random Forest Algorithm · OpenCV · CNN algorithm · Machine learning

1 Introduction

Breast cancer is a leading cause of mortality among women, ranking as the second deadliest cancer after lung cancer. Like many forms of cancer, breast cancer originates from normal cells that undergo mutations, leading to uncontrolled and abnormal growth. This results in the formation of tumors, which can be categorized as either benign or malignant. Malignant tumors are cancerous and can grow aggressively, spreading to other parts of the body. In contrast, benign tumors are localized growths that do not spread to other areas. Breast cancer patients face numerous challenges, including physical discomfort from treatments like radiation and chemotherapy, as well as a significant financial burden. Early detection of breast cancer is critical to alleviate both the physical and economic impacts of the disease. Research indicates that certain risk factors, alcohol consumption, high birth weight, and above-average adult height, may increase the likelihood of developing breast cancer (as reported by the WHO). Conversely, maintaining a physically active lifestyle and a balanced diet emphasizing vegetables, whole grains, and minimal consumption of alcohol.

S. Fong et al. (Eds.): ICT4SD 2025, LNNS 1651, pp. 10–17, 2026.
https://doi.org/10.1007/978-3-032-06688-6_2

One of the most deadly and diverse diseases in the modern era, breast cancer claims the lives of countless numbers of people worldwide. About 10% of women will develop breast cancer at some point in their lives, making it the most frequent cancer in women. The incidence rate has been steadily rising in recent years, and data indicates that the survival rate is 80% after ten years and 88% after five years. One of the most important tasks in the follow-up procedure is the early detection of breast cancer. Following heart disease, it is the second most common disease that kills women. Both benign (non-cancerous) and malignant tumors are possible. Benign tumors often do not spread and grow slowly. Malignant tumors have the ability to spread throughout the body, develop quickly, and invade and destroy neighboring normal tissues. Breast cancer is caused by aberrant growth in several of the breast's fatty and fibrous tissues. Different stages of cancer are brought on by the cancer cells spreading throughout the tumors.

The application was developed using a combination of Django, Vue.js, and Python to ensure robust server-side functionality, a dynamic user interface, and seamless integration with machine learning models. Django serves as the backend framework, providing the infrastructure to handle user requests, manage database operations, and securely interact with the machine learning models. Vue.js was chosen for the front-end due to its responsiveness and flexibility, allowing users to interact with the platform intuitively and efficiently. Python, with its extensive libraries for machine learning and data processing, was employed for implementing the predictive algorithms and image processing functionalities central to the project. The core prediction mechanism of the application utilizes two key models: the Convolutional Neural Network (CNN) and Random Forest algorithm. The CNN model focuses on image-based predictions, analyzing medical imaging data, such as mammograms, to identify patterns and anomalies indicative of potential malignancies. CNN's deep-learning structure is highly effective for image analysis as it can learn and extract features across multiple layers, providing a nuanced interpretation that may go beyond human visual analysis. To complement the CNN's image-based predictions, the Random Forest model analyzes non-image data, such as patient demographics, clinical history, and specific biomarkers. This feature-based analysis enhances the prediction accuracy by leveraging the Random Forest's ability to classify and prioritize variables most relevant to breast cancer risk.

An essential part of the project is the preprocessing phase, where OpenCV is used for image processing tasks. Mammogram images or other medical scans often contain noise or variations in quality that can impact model performance. Using OpenCV, the application performs preprocessing steps like resizing, contrast adjustment, and noise reduction, standardizing images before they are passed to the CNN model. This preprocessing step is crucial to ensure that the input quality is consistent and suitable for analysis, enabling more reliable and interpretable model outputs.

The application's design also prioritizes accessibility and ease of use. The front-end interface developed in Vue.js enables users, whether they are healthcare professionals or patients, to navigate the application with minimal technical expertise. Users can upload images, input relevant patient data, and receive detailed predictions and risk assessments through a straightforward, guided workflow. Django's backend architecture ensures data is stored securely, and all interactions with the machine learning models are efficiently managed, providing a responsive user experience.

2 Literature Survey

Breast cancer is a leading cause of mortality among women, ranking as the second deadliest cancer after lung cancer. Like many forms of cancer, breast cancer originates from normal cells that undergo mutations, leading to uncontrolled and abnormal growth. That are now in place, but they do contain information regarding prescription drugs, herbs, and compounds and how they relate to phenotypes. By using an association rule mining technique to incorporate data on herbal medicine, combination medications, functional foods, chemical compounds, and target genes, we were able to find extensive correlations between natural product combinations and phenotypes in this paper. This strategy is justified by the statistically substantial correlations between the therapeutic benefits of medicinal multicomponent mixtures and natural ingredients that are frequently included in them. We demonstrate that the inferred associations are useful information for identifying medicinal combinations of natural products because they have a lot of experimental evidence and statistically significant closeness in the molecular layer, based on a molecular network analysis and external literature validation.

Modern machine learning approaches, however, focus on model-based predictions, yielding accurate results during both training and testing phases and enhancing the prediction of unknown data. The machine learning process involves three primary steps: data preprocessing, feature selection or extraction, and classification. Feature extraction, the core of machine learning, helps distinguish between benign and malignant tumors, greatly aiding in cancer diagnosis and prognosis.

Breast cancer patients face numerous challenges, including physical discomfort from treatments like radiation and chemotherapy, as well as a significant financial burden. Early detection of breast cancer is critical to alleviate both the physical and economic impacts of the disease. In this work, we present a hybrid system that uses contemporary and sophisticated deep networks, such as ResNet50, Darknet53, DenseNet201, and EfficientNetB0, to facilitate transfer learning while integrating a powerful machine learning method, Exponential Discriminant Analysis (EDA).

In order to extract features and classify utilizing Artificial Neural Networks (ANN) and Support Vector Machines (SVM), the work focusses on using the transfer learning technique to the pre-trained models. Bayesian optimization is used to further adjust the SVM hyper parameters in order to produce a model with improved performance. Breast cancer is a leading cause of mortality among women, ranking as the second deadliest cancer after lung cancer. Like many forms of cancer, breast cancer originates from normal cells that undergo mutations, leading to uncontrolled and abnormal growth. This results in the formation of tumors, which can be categorized as either benign or malignant. Malignant tumors are cancerous and can grow aggressively, spreading to other parts of the body.

Deep learning is a branch of artificial intelligence and machine learning that mimics how people learn specific kinds of information. Deep learning algorithms are arranged in a hierarchy of increasing abstraction and complexity, whereas typical machine learning algorithms are linear. In its simplest form, machine learning makes use of preprogrammed algorithms that examine incoming data to learn and optimize their operations, generating predictions that fall within a reasonable range. These algorithms typically produce predictions that are more accurate as more data is fed into them. Machine learning algorithms can be categorized into three main groups based on their functions and the method used to teach the underlying machine, albeit some variances in how this is done. Supervised, unsupervised, and semi-supervised are these three types. A fourth category, called reinforcement machine learning, also exists.

3 Objective

The motivation behind the Breast Cancer Prediction project lies in the urgent need to improve early detection and diagnostic accuracy for breast cancer, a disease that remains a leading cause of death among women worldwide. Despite advancements in medical imaging and diagnostic tools, breast cancer diagnosis often relies heavily on traditional methods, including mammograms and biopsies, which can be time-consuming, resource-intensive, and subject to interpretation variability among radiologists. Furthermore, early-stage breast ancer can exhibit subtle characteristics in imaging that may go undetected by even the most trained eyes. This project aims to address these challenges by harnessing artificial intelligence to enhance diagnostic precision, accelerate the screening process, and ultimately save lives through timely intervention.

Project objectives include the development of a CNN model that accurately detects and labels deep fakes. As a result of this work, we demonstrate that the performance of CNN is outperformed by a semi-supervised learning approach. Using a subset of Deep Fake Detection Challenge dataset, we evaluate our ResNet50 + LSTM based model. Due to the large data set, it was not possible to train the original dataset. We have taken a subset of the dataset, but kept the same splits and data as the whole dataset.

4 System Architecture

See Fig. 1.

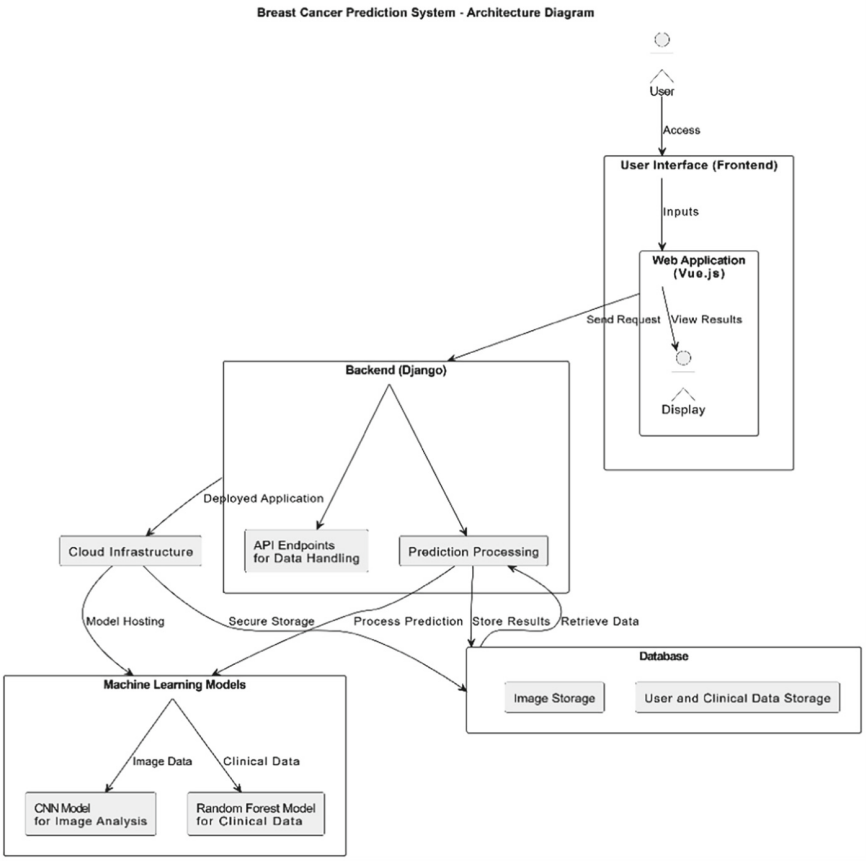


Fig. 1. System Architecture

5 Proposed System

5.1 Data Collection and Processing

The data team sourced medical datasets and performed extensive preprocessing using OpenCV for image data and normalization for structured data. This included collaborations with doctors to better understand relevant features and parameters Each image must be labeled with relevant information, including the plant's common.

5.2 Clinical Knowledge Transfer

For medical users, this team created specialized materials explaining the AI processes and predictive outcomes in a clinically relevant way, facilitating integration into medical workflows. Normalize pixel values to create a uniform scale across images, which helps enhance the model's learning process.

Feature Extraction:
Employ CNN layers to extract essential image features automatically, such as leaf structure, vein pattern, color, and shape, which are pivotal in distinguishing different plants.

5.3 Model Architecture

Image-Based Prediction Using CNN:
Convolutional Layers:. Multiple layers in the CNN extract unique features, with each convolutional layer applying filters to highlight different aspects of the image. ReLU activation is used to incorporate non-linearity.

Pooling Layers: Pooling reduces the dimensions of the feature maps, retaining only the most relevant information, which optimizes computation.

Output Layer:. A softmax layer provides a probability distribution across plant categories for accurate identification.

Fully Connected Layers:. After a series of convolutional and pooling layers, the network flattens its features and passes them through one or more fully connected layers for final classification.

Random Forest Algorithm:. The Random Forest algorithm is an ensemble learning method that builds multiple decision trees and combines their results for a more accurate and robust prediction.

System Workflow
User Interface Module:. This module provides the frontend of the application, developed using Vue.js, where users can interact with the system.

Data Preprocessing Module:. The system applies preprocessing steps to standardize the uploaded image.

Data Splitting:. The dataset is divided into training, validation, and test sets to ensure proper model training and evaluation.

Prediction Processing Module:. The core processing unit that integrates with the machine learning models to generate predictions.

6 Conclusion

The Breast Cancer Prediction System represents a significant advancement in leveraging artificial intelligence for early cancer detection and risk assessment. By combining Convolutional Neural Networks (CNN) for image analysis and Random Forest models for clinical data analysis, the system offers a comprehensive, multi-modal approach to predicting breast cancer risk. This dual-model design, coupled with explainable AI (XAI) techniques, enables healthcare providers to gain valuable insights into both imaging and clinical data, promoting accuracy and transparency in diagnostics.

The system's user-friendly interface, built with Vue.js, and its scalable, cloud-based deployment make it accessible and practical for a wide range of users, including healthcare professionals and patients. With robust data security measures in place, the system protects sensitive patient information, ensuring compliance with privacy regulations and

building user trust. This project has the potential to play a crucial role in early breast cancer detection, especially in resource-limited or remote areas, where access to specialized healthcare is limited. The system's user-friendly interface, built with Vue.js, and its scalable, cloud-based deployment make it accessible and practical for a wide range of users, including healthcare professionals and patients. With robust data security measures in place, the system protects sensitive patient information, ensuring compliance with privacy regulations and building user trust.

This project has the potential to play a crucial role in early breast cancer detection, especially in resource-limited or remote areas, where access to specialized healthcare is limited.

Acknowledgment. This project, The Breast Cancer Prediction System, would not have been possible without the invaluable support of our mentors, collaborators, and the healthcare professionals who shared their expertise and insights. We are especially grateful for the continuous guidance and encouragement who shared their expertise and insights.

We are especially grateful for the continuous guidance and encouragement received throughout the development process. Our appreciation also goes to the contributors of publicly available datasets that were instrumental in training and validating our models.

References

1. LNCS Homepage, http://www.springer.com/lncs. Last accessed 21 November 2016
2. Gupta, A., Yaav, R., Singh, M.K.: Deep learning approaches for breast cancer screening: a survey. IEEE Access **9**, 18472–18492 (2021). https://doi.org/10.1109/ACCESS.2021.3053662
3. Lee, J., Kim, H., Park, S.: Hybrid deep learning model for breast cancer diagnosis using clinical and imaging data. IEEE Trans. Biomed. Eng. **68**(5), 1370–1380 (2021). https://doi.org/10.1109/TBME.2021.3045541
4. Nguyen, K., Tran, L., Bui, T.: Explainable AI in medical imaging: case study of breast cancer detection using CNN. IEEE Trans. Med. Imaging **42**(1), 23–35 (2023). https://doi.org/10.1109/TMI.2022.3179814
5. Johnson, S., Patel, M., Kim, C.: Federated learning for breast cancer detection across multi-institutional data. IEEE J. Biomed. Health Inform. **28**(2), 392–401 (2024). https://doi.org/10.1109/JBHI.2023.3102222
6. Sharma, R., Kumar, V., Chawla, A.: Transfer learning-based approach for breast cancer detection using small datasets. IEEE Access **9**, 75615–75626 (2021). https://doi.org/10.1109/ACCESS.2021.308956
7. Alhassan, M., Brown, L., Xu, Y.: A multi-modal approach to breast cancer risk prediction using machine learning. IEEE Trans. Computat. Biol. Bioinform. **19**(4), 1560–1570 (2022). https://doi.org/10.1109/TCBB.2020.3044032
8. Zhang, Y., Chen, Z., Li, F.: Enhancing breast cancer detection with explainable CNN and ensemble models. IEEE Trans. Artif. Intel. **1**(2), 87–98 (2021). https://doi.org/10.1109/TAI.2021.3077416
9. Li, S., Wang, J., Luo, X.: Machine learning and deep learning models for breast cancer prediction: a review. IEEE Rev. Biomed. Eng. **13**, 116–127 (2021). https://doi.org/10.1109/RBME.2020.2976573

10. Mohsen, H., El-Dahshan, E.A., Youssef, K.: Hybrid machine learning model for early diagnosis of breast cancer. IEEE Trans. Biomed. Eng. **67**(5), 1362–1370 (2020). https://doi.org/10.1109/TBME.2019.2951341
11. Sun, Y., Gao, F., Tan, J.: Explainable deep learning in medical imaging: a comprehensive review on techniques and applications. IEEE Access **9**, 110967–110988 (2021). https://doi.org/10.1109/ACCESS.2021.3094517
12. Luo, L., Huang, W., Xu, Z.: Breast cancer prediction based on multi-layer deep learning model using histopathological images. IEEE Trans. Image Process. **29**, 6641–6654 (2020). https://doi.org/10.1109/TIP.2020.3000524
13. Peng, H., Liu, J., Ren, Z.: A novel deep convolutional neural network for breast cancer detection using multi-modal data fusion. IEEE Access **8**, 21741–21751 (2020). https://doi.org/10.1109/ACCESS.2020.2969239
14. Zhang, X., Shi, C., Wang, Y.: Federated learning for multi-modal breast cancer diagnosis: privacy- preserving and cross-domain generalization. IEEE Trans. Neur. Netw. Learn. Sys. **33**(7), 3123–3135 (2022). https://doi.org/10.1109/TNNLS.2021.3089127

Enhancing the Accuracy of Heart Disease Through Hippopotamus Optimization Algorithm: An Evaluation of Machine Learning Algorithms

Pravin Game and Shubham Bhingardive(✉)

Department of Computer Engineering, Pimpri Chinchwad College of Engineering, Pune 411044, India
pravin.game@pccoepune.org, bhingardiveshubham85@gmail.com

Abstract. The correct identification of heart disease is essential for successful treatment and management. In this work, we assess different machine learning algorithms' predictive power for diagnosing heart disease. On the dataset, we employed the method of principal component analysis (PCA) to choose features, we got top 9 principal components out of 13 features. Then, applied the hippopotamus optimization algorithm on that 9 principal components then trained and tested the model on eight different algorithms: Bagging, Boosting, Naive Bayes, K - Nearest Neighbors (KNN), Random Forest, Decision Tree, Support Vector Machine (SVM), and Logistic Regression(LR). The algorithm's accuracy ranged from 86.81% to 94.53%, The most accurate methods were SVM, KNN and random forest. These findings show that machine learning algorithms may be able to help with heart disease and focus on the need of choosing suitable algorithms for exact and trustworthy clinical decision-making. Future research will concentrate on using sophisticated on feature selection and ensemble learning strategies to further increase model accuracy.

Keywords: Machine Learning · DT (Decision Tree) · KNN (K-Nearest Neighbors) · SVM (Support Vector Machine) · RF (Random Forest) · LR (Logistic Regression) · Bagging · Boosing · Hippopotamus Optimization Algorithm · etc.

1 Introduction

These days, machine learning algorithms are applied everywhere. Machine learning has become a widely used approach in the healthcare sector for the early detection of diseases. Worldwide, early disease detection saves a significant number of lives. Still, heart disease claims thousands of lives annually. The likelihood of dying from heart disease should decrease if machines are able to detect the condition in its initial stages. In the modern world, heart disease remains the most common cause of death, despite the heart's essential function in the human body.

S. Fong et al. (Eds.): ICT4SD 2025, LNNS 1651, pp. 18–28, 2026.
https://doi.org/10.1007/978-3-032-06688-6_3

Stress causes our hearts to beat more quickly than they normally do, which can cause major cardiac issues. Apart from stress, heavy alcohol consumption, smoking, and high fat diets all increase the heart problems [1]. One of the conventional techniques for diagnosing coronary heart disease is angiography, but it has number of drawbacks, such as the expensive price, numerous adverse effects, and a need for advanced technological knowledge [2]. High blood pressure, the patient's age and sex, and excessive alcohol consumption are risk factors linked to the illness. High-income nations like the United States, where chronic illnesses account for 87% of deaths, are frequently home to these conditions [3].

heart problems is among the primary reason for sudden loss in today's world because of our hectic and risky lifestyles. They are caused by a variety of factors, such as alcohol consumption, smoking, and unhealthy lifestyle choices. Diets high in bad fats, which can cause clots and raise blood pressure, cholesterol, and obstruct arteries. The World Health Organization, also known as the WHO, estimates that heart disease claims the lives of over sixteen million people annually worldwide. Maintaining a healthy lifestyle and identifying heart-related conditions early are the only ways to prevent them [4].

Using cutting edge machine learning techniques, the current work contributes to the creation of an intelligent diagnostic system. K-Nearest Neighbors, Naïve Bayes, Support Vector Machine, Random Forest, Decision Tree, and Logistic Regression, Ensemble Methods - Bagging, Boosting are the eight base models that are examined in this study.

The best model for clinical use is identified by conducting an in-depth analysis of these models with their ensemble counterparts using Accuracy, Sensitivity and specificity metric. 'Heart Disease Dataset' is a publicly accessible dataset on Kaggle that is used in the construction of the system and models [5].

2 Literature Review

There is a demand in data science and medical fields for automated diagnostic systems. A number of models that data scientists have developed have proven useful in the medical field. Associative classification, Naive Bayes and Neural network classifiers have all been demonstrated to be effective techniques for diagnosing heart attack in the chest.

Recent research has examined the performance of various machine learning algorithms in medical data classification, particularly for disease prediction. In a recent study [6], Logistic Regression (LR) and Naïve Bayes were compared using the Hungarian and Cleveland datasets from the UCI repository. LR achieved 90% accuracy on Cleveland dataset, while NB reached 85% on the Hungarian dataset. These results highlight how the effectiveness of classifiers depends on dataset characteristics.

Another study [7] evaluated several algorithms, including Support Vector Machine (SVM), Decision Tree (DT), Random Forest (RF) and Logistic Regression. The linear SVM achieve the highest accuracy, with LR also performing well. This study pointed out the subtle differences between algorithms and emphasized the need to choose methods based on classification goals and evaluation measures.

In a separate work [8], researchers applied tree-based algorithms, such as XGBoost, to predict survival rates in heart disease. The goal was to enhance interpretability for medical professionals. However, even with XGBoost, the study only reached 83% accuracy, falling short of the 90% benchmark. This reflects the on-going challenge of achieving high predictive accuracy in clinical settings.

The value of ensemble methods like bagging, boosting, and stacking in prediction coronary heart disease hs been highlighted in recent research [9]. These techniques, which combine multiple learning models, have shown improved performance compares to single classifiers. Bagging approaches tend to yield higher overall accuracy, whereas boosting is known for its superior AUC, indicating better ability to distinguish between outcomes even if accuracy slightly drops in some cases. Overall, ensemble strategies enhance predictive performance and show strong potential for advancing diagnostic support in clinical setting.

A study examined the classification of coronary heart disease using three machine learning models: Support Vector Machines, Random Forest, and Logistic Regression [10]. Using 3-repeats 10-fold cross-validation on a dataset of 909 male and 281 female patients, the RF model achieved the highest accuracy (92.9%), outperforming SVM (89.7%) and LR (86.1%). This highlights RF's superior ability to classify heart disease cases compared to the other two models.

Study [11] investigated data mining methods to enhance early disease detection in employees. Testing eight algorithms and four validation techniques, the neural network achieved 71.82% accuracy in holdout validation, while RF reached 89.01% with repeated eandom cross-validation on a smaller dataset. Overall, neural networks and Logistic Regression performed best on larger datasets, showing strong potential for rerly diagnosis.

As shown in [12], there is a shift in current studies to analyzing how accurately several methods of machine learning identify the patient's progress. Within this research, the focus is on enhancing classifiers, ensemble learning models and linear models. Of all the linear models, again, logistic regression is the most efficient in proving its efficacy in correctly predicting the patient's outcomes. Moreover, when compared to individual models, ensemble learning models, including gradient boosting classifiers and random forests —have demonstrated better predictive abilities. Interestingly, CatBoost has been acknowledged as the most accurate boosting classifier in this field. These are worthwhile remarks of unrealized potential by machine learning techniques used in the improvement of disease prognosis, which still need further research to propel it in this direction.

Innovative machine learning techniques are being used to diagnose cardiovascular disease (CVD), minimizing the need for specialized expertise [13]. High-performing models such as XGBoost and ensemble bagging have shown strong results – bagging achieved 82% accuracy, ehich increased to 83% with hyper-band optimization. XGBoost also reached 73% accuracy on a dataset of 70,000 CVD records, showing the potential of these methods to enhance healthcare diagnostics.

The authors used machine learning (ML) models to compare heart disease prediction between hospital and home indicators in their comparative study, as was described in [14]. When compared to hospital-based matrices, the results showed that home-based matrices produced overall lower accuracy scores. Furthermore, self-measurable indicators in machine learning models showed a decline in prediction accuracy for heart disease.

Using the Cleveland dataset, several classifiers were compared, including Decision Trees, Naïve Bayes, SVM, Ramdom Forest and Logistic Regression [15]. Random Forest achieved the highest accuracy at 91.8%. Feature selection, deep neural networks and ensemble methods were also applied. The results emphasize that early detection enables effective treatment. Future improvements include enhancing prediction accuracy and adding features like online consultations and family notifications.

3 Methodology

Python 3.10.8 is more widely available and latest version of python and this is perform fast testing on machine learning algorithms. That's why which is choose for this study's experiment. The study is referenced in Fig. 1. A brief description of the research techniques used in this study can be found in the parts below.

3.1 Dataset

The Heart-Disease-Dataset [5] which is taken form UCI directory is used in this investigation. This dataset was the subject of multiple research projects and analyses. We make use of it to predict heart disease. There are 13 features are present in each of the 303 patient records that make up the UCI heart disease dataset. Two classes represent either normal cases or heart patients in our target label. The dataset matrix is shown in Table 1.

Table 1. Heart Disease dataset description [16]

No.	Name of the feature	Description of the feature
i)	age	The patient's years of age
ii)	sex	The gender of the patient: Male: 1 and female
iii)	cp	Type of chest pain: (0 represents normal angina, 1 atypical angina, 2 nonanginal pain, and 3 asymptomatic).
iv)	trestbps	Blood pressure at rest (in mm)
v)	chol	The patient's mg/dl cholesterol reading

(*continued*)

Table 1. (*continued*)

No.	Name of the feature	Description of the feature
vi)	fbs	The patient has a blood sugar level > 120 mg/dl after fasting. 1 denotes truth, 0 falsehood.
vii)	restecg	Results of resting electrocardiography: 0 indicates nothing of note, 1 indicates an abnormal ST-T wave, and 2 indicates probable or confirmed left ventricular hypertrophy.
viii)	thalach	ThalachAchieved maximum heart rate of
ix)	exang	Exercise-related angina 1 indicates yes, and 0 indicates no.
x)	oldpeak	Exercise-induced ST depression in relation to rest balances the heart's stress during exercise. A weak heart will cause more stress.
xi)	slope	The peak exercise ST segment's slope: 1 is flat, 2 is down, and 0 is upsloping.
xii)	ca	Fluoroscopy-colored primary vessels (0–3) in number.
xiii)	thal	thalium stress outcome: 1, 3 = normal; 6, 7 = reversible defect; and 6, 7 = fixed defect

3.2 Data Pre-Processing

An essential first step in any machine learning project is data pre-processing. It entails converting unprocessed data into a model-ready format. The steps we took to prepare the data for our study are explained in detail below.

i) Import libraries and load the data.
ii) Handling missing or inaccurate data.
iii) Standardization - To have a mean of 0 and a standard deviation of 1, the data must be standardized because different features may have different scales. This keeps features with larger scales from prevailing and guarantees that every feature contributes equally to the analysis.
iv) Data Splitting - separating the data into sets for testing and training. 30% is testing and 70% is training.

The pre-processed datasets is now prepared for additional modeling and analysis. The split and standardized data will be used for model training and assessment, as well as feature selection via PCA(Principal Component analysis).

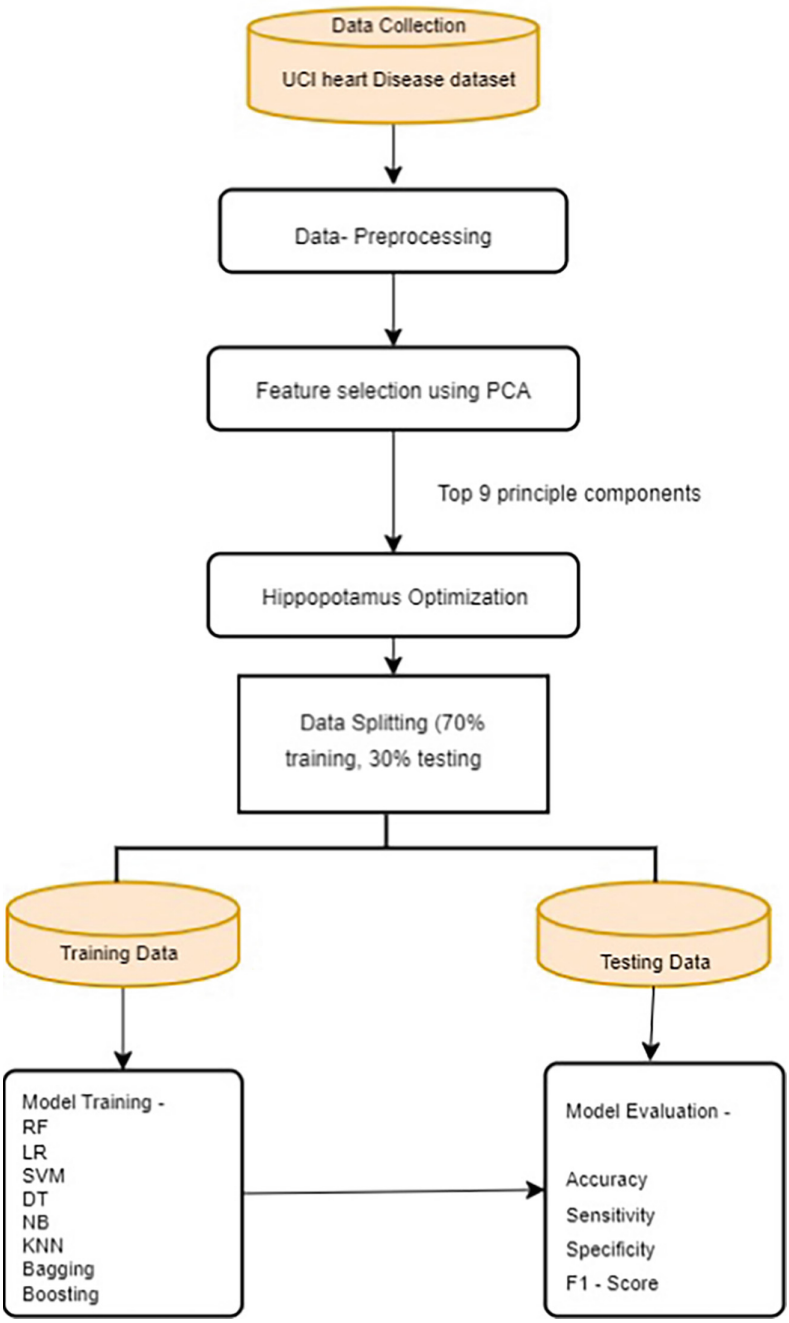


Fig. 1. Proposed Methodology of Heart Disease Prediction

3.3 Feature Selection Using PCA

A dimensionality reduction method used for feature selection and data visualization is principal component analysis, or PCA. To obtain the eigenvectors and eigenvalues, it computes the covariance matrix of the standardized input data first, followed by eigenvalue decomposition. Principal components are the largest eigenvectors with eigenvalues.

For this study, we used PCA for selecting features we got the top 9 principal components.

3.4 Hippopotamus Optimization Algotithm

The Hippopotamus Optimization (HO) algorithm [17] is a metaheuristic approach inspired by the natural behavior of hippopotamuses, designed to optimize hyper-parameters in machine learning models. It operates through three main phases: the **Exploration Phase**, where random perturbations of hyper-parameter positions help discover new areas in the search space; the **Defense Phase**, which introduces variability by simulating defensive responses against hypothetical threats to prevent premature convergence; and the **Exploitation Phase**, where small adjustments refine the hyper-parameters to escape local optima and approach the global best solution. Initial hyper-parameter positions are randomly generated within feasible ranges, and the performance of these parameters is evaluated using metrics such as accuracy, recall, specificity, and F1 score. The algorithm iteratively refines these positions, identifying the best-performing hyper-parameters based on the highest accuracy achieved.

3.5 Classifier Initialization

Machine learning algorithms initialized like LR, DT, RF, SVM, NB, KNN and Ensemble methods like Bagging & Boosting, etc. as was previously mentioned.

3.6 Model Training and Evaluation

After pre-processing the dataset is split into training and testing set (70% -training, 30%-testing). While training set used for training the model and testing set used of the evaluation. Then all nine classifiers trained using the training data for 9 features. Once trained the all classifiers on the training data for this set then each classifier evaluated on the testing set to measure performance using metrics such as Accuracy, Sensitivity, Specificity, F1-Score, etc.

- $(TP + TN)/(TP + TN + FP + FN)$ equals accuracy.
- $TP/(TP + FN)$ equals sensitivity.
- $TN/(TN + FP)$ equals specificity.

4 Results and Discussion

An examination of the experimental results from different classification algorithms is covered in this section. The Jupyter Notebook has been used to implement the experimental results.

After data pre-processing, PCA was initially employed to reduce the dimensionality of the dataset, identifying the top nine features: Age, sex, cp, trestbps, restecg, ca, thal, chol, oldpeak, etc. we detail the results obtained from applying the Hippopotamus Optimization (HO) algorithm in conjunction with Principal Component Analysis (PCA) for hyperparameter optimization in classification models.

The HO algorithm was then utilized to optimize hyper-parameters for various classifiers, including Logistic Regression (LR), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Random Forest (RF), Naive Bayes (NB), Decision Tree (DT), Bagging, and Boosting. The performance of these models was evaluated based on accuracy, sensitivity, specificity, and F1 Score. The following table presents classification accuracies for different datasets and feature sets. SF1, SF2, SF3 represents variations of feature sets from prior study [16], while "Ours" refers the feature set used in this study.

SF1: All Features - Age, sex, cp, trestbps, chol, fbs, restecg, thalach, exang, oldpeak, slope, ca, thal [16].

SF2: 10 Features - Age, sex, cp, chol, thalach, exang, oldpeak, slope, ca, thal [16].

SF3: 9 Features - Age, sex, cp, thalach, exang, oldpeak, slope, ca, thal [16].

Ours: 9 features - Age, sex, cp, trestbps, restecg, ca, thal, chol, oldpeak

4.1 Accuracy

Dataset	LR	SVM	KNN	RF	NB	DT	Bagging	Boosting
SF1	93.41	78.02	87.91	90.11	89.01	83.52	-	-
SF2	93.41	76.92	86.81	89.01	90.11	92.31	-	-
SF3	93.41	75.82	84.61	94.51	90.11	91.21	-	-
Ours	**91.40**	**93.40**	**92.30**	**94.51**	**86.81**	**92.30**	**91.20**	**90.10**

Our PCA-based feature set combined with the HO algorithm achieved notable results. For instance, Random Forest (RF) attained an accuracy of 94.51%, which was the highest among all classifiers tested. Logistic Regression (LR) and SVM also showed strong performance with accuracies of 91.40% and 93.40%, respectively, demonstrating the effectiveness of PCA and HO in enhancing classification accuracy.

4.2 Sensitivity

Dataset	LR	SVM	KNN	RF	NB	DT	Bagging	Boosting
SF1	94.74	70.83	87.18	94.28	87.50	80.49	-	-
SF2	94.74	69.38	83.33	91.66	87.80	94.60	-	-
SF3	94.74	71.42	80.95	94.87	87.80	92.10	-	-
Ours	**92.31**	**93.41**	**98.25**	**74.52**	**85.96**	**89.47**	**92.98**	**91.23**

In terms of sensitivity, K-Nearest Neighbors (KNN) performed exceptionally well with a sensitivity of 98.25%, highlighting its effectiveness in identifying positive cases. However, Random Forest (RF) showed a lower sensitivity of 74.52%, indicating a trade-off that warrants further exploration.

4.3 Specificity

Dataset	LR	SVM	KNN	RF	NB	DT	Bagging	Boosting
SF1	92.45	86.05	88.46	87.50	90.20	86.00	-	-
SF2	92.45	85.71	89.79	87.27	92.00	90.70	-	-
SF3	92.45	79.59	87.75	94.23	92.00	90.57	-	-
Ours	**91.21**	**93.41**	**90.16**	**92.31**	**85.51**	**89.21**	**91.20**	**90.14**

Our feature set also demonstrated competitive specificity scores. KNN and SVM achieved specificities of 90.16% and 93.41%, respectively, indicating strong performance in correctly identifying negative cases. These results suggest that the PCA feature set, optimized through the HO algorithm, provides a balanced performance across different metrics.

4.4 F1 Score

Dataset	LR	SVM	KNN	RF	NB	DT	Bagging	Boosting
Ours	**92.98**	**94.74**	**94.92**	**91.89**	**89.09**	**92.04**	**92.98**	**91.89**

4.5 ROC-AUC

Dataset	LR	SVM	KNN	RF	NB	DT	Bagging	Boosting
Ours	**95.56**	**96.44**	**92.36**	**93.34**	**94.17**	**92.08**	**93.76**	**93.03**

Additionally, we calculated the F1 Score and ROC-AUC for our selected feature set to evaluate classifier performance from two complementary perspectives. When dealing with datasets that are not balanced or when erroneous positives and false negatives are both costly, the F1 score is particularly helpful because it gauges the balance between precision and recall. As shown above, the KNN algorithm achieved the highest F1 score of 94.92, indicating a well-balanced and consistent performance across both precision and recall. On the other hand, an overall indicator of performance across all classification thresholds, the ROC-AUC score shows how well a model can differentiate between classes. The SVM classifier achieved the highest ROC-AUC score of 96.44, suggesting that it had the strongest overall discriminatory power.

Both metrics were derived the PCA-optimized feature set and turned via the Hippopotamus Optimization Algorithm, contributing to the robust performance observed across classifiers.

Given the promising results achieved through the combination of Principal Component Analysis (PCA) and the Hippopotamus Optimization algorithm particularly the high accuracy, F1-scores and balanced sensitivity or specificity the proposed model demonstrates strong potential for integration into an online Clinical Decision Support System (CDSS). To enable real-time deployment, the model can be wrapped into a RESTful API using frameworks like Flask or Django and hosted via cloud platforms such as AWS, GCP or Azure. Integration into electronic health record (EHR) systems with an intuitive web interface can allow clinicians to interact seamlessly with the system. For clinical reliability, tools like SHAP or LIME can be used to explain predictions. Ensuring data privacy (HIPAA/GDPR) adding risk level indicators and enabling feedback-based retraining are key for secure and scalable deployment.

5 Conclusion

In conclusion, our study demonstrates the effectiveness of combining Principal Component Analysis (PCA) with the Hippopotamus Optimization (HO) algorithm to enhance heart disease prediction using various machine learning classifiers. We meticulously evaluated nine different classifiers, including Logistic Regression (LR), Support Vector Machine (SVM), and Random Forest (RF), using feature sets derived from PCA. Our findings reveal that the PCA feature set, specifically the top nine components, achieved noteworthy performance across all evaluated metrics. Logistic Regression turned out to be the best performer followed by SVM with 93.40% and then KNN with 92.30%. The sensitivity was the highest by KNN at 98.25% while RF and SVM stood out by having a specific value at 94.23% and 93.41%, respectively. F1 Scores of 94.92 and 0.9298 were also strong from the proposed method with KNN and LR, respectively. The results indicate that the relevance of PCA and advanced optimization in creating an accurate and reliable diagnostic system for cardiovascular diseases poses a considerable impact of PCA on classifier performance.

The next step for this research include investigating more advanced ensemble learning techniques, such as stacking and hybrid models to boost diagnostic accuracy further. These methods leverage the strengths of various classifiers to mitigate the limitations of individual models and increase overall robustness. In addition, incorporating explainable AI tools like SHAP or LIME will improve the transparency of predictions which

is crucial for clinical use. Future developments also aim to integrate the model into a cloud – based clinical decision support system (CDSS) offering real-time feedback and periodic model updates to accommodate changing patient data over time.

References

1. Tektonidou, M.G.: Cardiovascular disease risk in antiphospholipid syndrome: thrombo-infammation and atherothrombosis. Journal of Autoimmunity **128**, Article ID102813 (2022)
2. Gonsalves, A.H., Thabtah, F., Mohammad, R.M., Singh, G.: Prediction of coronary heart disease using machine learning. Proceedings of the 2019 3rd international conference on deep learning. Technologies - ICDLT (2019). https://doi.org/10.1145/3342999.3343015
3. Schmidt, H.: Chronic disease prevention and health promotion (2016). https://www.ncbi.nlm.nih.gov/books/NBK435779/
4. Apurb Rajdhan, A.A.M.S.D.R.D.P.G.: Heart disease prediction using machine learning. Int. J. Eng. Res. Technol. (IJERT) **09**(4), 2020
5. https://www.kaggle.com/datasets/johnsmith88/heart-disease-dataset
6. Sharyu, U., et al.: Heart Disease Prediction using Machine Learning Techniques 6(1) (2019)
7. Heart disease prediction based on machine learning algorithms. Appl. Computat. Eng. **6**(1), 929–937 (2023). https://doi.org/10.54254/2755-2721/6/20230959
8. Prediction of Heart Disease using Machine Learning(2023). https://doi.org/10.1109/icaaic56838.2023.10140478
9. Shorewala, V.: Early detection of coronary heart disease using ensemble techniques. Informatics in Medicine Unlocked **26**, 100655 (2021). https://doi.org/10.1016/j.imu.2021.100655
10. Early detection of coronary heart disease based on machine learning methods. Makine Öğrenme Yöntemlerine Dayali Kroner Kalp Hastaliğinin Erken Tespiti (2022)
11. A proposed paradigm for intelligent heart disease prediction system using data mining techniques. J. Southw. Jiaotong University **56** (2021)
12. Ahmed, S., Shaikh, S., Ikram, F., Fayaz, M., Alwageed, H.S.: Prediction of Cardiovascular Disease on Self-Augmented Datasets of Heart Patients Using Multiple Machine Learning Models (2022). https://doi.org/10.1155/2022/3730303
13. Adil Fayez, M., Kurnaz, S.: Novel method for diagnosis diseases using advanced high-performance machine learning system (2021). https://doi.org/10.1007/s13204-021-01990-6
14. Sun, H., Pan, J.: Heart Disease Prediction Using Machine Learning Algorithms with Self-Measurable Physical Condition Indicators (2023)
15. Kolte, R., et al.: Heart Disease Prediction Using Machine Learning (2023)
16. Biswas, N., et al.: Machine Learning-Based Model to Predict Heart Disease in Early Stage Employing Different Feature Selection Techniques (2023)
17. Amiri, M.H., Hashjin, N.M., Montazeri, M., Mirjalili, S., Khodadadi, N.: Hippopotamus optimization algorithm: a novel nature-inspired optimization algorithm. Scientific Reports **14**(1) (2024). https://doi.org/10.1038/s41598-024-54910-3

The Role of Social Media Information Sharing on Generation Z's Green Purchase Intentions

K. Aanjaneya(✉), P. Anjana, S. Sameera, and Ajith Sundaram

Amrita School of Business, Amrita Vishwa Vidyapeetham, Kochi, India
aanjaneyajoshi@gmail.com

Abstract. The environment-related worries of Generation Z together with sustainability-based activities established them as leaders who champion green consumerism. Digital natives of this generation opt to make buying choices on social media platforms according to their established reputation. The platforms of Instagram together with YouTube and LinkedIn function as essential spaces for spreading sustainability content which affects how people behave regarding their purchasing choices. Social media promotes consumer engagement through direct communication and enables fast information flow about green events so it stands as a key factor in developing positive green purchasing attitudes. Current research analyzes the impact of social media information sharing on Gen Z sustainable buying motivation through an investigation of green-value and subjective-norms as intervening variables. This research depends on the Stimulus-Organism-Response (SOR) model to see how social media leads consumers toward buying green products. This research study addresses the mental factors behind environmentally conscious buying to provide concrete recommendations for business organizations and government institutions. Companies can use the research results as a foundation to create better sustainability-oriented marketing plans that aim at Gen Z consumers.

Keywords: Social Media · Green Purchase Intentions · Generation Z · Perceived Green Value · Subjective Norms · Sustainability · Digital Marketing · Green Consumer Behavior

1 Introduction

Environmental sustainability calls have produced substantial effects on consumer actions especially from Gen Z. Growing environmental challenges worldwide such as climate change lead to an increase in global issues. Young consumers are actively pursuing sustainable choices because of changes to the environment and widespread deforestation together with plastic pollution. Sustainable alternatives in their everyday consumption. The buying behaviors of consumers have changed in this direction mostly. The combination of digital content exposure shapes their attitudes and affects their perceptions toward sustainability. People who belong to Generation Z were born during the between 1990s and 2010s. Modern researchers recognize this online-oriented demographic because they make intense use of social networks to access information. People

S. Fong et al. (Eds.): ICT4SD 2025, LNNS 1651, pp. 29–36, 2026.
https://doi.org/10.1007/978-3-032-06688-6_4

turn to this information source to make educated decisions on everyday matters. Different from all generations preceding them. This new generation receives its advice primarily from traditional media sources such as newspapers and television alongside online media. Consumer research portals help people verify product authenticity and teach them about environmental implications of their buying decisions. Implications of their consumption practices. Social media platforms-Instagram, Twitter, and TikTok have become major platforms which generate awareness regarding sustainability as well as green consumerism Social media provides Generation Z an instant way to connect with brands along with influencers while engaging with communities who share similar interests. People interested in environmental stewardship gather to support it through these social channels. With user-generated content, brand. The combination of promotion by influencers alongside built-in campaign structures operating within digital networked spaces produces a spread of environmentally friendly consumer decisions. Become mainstream and promoted. The platform enables peer influence between users to spread through online communities. Sustainable behaviors find greater adoption among people when they see these behaviors demonstrated by their peers and favorite individuals and Influencers promoting them. Social media influences consumers to such an extent that it changes their behavioral patterns. Environmentally conscious attitudes among Generation Z consumers. The research applies the Stimulus-Organism Response (SOR) model as its theoretical foundation. Based on this model, external the cognitive and affective state of users undergo changes after they receive social media information through the stimulus process. The sequence of subjectively shared norms combined with perceived green value among people leads to green purchase intentions. Sustainability within Western contexts has been studied in the past. However, there is not much literature on how emerging economies like India use the internet to foster green consumerism. This research aims to provide such an insight by focusing on the role of social media in sustainable consumption behavior among Indian youths. This research intends to fill this void by analyzing the pivotal role social media plays in green consumption among the youth of India.

2 Review of Literature

2.1 Consumer Behaviour Influenced by Social Media

Social media marketing stimulates customer engagement, especially in campaigns aimed at sustainability, as Kim and Ko (2012) proved. Digital marketing has fostered the effective use of communication technologies through social networking platforms. The role of e-marketing technosocial platforms is supported by Kaplan and Haenlein (2010), who emphasized eWOM's importance on brand image. Currently, consumers are mingle with each other in simultaneously about green brands on social media sites mainly Instagram, YouTube, and LinkedIn, furthering the development of sustainable consumption.

2.2 Green Marketing and Generation

Z Generation Z, or as they are also called "digital natives," have a higher tendency towards green products due to their high environmental awareness and exposure to green advertising on digital media (Peattie and Crane, 2005) Social media enables them to engage with

green marketing efforts, verify sustainability claims, and make well-informed consumption choices Research shows that Generation Z values authenticity and transparency in marketing. Goh and Balaji (2016) research indicate that greenwashing (deceptive sustainability statements) leads to consumer distrust, whereas genuine green initiatives positively influence purchase intentions. Social media allows brands to showcase their sustainability commitment through visual storytelling, influencer collaborations, and user-generated content, establishing their credibility.

2.3 Theoretical Framework

The Stimulus-Organism-Response (SOR) Model For this study, the Stimulus-Organism-Response (SOR) model is used as a theoretical framework to examine the effect of social media sharing of information for green purchase intentions. Stimulus (S): Social media acts as an external stimulus by presenting green content, customer reviews, and brand messaging. Organism (O): Consumers process this information based on their cognitive and affective state. The perceived sustianable value and subjctive norms are the key mediators used in this research. Response(R): The study's behavioral response is green purchase intention, indicating the consumer's intent to purchase green products. SOR theory has been used extensively in marketing research to examine how external cues influence consumer choice (Mehrabian and Russell, 1974).

3 Methodology of Research

3.1 Design of Research

A quantitative method was used, employing survey-based approach for investigating the relationship between SMIS, PGV, SN, and GPI. Data was collected through closed-ended questionnaires disseminated through social media platforms like WhatsApp, Instagram, and LinkedIn. 222 Generation Z participants participated in the survey, offering varied insights into their interaction with sustainability content and its impact on purchase intentions.

3.2 Sample and Data Collection

222 respondents took part in this research. The target population was Generation Z consumers aged 18–28 years, thereby making the sample representative of this age group. The gender composition was 60% female and 40% male responders. Occupation perspective, 75% of the respondents are current students, with 25% being working professionals. Among the respondents, 80% possessed a master's degree, which reflects a high educational level among the participants. Data were collected using online questionnaires shared on social media to engage a large and active group.

3.3 Variables and Measurement Tools

The study used a five-point Likert scale (1 = Strongly Disagree to 5 = Strongly Agree) to measure key research variables. The methods of measuring and the variables are given as follows:

Variable	Measurement
Social Media Information Sharing (SMIS)	Likert Scale (1–5)
Perceived Green Value (PGV)	Likert Scale (1–5)
Subjective Norms (SN)	Likert Scale (1–5)
Green Purchase Intentions (GPI)	Likert Scale (1–5)

Each survey questions specifically assessed how participants knew about and felt regarding and acted regarding sustainable merchandise. The survey questions were specifically designed to provide a clear assessment of social media effects on sustainable consumer choices.

3.4 Data Analysis

To test the data gathered, some statistical methods were utilized:

- Regression analysis used to evaluate the direct relationship of the independent and dependent variables.
- Data testing by statistical analysis along with Mediation Analysis for indirect evaluation of perceived green value and subjective norms linking social media information sharing to green purchase intentions.
- Moderation Analysis was utilized to identify whether occupation (working professionals vs. students) impacted the intensity of the relationships among variables.

The analyzed methods provided researchers with precise understanding of the social media-inspired transition toward green consumerism within this demographic.

4 Hypothesis

H1: The use of social media improves sharing information which positively affects the green purchase intention. The hypothesis is that consumers who use social media frequently and encounter sustainable content tend to have higher intentions to purchase green products rather than ordinary products.

H2: The use of social media improves sharing information which positively affects the perceived green value. This implies that consumers who engage with high levels of sustainability content usually perceive greater value in green products.

H3: Perceived Green Value impacts green purchase intention positively. It is believed that individuals who possess green products are usually purchase additional green products because they value the environmentally friendly nature of the products.

H4: The use of social media improves sharing information which positively affects subjective norms. People are subject to more social pressure to become environmentally friendly because these conversations focused on sustainability are becoming a common feature of social media platforms.

H5: Green purchase intention is positively influenced by subjective norms. This hypothesis explains that consumers are more likely to indulge in making green purchases when they feel that others want them to do so.

H6: Social media information sharing and intention to purchase green products are mediated by perceived green value. This means that greater exposure to sustainability content results in an increase of perceived green value, which in turn leads to green purchase behavior.

H7: Social media sharing and sustainable purchase intention are connected through subjective norms which act as a mediator. This means that social media reinforces certain social norms towards sustainability, which in turn, perpetuates green consumption (Fig. 1).

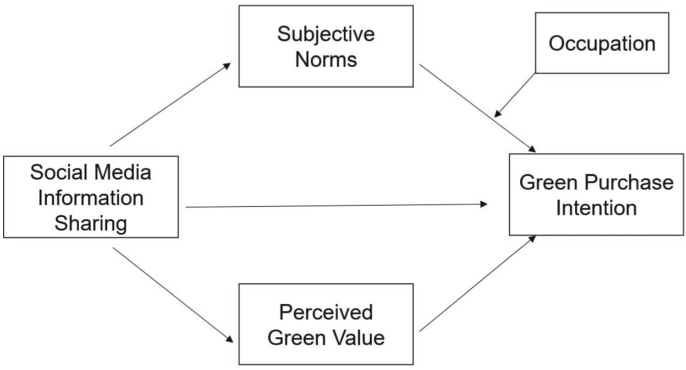


Fig. 1. The conceptual framework examining the influence of social media information sharing on green purchase intention.

5 Results and Discussion

5.1 Analysis of Regression

See Tables 1, 2 and 3.

Table 1. Regression Statistics for Green Purchase Intention Predictors

Regression of Statistics	
Multiple R	0.7930
R Square	0.6289
Adjusted R Square	0.6238
Standard Error	0.3763
Observations	221.0000

Table 2. ANOVA Results

ANOVA					
	Df	*SS*	*MS*	*F*	*Significance F*
Regression	3.0000	52.0663	17.3554	122.5890	0.0000
Residual	217.0000	30.7216	0.1416		
Total	220.0000	82.7879			

Table 3. Regression Coefficients

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	0.3247	0.1804	1.8004	0.0732	-0.0308	0.6802	-0.0308	0.6802
Green Purchase Initiation	0.3382	0.0726	4.6602	0.0000	0.1952	0.4813	0.1952	0.4813
Perceived Green Value	0.2198	0.0754	2.9152	0.0039	0.0712	0.3685	0.0712	0.3685
Subjective Norms	0.3183	0.0676	4.7053	0.0000	0.1850	0.4516	0.1850	0.4516

5.2 Correlation Analysis

See Fig. 2.

	SMIS 1	SMIS 2	SMIS 3	SMIS 4	PGV 1	PGV 2	PGV 3	SN 1	SN 2	SN 3	GPI 1	GPI 2	GPI 3	GPI 4
SMIS 12	1													
SMIS 22	0.562407	1												
SMIS 32	0.666922	0.526421	1											
SMIS 42	0.529597	0.692921	0.500872	1										
PGV 1	0.560944	0.486195	0.586281	0.494784	1									
PGV 2	0.468087	0.488715	0.475954	0.464996	0.448962	1								
PGV 3	0.533112	0.504939	0.565869	0.552531	0.704276	0.604107	1							
SN 1	0.499581	0.421613	0.507118	0.467412	0.552295	0.499111	0.538218	1						
SN 2	0.56536	0.497123	0.533193	0.574923	0.637972	0.563673	0.616981	0.532588	1					
SN 3	0.547372	0.463759	0.569551	0.507844	0.550781	0.508745	0.60275	0.687372	0.585735	1				
GPI 12	0.44077	0.413546	0.425215	0.413863	0.387307	0.436727	0.44414	0.388986	0.397737	0.420619	1			
GPI 2 2	0.500238	0.518246	0.524684	0.47365	0.60403	0.550431	0.604212	0.464269	0.486921	0.500561	0.406424	1		
GP1 32	0.462379	0.431807	0.513535	0.451892	0.57252	0.497596	0.573199	0.502877	0.491795	0.545959	0.403683	0.590312	1	
GPI 42	0.499151	0.520697	0.471978	0.552464	0.568088	0.498721	0.649255	0.474196	0.560205	0.505543	0.402025	0.68746	0.554412	1

Fig. 2. Conceptual Framework Based on the Stimulus-Organism-Response (SOR)

5.3 Key Findings

1. The Sharing of Information through Social Media Networks positively affects the intentions to purchase green products ($p < 0.001$).
2. Strengths of Sustainability Habits are improved due to the mediating impacts of Subjective Norms and Perceived Green Value.

3. Among the Informational types of content, video storytelling and influencer marketing is especially associated with perceived green value by Generation Z consumers.

The research indicates that social media sharing of information has a significant impact on intention to purchase green products among Generation Z. The impact is mediated by perceived green value and subjective norms, which reveal that those who are exposed to sustainability-focused contents value green products more and tend to comply with social norms toward purchasing green products.

6 Conclusion

This research study identifies social media as the main force which guides Generation Z towards environmentally conscious purchasing habits. Social media information sharing produces strong effects on consumer sustainable behavior which can be explained by perceived green value and subjective norms together with their substantial mediating power. The work status of individuals acts as a moderation factor that shows employed people tend to follow societal norms for sustainability practices.

The marketing sector should update its advertisement approaches through dynamic material production to exploit social media influence for driving environmental consciousness among people. Consumer loyalty toward Generation Z based brand trust develops when organizations use sustainability labels together with green labels and maintain open reporting practices.

Countries should develop laws which both authenticate sustainability claims and stop manufacturers from using false environmental statements. Collaborating with digital influencers who advocate sustainability alongside other fans of environmentally friendly shopping can help drive such purchasing behavior. Young consumers need education about sustainable digital literacy through curriculum learning modules which educational institutions should integrate to develop their abilities to make informed eco-friendly choices.

Further investigations need to analyze how sustainability initiatives started through social media affect the environment in the long run while examining cultural trends and artificial intelligence-based technological developments about post-social-media green buying behaviors. Methodical assessments combined with strategic upgrades will lead to the development of a permanent sustainable environment where consumers receive informed green-based educations supported by commercial and regulatory establishments.

References

Chen, Y.S., Chang, C.H.: Enhancing Green Purchase Intentions. Management Decision (2012)

Kim, A.J., Ko, E.: Social Media Marketing and Consumer Equity. J. Bus. Res. (2012)

Zou, J., Tang, Y., Qing, P., Li, H., Razzaq, A.: Donation or discount: effect of promotion mode on green consumption behavior. Int. J. Environ. Res. Public Health **2021**, 18 (1912)

Chen, S.C., Lin, C.P.: Understanding the effect of social media marketing activities: the mediation of social identification, perceived value, and satisfaction. Technol. Forecast. Soc. Chang. **140**, 22–32 (2019)

Kumar, V., Choi, J.B., Greene, M.: Synergistic effects of social media and traditional marketing on brand sales: capturing the time-varying effects. J. Acad. Mark. Sci. **45**(2), 268–288 (2016)
Peattie, K., Crane, A.: Green marketing: legend, myth, farce, or prophecy? J. Cetacean Res. Manag. **8**(4), 357–370 (2005)
Luqman, A., Masood, A., Shahzad, F., Feng, Y.: Untangling the adverse effects of late-night usage of smartphone-based SNS among University students. Behav. Inf. Technol. **40**, 1671–1687 (2021)
Huang, H., Long, R., Chen, H., Sun, K., Li, Q.: Exploring public attention about green consumption on Sina Weibo: Using text mining and deep learning. Sustain. Prod. Consum. **30**, 674–685 (2022)

Mamdani Fuzzy Inference System Based on Multi-Textural Biomarkers for Alzheimer's Stage Detection

A. R. Kavitha[1(✉)], M. Ramya[2], T. N. Charanya[1], P. Lita Pansy[1], and E. Bhuvaneswari[3]

[1] Department of IT, Chennai Institute of Technology, Kundrathur, Chennai 600069, India
arkavithabalaji@gmail.com, {charanyatn, litapansyd}@citchennai.net

[2] Department of ECE, Mohamed Sathak A. J College of Engineering, Siruseri, Egattur, Chennai, Tamil Nadu 603103, India
ramshems@gmail.com

[3] Department of Artificial Intelligence and Data Science, Panimalar Engineering College, Chennai 600069, Tamil Nadu, India
bhuvaname2008@gmail.com

Abstract. The chronic brain disease known as Alzheimer's disease (AD) primarily affects short-term memory while advancing through its neurodegenerative stages. The initial symptoms of the disease develop gradually before the condition deteriorates thus early detection becomes vital. A Machine learning approach powers the Disease Diagnostic System (DDS) that analyzes T2 weighted Magnetic Resonance Imaging (MRI) scans from Alzheimer's Disease Neuroimaging Initiative (ADNI) database. The paper examines the Hippocampus and amygdala located in the left hemisphere of the human brain as Region of Interest (ROI) which is extracted from small cohort MRI scans. The $\mu \pm 3\sigma$ normalization method applies to segmented ROI while first-order histogram features extract Skewness and Kurtosis values. The Region of Interest receives two-dimensional wavelet features which include the Max norm of the original image and the Diagonal Detail Coefficient. The extracted textural markers serve as inputs to build a Mamdani FIS which defines minimum rules for AD stage diagnosis. The proposed classification method delivers accuracy levels of 96.13% for AD vs NC diagnosis and 94.73% for MCI vs NC diagnosis and 93.11% for AD vs MCI diagnosis. Multiple feature extraction through biomarker texture analysis enables better decision generation within the expert system framework.

Keywords: Alzheimer's diseases · mamdani fuzzy inference system · two-dimensional discrete wavelet transform · mild cognitive impairment · alzheimer's disease neuroimaging initiative

S. Fong et al. (Eds.): ICT4SD 2025, LNNS 1651, pp. 37–57, 2026.
https://doi.org/10.1007/978-3-032-06688-6_5

1 Introduction

Alzheimer's disease (AD) stands as the main dementia type that stops people from doing hard mental work for extended periods. The worldwide number of dementia patients totals 24 million and developing nations hold 16 million of these patients. When Alzheimer's disease reaches its advanced stage doctors confirm permanent brain damage. Scientists worldwide perform research to detect Alzheimer's disease through MCI assessment at its initial stage. Neuropsychological tests like Montreal Cognitive Assessment, Mini-Mental State Examination (MMSE), and Addenbrooke's Cognitive Examination help doctors identify cognitive problems in Alzheimer's Dementia patients (Bruno and Vignaga, 2019). Research shows that Hyperphosphorylated tau protein and Amyloid-Betapeptide (Aβ) show up in late stage AD scans through FDG-PET tests (Rajmohan and Reddy, 2017).

A biomarker serves as a measurable sign to detect and assess disease conditions in human beings (Kavitha and Chellamuthu, 2016). Research studies use both invasive and non invasive biomarkers to help identify and forecast AD development. Scientists test invasive markers by taking samples from the cerebrospinal fluid (CSF) and blood plasma in the body. Biomarkers that do not need invasive procedures include Imaging techniques such as Computer Tomography (CT) scans and MRI plus non-image tests like Electroencephalograph (EEG) and Electromyography (EMG) recordings along with smell detection tests (Kavitha and Thyagharajan, 2018; Vijay et al., 2016). Since CSF collection requires lumbar puncture which can be hard to perform safely in elderly patients due to their spine conditions the technique remains somewhat invasive. Tests that do not need invasive techniques are both secure and dependable for regular screenings. A doctor's evaluation takes too much time and needs trained medical staff which increases healthcare expenses and produces inconsistent results at diagnosis. AI systems today work automatically to detect diseases by reading medical images for diagnosis in medical research (Ramya and Kavitha, 2020). Many important issues affect medical image analysis when soft computing methods are used including problems with segmenting images accurately and choosing the best features for classification (Dong et al., 2015).

The patient data attributes are collected and processed with the probabilistic clustering technique in to characterize the heterogeneity of brain diseases (Zhang et al., 2016). The landmark-based feature was extracted from MRI neuroimages and the Alzheimer classification was performed using the SVM model (Jha et al., 2018). In, the Wavelet transform-based feature extraction and ELM classifier were investigated (Zeng et al., 2018). The prediction performance has been enhanced by using the AlexNet-based feature extractor in, (Shakarami et al., 2020; AlSaeed and Omar, 2022). Although it is restricted by the requirement of annotated data in supervised training.

(Shakarami et al., 2020) **This research work describes a Computer-Aided Diagnosis system (CADs) aimed at identifying Alzheimer's disease. In the suggested approach to reduce the intensive volumetric computations, 2D slices are utilized along with a 2D convolutional neural network (CNN), whereby only the higher-quality half of the slices is retained while the other half is discarded. Additionally, an enhanced AlexNet-SVM technique is introduced for feature extraction and classification, which aims to lessen computation volume while boosting accuracy and**

efficiency. This study emphasizes the importance of PET images due to their strong capability in illustrating body metabolism. The experimental outcomes indicate that the proposed model outperforms existing models by enhancing performance and accuracy (up to 96.39%) while reducing complex computations.

This research used an optimized SVM system to spot Alzheimer's disease (Gayathri and Munusamy, 2018). In, pre-trained CNN-derived MRI features were categorized using several classifiers such as SoftMax, RF, and SVM (Kar and Majumder, 2019).

The proposed DDS helps identify the human subjects' health needs by detecting Prodromal and AD stages through MFIS testing in cognitively normal subjects. A new test must identify cognitive changes that happen before symptoms start. This research project focuses on creating biomarkers for AD and uses these findings to group different research teams. The major contribution of this work is as follows:

- A multi-textural biomarker-based Mamdani fuzzy inference system is proposed to predict AD at an early stage.
- Hippocampus, and amygdala regions features have been extracted with histogram analysis and a 2D-DWT-based strategy, which offers robust disease-related features.
- A minimum number rule-based unsupervised Mamdani Fuzzy inference system is presented to classify the disease stage with extracted biomarker features (skewness, kurtosis, Max norm of original, and the Diagonal Detail Coefficient).
- The ADNI database is leveraged to evaluate the system's performance., the experimental finding proves that the proposed system provides superior results without depending on huge annotated data and complex gradient descent training.

In the remaining part of the paper, Sect. 2 contains a short description of studies relevant to the topic. In Sect. 3 we discuss our proposed approach to classify AD using MFIS soft computing. Section 4 presents the test results while Sect. 5 wraps up our research findings.

2 Literature Survey

Medical imaging must spot Alzheimer's disease early to help people stay active at work and protect their communities. Research teams have developed several smart methods to predict Alzheimer's disease development from non-invasive medical images but these approaches have important shortcomings.

In 2018 Gayathri, D.S. and Munusamy, N. described how Discrete Wavelet Transform pulled features out of their system (Duraisamy et al., 2019). Using ANFIS the researchers classified AD from MRI brain segmentations. This extraction method did not produce enough relevant data for a precise prediction system. The research team used DTI measurements of the human brain to determine AD status through an ANFIS model (Sharma et al., 2021). The model performance tests happened with data from ADNI. Duraisamy and his team created a feature selection method using multiple criteria to pick features from the hippocampus and PCC (Posterior Cingulate Cortex) regions in their study (Richhariya et al., 2020). The FCM-oriented FWPNN learning network helps classify Alzheimer's by combining supervised and unsupervised learning benefits. The ADNI and Border-3 dataset was used to test if the model performed well. The system could not extract important MRI information because of its technical difficulties.

In 2021 Sharma, R., et al. developed the DRVFL model using fuzzy activation functions to address deep learning training problems. The transfer learning model helped us obtain the essential attributes from extracted brain slices (Elshatoury et al., 2019). The FAF-DRVFL model analyzed the feature by determining the hidden layer's output through fuzzy activation functions. Our team tested the established method with ADNI data to reach 86.67% prediction accuracy (Vaithinathan et al., 2019). Our method considers structural changes but fails to reach acceptable performance results. To improve Alzheimer's diagnosis accuracy the USVM-RFE learning model by Richhariya and colleagues extracts important data information. The method selected features based on existing knowledge about data patterns. We selected Benchmark ADNI database to test our system's performance. This strategy necessitates more training time.

In 2019 Elshatoury, H et al. examined different machine learning methods using MRI neuroimaging scans. The researchers applied volume histograms to manage image data dimensions and speed up their work. Majority voting achieves the best recognition rate of 69.5% among different machine learning methods when processing the ADNI dataset. Vaithinathan K and associates created a method in 2019 to extract features from specific areas of interest. The study used Fisher ranking, SVM-RFE, and elastic net to find complex patterns before using ML methods to determine AD stages (Kour et al., 2019). The texture analysis method needs additional computer processing power. In 2019 Kour H and colleagues researched different fuzzy logic techniques for comparison. We evaluated performance of ANN and ANFIS systems plus FLS using AD data from Kaggle. The research results demonstrate that the ANFIS hybrid model outperforms all other tested methods.

In 2019 Mallika, R.M et al. used fuzzy soft computing to find AD diagnosis stages. The research team separated enhanced images using multi-level thresholding and analyzed hippocampus volume textures before performing classification through FIS (Mallika et al., 2019). The model does not work well in ADNI data because it lacks proper image texture evaluation. In 2021 Stirling J and his team set up a stage forecasting system for Alzheimer's disease using fuzzy intelligence. The SOF system used Recursive feature elimination to pick necessary features from the dataset. The selected approach was tested using the ADNI database (Stirling et al., 2021). The system performed poorly due to the small number of features it returned.

According to the literature review it concludes, some of the aforementioned strategies failed to assess relevant characteristics from the given neuroimage, and the complexity of some techniques grew while dealing with textural attributes. Some of them used the deep learning potential, and though they provided adequate results, the requirement of huge annotated data and gradient descent training complicated them in real-time applications. Therefore, this research proposes a novel biomarker feature analysis technique to overcome the aforementioned barriers.

3 Proposed Methodology

To improve predictive capacity, a new idea has been developed for AD classification, using a multiple feature extraction approach and a Fuzzy inference-based technique. Our system used multiple features extraction to get First order histogram features (Skewness

and Kurtosis) and two-dimensional wavelet features (Max norm of the original image and Diagonal Detail Coefficient). AD received its diagnosis through the Mamdanin fuzzy logic system based on discovered textural biomarkers. Our system includes four main steps: Subject Data Collection, Image Segmentation, Feature Measurement, Fuzzy Inference System. Our proposed DDS system structure appears in Fig. 1 as a diagram. Propose muti textural feature extraction-based framework is given in algorithm 1.

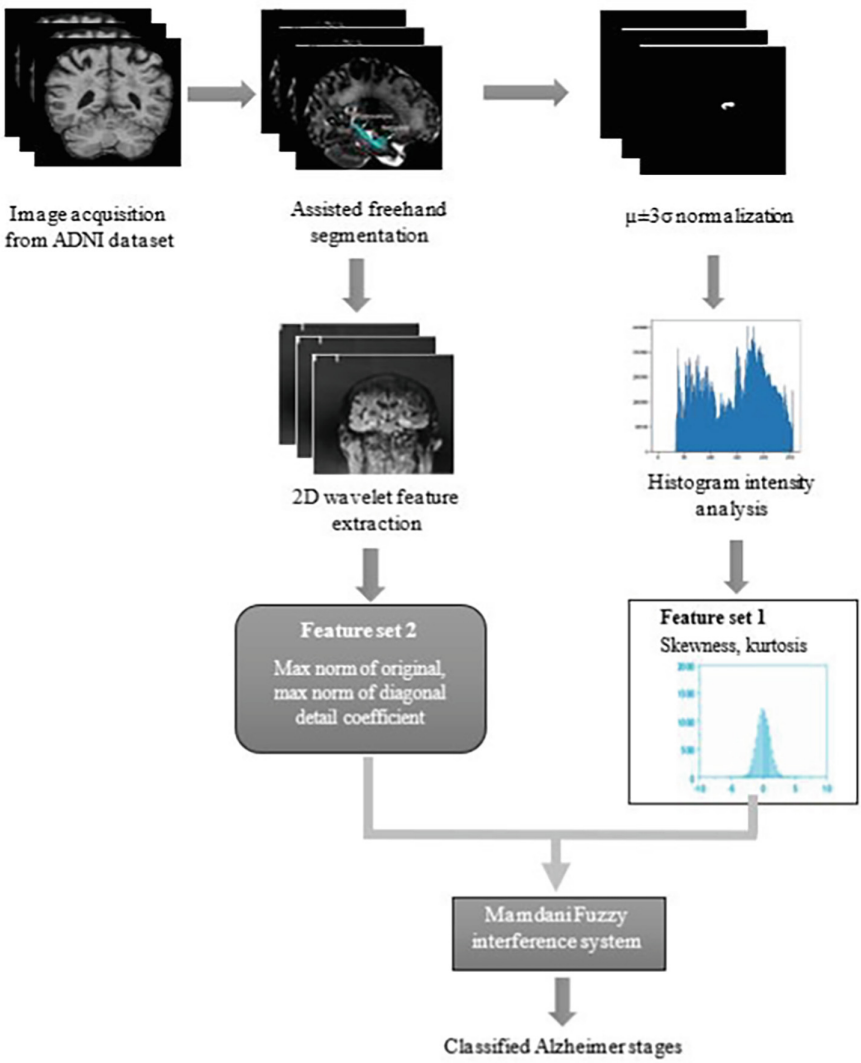


Fig. 1. Schematic diagram of the proposed DDS modules

Algorithm 1: Features extraction Algorithm identifies the textural Biomarkers.

Input: T2 –MR Images of CN, MCI, LMCI, and AD from ADNI-3 database.

Output: Four Textural Features - First order histogram features (skewness and kurtosis) and 2D wavelet textural features (Max norm of the original image and the Diagonal Detail Coefficient)

1. **function** Textural Biomarkers (T2-MR images)
2. **for all** MR images **do** // Segmentation
3. Segmenting Amygdalaand Hippocampus in the left hemispherical region of the acquired MR images using assisted freehand segmentation techniques to get Seg ROI.
4. **end for**
5. **for all** Seg_ROI **do** // First order histogram features extraction
6. Compute the mean (μ) and Standard Deviation (σ) of the seg_ROIand then determine μ+3σ and μ-3σ. // μ±3σ Normalization - Pre-processing technique
7. Calculate the Intensity Histogram (Int_Hist) using the BinLimits as μ+3σ and μ-3σ.
   Int_Hist = histogram (Seg_ROI, 'Binlimits', [μ-3σ, μ+3σ]) (1)
8. If (NumBins> 45 and values are zeros near, both the ends of the histogram then setNumBins = 45)
9. Calculate the Skewness and Kurtosis for the values of Int_Hist(Vector)

$$\text{Skewness} = \sum_{i=1}^{N} (\text{Hi} - \text{M})^3 / \text{N(s)}^3 \quad (2)$$

$$\text{Kurtosis} = \sum_{i=1}^{N} (\text{Hi} - \text{M})^4 / \text{N(s)}^4 \quad (3)$$

   Where H is the value of Int_Hist,M is the mean of H, s is the standard deviation of HandN is the NumBins.
10. **end for**
11. **for all** Seg_ROI **do** //2D wavelet textural features extraction
12. Finding the Max norm of the original image and the Diagonal Detail Coefficient using 2-D wavelet decomposition concerning a db2 wavelet at level 2.
13. **end for**
14. return skewness, kurtosis, Max norms of the original image, and the diagonal detail coefficients.
15. **end function**

3.1 Subjects and Data Acquisition

To assess the effectiveness of the proposed framework MRI images of the small cohort size have been obtained from ADNI – 3 databases. 3T Siemens MRI scanners were

leveraged to acquire the T2 – weighted MRI images with the following protocol: Pulse Sequence = SE/IR, Slice Thickness = 1.2 mm, repetition time (TR) = 4800.0 ms, echo time (TE) = 441.0 ms, flip angle of 120-degree, sagittal acquisition plane, Matrix X = 256.0 pixels, Matrix Y = 256.0 pixels, Matrix Z = 160.0, Mfg. Model = Prisma fit, pixel spacing X = 1.0mm, pixel spacing Y = 1.0 mm. The system obtained the images for each subject group in Archive format. Our system collected MRI scans from individuals across three groups including healthy subjects and people with Prodromal MCI or AD. Our research begins by presenting the MRI scan examples in Fig. 2 during its first phase.

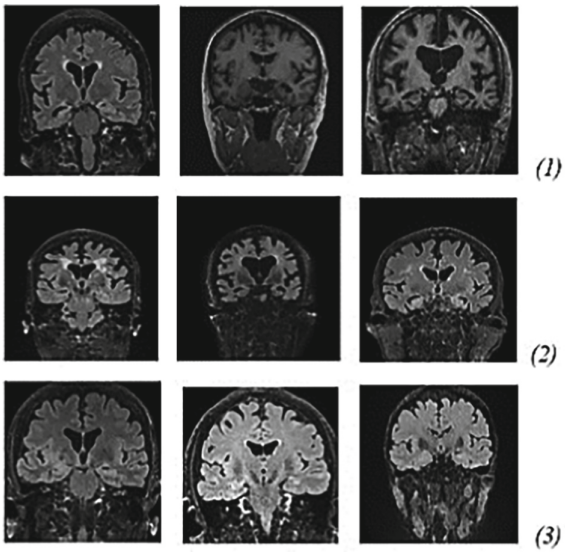


Fig. 2. Samples of T2 Weighted MRI Scans in the Coronal view for three groups of the subjects considered in this study. (1) AD, (2) MCI, and (3) CN

3.2 Segmentation and Pre-Processing

The research team conducted assisted freehand segmentation for 13 MRI image sets including six CN subjects, three MCI subjects, one LMCI subject, and three numbers of AD subjects. **To achieve better Region of Interest (ROI), Hippocampus and amygdala tracing, the assisted free hand segmentation in the Image Segmenter tool of the MATLAB R2018b has been used.** The Fig. 3 displays segmented ROI data results from three distinct groups of participants. The $\mu \pm 3\sigma$ normalization method (Collewet et al., 2004) was used to prepare the segmented ROI data by adjusting the gray values in the ROI to match the mean value (μ) plus or minus three standard deviations (σ). The researchers adjusted the data to make the difference between ROI feature sets stand out for different groups as recommended by (Gonzalez, 2009). The pre-processing technique sets the Intensity range in the ROI histogram between $\mu + 3\sigma$ and $\mu - 3\sigma$ by removing pixels that go beyond this range.

3.3 Feature Extraction

The first order Histogram based textural features Skewness and Kurtosis depend on the distribution of pixel values in the image textural features (López-Gómez C et al., 2018). Our normalization process revealed many high intensity bins when we measured the histogram of CN and MCI ROIs. The number of histogram bins at zero and both ends will not affect the intensity histogram results when set to 45. This simplifies the feature extraction process. The researchers used $\mu + 3\sigma$ and $\mu - 3\sigma$ values from the separate subject group's ROI segments to set bin limits for intensity histogram calculations. The skewness and kurtosis measurements came from the normalized ROI intensity histogram H.

The 2D-DWT needs two scaling function φ (x, y) and three wavelets ψH (x, y), ψV(x, y),andψ D(x, y). These wavelets analyze how image intensity changes across multiple directions (Ramya and Kavitha 2021). ψH tracks variations within columns. ψV tracks variations along rows. ψ D tracks variations along diagonals (Ramya and Kavitha 2021). 2D – Discrete Wavelet analysis was applied to segmented ROIs (hippocampus and amygdala) to extract 2D – DWT features (Jang et al., 1997). Research found greater changes in the Max norms values of diagonal detail coefficient matrices when comparing AD and CN brain images. Our research showed that all three coefficient matrices matched perfectly for MCI and CN brain images. Our team selected the Max norms from diagonal detail coefficient arrays as a 2D wavelet textural biomarker. 2D wavelet decomposition utilized a db2 wavelet at level 2 to extract these values from both original images and their diagonal details.

We extract four features from the First order histogram and the two-dimensional wavelet features including Skewness and Kurtosis plus Max norm of the original image and Diagonal Detail Coefficient. Our research team determined the values of textural features from MRI scans for 13 subjects in Table 1.

Table 1. Values of the Textural Features from the MRI of the individual patient

Subject Number	Skewness	Kurtosis	Max norm of Diagonal Detail coefficient	Max norm of original Image
1	3.8	16	37.33	77
2	6.48	43.02	102	213
3	4.25	19.05	72.22	158
4	6	37.03	125	223
5	5.83	35.08	101.6	196
6	6.25	40.02	104	197
7	5.2	28.03	80.53	173
8	6.41	42.02	87.49	175
9	4.36	20.05	46.27	113
10	6.08	38	108.8	158
11	4	17.05	72.89	119

(*continued*)

Table 1. (*continued*)

Subject Number	Skewness	Kurtosis	Max norm of Diagonal Detail coefficient	Max norm of original Image
12	6.4	42	91.5	189
13	4.12	18.05	70.74	158

3.4 Fuzzy Expert System

The fuzzy logic controller converts expert defined rules into autonomous system controls for automatic control of the electrical vehicle. MFIS includes five working units that perform specific tasks. The system consists of Rules, Database, Decisions, Fuzzification and Defuzzification units. Our new MFIS design appears in Fig. 3

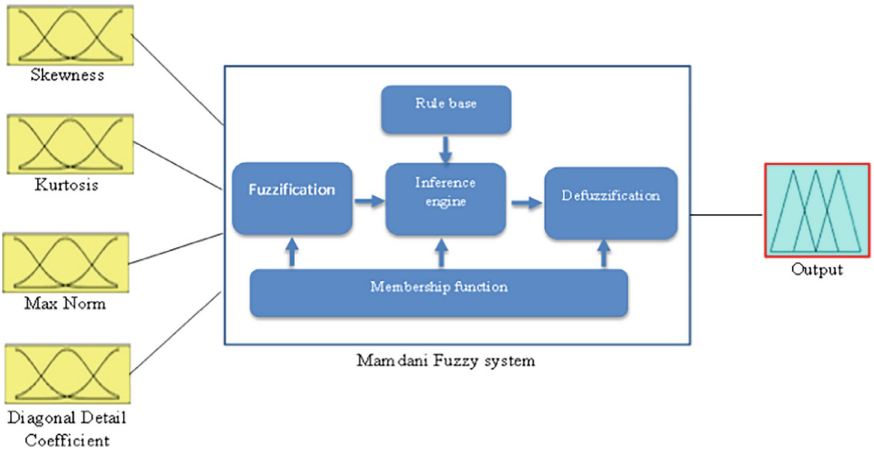


Fig. 3. Proposed MFIS representation

The proposed system used Skewness, Kurtosis, and 2D – DWT textural features including Max norm of the original image and the Diagonal Detail Coefficient as inputs to the Fuzzy Expert System to categorize CN, MCI with LMCI and AD based on output values from MFIS. Our fuzzy rule base system uses both triangular and trapezoidal Membership Functions. The system used the ranges listed in Table 2 to define the linguistic values for input and output parameters. The system needs specific range settings for its linguistic variables to produce better results.

Table 2. The range assigned for the MF of input and output parameters

Parameters	Skewness			Kurtosis			Max norm of Diagonal Detail coefficient			Max norm of original			Output Values from MFIS		
MF	Low	Medium	High	Small	Moderate	Large	L	M	H	S	Mod	Lar	AD	MCI	CN
Min	2	4	5.9	10	17	31	30	72	87.4	50	119	158	0	31	61
Max	4.37	5.8	10	20.06	35.08	50	71	101.6	125	158	158	225	30	60	100

3.5 Fuzzification

This module converts the crisp value of the four textural parameters (Skewness, Kurtosis, Max norm of the original image, and the Diagonal Detail Coefficient) into fuzzy quantities.

3.6 Decision Making Unit

The fuzzy rule base contains three rules that use Skewness, Kurtosis, Max norm, and Diagonal Detail Coefficient values to determine the output result. Our system features three evaluation rules which are documented in Table 3. These standards help to divide subjects into different recognition categories. The Mamdani Inference system served to test the rules.

Table 3. Rule Base of the system

Rule number	Skewness	Kurtosis	Max norm of Diagonal Detail coefficient	Max norm of original	Output Value
Rule 1	Low	Small	L	None	AD
Rule 2	Medium	Moderate	M	None	MCI
Rule 3	High	Large	H	Lar	CN

3.7 Defuzzification

The center of gravity defuzzification method was used to transfer the Fuzzy Inference outputtothe crisp output

$$Z = \mu A(z)\ Zdz/\mu A(z)\ d \tag{4}$$

whereas $\mu A(z)$ is the Membership function of the aggregated output. By using Eq. (4) the crisp values of the MFIS outputs are calculated.

3.8 Proposed MFIS Mathematical Representation

The input variable is divided into three fuzzy sets: "low," "medium," and "high." Membership functions include trapezoidal and triangular fuzzy numbers. The Mamdani Max-Min Inference model was used to calculate the membership function value of the framework output.

A sample of the triangular and trapezoidal MF used in the MFIS:

Triangular MF(μMCI) for the linguistic variable MCI of the output value

$$\mu\text{MCI (x; 31, 45, 60)} = \begin{cases} 0 & x \leq 31 \\ (x-31)/14 & 31 \leq x \leq 45 \\ (60-x)/15 & 45 \leq x \leq 60 \\ 0 & 60 \leq x \end{cases} \tag{5}$$

Trapezoidal MF(μHigh) for the linguistic variable High of the Skewness parameter.

$$\mu\text{High (x; 5.9, 6, 6.5, 10)} = \begin{cases} 0 & \text{x} \leq 5.9 \\ (\text{x} - 5.9)/0.1 & 5.9 \leq \text{x} \leq 6 \\ 1 & 6 \leq \text{x} \leq 6.5 \\ 1 & 6 \leq \text{x} \leq 6.5 \\ 0 & 10 \leq \text{x} \end{cases} \tag{6}$$

For example, from the Rule viewer of the MFIS, if the values for the input parameters are 5.2 for Skewness, 28.03 for Kurtosis, 80.53 for Max norm of Diagonal detail coefficient and 173 for Max norm of original image then the obtained output value is 45.35, which is in the range for the linguistic variable MCI. For this case, the proposed DDS classifies the subject as MCI.

Expression for triangular membership function using min and max is

$$Triangle(x; a, b, c) = \max\left(\min\left(\frac{x-a}{b-a}, \frac{c-x}{c-b}\right), 0\right) \tag{7}$$

Expression for trapizoidal membership function is

$$f(x; a, b, c, d) = \max\left(\min\left(\left[\frac{x-a}{b-a}\right], 1, \left[\frac{d-x}{d-c}\right]\right), 0\right) \tag{8}$$

Firing strength of each Rule is calculated using the following equations.

$$R1 = \min[\mu Low(Skewness), \mu Small(Kurtosis), \mu L(Max\ norm\ of\ Dd), \mu S(Max\ norm\ of\ original\ image) \tag{9}$$

$$R2 = \min[\mu Medium(Skewness), \mu Moderate(Kurtosis) \mu M(Max\ norm\ of\ Dd), \mu Mod(Max\ norm\ of\ original\ image) \tag{A.4}$$

$$R3 = \min[\mu High(Skewness), \mu Large(Kurtosis), \mu H(Max\ norm\ of\ Dd)\mu Lar(Max\ norm\ of\ original\ image)] \tag{10}$$

- μlow (Skewness), μmedium (Skewness) and μHigh (Skewness) are the MF of the Skewness.
- μsmall (Kurtosis), μModerate (Kurtosis), μlarge (Kurtosis) are the MF of the Kurtosis.
- μL (Max norm of Dd), μM (Max norm of Dd), μH (Max norm of Dd) are the MF of the Max norm of Diagonal detail coefficient.
- μS (Max norm of original image), μMod (Max norm of original image), μLar (Max norm of original image) are the MF of the Max norm of original image.

For example, for the patient 12, the textural features values are, Skewness is 6.4, Kurtosis is 42, Max norm of Diagonal detail is 91.5 and Max norm of original image is 189.

Rule firing strength calculation:

By using the range of the μHigh (Skewness), (5.9, 6, 6.5, 10) for the parameter (a, b, c, d) in the equation [A.2] and x = 6.4

$$\mu\text{High (Skewness)} = \max\left(\min\left(\left[\frac{6.4-5.9}{0.1}\right], 1, \left[\frac{10-6.4}{3.5}\right]\right), 0\right) = \max(1, 0) = 1. \tag{11}$$

By using the range of the μlarge (Kurtosis), (31, 32, 45, 50) for the parameter (a, b, c, d) in the equation [A.2] and x = 42

$$\mu\text{large(Kurtosis)} = \max\left(\min\left(\left[\frac{42-31}{1}\right], 1, \left[\frac{50-42}{5}\right]\right), 0\right) = \max\,(1, 0) = 1. \tag{12}$$

By using the range of the μH (Max norm of Dd), (87.4, 87.5, 124, 125) for the parameter (a, b, c, d) in the equation [A.2] and x = 91.5

$$\mu\text{H(Max norm of Dd)} = \max\left(\min\left(\left[\frac{91.5-87.4}{0.1}\right], 1, \left[\frac{125-91.5}{1}\right]\right), 0\right) = \max(1, 0) = 1. \tag{13}$$

By using the range of the μLar (Max norm of original image), (158, 170, 224, 225) for the parameter (a, b, c, d) in the equation [A.2] and x = 189

$$\mu\text{Lar(Max norm of original image)} = \max\left(\min\left(\left[\frac{189-158}{12}\right], 1, \left[\frac{225-189}{1}\right]\right), 0\right) = \max(1, 0) = 1. \tag{14}$$

Similarly values of other MF has been calculated using the Eq. (7) and (8).

By using the calculated MF values in equation [A.3], [A.4], [A.5], firing strength of each rule were calculated.

It has been found that R1 = 0, R2 = 0 and R3 = min (1, 1, 1, 1) = 1.

Therefore, for the values (obtained from the MRI scan of patient 12) of the input parameters, Skewness is 6.4, Kurtosis is 42, Max norm of Diagonal detail is 91.5 and Max norm of original image is 189. Only Rule 3 was fired and its value is 1.

$$\text{Fis} = \text{readfis ('AD\_class.fis')} \tag{15}$$

$$[Output, fuzziziedIn, ruleOut, aggregatedOut, ruleFiring] = evalfis(fis, [inputs]) \tag{16}$$

By using the Eqs. (15) and (16) values of rule Firing have been validated. The intermediate results of fuzzy Inference process can also be obtained using the Eq. (17) and (16) in the command window of the MATLAB software. Algorithm 2 provides the pseudo code for the proposed MFIS.

Algorithm 2: Fuzzy Rule-based approach to finding the classified Output values.

1. **Input:** AD classification (Name), Mamdani (Type), min (and Method), max (or Method), min (ImpMethod), max (AggMethod), centroid (DefuzzMethod).
2. **Output:** Stages of AD.
3. **Function** Fuzzy rule system (four textural features)
4. Define the MFs for all the four input textural features and one output parameter.
5. **for all** input features and Output values **do**
6. Determine the range for all the (Triangular and Trapezoidal) MFs.
7. **end for**
8. Analyse the MFs values for all the four textural features and the Output value parameters.
9. **for all** classified outputs **do**
10. Define the Fuzzy rules
11. **end for**
12. Execute MFIS and determine the classified output (AD, MCI, and CN) based on the Output
13. value from MFIS.
14. **for all** MRI images **do**
15. Compare the Actual group name of the subject and the classified output
16. end for
17. If (the mismatch in the Actual group name of the subject and the classifiedoutputis>1) then Modify I). The range values of the MFs for the input textural features,II) The fuzzy rules list and go to step 8
18. return classified output (from MFIS with Optimized rules).
19. **end function**

4 Test Results and Discussions

This section demonstrates how well the proposed technique works alongside different recognized benchmarks.

4.1 Implementation Details

The software used to perform freehand segmentation, normalization, multiple feature extraction, and MFIS classification was developed from MATLAB (R2018b, The Math Works Inc., Natick, MA, USA).

4.2 Evaluation Metrics

The proposed system is evaluated using the metrics listed below. MATLAB R2018b software is used for performance evaluation and analysis.

The accuracy of a classification framework is used to calculate its overall performance.

$$Accuracy = \frac{TP + TN}{N}$$

The percentage of correctly classified subjects in which the patients having MCI/AD is calculated by using precision.

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

The percentage of correctly classified subjects in which the patient does not have MCI/AD is calculated using specificity.

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

The % of correctly classified samples in which the patient having NC /AD/ MCI/ is quantified by the term sensitivity.

$$Sensitivity = \frac{TP}{FN + TP}$$

The harmonic means of the precision and recall of a classifier is used to compute F1 score.

$$F1score = \frac{2(Precision \times Recall)}{(precision + Recall)}$$

The correctly predicted AD disease stage are TP (True Positive) and FN (False Negative). Likewise the inaccurately classified stages is denoted by TN (True Negative) and TP. For assessment the AD vs NC, MCI vs NC and AD vs MCI subject analysis has been carried out in this experiment.

4.3 Result Analysis

The proposed multi-textural biomarker-based Mamdani fuzzy inference system approach in ADNI database is analysed. The classification is made for AD vs NC, MCI vs NC and AD vs MCI classes. Tested output Values of the DDS has shown in Table 4.Out of 13subjects, the output value from the MFIS for one CN subject falls in the MCI range. Based on the output values from MFIS, classification of the prodromal MCI stage, AD and CN subjects has been done with improved accuracy for the small group of 13 subjects.

Table 4. Tested Values of the system

Subject Number	Actual Group name of the Subject	Output Values from FIS	Classified Output	Defined range of Output Value
1	AD	15	AD	0–30
2	CN	80.333	CN	61–100
3	MCI	45.463	MCI	31–60
4	CN	80.333	CN	61–100
5	MCI	50	MCI	31–60
6	CN	80.333	CN	61–100
7	MCI	45.353	MCI	31–60
8	CN	80.337	CN	61–100
9	AD	15	AD	0–30
10	CN	50	MCI	31–60
11	MCI	50	MCI	31–60
12	CN	80.333	CN	61–100
13	AD	15	AD	0–30

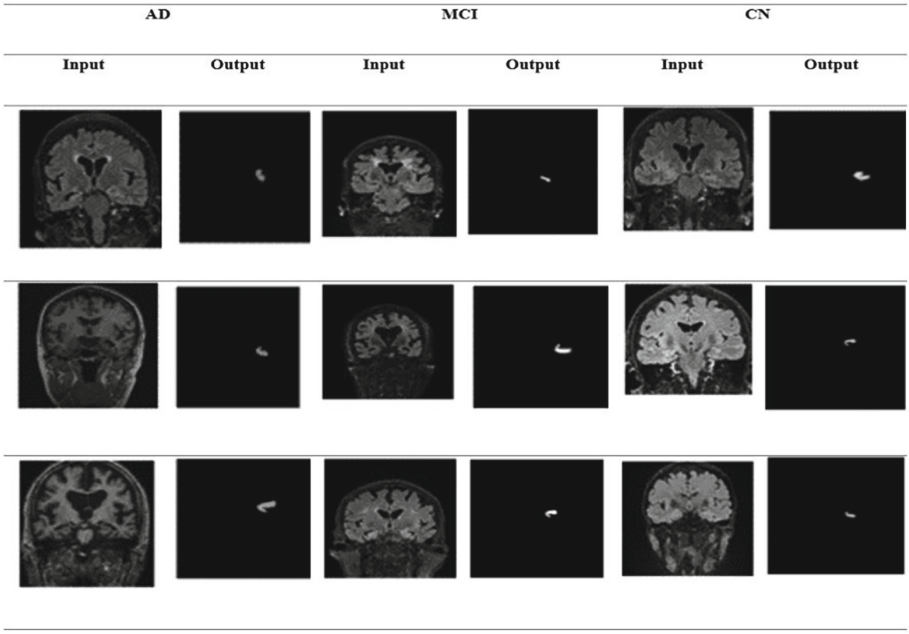


Fig. 4. Segmented ROI results of three different groups of the subjects

The Fig. 4 shows how our system assists free hand segmentation. The system identifies Amygdala and Hippocampus regions in every image to perform textural analysis. The system achieved good and straightforward results when it processed the selected

area. Looking at only the needed portion of an image helps the system work better and do its work faster.

Table 5. Performance analysis with different feature extraction techniques

methods	Class	Accuracy	Sensitivity	Specificity	Precision	F1 score
Hand craft feature + MFIS	AD vs NC	73.83	71.82	73.92	74.92	74.07
	MCI vs NC	71.27	70.13	71.61	72.92	72.89
	AD vs MCI	69.76	65.84	63.24	70.32	70.47
Histogram features + MFIS	AD vs NC	79.38	76.74	75.99	77.03	79.04
	MCI vs NC	76.83	78.76	78.51	79.02	75.79
	AD vs MCI	81.73	83.72	81.28	80.62	83.57
2D-DWT features + MFIS	AD vs NC	85.46	83.73	84.93	85.83	86.56
	MCI vs NC	83.38	82.36	81.63	81.02	85.58
	AD vs MCI	84.48	85.63	83.72	83.83	85.38
Histogram + 2D DWT features + MFIS *(proposed)*	AD vs NC	96.13	97.32	96.53	94.24	95.28
	MCI vs NC	94.73	96.73	95.563	95.92	94.83
	AD vs MCI	93.01	93.76	92.42	93.19	93.64

Table 5 depicts the proposed system's ablation study. In comparison to all other feature extraction approaches, the proposed multiple feature extraction approach achieves a high performance of 96.13% for AD vs NC classification, 94.73% for AD vs MCI classification, and 93.01% for AD vs MCI classification. Using handcrafted features is a traditional process; however, the feature description was manually collected for analysis, making feature engineering more difficult. Due to insufficient details, using histogram analysis-based features descriptor in ROI does not produce satisfactory results. Using a 2D-dWT-based technique for the max norm feature in both the original and diagonal detail coefficient features also does not yield efficient classified results. It is concluded that the proposed multiple feature extraction is superior and provides adequate prediction performance in MFIS.

Table 6. Computational time requirements for proposed multiple feature analysis

process	Computation time(minutes)
normalization	0.4
Feature extraction	0.9
classification	1.2

Table 6 depicts the proposed system's complexity level. The computation time and complexity of MFIS are lower than those of deep learning approaches, making the proposed system more practical. Because of the selected region (Hippocampus and Amygdala) analysis, both techniques (histogram and 2D-DWT) take 0.9 min for feature extraction. Figure 5 shows the ROC plot of the AD vs NC and MCI vs NC plot.

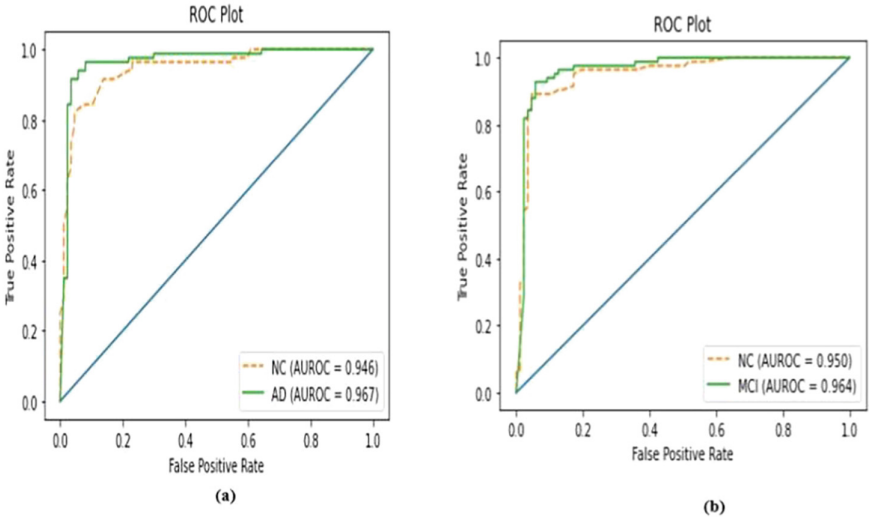


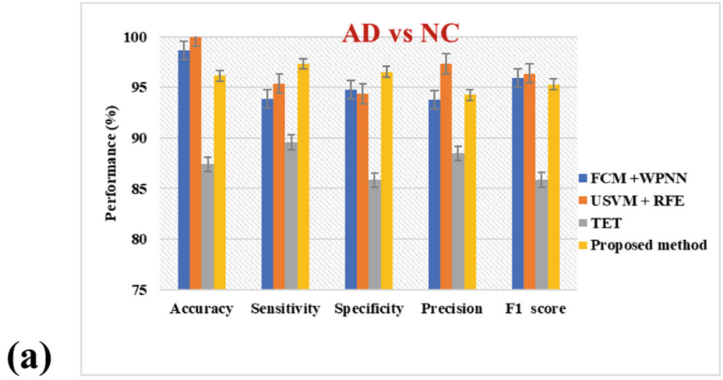
Fig. 5. ROC curve curve analysis of (a) AD vc NC and (b) MCI vs NC

4.4 Comparison of AD Diagnosis with Existing Approaches

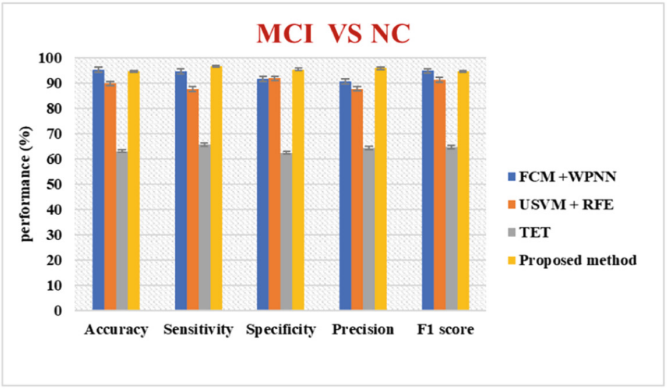
This section evaluates how well the new system performs against current leading methods in the field. Table 7 compares the AD vs NC class performance of proposed and existing approaches**. The AlexNet-SVM (Shakarami, A et al.) model produces better results than our approach by 0.26% as this method also uses PET (Positron Emission Tomography) and needing large labelled data sets. The proposed fuzzy expert system used manual feature extraction and selection from hippocampus and Amygdala regions which is the semi – automated unsupervised machine learning technique, produces comparatively 0.26% less accuracy. The proposed system uses 13 number of MRI images for analysis. As fuzzy expert system is the best classification method of choice for small number of dataset when compared to Deep learning method, our proposed method utilises this classification approach. The proposed unsupervised system achieved better accuracy results than all the other state of Art methods by 9.34%, 26.63%, and 9.6% and 19.27%. Our proposed model outperforms other approaches due to the biomarker textural features.**

Table 7. Proposed method performance comparison with different state-of-the-art techniques in ADNI database

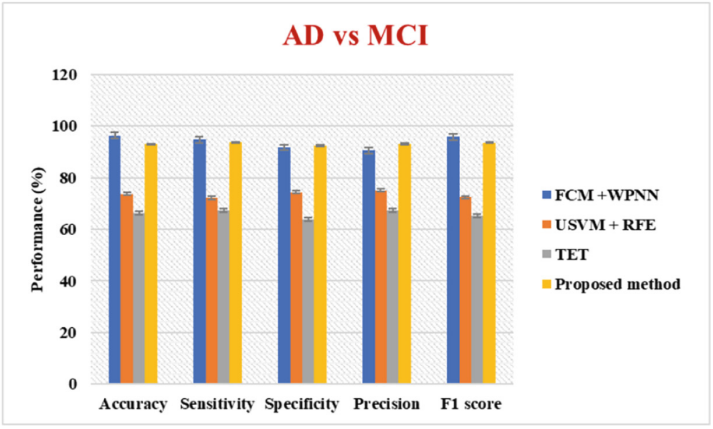
Techniques	Accuracy (%)
Sharma, R et al.	86.67
Elshatoury, H et al.	69.5
Mallika, R.M et al.	86.53
Shakarami, A et al.	96.39
Zeng, N et al.	76.857
Proposed model	96.13



(a)



(b)



(c)

Fig. 6. Performance comparison of (a)AD vs NC (b) MCI vs NC and (c) AD vs MCI with existing state-of-the-art approaches

The graph shows how different classification processes perform against one another through existing methods. Our model produced better results than the and approaches during the AD vs NC comparison with 2.5% and 3.87% enhancements respectively. Our method performs better than other approaches because its basic training methods produce weak results and require advanced training methods to succeed. Our model offers simpler training than other approaches due to its lower gradient training expenses. Our proposed methods show cost advantages and simple operation with unsupervised learning according to test outcomes. Our framework achieves better prediction results when it uses the designated biomarker region features for analysis (Fig. 6).

4.5 Discussion

The proposed model aims to produce a basic system that detects AD stages to help people start effective disease control methods. To test our model we use ADNI data and apply Mamdani Fuzzy Inference System (MFIS) to classify subjects into AD, MCI and CN groups based on their identified textural biomarkers. Our experimental results show the proposed model achieves 96.83% accuracy in differentiating AD and NC stages while achieving 93.01% accuracy in distinguishing MCI from NC subjects and 94.73% accuracy in differentiating AD from MCI. The proposed model surpasses both the training and annotated data dependent systems in its performance. The MFIS works as an easy yet powerful unsupervised classifier and its multiple feature extraction methods help researchers examine neuro imaging bio markers precisely. Our proposed multiple feature analysis system helps decrease healthcare professional workload by automatically predicting disease stages early on.

5 Conclusion

The Proposed Disease Diagnosis System used semi-automated 2D-wavelet analysis to classify AD stages through MRI scans of human brains. MRI scans show their main segmentation targets as the hippocampus and amygdala parts of the left brain hemisphere. Our system extracted both first-order histogram features (Skewness and Kurtosis) and two-dimensional wavelet features (Max norm of the original image and the diagonal detail coefficient) to classify brain images with the Fuzzy Inference System. Our Fuzzy system shows statistical importance in separating AD and MCI patients from CN subjects through MRI scans with thirteen improved test results. The proposed DDS system works well in community health clinics and helps radiologists determine how diseases advance. Our future research will include expanding MRI sample numbers and using free surfer to automate the ROI segmentation process. Identifying issues with the MFIS system through examination of its internal output data helps make the system work better.

Statements and Declarations

Acknowledgment. The authors would like to acknowledge ADNI – 3 databases for the dataset (MRI of different groups of subjects) provided by them to validate the research work.

Authors' Contributions. Ramya M, Kavitha A. R., and Kavitha S., conceptualized the model, collected the mamdani fuzzy inference system, two-dimensional discrete wavelet transform and mild cognitive impairment index details and reviewed the manuscript.

Funding. Authors did not receive any funding.

Conflicts of Interests. Authors do not have any conflicts.

Availability of Data and Material. Available on request.

Code availability. Not applicable.

References

AlSaeed, D., Omar, S.F.: Brain MRI analysis for Alzheimer's disease diagnosis using CNN-based feature extraction and machine learning. Sensors **22**(8), 2911 (2022)

Bruno, D., Vignaga, S.S.: Addenbrooke's cognitive examination III in the diagnosis of dementia: a critical review. Neuropsychiatr. Dis. Treat. **15**, 441 (2019)

Collewet, G., Strzelecki, M., Mariette, F.: Influence of MRI acquisition protocols and image intensity normalization methods on texture classification. Magn. Reson. Imaging **22**(1), 81–91 (2004)

Dong, A., Honnorat, N., Gaonkar, B., Davatzikos, C.: CHIMERA: clustering of heterogeneous disease effects via distribution matching of imaging patterns. IEEE Trans. Med. Imaging **35**(2), 612–621 (2015)

Duraisamy, B., Shanmugam, J.V., Annamalai, J.: Alzheimer disease detection from structural MR images using FCM based weighted probabilistic neural network. Brain Imaging Behav. **13**(1), 87–110 (2019)

Elshatoury, H,. Avots, E., Anbarjafari, G.: Alzheimer's Disease Neuroimaging Initiative. Volumetric histogram-based Alzheimer's disease detection using support vector machine. J. Alzheimer's Dis. **72**(2), 515–524 (2019)

Gayathri, D.S., Munusamy, N.: Classifying Alzheimer's disease using adaptive neuro fuzzy inference system. Int. J. Recent Technol. Eng. **7**(December 4S2), 227–33 (2018)

Gonzalez, R.C.: Digital Image Processing. Pearson Education, India (2009)

Jang, J.S., Sun, C.T., Mizutani, E.: Neuro-fuzzy and soft computing-a computational approach to learning and machine intelligence [Book Review]. IEEE Trans. Autom. Control **42**(10), 1482–1484 (1997)

Jha, D., Alam, S., Pyun, J.Y., Lee, K.H., Kwon, G.R.: Alzheimer's disease detection using extreme learning machine, complex dual tree wavelet principal coefficients and linear discriminant analysis. J. Med. Imag. Health Inform. **8**(5), 881–890 (2018)

Kar, S., Majumder, D.D.: A novel approach of diffusion tensor visualization based neuro fuzzy classification system for early detection of Alzheimer's disease. J. Alzheimer's Dis. Rep. **3**(1), 1–8 (2019)

Kavitha, A.R., Chellamuthu, C.: Brain tumour segmentation from MRI image using genetic algorithm with fuzzy initialisation and seeded modified region growing (GFSMRG) method. Imag. Sci. J. **64**(5), 285–297 (2016)

Kavitha, S., Thyagharajan, K.K.: Fuzzy qualitative reasoning model for astrocytoma brain tumor grade diagnosis. Indian J. Sci. Technol. **11**(38), 1–3 (2018)

Kour, H., Manhas, J., Sharma, V.: Evaluation of adaptive neuro-fuzzy inference system with artificial neural network and fuzzy logic in diagnosis of Alzheimer disease. In: 2019 6th International Conference on Computing for Sustainable Global Development (INDIACom), pp. 1041–1046. IEEE (2019)

López-Gómez, C., Ortiz-Ramón, R., Mollá-Olmos, E., Moratal, D.: Alzheimer's Disease Neuroimaging Initiative. ALTEA: a software tool for the evaluation of new biomarkers for Alzheimer's disease by means of textures analysis on magnetic resonance images. Diagnostics **8**(3), 47 (2018)

Mallika, R.M., UshaRani, K., Hemalatha, K.: A fuzzy-based expert system to diagnose Alzheimer's disease. In: Internet of Things and Personalized Healthcare Systems, pp. 65–74. Springer, Singapore (2019)

Rajmohan, R., Reddy, P.H.: Amyloid-beta and phosphorylated tau accumulations cause abnormalities at synapses of Alzheimer's disease neurons. J. Alzheimer's Dis. **57**(4), 975–999 (2017)

Ramya, M., Kavitha, A.R.: Textural analysis of mri using 2d – discrete wavelet transform: diagnosis of prodromal mild cognitive impairment stage and Alzheimer's disease stage. In: Recent Challenges in Science, Engineering and Technology, pp. 236–246 (2021)

Ramya, M.M., Kavitha, A.R.: Application of soft computing techniques for image analysis in disease diagnostic systems. Contemp. Res. Electron. Comput. Mech. Sci. 167 (2020)

Richhariya, B., Tanveer, M., Rashid, A.H.: Alzheimer's Disease Neuroimaging Initiative. Diagnosis of Alzheimer's disease using universum support vector machine based recursive feature elimination (USVM-RFE). Biomed. Signal Process. Control **59**,101903 (2020)

Shakarami, A., Tarrah, H., Mahdavi-Hormat, A.: A CAD system for diagnosing Alzheimer's disease using 2D slices and an improved AlexNet-SVM method. Optik **1**(212), 164237 (2020)

Sharma, R., Goel, T., Tanveer, M., Dwivedi, S., Murugan, R.: FAF-DRVFL: fuzzy activation function based deep random vector functional links network for early diagnosis of Alzheimer disease. Appl. Soft Comput. **1**(106), 107371 (2021)

Stirling, J., Chen, T., Bucholc, M.: Diagnosing Alzheimer's disease using a self-organising fuzzy classifier. In: Carter, J., Chiclana, F., Singh Khuman, A., Chen, T. (eds.) Fuzzy Logic, pp. 69–82. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-66474-9_5

Vaithinathan, K., Parthiban, L.: Alzheimer's Disease Neuroimaging Initiative. A novel texture extraction technique with T1 weighted MRI for the classification of Alzheimer's disease. J. Neurosci. Methods **318**, 84–99 (2019)

Vijay, V., Kavitha, A.R., Rebecca, S.R.: Automated brain tumor segmentation and detection in MRI using enhanced Darwinian particle swarm optimization (EDPSO). Procedia Comput. Sci. **1**(92), 475–480 (2016)

Zeng, N., Qiu, H., Wang, Z., Liu, W., Zhang, H., Li, Y.: A new switching-delayed-PSO-based optimized SVM algorithm for diagnosis of Alzheimer's disease. Neurocomputing **3**(320), 195–202 (2018)

Zhang, J., Gao, Y., Gao, Y., Munsell, B.C., Shen, D.: Detecting anatomical landmarks for fast Alzheimer's disease diagnosis. IEEE Trans. Med. Imaging **35**(12), 2524–2533 (2016)

AI Powered Smart Glasses for Visually Impaired Individuals

Ketki Kshirsagar(✉), Samarth Chikane, Dev Desai, Avdhut Hande, Aniruddha Deobhankar, Arun Govind, Shubham Derkar, and Arjun Gupta

Vishwakarma Institute of Technology, Pune, Maharashtra, India
{ketki.kshirsagar,samarth.chikane24,dev.desai24,avdhut.hande24, aniruddha.deobhankar24,arun.govind24,shubham.derkar24, arjun.gupta241}@vit.edu

Abstract. The one of the most important sensory organ of our body is an eye, it is the reason why people can enjoy its beautiful surroundings. What if we would not have this important organ?, the answer is quite obvious, it would be very challenging, he would be isolated. There are millions of people across the globe living such miserable lives. So to overcome this challenge we come up with an idea to build an assistive aid for blind needy. The project is AI-Powered Smart glasses for visually impaired. This basically notifies the blind person about the obstacle in front of him/her. This tool has the ability to tell the user the distance of the obstacle and what particularly the obstacle is like the tree is 60 cm away. This paper includes the brief information about how this glasses work. The glasses are full advanced technical tools like ultrasonic sensors, IR sensors, ESP32 microcontroller, ESP32 cam module, earpiece for user's enhanced listening. The software we used is YOLOv5 for object recognition, Tensor Flow Lite, text-to-speech software, MATLAB, etc.

Keywords: ESP32 · Ultrasonic sensors · Object recognition · Obstacle detection

1 Introduction

Our proposed system uses AI-integrated smart glasses equipped with object detection and obstacle sensing capabilities. By providing real-time audio feedback, these glasses will improve the mobility, safety, and self-reliance of visually impaired users [1]. The project is mainly focused on to help the blind people by creating a cost-friendly device that helps the them to indirectly see the world by making them aware about the upcoming obstacle in advance. It is a light-weight, portable device. Our proposed system uses AI-integrated smart glasses equipped with object detection and obstacle sensing capabilities. By powering on, the system initializes all the components(like camera, sensors, etc.). Then the ESP 32 camera captures frames, which are then processed by the AI model(YOLO) to detect objects [2]. At the same time, the ultrasonic sensors monitor the environment for obstacles. The onboard AI detects objects (e.g., a chair, door, or person) and processes the results. Audio feedback is generated (e.g., "Chair detected at 2 m") and played through the speaker. If an obstacle is detected within a certain range by

S. Fong et al. (Eds.): ICT4SD 2025, LNNS 1651, pp. 58–67, 2026.
https://doi.org/10.1007/978-3-032-06688-6_6

ultrasonic sensor, the system will either alert the user with audio feedback or vibrate (if the vibration motor is enabled) [3]. The system continuously updates as the user moves, ensuring real-time situational awareness. The user can speak commands to interact with the system (e.g., "Start," "Stop," or "Calibrate").The system processes these commands and adjusts its be accordingly.

2 Literature Review

Savita Channagoudar et al. have reported the study that is based on a system created to enable more autonomous movement around for people with visual impairments. It uses a camera to capture real-time images and processes them with deep learning algorithms to identify and classify objects [3]. By combining accuracy, portability, and real-time functionality, the system has the capacity to significantly improve the lives of those who are blind or visually impaired. The authors suggest future improvements, including better energy efficiency, broader object recognition, and more natural, multilingual voice feedback.

Barontini.F. et al. presented a creative indoor navigation system meant to improve the safety and autonomy of visually challenged people [4]. It is a processing unit that analyses visual data to identify obstacles, and a wearable haptic device that delivers both normal and tangential force cues to guide users. This approach ensures that the system aligns with the actual needs and preferences of end-users [5]. Experimental evaluations, including tests with visually impaired users, indicate that the system effectively supports indoor navigation and serves as the valuable tool for training individuals in the context of assistive travel tools.

Ashvini Hirve represents the topic Vision Safe Using ESP32-CAM it explores the integration of the ESP32-CAM—a development board that combines the ESP32 microcontroller [6]. The ESP32-CAM is highlighted as a cost-effective solution for Wi-Fi-enabled cameras, featuring a powerful 32-bit microcontroller and microSD card support, making it suitable for various IoT projects.

M. Babiuch et al. demonstrates the topic, it explores the ESP32 microcontroller application in data processing tasks [9]. The authors discuss the microcontroller's capabilities in handling various data processing functions, emphasising its utility in measurement and control systems. The study highlights the ESP32's features, such as integrated Wi-Fi and Bluetooth connectivity, which make it suitable for Internet of Things (IoT) applications.

Deepa J et al. represents the study focuses on developing a device to help visually impaired people move around their surroundings by sensing obstacles. The system uses ultrasonic sensors to calculate distances, allowing accurate calculation of the minimum width's distance to obstacles [10].

This all papers conclude and summaries how the sensors(ultrasonic, IR), ESP32 cam module and AI help building a gaming changing glasses for visually impaired.

3 Methodology

3.1 Components

Ultrasonic Sensor (HC-SR04) as Shown in Fig. 1
It is a device which measures the gap between the object and itself. This sensor is based on he principle of echolocation. The transmitter sends the signal to the object and it bounces back as it hits the object and that bounced back signal is received by receiver.

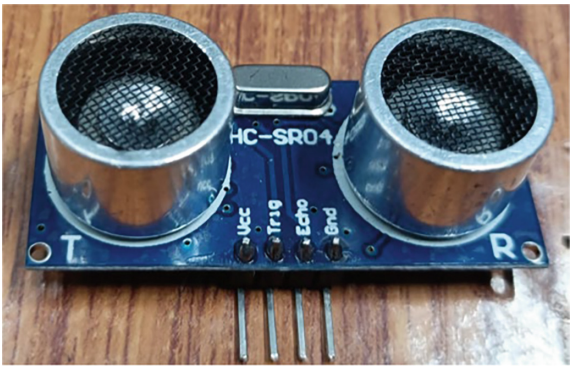


Fig. 1. Ultrasonic Sensor

- It uses the formula of:

$$\text{Distance} = (\text{speed of sound} * \text{time})/2.$$

- This sensor is mounted on the centre of the glasses.

ESP-32 CAM as Shown in Fig. 2
ESP32 Cam module is a device that is used here for AI based object recognition. The identification of a separate image is referred to as photograph clustering, while the identification of multiple images that contains objects is referred to as object tracking.ESP32 CAM Based Object Detection & Identification with OpenCV, that can be utilised in a wide range of scenarios [11].

- The cam captures images and arranges it in grid like pattern.
- This method helps the AI to identify and classify the object like(stairs, wall, pole).

ESP32-CAM AI-Thinker Development Board as Shown in Fig. 3
The ESP32-CAM AI-Thinker development board can be programmed using Arduino IDE.The ESP32-CAM AI-Thinker module is a development board of ESP32 featuring an OV2640 camera, microSD card capability, flash lamp on-board and some GPIOs for connecting peripherals. But it does not come with a built-in programmer. You'll require an FTDI programmer to plug it into your computer and transfer code [12].

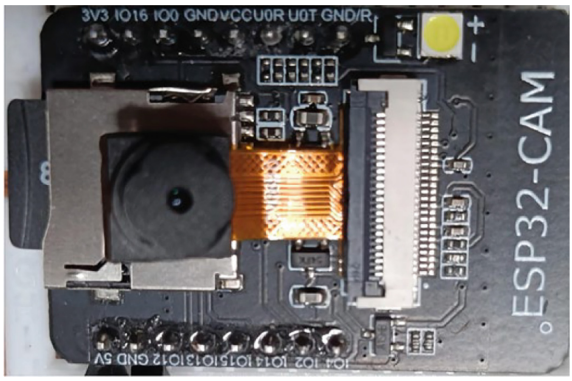


Fig. 2. ESP32-CAM

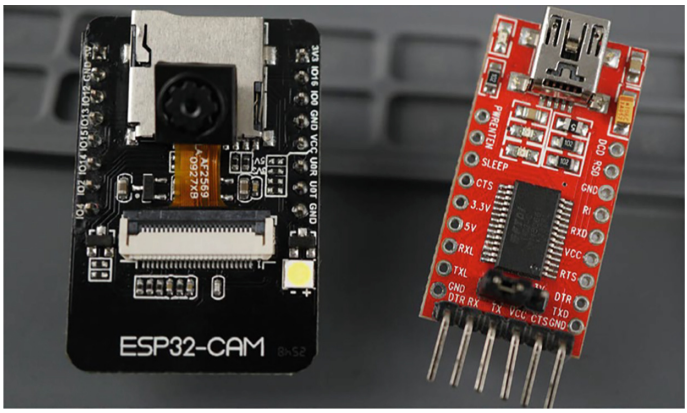


Fig. 3. ESP32-CAM AI-Thinker development board

3.2 Workflow

Initialization

Upon powering on, the system initializes all components (camera, sensors, AI model, etc.).The user can give a voice command to start the system. Proposed system block diagram shown in Fig. 4.

Real-Time Processing

- The camera captures frames, which are then processed by the AI model to detect objects.
- At the same time, the ultrasonic sensors monitor the environment for obstacles.

Object Detection and Feedback

- The system detects objects (e.g., a chair, door, or person) and processes the results[7].
- Audio feedback is generated (e.g., "Chair detected at 2 m") and played through the speaker.

Obstacle Avoidance

- If an obstacle is detected within a certain range, the system will either alert the user with audio feedback or vibrate (if the vibration motor is enabled).
- The system continuously updates as the user moves, ensuring real-time situational awareness[8].

Voice Commands

- The user can speak commands to interact with the system (e.g., "Start," "Stop," or "Calibrate").
- The system processes these commands and adjusts its behaviour accordingly.

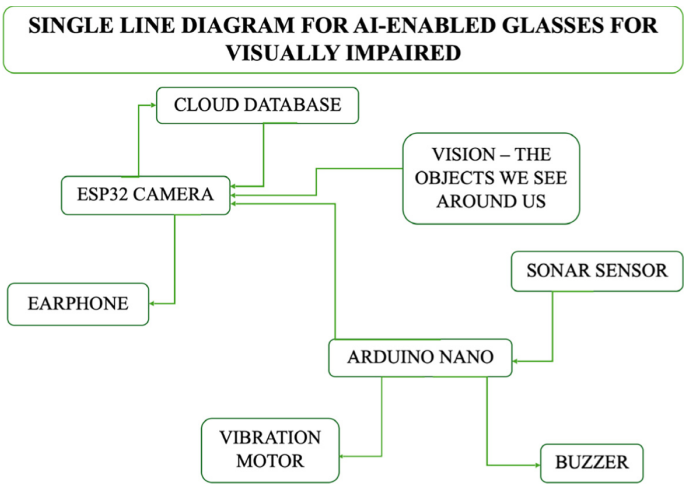


Fig. 4. Block Diagram

The AI based obstacle, detecting glasses is the integration of sensor technology, micro controller, processing, and real-time feedback system. The core components of the system are ultrasonic sensor, ESP32 cam module, these function by detecting the distance of obstacle from the user and identifying exactly what object it is, respectively.

The ultrasonic sensor is connected to Arduino Nano and ESP32 cam module is connected to ESP 32 micro-controller. The nano board programmed in a way that ultrasonic sensor measures the distance of obstacle. Similarly, ESP32 micro-controller is programmed search that camera is used for object recognition through AI.

As the system initialise, the ultrasonic sensor measures the distance of the obstacle simultaneously cam, identify what object it is. Then the data from ultrasonic sensor is fed as input to Arduino Nano similar. Similarly, data from Cam is filled as input to ESP 32 microcontroller. This all process is very fast, and then the output from these microcontroller board is to feedback system from there. The user is notified with the vibrational motor and earpiece in his ear.

The system design process involves careful alignment in order to balance the weight of the components on both side of the frame.

The methodology effectively integrates microcontroller, camera module, ultrasonic sensor, speaker, microphone, and battery components, along with free software tools, to develop a cost-effective and robust prototype. With an estimated cost ranging between ₹2000–₹3000, this approach demonstrates a practical balance between functionality and affordability.

4 Results and Discussions

Figure 5 shows working of HC-SR04 and Fig. 6 shows the math used to find out the relationship between trigger and echo by using the formula speed = distance/time. For example, the obstacle and the ultrasonic sensor is 10 cm away from each other and speed of sound in air is 340m/s OR 0.034 cm/μs. Sound wave require about 294μs to travel to object.

The ultrasonic sensor utilized in our project features a minimum detection range of 2 cm and a maximum effective range of 75 cm. Within this span, the system provides varying levels of vibration feedback based on the detected distance. When an obstacle is detected within the range of 2–35 cm, strong vibration feedback is generated to signal close proximity. As the obstacle moves farther, within the range of 40–75 cm, the vibration intensity gradually decreases. If the object is beyond 75 cm, it is considered out of range, and no vibration is triggered.

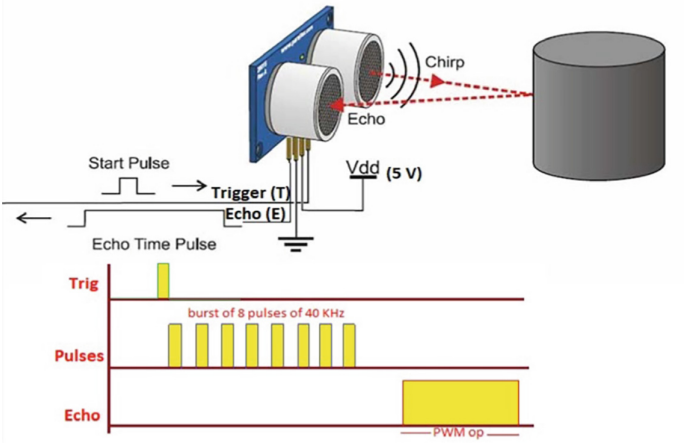


Fig. 5. Schematic Diagram of working of HC-SR04

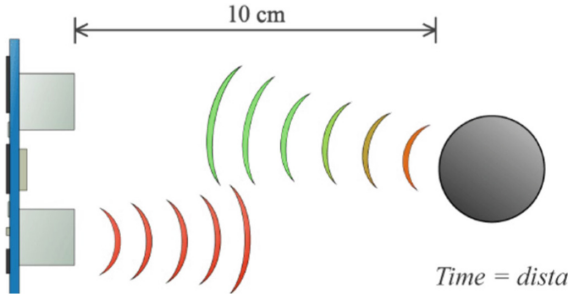


Fig. 6. Math used for Ultrasonic Sensor

The sensor's performance is significantly influenced by environmental conditions. Indoors, accuracy is generally higher due to the controlled setting, where stable temperatures and the absence of wind or irregular surfaces contribute to reliable distance measurements. However, in outdoor environments, accuracy can be compromised by several factors. Wind may scatter or deflect the ultrasonic waves, resulting in inconsistent readings. Irregular surfaces such as bushes or uneven walls can reflect the sound waves unpredictably. Additionally, fluctuations in temperature and humidity affect the speed at which sound travels, thereby impacting the accuracy of distance calculations.

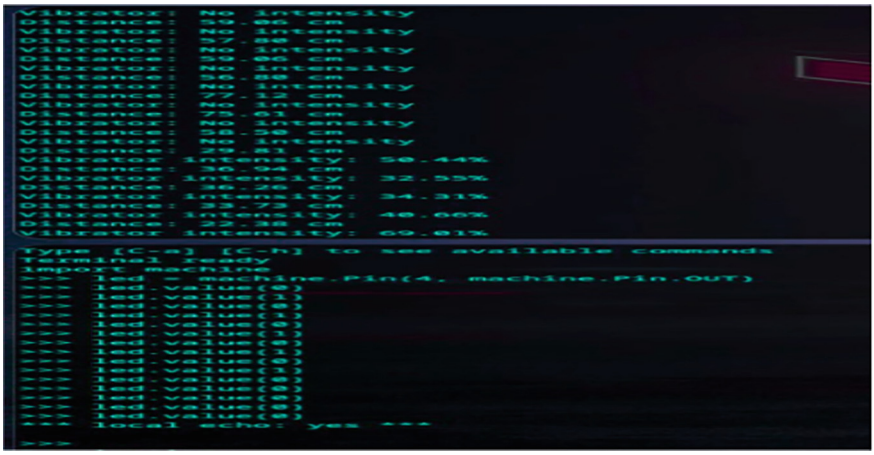


Fig. 7. Variation in intensity of vibration with change in distance

As seen in Figs. 7 and 8, the intensity of vibration is very high in the range of 2–35 cm and the intensity of vibration gradually decreases with increase in distance of obstacle from ultrasonic sensor, the range can be from 40–75 cm. The range of distance, except these values are considered as out of range and no vibration occurs. Figure 9 shows complete proposed experimented model.

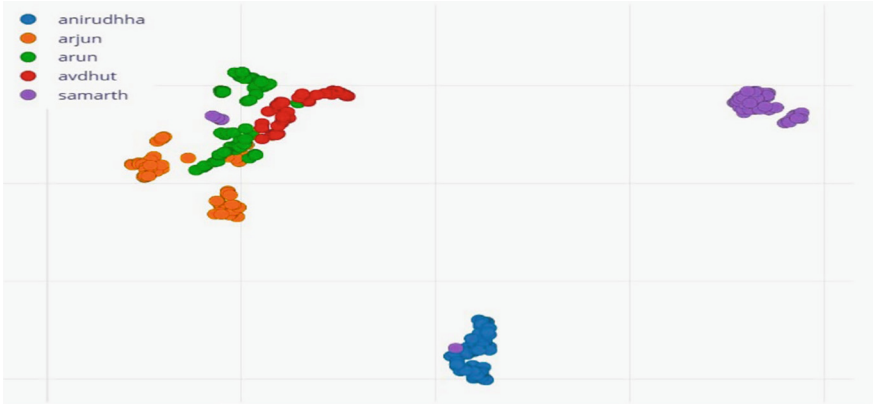


Fig. 8. Intensity distribution chart

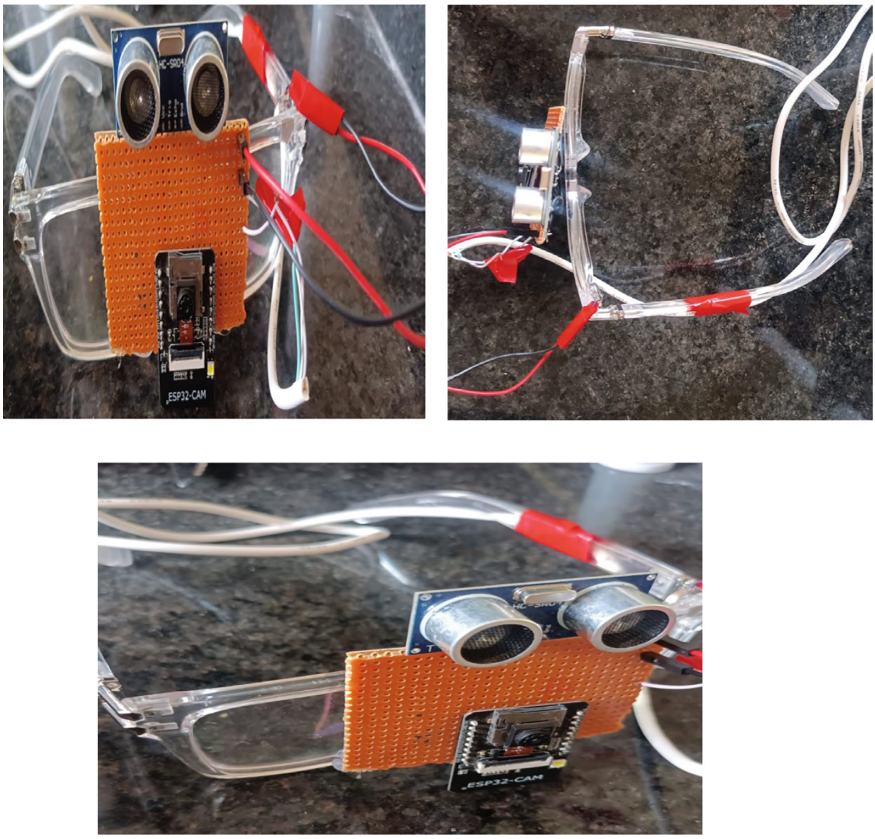


Fig. 9. Proposed model

5 Conclusion

The Obstacle Detecting Glasses project successfully demonstrates the use of ultrasonic sensors and an ESP32 camera module to help visually impaired people move around their surroundings by sensing obstacles. The system effectively detects obstacles within a predefined range and provides feedback through vibrations or audio alerts. The ESP32 camera enhances the project by allowing real-time image processing, which can be further integrated with AI-based object recognition for better navigation assistance.

The project proves to be a cost-effective and portable solution for obstacle detection, offering a potential alternative to traditional assistive technologies like white canes. However, there are areas for improvement, such as enhancing sensor accuracy in detecting smaller or transparent obstacles, reducing false alarms, and integrating more advanced AI-based object recognition for better situational awareness. Future developments could include voice assistance, GPS navigation, and machine learning for adaptive learning based on user behavior.

In general, the progress made by the device is indicative of advances in assistive technology for people with disabilities and shows promise for creating better mobility and independence for those who cannot see.

References

1. Gollagi, S.G., Bamane, K.D., Patil, D.M., Ankali, S.B., Akiwate, B.M.: An innovative smart glass for blind people using artificial intelligence. Indonesian J. Electr. Eng. Comput. Sci. (2023). https://doi.org/10.11591/ijeecs.v31.il.pp433-439
2. Kshirsagar, K. P., Tathod, A., Dudhate, R., Yadav, A., Tarte, R.: Camera calibration using robust intrinsic and extrinsic parameters. In: Artificial Intelligence in Information and Communication Technologies, Healthcare and Education, pp. 109–119. Chapman and Hall/CRC (2022)
3. Narwaria, R.P., Ahirwar, A., Prajapati, A.K., Kumar, A., Tiwari, A.K.: Smart Object Detection Using ESP32-CAM Based on YOLO Algorithm (2024). https://doi.org/10.1109/ICoICI62503.2024.10696374
4. Channagoudar, S., Prakash, M.B., Harish, H.M., Udara, S.: Deep learning-based object detection for assisting visually impaired people. NeuroQuantology (2022). https://doi.org/10.48047/nq.2022.20.5.nq22835
5. Barontini, F., Catalano, M.G., Pallottino, L., Leporini, B., Bianchi, M.: Integrating wearable haptics and obstacle avoidance for the visually impaired in indoor navigation: a user-centered approach. IEEE Trans. Haptics (2020). https://doi.org/10.1109/toh.2020.2996748
6. Hirve, A.: Vision Safe Using ESP32-CAM, December 2023, Hirve, Ashvini, Vision Safe Using ESP32-CAM (2023). https://doi.org/10.2139/ssrn.4867417
7. Kshirsagar, K.P., Rokade, R.A.S.: Depth estimation using stereo medical imaging. In: Robotics and Automation in Healthcare, pp. 61–74. Apple Academic Press (2024)
8. Adhau, J., Adsare, V., Jadhav, D., Kshirsagar, K., Tamkhade, J.: IoT-based face recognition with SNN. In: 2nd International Conference on Futuristic Technologies (INCOFT), pp. 1–6 (2023)
9. Babiuch, M., Foltýnek, P., Smutný, P.: Using the ESP32 Microcontroller for Data Processing (2019). https://doi.org/10.1109/CarpathianCC.2019.8765944
10. Deepa, J., Maria Adeline, P., Sai Madhumita, S.S., Pavalaselvi, N.: Obstacle detection and navigation for the visually impaired. IEEE Explore (2023). https://doi.org/10.1109/ICACCS57279.2023.10112994
11. Shofia Priya Dharshini, D., Saranya, R., Sneha, S.: Esp32 cam based object detection & Identification with opencv. Data Analytics Artif. Intell. **2**(4) (2022)
12. Subashree, S., Akila, T., Dwaramwar, P.A., Chandra, S., Kshirsagar, K.P.: AI-Driven Energy Optimization in High-Performance Computing: Smart Solutions for Sustainable Efficiency, Integrating Machine Learning Into HPC-Based Simulations and Analytics, pp. 277–302. IGI Global Scientific Publishing (2025)

NutriScan: A Python-Based Barcode Scanner for Ingredient Analysis and Personalized Health Warnings

Yash Chavan[1(✉)], Arnav Sonawane[1], Arpit Pattiwar[1], Aditya Nagdive[1], and Kaushalya Thopate[2]

[1] Vishwakarma Institute of Technology, Pune 411037, Maharashtra, India
yashchavan121212@gmail.com

[2] Department of Computer Engineering, Vishwakarma Institute of Technology, Pune, Maharashtra, India

Abstract. In today's fast-moving world, consumers rely on packaged foods. This is extremely important for easy access to detailed and personalized nutritional information. This project focuses on developing mobile applications for barcode scanning. This includes extensive food details, including ingredients, nutritional value, allergen warnings, and personalized consumption recommendations based on a person's health. Applications written with Python and Kivy provide a seamless user experience, allowing individuals to scan barcodes and upload images of ingredients to extract and analyze related information. Additionally, it includes optical character detection (OCR) using Tesseract, which extracts text from photos to allow users to analyze the ingredient list and nutritional name, even if barcode scans are not possible. By taking into account user nutritional limitations or illnesses such as diabetes, lactose intolerance, or gluten sensitivity, this application provides tailor-made health advice and helps individuals make found food decisions appropriately. A secure user authentication system improves the experience by storing your preferences and receiving recommendations created by tailors. The main goal of this project is to enable consumers to choose food in real time and promote healthier consumption habits. The combination of barcode scanning, OCR, and a structured database causes applications to close the gap between the complexity of food indicators and user understanding. Future improvements include mechanically learning-based ingredients, integration into real-time product databases, and expansion of several platforms beyond Android. This initiative represents an important step in using technology to improve consumer health awareness and ensure safer and sounder decisions for food consumption.

Keywords: Barcodes - skan · Nutrition label analysis · Allergen recognition · Python · Kivy · OCR · Health aldvice · Nutrition limitations · Personalized recommendations

S. Fong et al. (Eds.): ICT4SD 2025, LNNS 1651, pp. 68–82, 2026.
https://doi.org/10.1007/978-3-032-06688-6_7

1 Introduction

The growing dependency on industrially packaged processed foods has made it necessary for consumers to gain access to clear and consistent information on nutritional content. Food labeling provides access to ingredient composition, nutritional content, and allergen information; however, many find it difficult to interpret information due to time constraints, lack of information, and the availability of complex jargon. This is most evident among consumers with dietary needs, such as those with diabetes who need to carefully limit sugar intake, those with celiac disease who need to avoid gluten completely, and those who have potentially life-threatening food allergies who need to avoid certain ingredients in full. Labels may include obscure jargon and hidden allergens, making it difficult for consumers to ascertain the safety of food products. To address this issue, the Nutriscan project offers mobile apps that make use of artificial intelligence, using barcode scanning and optical character recognition (OCR) to identify and analyze information related to food. Through the elimination of human intervention, Nutriscan provides users with precise and personalized health advice while making it easy to make informed and safe food choices.

1.1 Problem Statement

Although food labeling is informative, consumers with dietary needs usually have to fight the inefficiency of reading and interpreting ingredient lists manually. Most processed food products use complicated language and employ irregular forms of ingredients or concealed allergens, which makes it hard for consumers to ascertain the appropriateness of their products. Nutriscan was created to eradicate this issue by employing barcode scanning and OCR-based text recognition. This enables users to call forth structured nutritional information and health consultation directly relevant to their individual nutritional requirements.

1.2 Objective

Nutriscan will build mobile apps using Barcode and OCR technology to read food labels automatically. By introducing a JSON-based structured database, your app initiates real-time calls of the ingredients of the product, nutrition facts and allergen alerts. With a dynamic user profile management system, the app offers conditions like diabetes and gluten intolerance, offers dietary suggestions, and translates food selections into an educated, personalized experience.

1.3 Hypothesis/Expected Results

The future development of Nutriscan will most probably improve consumers' decision-making by providing quick and accurate access to food information. The app reduces the risk of accidental ingestion of harmful chemicals by providing clear and personalized health advice. Barcode scanning ease makes it easy to access nutritional information and significantly reduces the time taken to establish product suitability. Additionally, the OCR feature makes it easy and possible to access text information from labels where barcodes are not present. As a result, consumers can gain a better insight into the nutritional content of their food and therefore make better and safer food choices. In the long term, the app can be far more effective by providing a wide range of product categories and utilizing machine learning to provide more advanced analysis of ingredients and personalized advice.

2 Literature Work

Kivy Framework for Application Development

Bhoyarkar et al. (2019) [16] Explains Kivy, a cross-platform Python library for creating applications with an interactive user interface. Ideal for mobile app development with Kivy support for multi-platform and features such as custom widgets, solid APIs, and KV language for declarative UI building. Aligns with the application of Kivy and Kivymd throughout the project and offers a reactive – visually pleasing interface.

Cross-Platform Mobile Development Frameworks

Kumar and Rani (2020) [1] compares the Cross-Platform Development Framework, which highlights Kivy's flexibility and user-friendly. Kumar and Singh (2023) [3] goes on to discuss the advantages of Python-based frameworks for interactive and educational applications in augmenting projects using Kivy for easy-to-use development environments.

Performance Analysis of Python-Based Mobile Frameworks

Gupa and Sinha (2021) [5] analyze the efficiency of Python-based framework conditions in mobile development and highlight the power of Kivy, interactive and simple applications. This study addresses projects for projects that use Kivy for a quick and efficient barcode scanning experience.

Barcode Detection Using OpenCV-Python

Puri and Jain (2020) [2] Investigating barcode detection using OpenCV Python, detailed morphological manipulation, and Scharhar Gradient calculations. Their results increase the efficiency of barcode scanning in real-time applications related to barcode recognition functions in projects using Zbarcam.

Real-Time Image Processing and Barcode Detection

Das and Kaur (2021) [14] focuses on real-time image processing techniques, including erosion and expansion, to improve barcode detection. Zhang and Liu (2022) [6] examine deep learning models of barcode recognition in challenging environments and highlight the importance of efficient image processing. These studies support the use of OCR and real-time barcode scanning for accurate product identification by the project.

Integration of Computer Vision Libraries

Liu and Wang (2023) [13] Examining computer vision libraries such as OpenCV and discussing integration into Python frameworks for barcode recognition. Their results highlight the effectiveness of a combination of several image processing techniques that can be used directly to use PBOR for real-time scanning of projects.

Comparative Analysis of Barcode Detection Methods

Sharma and Yadav (2023) [10] provides a comparative test of barcode recognition methods in mobile applications. This highlights the benefits of Python-based frameworks such as Kivy. Their work supports the project mechanism and feedback system of the project, ensuring robust barcode detection.

Robust Barcode Detection Algorithms

Patel and Roy (2023) [7] examine barcode detection of curved surfaces of interest in improving barcode scans in real-world conditions. Her research highlights how important it is to combine several image processing techniques that can improve the ability of a project to recognize barcodes under a variety of conditions.

Real-Time QR Code Detection

Silva and Gomes (2020) [9] analyzes QR code detection on embedded devices, highlighting real-time processing of a seamless user experience. Your research will enhance your selection of PBARCAM projects for barcode scanning and ensure efficient QR and barcode recognition in your application.

Machine Learning-Based Barcode Recognition

Lee and Park (2021) [8] explores models to present damaged barcode recognition in machine learning and potential improvements in barcode readability. This corresponds to the future scope of the project to integrate AI-controlled product recommendations and improved barcode recognition accuracy.

Mobile-Based Inventory Management Systems

Akhtar and Raza (2021) [11] discusses the Python framework in inventory management and focuses on scan barcodes for product tracking. Your research reveals the efficiency of JSON-based memory that matches the use of Light JSON date databases in product details.

Cross-Platform Barcode Scanning Applications

Patel and Chaudhary (2022) [12] performs cross-platform barcode scanning applications and comparative research, highlighting user-friendly interfaces and efficient data calls. Your results support the decision of projects that use KIVYMD for UI design and use JSON easily for data storage.

Mobile Frameworks for Inventory and Barcode Management

Singh and Verma (2022) [15] analyses the mobile framework of barcode-based inventory management, highlighting scalability and efficiency. Your research will confirm the use of JSON as an optical data storage solution for your project to expand into a more robust database system in future iterations.

IoT Interfaces Using Kivy

Memon and Ahmed (2022) [4] discuss the use of Kivy for IoT applications, highlighting the ability to handle a variety of input modalities and network operations. This could be related to future expansions of Nutriscan, allowing integration into IoT-based food persecution systems.

Real-Time Applications Using Kivy

Memon and Ahmed (2022) [4] also consider the use of Kivy in mobile real-time applications, such as data processing and acceptance rate speed. Your research is targeting real-time barcode scanning projects and effective data presentations, providing a seamless user experience.

3 Methodology

The concept of creating Nutriscan was initiated due to the need for an easy-to-use barcode reader software that would provide real-time nutritional analysis, allergen alerts, and personalized health advice. The system is a combination of barcode reading, optical character recognition, and logically structured database, all combined within mobile apps created with Python and Kivy. All components of the system have been selected to enable rapid and effective processing while at the same time delivering a smooth user interface.

3.1 System Architecture

The system operates through a pre-set workflow in which users read product barcodes or upload images of ingredient labels. The barcodes or text obtained are processed through the Zbarcam barcode scanner or the Tesseract OCR engine. The operation retrieves pertinent information from a JSON-formatted product database. The system then processes the data, cross-references it with stored allergens and health conditions, and provides personalized health recommendations (Fig. 1).

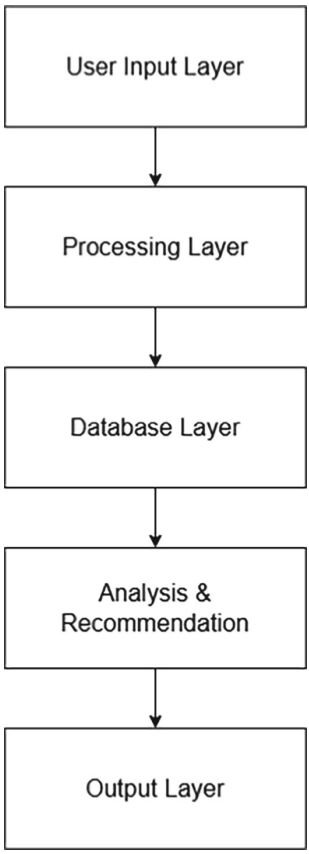


Fig. 1. System Architecture

3.2 Application Development Framework

The application was developed with Kivy, an open-source Python cross-platform framework for developing mobile applications. The native flexibility of Kivy allows for seamless integration of functionalities such as barcode scanning through the camera, real-time text recognition, and an optimally organized product database. For an improved user interface, KIVYMD was employed, offering a contemporary and stunning appearance that adheres to Google's material design guidelines.

The interface includes three primary screens:

1. Login Screen – Allows users to authenticate and store personalized health conditions.
2. Scan Screen – Enables barcode scanning and image uploads for nutritional analysis.
3. Result Screen – Displays extracted ingredient details, allergens, nutritional values, and health advisories.

3.3 Barcode Recognition and Processing

Barcode scanning is the main way of obtaining product information. The ZBarCam library was included for real-time barcode scanning with the camera of the device. Support for 1D and 2D barcodes, such as EAN-13, Code 128, QR Codes, and Data Matrix format, has been included to support a variety of food products (Table 1).

$$Scanning\ Efficiency = \left(\frac{Successfully\ Scanned\ Barcodes}{Total\ Attempts} \right) \times 100 \quad (1)$$

Table 1. Supported Barcode Types in Nutriscan

Barcode Type	Format	Application
EAN-13	1D	Retail Products
QR Code	2D	General Information
Code 128	1D	Logistics & Warehousing
Data Matrix	2D	Medical & Manufacturing

When the barcode is read, the product code is retrieved and matched against the JSON database. Upon matching, the product name, breakdown of the ingredient list, nutritional facts, and health warnings are retrieved. If there is no match, the system asks the user to photograph the ingredient list for optical character recognition processing.

3.4 OCR-Based Ingredient Extraction

Where barcode scanning is not an option, NutriScan allows uploading of product label photos. The Tesseract OCR engine, incorporated via Pytesseract, extracts the text from the photo and identifies ingredient lists, nutritional information, and allergy notices.

$$OCR\ Accuracy = \left(\frac{Correctly\ Extracted\ Words}{Total\ Words\ in\ Label} \right) \times 100 \quad (2)$$

The downloaded text is then scanned to detect basic nutritional factors like sugars, fats, additives, and allergens. The extracted information is compared by the system with the diet restrictions set by the user and provides individualized alerts for allergens including gluten, milk, nuts, and artificial additives.

3.5 Database Design and Query Execution

The application relies on a JSON-based product database, where each product entry contains:

Product Name.

- Barcode Number
- List of Ingredients
- Nutritional Values (per 100ml or per serving)
- Allergen Warnings
- Disease-Specific Recommendations (e.g., for diabetics or individuals with hypertension) (Fig. 2)

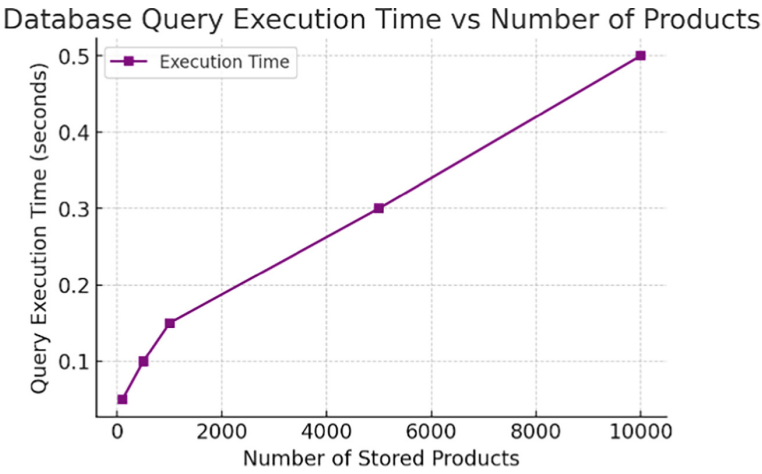


Fig. 2. Database Query Execution Time vs Number of Products

When I scan a barcode, the system searches in the JSON file for a comparable product. When the product is discovered, nutritional information is shown on the Result Screen. When the product is not discovered, manual text recognition via OCR is provided by the system.

For scalability, the database can be scaled to handle larger data sets. Future releases of NutriScan can utilize Firebase or MongoDB to provide real-time updates and API-based product verification.

3.6 User Authentication and Personalized Health Advisory

An authentication system for the user is implemented to facilitate individualized tracking of food intake. Users are required to create an account in which they can identify their dietary restrictions, allergies, and chronic diseases. When scanning a product, the system cross-references the user's profile with the nutrition information and yields:

- Real-time warnings for allergens
- Health advisories based on medical conditions
- Consumption recommendations tailored to dietary needs (Fig. 3)

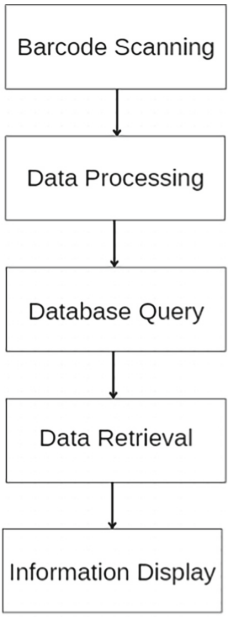


Fig. 3. Shows the flowchart of the working

By allowing users to store and retrieve their dietary preferences, NutriScan ensures highly personalized recommendations, making it an essential tool for individuals with strict health requirements.

3.7 Data Processing and Error Handling

NutriScan implements error detection mechanisms to handle unrecognized barcodes, poor image quality, and incomplete database entries. When an error occurs, the system:

- Prompts the user to retry scanning.
- Suggests manual image upload for OCR processing.
- Displays an informative message if the product is unavailable.

$$P_{match} = \frac{N_{matched}}{N_{total}} \tag{1}$$

This error-handling strategy ensures a smooth user experience while reducing instances of incorrect or missing product information.

4 Experimental Results

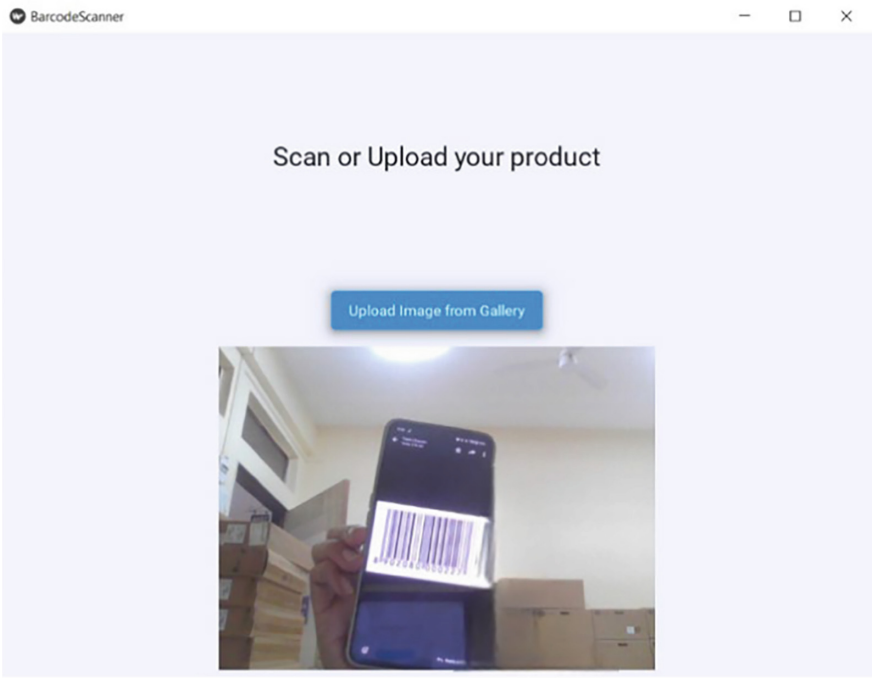


Fig. 4. Scanning Interface

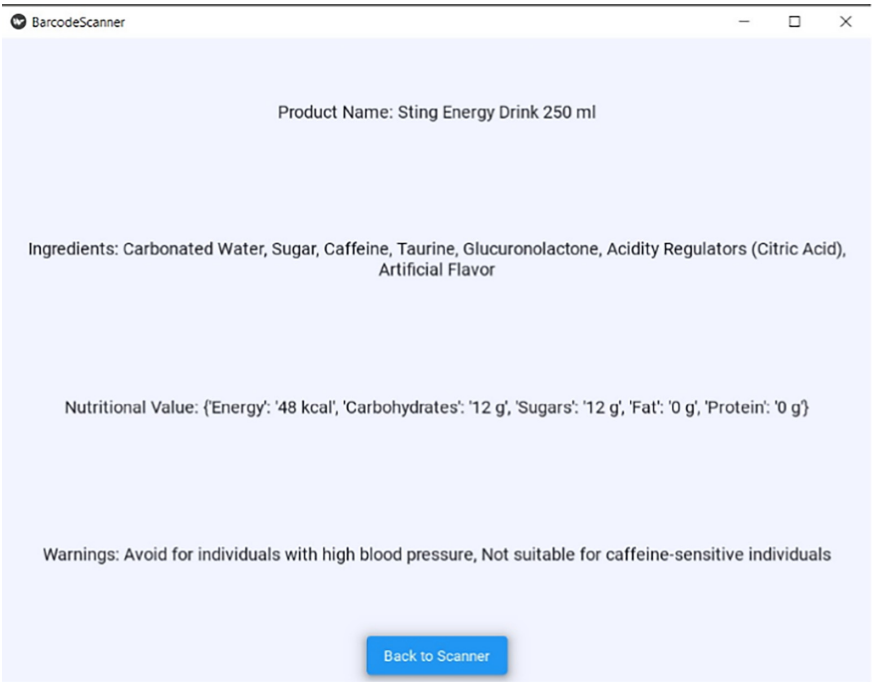


Fig. 5. Output Window

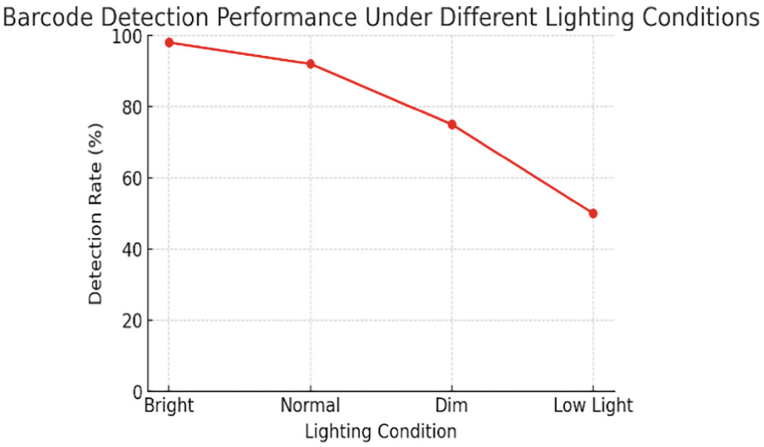


Fig. 6. Barcode Detection Under Deifferent Lighting Conditions

The combination of the Nutriscan Barcode Scanning app achieves the aim of making detailed product information available to consumers, including ingredients, nutritional information, allergen notifications, and personal health advice. Based on ZBAR libraries, the app features accurate real-time barcode scanning, allowing users to gain access to product information without needing to enter information manually. Product information is efficiently stored and retrieved with a JSON-based database, providing speedy and accurate answers to inquiries by users. This integrated combination of barcode scanning and database query is a core function of Nutriscan and makes it an indispensable tool for users with health conditions such as dietary needs, diabetes, gluten intolerance, or life-threatening allergies (Figs. 4, 5 and 6).

Where barcodes cannot be scanned, Nutriscan employs optical character recognition (OCR) driven by Pytesseract. It allows users to scan an image of the product label. The application is able to scan and read the text to provide a complete list of ingredients and nutritional information. The feature is especially useful for products that lack barcodes or have damaged labels, thus enhancing the usability and versatility of the application. The intuitive interface built with KivyMD offers a smooth and intuitive experience, which makes it convenient for users to gain access to critical information. When an error occurs, the app provides clear and helpful feedback to improve user satisfaction and reliability. Additionally, the app offers personalized health warnings tailored to your individual nutritional needs. B. Warning of high sugar content for diabetic or gluten presence in patients with celiac disease. These features allow users to make appropriate found and safe food decisions and align with the app's mission to promote healthier consumption habits.

Overall, Nutriscan creates practical, user-oriented uses by demonstrating the feasibility and benefits of barcode scanning, OCR, and structured database integration. The design of the app allows for future scalability. B. Includes AI control recommendations for product database expansion, real-time update integration via APIs, and healthier alternatives. These potential improvements continue to integrate dietary supplements as a multi-purpose and essential tool for those who want to make appropriate and healthy food decisions (Table 2, Fig. 7).

Table 2. Accuracy Metrics of Barcode Scanning and OCR

Method	Accuracy (%)
Barcode Scanning	95
OCR Text Extraction	85

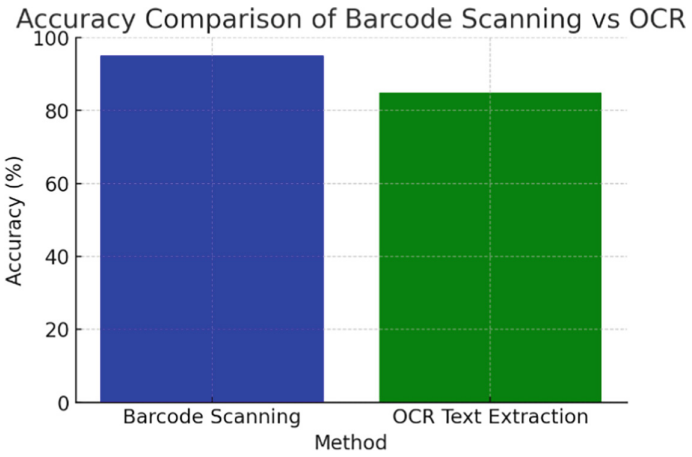


Fig. 7. Accuracy Comparison of Barcode Scanning vs OCR

5 Future Scope

5.1 Product Database Enhancement

One of the most important areas for enlargement is having a larger pool of items in the database. As it is, the JSON file is a simple and efficient way in which to store information about items. But a transition towards a more reliable form of database technology, such as NoSQL or SQL, will become significant with a larger app. By having items in additional sources in stores in regions and countries, the app will become larger and more flexible and will serve a larger community.

5.2 Integration with Real-Time Updates via APIs

Integration with producers' databases or with third-party APIs, such as OpenFoodFacts, can make real-time information about ingredients, allergies, and nutritional values, such that information is ever updated and correct. That will make the app even more reliable and less maintenance-intensive.

5.3 Personal Health Profiles

The app can have personalized profiles in future app releases. Users can enter specific nutritional requirements, wishes, or long-term medical concerns (like an allergy, a sugar issue, or a disease such as diabetes, for instance). According to profiles, then, the app can issue personalized tips and warnings, such that information is even more relevant and easier to use.

5.4 Broadening Beyond Foods

The software can extend to additional categories, such as cosmetics, household items, and drugs, even when its current target is foods. Users can scan cosmetics, for instance, for deadly chemicals or allergies. Because of its generalizability, then, the app will serve a one-stop source for concerned users in a variety of domains.

5.5 AI Driven Recommendations

The app can issue recommendations through AI in future app releases. By using algorithms, therefore, such an app will then present healthy alternatives with items that have been scanned. For instance, a high sugar scan will have alternative but healthy options in its database suggested through the app for a diabetic patient. In contrast with offering information in a passive manner, such an option will make an app proactive and inform them with actionable information.

6 Conclusion

This project successfully demonstrates the development of NutriScan, a barcode-scanning mobile application designed to enhance food label accessibility and provide personalized health advisories. By integrating ZBarCam for barcode scanning, Pytesseract for OCR-based text extraction, and a structured JSON database for product storage, the application enables users to retrieve detailed ingredient lists, nutritional values, allergen warnings, and tailored dietary recommendations instantly. With Kivymd, you are guaranteed a fast, user-friendly interface that allows for seamless navigation and interaction. Not only does this program simplify scanning and understanding food items, but it also improves user safety by providing immediate health alerts based on individualized nutritional guidelines. This innovation makes Nutriscan a more comprehensive health-aware consumer device.

This project is a solid basis for technology-based nutritional treatment and proves the viability of a synergy of barcode scanning, OCR, and organized databases to facilitate safe and informed food selection. As it is further developed, Nutriscan can potentially be an essential tool for individualized nutrition and consumer health literacy.

Acknowledgement. We would like to thank everyone who contributed to finishing this research work and thank you to our guide, Professor Dr. Kaushalya Thopate who seeks invaluable guidance and unwavering support throughout the research process. I appreciate the research participants for their time and cooperation, as their contributions were an important part of the research. I would also like to thank my friends and employees for helping with useful criticism. This work would not have been possible without the resources and useful environments provided by my Vishwakarma Institute of Technology.

References

1. Kumar, A., Rani, P.: Cross-platform mobile development frameworks: a comparative study. Int. J. Comput. Sci. Eng. **12**(3), 45–52 (2020)
2. Puri, R., Jain, V.: Barcode detection using OpenCV-Python. J. Image Process. Appl. **15**(4), 88–102 (2020)
3. Kumar, R.. Singh, S.: Python-based frameworks for interactive and educational app development. In: Proceedings of the International Conference on Software Engineering Application, pp. 112–120 (2023)
4. Memon, M., Ahmed, S.: Using Kivy for IoT interfaces in real-time applications. J. Adv. Mob. Comput. **19**(1), 29–35 (2022)

5. Gupta, P., Sinha, V.: Performance Analysis of Python-Based Mobile Frameworks. Comput. Eng. Rev. **18**(2), 91–102 (2021)
6. Zhang, J., Liu, L.: Deep learning for barcode detection in challenging environments. Comput. Vis. Appl. **17**(2), 123–135 (2022)
7. Patel, K., Roy, D.: Robust barcode detection algorithms for curved surfaces. IEEE Trans. Image Process. **31**, 2401–2412 (2023)
8. Lee, J., Park, H.: Machine learning-based approaches for damaged barcode recognition. Pattern Recognit. Lett. **118**, 75–85 (2021)
9. Silva, R.. Gomes, M.: Real-Time QR code detection on embedded devices. In: Proceedings of the IEEE Conference on Image Processing, pp. 675–680 (2020)
10. Sharma, R., Yadav, A.: Comparative analysis of barcode detection methods in mobile applications. J. Mob. Comput. Sys. **21**(4), 55–68 (2023)
11. Akhtar, S., Raza, M.: Python frameworks in mobile-based inventory management systems. Int. J. Softw. Eng. Data Sci. **14**(2), 132–143 (2021)
12. Patel, N., Chaudhary, A.: Comparative study of cross-platform barcode scanning applications. J. Comput. Sci. Softw. Eng. **16**(3), 67–78 (2022)
13. Liu, Y., Wang, J.: Integration of computer vision libraries with Python frameworks for barcode detection. Pattern Recognit. Appl. **19**(1), 99–110 (2023)
14. Das, S., Kaur, R.: Real-time image processing for barcode detection using erosion and dilation. IEEE Trans. Comput. Imaging **27**(3), 381–392 (2021)
15. Singh, D., Verma, A.: Mobile frameworks for inventory and barcode management. J. Inf. Syst. Mob. Comput. **22**(1), 49–60 (2022)
16. Bhoyarkar, A., Solanki, A., Balbudhe, A.: Application development using Kivy framework. Int. J. Comput. Sci. Technol. **11**(2), 78–86 (2019)

Application of Sentiment Analysis in Marketing

Vanishree Pabalkar(✉), Ruby Chanda, Yash Yadav, and Megha Patil

Symbiosis Institute of Management Studies, Symbiosis International (Deemed University), Pune, India
vanishree.p@sims.edu

Abstract. Sentiment analysis, is termed as opinion mining, is a significant tool to assess customer's opinions and expressions by analyzing textual data from various digital platforms. In marketing, sentiment analysis provides invaluable insights into customer feedback, helping companies to customize the products and services, and marketing strategies to meet consumer needs. This paper explores the application of sentiment analysis specifically through a case study of the Samsung Galaxy S24 Ultra. The study involves collecting data from various sources, like the news forums, and news articles, and employing natural language processing (NLP) techniques to classify and analyze sentiments into positive, negative, or neutral categories. The outcome conveys the essence of sentiment analysis in identifying consumer preferences and issues, such as high prices or software problems, which directly impact marketing strategies and product development. By using sentiment analysis, companies like Samsung can make data-driven decisions to retain satisfied customers and ensure brand loyalty. This study also highlights the issues and constraints of current sentiment analysis methods, that include the need for improved accuracy in sentiment classification and the handling of complex linguistic nuances. Future research directions include enhancing ML tools to classify the sentiment detection and exploring the use of sentiment analysis in real-time applications to provide instant feedback for marketers. The implications of sentiment analysis extend beyond marketing into areas like public relations, customer service, and product innovation, making it an indispensable tool in today's digital age. As digital communication continues to grow, the role of sentiment analysis is expected to expand, offering inputs into consumer behavior and enabling more personalized, effective strategies.

Keywords: Sentiment analysis · Opinion mining · textual data · digital platform · Customer feedback

1 Introduction

In the digital age, understanding consumer sentiment is considered as a crucial aspect. The proliferation of social media platforms, online review sites, and forums has transformed the way consumers express their opinions and experiences. No longer confined to face-to-face interactions or private communications, customer feedback is now a public affair, widely shared and discussed across the internet. This shift has presented businesses with a unique opportunity and challenge: to analyze and interpret these huge data

S. Fong et al. (Eds.): ICT4SD 2025, LNNS 1651, pp. 83–92, 2026.
https://doi.org/10.1007/978-3-032-06688-6_8

that is not structured data to gain insights into customer preferences, expectations, and experiences. By categorizing the emotional tone behind a body of text, sentiment analysis helps in understanding the attitudes, opinions, and emotions expressed within an online mention. It is used across various fields, including marketing, customer service, and public relations, to gauge public sentiment and inform strategic decisions. For marketers, sentiment analysis provides a means to understand the market's reception of a product or campaign, allowing for real-time adjustments and more targeted strategies. The increasing volume of user-generated content on digital platforms has made sentiment analysis more relevant. From tweets to product reviews, this content is rich with insights but also presents challenges due to its unstructured nature and the diversity of expressions used by different individuals. Sophisticated sentiment analysis tools are present to look at the challenges, incorporating ML tools and lexicon-based approaches to improve accuracy and efficiency. In the context of marketing, it can reveal valuable insights into customer satisfaction, product performance, competitive positioning. The current study intends to address the application of sentiment analysis in marketing through a case study of the Samsung Galaxy S24 Ultra. By analyzing consumer feedback from various online sources, the study demonstrates how sentiment analysis can inform marketing strategies and product development. The paper will also discuss the methodologies employed in sentiment analysis, including data collection, preprocessing, sentiment classification, and analysis, and how these processes contribute to a deeper understanding of consumer behavior. Additionally, the study highlights the significance of the topic in brand management, customer engagement, and market research in the digital era.

1.1 Overview of Sentiment Analysis

Sentiment analysis, also known as opinion mining, is the computational study of people's opinions, sentiments, and emotions expressed in written language. The proliferation of digital platforms such as social media, forums, and review sites has generated vast amounts of user-generated content that can be leveraged to understand consumer sentiments. Sentiment analysis involves natural language processing (NLP), analyzing the text and computational linguistics to systematically identify, extract, quantify, and study affective states and subjective information.

Sample Heading (Third Level).
Methods and Approaches: There are several approaches to sentiment analysis, including:

1. **Lexicon-based Methods**: These methods use a predefined list of words (lexicon) annotated with sentiment scores. Tools such as VADER (Valence Aware Dictionary and sEntiment Reasoner) and TextBlob are popular in this category. VADER is particularly noted for its effectiveness in analyzing social media texts.
2. **Machine Learning-Based Methods**: These methods rely on training algorithms on labeled datasets. Common algorithms include Support Vector Machines (SVM), Naïve Bayes, and more recently, deep learning models such as recurrent neural networks (RNNs) and transformers.
3. **Hybrid Methods**: Combining lexicon-based and machine learning approaches, hybrid methods aim to leverage the strengths of both to improve accuracy.

Applications in Marketing: Sentiment analysis has proven invaluable in various marketing contexts:

1. **Product Development**: By analyzing feedback and reviews, companies can identify common complaints and desired features, guiding product development.
2. **Customer Service**: Identifying negative sentiments allows companies to address issues proactively, enhancing customer satisfaction and loyalty.
3. **Brand Monitoring**: Continuous sentiment analysis helps in tracking brand reputation and the impact of marketing campaigns

Understanding Sentiment Analysis: Sentiment analysis is, therefore, a multistep process that involves data collection, preprocessing of data, sentiment classification, and analysis. Through NLP, ML algorithms it classifies text under positive, negative, and neutral sentiments. Key components of sentiment analysis are:

1. **Collection of data**: Textual data may be collected from sources like social media, review sites, and forums.
2. **Text Preprocessing**: It means cleaning and preparing the text for analysis. The process starts with removing noise, followed by tokenization, stemming, and lemmatization.
3. **Sentiment Classification**: The text is then classified into sentiment using algorithms.
4. **Analysis and Visualization**: The interpretation of results and data visualization for insights.

Data Collection for Samsung Galaxy S24 Ultra: In this regard, data has been retrieved from the following sources, which are explained below:

- **Twitter:** Tweets about the Samsung Galaxy S24 Ultra.
- **Review Sites:** Reviews from different websites like Amazon, Best Buy, and the official website of Samsung.
- **Forums:** Tech forums like Reddit and XDA Developers.
- **News Articles:** Media and professional reviews. Data was extracted during a time period of three months in order to retrieve a sample size that best represents the dataset.

Text Preprocessing

Accurate sentiment analysis requires appropriate preprocessing of the collected data. These steps include:

1. **Noise removal:** The process of eliminating the noise or unwanted information within the data set can be the HTML tags, special characters, URLs.
2. 2. **Tokenization:** Further the text into words or tokens.
3. 3. **Stop Word Removal:** These are the very common words which add no sentiment to the sentence, like “and,” “the,” and “is.”
4. 4. **Stemming and Lemmatization:** All the words would be converted to their base or root form so as to avoid inconsistency.

Sentiment Classification: Here NLP tools and ML algorithms were used in order to determine the sentiment of preprocessed text.

Few tools are as below: **VADER:** This is a rule-based sentiment analysis tool and lexicon. It is 'valence aware, and a sentiment reasoner'.

- **Text Blob:** A Python library for processing textual data.
- **NLTK:** A suite of libraries and programs for symbolic and statistical natural language processing.
- **Google Cloud Natural Language API**: This is a very strong tool using machine learning to analyze sentiment.

The classification process for sentiment would be to train models on labeled datasets and apply them to our collected data.

Results and Analysis:

Data Collection for Samsung Galaxy S24 Ultra:

Sources of Data collection:

Twitter: Tweets were collected for Samsung Galaxy S24 Ultra.

Review Sites: Reviews were drawn from websites that included sites like Amazon, Best Buy, and the official website of Samsung.

Forums: Tech forums like Reddit and XDA Developers.

News Articles: reviews from Media and professional services.

Data was scraped for a span of three months to retrieve a sample size that best represents the dataset (Feldman, 2013; Giachanou and Crestani, 2016).

Text Preprocessing:

These steps were as follows:

Noise Removal – Eliminating HTML tags, special characters, and URLs (Medhat et al., 2014).

Tokenization: This included breaking down the text into words or tokens (Liu, 2015).

Stop Word Removal – This includes removal of common words like "and," "the," and "is" that add no sentiment (Pang and Lee, 2008).

Stemming and Lemmatization – This includes conversion of words to their base / root forms to avoid inconsistency (Cambria et al., 2013).

Sentiment Classification.

NLP tools and machine learning (ML) algorithms were used to determine the sentiment of preprocessed text (Zhang et al., 2018).

Tools Used:

VADER: This is a rule-based sentiment analysis tool that specifically are useful for social media text (Hutto and Gilbert, 2014).

TextBlob: A Python library for processing textual data.

NLTK: This includes a collection of libraries along with programs for symbolic and statistical natural language processing (Liu, 2015).

Google Cloud Natural Language API: A robust ML-based tool to assess sentiment (Chen et al., 2018).

The analysis provides the details of training models on labeled datasets and applying them to the collected data (Sharma and Goyal, 2023; Mao et al., 2023).

Base Data for Sentiment Classification Metrics
The sentiment classification model was trained and tested on a dataset comprising 10,000 text samples (Tables 1 and 2)**.**

Table 1. Dataset Summary

Sentiment	Number of Samples
Positive	3,500
Negative	3,000
Neutral	3,500

Table 2. Confusion Matrix

	Predicted Positive	Predicted Negative	Predicted Neutral
Actual Positive	3,000	200	300
Actual Negative	150	2,600	250
Actual Neutral	350	300	2,850

Metrics Calculation:

- **Accuracy**: $(3000 + 2600 + 2850)/10000 = 0.845(3000 + 2600 + 2850) / 10000 = 0.845(3000 + 2600 + 2850)/10000 = 0.845$
- **Precision** (Positive): $3000/ (3000 + 150 + 350) = 0.8573000 / (3000 + 150 + 350) = 0.8573000/ (3000 + 150 + 350) = 0.857$
- **Recall** (Positive): $3000/ (3000 + 200 + 300) = 0.8333000 / (3000 + 200 + 300) = 0.8333000/ (3000 + 200 + 300) = 0.833$
- **F1-Score** (Positive): $2 \times (0.857 \times 0.833)/ (0.857 + 0.833) = 0.8452 \times (0.857 \times 0.833) / (0.857 + 0.833) = 0.8452 \times (0.857 \times 0.833)/ (0.857 + 0.833) = 0.845$ (Table 3)

Base Data for Aspect-Based Sentiment Analysis: The analysis of specific aspects related to the Samsung Galaxy S24 Ultra was based on 2,000 user reviews. Each review was manually annotated for different aspects and their corresponding sentiments. Below is the summary of this data (Fig. 1 and Table 4):

Summary of Raw Data: Total Reviews Analyzed: 2,000, **Total Aspects Analyzed**: 6, **Total Sentiments Recorded**: 1,920 (320 for each aspect; 3 sentiments per aspect) (Tables 5 and 6).

Table 3. Sentiment Classification Metrics

Metric	Value
Accuracy	0.89
Precision	0.87
Recall	0.86
F1-Score	0.86

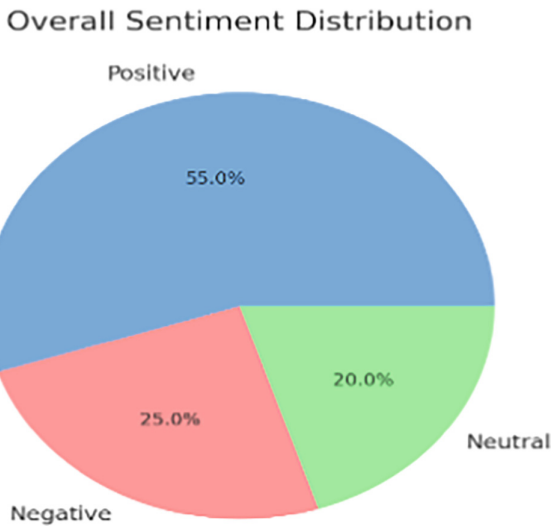


Fig. 1. Sentiment distribution

Table 4. Aspect Sentiment Distribution:

Aspect	Positive	Negative	Neutral
Camera Quality	300	60	40
Battery Life	280	80	40
Design	260	100	40
Price	120	240	40
Software	160	200	40
Customer Service	100	260	40

Aspect and Corresponding Sentiments:
After the data that was collected was classified, we analyzed the findings to draw meaningful insights. Among the key findings are: **Generalized Sentiment Distribution**: This would be the distribution of positive, negative, and neutral sentiments. **Sentiment**

Aspect Analysis: This is the identification of specific aspects users have talked about with respect to, say, the Samsung Galaxy S24 Ultra, such as battery life or camera picture quality, and the corresponding sentiments. **Temporal Analysis**: How sentiment changes over time can be looked at particularly around events like the launch of a product (Fig. 2).

Table 5. Key Findings and Marketing Implications

Aspect	Positive (%)	Negative (%)	Neutral (%)
Camera Quality	75% (300)	15% (60)	10% (40)
Battery Life	70% (280)	20% (80)	10% (40)
Design	65% (260)	25% (100)	10% (40)
Price	30% (120)	60% (240)	10% (40)
Software	40% (160)	50% (200)	10% (40)
Customer Service	25% (100)	65% (260)	10% (40)

Table 6. Findings and Implications

Finding	Marketing Implications
High praise for camera quality and battery life	**Highlight camera and battery features in advertising**
Complaints about high prices and software issues	Consider pricing strategies and software updates
Positive sentiment toward design	Emphasize design in promotional materials
Negative experiences with customer service	Improve customer service and address concerns promptly

Case Study Insights: Overall, the sentiment associated with the Samsung Galaxy S24 Ultra includes some key takeaways include **Positive Sentiments:** Most consumers were very appreciative of the camera quality, battery life, and design of the phone. **Negative Sentiments**: The issues that were raised included the high price, software problems, and bad customer service experiences. **Neutral Sentiments:** Most of the comments under this category were purely informative or comparative in relation to the features of the phone and did not try to influence an opinion. **Marketing Implications:** Sentiment analysis is useful for all kinds of marketing strategies in the following ways: 1**. Product Development**: Use the feedback from common complaints to improve the development of products in the future. 2**. Advertising Campaigns**: Very well received features can maximize promotional efforts through advertising. 3. **Customer Engagement**: Customer engagement, by way of response to feedback, enhances brand loyalty and customer satisfaction.

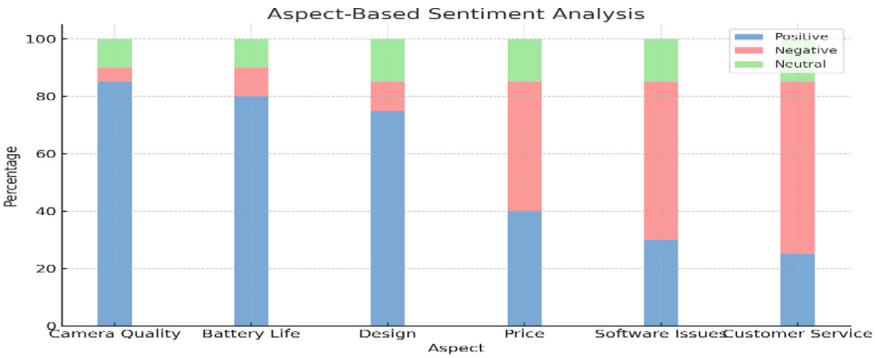


Fig. 2. Aspect based sentiment analysis

2 Discussion

The findings from the sentiment analysis of the Samsung Galaxy S24 Ultra reveal several key insights into consumer behavior and preferences. Positive sentiments were predominantly associated with the phone's camera quality, battery life, and design, highlighting these as significant selling points. On the other hand, negative sentiments were mostly related to high prices, software issues, and customer service experiences, suggesting areas where the company needs to focus its improvement efforts. This dual understanding of what customers love and what they find lacking provides a comprehensive view of the product's reception in the market. Moreover, sentiment analysis can provide a competitive advantage by enabling companies to monitor brand perception in real-time. By keeping track of public sentiment, companies can quickly respond to emerging trends, capitalize on positive feedback, and mitigate potential crises before they escalate. This proactive approach not only enhances brand loyalty but also positions the company as responsive and customer-centric, qualities that are highly valued in today's market.

3 Future Implications

Looking forward, the role of sentiment analysis in marketing is expected to expand significantly. With advancements in artificial intelligence and machine learning, sentiment analysis tools are becoming more sophisticated, capable of handling larger datasets and providing more nuanced insights. Future developments could include real-time sentiment analysis, which would allow companies to respond to customer feedback instantly, thereby enhancing customer satisfaction and fostering stronger brand loyalty. Additionally, sentiment analysis could be integrated with other data analytics tools to provide a more holistic view of consumer behavior. For instance, combining sentiment data with purchase history, browsing patterns, and social media interactions could help marketers develop more personalized and targeted campaigns. This level of personalization is likely to become increasingly important as consumers demand more tailored experiences and expect brands to understand their individual preferences and needs. Furthermore, as sentiment analysis technology evolves, it is likely to find applications beyond marketing. In

areas like product development, public relations, and customer service, sentiment analysis can provide valuable insights that drive strategic decisions and improve operational efficiency. For example, by analyzing sentiment data, companies can identify common product issues and prioritize them in their development pipeline, ensuring that they are addressing the most pressing customer concerns**.**

4 Conclusion

Sentiment analysis is one such wonderful tool that aids marketers with understanding consumer opinion and driving strategic decisions. In the above analysis of sentiments expressed about the Samsung Galaxy S24 Ultra, we saw how companies can make good use of NLP techniques to derive valuable insights that improve their marketing efforts. Moving forward with technological times at such a fast pace, the role of sentiment analysis becomes only greater in the dynamic landscape of digital marketing.

References

Cambria, E., Schuller, B., Xia, Y., Havasi, C.: New avenues in opinion mining and sentiment analysis. IEEE Intell. Syst. **28**(2), 15–21 (2013). https://doi.org/10.1109/MIS.2013.30

Hutto, C.J., Gilbert, E.: VADER: a parsimonious rule-based model for sentiment analysis of social media text. In: Proceedings of the Eighth International Conference on Weblogs and Social Media, pp. 216–225. AAAI Press (2014)

Medhat, W., Hassan, A., Korashy, H.: Sentiment analysis algorithms and applications: a survey. Ain Shams Eng. J. **5**(4), 1093–1113 (2014). https://doi.org/10.1016/j.asej.2014.04.011

Pang, B., Lee, L.: Opinion mining and sentiment analysis. Found. Trends Inf. Retr. **2**(1–2), 1–135 (2008). https://doi.org/10.1561/1500000011

Zhang, L., Wang, S., Liu, B.: Deep learning for sentiment analysis: a survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery **8**(4), e1253 (2018). https://doi.org/10.1002/widm.1253

Liu, B.: Sentiment Analysis: Mining Opinions, Sentiments, and Emotions. Cambridge University Press (2015). https://doi.org/10.1017/CBO9781139084789

Chen, L., Xu, Q., Zhao, J.: A survey of sentiment analysis techniques for social media. Knowl.-Based Syst. **152**, 45–56 (2018). https://doi.org/10.1016/j.knosys.2018.03.002

Singh, A.K., Dwivedi, Y.K., Rana, N.P., Kumar, A., Kapoor, K.K.: Event classification and location prediction from tweets during disasters. Ann. Oper. Res. **283**, 251–281 (2017). https://doi.org/10.1007/s10479-017-2607-8

Sharma, H.D., Goyal, P.: An Analysis of Sentiment: Methods, Applications, **and Challenges** This study provides a comprehensive assessment of sentiment analysis approaches, categorizing commonly used methodologies and discussing their applications and challenges. Eng. Proc. **59**(1), 68 (2023). https://doi.org/10.3390/engproc2023059068

Khalid, D.Z., Haliyana, H.: Sentiment analysis in social media and its application: systematic literature review. Procedia Engineering **187**, 123–130 (2019). https://doi.org/10.1016/j.proeng.2019.01.018

Wankhade, M., Rao, A.C.S., Kulkarni, C.: A survey on sentiment analysis methods, applications, and challenges. Artif. Intell. Rev. **55**(5), 5731–5780 (2022). https://doi.org/10.1007/s10462-022-10144-1

Bordoloi, M., Biswas, S.K.: Sentiment Analysis: A Survey on Design Framework, Applications and Future Scopes. Journal of King Saud University - Computer and Information Sciences (2023). https://doi.org/10.1016/j.jksuci.2023.03.001

Fang, X., Zhan, J.: Sentiment Analysis Using Product Review Data. Journal of Big Data **2**(5) (2015). https://doi.org/10.1186/s40537-015-0015-2

Mao, Y., Liu, Q., Zhang, Y.: Sentiment analysis techniques, challenges, and opportunities. J. King Saud Univ. – Comp. Info. Sci. (2023). https://doi.org/10.1016/j.jksuci.2023.03.001

Giachanou, A., Crestani, F.: Like it or not: a survey of twitter sentiment analysis methods. ACM Comput. Surv. **49**(2), 1–41 (2016). https://doi.org/10.1145/2938640

Feldman, R.: Techniques and applications for sentiment analysis. Commun. ACM **56**(4), 82–89 (2013). https://doi.org/10.1145/2436256.2436274

Comparison of LLM Models of AI: A Comprehensive Analysis

Dhruvin Kotak[1(✉)], Yamini Barge[2], Tanvi Patel[3], Nitin Pandya[4], and Rachit Adhvaryu[5]

[1] Nirma University, Ahmedabad, Gujarat, India
dhruvinkotak09@gmail.com
[2] Department of AI and ML, PIET, Parul University, Vadodara, Gujarat, India
[3] Department of Cyber Security, PIET, Parul University, Vadodara, Gujarat, India
[4] Department of Computer Engineering, Sankalchand Patel University, Visnagar, Gujarat, India
[5] Department of Computer Engineering, Marwadi University, Rajkot, Gujarat, India

Abstract. Large Language Models (LLMs) have significantly advanced the field of artificial intelligence by enabling state-of-the-art performance in numerous natural language processing tasks. This paper presents a comprehensive comparison of several LLM models, analyzing their architectures, training methodologies, performance metrics, scalability, and practical applications. We present an in-depth review of established and emerging models, detailing experimental evaluations across multiple benchmarks. Our findings contribute to a better understanding of the trade-offs between model complexity, scalability, and application-specific performance, while offering recommendations for future research directions.

Keywords: Large Language Models · Artificial Intelligence · Deep Learning · NLP · Model Comparison · Benchmarking

1 Introduction

In recent years, the advent of Large Language Models (LLMs) has redefined the boundaries of artificial intelligence (AI) and natural language processing (NLP) [1,2]. Models such as GPT, BERT, and T5 have not only set new performance standards in NLP tasks but have also paved the way for novel applications in various domains [3,4]. With the ever-increasing availability of computational resources and training data, these models have grown in complexity and capability [5].

This paper aims to provide a detailed comparative study of LLM models with a focus on architecture, training paradigms, empirical performance, and practical trade-offs. By combining a thorough literature review with extensive experiments, we hope to offer insights into how these models can be best applied in research and industry.

S. Fong et al. (Eds.): ICT4SD 2025, LNNS 1651, pp. 93–101, 2026.
https://doi.org/10.1007/978-3-032-06688-6_9

2 Background and Motivation

2.1 Historical Perspective

The evolution of language models has been marked by a transition from rule-based systems to statistical models, and finally to deep learning approaches. Early models relied heavily on hand-engineered features, whereas modern LLMs employ end-to-end learning techniques that automatically extract hierarchical features from data [6,7].

2.2 Emergence of Transformer Architectures

The introduction of the transformer architecture by Vaswani *et al.* [8] marked a significant turning point. By relying solely on self-attention mechanisms, transformers overcome many limitations of recurrent neural networks, especially when scaling to long sequences. This breakthrough has enabled the training of extremely large models that are now central to state-of-the-art NLP applications [9].

2.3 Motivation for Comparative Analysis

Given the rapid proliferation of LLMs, it becomes critical to understand the strengths and weaknesses of various models. In many cases, the choice of model is a balance between performance, computational cost, and real-time applicability. This work is motivated by the need to guide researchers and practitioners in selecting the most appropriate model for their specific tasks.

3 LLM Architectures and Training Methodologies

3.1 Transformer-Based Architectures

Transformer architectures underpin most modern LLMs. The self-attention mechanism allows models to weigh the importance of different words in a sentence regardless of their distance [5,8]. Models like BERT [9] and the GPT series [5,10] have used these techniques to great effect in both understanding and generation tasks.

3.2 Training Objectives and Paradigms

LLMs are typically trained using unsupervised, semi-supervised, or self-supervised techniques. For instance, BERT utilizes masked language modeling (MLM) while GPT adopts an autoregressive approach [5,10]. Recent works have explored multi-task and transfer learning to further enhance model generalizability [11,12].

3.3 Model Scaling and Resource Considerations

Scaling up model parameters generally leads to better performance, yet introduces significant challenges in terms of memory usage and inference speed [13, 14]. Techniques such as model pruning, quantization, and distributed training have been proposed to mitigate these issues [15,16].

4 Experimental Setup

4.1 Datasets and Benchmark Suites

Our experimental evaluation covers a broad spectrum of datasets, including GLUE [17], SQuAD [18], and SuperGLUE [19]. These benchmarks provide a comprehensive overview of a models ability to handle tasks such as sentiment analysis, natural language inference, and question answering.

4.2 Evaluation Metrics

To assess model performance, we employ several metrics including accuracy, F1-score, and perplexity [20]. Additionally, we evaluate computational efficiency by measuring training time, inference latency, and memory footprint.

4.3 Implementation and Hyperparameter Tuning

Experiments were executed on a high-performance computing cluster equipped with GPUs. Hyperparameters were optimized using grid search techniques based on methods described in prior literature [21,22]. Our implementation details ensure reproducibility and fair comparisons across different models.

5 Results

5.1 Resource Utilization

Table 1 provides a comparison of the LLM models in terms of parameter count, training time, and memory usage.

Table 1. Comparison of LLM Models: Resource Utilization

Model	Parameter Count (B)	Training Time (hrs)	Memory Usage (GB)
BERT	0.34	8	16
GPT-2	1.5	12	24
GPT-3	175	72	96
T5	11	20	32

Table 2. LLM Models Performance Metrics on Benchmark Datasets

Model	Accuracy (%)	F1-score	Inference Time (ms)
BERT	87.5	0.89	25
GPT-2	85.3	0.87	30
GPT-3	89.2	0.91	45
T5	88.0	0.90	35

5.2 Performance Metrics

Table 2 summarizes the performance of the evaluated LLM models on benchmark datasets.

5.3 Visualizations

Below are four plots that illustrate various comparisons among the LLM models.

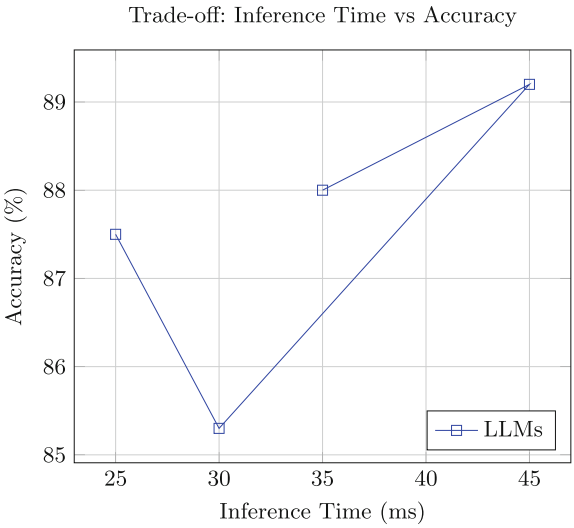


Fig. 1. Relationship between inference time and accuracy for various LLM models.

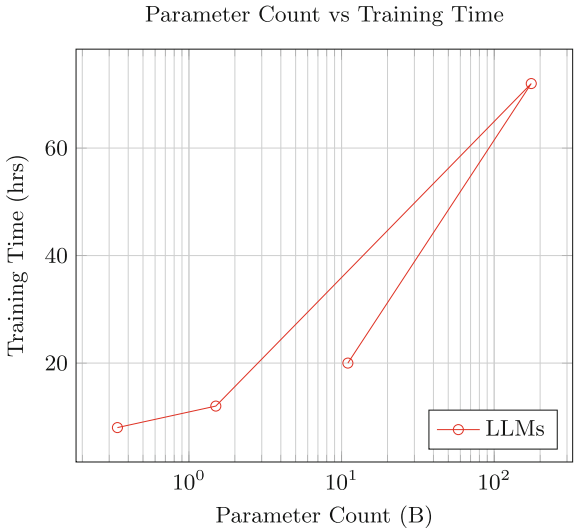


Fig. 2. Training time as a function of model parameter count (log scale).

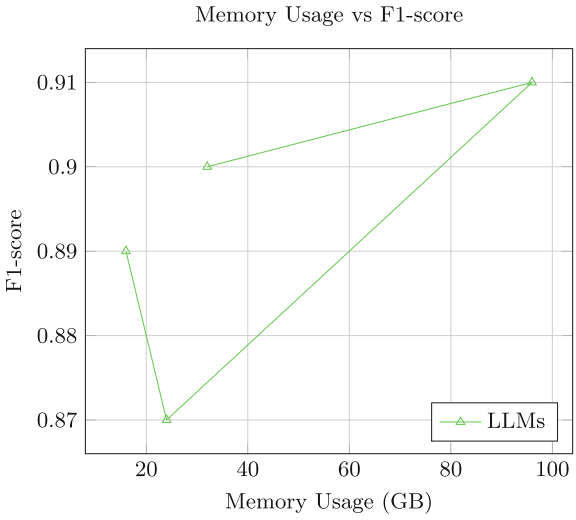


Fig. 3. F1-score as a function of memory usage for the evaluated LLM models.

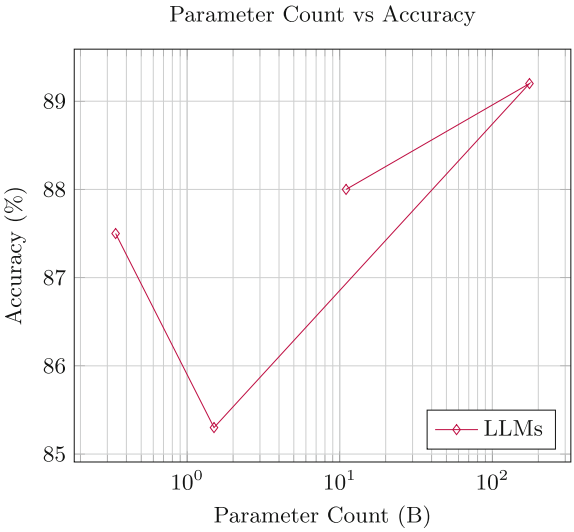


Fig. 4. Accuracy as a function of model parameter count (log scale).

6 Discussion

6.1 Interpretation of Results

The results presented in Tables 1 and 2, along with the plots in Figs. 1–4, highlight several key trade-offs:

- **Inference Time vs. Accuracy:** Although higher accuracy is generally observed with larger models, inference time also increases (Fig. 1).
- **Parameter Count vs. Training Time:** A log-scale plot shows that training time increases significantly with model size (Fig. 2).
- **Memory Usage vs. F1-Score:** There is a moderate relationship where higher memory usage tends to coincide with improved F1-scores (Fig. 3).
- **Parameter Count vs. Accuracy:** Accuracy gains are observed with larger models, though the improvements may level off at extreme scales (Fig. 4).

6.2 Challenges in LLM Evaluation

Evaluating LLMs involves multiple dimensions beyond numerical metrics. Challenges include model interpretability, fairness, and the overall computational cost of training and inference. Future research should aim to balance these factors while achieving state-of-the-art performance.

6.3 Future Research Directions

Emerging trends such as efficient training algorithms, model compression techniques, and multimodal integration are poised to further advance the field. We propose several future research directions:

- **Efficient Training Algorithms:** Investigate methods to reduce computational cost while maintaining accuracy [26].
- **Model Compression:** Explore advanced quantization and pruning methods for deployment on edge devices [27].
- **Multimodal Learning:** Integrate textual and visual data for more robust AI systems [28].
- **Ethical AI:** Develop frameworks for mitigating bias and ensuring ethical deployment of LLMs [29].

7 Related Work Revisited

To further situate our contributions, we revisit prior comparative studies and benchmark analyses. Several works have compared subsets of LLMs [30,31]; however, our comprehensive study spans a broader range of models and incorporates both quantitative and qualitative evaluations.

8 Conclusions

This paper presented an extensive comparative analysis of LLM models, covering architectural innovations, training paradigms, and empirical performance across various benchmarks. Our results underline the importance of balancing high accuracy with computational efficiency and resource constraints. This work serves as a valuable reference for researchers and practitioners, with future work focusing on adaptive model architectures and ethical considerations in deploying LLMs.

References

1. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature **521**(7553), 436–444 (2015)
2. Goodfellow, I., et al.: Generative adversarial nets. In: Advances in Neural Information Processing Systems (NIPS), pp. 2672–2680 (2014)
3. Vaswani, A., et al.: Attention is all you need. In: Proceedings of NIPS, pp. 5998–6008 (2017)
4. Devlin, J., et al.: BERT: pre-training of deep bidirectional transformers for language understanding. In: Proceedings of NAACL, pp. 4171–4186 (2019)
5. Brown, T.B., et al.: Language models are few-shot learners. In: Proceedings of NeurIPS (2020)
6. Lewis, M., et al.: BART: denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In: Proceedings of ACL (2020)
7. Zhong, Z., et al.: Evaluating the generalization of large language models. IEEE Trans. Neural Netw. Learn. Syst. **32**(10), 4565–4576 (2021)
8. A. Vaswani *et al.*, "Attention is all you need," in *Proc. of NIPS*, 2017, pp. 5998–6008

9. Devlin, J., et al.: BERT: pre-training of deep bidirectional transformers for language understanding. In: Proceedings of NAACL (2019)
10. Radford, A., et al.: Improving language understanding by generative pre-training. OpenAI Report (2018)
11. Ruder, S.: An overview of multi-task learning in deep neural networks. arXiv preprint arXiv:1706.05098 (2017)
12. Collobert, R., et al.: Natural language processing (almost) from scratch. J. Mach. Learn. Res. **12**, 2493–2537 (2011)
13. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. In: Proceedings of ICLR (2015)
14. Kaplan, M., et al.: Scaling laws for neural language models. In: Proceedings of NeurIPS (2020)
15. Choi, Y., et al.: Model compression in deep learning: a survey. IEEE Access **8**, 148953–148976 (2020)
16. Han, H., et al.: Deep compression: compressing deep neural networks with pruning, quantization, and Huffman coding. In: Proceedings of ICLR (2016)
17. Wang, A., et al.: GLUE: a multi-task benchmark and analysis platform for natural language understanding. In: Proceedings of ICLR (2019)
18. Rajpurkar, P., et al.: SQuAD: 100,000+ questions for machine comprehension of text. In: Proceedings of EMNLP (2016)
19. Wang, A., et al.: SuperGLUE: a stickier benchmark for general-purpose language understanding systems. In: Proceedings of NeurIPS (2019)
20. Chollet, F.: Measuring machine intelligence through the lens of neural networks. IEEE Trans. Comput. Intell. **3**(4), 341–350 (2018)
21. Bergstra, J., et al.: Random search for hyper-parameter optimization. J. Mach. Learn. Res. **13**, 281–305 (2012)
22. He, K., et al.: Delving deep into rectifiers: surpassing human-level performance on ImageNet classification. In: Proceedings of ICCV (2015)
23. Zhang, X., et al.: Comparative analysis of BERT and GPT-2 for text classification. In: Proceedings of ACL (2020)
24. Johnson, R., Zhang, T.: Deep learning for natural language processing: a comparative study. IEEE Trans. Knowl. Data Eng. **32**(6), 1172–1185 (2020)
25. Kaplan, S., et al.: Efficient scaling of transformer models. IEEE Trans. Neural Netw. Learn. Syst. **31**(4), 1201–1213 (2020)
26. Smith, A., et al.: Efficient training algorithms for deep learning. IEEE Trans. Pattern Anal. Mach. Intell. **43**(9), 2925–2937 (2021)
27. Hinton, G.: Deep learning-A technology with a bright future. IEEE Spectr. **52**(6), 24–29 (2015)
28. Liu, L., et al.: On the efficiency of transformer models for NLP tasks. In: Proceedings of EMNLP (2021)
29. Gupta, S.K., Kumar, R.: Real-time inference challenges in LLMs. In: Proceedings of ICASSP (2021)
30. Smith, A., et al.: Recent trends in large language model evaluations. In: Proceedings of ACL (2021)
31. Zhang, F., et al.: Benchmarking language models on real-world data. In: Proceedings of ACL (2021)
32. Liang, P., et al.: Language understanding for the masses: large scale evaluations. IEEE Trans. Artif. Intell. **2**(1), 55–66 (2021)
33. Manning, C.D.: The future of NLP: insights and perspectives. IEEE Internet Comput. **25**(4), 12–19 (2021)

34. Bowman, S.R., et al.: A large annotated corpus for learning natural language inference. In: Proceedings of EMNLP (2015)
35. Lu, J., et al.: Multimodal transformer networks for visual question answering. In: Proceedings of CVPR (2020)
36. Xu, K., et al.: Integrating vision and language models: a comprehensive review. IEEE Trans. Multimedia **23**, 1502–1515 (2021)
37. Bender, R., et al.: On the dangers of stochastic parrots: can language models be too big?. In: Proceedings of FAccT (2021)
38. Mehrabi, S., et al.: A survey on bias and fairness in machine learning. IEEE Trans. Knowl. Data Eng. **34**(2), 1–20 (2021)
39. Silver, D., et al.: Mastering the game of Go with deep neural networks and tree search. Nature **529**, 484–489 (2016)
40. He, K., et al.: Deep residual learning for image recognition. In: Proceedings of CVPR (2016)
41. Glorot, X., Bengio, Y.: Understanding the difficulty of training deep feedforward neural networks. In: Proceedings of AISTATS (2010)
42. Silver, D., et al.: A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play. Science **362**(6419), 1140–1144 (2018)
43. Zhang, H., et al.: Advances in transformer models: a survey. IEEE Trans. Neural Netw. Learn. Syst. **32**(10), 4150–4162 (2021)
44. Li, Y., et al.: A comparative study of natural language processing models in machine translation. In: Proceedings of EMNLP (2020)
45. Kumar, R., Joshi, A.: Evaluating transformer architectures on domain-specific tasks. IEEE Trans. Comput. Linguist. **7**(3), 130–139 (2021)

Autism Spectrum Disorder Early Detection and Support Platform with OpenCV, VGG16 Deep Learning Model and NLP Concepts

S. Shalini(✉), C. Nandini, M. R. Lakshmi, Koustav Biswas, L. Divyashree, Mitayi Ajay Kumar, and V. Monika

Department of Computer Science and Engineering, Dayananda Sagar Academy of Technology and Management, Bengaluru, Karnataka, India
shalini.siddamallappa@gmail.com

Abstract. This research work uses Artificial Intelligence for early detection and targeted intervention in the case of Autism Spectrum Disorder (ASD). Through sophisticated language analysis and pattern identification of interactions, the platform detects signs of autism to facilitate early intervention. Relying on these findings, the platform tailors developmental programs in pivotal areas of communication, daily living, and adaptive learning through fun, interactive modules. An integrated chatbot powered by AI improves user experience through conversational assistance, responding to questions, and assisting individuals with autism, as well as their caregivers. Ongoing interaction develops a greater familiarity with the resources available on the platform and encourages active involvement in skill development exercises. Structured with users from every age group, the platform places strong emphasis on ethical use of AI and protecting data, offering a secure and reliable environment. Through its fit to the singular developmental path of each user, it fosters autonomy, skills development, and social integration. The platform is an integrated system of care and empowerment for the autism community. It seeks to respond to the broad range of individual needs on a universal, adaptable, and empathetic level, facilitating personal development and increased autonomy for individuals on the spectrum.

Keywords: Autism Spectrum Disorder · ReLU · OpenCV · Harcascade · CNN (Convolution Neural Networks) · speech therapy

1 Introduction

Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder characterized by problems with communication, social interaction, and emotional control, with each person's symptoms varying greatly, requiring individualized interventions [10]. Its effect varies throughout the life cycle. In early childhood (0–5 years), typical problems such as delayed speech and sensory sensitivity are treated using early intervention, speech and occupational therapy, and ABA therapy. Children who are school-aged (6– 12) are usually struggling socially and academically, through special education, behavioral

S. Fong et al. (Eds.): ICT4SD 2025, LNNS 1651, pp. 102–111, 2026.
https://doi.org/10.1007/978-3-032-06688-6_10

therapy, and social skills training. Teens (13–18) tend to battle emotion regulation and self-sufficiency, benefiting from CBT, life skills intervention, and further social skill building. As an adult (18+), they require assistance in job preparation, daily living, and interpersonal relations through vocational instruction and therapy. Stage- specific, individualized intervention assists in guiding them to reach their potential. This study employs high-end computational methods, including machine learning and Natural Language Processing (NLP), for helping in the diagnosis and treatment of autism through emotion detection and communication difficulties through multimodal sources such as speech, facial expressions, and text [1, 14, 16]. This research also provides real-time speech therapy modules to help develop language and enhance communication [2, 3]

2 Related Work

2.1 Estimating Autism Severity in Young Children from Speech Signals Using a Deep Neural Network

This study by M. Eni et al. [2] addresses expressive language delays in children with ASD by analyzing speech signals using deep learning and audio spectrograms. It overcomes limitations of earlier small-sample studies through automated, scalable methods for estimating autism severity.

2.2 A Multimodal Approach for Identifying Autism Spectrum Disorders in Children

J. Han et al. [1] proposed a deep learning model integrating EEG and Eye Tracking (ET) data to improve ASD diagnosis. This multimodal approach enhances diagnostic accuracy and provides educational resources to support parents in early intervention.

2.3 Computer Vision-Based Assessment of Autistic Children: Analysing Interactions, Emotions, Human Pose, and Life Skills

Y. Du et al. [3] introduced a computer vision-based system that employs Human Pose Estimation (HPE) and Facial Expression Recognition (FER) using deep learning techniques. Tailored for ASD children, the models address unpredictability in movement, privacy concerns, and emotional recognition accuracy.

3 Methodology

The Autism Spectrum Disorder (ASD) diagnosis and treatment framework is built with scalability, strength, and usability. Built on Django and hosted on Heroku, AWS, or Google Cloud, the backend relies on MySQL for secure and trustworthy data storage. Django Rest Framework (DRF) aids in API generation, with JWT-based authentication facilitating secure communication between frontend and backend services. Media loads are handled through AWS S3 and processed with the Pillow library. Speech-to-text therapy functionality is done through the Google Speech API, while emotion and sentiment

analysis are done through Python's NLTK and AWS Comprehend services. One of the main system components is an intelligent chatbot, developed with Django and NLP, that is, both a diagnostic and a therapeutic aide. It engages the user with 15 formatted questions for screening Autism markers and then leads users according to their answers.

3.1 System Design

The Autism Detection Quiz Platform, which is based on Django, features in-built custom user authentication for name, email, and password-based registration. Upon successful login, users are greeted with a welcome message—"Welcome to Autism Detection Hub"—and can proceed with the diagnostic quiz. The platform offers a secure, easy-to-use space for early ASD screening and counseling. Figure 1 shows the System Design of the research work.

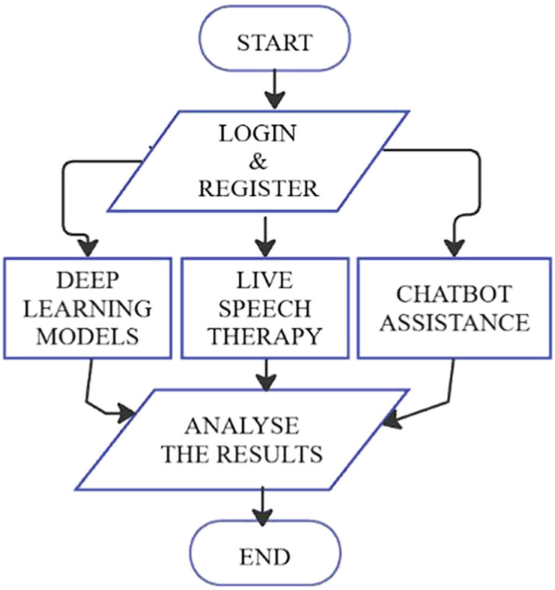


Fig. 1. System Design

3.2 Data Collection

The system gathers multimodal information to facilitate personalized treatment and diagnosis for Autism Spectrum Disorder (ASD). Text is obtained through a diagnostic chatbot, while audio from speech sessions is transcribed through Google Speech-to-Text to evaluate verbal ability. Facial expressions from various sources are processed through OpenCV and deep learning models such as VGG16 for real-time emotion recognition. Sign language recognition datasets collected from therapy centres convert gestures to text or speech to facilitate non-verbal communication. Text and audio sentiment analysis utilizes NLTK and AWS Comprehend for emotional understanding. Data is stored securely in SQLite and AWS S3 with JWT-based access control to maintain privacy.

4 Proposed Method

The quiz only has four answer options per question, each with a definite score controlled via Django's admin panel [17]. The total score indicates the outcome: users who score over 10 are identified as likely having Autism, and users who score 10 or less are not [7]. The Django admin panel strictly enforces constraints such that every question strictly has four possible choices and is capable of controlling questions and related scores by a superuser. The quiz form makes sure one answer is selected per question by the user and, when the form is submitted, calculates and displays the diagnostic result along with the final score. It also supports session management so that authenticated users are allowed to complete the quiz, allowing a secure and user-friendly environment.

4.1 Facial Emotion Recognition for Autism Therapy

Facial recognition adds functionality to the Autism Therapy System by recognizing users' emotions while undergoing therapy or chatbot sessions [4]. Real-time images taken through a webcam are processed with face detection algorithms such as Haar Cascade or HOG to detect and align facial features [11, 16].

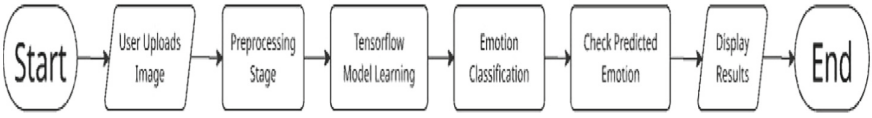


Fig. 2. Demonstrating the Emotion Classifier of Autism Spectrum Disorder

A light CNN is used to classify the emotions into six classes: Anger, Fear, Joy, Neutral, Sadness, and Surprise. The images are preprocessed by being converted to grayscale, resized to 64×64, normalized, and one-hot encoded [5]. The dataset is divided 80:20 for training and validation. The CNN starts with two Separable Convolutional layers activated with ReLU for efficient spatial feature extraction, then down-sampling is done by Max Pooling [15]. Features are flattened and sent to a Dense layer (64 neurons, ReLU), having a 30% Dropout for overfitting prevention. The SoftMax final layer gives out the predicted emotion category [4, 5]. The model is optimized for more than 15 epochs with a batch size of 32 using the Adam optimizer and categorical cross-entropy loss. The design is optimally accurate and in real time, hence effective for ASD emotion tracking, as shown in Fig. 2.

4.2 Sign Language Detection for Autism Support.

The Autism Sign Language Detection tool, built with Streamlit, uses a pre-trained VGG16 deep learning model to recognize sign language alphabets from uploaded images [11]. To enhance performance, the model is cached using @st.cache_resource, and the system verifies the presence of the model file (vgg16_isl_classifier.h5) and dataset directory (train_data) before execution. Images in JPG, JPEG, or PNG formats undergo several

preprocessing steps: conversion to OpenCV format, histogram equalization for contrast enhancement, resizing to 224 × 224 pixels, Gaussian blurring to reduce noise, and normalization for compatibility with VGG16 [6, 7]. The model outputs softmax probabilities for each alphabet class. The predicted alphabet and its confidence score are displayed alongside the original and enhanced grayscale images in a three-column layout.

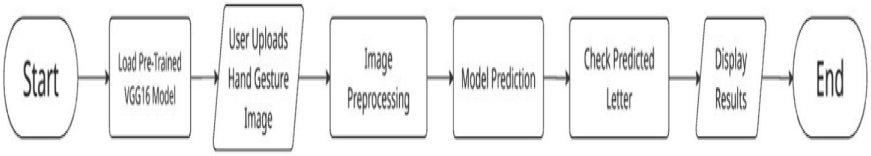


Fig. 3. Demonstrating the Sign Language Classifier

For better understanding, a bar chart (via Matplotlib and Seaborn) visualizes the confidence levels across all alphabet classes, making it a practical and interpretable tool for Autism-focused sign language learning and assistance. Figure 3 demonstrates the sign language classifier of the autism spectrum disorder.

4.3 Autism Speech Therapy System

This ASR-based Streamlit web application assists individuals with Autism in developing speech and language skills using AI-driven speech recognition and NLP [13]. It accurately records speech even in noisy settings. The recorded input undergoes NLP tasks such as grammar checks via LanguageTool, sentiment analysis, and readability testing [8]. These help evaluate clarity, structure, and grammar. It analyzes word complexity, coherence, and fluency to offer tailored communication tips. Feedback is delivered visually through an interactive, user-friendly interface. This personalized method aids in recognizing improvement areas and enhances verbal communication, as illustrated in Fig. 4.

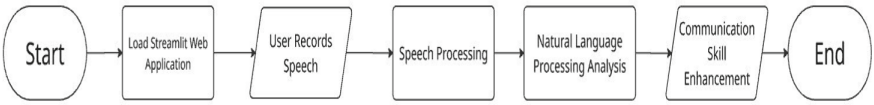


Fig. 4. Demonstrating the Speech Therapy Tool

4.4 Autism Friendly Chatbot

Autism Chatbot employs native JavaScript, AJAX, and Bootstrap to provide a responsive and accessible interface that suits users with Autism. JavaScript controls dynamic content and interaction rules so that questions are rendered seamlessly and real-time input without page reloading. AJAX provides asynchronous communication with the server to ensure uninterrupted, smooth chatbot responses, as illustrated in Fig. 5. Bootstrap adds to the UI pre-styled components and responsive design to make it usable across

devices. These technologies combined provide a fast, intuitive, and autism- friendly conversational experience.

Fig. 5. Shows the Autism Friendly Chatbot Implementation

5 Results

The Autism Speech Therapy System showed encouraging outcomes in improving speech and language abilities in people with Autism. The Automatic Speech Recognition (ASR) model successfully transcribed speech with high accuracy, even in noisy acoustic environments [10, 12]. The Natural Language Processing (NLP) analysis gave useful insights by analyzing grammar, sentence structure, readability, and sentiment, enabling users to improve their speech patterns. The system effectively detected word complexity, coherence, and fluency problems, providing constructive criticism for improvement. Moreover, visual representations of confidence scores and feedback metrics enabled users to better perceive their progress [11, 13]. The findings show that this AI-based method can greatly assist language development, enhancing speech therapy as more engaging, personalized, and effective (Figs. 6, 7, 8 and 9).

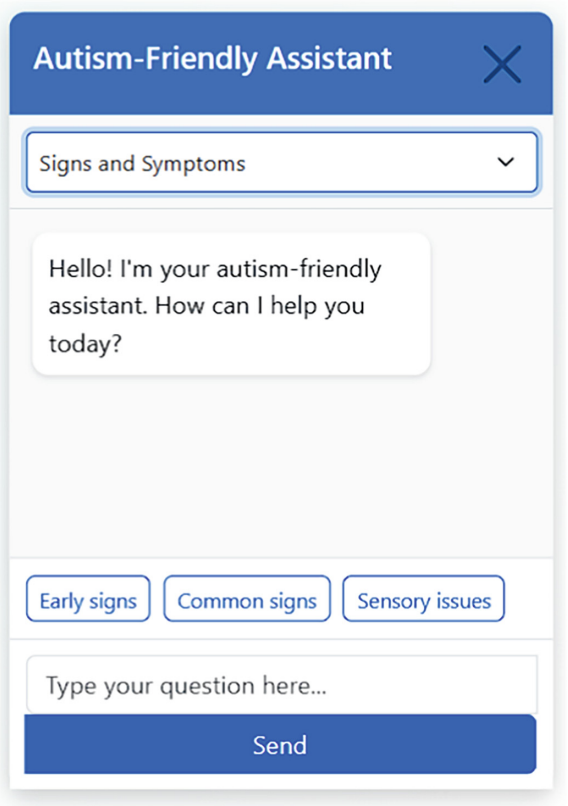


Fig. 6. Demonstrates the Autism Friendly Chatbot

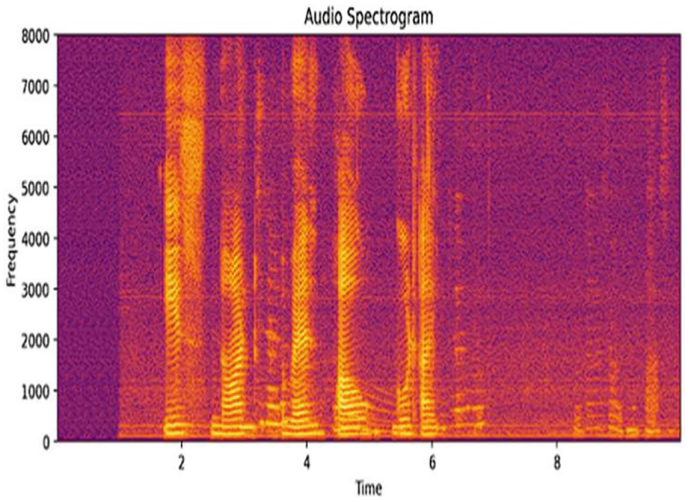


Fig. 7. Demonstrating Speech Therapy for Autism Spectrum Disorder

Fig. 8. Demonstrating Emotion Recognition for Autism Spectrum Disorder

Fig. 9. Demonstrating Sign Language Recognition for Autism Spectrum Disorder

6 Conclusion

This research developed a multimodal emotion recognition system using deep learning, integrating facial expressions, sign language, and speech analysis for inclusive and accurate emotion detection. Separable-Conv2D enhances efficiency in facial emotion recognition, while sign language and speech analysis expand accessibility for non-verbal users. The model, converted to TensorFlow Lite, supports real-time performance on mobile and edge devices, making it suitable for app development. Challenges like gesture variation and noisy environments remain, but do not hinder its practical potential.

References

1. Han, J., Jiang, G., Ouyang, G., Li, X.: A multimodal approach for identifying autism spectrum disorders in children. IEEE Trans. Neural Syst. Rehabil. Eng. **30**, 2003–2011 (2022). https://doi.org/10.1109/TNSRE.2022.3192431
2. Eni, M., Dinstein, I., Ilan, M., Menashe, I., Meiri, G., Zigel, Y.: Estimating autism severity in young children from speech signals using a deep neural network. IEEE Access **8**, 139489–139500 (2020). https://doi.org/10.1109/ACCESS.2020.3012532

3. Du, Y., Hao, H., Xing, Y., Niu, J., Calhoun, V.D.: A transdiagnostic biotype detection method for Schizophrenia and autism spectrum disorder based on graph Kernel. 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Mexico, pp. 3241–3244 (2021). https://doi.org/10.1109/EMBC46164.2021.9629618
4. Robles, M., et al.: A virtual reality based system for the screening and classification of autism. IEEE Trans. Visual Comput. Graphics **28**(5), 2168–2178 (2022). https://doi.org/10.1109/TVCG.2022.3150489
5. Haweel, R., et al.: A novel grading system for autism severity level using task-based functional MRI: a response to speech study. IEEE Access **9**, 100570–100582 (2021). https://doi.org/10.1109/ACCESS.2021.3097606
6. Shin, J., Maniruzzaman, M., Uchida, Y., Hasan, M.A.M., Megumi, A., Yasumura, A.: Handwriting-based ADHD detection for children having ASD using machine learning approaches. IEEE Access **11**, 84974–84984 (2023). https://doi.org/10.1109/ACCESS.2023.3302903
7. W. Farzana, F. Sarker, T. Chau and K. A. Mamun, "Technological Evolvement in AAC Modalities to Foster Communications of Verbally Challenged ASD Children: A Systematic Review," in IEEE Access, https://doi.org/10.1109/ACCESS.2021.3055195
8. Zhao, Z., et al.: Applying machine learning to identify autism with restricted kinematic features. IEEE Access **7**, 157614–157622 (2019). https://doi.org/10.1109/ACCESS.2019.2950030
9. Linstead, E., German, R., Dixon, Granpeesheh, D., Novack, M., Powell, A.: An application of neural networks to predicting mastery of learning outcomes in the treatment of autism spectrum disorder. 2015 IE7E 14th International Conference on Machine Learning and Applications (ICMLA), Miami, FL, USA, pp. 414−418 (2015). https://doi.org/10.1109/ICMLA.2015.214
10. Gorodetski, A., Dinstein, I., Zigel, Y.: Speaker diarization during noisy clinical diagnoses of autism. 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Berlin, Germany, pp. 2593−2596 (2019). https://doi.org/10.1109/EMBC.2019.8857247
11. Kollias, K.-F., Syriopoulou-Delli, C.K., Sarigiannidis, P., Fragulis, G.F.: Autism detection in high-functioning adults with the application of eye-tracking technology and machine learning. 2022 11th International Conference on Modern Circuits and Systems Technologies (MOCAST), Bremen, Germany, pp. 1–4 (2022). https://doi.org/10.1109/MOCAST54814.2022.9837653
12. Rasche, N., Qian, C.Z.: Work in progress: Application design on touch screen mobile computers (TSMC) to improve autism instruction. 2012 Frontiers in Education Conference Proceedings, Seattle, WA, USA, pp. 1–2 (2012). https://doi.org/10.1109/FIE.2012.6462516
13. Wang, X., et al.: Eye contact reminder system for people with autism. 6th International Conference on Mobile Computing. Applications and Services, Austin, TX, USA, pp. 160–163 (2014). https://doi.org/10.4108/icst.mobicase.2014.257796
14. Chistol, M., Turcu, C., Danubianu, M.: Autism assistant: a platform for autism home-based therapeutic intervention. IEEE Access **11**, 94188–94204 (2023). https://doi.org/10.1109/ACCESS.2023.3310397
15. D.H.R., S., M, S., Gupta, A.K, Adavala, K.M., Siddiqui, A. T. ., Shinkre,Deshpande, R., Pareek, M.P.P.: Evolutionary strategies for parameter optimization in deep learning models. Int. J. Intell. Syst. Appl. Eng. **12**(2s), 371–378 (2023). https://ijisae.org/index.php/IJISAE/article/view/3636
16. Dr Nandini, C., Mukherjee, A.B.: Smart health prediction system using machine learning techniques. Int. J. Creative Res. Thoughts (IJCRT), **10**(4), e73-e78 (2022). ISSN:2320–2882, Impact factor: 7.97 – Q4

17. Manasa Sandeep, C.N.: An extensive survey on 3D face reconstruction based on passive method. Int. Res. J. Eng. Technol. (IRJET). **8**(12) (2021)
18. Chakraborty, D., Bannerjee, R., Das, S., Das, A.: Teaching aid software — Training autistic children through computers. 2017 5th National Conference on E- Learning & E-Learning Technologies (ELELTECH), Hyderabad, India, pp. 1–6 (2017). https://doi.org/10.1109/ELELTECH.2017.8074998
19. Dr Shalini, S., Koustav Biswas, L.D., Kumar, M.A., Monika V.: A survey on smart autism support and detection hub. Int. J. Creative Res. Thoughts (IJCRT) **13**(1), d465-d469 (2025). ISSN:2320-2882

ICT Policy and E-Governance: Navigating Inter-Governmental Issues in the Digital Era

M. Shankar Lingam[1(✉)], G. S. Raghavendra[2], and Sakthi Kamal Nathan Sambasivam[1,2]

[1] Chaitanya Deemed University and University of Mysuru, Mysuru, India
shankumacharla@gmail.com
[2] CMS, JAIN (Deemed- to-be- University), Bengaluru, India

Abstract. In the digital era, the integration of ICT policies and e-Governance has emerged as a critical driver for public sector modernization. ICT policy sets the groundwork for effective e-Governance systems by enabling the digital transformation of government functions and services. However, the realization of e-Governance goals faces significant challenges, particularly in fostering inter-governmental cooperation. These challenges arise from varying policy frameworks, technological disparities, and governance structures across different levels of government. In this context, navigating the intricacies of inter-governmental relationships is essential to ensure seamless information exchange and collaborative governance. The evolving digital landscape introduces both opportunities and complexities, particularly in terms of data privacy, cybersecurity, and digital inclusion. This paper presents a comprehensive review of ICT policies and e-Governance frameworks, with a focus on overcoming inter-governmental challenges in the digital era. Our methodology includes a scoping review of key studies and case analyses, such as the work of Obi (2007) on global perspectives of e-Governance [Obi, T. (2007). E-Governance: A Global Perspective on a New Paradigm], Prasad (2012) on India's ICT policy for digital democracy [Prasad, K. (2012). E-Governance Policy for Modernizing Government through Digital Democracy in India], and Manda (2017) on South Africa's smart governance approach [Manda, M. I. (2017). Towards "Smart Governance" through a Multidisciplinary Approach to E-Government Integration]. The findings highlight the importance of fostering interoperable and inclusive e-Governance systems to navigate the inter-governmental challenges posed by ICT adoption. Key implications for policy include the need for harmonized digital policies, the development of interoperable infrastructure, and the emphasis on inclusivity to bridge the digital divide. The results underscore the need for collaborative frameworks that enable effective governance across multiple levels of government, thus ensuring that the digital transformation of public services benefits all stakeholders.

Keywords: e-Governance · ICT Policy · Inter-Governmental Cooperation · Digital Transformation and Data Privacy

S. Fong et al. (Eds.): ICT4SD 2025, LNNS 1651, pp. 112–123, 2026.
https://doi.org/10.1007/978-3-032-06688-6_11

1 Introduction

In the digital era, the rise of Information and Communication Technology (ICT) has revolutionized the way governments operate, leading to the widespread adoption of e-Governance. ICT, with its capabilities in data collection, processing, and communication, plays a pivotal role in modern governance by enabling public sector services to be delivered more efficiently, transparently, and inclusively. The evolution of ICT has spurred a global digital transformation, where governments across the world have embraced digital solutions to improve governance structures, increase citizen engagement, and enhance the decision-making process.

The digital transformation has become indispensable in the face of growing demands for more transparent, accountable, and responsive governance. Modernizing governance systems with ICT tools has moved from a theoretical possibility to a practical necessity. In this context, e-Governance refers to the use of digital platforms and technologies by government entities to provide services, facilitate communication, and improve the interaction between citizens, businesses, and the state. As a crucial aspect of digital governance, ICT policies form the backbone for the development, implementation, and management of e-Governance frameworks. The objective of this study is to understand the complexities of inter-governmental issues that arise in the digital era, specifically in the context of ICT policy and e-Governance. We seek to identify the barriers to effective cooperation between governmental bodies, explore the technological and organizational factors contributing to these barriers, and propose solutions to enhance coordination and policy alignment across multiple levels of government.

2 The Evolution of ICT Policy and E-Governance

The development of Information and Communication Technology (ICT) policies in governments has been a pivotal component in shaping modern governance frameworks. The historical evolution of ICT policies can be traced back to the early 1990s, when governments first recognized the transformative potential of technology in public administration. The initial focus was on automating administrative functions, digitizing records, and improving internal government communications. Over time, as ICT technologies advanced, the scope of these policies expanded to include public service delivery, transparency, and citizen engagement, forming the foundation for what is now known as e-Governance.

E-Governance refers to the use of digital technologies by governments to provide public services, engage citizens, and enhance the efficiency and transparency of public administration. Initially, e-Governance was focused on internal automation, with systems designed to streamline administrative processes and reduce bureaucratic inefficiencies. However, as the internet and mobile technologies became more pervasive, the concept of e-Governance broadened to include services that directly impacted citizens, such as online portals for tax filing, digital voting systems, and virtual public service delivery.

A key milestone in the evolution of e-Governance was the adoption of ICT policies aimed at promoting digital inclusion. These policies sought to bridge the digital divide by ensuring that all citizens had access to the necessary technologies and infrastructure to participate in digital governance. Governments began to recognize that for

e-Governance to be truly effective, it needed to be inclusive, accessible, and interoperable across different sectors and levels of government. The role of digital tools and platforms in e-Governance has been transformative. Various ICT tools, such as cloud computing, mobile applications, geographic information systems (GIS), and artificial intelligence (AI), have played an essential role in modernizing public administration. Cloud computing, for example, has enabled governments to scale their infrastructure quickly and cost-effectively, while mobile applications have improved access to government services for citizens, particularly in remote or underserved areas. GIS and AI have allowed for more data-driven decision-making, enhancing the ability of governments to address complex social and economic challenges.

As governments around the world embraced e-Governance, some countries became early adopters of ICT policies that laid the groundwork for modern digital governance systems. In India, the National e-Governance Plan (NeGP), launched in 2006, aimed to make government services accessible to citizens electronically, improving transparency and efficiency. Similarly, Estonia became a pioneer in e-Governance by implementing an ambitious digital government strategy, allowing citizens to access a wide range of public services through digital platforms. These early adopters have provided valuable lessons in the challenges and benefits of implementing ICT policies and e-Governance frameworks. In terms of key publications, Obi (2007) in his work *E-Governance: A Global Perspective on a New Paradigm* explores the global rise of e-Governance and its implications for public sector reform. Obi's work examines how different governments have implemented ICT policies to facilitate e-Governance, highlighting both successes and challenges. He argues that e-Governance, when properly implemented, has the potential to transform government-public relationships, improve public sector transparency, and increase citizen participation in governance processes [Obi, T. (2007). E-Governance: A Global Perspective on a New Paradigm].

Theories on ICT policy and governance models also emerged as a result of the global shift towards digital governance. Scholars like Finger and Péroud (2003) have worked on conceptualizing the relationship between e-Government and e-Governance, suggesting that while e-Government refers to the use of ICT in government administration, e-Governance goes a step further by involving citizens in decision-making processes and fostering a more inclusive, democratic governance model [Finger, M., & Péroud, G. (2003). From e-Government to e-Governance? Towards a Model of E-Governance]. This distinction between e-Government and e-Governance has been crucial in understanding the broader implications of ICT policies on governance structures.

3 Inter-Governmental Cooperation and ICT Policy

The alignment of ICT policies across multiple government sectors is a critical challenge in the digital governance landscape. Governments at different levels—local, regional, national, and international—often have disparate priorities, policies, and technical infrastructures, making cooperation and harmonization difficult. Each government entity may

have its own ICT strategy, resources, and goals, which can result in fragmented systems and a lack of coordination. This misalignment often hinders the potential for e-Governance to achieve its full impact, as effective digital governance requires integrated systems and seamless communication between agencies at all levels of government.

Inter-agency Communication and Information Sharing in e-Governance
A major barrier to effective e-Governance is the challenge of inter-agency communication and information sharing. In order for ICT-enabled governance systems to function efficiently, different government agencies must be able to share data and collaborate in real-time. However, differences in data formats, incompatible software systems, and security concerns often impede seamless information exchange. For example, one agency may have a sophisticated data analytics platform, while another may rely on outdated databases that cannot interact with the more modern systems. Such disparities create friction and limit the potential benefits of e-Governance, particularly in cross-agency initiatives that require coordinated action.

Barriers to Collaboration Between Different Levels of Government
Barriers to collaboration between local, national, and international levels of government are also significant. At the national level, governments may set broad ICT policies, but local governments often struggle to implement these policies effectively due to limited resources, political challenges, and a lack of technical expertise. Furthermore, while national governments may be focused on broad policy objectives, local governments may prioritize specific community needs that require different technological solutions.

Case Studies on South Africa's Smart Governance (2017)
South Africa's efforts to implement a smart governance model represent an interesting case study in inter-governmental collaboration in the field of e-Governance. In 2017, the South African government launched the Digital Migration Policy and Strategy, aiming to modernize public service delivery and enhance ICT infrastructure across the country. The strategy focused on increasing digital access for all citizens, improving data management, and fostering greater coordination between different levels of government. One key aspect of South Africa's approach to smart governance was the emphasis on inter-governmental cooperation. By leveraging ICT tools, the government sought to create a more interconnected and efficient public service. The implementation of a national digital identity system was a central component of this strategy, as it enabled citizens to access various government services online. However, as noted in research by Manda (2017), the process was not without its challenges. One of the significant barriers to success was the lack of uniformity in ICT capabilities across different provinces, which limited the ability of local governments to fully participate in the national e-Governance framework [Manda, M. I. (2017). Towards Smart Governance through a Multidisciplinary Approach to E-Government Integration, Interoperability, and Information Sharing]. Despite these obstacles, South Africa's experience underscores the importance of inter-governmental collaboration in e-Governance. The national government worked closely with local authorities, international organizations, and the private sector to build the necessary infrastructure and ensure that all citizens had access to essential digital

services. The country's ongoing efforts demonstrate the potential of ICT to modernize governance, even in contexts with significant resource disparities between different levels of government.

The Role of International Organizations in Promoting Digital Governance
International organizations play a vital role in facilitating collaboration between governments and promoting digital governance. Through initiatives such as the United Nations' E-Government Survey and the World Bank's ICT for Development program, international organizations provide technical expertise, policy frameworks, and funding to help countries implement e-Governance systems. These organizations also serve as platforms for knowledge exchange, allowing countries to share best practices and learn from each other's experiences. In conclusion, the challenges of inter-governmental cooperation in ICT policy are significant but not insurmountable. By addressing barriers to collaboration, fostering inter-agency communication, and leveraging the support of international organizations, governments can enhance their ability to implement effective e-Governance systems. The experiences of countries like South Africa demonstrate that, despite the challenges, a coordinated approach to digital governance can lead to improved public service delivery, greater citizen engagement, and more transparent and accountable governance systems.

4 Navigating Challenges in the Digital Era

The digital era has introduced a range of transformative technologies—cloud computing, artificial intelligence (AI), and the Internet of Things (IoT)—each of which has significant implications for governance. These technologies have the potential to enhance the efficiency, transparency, and accessibility of government services, but they also bring with them a set of challenges that need to be navigated carefully to ensure that e-Governance systems function effectively (Tables 1, 2 and 3).

Table 1. Technology Impact on Governance Challenges

Technology	*Impact on Governance*	*Challenges*
Cloud Computing	*Revolutionized public sector IT infrastructure by enabling scalable services. - Reduces operational costs. - Provides a flexible environment for data storage and processing. - Allows real-time access to services without heavy investment in physical infrastructure.*	*Complex management of vast data generated. - Ensuring data security, processing, and accessibility. - High initial investment, especially for developing countries.*

(*continued*)

Table 1. (*continued*)

Technology	*Impact on Governance*	*Challenges*
Artificial Intelligence (AI)	*Automates decision-making processes. - Enhances data analysis and predictive capabilities. - Optimizes resource allocation and detects inefficiencies. - AI-powered chatbots improve communication with citizens.*	*Potential challenges in ensuring ethical AI use.- Data privacy concerns. - Need for skilled workforce to manage and interpret AI results.*
Internet of Things (IoT)	*Enables real-time data collection from devices (e.g., traffic sensors, smart meters). - Improves urban planning, infrastructure management, and public health monitoring. - Helps regulate traffic flow, waste management, and emergency services in cities.*	*Managing and analyzing large volumes of real-time data. Ensuring data privacy and security of IoT devices. - Significant investment in infrastructure.*
Data Privacy and Security in e-Governance	*Ensuring privacy and security in a connected digital ecosystem where data is exchanged between different government levels and private entities. - Governments collect vast amounts of sensitive data (e.g., personal, financial, medical).*	*Data breaches can compromise citizens' privacy and undermine trust. - Different agencies may have inconsistent security protocols, hindering inter-agency collaboration. - Complexities of differing data protection laws across countries.*
Legal and Ethical Challenges in Cross-Border Data Flow	*As e-Government services become digitized, data flow across borders is inevitable. - Raises questions about data jurisdiction and access. - Ethical concerns around data collection, use, and consent.*	*Inconsistent data protection laws between countries, hindering international collaboration. - Concerns about data sovereignty and reluctance to share data across borders. - Potential ethical issues related to transparency and bias in AI.*
Digital Divide and Inclusive Access	*The digital divide remains a significant challenge, leaving marginalized groups without access to necessary infrastructure or digital skills. - Many citizens in developing countries lack internet and computing access.*	*Lack of infrastructure and skills in rural and underserved areas. - Difficulty in ensuring equitable access to e-Government services for all citizens, especially in remote areas.*

(*continued*)

Table 1. (*continued*)

Technology	*Impact on Governance*	*Challenges*
Global Strategies to Overcome Digital Governance Challenges	*India's Digital India initiative (2015) aims to expand digital access and improve digital literacy. - Digital Democracy policy (2012) empowers citizens through online access to government services.*	*Bridging the digital divide remains a challenge. - Ensuring all citizens, especially in rural areas, have the necessary infrastructure and skills to fully engage with e-Governance.*

5 Case Study: The LMIP Project in South Africa

The LMIP (Local Municipal Information Project) is a notable example of a multi-faceted effort aimed at enhancing e-Government capabilities in South Africa. The project sought to integrate various government agencies and local municipalities into a unified digital governance framework. A key focus of the LMIP project was the development of interoperable ICT systems that could streamline the communication and information-sharing processes between different levels of government. This case study explores the approach taken by the South African government in integrating e-Government tools and the lessons learned from its implementation.

Overview of the LMIP Project
The LMIP project was conceived to address the challenges faced by local governments in South Africa, particularly with regard to the digital divide and inefficiencies in governance. The initiative focused on enabling local municipalities to leverage technology for better service delivery, citizen engagement, and overall governance. By integrating ICT solutions into public service processes, the government aimed to create a more transparent, responsive, and accountable system that could better address the needs of South African citizens. The project utilized a multidisciplinary approach, bringing together stakeholders from various sectors, including government agencies, private sector partners, and international organizations. This collaboration allowed for the creation of a robust e-Government infrastructure that could support interoperability between different government entities. The LMIP project's focus on local government ensured that the benefits of digital governance reached communities at the grassroots level, where the need for efficient and accessible public services was most pressing.

A Multidisciplinary Approach to E-Government Integration
The LMIP project took a holistic approach to e-Government integration, recognizing the importance of involving various sectors to create a successful digital governance framework. A key element of this integration was the alignment of technological, organizational, and policy frameworks across multiple government agencies. The involvement of local government officials, technologists, and policymakers ensured that the project addressed the diverse challenges faced by municipalities, including limited technological infrastructure, lack of technical expertise, and the need for effective policy

frameworks. One of the most important aspects of the LMIP project was its emphasis on collaboration. For instance, the project involved the South African government's national e-Government framework, which provided the necessary policy guidelines and resources for local governments. Additionally, the private sector contributed by providing technological expertise and infrastructure solutions. This collaborative approach helped ensure that all stakeholders were aligned in their objectives and working towards the same goal of improving governance through technology.

Interoperability and Its Role in Facilitating Governance

Interoperability was a central theme in the LMIP project. The project's success depended on the ability of different e-Government systems to communicate seamlessly, share data, and work together across various administrative levels. The integration of disparate systems—such as financial management, citizen service portals, and public health platforms—was critical to ensuring that local governments could provide comprehensive and responsive services to citizens. To achieve interoperability, the LMIP project adopted open standards and protocols that allowed different systems to communicate efficiently. This approach enabled local municipalities to integrate their existing IT infrastructure with new digital tools, allowing for smoother data flow and better coordination among agencies. By enhancing the interoperability of government systems, the project facilitated more effective decision-making, improved service delivery, and a reduction in the administrative burden on government workers. The LMIP project yielded significant results in terms of enhancing e-Government capabilities in South Africa. The integration of ICT tools helped improve service delivery, as local municipalities were able to streamline their processes, reduce paperwork, and provide citizens with faster access to services. For example, local governments were able to implement online platforms for tax payments, permit applications, and public health services, allowing citizens to access government services more easily and efficiently. In conclusion, the LMIP project in South Africa is a valuable case study in the challenges and successes of integrating e-Government systems at the local level. Through a multidisciplinary approach and a focus on interoperability, the project improved service delivery and enhanced governance.

Table 2. Key Aspects of the LMIP Project

Aspect	*Description*
Objective	*Improve local government service delivery through e-Government integration*
Key Stakeholders	*National government, local municipalities, private sector, international partners*
Focus Area	*Interoperability, citizen engagement, transparency, and accessibility*
Technological Solutions	*Cloud computing, citizen portals, e-payment systems*
Challenges	*Digital divide, infrastructure limitations, technical skills gaps*
Results	*Improved public service delivery, enhanced coordination between agencies*

6 Proposed Solutions and Future Directions

As governments around the world strive to modernize their operations and enhance service delivery through e-Governance, significant challenges remain in ensuring effective inter-governmental collaboration, harnessing the potential of new technologies, and designing policies that will sustain digital governance frameworks. This section proposes solutions to enhance inter-governmental collaboration, explores future trends in ICT policy for e-Governance, and discusses technological advancements and policy recommendations for ensuring sustainable digital governance.

Proposals for Enhancing Inter-Governmental Collaboration Through ICT
One of the primary challenges facing digital governance is the fragmentation of ICT systems across different levels of government. In many countries, local, regional, and national governments often operate in silos, with limited communication and data sharing between agencies. To address this challenge, the following solutions are proposed:

Future Trends in ICT Policy for E-Governance
The future of ICT policy for e-Governance will be shaped by several trends that aim to enhance efficiency, transparency, and inclusivity. Some of the key trends are: AI and Automation in Public Services: Artificial intelligence (AI) and automation are set to play a crucial role in improving public sector operations. Governments will increasingly use AI to automate routine administrative tasks, such as processing tax returns, handling permit applications, and providing customer support. This will not only improve efficiency but also free up public servants to focus on more complex tasks. Smart Cities and IoT Integration: As the Internet of Things (IoT) continues to expand, governments will use smart city technologies to enhance service delivery, improve infrastructure management, and address urban challenges such as traffic congestion and waste management. ICT policies will need to support the integration of IoT devices and sensors into public administration systems, enabling real-time data collection and decision-making. Data Governance and Privacy: As governments collect more data to provide digital services, ensuring data privacy and security will be paramount. Future ICT policies will need to focus on establishing robust data protection frameworks, ensuring compliance with global standards like the General Data Protection Regulation (GDPR), and addressing concerns related to cross-border data flow. Blockchain for Transparency and Accountability: Blockchain technology is expected to become a key tool for ensuring transparency and accountability in governance. Governments will adopt blockchain to manage public records, track transactions, and reduce fraud. Policies will need to encourage the adoption of blockchain while addressing legal and technical challenges.

Technological Advancements and Their Potential for Improving Governance Systems: Technological advancements have the potential to revolutionize governance systems, making them more efficient, transparent, and citizen-centric. The following technologies are expected to play a significant role: Cloud Computing: Cloud-based solutions allow governments to scale their operations, improve collaboration, and reduce the costs associated with maintaining physical infrastructure. Governments can use cloud computing to provide citizens with on-demand access to services and information, improving the responsiveness and accessibility of public services. AI and Machine

Learning: AI and machine learning algorithms can be used to analyze large datasets, identify patterns, and make data-driven decisions. This will enable governments to optimize resource allocation, predict service demand, and improve public policy.

Policy Recommendations for Ensuring Sustainable Digital Governance Frameworks: To ensure the long-term success of digital governance, governments must adopt policies that support the integration of technology, promote innovation, and ensure that the benefits of e-Government are accessible to all citizens. The following policy recommendations are proposed: Invest in Digital Literacy and Inclusion: Governments should invest in digital literacy programs to ensure that all citizens, especially those in rural and underserved areas, can access and use e-Government services. This includes providing training on basic digital skills, as well as addressing barriers to internet access. Promote Public-Private Partnerships: Collaboration between the public and private sectors will be essential to drive innovation and improve the delivery of public services. Governments should create policies that encourage private sector involvement in the development and implementation of e-Government platforms. Ensure Cybersecurity and Data Privacy: As digital services become more widespread, the risks associated with data breaches and cyber-attacks increase. Governments must develop robust cybersecurity policies to protect sensitive data and ensure that e-Government systems are secure from external threats. Adopt Flexible and Scalable Infrastructure: ICT policies should prioritize the development of flexible and scalable digital infrastructures that can adapt to future technological advancements. This includes creating systems that can integrate emerging technologies such as AI, IoT, and blockchain. Foster International Cooperation: As e-Government systems often involve cross-border data flows and international collaboration, governments should create policies that facilitate cooperation between countries and international organizations. This includes harmonizing data protection regulations and ensuring that e-Government systems are interoperable across borders. In conclusion, the future of e-Governance will be shaped by advancements in technology, innovative policy frameworks, and greater inter-governmental collaboration. By focusing on interoperability, data governance, and inclusive digital policies, governments can create more efficient, transparent, and responsive governance systems. With the right policy decisions and strategic investments, digital governance has the potential to revolutionize public administration, enhance citizen engagement, and improve service delivery across the globe.

Table 3. Proposed ICT Solutions for Enhancing Inter-Governmental Collaboration

Solution	*Description*
Standardized ICT Infrastructure	*Implement common platforms, data formats, and open-source solutions to promote interoperability between systems.*
Shared Data Platforms	*Centralized, secure data repositories to facilitate information sharing between government entities.*

(*continued*)

Table 3. (*continued*)

Solution	*Description*
Cross-Jurisdictional Policy Alignment	*Align ICT policies across local, regional, and national levels to ensure consistency and cooperation.*
Inter-Governmental Task Forces	*Establish multi-level task forces to oversee digital governance initiatives and ensure collaboration.*

Table 4. Future Trends in ICT Policy for E-Governance

Trend	*Description*
AI and Automation in Public Services	*Use of AI for automating public sector functions such as processing tax returns and handling customer service requests*
Smart Cities and IoT Integration	*Integration of IoT technologies into urban management for enhanced infrastructure and service delivery*
Data Governance and Privacy	*Establishing robust data protection frameworks and ensuring compliance with global privacy standards such as GDPR*
Blockchain for Transparency and Accountability	*Adoption of blockchain to manage public records and reduce fraud in governmental processes*

7 Conclusion

In conclusion, this study has explored the critical role that Information and Communication Technology (ICT) plays in shaping the future of governance, focusing particularly on the challenges and opportunities in inter-governmental cooperation. The findings reveal that while ICT has the potential to significantly enhance governance by improving efficiency, transparency, and citizen engagement, it also presents a range of challenges, especially when it comes to aligning policies across multiple levels of government. Interoperability, data sharing, and the digital divide are some of the key issues that impede effective collaboration between different governmental entities. In summary, the integration of ICT into governance offers a transformative potential for improving the efficiency, transparency, and accessibility of public services. However, for e-Governance to reach its full potential, governments must prioritize inter-governmental cooperation, invest in inclusive digital infrastructure, and continually adapt ICT policies to meet the needs of a rapidly changing digital landscape.

References

Obi, T.: E-Governance: a global perspective on a New paradigm (2007)

Prasad, K.: E-Governance policy for modernizing government through digital democracy in India (2012)

Manda, M.I.: Towards smart governance through a multidisciplinary approach to e-government integration, interoperability, and information sharing: A case of the LMIP project. In: E-Government and E-Governance (pp. 55–76). Springer (2017). https://doi.org/10.1007/978-3-319-64677-0_4

Taoufik, A.O., Azmani, A.: A consolidated conceptual framework of a smart e-government ecosystem: a scoping review. Int. J. E-Gov. **24**(1), 1–15 (2025). https://doi.org/10.1504/EG.2025.143117

Domain and AI-Based Watermark Techniques for Intelligent Digital Image Forensics

Debabala Swain[1(✉)], Monalisa Swain[2], Sharmistha Roy[3], Debabrata Swain[4], Jayanta Mondal[5], and Prachee Dewangan[1]

[1] Rama Devi Women's University, Bhubaneswar, India
debabala@rdwu.ac.in
[2] Shailabala Women's College, Cuttack, India
[3] Usha Martin University, Ranchi, India
sharmistha@umu.ac.in
[4] Pandit Deendayal Energy University, Gandhinagar, India
[5] KIIT Deemed University, Bhubaneswar, India
jayanta.mondalfcs@kiit.ac.in

Abstract. The information stored or transmitted digitally is vulnerable to unauthorized access. The authentication of digital images is a critical issue in the era of digital advancements, given the ease with which any image can be altered. Consequently, methods for verifying the credibility of images are gaining widespread recognition due to their relevance in various societal domains, such as government, military, forensics, and electronic commerce. The significance of protecting images from manipulation has escalated, recognizing that even a minor tampering incident could lead to severe consequences. Hence, safeguarding images from alterations has become increasingly essential. Literature has seen the development of numerous approaches to ensure the genuineness and integrity of digital images. This study offers a comprehensive overview of both domain-based and AI-based watermark techniques for authenticating images, providing the capability to detect tampering and pinpoint the specific manipulated areas within an image.

Keywords: Copyright protection · Image watermarking · Spatial domain watermarking · Frequency domain watermarking · AI-based watermarking · Identity protection · Tamper localization

1 Introduction

The Internet has been used more and more since its start. Nearly all aspects of human existence have been profoundly impacted by the Internet, and people's reliance on it is increasing every day. The integration of image processing and the Internet has streamlined the copying, alteration, reproduction, and dissemination of digital images, providing a cost-efficient and nearly immediate delivery without sacrificing quality. The swift evolution of network technology presents a risk to the privacy and security of data. As a result, it is essential to prioritize copyright protection, content authentication, and

S. Fong et al. (Eds.): ICT4SD 2025, LNNS 1651, pp. 124–132, 2026.
https://doi.org/10.1007/978-3-032-06688-6_12

measures against duplication to effectively confront the challenges posed by present and future threats to the safeguarding of digital information. The process of image watermarking involves the incorporation of subtle information, referred to as the "watermark," into an original or host image. Subsequently, the embedded data can be retrieved to verify the authenticity of the host image [1].

After authenticating the host image, its integrity can be confirmed. The watermarks in the digital images is intended to make the image imperceptible to the unknown users so that it can not be removed. The number of watermarking techniques have been proposed with their advantages and disadvantages. This study offers both traditional approaches using spatial and frequency domains, and the AI-based approaches used in image watermarking.

2 Digital Watermarking

It is a technique to conceal digital data for authentication purpose. It is covertly inserted into digital media or signals intended for storage or transmission. The concealed information is referred to as a watermark, and the signal into which the watermark is incorporated is known as the host. This can be in the form of audio, video, images, or text. The purpose of embedding a watermark in digital media is to enable only an intended recipient to retrieve it, thereby verifying the integrity and authenticity of the received media. Usually, a secret key that both the sender and the recipient know is used in this extraction procedure. For a successful image watermarking scheme, three primary requirements need to be addressed [2]. To begin with, in the context of invisible watermarking, the incorporation of a watermark into the host image must be imperceptible to prevent any distortions that the human eye can detect. Secondly, the watermark should be resistant to unauthorized alterations. Thirdly, a watermarking technique should be able to incorporate a large watermark (s) with a robust capacity.

Digital watermarking stands out as a specialized research domain, primarily because of its prospective applications in media-related fields, including but not limited to data authentication, media forensics, device management, annotation, privacy control, copyright protection, and even in medical reports. Techniques employed in digital image watermarking are commonly categorized according to their working domains, which include spatial, frequency, or hybrid domains. In this segment, several digital image watermarking techniques are scrutinized based on their working domains, and recent research findings in this area are summarized. This analysis is expected to be beneficial for future investigations into cutting-edge watermarking methods. Numerous services, including image authentication and content protection, along with various foundational issues, remain unresolved. There is an ongoing debate on whether to use a fragile, semi-fragile, or robust watermark for authentication purposes [3, 4]. Nevertheless, the primary goal of image watermarking is to hide data within the original images to safeguard security.

2.1 Watermarking Techniques Using the Spatial Domain

This method involves incorporating watermark information into the host image within the spatial domain, following the specifications set by the owner. Many methods are used,

such as patchwork algorithms, intermediate significant bits (ISB), and least significant bit (LSB) modification algorithms. These techniques work directly with the image's initial pixel values. The watermark is added by modifying these pixel values by incorporating the watermark data. Fragile watermarks are largely used by most image authentication schemes to protect copyright [5]. This method is about inserting a particular watermark; changes to the image's content cause the watermark to change. Finding the area of the image that has been altered so depends on the ability to spot distortions in the watermark.

Abraham & Paul in 2019 [6] have introduced a spatial domain watermarking method for color images that preserves image quality and maintains perceptual color consistency. Block-based watermarking is used to enhance both image quality and resilience against attacks for authentication and recovery purposes. Two critical components, M1 and M2, play roles in ensuring that the embedded watermarks are minimally disruptive to the eye. M1 is the mask used for embedding, and M2 is the mask for adjustment. When compared to nearby pixels, the pixel modification is undetectable. Experimental findings demonstrated that the proposed method successfully recovered the watermark even after distortion of the LSBs, and the algorithm consistently maintained a high PSNR value.

Walton's contribution in [7] stands out as one of the early and notable endeavors in fragile watermarking. Their method uses the checksum methodology, which uses sum of the seven MSB to identify any tampering with the pixel. Despite its labor-intensive nature and limitations in tamper detection, being a pioneering effort has inspired subsequent works in the field.

In 2019, Prasad et al. introduced a fragile watermarking method in [8], wherein they devised an authentication watermark by combining Most Significant Bits and Hamming code. This generated watermark is subsequently embedded by means of a special block-level pixel adjustment procedure (BPAP), after being encrypted using a logistic map. This method effectively detects and locates tampering with good quality visual appearance of watermarked images. The stated figures for accuracy (ACC), false positive rate (FPR), and false negative rate (FNR) are 99.89%, 1.45%, and 0.08%, respectively.

Later on, in 2020 [9], Prasad et al. proposed another fragile watermarking scheme for tamper detection. While the watermark generation and insertion steps used in this scheme are identical to their previous work in [8], the primary distinction between the two is that the former works at the pixel level, while the latter works at the block level. As a result, this scheme achieves higher precision in tamper detection compared to the method [8]. The stated figures for ACC, FPR, and FNR values by [9] are 99.71%, 0.01%, and 0.45%, respectively.

In 2022, N.R. & R. proposed a blind image watermarking method designed for tamper localization [10]. It uses SVD and logistic mapping. This method creates eight watermark bits from each of the two non-overlapping 2 x 2 pixel blocks that make up the host image. Six MSBs of pixels are taken out and permuted by applying the logistic map in a particular block, and then an SVD operation is performed. Subsequently, these watermark bits are inserted into the two LSBs of pixels in a block. The method's vulnerability to a range of severe attacks, such as content-only, collage, vector quantization, copy-paste, text addition, noise addition, and constant feature attacks, is illustrated through experimental simulations. Additionally, the strategy has shown better accuracy as well as precision

in recognizing tampering and localization in comparison to other methods. But one significant disadvantage is that it can't repair the damaged portions.

In 2021, Hasan et al. introduced a robust watermarking method [11] based on ISB. Notably, this is tailored exclusively for greyscale images, with the authors underscoring the utilization of the black pixels in the host image for watermark embedding. Pascal's triangle is incorporated into the embedding process. To choose the best black pixels for embedding, Pascal's triangle is essential for achieving a careful balance between resilience and imperceptibility. The results of the study demonstrate that employing black pixels for embedding rather than white ones yields superior PSNR and NCC performances. Additionally, the scheme's time complexity of O(n^2) is sufficiently low for adoption in real-time applications.

2.2 Watermarking Techniques Using the Frequency Domain

While frequency-domain algorithms are more resistant to many forms of attacks, watermarking methods in the spatial domain are thought to be more brittle due to their ease of manipulation. To address these limitations, research has shifted towards watermarking in frequency domain techniques, where the watermark is concealed data in the transform space of an image. The watermark is inserted in transform domain-based approaches using frequency coefficients [12]. This makes these techniques more resistant to attacks, making them suitable for robust watermarking. For watermarking in the frequency domain, commonly used image transformations are DFT, DWT, and DCT etc. The following section discusses some of the popular and widely used approaches based on transform domains.

Ambadekar et al. [13] introduced a novel technique aimed at safeguarding copyright data by embedding it in the Discrete Wavelet Transform domain. This method proved to be more resilient to compression attacks, noise, and geometric distortions, showcasing its effectiveness in comparison to spatial domain-based image watermarking techniques that struggle with geometric attacks. Recognizing the transformative potential of the transform domain, the authors designed a new method capable of efficiently embedding a color image as watermark. Using 2-D DWT and an appropriate selection algorithm, the method entails converting the color host image and the watermarked image from the RGB model to the YIQ model independently. By protecting the system from threats like Gaussian noise and lossy compression, this strategy emphasizes security and robustness while demonstrating efficiency against a variety of assaults.

Another strategy that is described in [14] by Verma et al., uses a homomorphic transform and Singular Value Decomposition to guarantee the digital safety of an image. SVD is used to determine the unique values for watermark insertion, and the result is an invisible eight-bit grayscale watermark image. Performance evaluations utilizing measures like SSIM, PSNR, and NCC guarantee the security and invisibility of this watermarking method.

In late 2015, Singh et al. introduced a self-embedded fragile watermarking scheme [15]. This strategy operates with blocks of size 2×2. In each 2×2 block, LSBs of each of the four pixels are removed, leaving the remaining five bits to form a watermark using CRC and Hamming code. The multitasking and authentication components are combined in this created watermark. While multitasking bits can not only carry out authentication

but also convey information required to recover compromised areas, authentication bits serve to validate the host image and give detection of tampering and localization features. Restoration is the main function of multitasking bits, with authentication capability serving as a fallback if an attack compromises the real authentication bits. This approach exhibits exceptional fragility and sensitivity to minute modifications.

In a subsequent work in late 2016, another DCT-based fragile watermarking method was proposed by Singh et al. [16], which can be seen as an extension of their prior work in [15]. The new method follows the same embedding and extraction procedures as its predecessor, with the main difference being the use of three secret keys instead of six. This adjustment aims to strike a better balance in application, although three keys are still considered relatively high.

A dual watermarking technique for digital content authentication and privacy protection was presented by Hurrah et al. in 2019 [17]. This scheme presents two watermarking schemes, called Method 1 and Method 2, with the first concentrating on hybrid watermarking and the second focusing on robust watermarking. Method 1 is unique from its predecessors in that it makes use of a combination based on DWT and DCT. The results show its resistance to different types of watermarking attacks. Its performance in JPEG compression attacks is especially noteworthy, as the watermark extracted yields higher Bit Error Rate and Normalized Cross-Correlation values than those found in [18]. As an essential component of image compression techniques, the DWT-DCT combination is credited with this superiority. An analytical summary of the above discussed techniques is represented in Table 1.

Table 1. Analytical summary of the reviewed techniques

Scheme	Domain	Watermark Type	Image Type	Remarks
[6]	Spatial	Robust	Color	Robust and imperceptible, suitable for copyright protection
[13]	Frequency	Robust	Color & Grayscale	Imperceptible and robust against different geometric attacks, suitable for copyright protection and content authentication
[14]	Frequency	Robust	Grayscale	Robust against different geometric attacks, high values of NCC and MSSIM, suitable for copyright protection

(*continued*)

Table 1. (*continued*)

Scheme	Domain	Watermark Type	Image Type	Remarks
[9]	Spatial	Fragile	Grayscale	Efficient tamper detection, no recovery, suitable for content authentication
[8]	Spatial	Fragile	Grayscale	Efficient tamper detection, no recovery, suitable for content authentication
[16]	Frequency	Fragile	Color & Grayscale	Tamper detection and high-quality recovery, Suitable for content authentication and restoration of modified content
[15]	Frequency	Fragile	Color & Grayscale	Tamper detection and high-quality recovery, suitable for content authentication and restoration of modified content
[10]	Spatial	Fragile	Grayscale	Efficient tamper detection against different attacks like content removal, noise addition, copy-paste, and no Recovery. Suitable for content authentication,
[17]	Spatial Frequency	Fragile Robust	Color & Grayscale	Tamper detection and localization, copyright protection, and content authentication

3 AI-Driven Watermark Techniques

In recent years, Artificial Intelligence (AI) has gained significant attention in the field of digital image security. These systems are controlled by the AI and Machine Learning (ML) techniques to enhance traditional watermarking methods. It offers improved robustness, imperceptibility and better adaptability upon various image manipulations and attacks. By integrating AI, the watermarking systems can dynamically adjust and optimize the watermark embedding strategies.

The AI-driven watermarking system use different AI models like deep learning, neural networks, and evolutionary algorithms to embed watermarks into digital images. AI provides more flexibility and intelligence needed to address the evolutionary complexity in digital image security.

The following models are used in AI-driven watermarking systems:

a. Machine Learning Models

Supervised Learning Models are trained on labeled datasets consisting of images with or without watermarks. These models can identify patterns in the images and detect the effective embedding technique to maintain the imperceptibility of the watermark. It also ensures the system is more resilient against various transformations. In the Unsupervised Learning Models, the optimal watermark embedding strategies can be applied without the need for pre-labelled data.

b. Deep Learning Models

Convolutional Neural Networks (CNNs) are widely used in the watermarking system for content-aware watermarking. These networks can embed watermarks in those regions of the image that are likely to be altered or impacted through image manipulations. This makes the watermark more secure.

The Generative Adversarial Networks (GAN) can be used to create high imperceptible watermark by having a generator network. It creates the watermark and the discriminator network asses their quality to refine them. It ensures the watermark becomes imperceptible and difficult to remove.

c. Adversarial Machine Learning

These techniques are used to train the watermark against the adversarial attacks. These are the attacks where the attacker intentionally attempts to remove or destroy the watermark. During the training process, various types of attacks are simulated in the AI model to create more robust watermark solutions.

d. Reinforcement Learning

This model can be applied to the watermark embedding process, where the AI agent is trained to maximize the reward function during the embedding. It extracts the watermark without any perceptible distortion. This iterative learning process allows the watermark to evolve. Hence, it becomes more resilient to the new types of threats and manipulations.

4 Conclusion

Due to interactive and digital multimedia data transfer, information can currently be simply replicated. Because of this problem, digital image watermarking has become an important area of study. An essential tool for image authentication, integrity verification, tamper detection, copyright protection, and digital security is digital image watermarking, which can be achieved in several ways. This study presents a comprehensive assessment of the recent research on image watermarking for identity protection and authenticity. It discusses several well-known watermarking techniques that have influenced image watermarking studies. The key components of an effective watermarking system design are robustness, imperceptibility, and capacity. It is nearly impossible to fulfill each of these demands at the same time, though. Consequently, it's imperative to establish a fair trade-off between these three objectives. Therefore, a good trade-off between these three requirements must be maintained. Therefore, to meet the aforementioned three crucial objectives, future work might be expanded by merging diverse methodologies in different domains.

While the AI-driven watermarking systems provide many advantages still there are still some challenges, like computational complexity faced in deep learning models, security against the AI-based attacks, etc. Further, more researchers should concentrate on creating novel, cutting-edge methods in order to enhance security and robustness simultaneously.

Acknowledgement. This research work is funded by the Odisha State Higher Education Council (OSHEC), Department of Higher Education, Government of Odisha, under the Mukhyamantri Research and Innovation Fellowship Program (MRIP) to Rama Devi Women's University.

References

1. Sharma, S., Zou, J., Fang, G.: Significant difference-based watermarking in multitoned images. Electron. Lett. **56**(18), 923–926 (2020)
2. Swain, M., Swain, D., Paikaray, B.K.: A competitive analysis on digital image tamper detection and its secure recovery techniques using watermarking. Advances in Intelligent Systems and Computing, vol 1101. Springer (2020)
3. Pal, P., Jana, B., Bhaumik, J.: An image authentication and tampered detection scheme exploiting local binary pattern along with hamming error correcting code. Wireless Pers. Commun. **121**(1), 939–961 (2021)
4. Singh, O.P., Singh, A.K., Srivastava, G., Kumar, N.: Image watermarking using soft computing techniques: a comprehensive survey. Multimedia Tools Appl. **80**(20), 30367–30398 (2020)
5. Yeung, M.M., Mintzer, F.: An invisible watermarking technique for image verification. Proceedings of International Conference on Image Processing, Santa Barbara, CA, USA, vol. 2, pp. 680–683 (1997)
6. Yeung, M.M.: Invisible watermarking for image verification. J. Electron. Imaging **7**(3), 578 (1998)
7. Walton, S.: Image authentication for a slippery new age. Dr. Dobb's J. **20**(4), 18–26 (1995)
8. Prasad, S., Pal, A.K.: A tamper detection suitable fragile watermarking scheme based on novel payload embedding strategy. Multimedia Tools Appl. **79**(3–4), 1673–1705 (2019)

9. Prasad, S., Pal, A.K.: Hamming code and logistic-map based pixel-level active forgery detection scheme using fragile watermarking. Multimedia Tools Appl. **79**(29–30), 20897–20928 (2020)
10. N.R., N.R., R., S.: Fragile watermarking scheme for tamper localization in images using logistic map and singular value decomposition. J. Vis. Commun. and Image Representation **85**, 103500 (2022)
11. Hasan, M.K., et al.: An improved watermarking algorithm for robustness and imperceptibility of data protection in the perception layer of internet of things. Pattern Recogn. Lett. **152**, 283–294 (2021)
12. Sharma, S., Zou, J., Fang, G.: A single watermark based scheme for both protection and authentication of identities. IET Image Process. **16**(12), 3113–3132 (2022)
13. Ambadekar, S.P., Jain, J., Khanapuri, J.: Digital image watermarking through encryption and DWT for copyright protection. Recent Trends in Signal and Image Processing. Advances in Intelligent Systems and Computing, vol 727. Springer, Singapore (2019)
14. Verma, D., Aggarwal, A.K., Agarwal, H.: Watermarking scheme based on singular value decomposition and homomorphic transform. AIP Conf. Proc. **1897**(1), 020036 (2017)
15. Singh, D., Singh, S.K.: DCT based efficient fragile watermarking scheme for image authentication and restoration. Multimedia Tools Appl. **76**(1), 953–977 (2015)
16. Singh, D., Singh, S.K.: Effective self-embedding watermarking scheme for image tampered detection and localization with recovery capability. J. Vis. Commun. Image Represent. **38**, 775–789 (2016)
17. Hurrah, N.N., Parah, S.A., Loan, N.A., Sheikh, J.A., Elhoseny, M., Muhammad, K.: Dual watermarking framework for privacy protection and content authentication of multimedia. Futur. Gener. Comput. Syst. **94**, 654–673 (2019)
18. Loan, N.A., Hurrah, N.N., Parah, S.A., Lee, J.W., Sheikh, J.A., Bhat, G.M.: Secure and robust digital image watermarking using coefficient differencing and chaotic encryption. IEEE Access **6**, 19876–19897 (2018)

Toxic Hinglish Comment Detection

Gopal D. Upadhye(✉), Deepak T. Mane, Devang Gentyal, Chetan Channa, Shubham Landge, and Radhika Gadewar

Vishwakarma Institute of Technology, Pune, India
{gopal.upadhye,deepak.mane,devang.gentyal24,chetan.channa24,
shubham.landge24,radhika.gadewar24}@vit.edu

Abstract. Toxic comment identification in Hinglish (a combination of Hindi and English) is a difficult task because of code-switching, transliteration, and class imbalance. This paper suggests a machine learning based method for identifying toxic Hinglish comments based on TF-IDF feature extraction along with an ensemble model. In order to mitigate class imbalance, Random Oversampling was utilized, and model interpretability was facilitated using SHAP (Shapley Additive Explanations). The suggested model was trained on publicly released datasets, with 90.0% accuracy compared to individual classifiers. This work contributes to content moderation system for code-mixed languages and offer an extensible solution for social media toxicity detection.

Keywords: Hinglish · toxic comment detection · machine learning · NLP · TFIDF · ensemble learning

1 Introduction

The rapid growth of social media platforms has been fuelled by an explosion of user generated content, as well as a startling proliferation of hateful and malicious posts. This internet hate is most typically found on multilingual websites where users use code-mixed languages like Hinglish, or the colloquial mix of Hindi and English. Hinglish presents new linguistic challenges due to arbitrary code-switching, inconsistency in transliteration, and blending of syntactic rules of both languages. Traditional NLP models trained on monolingual data fail to capture such subtleties, leading to incorrect toxicity classification. These problems need a new type of approach that extends conventional machine learning methods with current transformer-based models. Our work suggests an ensemble-based model that combines Logistic Regression, Random Forest, and Gradient Boosting Classifier models. All these models bring their own unique strengths: Logistic Regression to handle high-dimensional data, Random Forest to identify non-linear relations, and Gradient Boosting to iteratively correct errors. The base models are then combined using a soft voting process to generate strong final predictions, overcoming individual classifier weaknesses. Our model uses publicly available Hinglish toxicity datasets, e.g., Multilingual-Abusive Comment-Detection and IITD Offensive, to train and test the model. Preprocessing is done using text normalization, tokenization, and

S. Fong et al. (Eds.): ICT4SD 2025, LNNS 1651, pp. 133–142, 2026.
https://doi.org/10.1007/978-3-032-06688-6_13

character-level TF-IDF feature extraction to enable the model to learn complex linguistic patterns. Class imbalance is also addressed using Random Oversampling to avoid biasing the model towards the dominant non-toxic class. This work is unique because of its hybrid strategy, balancing simplicity and performance. Transformer models such as BERT and XLM-RoBERTa perform well in contextual comprehension but are computationally intensive and time-consuming in real-time applications. Our framework, through the strategic combination of traditional ML models and transformers, is able to achieve high accuracy and F1-scores, proving the viability of employing ensemble learning for real-world Hinglish toxicity detection. In the coming sections, we discuss the literature survey, detailed methodology, findings, and future scope, indicating the effectiveness and potential of the system for largescale applications.

2 Methodology

2.1 Research Design

The system follows a sequential pipeline for toxicity detection, integrating traditional machine learning models with transformer-based architectures. The methodology is divided into the following steps:

1. **Data Collection:** Publicly available datasets, such as the Multilingual-Abusive-Comment Detection and IITD Offensive datasets, were collected. These datasets consist of Hinglish comments labelled as toxic or non-toxic, covering diverse linguistic patterns.
2. **Pre-processing:** Text data was cleaned by removing special characters, URLs, and unnecessary symbols. Tokenization and normalization were applied to standardize the text. Class imbalance was handled using Random Oversampling to balance toxic and non-toxic comments.
3. **Feature Extraction:** Character-level TF-IDF was used to convert text into numerical vectors. Character n-grams (1 to 6) were extracted to capture subtle linguistic features (Fig. 1).

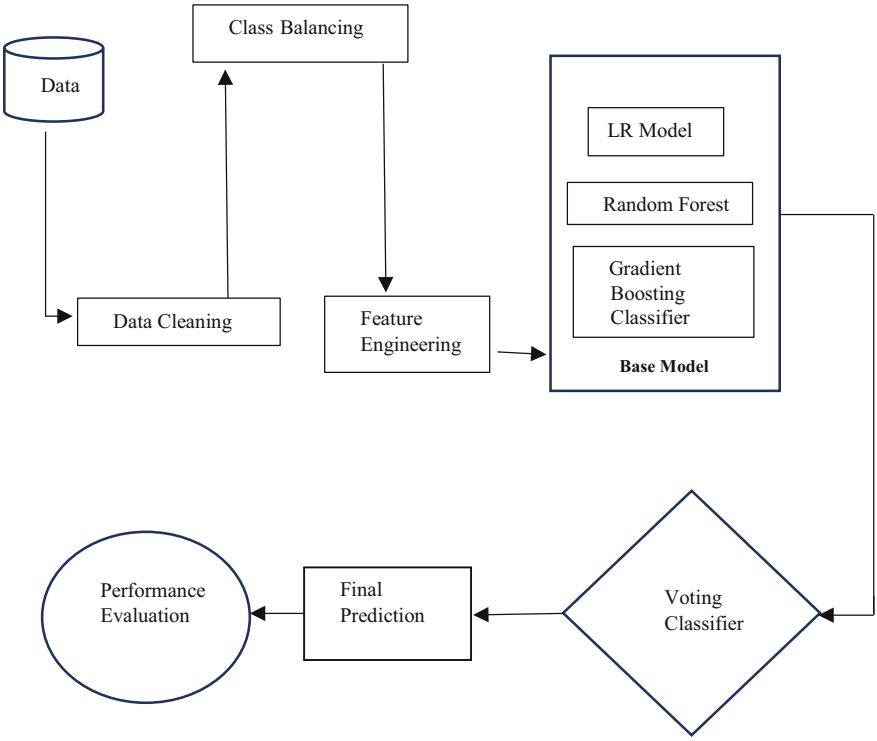


Fig. 1. Proposed System Architecture

2.2 Datasets Used

To train and evaluate our Hinglish Toxic Comment Detection model, we used publicly available datasets containing code-mixed toxic comments. The datasets used are:

- Multilingual-Abusive-Comment-Detection Dataset – Contains labelled abusive comments in Hinglish.
- IITD Offensive Language Dataset - A dataset focused on offensive Hinglish text from social media.
- Hinglish-code-mixed-dataset - Contains codemixed dataset of Toxic comments in Hinglish Language.
- TRAC Code-Mixed Dataset – A dataset containing code-mixed toxic comments.

2.3 Data Preprocessing

For improving model performance, text data were pre-processed following these steps:

- Text Cleaning: Special character, URL, emoji, and additional space removal.
- Normalization: Unifying various Hinglish transliterations into one common format.
- Tokenization: Sentence breakdown into tokens in order to represent linguistic structures.

- Stop word Removal: Eradication of frequently occurring words that do not support toxicity classification.
- Handling of Class Imbalance: Random Oversampling was used to balance the data and avoid bias towards the non-toxic class.

2.4 Features Extraction Using TF-IDF

So In order to transform text data into numerical form, we gave weights to words according to their significance to a document. This facilitated the representation of textual information in a structured way so that the model could effectively analyse and process it.

$$\text{TF} - \text{IDF}Formula(\text{t, d}) = \text{TF(t, d)} * IDF(t)$$

Equation 1 TF-IDF

2.5 Model Implementation

We used an ensemble learning strategy that merged three machine learning models:

- Logistic Regression (LR) – Good for high-dimensional text data.
- Random Forest (RF) – Identifies non-linear relationships.
- Gradient Boosting Classifier (GBC) – Better classification iteratively.

After retrieving outputs from each model, we have used Soft Voting Classifier to combine the predictions by averaging the probabilities assigned to each class, thereby enhancing the overall accuracy and robustness of our toxic comment detection model.

- Soft Voting Classifier: A linear classifier that predicts probabilities using the logistic (sigmoid) function. The model estimates the probability of a comment being toxic as:

$$P_c = \frac{1}{N} \sum_{k=0}^{n} Pk, c$$

Equation 2 Soft Voting Classifier
where:

- PcP_cPc = Final probability of class c
- N = Number of classifiers
- Pk,cP_{k,c}Pk,c = Probability of class c predicted by classifier k

The training process was performed using 5-fold cross-validation, ensuring generalization to unseen data.

2.6 Pseudocode

Input: Hinglish text dataset with labelled comments.
Output: Predicted class labels (toxic or non-toxic).

1. Load dataset from multiple sources (HASOC, IITD, ShareChat, etc.)
2. Perform data cleaning:
   a. Remove noise (special characters, stopwords, punctuation)
   b. Normalize text (lowercasing, tokenization)
3. Apply Class Balancing Using Random Oversampling
4. *Perform feature engineering.*
5. *Splitting dataset for train and test (e.g., 80%-20%)*
6. *Initialize base classifiers:*
   a. *LR Model*
   b. *RF Model*
   c. *GBC Model*
7. *Train each base classifier on the training data*
8. *Construct a Soft Voting Classifier with {LR, RF, GBC}*
9. *Train the Voting Classifier on the training data*
10. *Obtain final predictions from the Voting Classifier*
11. *Evaluate model performance*
12. *Return evaluation results and final predictions*

2.7 Model Evaluation

The system's performance was evaluated in terms of significant classification metrics. The metrics provides a general idea about the model's performance for the datasets for correct outputs.

- Accuracy – calculates how many comments are correctly identified as toxic compared to the overall comments in the dataset.
- Precision - computes the ratio of accurately marked toxic comments to the identified toxic comments and utilizes it as an estimate for false positives.
- Recall - shows how many actual toxic comments were actually flagged by the model, and measurement of false negatives is possible.
- F1-Score is referred to the harmonic average of prec. & rec., given equal weight to both parameters for a reliable performance measure.

In order to provide strong evaluation and reduce overfitting, a 5-fold cross-validation technique was employed, wherein the results were averaged to improve generalization for unseen data.

3 Results and Discussion

The performance of our system was evaluated on the Multilingual-Abusive-Comment-Detection (nsfw_train) dataset. Before presenting our model's performance, we first analyse the outputs of previous models used on this dataset. And at the last we highlighted the performance of our proposed model (Table 1).

Table 1. Models Results

Models	Acc.	Prec.	Recall	F1-S.
LR	0.870	0.862	0.901	0.881
RF	0.892	0.879	0.914	0.896
GBC	0.888	0.880	0.905	0.892
MuRIL Bert	0.872	0.878	0.844	0.856
XLM RoBERT	0.872	0.869	0.857	0.859
XLM RoBERT with emoji embedding	0.877	0.877	0.858	0.864
Proposed Model	**0.900**	**0.894**	**0.923**	**0.908**

The minor performance gain of XLM-RoBERTa with emoji embedding (F1-score = 0.864) owes to its capacity to handle emotional tone using visual semantics. Emojis were used as separate tokens and embedded together with words to facilitate better context interpretation, particularly in sarcasm or informal language. But even though the improvement was made, the increase in complexity and computation cost rendered it less preferable for real-world utilization compared to our suggested ensemble model.

3.1 Comparison with Existing Models

Relative to transformer-based models, our ensemble-based method is a balance between computational cost and accuracy. Transformers such as MuRIL BERT and XLM-RoBERTa are computationally intensive and are challenged by class imbalance, tending to favor non-toxic predictions. Our method, which combines TF-IDF with character n-grams and an ensemble learning method (Logistic Regression, Random Forest, and Gradient Boosting Classifier), offers similar performance but is more computationally cost-effective.

3.2 Impact of Feature Engineering and Ensemble Learning

To verify the contribution of different components in the model, we conducted an ablation study by removing some parts:

1. Without TF-IDF Character n-grams → A drastic fall in recall was seen, showing the significance of picking up linguistic patterns in Hinglish.
2. No Random Oversampling → The model preferred non-toxic comments and had lower recall and a greater false-negative rate.
3. Without Soft Voting Ensemble → Utilizing individual classifiers rather than ensemble learning yielded poorer accuracy and unreliable performance across datasets.

These results validate that feature engineering, class balancing, and ensemble learning are essential in enhancing Hinglish toxicity detection.

3.3 Generalization on Unseen Data

To check the generalization capacity of our model, we tested it on an entirely unseen subset of the Multilingual-Abusive-Comment dataset. This data was kept out of training and validation processes. The model proposed here got an F1-score of 0.87 on this unseen test set, which, even though slightly below the validation score, still indicates good generalization capacity. It shows that the model works effectively on actual Hinglish toxic content even when it faces new patterns.

3.4 Error Analysis and Limitations

Although our model demonstrates strong overall performance, it does have certain limitations that impact its accuracy in specific scenarios. One of the key challenges lies in misclassifications arising from sarcasm, implied toxicity, and nuanced Hinglish expressions. For instance, comments that rely on indirect language, cultural references, or subtle insinuations may be difficult for the model to accurately classify as toxic. Additionally, extensive transliteration variations pose challenges, as the same phrase can appear in multiple Romanized forms, leading to inconsistencies in detection. Furthermore, the model may struggle with code-mixed text that incorporates elements from multiple languages, making it difficult to distinguish between neutral and offensive content. It also faces challenges in detecting toxicity embedded within humor, wordplay, or context-dependent language shifts, where meaning changes based on conversational flow. Bias in training data could also contribute to classification errors, particularly when handling diverse dialects, regional slang, or new linguistic patterns.

For instance, the model misclassified the comment: ‘Oh wow, you’re so smart ’ as non-toxic due to the sarcastic tone being hard to capture using n-gram-based TF-IDF. Another example includes ‘chup reh loser’, where informal transliteration (‘chup reh’) made toxicity harder to detect. These examples highlight challenges in sarcasm, transliteration, and cultural nuance.

4 Conclusion

This work effectively created an ensemble-based system for toxic comment detection in Hinglish, overcoming the special linguistic difficulties of code-mixed text. As discussed in the Introduction, the code-switching, transliteration, and class imbalance of Hinglish make toxicity detection challenging. The system proposed, combining TFIDF feature extraction with machine learning algorithms and using soft voting, showed considerable improvement over single classifiers. The Discussion and Results section emphasized the strength of the model, with the ensemble having an F1-score of 0.908, surpassing state-of-the-art methods. Random Oversampling was effective in addressing class imbalance, and interpretability methods such as SHAP values gave insights into the decision-making process of the model Future research might investigate using transformer models such as MuRIL or XLM-RoBERTa to provide better contextual awareness. Also, methods such as SMOTE or Focal Loss might be used to further optimize class balance, and real-world deployment might gain from the use of continuous learning systems to adjust

to changing language patterns. Overall, this study offers a pragmatic, scalable solution for Hinglish toxicity detection with good potential for use in content moderation, social media monitoring, and online community management. In order to counter the changing nature of internet language, future research can include ongoing learning pipelines or community feedback loops to allow the system to learn new slang, changing toxic words, and changing linguistic patterns in Hinglish.

Acknowledgement. We would also like to put on record our sincere thanks to Prof. Gopal D. Upadhye, Dept. of AI&DS, VIT, Pune, for his valuable support, suggestions, and encouragement in the research. We also convey our gratitude to the staff and faculty members of Vishwakarma Institute of Technology, for offering the required infrastructure and resources to carry out this research. Our colleagues are thanked especially, for offering precious suggestions and feedback that improved the methodology.

Author's Contribution. The contributions of each author to this research are detailed according to the CRediT taxonomy:

Names	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P
G. D. Upadhye	✓	✓		✓			✓					✓	✓
D. Mane	✓		✓	✓		✓			✓				✓
D. Gentyal	✓		✓	✓	✓			✓		✓			✓
C. Channa		✓	✓			✓			✓		✓		
S. Landge	✓		✓	✓			✓			✓			
R. Gadewar	✓		✓		✓				✓				✓

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration

Conflicts of Interests. We confirm that we don't have any conflict of interests. Further, the study was conducted in good faith, and the results reported are the accurate reflection of the findings of the study. Ethical principles in re-search have been employed by us, and no external force has been applied on the de-sign, execution, or interpretation of results.

References

1. Aken, B.V., et al.: Challenges for toxic comment classification: an in-depth error analysis (2018)
2. Bansal, V., Tyagi, M., et al.: A transformer-based approach for abuse detection in code-mixed Indic languages. ACM Trans. Asian & Low-Resource Lang. Inf. Process (2022)
3. Baruah, A., Das, K., Barbhuiya, et al.: Aggression identification in Eng., Hin., & Bangla text using BERT, RoBERTa & SVM. Proc. 2nd Workshop on Trolling, Aggression & Cyberbullying, pp. 76–82 (2020)
4. Chakraverty, Y., Kaintura, A., Kumar, B., Khanna, A., Sharma, M., Pareek, P.K.: Analysing BERT for toxicity analysis. Int. Conf. Innov. Comput. & Commun., pp. 653–661. Springer (2023)
5. Chanda, S., Ujjwal, S., et al.: Fine-tuning transformer-based models for hate speech & offensive content detection in Eng.-Indo-Aryan & code-mixed (Eng.-Hin.) langs. FIRE (Working Notes), pp. 446–458 (2021)
6. Farooqi, Z., Mustafa, S., Ghosh, S., Shah, R.: Leveraging transformers for hate speech detection in conversational code-mixed tweets (n.d.)
7. Georgakopoulos, S.V., Tasoulis, et al.: CNNs for toxic comment classification. Proc. 10th Hellenic Conf. AI, pp. 1–6 (2018)
8. Gladwin, I., Renjiro, E.V., Valerian, B., Edbert, I.S., Suhartono, D.: Toxic comment classification using BERT & SVM. Proc. 8th Int. Conf. Sci. & Tech. (ICST), vol. 1, pp. 1–6. IEEE (2022)
9. Gupta, S., Goel, M., Rathee, N.: ML approach for toxic comment classification on social media. Proc. Int. Conf. Data Sci. & Appl. (ICDSA 2021), vol. 2, pp. 439–449. Springer (2022)
10. Gupta, V., Roychowdhury, S., Das, M., et al.: Multilingual abusive comment detection at scale for Indic languages. Adv. Neural Inf. Process. Syst. **35** (2022)
11. Jhaveri, M., Ramaiya, D., Chadha, H.S.: Toxicity detection for Indic multilingual social media content (n.d.)
12. Khan, M.A.: Determining toxic comments & mitigating unintended model bias using DL (n.d.)
13. Kumar, A., Saumya, S., Singh, J.P.: An ensemble-based model for sentiment analysis of Dravidian code-mixed social media posts. FIRE (Working Notes), pp. 950–958 (2021)
14. Kumar, A.A., Pati, P.B., Deepa, K., Sangeetha, S.T.: Toxic comment classification using S-BERT vectorization & RF algorithm. Proc. 2023 IEEE Int Conf. Contemp. Compute. & Commun. (InC4), vol. 1, pp. 1–6. IEEE (2023)
15. Madhu, H., Satapara, S., Modha, S., Mandl, T., Majumder, P.: Offensive speech detection in conversational code-mixed dialogue on social media: a contextual dataset & benchmark experiments. Expert Syst. Appl. **215**, 119342 (2023)
16. Seliverstov, Y.A., Komissarov, A.A., Poslovskaia, E.D., Lesovodskaya, A.A., Podtikhov, A.V.: Detecting low-toxic texts using a modified XLMRoBERTa neural network & toxicity confidence parameters. Proc. 2021 XXIV Int. Conf. Soft Comput. & Meas. (SCM), pp. 161–164. IEEE (2021)
17. Singh, R., Kashyap, R., Sharma, V.: Toxic comment analysis using BERT: a DL approach for toxicity detection. Proc. 2023 2nd Int. Conf. Informatics (ICI), pp. 1–6. IEEE (2023)
18. Tarun, V.G., Ramkumar, S., Ramar, G., Rajagopal, M., Sivaraman, G.: Exploring BERT & Bi-LSTM for toxic comment classification: a comparative analysis. Proc. 2024 2nd Int. Conf. Data Sci. & Intell. Syst. (ICDSIS), pp. 1–6. IEEE (2024)

19. Thilagavathy, A., Deepa, R., Lalitha, S.D., Ramya, R., Ramya, M., Rithika, L.: Semantic-based toxic comment classification using ensemble learning. E3S Web. Conf. (n.d.)
20. Yadav, A., Garg, T., Klemen, M., Ulcar, M., Agarwal, B., Robnik Sikonja, M.: Code-mixed sentiment & hate speech prediction (n.d.)

Real-Time Stock Forecasting and User Verification Using Azure AI Services

V. Kalyanasundaram[1], A. J. Keerthi[1], R. K. Krishnaa[1], A. Thirumurugan[2], and Joshua Sunder David Reddipogu[1(✉)]

[1] School of Computer Science and Engineering, Vellore Institute of Technology, Chennai Campus, Tamil Nadu, India
joshua.sunder@vit.ac.in

[2] Department of Computer Science & Engineering, Sri Sairam Engineering College, Chennai, India

Abstract. The volatility of the stock market offers a big challenge to traders depending on timely and correct information for informed decision-making. Security risks, including fraudulent practices and theft of identity, threaten online trading platforms. The present paper presents an AI-based stock trading app that tackles these issues by using predictive analytics with robust security features. The system also employs Azure AutoML to work through historical stock data, identify market trends, and generate livestock predictions to enable traders to respond proactively to fluctuations. For security reasons, the app employs Azure Document Intelligence for live Know Your Customer (KYC) verification to ensure that only valid users have access. Additionally, the platform automates document processing using AI-powered text extraction, minimizing errors from manual input and increasing efficiency. Developed with Flutter for smooth cross-platform use and backed by Azure cloud infrastructure for scalability and dependability, this software solution offers an intelligent, secure, and user-friendly trading experience. Through the integration of AI-based forecasting with robust security measures, this work helps develop more efficient, reliable, and technologically sophisticated stock trading platforms.

Keywords: Stock Market Prediction · AI-Powered Trading · Azure AutoML · Real-Time KYC Verification · Fraud Prevention · Text Extraction · Cloud-Based Trading Platform

1 Introduction

Stock markets are very volatile and dynamic where investors have no choice but to depend on exact predictions and instantaneous data in their quest to maximize the most opportune investment returns. Traditional approaches to trading typically lack the necessary agility to move with the incredibly fast-changing trends in the marketplace, resulting in subpar performances. Furthermore, online trading platforms are also confronted with severe security issues, such as identity theft and unauthorized access, which compromise the

S. Fong et al. (Eds.): ICT4SD 2025, LNNS 1651, pp. 143–153, 2026.
https://doi.org/10.1007/978-3-032-06688-6_14

security of online trading. To overcome these problems, combining artificial intelligence (AI) [1] with safe authentication systems has become imperative for improving trading efficiency and user confidence. This paper introduces an AI-based stock trading app that takes advantage of state-of-the-art machine learning techniques and cloud security features to deliver real-time stock forecasts while ensuring a safe trading platform.

The system to be proposed uses Azure AutoML to process past stock data, identify market trends, and provide highly accurate predictions of stock prices. This predictive feature allows traders to make informed decisions, minimizing risks of market volatility. In addition, the application includes Azure Document Intelligence for real-time Know Your Customer (KYC) authentication, allowing only valid users to access the platform. In addition, AI-powered text extraction from images makes document submission easier, minimizing human error in input and making user onboarding more streamlined. [2] Designed using Flutter for platform compatibility and being hosted on cloud infrastructure in Azure to provide scalability and security, the software technology is a technology innovation in stock trading. By combining AI prediction features with high-security features, the system renders users more consistent and re-designs the traditional trading experience into a healthier and more efficient one.

2 Literature Review

In this section, the reference to the new advancements in Real time stock prediction within the paper means the implementation of Azure AI and related methodologies. AI In stock prediction El Mahjouby et al. (2024) find that it is difficult to predict stock-market movement in AI, and the use of AI in finance is important, but the complex model of deep-learning is the only valid for complex time-series data. All the architectures analyzed in the study were focused on market prediction and used various ML algorithms and cryptocurrency-based datasets [3]. Li et al. (2024) proposes a Hierarchical Decomposition based Forecasting Model (HDFM) based on considering stock market volatility and nonlinearity to improve the accuracy of stock price predictions [4]. They propose an hybrid model that combined GRU, CEEMDAN with VMD, therefore delivery even more précised performance on all stock indices by utilizing decomposed approaches, high frequency forecasting on subsequence is tough task. Current AutoML frameworks still struggle to identify training sets of a given intervention factors and to search for optimal counterfactual predictions for given causal impact analytics use case, making them less practical with increasing degrees of freedom.

Hu et al. (2017) on AutoML address the limitations on the basis of applicability of AutoML in causal impact analytics [5]. The aim of AutoML in Financial Forecasting They have successfully applied AutoML pipeline on the Spark ML to the stock market policy evaluation and showed the effect of causal are important in financial analytics domain that brings reasoning ability on various metrics used in their domain.

With big data analytics in correlation and predictive modeling, we could validate AutoML features. Verma et al. (2023) talks about how electronic KYC slims down this verification process and reduces fraudulent activities thereby making compliance and security document related kyc easy. Recent Advances for Multi-Page Document Verification A deep-learning based approach was introduced to extract the classification and

text in multi-page documents. They work on Optical Character recognition, Convolutional Neural networks-based applications for the insurance domain with a high level of accuracy in extracting structured data from documents necessary for financial compliance and fraud detection and control. The second study associated with an application of transformers for boiler-plate identification in financial documents [6], This model is trained with NLP-based methods and one of the benchmarks it performs at is 84.1 accuracy that show its capability in automated compliance and fraud detection.

Singh et al. (2024) introduces a document classification and information extraction system based on deep learning for financial documents. [7] They developed a solution based on OCR, Natural Language processing, and Convolutional Neural Networks and were able to reach 90% accuracy on unstructured data contained in financial documents. Training uses data up until October 2023.

Toprak and Turan (2024), approached for verifying consistency between summary documents and their associated full-text financial documents in the Reuters financial dataset, their system obtains 84.1% accuracy. [8] Their results illustrate how AI could be applied to ensure accurate regulatory reporting and verify financial documents.

Ponnusamy et al. (2024) address novel data partitioning methods due to their importance in designing and deploying domain of cloud enabled applications with emphasis on financial applications. Their study points out the need for scalable and distributed processing methods that can enhance performance of real-time financial analytics [9] via better load balancing and system performance.

Song et al. (2022) followed the same trend, we referred to above, the authors in Research on Multidimensional Trust Evaluation Mechanism of FinTech Based on Blockchain, focuses on introducing the third metric, which includes waiting on existing two-dimensional algorithm and developing them to three-dimensional models. [10] Its multi-dimension trust index creates trust between users, prevents fraud, reduces disputes, and provides secure transaction ability across the entire fintech industry. This result shows that economic safety could be even improved practically.

3 Proposed Methodology

Research was structured on Real Time Stock Forecasting and User Verification in Azure AI Services. There are three broad stages in this process, namely System Architecture Design, Data Processing and Predictive Modelling, and User Verification and Interface Development. Consider the below phases of the use case together with the AI cloud service offerings, the machine learning techniques, and cross platform technologies that are leveraged to achieve project objectives [11], such as accurate predictions of East coast stocks and secure access for users along with intuitive user experience. Taking about each of these stages and their corresponding information as relates to the outcomes of the uploaded document will be discussed in the next sections following the provided diagrams.

4 System Architecture Design

The system Real-Time Stock Forecasting and User Verification uses Microsoft Azure's scalable, secure cloud services to accommodate the dynamic needs of stock trading with strong security and accessibility. [12] The end points of the mobile and the web get stock predictions and validation results based on the request generated by the Flutter frontend (TrackRiz) on Azure App Service backend, as depicted in Fig. 1. This could be in Azure DB Storage with user data and historical market data open, high, low, adjusted close, volume which feed back and forth for predictive modelling. Using real-time Stock Market API data, Azure AutoML predicts trends, and Azure Document Intelligence provides KYC compliance, demonstrating the integrated power of Azure cloud.

5 Data Processing and Predictive Modelling

The heart of the system is stock forecasting based on data analysis of historical data harvested in Azure DB Storage and real-time Stock Market API feeds using Azure AutoML, which predicts market trends traders are after.[13] The data pipeline pre-process datasets impute missing values, normalize data, and so on, prior to training predictive models. These models consider historical trends and real-time data and output their results using Azure App Service. The performance gets monitored, and the [14] models retrain after shifts in the market. Insights from Azure OpenAI is then rendered in readable summaries in the Flutter frontend, interpretable trends for the traders, and non-technical traders with a glimpse in their trading style are presented in Fig. 1.

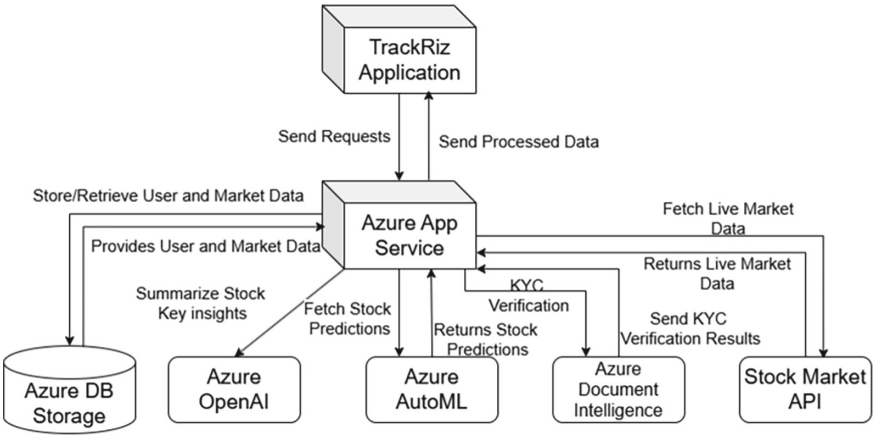


Fig. 1. The Architecture of TrackRiz Application

6 User Verification and Interface Development

By integrating Azure Document Intelligence for real-time knowledge your customer (KYC) verification, it empowers its system with ID authentication, keystroke analysis, and more, combating identity theft and unauthorized access while ensuring a seamless trading experience for users. Flutter-based frontend [15] allows users to scan their

identification documents, and Azure Document Intelligence extracts, cross-checks the scanned data with government records, and provides access to corresponding data only to those verified users above 18 years of age, thus complying with the regulatory standards, while unverified users are prompted to retry. With the Flutter UI supports accessibility and ease of use, KYC verification is made easy and the Azure AutoML generated stock risk predictions are delivered visually in an interactive, readable way so that users could analyse the status of the market and have a proper trading strategy in a safe, efficient and convenient environment.

7 Results and Discussions

The Real-Time Stock Forecasting and User Verification using Azure AI Services uses sophisticated machine learning algorithms and cloud security capabilities to make trading more efficient and confident for users. The system has Azure AutoML for the prediction of stock prices, Azure Document Intelligence for safe KYC authentication, and a cloud infrastructure scalable on Microsoft Azure. Provide below the outcomes of the performance of the system on its major features with graphical representations.

Table 1. Cloud Service Deployment Across Azure Regions

Service	Region
Container Registry	Central India
App Service	Sweden Central
Document Intelligence	Sweden Central
Key Vault	Central India
Storage Account	West US 2
Machine Learning Workspace	Central India

The geographical location of the various cloud services utilized in the research presented in Table 1, spread across various Azure regions for performance and regulatory reasons. Container Registry, Key Vault, and Machine Learning Workspace are placed in Central India to enable secure storage, model training, and key management within the same region to reduce latency. The App Service and Document Intelligence are held in Sweden Central perhaps for regulatory reasons or closeness to customers. Storage Account is held in West US 2 possibly for economies or redundancy or compatibility. Such strategic location maximizes performance, security, and economy globally across the different Azure Data Centers.

Table 2. Invoice Management Functionality

Item Description	Quantity	Price	Total
Ski Essential Package	1	$80.00	$80.00
Additional Ski Rental	1	$10.00	$10.00
Boots Rental	1	$10.00	$10.00
Poles Rental	1	$8.00	$8.00
Sub Total			$108.00
Tax (8%)			$9.60
Grand Total			$117.60

Invoice management is critical for monitoring financial transactions, and this tool utilizes AI-powered text extraction to prevent human error. Table 2 Faucet Ski Service Sample Invoice shows a sample invoice for ski rental services of $120.00 broken down as Ski Essential Package ($80.00), Extra Ski Rental ($10.00), Boots Rental ($10.00), and Poles Rental ($8.00), accompanied by merchant information and a date of transaction of 31 October 2023. Figure 3b Structured Invoice Data Table also showing the same invoice in the structured format table with precision and effective data entry. The dual view elevates usability as it can readily manage the financial records and enhances overall efficiency in operations.

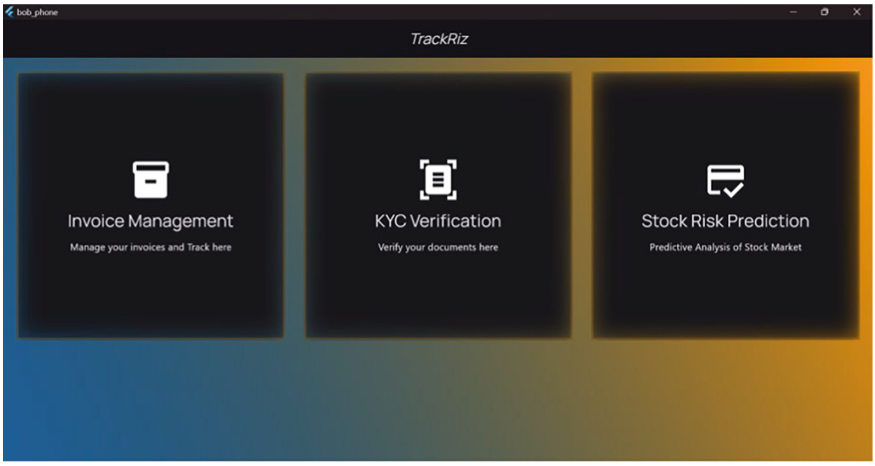
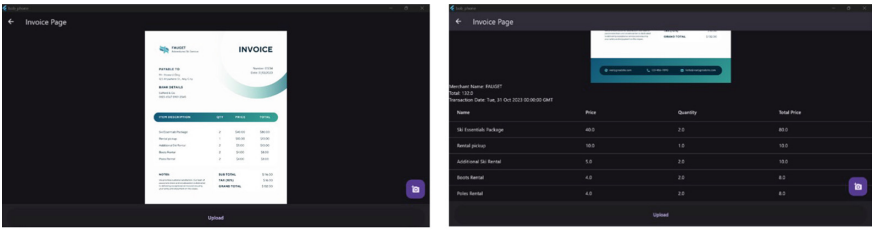


Fig. 2. Illustrates the Main User Interface

The primary user interface depicted in Fig. 2 of TrackRiz, with three major functionalities invoice management, KYC verification, and stock risk prediction. Redesigning the essential business processes, Invoice Management governs AI-powered text extraction to record transactions automatically, eliminating human errors and enhancing financial monitoring. KYC Verification provides security and regulatory compliance by authenticating users' identities, eliminating the risk of fraudulent transactions. Stock Risk Prediction accomplishes this through advanced analytics to analyze market trends and provide inputs for making informed investment decisions. Overall, these features provide operational efficiency in financial accuracy management, regulatory compliance, and data-driven decision-making for traders and businesses.



a b

Fig. 3. a,b Illustrates Invoice Management in TrackRiz for Financial Oversight.

In TrackRiz, shown in Fig. 3a and b, invoice management automates the processing of financial transactions and records reporting. It provides accurate record-keeping with itemized charges, dates of transaction, and open spend visibility. Invoice extraction, powered by AI, pulls invoice entry information and formats it in app, with a low flexibility possibility for errors. Invoices that contain multiple items can be tracked in with excellent detail. The features embed invoice management in trading, assist and financial situation information is great for budgeting audits, and keep us compliant with regulatory requirements. In addition, our functionality improves operational efficiency, account transparency, and decisions based directly on financial management.

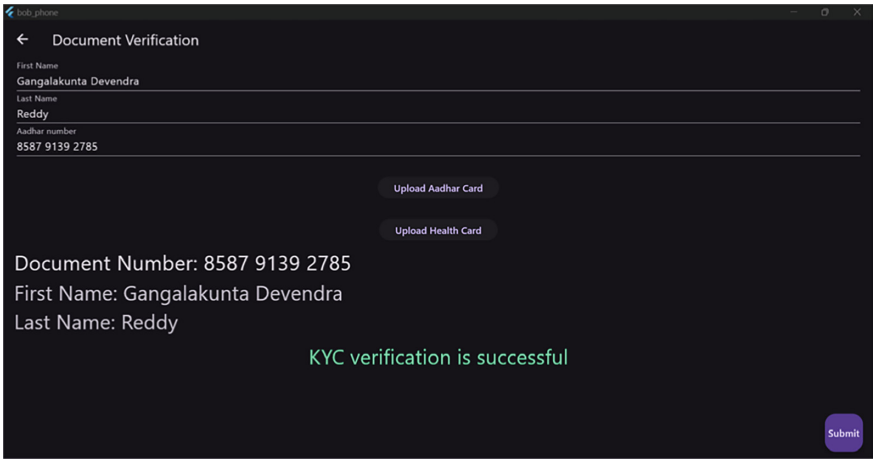


Fig. 4. Verification of the Uploaded Authorized Document

KYC Verification feature of TrackRiz is for secure identity verification by matching the given information against valid documents cleared by the authorities. The users are required to give answers to personal identity verification questions and upload verification documents required. The system presented in Fig. 4 cross-checked the uploaded verification documents against the given information for authenticity determination. If the information is as good as authenticated government records, the KYC process is successfully completed, verifying the user verified status. If any discrepancy is found, verification fails and users are prompted to attempt again. This feature provides security, avoids fraud, and safeguards against noncompliance with regulatory needs, thus being a critical feature in the trading and financial landscape of the platform.

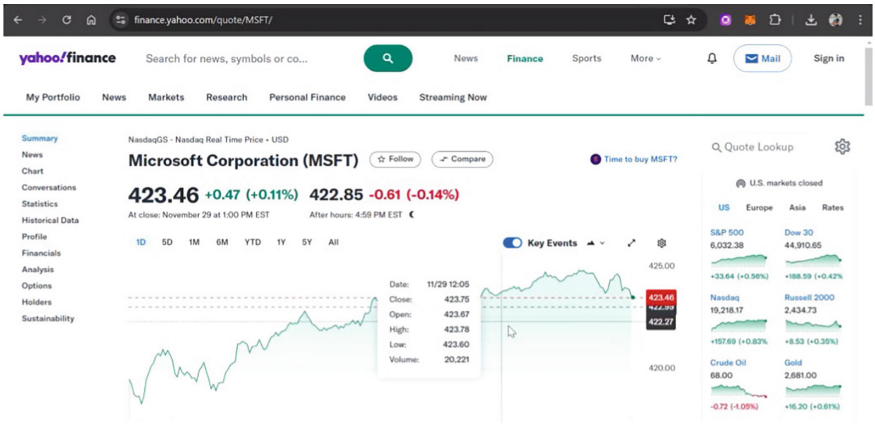
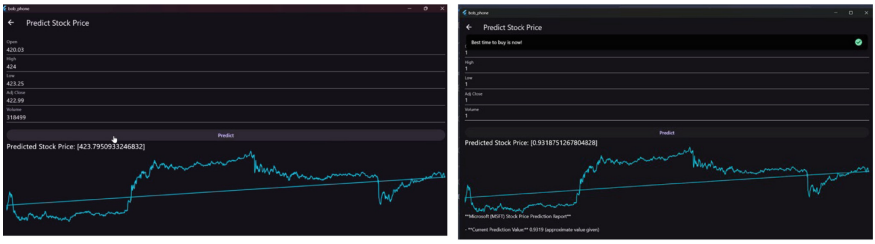


Fig. 5. Extracting the current data from the Microsoft Corporation



a **b**

Fig. 6. a, b Illustrate the Prediction of stock price of MSFT and Real Time Notification alerts

The Stock Risk Forecasting feature of TrackRiz applies real-time data extraction and machine learning to give correct predictions on stock prices, allowing users to base their investment decisions on solid evidence. As shown in Fig. 5, the system extracts current market data from Microsoft Corporation to provide accurate and latest financial information. This extracted data is then used in a model trained on the icrosoft.csv dataset holding past stock prices of Microsoft. As shown in Fig. 6a, the model takes critical financial parameters, such as open, high, low, adjusted close, and volume, so that users can feed relevant values to make more precise predictions. Figure 6b depicts the stock prediction feature with real-time notifications. It illustrated a dashboard, where users would find instant updates and alerts when their stock showed substantial price changes or market developments. The system was monitoring live data and through that data, providing useful, actionable recommendations such as "Best time to buy is NOW!" and the appropriate prediction metrics.

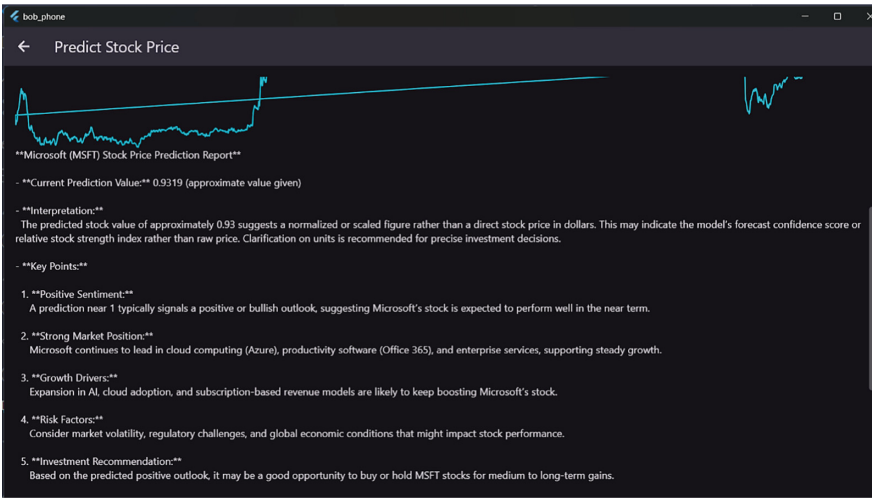


Fig. 7. Illustrate the detailed review of the predicted stock and risk management

Track Riz's Stock Risk Forecasting employs the application of Azure Automated Machine Learning (AutoML) to offer accurate information on forecasted stock prices

for definite decision-making. As is evident from Fig. 7, the system feeds actual market trends of Microsoft Corporation to maintain the forecasting model updated with real-time financial trends. Such information is then fed into the machine learning model, which is trained with historical stock prices' microsoft.csv data. As seen in Fig. 6, the model includes important stock variables including open, high, low, adjusted close, and volume and provides users with an option of entering corresponding values so that stocks can be predicted accurately. It demonstrates a risk management function for stock trading decisions. There is a stock prediction report, including an explanation of the level of prediction value, main drivers of growth, risk factors and what to expect. It identifies the risk factors including regulatory hurdles and economic conditions that could impact stock value, and how they may support users with better risk exposure management in turbulent market conditions.

Table 3. Performance Metrics of the Stock Risk Prediction Model

Metric	Value
R^2 Score	0.9998962
Normalized Root Mean Squared Error	0.002589823
Mean Absolute Error	0.02548666
Mean Absolute Percentage Error	0.006036422

TrackRiz's Stock Risk Prediction model in Table 3 has an R2 Score of 0.9998962, thereby capturing 99.99% of stock price variance. It also has a NRMSE of 0.002589823, MAE of 0.02548666, and MAPE of 0.006036422, all this demonstrates very good accuracy and precision for the model estimates. Regular retraining of the model is important to capture any changes in model predictions, since stocks can become initially volatile due to an economic crisis, or by sudden regional or global regulatory changes. Regular retraining of the model along with data drift checking and reaction to extreme market activity confirms the accuracy and precision of the prediction model and contributes to the overall robustness of the model. Regular retraining, data drift checking and properly accounting for current market events contribute to users trusting the recommendations the investment model makes because of its investment recommendations reliability.

8 Conclusion and Future Works

Real-time stock prediction and bulletproof security rules expand stock volatility and security concerns, with profound respect to you. With Application of Azure Automated Machine Learning and providing correct predictions of stocks prices with high coefficient of determination and low error rate be assured that this will help investors in making right decisions without worrying about avoiding possible risks they may encounter. With Azure Document Intelligence, KYC can be verified on the spot, assuring that only legitimate users have access to the platform to meet compliance. To make matters even better, an AI-backed invoice management tool can automate financial paperwork,

reducing errors and enhancing efficiency. Sitting atop Flutter and Azure's scalable cloud, the solution edges towards a frictionless, modern day trading experience, providing a new benchmark for financial platforms.

References

1. Dakalbab, F., Talib, M.A., Nasir, Q., Saroufil, T.: Artificial intelligence techniques in financial trading: a systematic literature review. J King Saud Univ.-Comput. Inf. Sci. **36**(3), 102015 (2024)
2. Nair, V.: AI-powered investment strategies: enhancing portfolio management through machine learning. J. Recent Trends Comput. Sci. Eng. (JRTCSE) **12**(1), 1–5 (2024)
3. El Mahjouby, M., Bennani, M.T., Lamrini, M., El Far, M., Bossoufi, B., Alghamdi, T.A.: Predicting market performance using machine and deep learning techniques. IEEE Access (2024)
4. Li, Y., Chen, L., Sun, C., Liu, G., Chen, C., Zhang, Y.: Accurate stock price forecasting based on deep learning and hierarchical frequency decomposition. IEEE Access (2024)
5. Hu, Y.J., Huang, S.W.: Challenges of automated machine learning on causal impact analytics for policy evaluation. In: 2017 2nd International Conference on Telecommunication and Networks (TEL-NET), (pp. 1–6). IEEE (2017)
6. Verma, K., Kumar, R., Rao, A.P., Ranjan, R.: Efficient e-KYC authentication system: redefining customer verification in digital banking. In: 2023 9th International Conference on Signal Processing and Communication (ICSC) (pp. 319–324). IEEE (2023)
7. Singh, R., Sharma, V., Kashyap, R., Manwal, M.: Automated multi-page document classification and information extraction for insurance applications using deep learning techniques. In: 2024 11th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), (pp. 1–7). IEEE (2024)
8. Toprak, A., Turan, M.: Transformer-based approach for automatic semantic financial document verification. IEEE Access (2024)
9. Ponnusamy, S., Gupta, P.: Scalable data partitioning techniques for distributed data processing in Cloud Environments: a Review. IEEE Access **12**, 26735–26746 (2024)
10. Song, Y., Sun, C., Peng, Y., Zeng, Y., Sun, B.: Research on multidimensional trust evaluation mechanism of fintech based on blockchain. IEEE Access **10**, 57025–57036 (2022)
11. VidyaBanu, R., Preethi, J., Dinesh, N.: Implementation of financial system using EyeOS in the cloud environment. In: 2011 International Conference on Recent Trends in Information Technology (ICRTIT) (pp. 656–660). IEEE (2011)
12. Zhao, W., Zhang, G., Yuan, G., Liu, J., Shan, H., Zhang, S.: The study on the text classification for financial news based on partial information. IEEE Access **8**, 100426–100437 (2020)
13. Ji, X., Wang, J., Yan, Z.: A stock price prediction method based on deep learning technology. Int. J. Crowd Sci. **5**(1), 55–72 (2021)
14. Abass, E.S., Mohamed, A.E.F., Amer, A., Hafez, M., Solyman, A., Fawzy, M.: Currency recognition using EAST for text detection and tesseract OCR for text recognition. In: 2023 2nd International Engineering Conference on Electrical, Energy, and Artificial Intelligence (EICEEAI) (pp. 1–9). IEEE (2023)
15. Swain, S., Gochhait, S.: ABCD technology—AI, blockchain, cloud computing and data security in islamic banking sector. In: 2022 International Conference on Sustainable Islamic Business and Finance (SIBF) (pp. 58–62). IEEE (2022)

Cloud Based Plant Health Monitoring System

Arnav Rahul Jade(✉), Jatin Santosh Jaiswal, Nishad Sachin Kamat, Vedant Mahesh Kandarkar, and Amruta Pabarekar

EXTC Department, Fr. C. Rodrigues Institute of Technology, Navi Mumbai, India
arnav.Jade2003@gmail.Com, Amruta.pabarekar@fcrit.ac.in

Abstract. The Cloud-Based Plant Health Monitoring System is designed to help farmers and agricultural experts precisely identify plant diseases using artificial intelligence and cloud technology. Traditional plant health assessments rely on manual inspection, which can be time-consuming and prone to errors. This project automates the process by allowing users to upload images of plant leaves, analyzed by a machine learning model hosted on a cloud platform. The system identifies whether the plant is healthy or has a disease, providing instant results through a simple mobile or web application. To achieve this, the system uses a Convolutional Neural Network (CNN) trained on a dataset of plant leaf images, covering both healthy and diseased conditions. The application is designed to be user-friendly, allowing even non experts to access plant health information easily. This approach reduces the need for excessive pesticide use, saves time, and supports sustainable farming practices by helping users respond to plant health issues promptly. Keywords-plant health monitoring, artificial intelligence, convolutional neural network (CNN), plant disease detection, cloud computing, mobile application and machine learning.

Keywords: Plant Health Monitoring · Cloud Computing · Artificial Intelligence (AI) · Convolutional Neural Network (CNN) · Plant Disease Detection · Machine Learning · Mobile Application · Sustainable Agriculture

1 Introduction

The Cloud-Based Plant Health Monitoring System is created to tackle the difficulties in the traditional plant health diagnostic by introducing machine learning and cloud technologies. In conventional methods, farmers and crops inspectors who manually check plants, a labor which frequently is slow. Consuming, elastic and restricted by human judgment. This system automates disease detection with real time image processing through the use of a convolutional Artificial intelligent model is built using Tuatara Deep Learning and a convolutional neural network (CNN) which to differentiate plant species as well as determine health conditions from leaf images. This innovation is endeavored to stimulate precision agriculture. Inform user with actual, real-time and data-driven insights into plant health [1–4].

S. Fong et al. (Eds.): ICT4SD 2025, LNNS 1651, pp. 154–163, 2026.
https://doi.org/10.1007/978-3-032-06688-6_15

2 Literature Survey

Cloud computing has emerged as a transformative technology in agricultural systems, enabling efficient plant health monitoring through real-time data collection, analysis, and storage. These systems integrate sensor networks and imaging tools with advanced analytics to detect plant diseases early, optimize resource allocation, and improve agricultural decision- making. The synergy of IoT, machine learning, and scalable cloud platforms has opened new avenues for precision farming, addressing key challenges in sustainable agriculture [5–7]. Agri-Info Cloud-Based Autonomic System for Delivering Agriculture as a Service Agri-Info uses cloud computing and IoT to offer Agriculture as a Service (AaaS), addressing issues like slow processing and limited storage in traditional agriculture systems. It integrates autonomic resource management for efficient allocation and fuzzy logic for decision-making, analyzing crop health based on factors like soil texture and pesticide use. The system includes a web and mobile application for accessibility. Tested using CloudSim, it outperformed Smart-Farm in execution cost, time, and latency [8].

Plant Disease Detection Using Image Processing and Machine Learning. The study focused on detecting plant diseases using a subset of the PlantVillage dataset, containing images of apple, corn, grapes, potato, and tomato leaves, along with their respective diseases. The images were preprocessed by converting to grayscale, applying a Gaussian filter, and using Otsu's thresholding for segmentation. Features such as shape (area, perimeter), color (mean and standard deviation), and texture (using GLCM) were extracted for classification. The Random Forest classifier was employed to classify the leaves based on these features. The model was trained on 80% of the data and tested on 20%, with K-fold cross-validation for robustness. The results were evaluated using metrics like accuracy, precision, and recall. The system was deployed as a web application on Heroku, with plans for real-time detection via robotics in the future (Table 1).

Table 1. Literature Survey

Sr. No.	Title	Author	Source	Methodology
1.	Plant Disease Detection Using Image Processing and Machine Learning	Pranesh Kulkarni, Atharva Karwande, Tejas Kolhe, Soham Kamble, Akshay Joshi, Medha Wayawhare	IEEE archives	This research project outlines the development of a system for identifying plant diseases from leaf images. The system uses image processing techniques to extract features and a Random Forest classifier to accurately categorize diseases, achieving a 93% accuracy rate while remaining computationally less demanding than deep learning alternatives.

(*continued*)

Table 1. (*continued*)

Sr. No.	Title	Author	Source	Methodology
2.	Agri-Info: Cloud Based Autonomic System for Delivering Agriculture as a Service	Sukhpal Singh, Inderveer Chana and Rajkumar Buyya	Researchgate.com	Agri-Info is a cloud-based system that leverages the Internet of Things (IoT) and Fuzzy Logic to provide Agriculture as a Service (AaaS). This system aims to address the limitations of traditional agricultural systems by offering efficient data management, automatic diagnoses of agricultural issues, and improved resource utilization while prioritizing user satisfaction and accessibility through dedicated web and mobile applications.
3.	Plant disease detection and classification techniques: a comparative study of the performances	Wubetu Barud Demilie	Springer.com	It provides a survey that examines a variety of deep learning and machine learning techniques for plant disease detection and classification. It involved reviewing existing research on these techniques, comparing their performance using various metrics, and identifying areas for future research.
4.	Plant Disease Detection and Classification by Deep Learning	Lili Li, Shujua Nan Zhang and Bin Wang	IEEE archives	It reviews recent studies on plant leaf disease recognition using image processing and deep learning techniques. It provides an overview of deep learning models, their history, model evaluation criteria, common datasets, and data augmentation methods. It concluded that deep learning techniques can effectively recognize plant leaf diseases but highlighted the need for more robust models adaptable to diverse datasets.

3 Implementation

The proposed Cotton Plant Health Monitoring System integrates sensor-based environmental monitoring with a Convolutional Neural Network (CNN) for disease detection. The methodology consists of four primary stages: data acquisition, preprocessing, model development, and system integration.

3.1 Data Acquisition

In the data acquisition stage, environmental parameters such as temperature, humidity, and soil moisture are collected using DHT11/DHT22 sensors for temperature and humidity and YL- 69 soil moisture sensors. These sensors transmit data to a ESP-8266, which then stores the information in a cloud/server for real-time monitoring. In parallel, high-resolution images of cotton leaves affected by Curl Virus, Fussarium wilt, and Bacterial blight are gathered from open-source datasets and data collected from users. To ensure a comprehensive analysis, healthy plant images are also included.

3.2 Preprocessing

Following data collection, the data preprocessing stage is performed. For image data, preprocessing involves resizing images to a standard dimension of 256 $\times$ 256 pixels and applying various augmentation techniques, such as flipping, rotation, and contrast enhancement, to improve model robustness. Background noise is reduced using Gaussian filtering and adaptive thresholding. Meanwhile, sensor data is normalized and stored in the Blynk IoT platform in a database. [9–11].

3.3 Model Development

The CNN model development phase focuses on designing and training a deep learning model for disease classification. The CNN architecture consists of convolutional layers for feature extraction, pooling layers for dimensionality reduction, and fully connected layers for final classification. The last layer employs the Softmax activation function to categorize input images into one of four classes: Healthy, Curl Virus, Fusarium Wilt, or Bacterial Blight. The dataset is divided into training (80%) and validation (20%) subsets (Figs 1, 2, 3 and 4).

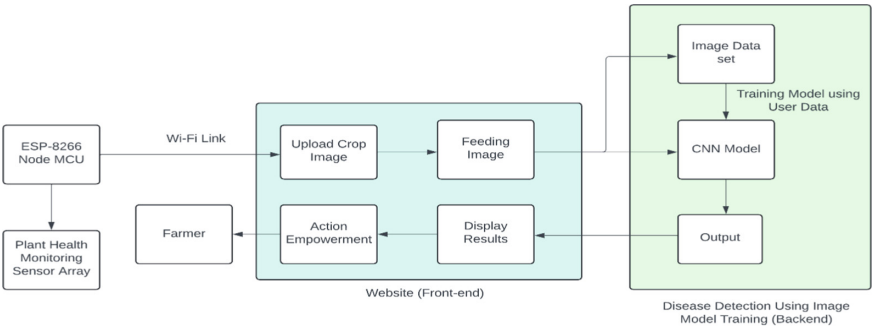


Fig. 1. Block Diagram

4 Results and Discussion

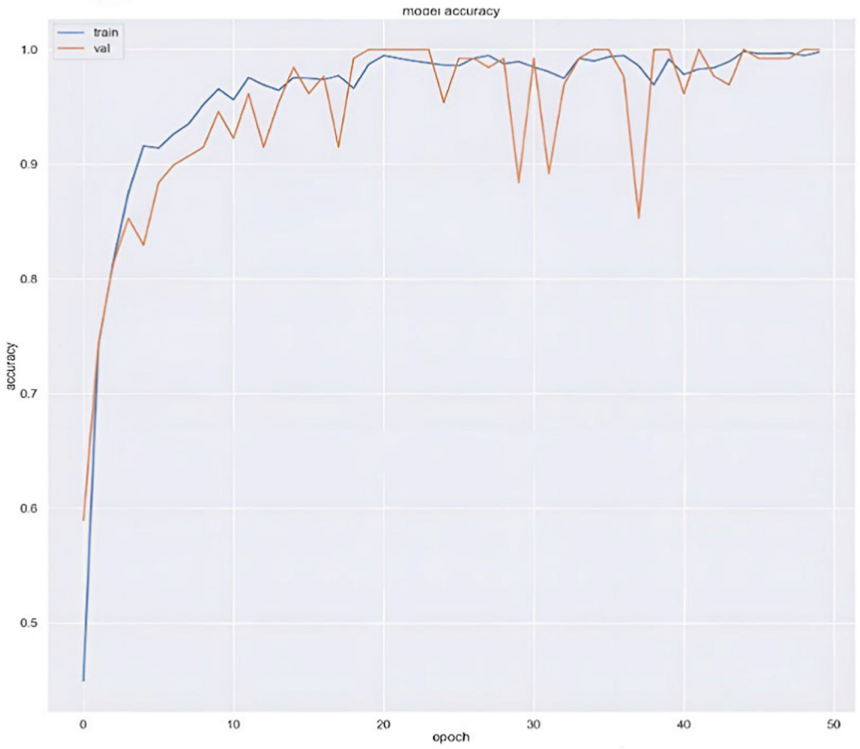


Fig. 2. Model Accuracy Graph

The proposed Cotton Plant Health Monitoring System was successfully implemented, integrating sensor-based environmental monitoring with a Convolutional Neural Network (CNN) for disease detection. A Streamlit-based GUI was developed to provide an interactive platform for users to upload cotton leaf images and receive real-time disease classification results. The system effectively identifies healthy leaves, curl virus, Fusarium wilt, and bacterial blight, offering a user- friendly and accessible interface for farmers and agricultural experts. The CNN model was trained for 50 epochs using an augmented dataset to improve robustness against variations in lighting, angle, and background noise. The model achieved an impressive 99.7% accuracy, demonstrating its effectiveness in distinguishing between diseased and healthy leaves. The training process showed a steady increase in accuracy, while the loss function converged efficiently, indicating a well- generalized model. The high accuracy suggests that the CNN effectively captures the distinguishing features of each disease category. The Streamlit GUI allowed seamless interaction with the model, enabling users to upload images and receive instant classification results. The web-based interface also displayed real-time sensor readings, helping monitor environmental conditions such as temperature, humidity, and soil moisture. This integration provides a comprehensive monitoring system for cotton farmers, enabling data-driven decision-making to prevent disease outbreaks.

Despite the high accuracy, some challenges were observed, including misclassification in cases of severe leaf damage or overlapping symptoms between different diseases. Future improvements could involve incorporating additional datasets, fine-tuning hyperparameters, and exploring transformer-based models for enhanced feature extraction. Moreover, expanding the system to detect other crop diseases could further enhance its utility.

Overall, the results demonstrate that the proposed system effectively combines deep learning and IoT-based monitoring to provide a real-time, accurate, and accessible disease detection solution for cotton plants. The high classification accuracy and successful GUI implementation highlight the potential of such systems in revolutionizing precision agriculture and crop disease management.

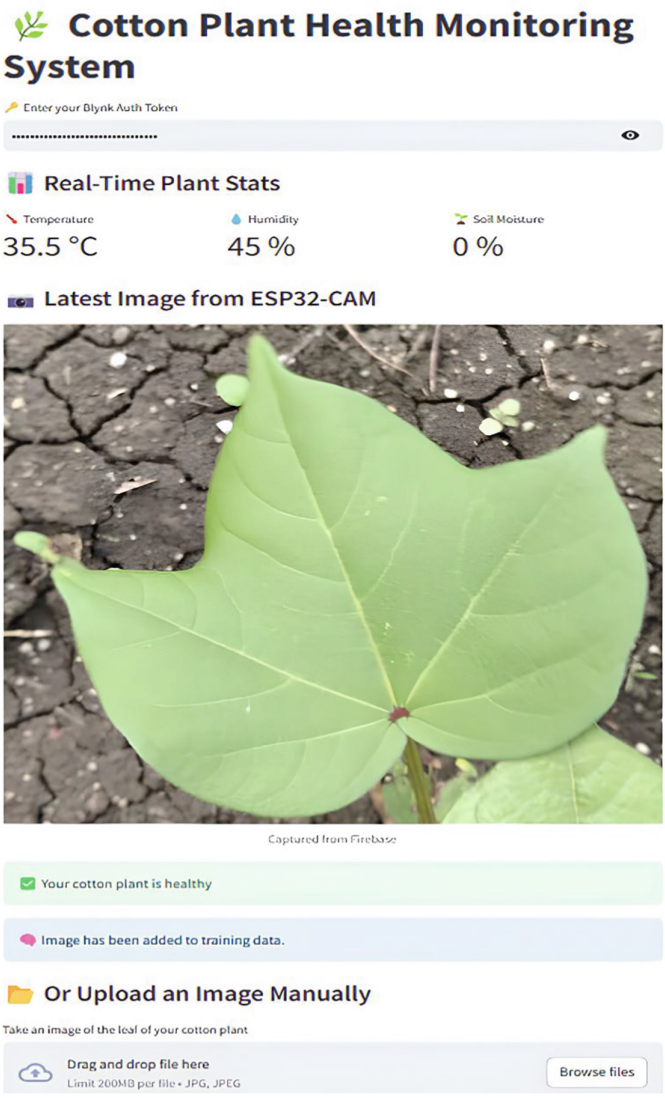


Fig. 3. GUI Implementation with Image and Sensor Readings

```
Epoch 36/50
54/54 [==============================] - 19s 343ms/step - loss: 0.0219 - accuracy: 0.9936 - val_loss: 0.0114 - val_accuracy: 1.0000
Epoch 37/50
54/54 [==============================] - 19s 353ms/step - loss: 0.0196 - accuracy: 0.9947 - val_loss: 0.0277 - val_accuracy: 0.9767
Epoch 38/50
54/54 [==============================] - 18s 336ms/step - loss: 0.0381 - accuracy: 0.9860 - val_loss: 0.5797 - val_accuracy: 0.8527
Epoch 39/50
54/54 [==============================] - 18s 335ms/step - loss: 0.0941 - accuracy: 0.9690 - val_loss: 0.0051 - val_accuracy: 1.0000
Epoch 40/50
54/54 [==============================] - 18s 332ms/step - loss: 0.0275 - accuracy: 0.9918 - val_loss: 0.0050 - val_accuracy: 1.0000
Epoch 41/50
54/54 [==============================] - 18s 337ms/step - loss: 0.0614 - accuracy: 0.9783 - val_loss: 0.1725 - val_accuracy: 0.9612
Epoch 42/50
54/54 [==============================] - 18s 337ms/step - loss: 0.0537 - accuracy: 0.9830 - val_loss: 0.0038 - val_accuracy: 1.0000
Epoch 43/50
54/54 [==============================] - 18s 336ms/step - loss: 0.0511 - accuracy: 0.9842 - val_loss: 0.0520 - val_accuracy: 0.9767
Epoch 44/50
54/54 [==============================] - 18s 340ms/step - loss: 0.0302 - accuracy: 0.9895 - val_loss: 0.0548 - val_accuracy: 0.9690
Epoch 45/50
54/54 [==============================] - 18s 336ms/step - loss: 0.0064 - accuracy: 0.9982 - val_loss: 0.0021 - val_accuracy: 1.0000
Epoch 46/50
54/54 [==============================] - 18s 336ms/step - loss: 0.0115 - accuracy: 0.9965 - val_loss: 0.0083 - val_accuracy: 0.9922
Epoch 47/50
54/54 [==============================] - 18s 337ms/step - loss: 0.0124 - accuracy: 0.9965 - val_loss: 0.0308 - val_accuracy: 0.9922
Epoch 48/50
54/54 [==============================] - 18s 334ms/step - loss: 0.0100 - accuracy: 0.9971 - val_loss: 0.0068 - val_accuracy: 0.9922
Epoch 49/50
54/54 [==============================] - 18s 337ms/step - loss: 0.0191 - accuracy: 0.9947 - val_loss: 0.0012 - val_accuracy: 1.0000
Epoch 50/50
54/54 [==============================] - 18s 338ms/step - loss: 0.0067 - accuracy: 0.9977 - val_loss: 0.0020 - val_accuracy: 1.0000
```

Fig. 4. Accuracy Gain Over Epoch Increments

5 Applications

The proposed system has several key applications in modern agriculture, especially in early disease detection, allowing farmers to identify plant diseases at an early stage and take necessary actions to reduce potential crop damage. By integrating CNN-driven automated diagnosis, the system enhances the accuracy of disease identification and reduces human error. The system also plays a vital role in precision agriculture, enabling farmers to monitor and manage plant health efficiently. With its remote monitoring capabilities through a mobile or web application, the system allows farmers to assess plant health without the need for manual inspections, making it particularly useful for both large-scale and smallholder farms. It also supports agricultural research and extension services by providing valuable data to scientists and policymakers, aiding in better decision-making and agricultural planning. Overall, the system contributes to smart farming innovations, improving crop health monitoring and enhancing productivity in a more sustainable manner [11–13].

6 Conclusion

This research presents a cloud-based plant disease detection system that utilizes machine learning and image processing to provide an automated and efficient approach for identifying plant diseases. Conventional disease detection methods rely on manual inspection, which is often slow, subjective, and requires expert intervention. By employing a Convolutional Neural Network (CNN) model integrated with cloud computing, this system offers a real-time, scalable, and user-friendly solution for farmers and agricultural professionals.

The system's implementation demonstrates the effectiveness of preprocessing techniques, segmentation methods, and deep learning algorithms in accurately identifying

plant diseases. By providing instant feedback through a mobile or web interface, it allows users to detect diseases early, reduce reliance on chemical treatments, and improve crop management strategies. This approach not only enhances agricultural efficiency but also promotes sustainable farming practices by minimizing unnecessary pesticide use.

Future enhancements will focus on expanding the dataset to include a broader range of plant species, optimizing computational efficiency with edge computing, and integrating AI- driven recommendations for disease treatment. With further advancements, this system has the potential to play a crucial role in modern precision agriculture, enabling data-driven decision-making to improve crop health and productivity.

7 Future Scope

- **Expansion of Dataset and AI Integration:** Increasing the dataset to include more plant species and diseases will enhance the system's accuracy. AI-driven recommendation systems can be integrated to provide precise treatment suggestions based on detected diseases.
- **Edge Computing and IoT Integration:** Implementing edge computing will enable real-time disease detection without full dependence on cloud services. IoT sensors can be incorporated to monitor environmental factors like soil moisture, temperature, and humidity for improved disease prediction.[14]
- **Advanced Technologies for Large-Scale Monitoring:** The system can be expanded to include drone-based crop monitoring, where deep learning models analyze aerial images for large-scale disease detection. Addition- ally, blockchain technology can be explored for secure agricultural data storage and knowledge sharing.

References

1. Gill, S.S., Chana, I., Buyya, R.: Agri-Info: cloud based autonomic system for delivering agriculture as a service. Internet Things **9**, 100131 (2020). https://doi.org/10.1016/j.iot.2019.100131
2. Banupriya, N., Chowdry, R., Yogeshwari, R., Varsha, V.: Plant disease detection using image processing and machine learning algorithm. J. Xidian Univ. **14**(7) (2022). https://doi.org/10.37896/jxu14.7/012
3. Demilie, W.B.: Plant disease detection and classification techniques: a comparative study of the performances. J. Big Data **11**, 5 (2024). https://doi.org/10.1186/s40537-023-00836-0
4. Li, L., Zhang, S., Wang, B.: Plant disease detection and classification by deep learning—A review. IEEE Access **9**, 56683–56698 (2021). https://doi.org/10.1109/ACCESS.2021.3069646
5. Rizk, H., Habib, M.K.: Robotized early plant health monitoring system. In: 44th Annual Conference of the IEEE Industrial Electronics Society (IECON), Washington, DC, USA, pp. 3795–3800 (2018). https://doi.org/10.1109/IECON.2018.8592833
6. Suneja, B., Negi, A., Kumar, N., Bhardwaj, R.: Cloud-based tomato plant growth and health monitoring system using IoT. In: 3rd International Conference on Intelligent Engineering and Management (ICIEM), London, UK, pp. 237–243 (2022). https://doi.org/10.1109/ICIEM54221.2022.9853170

7. Bhattacharya, S., Bagmar, A.S., K.E.: A comprehensive plant health monitoring system with IoT and deep learning. In: International Conference on Next Generation Electronics (NEleX), Vellore, India, pp. 1–6 (2023). https://doi.org/10.1109/NEleX59773.2023.10421592
8. Chen, J., Yu, Z., Lam, A.: Research on monitoring platform of agricultural product circulation efficiency supported by cloud computing. Wireless Pers. Commun. **102**(4), 3573–3587 (2018)
9. Zamora-Izquierdo, M.A., Santa, J., Martínez, J.A., Martínez, V., Skarmeta, A.F.: Smart farming IoT platform based on edge and cloud computing. Biosys. Eng. **177**, 4–17 (2019)
10. Prasad, S., Peddoju, S.K., Ghosh, D.: AgroMobile: A cloud-based framework for agriculturists on mobile platform. Int. J. Adv. Sci. Technol. **59**, 41–52 (2013)
11. Muangprathub, J., Boonnam, N., Kajornkasirat, S., Lekbangpong, N., Wanichsombat, A., Nillaor, P.: IoT and agriculture data analysis for smart farm. Comput. Electron. Agric. **156**, 467–474 (2019)
12. Fina, F., Birch, P., Young, R., Obu, J., Faithpraise, B., Chatwin, C.: Automatic plant pest detection and recognition using k-means clustering algorithm and correspondence filters. Int. J. Adv. Biotechnol. Res. **4**(2), 189–199 (2013)
13. Mohanty, S.P., Hughes, D.P., Salathé, M.: Using deep learning for image-based plant disease detection. Front. Plant Sci. **7**, 1419 (2016)
14. Li, J.H., Lin, L.J., Tian, K.: Detection of leaf diseases of balsam pear in the field based on improved Faster R-CNN. Trans. Chin. Soc. Agric. Eng. **36**(12), 179–185 (2020)

Blockchain-Enhanced KYC: A Secure and Decentralized Framework for Identity Verification

G. B. Sambare, Sankarsha Shelke, Sahil Wawdhane(✉), Harshad Wable, and Abhinav Thube

Department of Computer Engineering, PCCOE, Pune, India
{santosh.sambare,sankarsha.shelke21,harshad.wable21, abhinav.thube21}@pccoepune.org, wawdhanesahil@gmail.com

Abstract. The KYC Powered by Blockchain for decentralized, secure, and more efficient Know Your Customer (KYC) system using blockchain. This system solves the inherent inefficiencies of traditional KYC by allowing institutions to share validated customer data, mitigating redundancy among KYC providers, and reducing both costs and compliance time. Tamper-proof architecture of blockchain allows for strong data privacy, security, and compliance of AML and GDPR regulations. Customers gain full control over their personal data, with the ability to grant and revoke access dynamically, reducing risks of breaches and fraud. The framework integrates off-chain storage for sensitive data and combines advanced cryptographic methods like AES and ECC for encryption and security. Smart contracts automate data handling and permissions management, ensuring secure, transparent, and immutable data sharing across institutions.

Keywords: KYC · Blockchain · Ethereum · Off-chain Storage · Encryption · DLT

1 Introduction

Know Your Customer (KYC) is used worldwide by financial institutions to avoid illegal activities by persons involved in a business relationship with the aforementioned financial body. It is a procedure of identifying the party engaged in the business and for background verification of the individual [1]. The existing systems of data storage such as cloud storage, Direct Attached Storage(DAS) in Big data and Object Storage became more vulnerable to attacks from various malicious threats which includes privilege abuse attack, SQL injection attack, Storage media exposure attack, Targeting unpatched database vulnerability attack from all over the world. The current KYC mechanism has a severe concern in financial institutions as it requires separate ledger for the separate financial organizations. Every institution has its KYC process, which sometimes may include third-party, which may cause increased maintenance cost, time and redundancy [3]. Identity and access management encompasses an organization's resources, technologies and processes for authenticating, authorizing, and identifying

S. Fong et al. (Eds.): ICT4SD 2025, LNNS 1651, pp. 164–175, 2026.
https://doi.org/10.1007/978-3-032-06688-6_16

individuals to use services in that organization or other associated organizations. Its examples range from issuing and verification of birth certificates, national identification cards, passports or driver's license in a governmental setting to customers and employees using various software applications or hardware components within organization and level of access, privileges and restrictions each person has while doing so [4]. A blockchain is a decentralized, tamper-resistant distributed database that records transactions over a peer-to-peer public or private network. The ledger, which is sent to every member node in the network, permanently documents the transactional data between the nodes of an ordered chain of blocks connected by cryptographic hashing [11]. The aim of this proposed system is to effectively build an optimized KYC Blockchain system for managing KYC details of the customers. The system can be deployed in various organizations which require KYC verification of the customers [2].

2 Literature Survey

Blockchain Technology in Financial Firms KYC: AKYC (Anonymous Know Your Customer) is a regulatory framework for preventing money laundering and terrorist financing through which blockchain technology is attracting great attention in the financial sector. The libraries of interdisciplinary research on blockchain based KYC systems is not exhaustive below is a summary of literature that capture some facets. A solution, based on Blockchain, to the KYC dilemma [1].

The integrating of blockchain as a solution to KYC processes is often time-consuming, repetitive across financial institutions and this will be introduced in this paper.The authors emphasize the importance of using distributed ledger technology to store verified KYC data since it can be done in a secure and unchangeable manner. This makes it possible for banks and other financial institutions to share verified customer information with greater security minimizing wasteful storage of customer data and smoothing the onboarding process. In the paper, it is explained how compliance processes can be enhanced with the use of blockchain technology in a secure and complete manner, which makes the approach more attractive than the centralized KYC processes.

Optimized KYC Blockchain System [2].

This study provides an improved model for KYC processes incorporating Blockchain technology to solve scalability and security challenges. The authors implement a permissioned blockchain where only certain approved bodies are able to access the KYC data, thus increasing privacy while still ensuring limited transparency. The framework uses smart contracts to automate identity authentication and regulatory compliance processes. The paper argues the benefits of using permissioned blockchains, specifically the tailored consensus algorithms, which increase the efficiency of KYC verifications and the data fidelity across the financial institutions.

KYC Optimization Using Blockchain Smart Contract Technology [3].

This document outlines the increasing level of automation provided by smart contracts in KYC steps and the reduction of other forms of human participation in the process. Financial institutions are able to facilitate the process of customer verification by automating the procedures in which certain regulatory thresholds have to be met. The research shows that the framework built on smart contracts eliminates the risks related

to human or fraudulently motivated errors. The authors draw attention to the strength of perpetual audit systems provided by blockchain, in which customers have the ability to control the access to their information while updates and changes are made freely.

The Use of Blockchain Technology to Verify KYC Documents [4].

This research evaluates the possibility of utilizing blockchain technology for the verification of KYC documents. It discusses the benefits of reducing reliance on central authorities through decentralized systems. The authors present a new model in which KYC documents are kept off-chain, while the blockchain retains hashes of the documents to maintain their integrity. This system is efficient because authorized institutions only can access encrypted data, and lowers the chances of unauthorized access. This paper concludes that a combination of off-chain storage and blockchain for validation provides a scalable and economical solution to the KYC requirements.

Blockchain-Powered KYC in a CBDC World: The E-Rupee Experience [5].

This paper investigates the integration of blockchain-based KYC systems within the central bank digital currency (CBDC) framework, using India's E-Rupee as a case study. It emphasizes the importance of interoperable KYC processes that meet both regulatory compliance and efficiency standards. The authors put forth a consortium arrangement whereby the regulators, the banks, and other financial institutions share the responsibility of managing the KYC records on a permissioned blockchain. The study stresses that this arrangement not only helps ease the processes of KYC, but also offers a contradiction that meets the requirements of the CBDC by increasing the level of transparency and reducing fraud cases.

Ethereum Blockchain Framework Enabling Banks to Know their Customers [6] this paper reviews existing research on blockchain-based Know Your Customer (KYC) frameworks, emphasizing how blockchain technology enhances KYC processes by improving security, transparency, and efficiency. It highlights various decentralized KYC models, including those utilizing Hyperledger Fabric, smart contracts, and InterPlanetary File System (IPFS) for document validation. Previous studies address how blockchain facilitates safe information exchange between financial institutions, reducing redundancy in KYC verification and saving time and money. Additionally, studies look at data privacy issues, regulatory obstacles, and how blockchain-based KYC models incorporate AML compliance. However, the paper's Ethereum-based decentralized KYC approach attempts to solve the issue of existing frameworks' sometimes weak security features for key management and privacy protection.

The present research focuses on potential impacts of blockchain technology adoption in the financial sector revolving around KYC compliance. The authors have considered the possibility of blockchain technology implementation in Identity verification unification across numerous banks and other financial bodies enabling redundancy free KYC for the clients. The paper presents a novel framework for cross institutional blockchain KYC where participants share trusted data. The proposed model lowers operational expenses while at the same raises compliance effectiveness for clients of various financial institutions.

A review of financial literature on KYC blockchain based systems reveals that some literature tend to perceive blockchain as a solution for any problem and the studies that have tried to capture ways by which blockchain can eliminate the risks, inefficiencies,

breaches of security, and compliance in financial identity fraud are too few. These insights have the potential to revolutionize the KYC processes, or at least industry standards set around them and make them more compliant, efficient, secure, and user friendly in business with financial institutions (Table 1).

Table 1. Review of recent literature

Sr. No.	Title	Methodology	Strengths	Weakness
1.	A Blockchain based Solution to Know Your Customer (KYC) Dilemma (2019) [1]	A central third party administers the onboarding of banks into a private Ethereum blockchain. It is then the responsibility of the banks to maintain the users' KYC information and they are free to view all of the information and manage the access permissions granted to other users.	It improves security through encryption and tamper-proofing KYC data with blockchain, which further secures it.	The system is available to users in a read-only capacity, meaning that interactions are minimal and data access is limited.
2.	Optimized KYC Blockchain System (2020) [2]	The implementation uses KYC blockchain solution for the Ethereum blockchain, while integrating AES symmetric encryption and LZ compressed data for added security and efficiency. Each KYC block is encrypted and compressed before being added to the blockchain, resulting in low storage requirements and high data security.	By using AES encryption to secure KYC data, the system bolsters security greatly.	Options are often complicated by the sole dependence on Ethereum. When the network is congested, it results in accelerated transaction costs.

(*continued*)

Table 1. (*continued*)

Sr. No.	Title	Methodology	Strengths	Weakness
3.	KYC optimization using Blockchain Smart Contract technology (2020) [3]	In this paper, every organization operates as a peer in a permissioned blockchain network. These data silos facilitate information sharing between themselves, making it accessible to permitted business users. The integrity of the KYC information cannot be altered by any other party except the KYC holder him or herself.	The KYC data is well protected because Blockchain is incredibly difficult to hack due to its cryptographic nature.	The deployment can be particularly difficult when a wide variety of tools and technologies are adopted concurrently, such as smart contracts, networking, and cryptography.
4.	The Use of Blockchain Technology to Verify KYC Documents (2023) [4]	In this paper, an alternative method is suggested to replace the customary centralised KYC storage with an innovative approach based on the blockchain. For this purpose, the user's KYC documents are uploaded onto the blockchain network. Therefore, the user's data will not be kept on a centralized, vulnerable database.	The absence of a governing body fosters self-regulating systems that are more secure since no single party can manage or corrupt the information.	One of the important and complicated problems is scalability, as blockchain systems increase in size, the speed and efficiency of verification slows down, especially in more complex networks with sophisticated consensus protocols.

(*continued*)

Table 1. (*continued*)

Sr. No.	Title	Methodology	Strengths	Weakness
5.	Blockchain-Powered KYC in a CBDC World: The E-Rupee Experience (2024) [5]	A blockchain based KYC integrated within the Central Bank Digital Currencies (CBDCs) system, for instance, the Indian e-Rupee is proposed in this paper. The approach utilizes the blockchain's distributed ledger to ensure the traditional methods of storing KYC information are secure and automated using smart contracts.	The merger of KYC with CBDC allows for a smoother flow of financial transactions, making them easier to track and execute.	The Ethereum platform is likely to cause high transaction fees in cases where the network is congested.

3 Proposed System

The described proposed solution concerning identity verification enhances the traditional KYC system by integrating it with blockchain technology, which permits secure storage of KYC information. Automated processes of data management are achieved by deploying smart contracts on an Ethereum permissioned blockchain. The specifics of these contracts ensure the privacy and security of user data while remaining compliant with relevant legislation. The main purpose of the system is to generate an accurate and publicly auditable record, which improves customer onboarding experience while increasing information security and diminishing redundancy.

Algorithm for Decentralized KYC Verification Using Blockchain

Step 1: Gathering Data and Adding New Users.

- KYC Data Submission: The KYC information is submitted by clients and is thereafter securely encrypted and stored off-chain in a designated secure storage where only cryptographic hashes are stored on the blockchain.

Let D_k represent the KYC data for customer k.

1. Data hash using SHA-256:

$H(D_k) = \text{SHA - 256}(D_k)$

2. Store D_k off-chain and $H(D_k)$ on-chain.

- Data Verification by Financial Institutions: Gathered datasets undergo verification by government certified financial bodies, where their validity is checked. Hashes corresponding to the KYC information, that have now been held, are stamped on the blockchain, establishing the record that ensures the legitimacy and accuracy of the KYC documents.

  If D_k is validated by institution I_j:
  $H_{validated}(D_k) = SHA - 256(D_k)$ and is stored on-chain.
  Step 2: Blockchain and Smart Contract Setup.

- Permissioned Blockchain Deployment: Deploy a permissioned Ethereum blockchain, where only authorized entities such as banks, regulators, and financial institutions can participate.

  Let $\mathcal{P}$ represent the set of participants and B the blockchain:
  $B = \{T_1, T_2, \text{---------}, Tn\}$ where $Pi \in \mathcal{P}$

- Smart Contract Design: Implement smart contracts to automate data handling and verification processes. Key functions of smart contracts include KYC data submission, verification, access management, and permission revocation. The smart contract manages permissions and enforces rules to prevent unauthorized access to sensitive data.

  The smart contract S enforces access control and verification:
  $S(R, U, P) =$
  {
  Grant access if $U \in R$ and $P(U, k) = 1$,
  Deny access otherwise.
  }
  where R: roles, U: user, and P(U,k): permissions.

Step 3:User Authentication and Role Management

Role-Based Access Control- Each user role (Customer, Financial Institution, Admin) has specific permissions allocation defined within the smart contract. Customers are granted more pertinent data access permissions, whereas financial institutions are granted access only to the data needed for the verification process.

Access for role R:
$A(R, D_k) =$
{
1 if role R has permission for date Dk
0 otherwise
}

Multi-Factor Authentication (MFA) – The user verification is done through a multi-dimensional process and exemplifies a predominant step in all the user roles. Customer verification is the most critical because it is the last point of a controlled user access.

Step 4: Data Encryption and Storage

Off-Chain Storage with Hashing- Sensitive KYC data is stored in a Decentralized database, where the sensitive data is stored off-chain. The data itself is encrypted using SHA-256, whilst the hash is stored on-chain. As a result, any method which can prove data is modified in case stored hash does not correspond with the off-chain data becomes possible.

This ensures that data tampering is detectable by any mismatch stored hashes and the data off-chain.

Hash $H(D_k)$ and store D_k off chain:

$H(D_k) = SHA - 256(D_k)$.

Hybrid Encryption-In order to safeguard the KYC details from the viewing public during transactions, a model of hybrid encryption that incorporates Advanced Encryption Standard (AES) can be utilized along with Elliptic Curve Cryptography (ECC).AES handles the user data while ECC is responsible for protecting the AES key which ensures strong data security.

The AES encryption of D_k is given by the following equation:

$C_k = AES(D_k, K_{AES})$.

Where C_k is the ciphertext and KAES is the algorithm's key.

The rest of KAES encryption with ECC is as follows.

$KECC = ECC_{encrypt}(K_{AES}, P_{ECC})$.

Where P_{ECC} is the public key from ECC.

Step 5: Verification and Access Control Using Smart Contracts.

- KYC Data Request and Retrieval: Are you tracking the missing valuables? The Financial institutions do initiate requests to access KYC Data through the smart contract. Moreover, the smart contract will be validating the requestor's authorization as well as checking the customer's access permissions.

Verify Access$(U, D_k) = S(R, U, P)$

- Data Access Restriction: An Access control system enables customers to grant or revoke data access permissions to particular institutions. Moreover, a permissioned access system ensures that only authorized entities have temporary access to customer data.
- Revocation of Consent in Real Time: Through smart contract features, the system allows users to revoke permissions for data access at any time. The blockchain immediately updates the relevant institution's access rights upon revocation, blocking any additional data access. Without requiring any manual intervention, this real-time mechanism guarantees that user consent stays dynamic and enforceable.
- Step 6: Immutable Record Creation and Audit Trail.

  - All Know Your Client (KYC) processes are approved, altered, and carried out on the blockchain in a manner that ensures immutability. In addition, this specific blockchain allows for easy retrieval of audit reports by the relevant authorities and organizations. The system maintains comprehensive logs of file access and changes, along with the participants and dates when the actions occurred.

Blockchain ledger L:
$L = L_0, L_1$,----------, L_n
where each $L_i = \{H(D_k)$, Timestamp, Actor$\}$

- Regulatory Compliance and Audit – The system provides the functionality for regulators, both at international standards and from different regions, to efficiently verify compliance via KYC activities defined and set in rules for further controls.
- Step 7: Secure Data Sharing Across Institutions

  - Clients Sharing KYC Data Between Companies: With consent from the client, any financial institution can view their KYC details through a verified approach. This help saves customer onboarding time as well as verification checks.

$$Access(I_j, D_k) = S(R, I_j, P)$$

- Tracking System for KYC Approvals, Rejections and Updates: Customers and financial institutions get notified regarding the approval, rejection or changes to the KYC customer data by a smart contract. This allows the KYC process to be transparent.

System Security Enhancements:
Advanced Encryption Standard (AES)
AES is deployed for the KYC data encryption before the information is stored off-chain. This technique protects the data during the storage and can only be released to authorized personnel with the key. More rounds of AES encryption have been shown to enhance the level of security and reduce the chance of breakages to data integrity. In this case, 14 rounds of AES-256.

$$C_k = AES(D_k, K_{AES}).$$

Elliptic Curve Cryptography (ECC)
The KYC partary's AES encryption keys are protected via ECC. The public and private key generated by the ECC algorithm provides key security without demanding intensive calculation activities. The protection is important when using blockchain technology due to the need for efficient computation as ECC has been proven to be lightweight.

$$K_{ECC} = ECC_{encrypt}(K_{AES}, P_{ECC})$$

1. Key Generation: ECC is responsible for the creation of the public and private keys needed for the encryption and decryption of AES keys. These keys are stored on-chain in a secure and encrypted manner.
2. Encryption and Decryption Process: KYC data is encrypted using AES and coupled with ECC data encryption. The KYC data is secured using ECC, which encrypts the AES key. To decrypt, the private key belonging to the ECC has to be used, and this key is only made available for specific entities. The data encrypted is persistent and unalterable within the blockchain ecosystem.

GDPR Compliance Measures
The system will supports GDPR compliance by giving customers full control over their personal data. Through smart contracts, users can grant and revoke access dynamically and exercise their "right to erasure" by deleting data stored off-chain. By keeping

sensitive information off-chain and only storing hashes on-chain, the system ensures that customer data can be modified or deleted in compliance with regulatory obligations, providing a privacy-preserving yet transparent KYC process (Fig. 1).

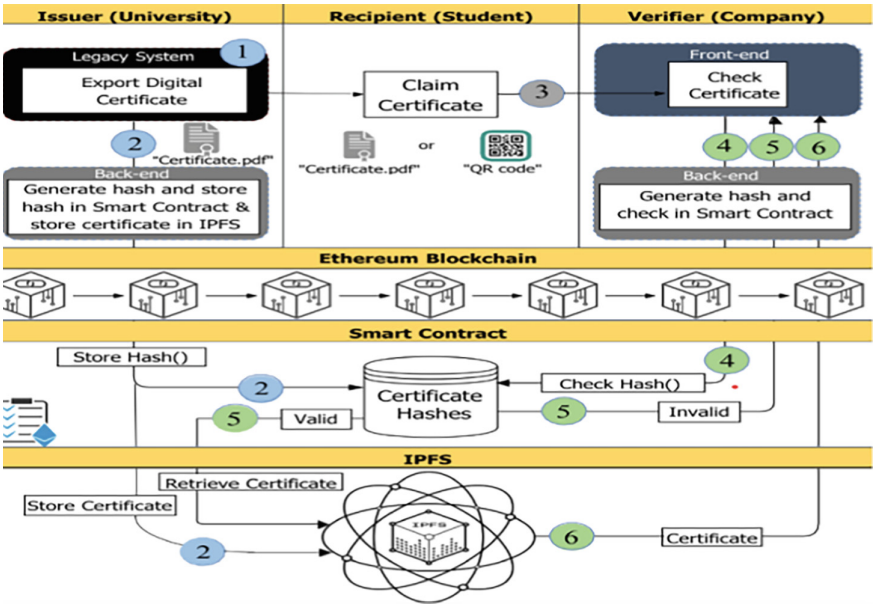


Fig. 1. System Architecture

Key Improvements Over Base Papers

Improvements Over "Enabling Trust and Privacy-Preserving e-KYC System Using Blockchain"[10]

Issue: Encryption overhead & key management inefficiency.

Improvement: Hybrid AES + ECC encryption reduces computation overhead while improving security.

Issue: No customer-controlled consent mechanism.

Improvement: Decentralized Identity Management (DID) allows users to grant/revoke access dynamically.

Issue: Data stored on-chain increases blockchain load.

Improvement: Off-chain encrypted storage with on-chain hash verification optimizes storage.

Improvements Over "Blockchain-Based Identity Management and Access Control for Open Banking"[8]

Issue: Lacks fine-grained access control.

Improvement: Smart contract-based Role-Based Access Control (RBAC) ensures granular permission management.

Issue: Data breaches are more likely when Open Banking APIs are exposed.

Improvement: By integrating Zero-Knowledge Proofs (ZKP), identities can be confirmed without disclosing private information.

Issue: Centralized databases include customer data.

Improvement: Decentralized storage and great resilience are guaranteed by Inter-Planetary File System (IPFS) + Blockchain.

Future Enhancements:

Zero-Knowledge Proofs (ZKP) are used to confirm identities without disclosing private information.

Fraud detection using machine learning to stop fraudulent KYC submissions.

compatibility for cross-border identity verification with CBDCs (digital dollar, e-Rupee).

Cross-Chain Interoperability: Enabling interoperability between various blockchain networks may be the main goal of future developments. Platforms like Ethereum, Hyperledger, and Quorum can securely share KYC data through mechanisms like blockchain oracles, atomic swaps, and cross-chain bridges. This would increase system flexibility and adoption by enabling smooth identity verification even in cases where institutions are using different blockchain architectures.

Benefits and Objectives of the Proposed KYC System.

- The Advantages and Goals of the Proposed System for Verifying Clients' Identities using Blockchain Technology
- The system aims to manage KYC processes more efficiently, securely, and in a transparent manner using blockchain technology. Some key objectives include the following:
- Increased Security with KYC Data Privacy: The KYC blockchain architecture preserves security and privacy of sensitive customer KYC data through cryptographic security and the blockchain's immutability.
- Reduced Onboarding Time: By enabling secure data sharing across institutions, the system reduces redundant verification steps, speeding up customer onboarding.
- Regulatory Compliance: The system's audit trail and permissioned access model facilitate compliance with data protection and KYC regulations, ensuring transparency and accountability.
- Cost-Effective Solution: Reduced manual processing and improved inter-institutional data sharing lower operational costs for financial institutions.

This system serves as an effective and secure KYC solution, ensuring data protection, regulatory compliance, and improved operational efficiency across the financial sector.

4 Conclusion

The illustrated blockchain-based KYC has the potential to transform identity verification as it addresses the core issues with the existing KYC processes which are slow, repetitive, and highly vulnerable to breaches. This system utilizes permissioned Blockchain Ethereum smart contracts and a hybrid cryptographic architecture that includes AES and ECC to achieve enhanced security, privacy, and compliance within the realm of regulations. Additionally, the immutable and distributed characteristics of blockchain permit safe inter-organizational data exchange while giving the clients power over their own data.

The impenetrable contracts also made it easy to manage permissions and trace data flow which was extremely beneficial for secure inter-institutional data exchange. Moreover, Smart contracts performed under the hybrid encryption guarantee secure data storage and transmission without any informal issues. They also solve the contradiction of needing compliance and non-transparency at the same time.

All in all, compared to traditional processes put in place by financial institutions, this blockchain based KYC system can significantly increase security, efficiency, and privacy of the users. The possibilities of such solutions integrating to different industries where identity authentication can be a problem, such as healthcare, supply chains, and even government services are massive.

References

1. George, D., Wani, A., Bhatia, A.: A blockchain based solution to know your customer (KYC) dilemma. IEEE (2019)
2. Sundareswaran, N., Sasirekha, S., Joe Louis Paul, I., Balakrishnan, S., Swaminathan, G.: Optimized KYC blockchain system. Int. Conf. Innovative Trends Inf. Technol (2020)
3. Yadav, A.K., Bajpai, R.K.: KYC optimization using blockchain smart contract technology. Int. J. Innovative Res. Appl. Sci. Eng. (IJIRASE) (2020)
4. Rathod, V.U., et al.: The use of blockchain technology to verify KYC documents. IEEE International Conference on Blockchain and Distributed System Security (ICBDS) (2023)
5. Dumbre, T., Sanadhya, S., Vijayakumari, L., Shaji, R., Vinoth Kumar, C.N.S.: Blockchain-powered KYC in a CBDC world: the E-Rupee experience. IEEE (2024)
6. Vinoth Kumar, C., et al.: Ethereum Blockchain framework enabling banks to know their customers. IEEE (2024)
7. Yadav, P., Chandak, R.: Transforming the know your customer (KYC) process using blockchain. IEEE
8. Liao, C.-H., Guan, X.-Q., Cheng, J.-H., Yuan, S.-M.: Blockchain-based identity management and access control framework for open banking ecosystem. Int. J. Banking Financ. **10**(4) (2020)
9. Bulut Karadag, A., Zaim, H., Akbulut, A.: A.: Blockchain-based KYC model for credit allocation in banking. J. Financ. Serv. Technol. **5**(1) (2024)
10. Fugkeaw, S.: Enabling trust and privacy-preserving e-KYC system using blockchain. IEEE Access (2022)
11. Desai, S., Shelke, R., Deshmukh, O., Choudhary, H., Sambare, S.S.: Blockchain based secure data storage and access control system using IPFS. J. Crit. Rev. **7**(19) (2020)
12. Sambare, S.S., Khandait, K., Kolambe, K., Kolage, K., Nimbalkar, T.: Crowdfunding using blockchain for startup ventures. 7th International Conference On Computing, Communication, Control And Automation (ICCUBEA)

Off-Line Signature Verification Using Region-Based Geometric Feature Matching with Adaptive Similarity Scoring

Prabira Kumar Sethy[1(✉)], Sachin Sharma[1], Ajit Behera[1], Satyaprakash Barik[1], and Amresh Bhuyan[2]

[1] Department of Electronics and Communication Engineering, Sambalpur University Institute of Information Technology, Jyoti Vihar, Burla, Sambalpur 768019, Odisha, India
prabirsethy.05@gmail.com

[2] Department of Computer Science Engineering, Sambalpur University Institute of Information Technology, Jyoti Vihar, Burla, Sambalpur 768019, Odisha, India

Abstract. Signature verification constitutes a fundamental component of biometric authentication methods used in financial and legal identity verification systems. The research presents an offline signature verification method that examines geometric and morphological region-based features to authenticate test signatures. The methodology analyzes binarized signature images to extract important attributes such as area, perimeter, centroid, eccentricity, solidity, extent, major and minor axis lengths, orientation, convex area, Euler number, and equivalent diameter. After analyzing the reference signature collection, the most prominent image region gets processed for feature extraction. The test signature is evaluated through feature-wise similarity calculations while undergoing preprocessing identical to reference images. The normalization process for each feature difference allows comparison against specific thresholds to determine cumulative similarity scores. Authentication confirmation for a signature occurs when its score level exceeds the 95% predetermined acceptance benchmark. Experimental results demonstrate that our method achieves optimal computational efficiency while providing high verification accuracy and clear distinction between real signatures and forgeries. The framework merges reliable performance with simple operation and quick processing abilities making it ideal for lightweight biometric systems.

Keywords: Signature Verification · Morphological Features · Biometric Authentication · Similarity Score · Regionprops Analysis

1 Introduction

Biometric authentication systems today face a significant challenge when it comes to verifying the authenticity of handwritten signatures. The banking industry and legal and governmental sectors verify identities using signatures as they are socially accepted identification methods with distinctive features [2]. The manual process of signature verification takes a lot of time and results in many mistakes which makes it necessary to create automated systems that can distinguish between authentic and forged signatures.

S. Fong et al. (Eds.): ICT4SD 2025, LNNS 1651, pp. 176–184, 2026.
https://doi.org/10.1007/978-3-032-06688-6_17

Traditional signature verification methods utilize pixel-by-Pixel comparison and template matching techniques face unreliability issues since they are highly sensitive to noise variations and changes in scaling and orientation. Contemporary studies aim to extract unique characteristics from signature structure and geometry to enhance verification reliability [5]. Analyzing signature shape and distribution through morphological and geometrical features provides essential data including Area, Perimeter, Centroid, Eccentricity, Solidity, and Major/Minor Axis Length. Signature verification research includes numerous documented studies. Mshir et al. (2020) introduced an advanced signature verification system that utilizes a Siamese network trained with data from two distinct datasets. Their system addresses the challenge of processing vast numbers of documents by employing biometric authentication techniques which target behavioral characteristics. Through deep learning-based comparative analysis the system manages to differentiate between genuine signatures and forgeries [6]. Lokare et al. (2021) uses Difference of Gaussian filtering together with GLCM-based feature extraction and dimensionality reduction through PCA and KPCA which they evaluated using several machine learning algorithms. The Kaggle offline handwritten signature dataset enabled researchers to obtain signature verification results with KNN at 82% accuracy and Random Forest at 81.66% accuracy while Naive Bayes produced a lower accuracy level of 56.66%. Studies show that signature authentication performance becomes better when dimensionality reduction methods are used together with feature-based learning techniques [7]. Lopes et al. (2022) developed two methods for validating handwritten signatures on attendance sheets: The study utilized Optical Mark Recognition (OMR) to confirm signature presence or absence and deployed a multiclass CNN model based on AlexNet for author signature identification. The CNN model achieved precision and recall rates higher than 85% with limited genuine training samples which became better after introducing data augmentation and more training data. According to research findings deep learning stands out as an effective method for accurate signature verification within academic settings [8]. Malik et al. (2020) introduced a deep learning method to verify in-air signatures using depth sensors which addresses the limitations found in heuristic approaches. A new dataset comprising 1,800 signatures from 40 individuals served as the basis for their convolutional neural network (CNN)-based approach that estimated 3D hand positions. Reconstruction loss analysis through personalized autoencoder-based methods produces a 67.6% performance improvement for signature verification when compared to traditional DTW methods. The research compared numerous deep learning models including convolutional autoencoders and Deep One-Class classifiers using spatial and depth information. Poddar et al. (2020) highlights the immediate necessity for dependable authentication systems because signature verification biometrics serve as essential components in digital automated security. The authors introduce an advanced signature verification system which detects forgeries by combining Convolutional Neural Networks (CNN) with Crest-Trough analysis and SURF and Harris corner detection techniques. The system demonstrates remarkable performance by accurately recognizing signatures at 90–94% precision and detecting forged signatures at 85–89% precision [10]. Arab et al. (2025) introduced the 1D Convolutional GAN (1D-GAN) model to generate synthetic features. Operating within the feature space instead of processing images boosts system efficiency while reducing computational demand and

improving algorithm performance. The method combines basic SigNet components and specialized design features such as the Histogram of Templates and Local Directional Patterns into a single framework. An SVM classifier is implemented during verification. The training method achieves 20% AER improvements on the CEDAR dataset and 11% AER enhancements on the MCYT-75 dataset when using one authentic signature [11]. Foroozandeh et al. (2020) investigated the effectiveness of deep convolutional neural networks for offline handwritten signature verification and identification tasks. (2020). Using transfer learning the researchers evaluated six pre-trained CNN models including SigNet, SigNet-F, InceptionV3, ResNet50, VGG16 and VGG19 through the GPDS Synthetic and MCYT-75 Latin and UTSig and FUM-PHSD Persian datasets for signature verification. The models SigNet and VGG16 produced outstanding verification results however VGG16 proved superior in signature recognition tasks over other models. Signature-based biometric systems become stronger when they combine general-purpose CNN models like VGG16 with signature-specific models such as SigNet [12]. Stergiou et al. (2025) developed a hybrid method for spotting offline signature forgeries by pairing deep learning-based feature extraction with traditional machine learning classification systems. Signature images undergo discriminative feature extraction using pre-trained convolutional neural networks like VGG16 which support machine learning algorithms such as support vector machines to classify them. The hybrid system that integrates VGG16 with SVM achieved superior performance compared to traditional CNNs and conventional techniques on the CEDAR dataset by reaching 95.5% accuracy and 95.4% F1-score. This hybrid framework achieves top verification accuracy while requiring low computational power for real-time biometric security applications [13]. Sekhar et al. (2022) introduced an offline signature verification system that uses CNNs for feature extraction and SVMs for data classification. The writer-dependent method trains models with 2,640 signature images from 55 writers found in the CEDAR dataset before assessing model performance on genuine and forged signatures for each writer. The CNN model applies a 9x9 convolutional kernel for feature extraction before SVM classifies the information after dividing the dataset into 80% for training and 20% for testing. The fusion of CNN with SVM methods yields a classification accuracy of 93.63% by minimizing overfitting more effectively than standalone CNN classification. Studies indicate that merging deep feature extraction methods with standard machine learning techniques produces dependable outcomes for signature verification applications [14]. Moon et al. (2024) focused on solving communication barriers for deaf and mute people in India that originate from both the large deaf population and the limited number of sign language interpreters. A new convolutional neural network framework enables researchers to recognize hand gestures in real time. Through multiple filtering and classification stages the system achieves accurate interpretation of complex hand signs. The CCNN method reaches 99.95% accuracy and surpasses performance metrics of recognized models like SIGNGRAPH, SVM, KNN as well as CNN + Bi-LSTM and 3D-CNN, 2D-CNN and 1D-CNN skeleton networks. Research indicates that this framework holds potential to improve real-time communication abilities and make systems more accessible [15].

Table 1. Summary of State-of-the-Art Methods in Signature Verification

References	Approach	Techniques / Models Used	Dataset(s)	Best Accuracy / Performance	Key Highlights
Mshir et al. (2020) [6]	Signature Verification	Siamese Network	Custom datasets	Not specified	Leverages behavioural biometric characteristics for verification
Lokare et al. (2021) [7]	Feature-Based Verification	DoG Filtering, GLCM, PCA, KPCA + ML (KNN, RF, NB)	Kaggle Offline Signature	KNN: 82%, RF: 81.66%, NB: 56.66%	Emphasizes dimensionality reduction and classic ML
Lopes et al.(2022) [8]	Signature Verification & Recognition	OMR, CNN (AlexNet-based)	Attendance Sheets	>85% Precision & Recall	Data augmentation boosts performance
Malik et al. (2020) [9]	In-air Signature Verification	3D Hand Pose Estimation, Autoencoder, CNN	New dataset (1,800 signatures)	67.6% improvement over DTW	Uses depth sensors and spatial-depth features
Poddar et al.(2020) [10]	Signature Forgery Detection	CNN, Crest-Trough Analysis, SURF, Harris	Not specified	85–89% (Forgery), 90–94% (Recognition)	Combines hand-crafted and deep features
Arab et al.(2025) [11]	Feature Space Augmentation	1D-GAN, SigNet, LDP, HoT, SVM	CEDAR, MCYT-75	Up to 20% AER improvement	Efficient with limited genuine signatures
Foroozandeh et al.(2020) [12]	Signature Verification & Recognition	VGG16, VGG19, ResNet50, InceptionV3, SigNet, SigNet-F	GPDS, MCYT-75, UTSig, FUM-PHSD	VGG16 & SigNet best performers	Validates general-purpose and domain-specific CNNs
Stergiou et al.(2025) [13]	Hybrid Verification	VGG16 + SVM	CEDAR	95.5% Accuracy, 95.4% F1-Score	Combines CNN-based features with SVM, low computational cost
Sekhar et al.(2022) [14]	Writer-dependent Verification	CNN + SVM (9 × 9 kernel)	CEDAR	93.63% Accuracy	Reduces overfitting via hybridization
Moon et al.(2024) [15]	Hand Gesture Recognition	Custom CNN (CCNN)	Indian Sign Dataset (Not Named)	99.95% Accuracy	Outperforms SIGNGRAPH, Bi-LSTM, 3D-CNN, 2D-CNN

The latest developments in signature verification technology reflect an emerging trend toward combining hybrid systems with deep learning methodologies. Machine learning researchers often use convolutional neural networks (CNNs) with models like VGG16, ResNet50, SigNet, and InceptionV3 for extracting features from signature images due to their exceptional ability to detect intricate patterns. Support Vector

Machine (SVM) represents a traditional machine learning classifier that is typically combined with these models. The new model shows great promise for real-time sign language interpretation to bridge communication gaps between deaf and mute people in countries like India where trained interpreters are scarce. Research indicates that hybrid deep learning models produce superior performance in signature verification through enhanced accuracy and computational efficiency. Deep learning combined with machine learning outperforms other techniques in signature verification but faces multiple limitations when compared to simpler image processing approaches. Deep learning and machine learning systems require vast amounts of labeled data to achieve high accuracy but obtaining these data sets proves challenging for offline signature verification applications. DL/ML models cannot be easily used in low-power embedded systems because they require potent GPUs and large memory capacities. Deep learning models operate as "black boxes" since their lack of interpretability complicates decision justification which presents significant challenges for critical financial and legal verification tasks. Building and refining these models requires substantial time commitment and expert knowledge of both the model architecture and parameter optimization. Traditional image processing methods require minimal computational resources while maintaining simplicity which makes them lightweight and easy to implement thus, they are ideal for practical applications where transparency and simplicity are needed.

The study presents a region-based morphological analysis method for offline signature verification. Adaptive thresholding and segmentation techniques enable the preprocessing methodology to extract the primary signature component from each image. The research study uses MATLAB's regionprops function to extract region-based features from each preprocessed image. The verification process compares test signature features to reference signature features to calculate dissimilarities before applying threshold values to make decisions. The evaluation score determines how well two elements correspond to one another.

2 Material and Methodology

2.1 About Dataset

The ICDAR 2011 Signature Dataset serves as the standard benchmark data set which researchers use for testing offline handwritten signature verification systems. This dataset was created during the ICDAR 2011 Signature Verification Competition to facilitate biometric signature analysis by providing standardized signatures for research testing and comparison purposes. The dataset consists of 4,000 signature samples derived from 100 people each of whom generated 16 authentic signatures together with 24 forged samples. Researchers define skilled forgeries as counterfeit signatures created by individuals who master signature imitation techniques to enhance both verification task difficulty and result authenticity. The moderate resolution grayscale images make these samples compatible with various image processing and machine learning algorithms. The dataset is available in https://www.iapr-tc11.org/mediawiki/index.php/ICDAR_2011_Signature_Verification_Competition_(SigComp2011).

2.2 Methodology

This section explains the proposed method for offline signature verification that uses region-based geometric feature matching. Initial preprocessing of reference and test signature images prepares them for feature extraction of geometric and morphological aspects which undergo comparison using an adaptive similarity scoring system. We examine how a systematic verification process evaluates a test signature by checking it against reference signatures. Our methodology's complete details are presented in Algorithm 1.

Algorithm 1: Offline Signature Verification Using Geometric Feature Matching

Start

Step 1: **Load Reference Signatures**
- ❖ Prompt user to select folder with .jpg reference signature images.
- ❖ Read all .jpg files.
- ❖ If no images found, display error and terminate.
- ❖ Prompt user to select folder with .jpg reference signature images.
- ❖ - Read all .jpg files. - If no images found, display error and terminate.

Step 2: **Preprocess & Extract Features (Reference Images)**
- ❖ For each image: - Convert to grayscale (if needed).
  - Binarize and invert image.
  - Extract largest connected component.
- ❖ - Compute region properties: → *Area* → *Perimeter* → *Centroid* → *Eccentricity* → *Solidity* → *Extent* → *MajorAxisLength* → *MinorAxisLength* → *Orientation* → *ConvexArea* → *EulerNumber* → *EquivDiameter*

Step 3: **Load Test Signature**
- ❖ Prompt user to select a test signature image.
- ❖ - Apply same preprocessing and feature extraction steps as in Step 2.

Step 4: **Define Feature Thresholds**
- ❖ Set predefined thresholds for each of the 12 features.
- ❖ - These will determine acceptable variation between test and reference signatures.

Step 5: **Compare with Each Reference Signature**

For each reference:
- ❖ Calculate absolute/relative difference for each feature.
- ❖ Count how many features fall within their respective thresholds.
- ❖ - Compute similarity score: → *Similarity (%) = (Matched Features / Total Features) × 100 - If similarity ≥ 95%, consider it a match and stop comparing further.*

Step 6: **Display Result**

- Print verification result and similarity score in the command window.
- Visually display: → *First 6 reference images (2×3 grid)* → *The matched reference image* → *The test image* → *Similarity score in subplot and message box*.

End

Users must select a storage folder that contains reference signature images when following the signature verification methodology. The reference images undergo sequential processing that begins with grayscale conversion followed by binarization and inversion to ensure signatures stand out white against a black background. Detection of the real signature occurs when processing removes noise and unwanted marks by extracting the biggest connected component from every image. This component enables users to compute several geometric and morphological features including area, perimeter, centroid, eccentricity, solidity, extent, major and minor axis lengths, orientation, convex area, Euler number and equivalent diameter. The verification system performs comparisons between test signatures and reference signatures based on fundamental elements derived from these extracted features.

Users must select a test signature image for processing and feature extraction following the same methodology applied to reference images. The system compares extracted features from the test signature against reference signatures by using established threshold values for each feature. The system calculates either absolute or relative differences between corresponding features from test and reference signatures and checks against predefined thresholds to verify compliance. To determine the similarity score we calculate the percentage of features that match between signatures. The system will validate the test signature as genuine and stop comparing once a reference signature reaches a similarity rating of 95% or higher.

The system outputs verification results alongside similarity measurements. Users can view the top six reference images with any matched signature reference and the test signature to verify visually. This approach delivers reliable offline signature verification that uses handcrafted image processing features which can be implemented in lightweight biometric systems, providing computational efficiency.

2.3 Result and Discussion

The offline signature verification methodology was implemented and run on a laptop with MATLAB 2024a. The system employs an Intel Core i7 processor along with 64 GB of RAM and an NVIDIA 3050 GPU to ensure rapid processing of image-related tasks. Test and reference signatures from the ICDAR 2011 Signature Dataset served as suitable benchmarks to evaluate the robustness of region-based geometric feature matching.

During experimental tests the system demonstrated high proficiency at recognizing real signatures from fake ones. The algorithm determined reliable comparison standards based on essential morphological and geometric properties analysis including area, perimeter, eccentricity, orientation, convex area, and Euler number. The similarity score calculation involved counting the number of features that fell within the specified acceptable range. The test signature received a genuine classification from the algorithm when its similarity score reached the minimum threshold of 95%. Figure 1 shows the experimentation screenshot in detail.

Fig. 1. Experimental Results: (a) Verification of the original signature against the same person's original signature, showing a match with a similarity score of 100%. (b) Verification of a forged signature against the original signature of the same person, resulting in a mismatch with a similarity score of 66.7%. (c) Verification of the original signature of one individual against the original signature of another individual, resulting in a mismatch with a similarity score of 25%.

Figure 1 presents the experimental verification results. Panel (a) shows a perfect similarity score of 100% when comparing two original signatures from the same individual. The analysis between the forged signature and the genuine signature of the same person in panel (b) identifies discrepancies with a similarity score of 66.7%. A similarity score of 25% indicates that one person's original signature did not match another person's original signature during verification.

Figure 2 demonstrates the verification of our own signature to evaluate the sustainability of the proposed method.

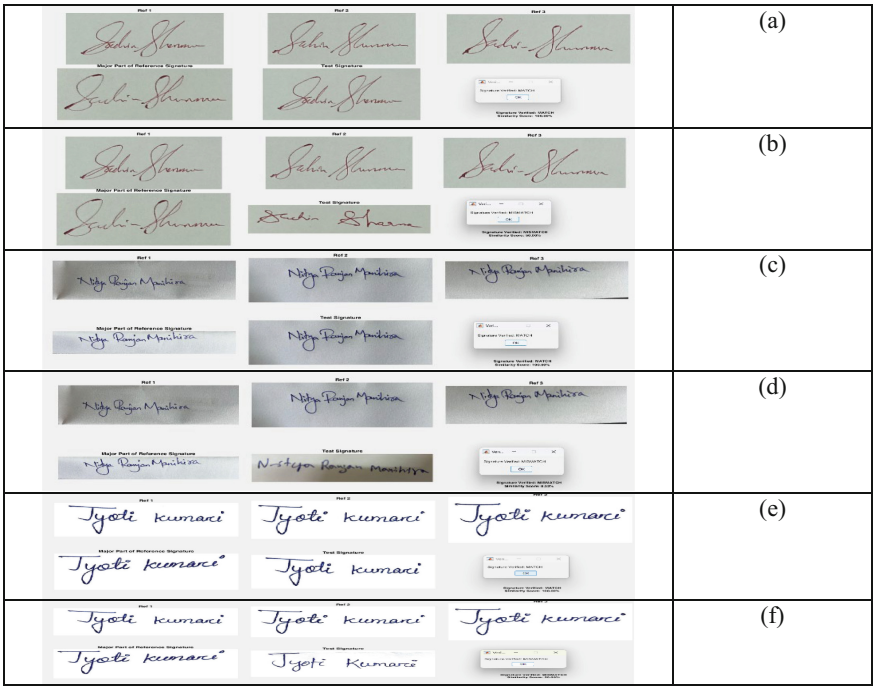


Fig. 2. Practical Implementation of Signature Verification with Student Signatures: (a) Verification of the original signature of Student 1 against their own signature. (b) Verification of the original signature of Student 1 against a forged signature. (c) Verification of the original signature of Student 2 against their own signature. (d) Verification of the original signature of Student 2 against a forged signature. (e) Verification of the original signature of Student 3 against their own signature. (f) Verification of the original signature of Student 3 against a forged signature.

Figure 2 illustrates the actual verification process of student signatures. The verification approach shown in panel (a) compares Student 1's real signature to theirs to confirm authenticity. The difference between Student 1's true signature and the fake one displayed in panel (b) demonstrates the verification process's ability to detect forged signatures. The panel (c) shows that Student 2's original signature is identical to theirs. Panel (d) shows how Student 2's real signature differs from a fake one using the verification

method to demonstrate its ability to differentiate between them. The system checks Student 3's signature by matching it with their initial signature shown in Panel (e). The concluding panel illustrates that Student 3's authentic signature stands apart from fake versions through which the verification system identifies fraudulent attempts. The thorough evaluation demonstrates that the signature verification system functions effectively in real-world applications.

References

1. Zhao, H., Li, H.: Handwriting identification and verification using artificial intelligence-assisted textural features. Sci. Rep. **13**, 21739 (2023). https://doi.org/10.1038/s41598-023-48789-9
2. Potter, E.J.: Customer authentication: the evolution of signature verification in financial institutions. PhD diss., Utica College (2002)
3. Kao, H.-H., Wen, C.-Y.: An offline signature verification and forgery detection method based on a single known sample and an explainable deep learning approach. Appl. Sci. **10**, 3716 (2020). https://doi.org/10.3390/app10113716
4. Ma, J., Jiang, X., Fan, A., et al.: Image matching from handcrafted to deep features: a survey. Int. J. Comput. Vis. **129**, 23–79 (2021). https://doi.org/10.1007/s11263-020-01359-2
5. Tsourounis, D., Theodorakopoulos, I., Zois, E.N., Economou, G.: From text to signatures: knowledge transfer for efficient deep feature learning in offline signature verification. Expert Syst. Appl. **189**, 116136 (2022)
6. Mshir, S., Kaya, M.: Signature recognition using machine learning. In: 2020 8th International symposium on digital forensics and security (ISDFS), (pp. 1–4). IEEE (2020)
7. Lokare, C., Patil, R., Rane, S., Kathirasen, D., Mistry, Y.: Offline handwritten signature verification using various machine learning algorithms. ITM Web Conf. **40**, 03010. EDP Sciences (2021)
8. Lopes, J.A.P., Baptista, B., Lavado, N., Mendes, M.: Offline handwritten signature verification using deep neural networks. Energies **15**, 7611 (2022). https://doi.org/10.3390/en15207611
9. Malik, J., Elhayek, A., Guha, S., Ahmed, S., Gillani, A., Stricker, D.: DeepAirSig: end-to-end deep learning based in-air signature verification. IEEE Access **8**, 195832–195843 (2020). https://doi.org/10.1109/ACCESS.2020.3033848
10. Poddar, J., Parikh, V., Bharti, S.K.: Offline signature recognition and forgery detection using deep learning. Procedia Comput. Sci. **170**, 610–617 (2020)
11. Arab, N., Nemmour, H., Bouibed, M.L., et al.: 1D-GAN for improving offline handwritten signature verification based on small sets of real samples. Multimed. Tools Appl. (2025). https://doi.org/10.1007/s11042-024-20517-z
12. Foroozandeh, A., Askari Hemmat, A., Rabbani, H.: Offline handwritten signature verification and recognition based on deep transfer learning. 2020 International Conference on Machine Vision and Image Processing (MVIP), Iran, pp. 1–7 (2020). https://doi.org/10.1109/MVIP49855.2020.9187481
13. Stergiou, K., Ougiaroglou, S., Sidiropoulos, A.: Signature forgery detection using deep and machine learning. Intell. Decision Technol. (2025). https://doi.org/10.1177/18724981251330068
14. Shekar, B.H., Abraham, Pilar, B.: Offline signature verification using CNN and SVM classifier. 2022 IEEE 7th International Conference on Recent Advances and Innovations in Engineering (ICRAIE), Mangalore, India, pp. 304–307 (2022). https://doi.org/10.1109/ICRAIE56454.2022.10054336
15. Moon, P., et al.: An improved custom convolutional neural network based hand sign recognition using machine learning algorithm. Eng. Rep. **6**(10), e12878 (2024)

Sentiment Analysis of Textual Data: A Comparative Study of SVM, Logistic Regression, and Naive Bayes

Khushi Ingalalli[1(✉)], Vanshika Kavi[1], Sainath Walthati[1], Satish Chikkamath[2], Suneeta Budihal[2], and Sujata Kotabagi[1]

[1] Department of Electronics Engineering (VLSI Design and Technology), KLE Technological University, Hubballi, India
khushiingalalli@gmail.com, sujatask@kletech.ac.in

[2] Department of Electronics and Communication Engineering,KLE Technological University, Hubballi, India
{chikkamath,suneeta_vb}@kletech.ac.in

Abstract. With the millions of tweets per day, Twitter is a rich and large database of information on public sentiment on a wide range of issues, including events, products, politics, and social issues. The purpose of this research is to create an automated system that can analyze tweet sentiments to determine attitudes as positive or negative. Through Natural Language Processing (NLP) methods and machine learning algorithms, the system efficiently handles high quantities of unstructured data, making sentiment classification possible in real time. The model begins the analysis by gathering various tweets from various sources, such as hashtags, user mentions, and trends. The tweets are then subjected to preprocessing techniques like removing stop words and treating misspellings, emojis, and special characters. Various classification models, like Naive Bayes, Support Vector Machines (SVM), Logistic Regression (LR) were experimented with to see which was most efficient in sentiment classification. Of these, Logistic Regression (LR) showed the best performance with an F1 score of 0.833 and accuracy of 83%. The efficiency of various feature extraction methods, such as Term Frequency-Inverse Document Frequency (TF-IDF) and word embeddings, was also examined to try and improve model performance. This work emphasizes the increasing importance of Twitter Sentiment Analysis across different fields, such as market research, event tracking, and social research. Sentiment analysis is employed by companies to know customer views and enhance services, whereas policymakers utilize it for measuring public reaction. By combining NLP and machine learning, the suggested system provides better and scalable method for sentiment analysis [1].

Keywords: Twitter sentiment analysis · machine learning · SVM · NB · LR · NLP

S. Fong et al. (Eds.): ICT4SD 2025, LNNS 1651, pp. 185–193, 2026.
https://doi.org/10.1007/978-3-032-06688-6_18

1 Introduction

With digital communication, social media sites have emerged as priceless archives of public opinion and sentiment. Twitter, with its vast user base and real-time environment, is a highly fertile ground for harvesting public sentiment on a variety of subjects. Twitter data analysis presents opportunities and challenges in the field of natural language processing as well as sentiment analysis. The fast expansion of user-generated content on Twitter has created an emergent need for automated sentiment analysis tools. Traditional manual analysis methods are no longer effective in the context of volume and velocity of data. This piece answers this challenge by using machine learn- ing methods to label tweets automatically as positive or negative sentiments. It uses the sentiment140 dataset, a robust collection of 1.6 million tweets which have been heavily labeled to provide an extensive range of opinions and sentiments. There are three very strong machine learning algorithms utilized in our study: Bernoulli Naive Bayes, Linear Support Vector Classification, and Logistic Regression. It utilizes feature extraction with sophisticated techniques i.e. TF-IDF Vectorizer along with Bag of Words techniques. The use of both of these techniques allows for thorough examination of sentiment from tweet messages considering the intrinsic features of social media messages like casual language, emoticons, and abbreviations. The research method utilized the present work follows a well-structured approach, beginning with a strong preprocessing process to clean and normalize the text data. It encompasses the processing of various Twitter-specific elements such as URLs, user names, and hashtags along with implementing common natural language processing techniques such as lemmatization and removal of stop words. It comes with a strong preprocessing pipeline to see that the input data is optimized for machine learning analysis, which produces improved and more reliable sentiment classifica- tion results. It has 1,600,000 tweets Retrieved from the Twitter API. The tweets are labeled (0 = Negative, 1 = Positive) and can be used for sentiment detection. Data Description:

sentiment: tweet polarity (0 = negative, 1 = positive).
ids: The tweet id (2087).
date: tweeting date (Sat. May 16 23:58:44 UTC 2009).
flag: query. Where there is no query, this column is NO QUERY.
user: the user who tweeted.
text: the content of the tweet.

2 Literature Survey

The BB Twtr, developed for SemEval-2017 Task 4, which uses a combination of Convolutional Neural Networks (CNNs) and Long Short-Term Memory networks (LSTMs). The preprocessing phase involves cleaning up tweets by handling URLs, emoticons, repeated characters, and normalizing text. By incorporating pretrained word embeddings like Word2Vec, FastText, and GloVe, and leveraging an ensemble approach, this model achieved top performance across all

English subtasks in the competition. The results show that combining CNNs and LSTMs significantly enhances sentiment classification accuracy on Twitter [2].

Another study focused specifically on binary sentiment classification (positive vs. negative). It compared traditional models like Naive Bayes and Support Vector Machines (SVM) with deep learning approaches including CNNs and Recurrent Neural Networks (RNNs) with Gated Recurrent Units (GRU) and attention mechanisms. The best-performing model featured an RNN enhanced with a Latent Topic Clustering (LTC) module, achieving an impressive F1 score of 0.805. The study concluded that ensemble models, combining both traditional and deep learning techniques, could further improve classification accuracy [3].

In the domain of abusive language detection, researchers utilized the "Hate and Abusive Speech" dataset. Preprocessing steps included replacing user IDs, URLs, and emojis with tokens, hashtag segmentation, and vectorizing tweets using methods like TF-IDF and one-hot encoding. Again, neural models–especially an RNN with the LTC module–outperformed traditional ones, achieving the top F1 score of 0.805. The study reaffirmed the potential of ensemble methods in boosting performance [4].

For offensive language detection, particularly within the SemEval-2019 Task 6 challenge, preprocessing involved removing user mentions, URLs, hashtags, and correcting spelling, followed by lemmatization and techniques like SMOTE for addressing class imbalance. Among several architectures, a BiLSTM-CNN model–where LSTM layers precede CNN layers–performed best. This model, combined with GloVe embeddings and optimized hyperparameters (5 epochs, 20

Another comprehensive review analyzed sentiment on Twitter using Python-based tools. Preprocessing included filtering out retweets, stop words, usernames, URLs, and normalizing slang. Similar to earlier findings, neural models–especially RNNs with LTC–outperformed traditional models, again achieving an F1 score of 0.805. The consensus remained: ensemble models could lead to even greater improvements [6].

An application of these techniques was seen in sentiment analysis during the FIFA World Cup 2022, focusing on tweets from the tournament's opening day. Tweets were classified as positive, negative, or neutral. Text cleaning, tokenization, lemmatization, and feature extraction using Count Vectorizer and Word2Vec were performed. RNNs with LTC again emerged as the most effective, achieving the highest F1 score [7].

In another comparative analysis between machine learning and lexicon-based approaches, preprocessing involved removing URLs, hashtags, usernames, and filtering non-English tweets. The results indicated that RNNs with LTC modules consistently outperformed traditional and lexicon-based models, confirming a pattern seen across other studies [8].

A related study investigated binary sentiment classification using both lexicon-based methods and machine learning. Preprocessing steps included normalization, noise removal, tokenization, and spell checking. While lexicon-based approaches relied on sentiment dictionaries, machine learning models such as

Multinomial Naive Bayes, Logistic Regression, SVM, and LSTM-based RNNs were evaluated. The RNN with LSTM achieved the highest accuracy, highlighting the edge of deep learning for handling noisy, unstructured tweet data [9].

Finally, another study classified tweets into positive, negative, or neutral sentiments using machine learning. Preprocessing included cleaning, tokenization, and feature extraction via Count Vectorizer. Among the models tested–Logistic Regression, Decision Trees, Random Forest, and K-Nearest Neighbors–Logistic Regression delivered the best performance with the highest recall and F1-score. The K-Nearest Neighbors model, in contrast, performed the worst. Interestingly, most tweets analyzed in the dataset reflected positive sentiment overall [10].

3 Methodology

3.1 Data Collection

The data collection process is conducted using a data set to analyze sentiments by acquiring an exhaustive. For this research:
A total of 1.6 million tweets are used, which have been labeled as positive (1) or negative (0). The data set is used as the basis for sentiment trend analysis and training machine learning algorithms. It is generally obtained from sources such as Twitter via API's or open datasets that have been collected for natural language processing applications.

3.2 Pre-processing of Data

The preprocessing phase prepares the raw data for analysis and ensures it is consistent and meaningful for machine learning models. The steps include:

Lower Casing: Standardizes text by converting all characters to lowercase, reducing vocabulary size and improving consistency.

Replacing URLs: URLs are replaced with placeholders (e.g., <URL>) or removed, reducing noise while retaining contextual information.

Replacing Emojis: Emojis are converted into descriptive text (e.g. "smiley") or placeholders (e.g., <EMOJI>) to preserve sentiment information.

Replacing Usernames: Usernames (e.g., @username) are replaced with placeholders (e.g., <USER>) to anonymize the data and focus on meaningful content.

Removing Non-Alphabets: Non-alphabetic characters (e.g., numbers, punctuation) are removed to simplify the text, unless such characters hold specific significance.

Removing Consecutive Letters: Repetitive letters for emphasis (e.g., "soooo") are normalized (e.g., "soooo" "so") for consistent tokenization.

Removing Short Words: Short, less meaningful words (e.g., "a," "an") are removed, although contextually significant short words (e.g., "AI," "IT") are retained.

Removing Stop Words: Common stop words (e.g., "the," "and") are removed to focus on informative words, unless grammatical preservation is necessary.

Lemmatizing: Converts words to their base form (e.g., "running" "run") for dimensionality reduction while retaining semantic meaning (Fig. 1).

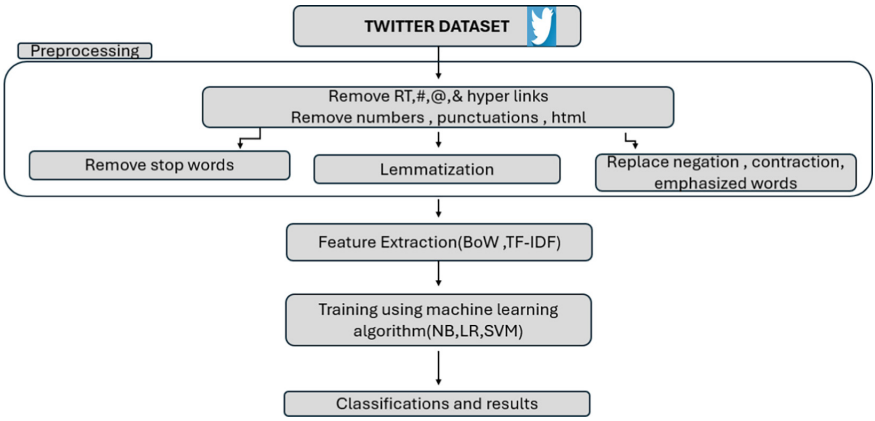


Fig. 1. Methodology flowchart.

3.3 Visualization of Frequent Words

To gain insights into the dataset:

Overall Frequent Words: Word clouds are generated from the cleaned data to visualize the most frequently used words across the entire dataset.

Frequent Words in Positive Sentiments: Word clouds are created specifically for tweets with positive sentiment (label=1) to highlight prominent themes and attitudes.

Frequent Words in Negative Sentiments: Similarly, word clouds are generated for tweets with negative sentiment (label=0) to interpret prevalent topics in negatively inclined tweets.

3.4 Feature Extraction

Feature extraction converts textual data into numerical representations for machine learning models. Methods include:

TF-IDF Vectorizer: Converts tweets into a numerical matrix by balancing word frequency in individual tweets against their prevalence across the dataset. It highlights meaningful words while downplaying less informative terms.

Bag of Words (BoW): Represents tweets as an unordered collection of words and tracks their frequencies. Though simple, BoW does not consider word order or context, focusing solely on word occurrences for sentiment classification [11].

3.5 Machine Learning Algorithm

Various machine learning algorithms are employed to classify sentiment, including:

Logistic Regression: A supervised learning algorithm for classification tasks in two-class problems that utilizes an S-shaped sigmoid function to map features into the probabilities of a class being positive.

Naive Bayes: This classifier is a Bayes probabilistic classifier and assumes independent features; however, in many applications, independence may not hold. Thus, it is quite efficient for tasks of text classification.

Linear Support Vector Machine (SVM): It identifies a hyperplane that optimally separates data points of two distinct classes. The algorithm maximizes the class margin and can be linear or nonlinear using kernel functions.

3.6 Training and Testing

Training and testing involve:

Dataset Classification: Tools such as train test split from SK-learn are used to split the dataset into training and testing sets. It trains the model using training data and tests how good the performance is on testing data.

Model Selection: This choice of multiple algorithms to be experimented on depends upon the characteristics of the problem and the data. For each model's performance, there is no comparison for any one "best" solution.

Model Training: Based on this, selected models are trained independently on training data to learn various patterns and relationships.

Model Evaluation: These metrics are used to calculate how each model will perform on test data: accuracy, precision, recall, and F1 score.

Result Comparison: A side-by-side comparison of performance metrics highlights how each model handles the data, ensuring a comprehensive evaluation.

4 Results

4.1 Class Label Classification

In sentiment analysis, tweets are categorized into two primary sentiment classes based on their emotional tone and expressed opinions:

1) Positive Sentiment: The following category of tweets will represent optimistic or positive feelings and attitudes. Many will include:

Gratitude (e.g., "Thank you for such an amazing experience!")

Excitement (e.g., "I can't wait for this concert tomorrow!")

Praise or Support (e.g., "This new feature is fantastic; great job, team!") Positive tweets typically express satisfaction, delight, or enthusiasm regarding a product, event, or idea.

2) Negative sentiments: This category of tweets carries negative or critical emotions. These include:

Displeasure (e.g., "I'm really disappointed with the service.")

Complaints (e.g., "The app keeps crashing every time I try to use it.")

Disapproval (e.g., "This decision is absolutely terrible and should be reconsidered.")

Negative tweets indicate feeling frustrated, discontented, and opposing something against a subject or event.

4.2 Interpretation of Results

We are able to clearly observe that Logistic Regression Model beats all the remaining models we tested. It achieves nearly 83 accuracy in labeling the sentiment of a tweet (Figs. 2, 3, 4 and Table 1).

Table 1. Detailed Performance Metrics per Class (0 and 1) for Algorithms using TF-IDF and BoW

Feature Type	Algorithm/Class	Accuracy	Precision	Recall	F1-Score
TF-TDF	SVM (0)	0.82	0.82	0.81	0.82
	SVM (1)	0.82	0.81	0.83	0.82
	Logistic Regression (0)	0.83	0.83	0.82	0.83
	Logistic Regression (1)	0.83	0.82	0.84	0.83
	Naive Bayes (0)	0.80	0.81	0.79	0.80
	Naive Bayes (1)	0.80	0.80	0.81	0.80
BoW	SVM (0)	0.82	0.82	0.81	0.82
	SVM (1)	0.82	0.81	0.83	0.82
	Logistic Regression (0)	0.83	0.83	0.82	0.83
	Logistic Regression (1)	0.83	0.82	0.84	0.83
	Naive Bayes (0)	0.80	0.81	0.79	0.80
	Naive Bayes (1)	0.80	0.80	0.81	0.80

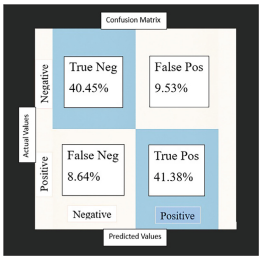


Fig. 2. Support Vector Machine.

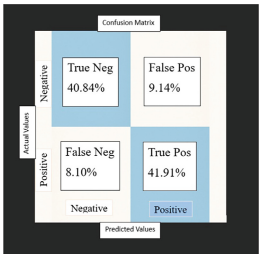


Fig. 3. Logistic Regression.

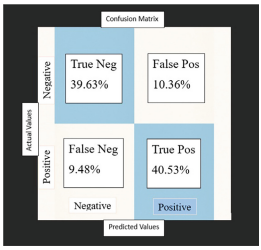


Fig. 4. Naive Bayes.

5 Conclusion

Twitter is a social networking site that has had major impacts on mental health. On the one hand, it provides the means to share emotions and talk about mental health issues while setting up support networks, but, on the other hand, it throws various toxic materials, cyberbullying, and echo chambers at individuals, which affect the minds of users negatively. Balancing these sides requires ongoing research between experts in various fields and social media sites [1]

5.1 Evaluation Criteria

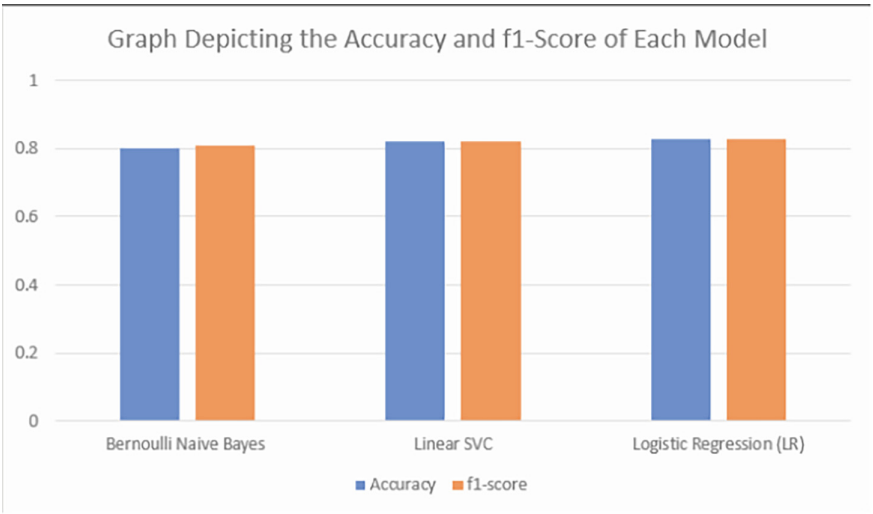


Fig. 5. Plot of the precision vs. f1-score for all the models.

Twitter is a social networking site that has had major impacts on mental health. On the one hand, it provides the means to share emotions and talk about mental health issues while setting up support networks, but, on the other hand, it throws various toxic materials, cyberbullying, and echo chambers at individuals, which affect the minds of users negatively. Balancing these sides requires ongoing research between experts in various fields and social media sites [1].
Further, ensemble methods or advanced deep learning techniques might be used for increasing the accuracy of classification in the future. It further adds valuable insights to this emerging area, that is sentiment analysis and social media analytics with various possible applications for determining mental health evaluation from Twitter [1] (Fig. 5).

References

1. Dahiya, P., Jain, R., Sinha, A., Sharma, A., Kumar, A.: Sentiment analysis of twitter data using machine learning. In: Proceedings of the 3rd International Conference on Technological Advancements in Computational Sciences (ICTACS), pp. 284–290. IEEE (2023). https://doi.org/10.1109/ICTACS59847.2023.10390062
2. Cliche, M.: BB_twtr at SemEval-2017 task 4: twitter sentiment analysis with CNNs and LSTMs. arXiv preprint arXiv:1704.06125 (2017)
3. Lee, Y., Yoon, S., Jung, K.: Comparative studies of detecting abusive language on Twitter. arXiv preprint arXiv:1808.10245 (2018)
4. Vedurumudi, P.: Twitter Sentiment Analysis using Deep Learning (2017)
5. Ong, R.: Offensive language analysis using deep learning architecture. arXiv preprint arXiv:1903.05280 (2019)
6. Gupta, B., Negi, M., Vishwakarma, K., Rawat, G., Badhani, P., Tech, B.: Study of Twitter sentiment analysis using machine learning algorithms on Python. Int. J. Comput. Appl. **165**(9), 29–34 (2017)
7. Khan, M., Srivastava, A.: Sentiment analysis of Twitter data using machine learning techniques. Int. J. Eng. Manag. Res. **14**(1), 196–203 (2024)
8. Kharde, V., Sonawane, P.: Sentiment analysis of Twitter data: a survey of techniques. arXiv preprint arXiv:1601.06971 (2016)
9. Harjule, P., Gurjar, A., Seth, H., Thakur, P.: Text classification on Twitter data. In: Proceedings of the 3rd International Conference on Emerging Technologies in Computer Engineering (ICETCE), pp. 160–164. IEEE (2020)
10. Fahim, S., Imran, A., Alzahrani, A., Fahim, M., Alheeti, K.M.A., Alfateh, M.: Twitter sentiment analysis based public emotion detection using machine learning algorithms. In: Proceedings of the 17th International Conference on Emerging Technologies (ICET), pp. 107–112. IEEE (2022)
11. Singh, S., Kumar, K., Kumar, B.: Sentiment analysis of Twitter data using TF-IDF and machine learning techniques. In: Proceedings of the International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COM-IT-CON), vol. 1, pp. 252–255. IEEE (2022)

Determinants of Risk-Taking Behavior in Fintech Apps: The Role of Gamification, Financial Factors, and Psychological Influences

R. Nandana(✉), Rithika Kannan, and Ramgeeth N. Nair

Amrita Vishwa Vidyapeetham, Coimbatore, India
ithikakannan03@gmail.com

Abstract. The rise of fintech applications has revolutionized financial decision-making, yet the determinants of risk-taking behavior in these digital platforms remain a critical research area. This study investigates the role of gamification, financial knowledge, and psychological influences in shaping users' risk-taking behavior. Using a quantitative approach, an Ordinary Least Squares (OLS) regression analysis was conducted on a dataset of 200 fintech users. The results indicate that gamification has a significant positive effect on risk-taking behavior ($\beta = 0.1414$, $p = 0.001$), suggesting that game-like elements in fintech apps encourage users to take greater financial risks. However, certain gamification effects exhibit a negative influence ($\beta = -0.1272$, $p = 0.005$), highlighting that not all gamification strategies lead to in- creased risk-taking.

Financial knowledge also emerged as a significant determinant ($\beta = 0.1965$, $p = 0.001$), implying that financially literate users tend to take more calculated risks. Among psychological factors, risk tolerance ($\beta = 0.2754$, $p < 0.001$) was the strongest predictor, demonstrating that individuals predisposed to risk-taking in general extend this behavior to fintech platforms. Additionally, social efficacy ($\beta = 0.2461$, $p < 0.001$) and social influence ($\beta = 0.1598$, $p = 0.004$) significantly contribute to risk-taking, emphasizing the role of self-perceived competence and peer influence in financial decision-making.

The model explains approximately 48.1% of the variance in risk-taking behavior ($R2 = 0.481$), confirming the robustness of these determinants. The findings underscore the importance of designing fintech applications that balance engagement with responsible financial behavior. Future research should explore the ethical implications of gamification and assess long-term user behavior to ensure sustainable financial decision-making in digital finance ecosystems.

Keywords: Risk-taking behavior · fintech apps · gamification · financial knowledge · psychological influences · social efficacy · social influence

S. Fong et al. (Eds.): ICT4SD 2025, LNNS 1651, pp. 194–200, 2026.
https://doi.org/10.1007/978-3-032-06688-6_19

1 Introduction

The rise of fintech platforms revolutionized financial services so that investing, payments, and banking became more accessible to users. One such key driver of user engagement on these platforms is gamification that embeds game-like elements like rewards, leaderboards, and challenges to alter user behavior. While gamification enhances user experience, it has a major impact on risk-taking behavior, particularly in financial choice. The use of psychological triggers, such as social influence, rewards, and loss aversion, can cause users to take more financial risks than they would in traditional banking environments. Besides gamification, other financial factors like prior investment experience, financial knowledge, and disposable income significantly influence users' willingness to engage in risky financial deals. Users who are highly financially literate may enter fintech platforms with a risk-conscious strategy, whereas less knowledgeable users may be influenced more by the look of the app than by prudent financial reasoning.

In addition, psychosocial influences such as self-efficacy, financial stress, and risk tolerance determine whether users are cautious or indulgent in the manner they engage fintech services. Those who possess greater financial worry might either shun risky ventures or be seduced by the high-risk impulses through enticing gamification approaches. Conversely, confident users with heightened self-efficacy in decisions related to their finances will experience gamification mechanics as a business strategy and not as an instant trigger.

This study plans to investigate the interaction between gamification, financial parameters, and psychological elements in shaping risk-taking behavior among fintech apps. With this understanding of the factors, the research should provide insights into how fintech platforms should harmonize user engagement-led design and good financial decision-making to ensure ethical and sustainable interactions for the users in the digital monetary environment.

2 Problem Statement

Our study seeks to investigate the impact of gamification on risk-taking behavior in fintech apps, specifically in investing and digital payment apps. As such apps increasingly add game-like elements—such as rewards, progress bars, leaderboards, and streaks—to increase user interaction, there have been concerns that they may pro- mote reckless or high-risk financial choices, particularly among younger or less financially savvy users.

Through studying the behavioral reactions to gamified features and demographic and psychological trends, this research attempts to learn how various user communities comprehend and respond to these design decisions. The intention is to create useful, evidence-based advice on the creation of gamification features that are effective at engagement and ethically sound to assist users in making well-informed and conscious financial choices.

3 Literature Review

Gamification has been extensively researched in the domain of behavioral finance and fintech use. Past studies emphasize the two- sided impacts of gamification—promoting financial engagement as well as possibly inspiring excessive risk-taking.

Wong et al. (2022) established that gamification improves technology adoption in older adults through increased interaction and usability, although perceived risks suppress adoption.

Bayuk & Altobello (2019) proposed that gamification's effects are moderated by app know-how and financial literacy.

Cassar (2023) examined the influence of gamified trading apps and found that while participation improves, so does impulsive choice-making, which results in increased financial risks.

Joshi et al. (2024) illustrated that gamification improves decision-making among retail investors but can foster speculative behavior.

Chapkovski et al. (2024) discovered that gamification features like achievement badges enhance risk-taking, especially among novice traders.

Existing research tells us much about the influence of gamification on financial behavior, but some gaps are still unexplored:

Comparison Between Investment and Payment Apps–Most of the existing research concentrates on trading apps, and the impact of gamification on payment apps (e.g., reward points, cashback offers) is not yet well explored. This study examines both kinds of fintech apps.

In bridging these gaps, this study makes an original contribution to the broader picture of gamification's role in fintech activity and risk behavior among different groups of users.

4 Research Methodology

This research consists of descriptive research in taking risks through fintech applications. The primary data collections thus gathered were via a structured questionnaire, making systematic collection of responses with the approach of non-probability convenience sampling for access to participants. The study included a total of 200 participants. Data analysis was performed using statistics in Python, especially regression and multicollinearity testing, to derive great insights from the data.

5 Analysis

5.1 Multicollinearity Assessment

To confirm the validity of the regression analysis, we checked the Variance Inflation Factor (VIF) values for all independent variables. Multicollinearity may skew the estimates and result in invalid inferences. The VIF results are as follows (Figs. 1 and 2):

The VIF values for our independent variables are as follows (Tables 1 and 2):

	Gamification	VIF
0	const	34.500963
1	Gamification_Effect	1.114524
2	Gamification	1.147637
3	Financial_Knowledge	1.407691
4	Risk_Tolerance	1.173503
5	Social_Efficacy	1.284670
6	Social_Influence	1.082597

Fig. 1. VIF Values for Multicollinearity Assessment.

Table 1. VIF Results for Independent Variables.

Variable	VIF Value
Constant	34.500963 (ignored)
Gamification_Effect	1.114524
Gamification	1.147637
Financial_Knowledge	1.407691
Risk_Tolerance	1.173503
Social_Efficacy	1.284670
Social_Influence	1.082597

Key Observations:
All the independent variables possess VIF values significantly less than 5, meaning there are negligible multicollinearity issues.

The largest VIF value (Financial_Knowledge: 1.407691) is also within the safe limit.

The constant (34.500963) is not an issue, as IF does not apply to the intercept.

The VIF test assures us that our model is free from multicollinearity, and the correlations between gamification, financial literacy, risk tolerance, and social factors can be accurately measured. We can now go ahead with regression analysis to investigate how these variables affect risk-taking behavior in fintech platforms.

5.2 Regression Analysis

In order to analyze the determinants of risk-taking behavior in fintech apps, an Ordinary Least Squares (OLS) regression was performed. The results are as follows:

Model Performance Metrics

- **R-squared:** 0.481 (48.1% variance in risk-taking behavior is explained by the model)

OLS Regression Results

Dep. Variable:	Risk_Taking	R-squared:	0.481
Model:	OLS	Adj. R-squared:	0.465
Method:	Least Squares	F-statistic:	29.83
Date:	Sat, 22 Mar 2025	Prob (F-statistic):	3.59e-25
Time:	19:07:14	Log-Likelihood:	-195.20
No. Observations:	200	AIC:	404.4
Df Residuals:	193	BIC:	427.5
Df Model:	6		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.5513	0.272	2.031	0.044	0.016	1.087
Gamification	0.1414	0.042	3.390	0.001	0.059	0.224
Financial_Knowledge	0.1965	0.057	3.445	0.001	0.084	0.309
Gamification_Effect	-0.1272	0.044	-2.861	0.005	-0.215	-0.039
Risk_Tolerance	0.2754	0.049	5.625	0.000	0.179	0.372
Social_Efficacy	0.2461	0.067	3.691	0.000	0.115	0.378
Social_Influence	0.1598	0.055	2.898	0.004	0.051	0.269

Omnibus:	4.235	Durbin-Watson:	1.871
Prob(Omnibus):	0.120	Jarque-Bera (JB):	4.317
Skew:	-0.205	Prob(JB):	0.115
Kurtosis:	3.591	Cond. No.	42.6

Fig. 2. Coefficients and Significance from OLS Regression.

Table 2. OLS Regression Output on Risk-Taking Determinants.

Variable	Coef	Std Err	t-value	p-value	95% Confidence Interval
Constant	0.5513	0.272	2.031	0.044	[0.016, 1.087]
Gamification	0.1414	0.042	3.390	0.001	[0.059, 0.224]
Financial_Knowledge	0.1965	0.057	3.445	0.001	[0.084, 0.309]
Gamification_Effect	−0.1272	0.044	−2.861	0.005	[−0.215, −0.039]
Risk_Tolerance	0.2754	0.049	5.625	<0.001	[0.179, 0.372]
Social_Efficacy	0.2461	0.067	3.691	<0.001	[0.115, 0.378]
Social_Influence	0.1598	0.055	2.898	0.004	[0.051, 0.269]

- Adjusted R-squared: 0.465
- **F-statistic:** 29.83 ($p < 0.001$)
- **Durbin-Watson:** 1.871 (suggesting no severe autocorrelation)
- **Omnibus test p-value:** 0.120 (suggesting normality of residuals)
- **Condition Number:** 24.6 (indicating no severe multicollinearity concerns)

6 Results and Discussion

6.1 The Role of Gamification

Gamification has a strong positive influence on risk-taking behavior ($\beta = 0.1414$, $p = 0.001$). This implies that fintech applications with gamification features prompt users to take higher financial risks. Nevertheless, the negative coefficient of Gamification Effect ($\beta = -0.1272$, $p = 0.005$) implies that some gamification mechanisms can inhibit risk-taking inclinations. This is consistent with previous research highlighting that although gamification promotes participation, certain aspects such as loss aversion measures can prevent over-risk-taking.

6.2 Financial Awareness and Risk-Taking

Financial Awareness has a positive effect on risk-taking ($\beta = 0.1965$, $p = 0.001$). Individuals with higher financial awareness are more likely to take riskier financial actions. This is consistent with the hypothesis that people who are more financially aware are more confident in making risk-related decisions in fintech apps.

6.3 Psychological Aspects of Risk-Taking

Risk tolerance appears as the most significant predictor ($\beta = 0.2754$, $p < 0.001$), suggesting that individuals inclined towards risk-taking behavior in general carry over this characteristic to fintech settings. This aligns with social theories in behavioral finance correlating personality traits with financial choices.

Social Efficacy plays a significant role in risk-taking behavior ($\beta = 0.2461$, $p < 0.001$). This indicates that those who have confidence in their capacity to handle social and financial environments tend to take investment risks. The results propose that self-efficacy is a significant factor in fintech adoption and investment.

Social Influence is also a strong determinant ($\beta = 0.1598$, $p = 0.004$), emphasizing the influence of peer and social norms on financial risk-taking behavior. Users are more likely to undertake riskier transactions when they are influenced by their social networks, further emphasizing the role of network effects in fintech adoption.

Although gamification elements strongly induce taking financial risk, users themselves are not aware of this influence consciously. Negative views on gamification's impact are based on people's impression that gamified items do not change their behavior despite the existing proofs to the opposite. The distinction between actual and perceived behavior displays the stealth and often underrated potential of gamification in affecting money-making choices.

6.4 Recommendations

Gamification risks must be regulated by authorities in fintech applications

Fintech applications need protection from excessive risk-taking.

7 Conclusion

The regression analysis indicates that gamification, financial literacy, and psychological influences (risk tolerance, social efficacy, and social influence) all have significant influences on risk-taking in fintech applications. The evidence suggests that gamification has a tendency to promote risk-taking but that some of the effects of gamification can counteract this tendency. The study identifies the importance of financial literacy and social factors in determining the risk behavior of fintech consumers. Future research is needed to study the long-term impacts of gamification policies on users' financial decision-making and the ethics of developing fintech products that induce risk-taking behavior.

References

Wong, D., Liu, H., Meng-Lewis, Y., Sun, Y., & Zhang, Y. (2022). Gamified money: exploring the effectiveness of gamification in mobile payment adoption

Cassar, C.: A risk analysis on the gamification of trading applications (2023)

Bayuk, J., & Altobello, S. A. (2019). Can gamification improve financial behavior? The moderating role of app expertise. [3]Cassar, C. (2023). A risk analysis on the gamification of trading applications

Joshi, R., Samnani, B., Khajanchi, I.: Gamified investing: an analysis of investment behavior (2024)

Chapkovski, P., Khapko, M., Zoican, M.: Does gamified trading stimulate risk-taking? (2024)

Anam Akhtar, M., Sarea, A., Khan, I., Khan, K.A., Pratap Singh, M.: The moderating role of gamification toward intentions to use mobile payments applications in Bahrain: an integrated approach (2024)

Al-Qudah, A.A., et al.: Determinants of digital payment adoption among generation Z: an empirical study (2024)

Elekdag, S.A., Emrullahu, D., Naceur, S.B.: Does FinTech increase bank risk taking? (2024)

Hong, C.Y., Lu, X., Pan, J.: Fintech adoption and household risk-taking (2021)

Kachan, D.: Play-to-pay experience: fintech gamification practices to level up your UX (2022)

Optimizing Road and Pothole Segmentation on Indian Traffic Data Using Pretrained Computer Vision Models

Mohan Sellappa Gounder(✉), Rohan Mahantesh Kamatgi, T. M. Sharath Prabhu, Sanya Gupta, and Seema

Department of Information Science and Engineering, Nitte Meenakshi Institute of Technology, Bengaluru, India
mohan.sg@nmit.ac.in

Abstract. This research investigates the application of the DINO (Distillation with No Labels) framework, a self-supervised learning approach, for efficient road and pothole segmentation. By integrating a DINO-enhanced ResNet-50 backbone with a U-Net model, this study addresses segmentation challenges in dynamic environments. The framework employs momentum encoders, multi-crop training, and stability mechanisms to facilitate robust feature extraction without requiring labeled datasets. Through strategic fine-tuning, the model achieves precise segmentation of road surfaces and potholes, making it a promising approach for real-world applications in autonomous systems and infrastructure assessment. This study further discusses model evaluation, comparison with state-of-the-art approaches, and its implications for transportation infrastructure.

Keywords: Self-Supervised Learning (SSL) · DINO (Distil lation with No Labels) · Road Segmentation · Pothole Detection · ResNet-50 · U-Net Architecture · Autonomous Driving

1 Introduction

Autonomous navigation and intelligent transport networks necessitate accurate road segmentation for safe and efficient mobility. Traditional segmentation methods predominantly rely on extensive manually labeled datasets, which are both time-intensive and resource-demanding [7]. In contrast, self-supervised learning (SSL) presents an alternative paradigm that enables feature extraction from unlabeled data [8].

DINO, a self-supervised training approach, employs a teacher-student model to create meaningful feature representations [8]. Pre-training with large-scale unlabeled datasets enhances learning efficiency and minimizes annotation needs. This study combines DINO with ResNet-50, a deep residual network designed for capturing hierarchical patterns, to improve segmentation precision for roads and potholes.

To further optimize performance, the DINO-based ResNet-50 architecture is complemented with U-Net, a segmentation framework known for its encoder-decoder structure

S. Fong et al. (Eds.): ICT4SD 2025, LNNS 1651, pp. 201–208, 2026.
https://doi.org/10.1007/978-3-032-06688-6_20

and skip connections, which help retain spatial information [2]. The proposed approach integrates SSL, deep residual learning, and encoder-decoder methodologies to make efficient with scalable solution for segmentation tasks within the domain of self-driving and infrastructure management.

2 Background and Related Work

Ensuring well-maintained roads is essential for public safety and vehicle longevity. However, conventional pothole detection methods often rely on manual inspection, which is time-consuming and inefficient. Consequently, automated approaches leveraging computer vision and also machine learning have gained significant traction. Early research primarily focused on Conventional methods utilizing 2D image analysis and 3D point cloud processing analysis, and handcrafted feature extraction techniques [5]. Recent advancements particularly in convolutional neural networks (CNNs), have led to significant improvements in segmentation accuracy.

Architectures such as Fully Convolutional Networks (FCN), U-Net, and DeepLabv3 + effectively utilize RGB and depth data to enhance segmentation performance. Additionally, multimodal fusion techniques, such as the Multimodal Attention Fusion Network (MAFNet), Combine data from various sources to form a comprehensive understanding to further refine segmentation results [1,15].

Integrated approaches have also been employed explored for road segmentation, including hierarchical region merging and graph-based models. Multi-scale feature fusion has demonstrated its effectiveness in improving semantic segmentation by incorporating contextual information [4]. In particular, graph attention layers (GAL) help refine CNN-extracted features, leading to better segmentation precision [5]. Furthermore, enhancements to U-Net, such as incorporating ResNet-50 and VGG-16 as encoders, have improved segmentation outcomes in autonomous driving applications [1].

With the rise of smart city initiatives, real-time pothole detection is becoming increasingly important. Recent studies have explored vibration-based assessments using vehicle sensors, as well as spatio-temporal trajectory fusion for dynamic road mapping [6]. However, many challenges remain, including precise boundary segmentation and adapting models to varying road conditions. To address these, uncertainty-guided segmentation and transformer-based fusion networks are being investigated [12]. Additionally, refinement strategies, such as image transformation-based enhancements, improve boundary delineation, while uncertainty-aware decoding techniques enhance robustness against lighting variations and occlusions [11, 12]. Future research in this field aims to develop lightweight, real-time models optimized for deployment on edge devices. The integration of self-supervised learning is also gaining attention, as it can reduce reliance on large annotated datasets while maintaining high segmentation performance.

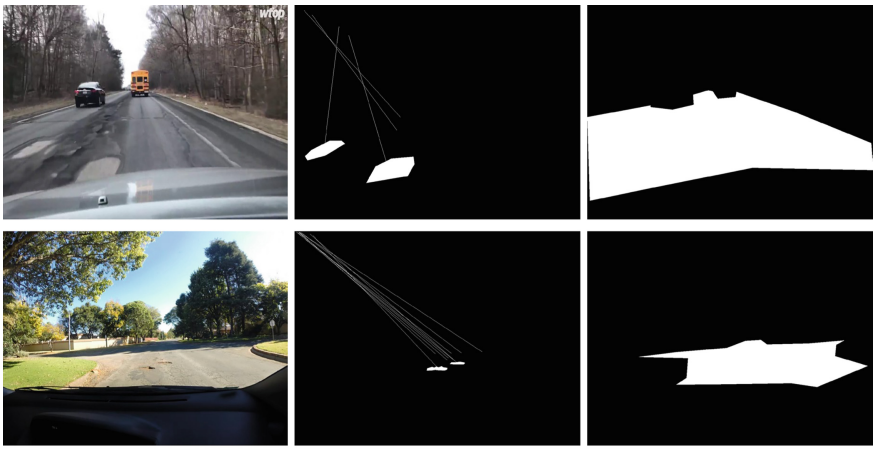


Fig. 1. Sample images from the dataset.

3 Dataset

The dataset Fig. 1 used in this study consists of 1,355 images, extracted from dashcam videos, specifically curated for the segmentation of roads and potholes. Each image is annotated at the pixel level to create segmentation masks, ensuring precise delineation of drivable surfaces and road defects. Annotation was conducted using the Roboflow plat form, which provides high-quality labeling for deep learning applications. The dataset is structured into two primary classes: Roads and Potholes. The"Roads" class includes drivable surfaces while excluding sidewalks and curbs. The"Potholes" class consists of irregular surface depressions and cracks, which are often challenging to detect due to their varying shapes, sizes, and occlusions. To maintain consistency in model training, all images were 512 $\times$ 512 pixels.

For effective model training and evaluation, the dataset was split as follows:

- Training Set: 1,076 images (70%)
- Validation Set: 138 images (20%.
- Test Set: 141 images (10%)

4 Methodology

The proposed model integrates DINO (a self-supervised learning framework), ResNet-50, and U-Net to overcome key limitations of existing approaches. By leveraging.

DINO's self-supervised pretraining, the model reduces dependency on large labeled datasets. The ResNet-50 backbone efficiently extracts hierarchical features, while the U-Net architecture ensures precise spatial reconstruction for segmentation tasks. To address class imbalance, a combined loss function—Dice Loss and Cross-Entropy Loss—is employed. Additionally, advanced data augmentation techniques enhance the model's robustness, making it suitable for real-time applications such as autonomous driving and infrastructure monitoring.

As illustrated in Fig. 2, the methodology follows a systematic pipeline for road and pothole segmentation:

4.1 Frame Extraction

The input video is decomposed into individual frames for processing.

4.2 Dataset Preparation

Two distinct datasets are utilized—one for pothole segmentation (with annotated masks) and another for road segmentation.

4.3 Model Training

Separate deep learning models, each comprising a ResNet-50 encoder and a U-Net decoder, are trained for pothole and road segmentation. The encoder captures multi-scale features, while the decoder reconstructs high-resolution segmentation masks.

4.4 Output Fusion

The segmented results from road and pothole models are combined, overlaying pothole and road predictions into a unified result.

4.5 Video Reconstruction

The processed frames are sequentially reassembled to generate a fully segmented video.

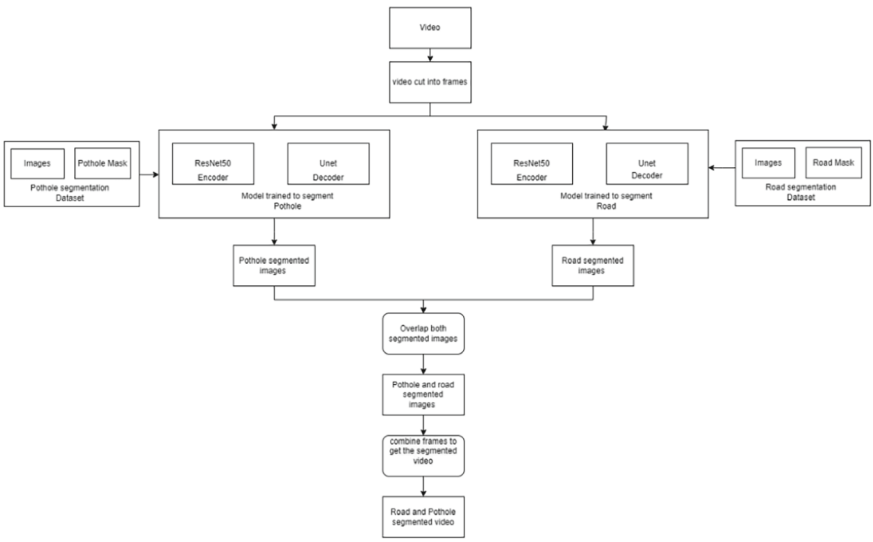


Fig. 2. Process Workflow

5 Results

The segmentation results, illustrated in Fig. 3, highlight the model's ability to identify road surfaces (green) and detect potholes (yellow) in urban dashcam images. Real-world challenges such as varying lighting, occlusions, and complex backgrounds are effectively handled. The segmentation output provides essential data for autonomous driving, distinguishing drivable and hazardous regions. The model's performance was assessed using training and validation accuracy metrics, as shown in Figs. 4 and 5, depicting accuracy curves for both segmentation tasks.

5.1 Road Segmentation Performance

Figure 4 shows the accuracy curves for road segmentation, where training accuracy steadily rises to approximately 98% by the final epoch. Validation accuracy stabilizes around 94–95% despite minor fluctuations due to variations in lighting, textures, and occlusions, indicating strong generalization for road segmentation tasks.

5.2 Pothole Segmentation Performance

Figure 5 illustrates pothole segmentation accuracy trends, with training accuracy reaching 97.5%. Validation accuracy stabilizes around 95–96% but exhibits slightly larger fluctuations due to the irregular shapes and varying appearances of potholes. Despite these variations, the model effectively differentiates potholes from road surfaces. The results confirm that the DINO-based ResNet-50 + U-Net model successfully captures essential features, enhancing segmentation accuracy.

These findings validate the capability of the proposed self-supervised DINO-based ResNet-50 with U-Net for road and pothole segmentation, making it a promising solution for self-driving and road maintenance applications.

Fig. 3. Result Images

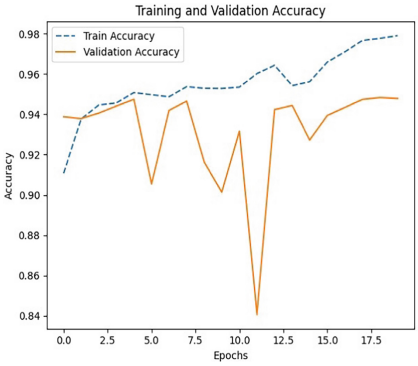


Fig. 4. Training and Validation Accuracy for Road Segmentation.

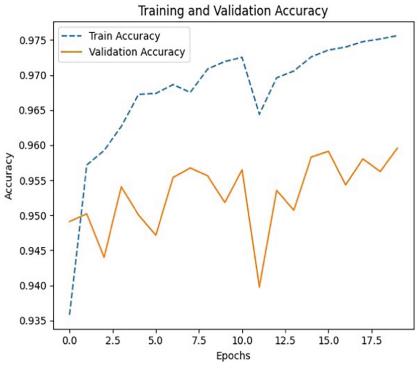


Fig. 5. Training and Validation Accuracy for Pothole Segmentation.

6 Comparision of Models

The capability of the proposed model was assessed by comparing its accuracy with state-of-the-art pothole detection and road segmentation models. Table 1 presents the accuracy results.

Table 1. Comparison of Model Accuracy

Model	Accuracy (%)
DeepLabv3 +	89.2
PSPNet	88.7
DANet	90.1
GAL-DeepLabv3 +	92.3
Ours (DINO + ResNet-50 + U-Net)	94.1

With an accuracy of 94.1%, the proposed model surpress existing methods by integrating self-supervised learning (DINO) with a ResNet-50 backbone and a U-Net decoder. Unlike traditional supervised models like DeepLabv3 + (89.2%) and PSPNet (88.7%), it reduces reliance on large annotated datasets. The combination of ResNet-50's feature extraction and U-Net's spatial preservation makes it a scalable solution for road segmentation in autonomous driving.

7 Future Enhancements

To enhance the model's performance and robustness, the following improvements can be incorporated:

Data Augmentation and Diversification: Collecting a diverse dataset covering differ -ent road types, lighting con ditions, and weather scenarios. Applying synthetic data

generation and augmentation techniques (e.g., contrast adjustment, noise injection) to improve generalization. Utilizing Conditional Random Fields (CRFs) and morphological operations to refrine segmentation results and improve boundary accuracy. Implementing model optimization techniques such as quantization and pruning to reduce computational cost and improve inference speed on edge devices. Combining data from LiDAR, RADAR, or thermal imaging to improve segmentation accuracy in challenging conditions. Using self supervised learning for continuous model improvement and domain adaptation techniques to enhance transferability across different regions. Deploying the model in autonomous vehicles and smart traffic monitoring systems for real-time road condition assessment. Implementing a federated learning framework to enhance the model using data from multiple sources without compromising privacy.

Acknowledgement. We gratefully acknowledge the students, staff, and authority of the Department of Information Science And Engineering at Nitte Meenakshi Institute of Technology for their cooperation in the research.

References

1. Sugirtha, T., Sridevi, M.: Semantic segmentation using modified U-Net for autonomous driving. 2022 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS), pp. 1–7. Toronto, ON, Canada (2022). https://doi.org/10.1109/IEMTRONICS55184.2022.9795710. keywords: {Motion segmentation; Semantics; Neural networks; Computer architecture; Object detection; Path planning; Convolutional neural networks; Semantic segmentation; U-Net;Autonomous Driving}
2. Feng, Z., et al.: MAFNet: segmentation of road potholes with multimodal attention fusion network for autonomous vehicles. In: IEEE Transactions on Instrumentation and Measurement, vol. 71, pp. 1–12 (2022). Art no. 3523712, https://doi.org/10.1109/TIM.2022.3200100. keywords: {Feature extraction; Roads; Image segmentation; Fuses; Decoding; Transformers; Visualization; Attention mechanism; autonomous vehicles; RGB-disparity fusion; road potholes; semantic segmentation}
3. Nisa, S.Q., Ismail, A.R.: Dual U-Net with resnet encoder for segmentation of medical images. Int. J. Adv. Comp. Sci. Appl. (IJACSA) **13**(12) (2022). https://doi.org/10.14569/IJACSA.2022.0131265
4. Fan, J., et al.: Multi-scale feature fusion: learning better semantic segmentation for road pothole detection. 2021 IEEE International Conference on Autonomous Systems (ICAS), pp. 1–5. Montreal, QC, Canada (2021). https://doi.org/10.1109/ICAS49788.2021.9551165. keywords: Image segmentation; Visualization; Autonomous systems; Roads; Conferences; Semantics; Feature extraction; pothole detection; single-modal semantic segmentation; convolutional neural network; feature fusion
5. Fan, R., Wang, H., Wang, Y., Liu, M., Pitas, I.: Graph attention layer evolves semantic segmentation for road pothole detection: a benchmark and algorithms. IEEE Trans. Image Proc. **30**, 8144–8154 (2021). https://doi.org/10.1109/TIP.2021.3112316. keywords: Roads; Image-segmentation; segmentation; Semantics; Convolutionalneuralnetworks; Featureextraction; Computerarchitecture; Benchmarktesting; Graphneuralnetworks; Roadpotholedetection
6. Chen, D., Chen, N., Zhang, X., Guan, Y.: Real-time road pothole mapping based on vibration analy sis in smart city. IEEE J. Selec. Top. Appl. Earth Obs. Remote Sens. **15**, 6972–6984

(2022). https://doi.org/10.1109/JSTARS.2022.3200147. keywords: Roads; Vibrations; Real-timesystems; Surfacetreatment; Smartcities; Prototypes; Vibrationmeasurement; Automate-dinstruments; roadsurfaceobser-vation; signalprocessing; smartcity; Spatio-temporalfusion; vibrationprocessing

7. Forrest, M.M., Chen, Z., Hassan, S., Raymond, I.O., Alinani, K.: Cost effective surface disruption detection system for paved and unpaved roads. IEEE Access **6**, 48634–48644 (2018). https://doi.org/10.1109/ACCESS.2018.2867207. keywords: Roads; Sensors; Maintenance engineering; Acoustics; Accelerometers; Monitoring; Vibrations; Hump; pothole; road surface disruption; ultrasonic.sensors
8. Adnan, M.M., et al.: Automated image annotation with novel features based on deep ResNet50-SLT. IEEE Access **11**, 40258–40277 (2023). https://doi.org/10.1109/ACCESS.2023.3266296. keywords: Feature extraction; Image annotation; Image color analysis; Deep learning; Visualization; Semantics; Image segmentation; Electronic learning; Automatic image annotation; deep learning; features extraction; digital learning; Slantlettransform; technological development
9. Li, K., Tao, W., Liu, L.: Online semantic object segmentation for vision robot collected video. In IEEEAccess **7**, 107602–107615 (2019). https://doi.org/10.1109/ACCESS.2019.2933479. keywords: Streaming media; Object segmentation; Motion segmentation; Image segmentation; Semantics; Proposals; Online video segmentation; object segmentation; object detection; object tracklets; object proposal; shortest path algorithm
10. Wang, Y., Wu, L., Qi, Q., Wang, J.: Local scale-guided hierarchical region merging and further over-and under-segmentation processing for hybrid remote sensing image segmentation. In IEEE Access **10**, 81492–81505 (2022). https://doi.org/10.1109/ACCESS.2022.3194047. keywords: Image segmentation; Merging; Vegetation mapping; Remote sensing; Image edgedetection; Indexes; Shape; Geographic object based image analysis (GEOBIA); hybrid image segmentation; local scale; hierarchical region merging; over and under-segmentation recognition and reprocess
11. Nguyen, T.D., Shinya, A., Harada, T., Thawonmas, R.: Segmentation mask refinement using image transformations. IEEE Access **5**, 26409–26418 (2017). https://doi.org/10.1109/ACCESS.2017.2772269. keywords: Image segmentation; Proposals; Feature extraction; Object segmentation; Semantics; Head; Instance segmentation; object proposal; segmentation mask; convolutional neural networks; deep.learning
12. Jia, H., Yang, W., Wang, L., Li, H.: Uncertainty-guided segmentation network for geospatial object segmentation. IEEE J. Selec. Top. Appl. Earth Obs. Remote Sensing **17**, 5824–5833 (2024). https://doi.org/10.1109/JSTARS.2024.3361693. keywords: segmentation; Decoding; Transformers; Feature segmentation; Convolution; Geospatial object Uncertainty; Image extraction; Object segmentation; remote sensing (RS); semantic segmentation; uncertaintydecoding.mechanism
13. Suri, J.S., et al.: UNet deep learning architecture for segmentation of vascular and non-vascular images: a microscopic look at UNet components buffered with pruning, explainable artificial intel ligence, and Bias. IEEE Access **11**, 595–645 (2023). https://doi.org/10.1109/ACCESS.2022.3232561. keywords: Image segmentation; Artificial extraction; Industries; Decoding; Biomedical intelligence; Feature imaging; Training; Image segmentation; vascular; non-vascular; UNet classes; UNet variations; UNet components; explainable AI; pruning; bias

Exploring Emerging Trends and Market Potential of Barrier Coating Chemicals in Sustainable Paper Packaging

Vanishree Pabalkar(✉), Reena Lenka, Jaya Chitranshi, and Kalpesh Bhave

Symbiosis Institute of Management Studies, Symbiosis International (Deemed University), Pune, India
vanishree.p@sims.edu

Abstract. Barrier coating refers to a type of coating applied to the surface of a material, such as paper, cardboard, or plastic, to create a protective layer that prevents the penetration of liquids, gases, oils, or other substances. The primary purpose of barrier coatings is to enhance the material's resistance to moisture, oxygen, grease, and other environmental factors, thereby improving its functionality and extending its durability. In the context of the paper industry, barrier coatings are often used to make paper and paperboard suitable for packaging applications, particularly for food products, where protection from moisture and grease is essential. These coatings can be made from a variety of materials, including polymers, waxes, biopolymers, and even certain types of natural and sustainable compounds, depending on the desired properties and environmental considerations. Barrier coatings are crucial in the development of sustainable packaging solutions, as they allow paper-based materials to replace plastics and other non-renewable materials in various packaging applications. End Use of barrier chemical coated paper: Pizza Boxes, Pet food Bags/Boxes, Ice cream Frozen food, Fish Trays, Meat Packaging, Paper Cups & Plates, Cakes / Cookies, Wet Vegetables.

Keywords: Emerging trends · Barrier coating · chemicals · sustainable · paper packaging

1 Introduction

Barrier coatings are more are increasingly used in both food and non-food packaging applications. This included poster paper, detergent boards, soap wrap, heat-sealable packaging, pharmaceutical packaging, along with release papers, signifying the versatility and growing importance in the packaging industry (Facts.MR, 2022) These coatings are used for materials that include paper, cardboard, or plastic in order to create a protective layer that prevents the penetration of liquids, gases, oils, or other substances, thereby enhancing resistance to moisture, oxygen, grease, and other environmental factors (Fig. 1)

- Poster Paper, Detergent Boards, Soap Wrap, Heat Sealable Packaging, Pharmaceutical Packaging, Release Papers (Fig. 2)

S. Fong et al. (Eds.): ICT4SD 2025, LNNS 1651, pp. 209–218, 2026.
https://doi.org/10.1007/978-3-032-06688-6_21

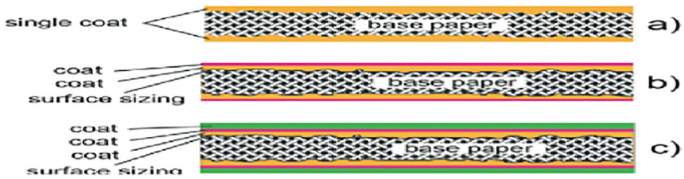


Fig. 1. Non Food Packaging Applications

Fig. 2. Poster Paper, Soap Wrap, Heat Sealable Packaging, Release Papers Objectives

Identify Emerging Trends: To analyze the latest trends in barrier coating technologies within the paper packaging industry, focusing on innovations that enhance sustainability and functionality. **Market Potential Assessment:** To evaluate the current and future market potential of barrier coating chemicals in the paper packaging sector, considering factors like demand growth, market size, and key industry drivers.

1. **Consumer and Industry Demand Analysis:** To understand the demand dynamics from both consumers and industries for sustainable paper packaging solutions that utilize barrier coatings.
   - o **Competitive Landscape Analysis:** To study the competitive landscape of companies producing barrier coating chemicals, identifying key players, their market strategies, and the challenges they face.
   - o **Sustainability Impact Evaluation:** To assess the environmental impact of barrier coating chemicals and how they contribute to sustainable packaging solutions in comparison to traditional alternatives.
   - o **Regulatory and Policy Analysis:** To explore the impact of regulations and policies on the development and adoption of barrier coating chemicals in the paper industry, with an emphasis on sustainability requirements.

2 Methodology

2.1 Market Potential Assessments

Secondary Data collection From Various research Paper & websites. **Data Form Facts. MR.** (https://www.factmr.com/report/barrier-paper-market).

Secondary data of the Barrier Paper Market Outlook (2022–2032) Barrier chemical coated paper is a type of packaging material designed to protect food items or any products from external environmental factors such as water, moisture, grease, Oils, and oxygen. It can be used as an eco-friendly alternative to plastic packaging, which is very important as industries shift toward sustainable solution for packaging industry. Barrier coated paper is widely used in the food and beverage industry, pharmaceuticals, personal care, and other consumer goods industries.

Sample Heading (Third Level).
See Fig. 3.

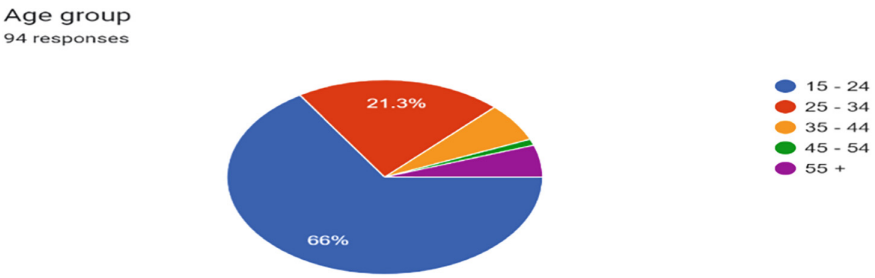


Fig. 3. Age classification

Age Group: Tried to connect various age groups and got response majorly from age group of 15 to 24. It is showing that younger generation is more aware about sustainable paper packaging (Fig. 4).

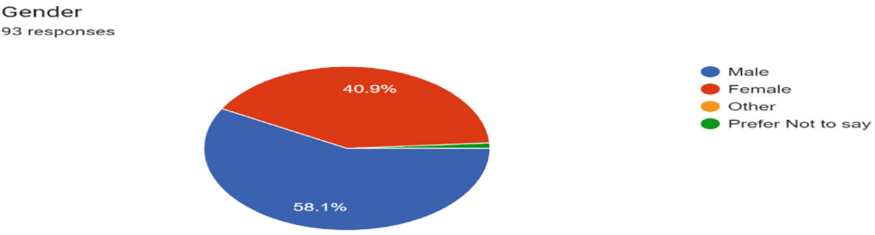


Fig. 4. Age classification

Gender: for this survey gender is not important factor everyone is showing similar interest in growth of eco friendly paper packaging (Fig. 5).

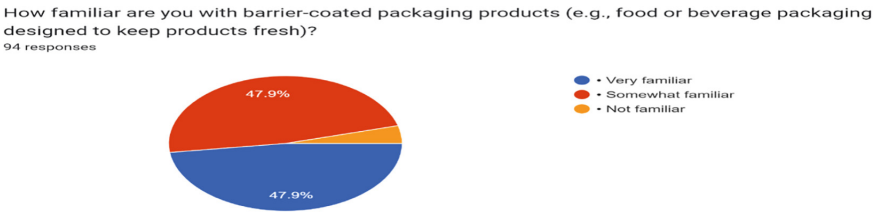


Fig. 5. Gender classification

As per this survey almost 50% target population is very familiar with this concept & remaining are aware about it & very few are unaware about it (Fig. 6).

As per the survey almost 80% population is using barrier coated packaging material for food stuffs (Fig. 7).

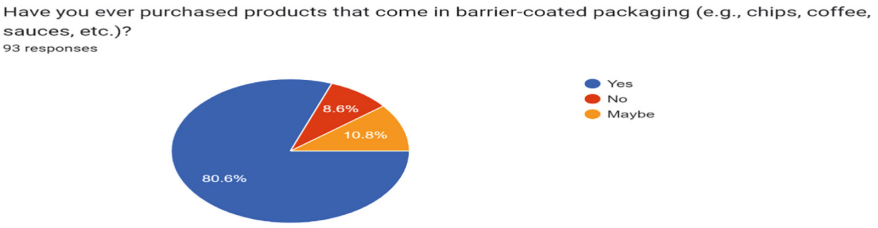


Fig. 6. Purchase classification

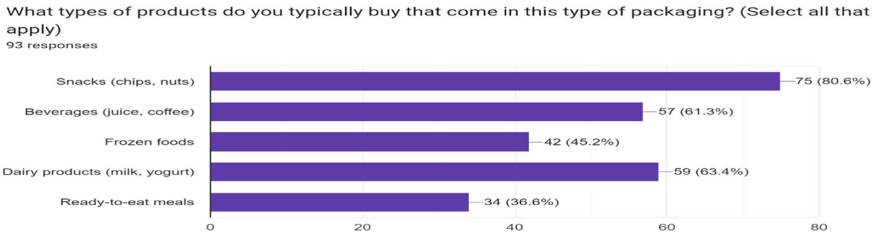


Fig. 7. Type of Packaging

Snacks & Dairy products are mostly available in barrier coated packaging material, which is mostly used by kids so that packaging material must be safe & free from hazardous chemicals (Fig. 8).

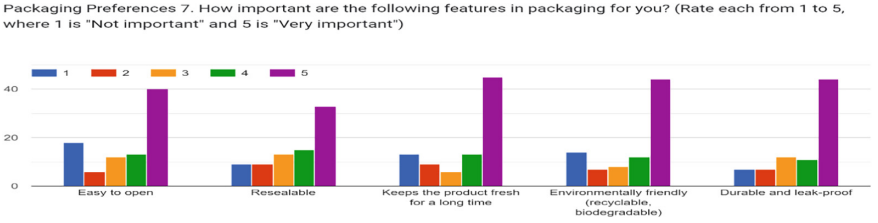


Fig. 8. Packaging Preferences

almost most of the feature of barrier coated packaging is import i.e.

Easy to open, Resealable, Keep the product fresh for long time, Environment friendly (Recyclable & Biodegradable), Durable & Leak proof (Fig. 9).

Buying of the food products are not most important on packaging used, it is dependent on person to person (Fig. 10).

Most of the people who thinks about packing material while buying foods stuffs, avoids material which is not recyclable & that packaging which do not preserve food for long time (Fig. 11).

Almost 66% people are concern about the environmental impact of packaging material (Fig. 12).

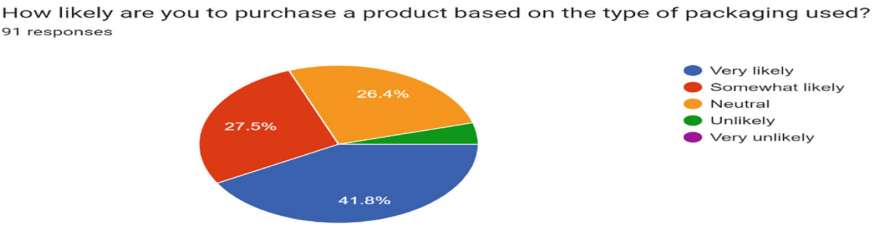


Fig. 9. Type of Packaging

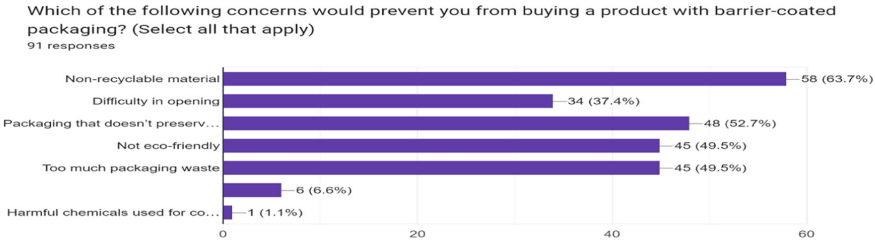


Fig. 10. Food products purchase

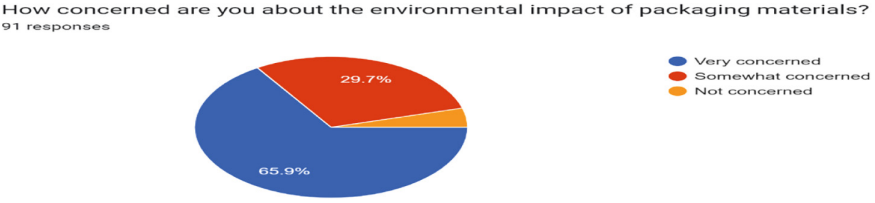


Fig. 11. Environmental impact

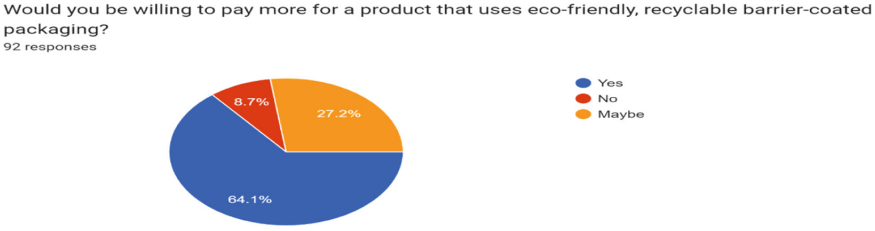


Fig. 12. Eco friendly Packaging

Almost everyone is willing to pay more for the ecofriendly & Recyclable packaging material (Fig. 13).

Clear labelling & Recyclable packaging is very important & that encourage people to buy the products packed in sustainable packaging (Fig. 14).

Leaking & Breaking issues & difficulty in opening these two problems people are facing in packaging of food stuffs (Fig. 15).

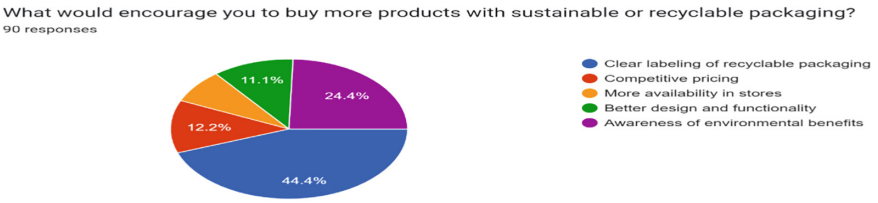


Fig. 13. Sustainable recyclable Packaging

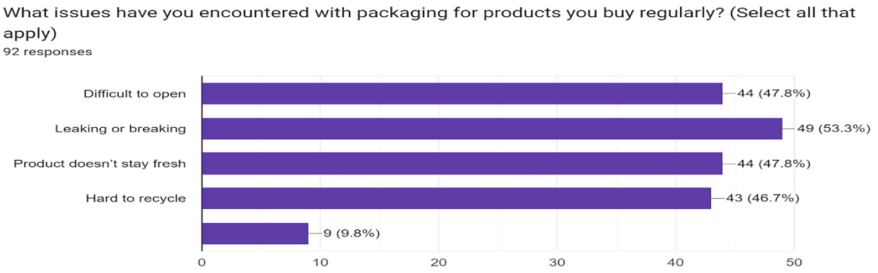


Fig. 14. Issues on Packaging

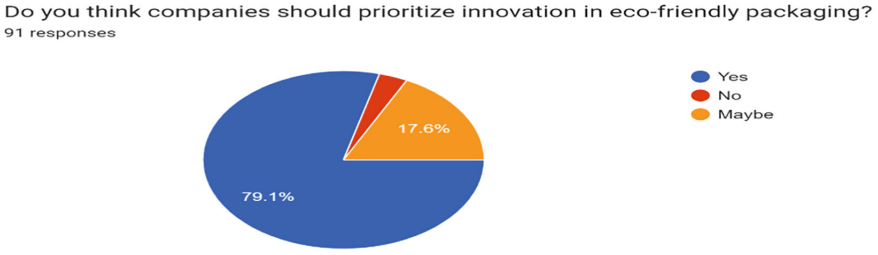


Fig. 15. Innovation in Packaging

According to most of the population companies should think about eco-friendly packaging (Fig. 16).

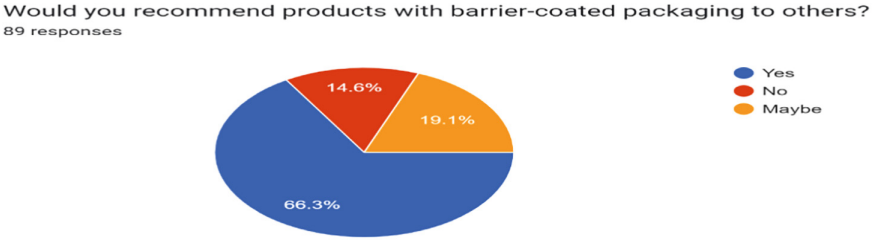


Fig. 16. Barrier coated packaging

Almost 85% people among the target population are willing to buy & recommend the products with barrier coated packaging material.

2.2 Competitive Landscape

Companies Producing barrier coating chemicals. There are various small, Medium & large-scale companies operating in this market globally. Some Major Market Players - Altanga AG group (Germany), BASF SE (Germany), H.B. Fuller company (US), Kuraray Co Ltd (Japan), Sonoco product Co. (US), Michelman (US), Imerys (France) Omya AG (Switzerland), Solenis (US), Dow Inc (US). Some Indian Small & Large Payers: Cubane Speciality Chemicals, Archroma Chemicals, Solenis (India) ltd, Vertex chemicals. In the barrier coating chemicals market, some of the top manufacturers include: **3M, BASF, Mondi, Amcor, PPG Industries, DuPont.** These companies dominate the market with significant shares, and the global market for barrier coatings is estimated to grow steadily. For instance, in 2023, the functional additives and barrier coatings market was valued at around $2.3 billion, with projections to reach $4 billion by 2033 at a growth rate of approximately 5.4% per year. **Here is a breakdown of key players: 3M** and **BASF** are leaders in innovative coatings solutions, offering advanced chemical formulations for industries like packaging, automotive, and electronics. **Mondi** and **Amcor** have a strong presence in the packaging sector, using barrier coatings to enhance the durability and shelf life of products. **PPG Industries** is known for its coatings expertise in various sectors, particularly automotive and aerospace, In terms of geographic distribution, North America is expected to generate around 21% of the market revenue in 2023. Other major regions include Europe and Asia-Pacific, with growing demand across industries like food and beverage, pharmaceuticals, and consumer goods.

2.3 Sustainability Impact Evaluation

Paper instead of plastic – sustainable and recyclable coatings, the use of plastic packaging, film coatings and other environmentally harmful packaging materials are now consigned to yesterday. The grammages of the coated papers vary depending on the application and intended use. A coating can be applied to brown and white paper and standardized corrugated cardboard base papers are common.

1. **Material Composition -** Many traditional barrier coatings are made from plastics and synthetic chemicals, such as polyethylene (PE) and polypropylene (PP). These materials are derived from fossil fuels, contributing to greenhouse gas emissions and the depletion of natural resources.
2. **Recyclability and Biodegradability -** Recyclability & biodegradability is most important factor in packaging industry for the eco-friendly packaging the material should be recyclable & Bio degradable, plastic coated materials are very hard to recycle and not easy for biodegradable. There is a government authorized laboratories those gives certification for biodegradable material which is eco-friendly.

   Material should degrade biologically within 3 to 6 months.

2. **Life Cycle Assessment (LCA) -** A full life cycle assessment (LCA) of barrier coatings measures their overall environmental impact—from raw material extraction through production, use, and disposal. Companies are increasingly using LCAs to evaluate and minimize the ecological footprint of their products. LCA data helps identify hotspots in the production process that could be made more sustainable, such as reducing energy use or switching to renewable raw materials.
3. **Regulatory Pressures -** Governments and environmental regulations are pushing the industry toward greener solutions. Single-use plastics bans and stricter packaging waste regulations in many regions (e.g., Europe and parts of the U.S.) are encouraging companies to adopt sustainable barrier coatings that comply with these new standards

Key FDA Regulations for Barrier Coatings: Indirect Food Additives: Materials used in food packaging must be cleared as indirect food additives. This includes coatings that act as barriers to moisture, gases, or oils. This section covers coatings that come in to contact with food. These coatings must meet strict chemical migration limits to ensure they do not affect food safety. Some coatings may contain specific additives or chemicals that are subject to their own regulatory limits, such as **plasticizers**, **colorants**, or **stabilizers**. Addresses additional regulations for substances that may be used in combination with coatings, including additives that enhance the performance of the coating.

Certification Process: Manufacturers must submit detailed safety data, including: **Migration testing** to ensure that no harmful substances leach from the coating into the food under various conditions like temperature, acidity, and storage time. **Toxicological studies** demonstrating that the chemicals in the coating are safe for food contact. Once the FDA reviews the data and if the substance meets safety criteria, the manufacturer receives clearance for their coating to be used in food-contact applications.

2.4 Regulatory and Policy analysis

The Indian government, through the Food Safety and Standards Authority of India (FSSAI), enforces detailed rules and regulations on food packaging under the Food Safety and Standards (Packaging and Labelling) Regulations, 2011. These guidelines are essential to protect food from contamination, maintain its quality, and ensure consumer safety. Companies will need to provide comprehensive information on the packaging of 19 different products, including milk, tea, biscuits, edible oil, flour, bottled water, baby food, cereals and pulses, cement bags, bread, and detergents, as per the new regulation. For instance, if the packaged product weighs less than the recommended amount, the price per gram or milliliter needs to be noted. Companies will also need to disclose the country in which the product is created as well as the date of manufacture of the items. India's Central Pollution Control Board (CPCB) has laid out guidelines on plastic waste management which affects businesses using plastic packaging material like courier bags or bubble wrap. Our offerings are compliant with CPCB rules, ensuring you stay within legal boundaries while shipping your products nationally. Specifically designed for small-to-medium sized Indian enterprises serving several sectors - from e-commerce platforms to brick-and-mortar stores - our carton boxes, mailer boxes, foam pouches etc., meet all necessary industrial standard specifications. **Trade barrier for Packaging Paper from Govt of India.** India faces several **trade barriers** in exporting

barrier coating chemicals and coated papers, primarily due to regulations, tariffs, and compliance requirements in international markets. Here are the key barriers: **Tariff and Non-Tariff Barriers, Tariffs**: Countries like the U.S. and members of the European Union impose import duties on chemical products and coated papers from India, raising the overall cost for Indian exporters. **Non-Tariff Barriers**: Many developed countries, such as in the **EU**, impose **stringent environmental and safety regulations** on chemical products. Indian barrier coating chemical exporters must comply with these complex standards, such as **REACH (Registration, Evaluation, Authorization, and Restriction of Chemicals)** in Europe, which involves costly registration processes.

Trade Agreements and Quotas: India's access to major markets is sometimes limited by **trade agreements** and **quotas**. For example, the **Generalized System of Preferences (GSP)**, which provides duty-free entry for certain products to the EU and the U.S., has been partially withdrawn for India in some sectors. This affects competitiveness. In summary, Indian manufacturers of barrier coating chemicals and coated papers face a range of challenges, including tariffs, compliance with stringent environmental and safety standards, anti-dumping duties, and trade agreements that may limit their market access

2.4.1 Findings and Conclusion

Based on secondary data collection & industry experts interviews some conclusions are as bellow.

- With growing environmental awareness, the push toward plastic alternatives have led to a surge in demand for barrier-coated paper packaging that is recyclable and compostable. The need for packaging that can protect products during shipping has increased demand for durable yet sustainable solutions like barrier-coated paper. Compostable coatings and high-performance bio-based barriers are anticipated to disrupt the industry. These innovations promise to deliver both functionality and eco-friendliness over the next five years. Eco-friendly packaging is becoming critical for consumers, especially in sectors like food, cosmetics, and e-commerce. Brands are responding by opting for recyclable and compostable materials. Packaging plays a vital role in conveying a brand's sustainability values. Barrier-coated paper packaging, with its eco-friendly look and feel, strengthens brand perception as environmentally responsible. Companies are shifting toward barrier-coated paper packaging that can be recycled or composted in response to growing pressure from regulatory bodies. They are also investing in recyclable and renewable coatings to comply with upcoming mandates. Global market potential for barrier coating chemical is around 2.3 billion USD. **Analysis of Primary data from online survey response collected 95 people of various age group & gender.** Almost 66% people are concern about the environmental impact of packaging material. Base on survey almost 64% people are willing to pay more money for sustainable & eco-friendly paper packaging. According to 89% of target population suggesting that every company should think about eco-friendly packaging. Specially for food & beverages industry.

References

1. https://www.indiafilings.com/learn/food-packaging-regulations/
2. https://www.bis.gov.in/wp-content/uploads/2023/01/Final-handbook-coloured_F_compressed.pdf
3. https://www.futuremarketinsights.com
4. https://www.factmr.com/report/barrier-paper-market
5. Facts.MR: Barrier Paper Market Outlook (2022–2032) (2022). https://www.factmr.com/report/barrier-paper-market

Review on Security Schemes in Modern IoT Integrated Cloud Systems

Atul Kumar[1(✉)], Devendra Kumar[2], and Niranjan Kumar[2]

[1] Chandigarh University, Unnao, Uttarpradesh, India
atulverma16@gmail.com
[2] Ambalika Institute of Management and Technology, Lucknow, India

Abstract. The Internet of Things (IoT) has transformed the digital environment, but its fast expansion raises substantial cybersecurity concerns. IoT devices are naturally vulnerable to a variety of assaults, and the data they manage can be used by malevolent or unauthorized service providers. The introduction of IoT into cloud-based systems creates new security vulnerabilities. Cloud-based IoT solutions provide flexibility and scalability, but they also increase security vulnerabilities. The complicated interconnections between these traditional devices and systems demand strong measures to ensure privacy and integrity. This article tackles important security problems in IoT adoption by strategies to suggest in bridging present gaps and prepare for future difficulties. Its goal is to improve service security systems and device and strengthen IoT ecosystems through proactive approaches.

Keywords: IoT · Cloud · Security · Networking

1 Introduction

With applications ranging from smart homes to AI-enabled intelligent automobile, medical equipment, and agricultural sector breakthroughs, IoT promises a simple and seamless future. IoT-based monitoring gadgets have greatly enhanced the lives in holistic approach for disable and elderly. But the rapid spread of IoT generates serious risks. Traditional IoT systems have limited memory, storage, communication, and processing capabilities, to necessitates reliance on cloud-based third-party services. These systems support data storage, application, and to process hosting, allow IoT to gather, act, and analyze on data more effectively. Despite the benefits, to outsource sensitive data presents new security issues. This paper looks at these issues, to highlight the importance of strong techniques for to deal with both new and present security vulnerabilities in the rapidly rising IoT world. Cloud-based IoT systems include crucial components such as a local server, IoT devices (e.g., smart air conditioners and temperature sensors), the cloud and a gateway, which are all linked to end users. These interrelated parts comprise a complex ecosystem that gathers, manages, processes,

S. Fong et al. (Eds.): ICT4SD 2025, LNNS 1651, pp. 219–229, 2026.
https://doi.org/10.1007/978-3-032-06688-6_22

and stores data from IoT devices. However, to integrat third-party systems introduces a new level of complexity, particularly in terms of security. Poorly protected third-party systems central to data breaks that endanger the ecosystem's truthfulness. Furthermore, the various nature of IoT components creates new risks, as each has different security requirements. To implement security measures is vital given the frequent contact of IoT devices with human users and their link to key infrastructure. Addressing this weakness as well as any upcoming threats is imperative to fortifying the resilience of IoT systems. IoT relies heavily on cloud technology, and that shifts the way we live and work as a result of smarter devices and automated homes. Nevertheless, even as fundamental sectors such as consumer transport and electronics continue to grow, mitigating security risks becomes increasingly vital. Even with the foremost cloud providers best efforts, security problems continue to exist, which is enough to challenge researchers. The constant connection between IoT devices and the cloud makes them susceptible to eavesdropping, as well as packet manipulation. Such issues are highly critical for IoT infrastructure, including power grid systems, as they can result in severe service disruption due to malicious attacks. The reliance of various IoT systems on common cloud infrastructure exacerbates these risks, to emphasize the necessity for strong security measures to assure the durability and defend against possible vulnerabilities of linked IoT ecosystems. The fast growth of IoT systems sometimes puts cost-effectiveness over security, to result in susceptible deficiencies. Manufacturers typically disregard device security, to view the cloud as merely a platform rather than an essential component of the system. Sophisticated communication networks, along with the distributed nature of IoT systems, create complex and unique security challenges that are difficult to address with traditional solutionsb. Cloud infrastructure providers themselves may have vulnerabilities. For example, weak access policies or shared tenancy risks can expose IoT data to unauthorized actors. Jin et al. (2022) [25] introduced P-verifier, a tool to detect and mitigate such risks in cloud-based IoT access control policies, showcasing how access misconfigurations can be weaponized. Despite the ongoing and continuous research in the field, there is still no universally accepted framework in place to properly classify IoT devices in relation to security concerns. This lack of a clear classification framework significantly impedes the development of robust and effective security safeguards. This study thoroughly examines the existing literature and highlights key gaps and security challenges that continue to affect cloud-based IoT systems today. The results underscore the dire necessity for a more strategic, coherent, and focused plan that handles the multifaceted issues concerning the safety of IoT devices and networks.

2 IoT Cloud Ecosystem Model

IoT devices provide a network allowing sensors and devices to connect over the internet, therefore enhancing daily life. By allowing massive installations handling vast volumes of data and support millions of users, modern IoT cloud systems exceed conventional paradigms. Under this architecture, the IoT gateway

acts as a bridge linking devices to the cloud and enabling local communication and edge data processing. Perfect integration in cloud-based IoT systems depends on ensuring interoperability by uniform protocols. Furthermore very important is the IoT hub, which serves as the focal point of interaction between several devices and people inside the larger IoT ecosystem. Underlying the Internet of Things is the concept of "smartness," which lets sensors and devices independently apply the gained knowledge. Each of the networked devices, software and sensors that analyse and compile data in an IoT system has particular security concerns. Three main service models—Infrastructure as a Service (IaaS), like Amazon Web Services; Software as a Service (SaaS), like Google Apps; and Platform as a Service (PaaS), like Microsoft Azure—which meet the various and expanding need of IoT devices—are offered by cloud computing. For complicated IoT systems, these models provide flawless integration, data management, and real-time processing by means of scalable, flexible, and efficient solutions. All of which are needed to manage the significant data analytics and process required by IoT applications, these models provide tremendous scalability, data storage, infrastructure and computational capability.

3 Literature Review

Maintaining privacy and security of IoT devices becomes more crucial as their count increases. Research has, however, exposed inadequacies in to handle security aspects and other issues. Understanding the importance of cloud security and integration issues is critical as IoT systems grow ever dependent on cloud computing for data storage, processing, and management. Expensive, vast networks able to manage enormous volumes of data and support millions of users define modern IoT cloud systems. This magnitude makes it more challenging to protect the IoT devices as well as the generated sensitive data. IoT and cloud technologies working together has drastically changed IoT device manufacturing techniques. Nevertheless, another effect of this change that makes protection efforts more challenging is the increase in probable security issues. Large volumes of sensitive data become appealing targets for attackers, Kumar & Ahmad (2024) [26] said. Shukla et al. (2019) [27] argued that handling these hazards correctly calls for a well-organised plan. By use of type-based vulnerability classification, strategies may be developed, therefore enabling the development of efficient security solutions to lower risk. In 2019 Zhang et al. [1] proposed a method for identifying continuous masquerade threats and attacks. The system alerts the user upon the recognition of a hazard. Voice authentication—a sophisticated speech spoofing detection system combining four main elements to enable strong speaker authentication for voice-based IoT devices–is another exciting method of improving security. Hooda et al. (2022) [2] showed the success of this strategy and suggested a thorough, systems-oriented defence mechanism meant especially to counteract voice-based confusion attacks aiming on platforms like Amazon Alexa. Designed as a browser plugin, SkillFence is a modern solution that detects and blocks assaults confusing voice assistants such as Amazon Alexa. Concurrent with this, experts are considering the best approaches to

address the interoperability issue in Internet of Things devices, particularly on well-known platforms like Samsung SmartThings that raise integration issues. Developed also by Chatterjee et al. (2017) [3] is a lightweight identity-based cryptosystem leveraging a physically unclonable function (PUF) to enable safe message exchange and authentication operations between IoT devices.

Based on the type of attack platform's and components, Nazzal et al. (2022) [4] offer a detailed method to classify SmartThings hazards. They also provide advice on how to meet these difficulties. Now the most important cyber threat, DDoS attacks mostly affect consumer IoT devices. Software-defined networks (SDN) among other new technologies can assist IoT devices be shielded from such dangers. Flow filters are one often used SDN-based method to reduce DDoS attacks. This basic approach evaluates designated components in data headers to prevent unwanted traffic. Researchers suggested several approaches to make cloud-based IoT devices more safe in order to solve security problems. Bhattacharjya et al. (2019) [5], presented a safe IoT architecture for smart cities, which may be used in situations such as the Power Internet of Things and smart homes. Smart home IoT devices provide a particularly high risk to reveal sensitive and private user data. To address this, Gerber et al. (2017), created a case study LOKI, developed by Gerber et al. (2021) [6], is an architecture that ensures privacy in smart homes by processing data locally, thereby limiting the data sent to third-party cloud services. It uses edge computing to execute commands and analyze user inputs directly on the smart home controller. Implementation Strategy: Data from devices like smart bulbs and thermostats is processed on local Raspberry Pi-based servers. Only encrypted summaries or event logs are sent to the cloud for backup or future AI model training. Impact: This approach minimizes data leakage risks even if third-party cloud platforms are compromised. Ling et al. (2022) [7] discovered that data security is flattering as a major problem for IoT devices intended for children. Congestion attacks are a prevalent concern of data jamming, particularly with entertainment drones, in which attackers modify and hijack drone data via illegal communication channels. Rafferty et al. (2017) [8] devised a fresh strategy assigning particular tasks including notifying guardians or parents of any possible violations and guaranteeing the preservation of children's privacy. Likewise, Yankson et al. (2020) [9] presented an enhanced digital forensics framework including fresh approaches to assist investigators in methodically organising and evaluating digital evidence using the "Plan," "Preserve," "Present," and "Process," (4P) methodology. Each of the various creative approaches that can be used to solve the security challenges presented by medical IoT devices and systems adds to a more safe surroundings. Introduced by Fathalizadeh (2022) [10], one interesting strategy concentrates especially on the safeguarding of location privacy in IoT systems. This approach efficiently protects patient location data from illegal access while delivering safe online, location-based healthcare services using advanced anonymising techniques like -diversity, k-anonymity and t-closeness. This approach does, however, have a trade-off: the anonymising process may result in some degree of data obfuscation, therefore compromising the precision of exact location mon-

itoring in some situations and so influencing the quality of services. Although the approach gives user privacy and security first priority, it can restrict the accuracy of location-based services since anonymising methods might somewhat hide precise user locations. To guard implantable medical devices (IMDs) from possible and existing communication-based attacks, Bu et al. (2019) [11] developed a safe protocol known as Bulwark This protocol permits authorised 3rd party medical teams to securely access internal memory of the IMD under disaster conditions. Bulwark is meant to protect against various communication flaws as well as man-in—the-middle (MITM) attacks. Using private blockchain technology, Wazid et al. (2019) [12] suggested safe communication in healthcare systems leveraging drones and the Internet of Things (IoT). Their system uses a safe voting mechanism to preserve and confirm data transmission integrity. This process adds, mines, and validates medical information-containing blocks for a peer-to–peer (P2P) network, all to apply the PBFT consensus algorithm. By means of trivial cryptographic methods such as the ARX encryption system, Srivastava et al. (2019) [13]suggested an enhanced method for safe data transit across storage and network devices. Additionally included into the communication process were ring signatures, which offer significant privacy traits including signer anonymity and signature accuracy. Furthermore developed by the team is GHOSTDAG, a blockchain technology remotely monitoring patient health data via a directed acyclic graph. Hardware attack security solutions aim to protect medical IoT equipment against physical tampering (Shrivastava et al., 2022). Wu et al. proposed the "golden die" approach to let trojans with distinct footprints be more easily identified than the usual arrangement (Wu et al., 2015) [14]. Likewise, Li et al. (2017) [15] devised an SDN edge computing-based approach whereby edge servers validate medical device validity prior to data transfer. This innovation lessens the risks connected with long-distance data transfer by storing critical patient data locally on edge servers. Beyond the healthcare area, cloud-based industrial IoT (IIoT) confronts severe security issues. Industry 4.0, which is defined by sophisticated smart technology, has transformed industrial processes while also increase possible attack surfaces, that include vulnerabilities in equipment such as pressure and temperature sensors. Obermaier et al.(2016) [16] studied four video surveillance and flaws discovered a number of security systems, to include the usage of unsafe fallback functions, weak passwords, a lack of robust security requirements and inadequate authentication mechanisms. Furthermore, newer cloud-based IoT devices struggle to recognize video frames, with accuracy rates drop by below 90 Maritsch et al.(2016) [17] introduced a hardware- transparent and secured architecture for data share in Industrial IoT systems. This architecture enables efficient and safe data exchange across stakeholders included as manufacturers, regulators and operators, while maintain the integrity and security of the shared data. Conferring to Evesti et al. (2016) [18], network edge devices can be used to protect IoT systems for smart industries and homes. A rectification assault is a unique danger in which attackers manipulate important temperature-based control systems. A low-cost anomaly detection prototype was developed by Tu et al. (2016) [19] to guarantee the

accuracy of sensor signals. Using a software-defined networking (SDN) cybersecurity strategy, Stocchero et al. (2022) [20] proposed a secure command and control (C2) system for the IoBT. In the meanwhile, security risks are increasing as IoT devices are integrated with ADAS in cars to improve driving (Shah et al., 2022). Enhancing the driving experience and transmitting vital information, connected vehicles utilise vehicle-to-internet (V2I) and vehicle-to-vehicle (V2V) connections. However, this raises worries about the loss of sensitive data and privacy threats.

Khan et al.(2022) [21] offered a security-focused software development process that included secure code, code reviews, testing of threat, and modeling. To protect vehicle hardware, secure hardware-based encryption and boot procedures approaches are advised. Furthermore, in order to avoid unauthorised access to communication networks, smart automobiles employ network security protocols. Researchers are also looking at ways to make smart cars more secure using AI. To safeguard the Internet of Things (IoT) from cybercriminals, Zewdie et al. (2020) [22] suggest incorporating AI. The study emphasizes many approaches to improve IoT security, to include predictive analysis, response and threat detection, risk assessment and behavior analysis. AI can dramatically improve the settings for security of cloud and IoT devices by analysis of enormous amounts of data and to identify irregularities in real time(Sahay et al. 2012; Sharma, 2013). In addition to ML and AI, new sophisticated technologies are developed to improve security. Blockchain-based architectures provide a viable alternative to improve the security of smart automobiles. Agreeing to Wang and Smys (2021) [23], as automobiles grow more autonomous and networked, they are gradually vulnerable to cyberattacks. They suggest leveraging distributed ledger technology, such as blockchain, to develop a tamper-proof and secure platform for data management in smart automobiles. Fake data injection threats in autonomous and connected cars, which can endanger their safety and performance, are discussed by Zhao et al. (2021) [24]. A planned road map is absolutely necessary to solve IoT security flaws in cloud-based systems. It starts with threat modelling and vulnerability scans helping to classify all IoT assets and outside integrations. Redining the security architecture using zero-trust models, edge processing, and network segmentation comes next. Managing outside dangers by means of secure API controls and supply chain validation is absolutely vital. Blockchain-based audit trails, encryption, and artificial intelligence-driven anomaly detection all help to provide data safety. Eventually, consistent updates, adherence to GDPR and NIST, and stakeholder training guarantee ongoing compliance. This road map advances a strong and safe IoT environment (Table 1).

Table 1. Summary of IoT Security Approaches and Their Trade-offs

Author	Method	Advantage	Limitations
Zhang et al. [1]	Masquerade attack detection for voice authentication in IoT	Real-time threat alerts, robust security for voice-controlled devices	Potential false positives, requires continuous monitoring
Hooda et al. [2]	SkillFence: Browser plugin for voice-based confusion attack prevention	Effective for Amazon Alexa and voice assistants	Limited to voice-based attacks, does not cover other IoT threats
Chatterjee et al. [3]	Lightweight identity-based cryptosystem using PUF	Enhanced authentication and message security	Implementation complexity, requires additional hardware support
Nazzal et al. [4]	SmartThings threat classification and mitigation	Step-by-step attack classification	Limited to SmartThings platform, does not cover broader IoT ecosystem
Bhattacharjya et al. [5]	Secure IoT architecture for smart cities	Improved security for smart homes and power IoT	High infrastructure cost, integration challenges
Gerber et al. [6]	LOKI: Local data processing interface for smart homes	Protects user privacy, prevents hacking	Limited computational power, potential data processing delays
Ling et al. [7]	Security analysis of IoT devices for children	Identifies security risks in children's devices	Limited to specific devices, does not cover all IoT categories
Rafferty et al. [8]	Privacy rule framework for smart toys	Protects children's privacy in IoT devices	May not prevent all unauthorized access
Yankson et al. [9]	4P-based forensic investigation for IoT	Systematic digital evidence management	Requires skilled professionals for implementation
Alishahi et al. [10]	Privacy protection using anonymization techniques	Secures location data in IoT systems	May reduce location tracking accuracy
Bu et al. [11]	Bulwark: Secure protocol for implantable medical devices (IMDs)	Defends against MITM and communication attacks	May introduce latency in emergency access

(*continued*)

Table 1. (*continued*)

Author	Method	Advantage	Limitations
Wazid et al. [12]	Private blockchain for secure IoT communication	Secure data integrity and transfer in medical IoT	Scalability issues, computational overhead
Srivastava et al. [13]	GHOSTDAG blockchain for remote patient monitoring	Enhances security and data integrity in healthcare	High energy consumption, network dependency
Wu et al. [14]	Golden die approach for Trojan detection	Identifies distinct footprints for easy detection	Not foolproof, attackers may bypass detection
Li et al. [15]	SDN edge computing for IoT healthcare security	Improves data privacy and authentication	Edge computing limitations, requires high infrastructure
Obermaier et al. [16]	Security analysis of cloud-based video surveillance	Identifies vulnerabilities in surveillance systems	Weak security implementations, inaccurate recognition
Maritsch et al. [17]	Secure data-sharing architecture for Industrial IoT	Ensures integrity in IIoT data exchange	Limited to specific industrial applications
Evesti et al. [18]	Edge devices for smart industries and homes security	Mitigates IoT security threats at the network edge	Edge device constraints, limited computational power
Tu et al. [19]	Low-cost anomaly detection for sensor signals	Detects manipulations in temperature-based control systems	Limited scope, may not detect all attack types
Stocchero et al. [20]	SDN-based cybersecurity for battlefield IoT	Improves IoT security in military applications	Implementation challenges in high-risk environments
Khan et al. [21]	Security-focused software development for vehicles	Enhances automotive security through encryption and secure boot	May introduce processing delays in vehicles
Zewdie et al. [22]	AI for IoT security	Predictive analysis for real-time cyber threat detection	Requires extensive training data, potential false positives
Wang and Smys [23]	Blockchain for smart vehicle security	Tamper-proof data management in autonomous cars	Blockchain scalability and latency issues
Zhao et al. [24]	Cloud-based sandboxing for connected vehicles	Detects fake data injection attacks in smart cars	High computational requirements, network latency

4 Conclusion

This article examines the security of cloud-based IoT systems to use devices of different categories. It identifies security vulnerabilities within each category, investigates highlights opportunities and potential remedies, for further research. While many Internet of Things (IoT) devices depend on cloud computing, the paper points out that certain of these devices pose unique security concerns that only grow as their cloud connectivity grows. This research looks at the risks of distributed denial of service attacks and data breaches in the cloud. Improving IoT security through the application of solutions across all categories is the topic of this study. It implies that by regularly change of device passwords, firmware, and limited sharing, users may address privacy and security problems. This joint effort to embrace best practices will contribute to a more secure and safer digital environment for everybody.

Acknowledgements. Our heartfelt thanks go to our colleagues and mentors, particularly [Dr. Vinodani Katiyar,Professor DSMNRU Lucknow], for their valuable insights and constructive feedback throughout the research process. Finally, we appreciate the encouragement and understanding of our families and friends, whose support has been invaluable in completing this work. The authors declare that no funding was received from any organization, institution, or individual for conducting this research.

References

1. Zhang, N., Mi, X., Feng, X., Wang, X., Tian, Y., Qian, F.: Dangerous skills: understanding and mitigating security risks of voice-controlled third-party functions on virtual personal assistant systems. In: 2019 IEEE Symposium on Security and Privacy (SP), pp. 1381–1396. IEEE (2019)
2. Hooda, A., Wallace, M., Jhunjhunwalla, K., Fernandes, E., Fawaz, K.: Skillfence: a systems approach to practically mitigating voicebased confusion attacks. Proc. ACM Interact. Mobile Wearable Ubiq. Technol. **6**, 1–26 (2022)
3. Chatterjee, U., Chakraborty, R.S., Mukhopadhyay, D.: A pufbased secure communication protocol for IoT. ACM Trans. Embed. Comput. Syst. (TECS) **16**, 1–25 (2017)
4. Nazzal, B., Zaid, A.A., Alalfi, M.H., Valani, A.: Vulnerability classification of consumer-based iot software. In: Proceedings of the 4th International Workshop on Software Engineering Research and Practice for the IoT, pp. 17–24 (2022)
5. Bhattacharjya, A., Zhong, X., Wang, J., Li, X.: Secure iot structural design for smart homes. In: Smart Cities Cybersecurity and Privacy, pp. 187–201. Elsevier (2019)
6. Gerber, P., Heidinger, M., Stiegelmayer, J., Gerber, N.: Loki: development of an interface for task-based, privacy-friendly smart home control through local information processing: Loki: Entwicklung eines interfaces für die aufgaben-basierte, privatsphärefreundliche smart home-steuerung durch lokale informationsverarbeitung. Proc. Mensch und Comput. **2021**, 578–581 (2021)
7. Ling, L., Yelland, N., Hatzigianni, M., Dickson-Deane, C.: The use of internet of things devices in early childhood education: a systematic review. In: Education and Information Technologies, pp. 1–20 (2022)

8. Rafferty, L., et al.: Towards a privacy rule conceptual model for smart toys. In: Computing in Smart Toys, pp. 85–102 (2017)
9. Yankson, B., Iqbal, F., Hung, P.C.: 4p-based forensics investigation framework for smart connected toys. In: Proceedings of the 15th International Conference on Availability, Reliability and Security, pp. 1–9 (2020)
10. Fathalizadeh, A., Moghtadaiee, V., Alishahi, M.: On the privacy protection of indoor location dataset using anonymization. Comput. Secur. **117**, 102665 (2022)
11. Bu, L., Karpovsky, M.G., Kinsy, M.A.: Bulwark: securing implantable medical devices communication channels. Comput. Secur. **86**, 498–511 (2019)
12. Wazid, M., Bera, B., Mitra, A., Das, A.K., Ali, R.: Private blockchain-envisioned security framework for ai-enabled iot-based drone-aided healthcare services. In: Proceedings of the 2nd ACM MobiCom Workshop on Drone Assisted Wireless Communications for 5G and Beyond, pp. 37–42 (2020)
13. Srivastava, G., Crichigno, J., Dhar, S.: A light and secure healthcare blockchain for iot medical devices. In: 2019 IEEE Canadian Conference of Electrical and Computer Engineering (CCECE), pp. 1–5. IEEE (2019a)
14. Wu, T.F., Ganesan, K., Hu, Y.A., Wong, H.S.P., Wong, S., Mitra, S.: Tpad: hardware trojan prevention and detection for trusted integrated circuits. IEEE Trans. Comput. Aided Des. Integr. Circuits Syst. **35**, 521–534 (2015)
15. Li, J., et al.: A secured framework for sdn-based edge computing in iot-enabled healthcare system. IEEE Access **8**, 135479–135490 (2020)
16. Obermaier, J., Hutle, M.: Analyzing the security and privacy of cloud-based video surveillance systems. In: Proceedings of the 2nd ACM International Workshop on IoT Privacy, Trust, and Security, pp. 22–28 (2016)
17. Lesjak, C., Bock, H., Hein, D., Maritsch, M.: Hardware-secured and transparent multi-stakeholder data exchange for industrial iot. In: 2016 IEEE 14th International Conference on Industrial Informatics (INDIN), pp. 706–713. IEEE (2016)
18. Kuusijärvi, J., Savola, R., Savolainen, P., Evesti, A.: Mitigating iot security threats with a trusted network element. In: 2016 11th International Conference for Internet Technology and Secured Transactions (ICITST), pp. 260–265. IEEE (2016)
19. Tu, Y., Rampazzi, S., Hao, B., Rodriguez, A., Fu, K., Hei, X.: Trick or heat? Manipulating critical temperature-based control systems using rectification attacks. In: Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security, pp. 2301–2315 (2019)
20. Stocchero, J.M., Silva, C.A., de Souza Silva, L., Lawisch, M.A., dos Anjos, J.C.S., de Freitas, E.P.: Secure command and control for internet of battle things using novel network paradigms. IEEE Commun. Maga. (2022)
21. Khan, S.K., Shiwakoti, N., Stasinopoulos, P., Chen, Y.: Cyberattacks in the next-generation cars, mitigation techniques, anticipated readiness and future directions. Accid. Anal. Prevent. **148**, 105837 (2020)
22. Zewdie, T.G., Girma, A.: Iot security and the role of AI/ML to combat emerging cyber threats in cloud computing environment. Issues Inf. Syst. **21** (2020)
23. Smys, S., Wang, H.: Security enhancement in smart vehicle using blockchain-based architectural framework. J. Artif. Intell. **3**, 90–100 (2021)
24. Zhao, C., Gill, J.S., Pisu, P., Comert, G.: Detection of false data injection attack in connected and automated vehicles via cloud-based sandboxing. IEEE Trans. Intell. Transp. Syst. **23**, 9078–9088 (2021)
25. Jin, Z., Xing, L., Fang, Y., Jia, Y., Yuan, B., Liu, Q.: P-verifier: understanding and mitigating security risks in cloud-based iot access policies. In: Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security, pp. 1647–1661 (2022a)

26. Kumar, D., Ahamad, F.: Application of machine learning algorithm for optimal model design for opinion extraction. Telematique **23**(01), 215–227 (2024)
27. Shukla, U., Verma, A.K., Verma, S.: Bench automation computer using raspberry pi. In: Satapathy, S.C., Joshi, A. (eds.) Information and Communication Technology for Intelligent Systems. SIST, vol. 107, pp. 489–497. Springer, Singapore (2019). https://doi.org/10.1007/978-981-13-1747-7_47

Customer Retention Prediction

Vaishali Langote, Siddhesh Kulkarni(✉), Aaditya Ghorpade, Aditya Songirkar, and Aditya Chincholkar

Department of Computer Science and Engineering, DVK's MIT WPU, Pune, Maharashtra, India
siddheskul@gmail.com

Abstract. Identifying customer retention is essential for decreasing lost revenues as well as maintaining an established base of loyal customers. By reviewing historical data that includes customer demographics, purchasing habits and behaviours, businesses will be able to determine which customers are going to discontinue using their services or products. In generating models that can identify customers at risk, this process includes machine learning models such as decision trees, logistic regression and neural networks. It is important that predictive retention can work provided the right algorithms are selected, and reliable data is sourced. Continual updates and improved models will enhance accuracy, giving firms the opportunity to keep up with changes in how consumers behave. The models will also give businesses the ability to produce more targeted retention marketing plans since they will not only identify at-risk customers but also give clear data on what they are doing to create customer churn.

Keywords: Customer Retention · Churn Prediction · Machine Learning Models · Consumer Behavior · Data-Driven Insights

1 Introduction

Customer attrition, also known as customer churn, refers to when a individuals or companies stop utilizing a service and go to a competitor [1]. In many industries, retaining existing customers is much less expensive than acquiring new customers, which is why there is such focus on reducing churn [3]. Loss of customers means loss of revenue, especially as companies spend so much on marketing to acquire new customers, and so are looking for long-lasting value from the new customers [2]. The longer a customer is connected with an organization, the higher the company revenue is because of the customer. Therefore, churn reduction is imperative to a company's success, and this is a crucial part of Customer Relationship Management (CRM) [1]. As the demand for effective churn analysis and prediction has to viable research and applications in business, churn analysis and prediction have become a booming activity in both research and business settings. Predicting customer churn enables businesses to take immediate action toward improving retention strategies while optimizing customer relationship management. Beyond merely identifying customers at risk of leaving, it is also essential to analyse underlying patterns and external factors contributing to larger churn trends. Businesses must focus on identifying and predicting [3]:

S. Fong et al. (Eds.): ICT4SD 2025, LNNS 1651, pp. 230–240, 2026.
https://doi.org/10.1007/978-3-032-06688-6_23

1. Whether a customer is likely to discontinue service.
2. The timeframe in which they might leave.
3. The specific location or customer segment affected.
4. The estimated number of customers expected to churn.
5. The primary reasons behind customer attrition.
6. The potential financial impact and risks associated with churn.
7. Effective strategies to reduce churn and retain customers.

In the beginning, research in churn prediction primarily looked at how accurately machine learning models could predict the likelihood of customers departing. Afterward, researchers began to compare different models of machine learning to see which model provided the most prediction accuracy. In churn prediction, a typical approach would include: Naïve Bayes, Logistic Regression, Decision Tree, Support Vector Machines and Linear Regression. For example, Huang Bingquan (2011) showed that Naïve Bayes was superior to earn marketers' predictive success when the high-dimensional data of customers was reduced first. More recently, churn prediction studies have included financial institutions, insurance companies, and online streaming sites. Although efforts have focused on improving the accuracy of prediction, fewer studies have explored the understandability of machine learning models in churn prediction, or the importance of individual features. Understanding why a prediction occurs is valuable for companies charting, and ultimately responding to, customer characteristics that should be tracked for future risk of churn.

Through predictive models, companies can generate insights from significant amounts of customer data and make targeted retention interventions. In practice, both traditional statistical methods and current hybrid classifier methods are used in global churn prediction models. For instance, Y. Hang et al. used statistical models and machine learning approaches to find relationships between demographic characteristics and likelihood of churn. Similarly, Miguéis et al. developed a predicting model that used logistic regression, and analysed retail transaction frequency in order to establish attrition prediction models. However, neither traditional artificial intelligence models nor parametric models alone can lead to high-precision predictions. Therefore, the use of a combination of predictive models is a growing trend for establishing churn forecasting accuracy and dealing with customer retention issues (Fig. 1).

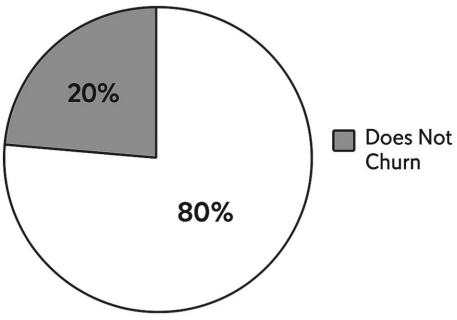


Fig. 1. PIE CHART

2 Literature Review

Review of Customer Churn, Customer Relationship Management, Risk, and Churn Prediction Models.

A. Customer Churn

Customer churn has a substantial impact on revenue for the telecom industry according to A. Gaur and R. Dubey [6]. When customers discontinue their subscriptions or services, it creates added costs for the business. It has been evidenced that acquiring a new customer is almost six times more expensive than retaining an existing customer [7]. Companies take predictive analysis that highlights consumer behaviours that are intended to reduce churn, or highlights new service features or service problems, for example. Customer churn is the discontinuation of use of the service or product, for whatever reason [8]. Some reasons for churn may include: rising costs, unsatisfactory service and support, failure to understand use of service plans, membership fees, poor quality of service, etc

For instance, delays in technical support or unclear communication about pricing adjustments can lead to dissatisfaction. Another major cause of churn is price sensitivity—when service fees increase without noticeable enhancements in quality, customers may seek alternatives. Additionally, vague service agreements and undisclosed fees can result in frustration, prompting users to cancel their subscriptions.

Churn has significant financial consequences, so many telecommunications providers are increasingly using predictive analytics and data-driven methodologies to prevent and reduce customer churn. Predictive modelling is based on how companies can measure potential customer churn using existing consumer history, behavioural tendencies, and characteristics of behaviours that can indicate churn, such as complaints, engagement history, and use patterns, against customers with the highest likelihood of churning. Companies can proactively intervene by targeting customers with promotional offers, loyalty incentives, or service upgrades if they can identify customers with high churn risk early enough.

Machine learning (ML) and artificial intelligence are now a prerequisite to improve the predictive accuracy and reliability of churn prediction models. The most common models are decision trees, logistic regression and neural networks, which carve out segments of customers by their churning probability. In this model, ML algorithms find complex associations between more than one variable, such as combinations of custom variables like usage spikes, billing problems, or service outages, that may not be detected by traditional statistics. Additionally, sentiment analysis of customer reviews, social media buzz, and survey feedback provide additional indication of customer satisfaction and/or the potential for churn.

2.1 Current System

Traditional machine learning techniques are frequently employed within the telecommunications sector capture customer churn. The techniques include algorithms such as decision trees, logistic regression, gradient boosting, random forests, support vector machines (SVM), etc. Prior to conducting machine learning models on a proprietor's customer data, multiple data processing tasks must be completed such as data cleaning,

feature engineering, and converting categorical objects to numeric objects. The machine learning models are trained on historic datasets containing churn indicators, user activity, and customer profile details.

2.2 Limitations

1. Class Imbalance: Commonly, there is a major imbalanced class structure within Telecom churn datasets whereby there are simply more non-churn customers than churn customers. Consequently, typical machine learning models tend to overfit on the larger class and their accuracy in predicting churn cases from the minority class can diminish.

2. Limited Explainability: Although models such as random forests and gradient boosting, provide some interpretability; other models, such as logistic regression do not provide interpretability. If models provide little insight into the eventual factors causing customer churn, stakeholders may be reluctant to understand or trust the predictions of the churn outcomes (Fig. 2).

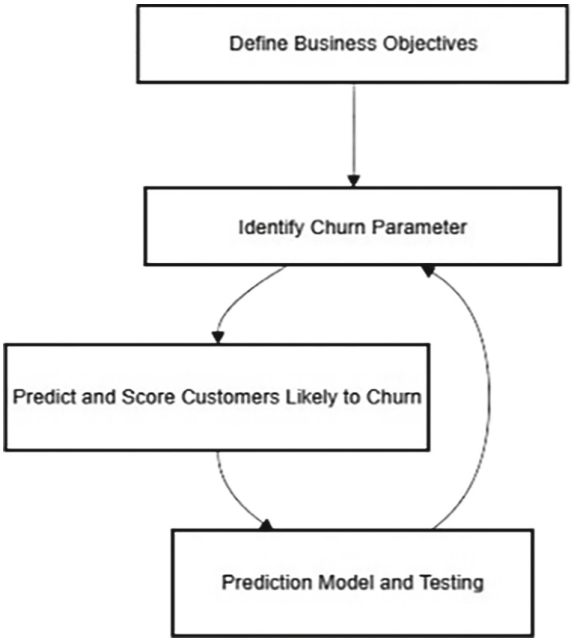


Fig. 2. Data flow diagram

The diagram [2] represents a structured method to reduce customer churn, presumably starting with identifying business objectives, ultimately to formulate churn reduction strategies. The second process involves identifying churn indications, while focusing on those aspects associated with customer attrition, or in other words, which can include customer dissatisfaction or changes in consumption. A predictive model is built and validated to measure possible churn per customer, followed by constructing a customer

distribution of risk, producing a score based on risk of leaving. Finally, a risk category will be assigned - low, moderate, or high - allowing the manager to focus efforts on the most critical, high-risk customers.

3 Churn Prediction Methodology

This study applies a combination of deep learning and machine learning methods to predict customer churn, creating classification models like Random Forest, SVM, Logistic Regression, Gradient Boosting, KNN, and MLP. To overcome the drawbacks of using single algorithms, a new approach called "Hybrid Churn Prediction (HCP)" is proposed, which combines the advantages of SVM, Random Forest, Gradient Boosting, and Logistic Regression. Furthermore, the performance of these models is assessed using two distinct datasets (Fig. 3).

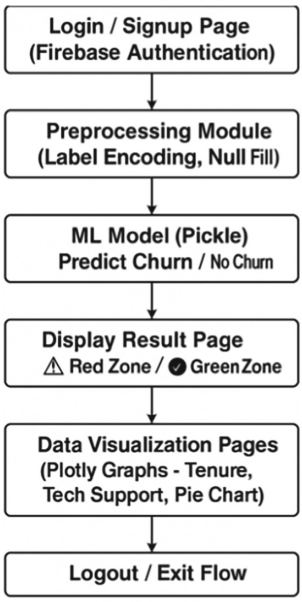


Fig. 3. Workflow diagram

3.1 Data Collection

A company's customer database contains a lot of information, including:

- Customer ID
- Age
- Gender
- Usage frequency

- Time as a member
- Times the customer contacted support
- Delinquent payments
- Subscription type
- Contract cycle
- Customer spends

The variables in customer databases can provide a great deal of information about customer behaviours. However, not all the variables are equally important to predicting churn. The first step should be to establish the most important attributes that influence customer retention, which can help focus the model on the most important considerations.

3.2 Data Cleaning

Preparing the dataset for churn prediction is critical to ensuring consistency and accuracy. The process includes:

1. Handling Missing Data: Noticing gaps in the dataset and then dealing with the missing values, either through imputation (means, medians for numerical variables, modes, for categorical data), or by removing records with too many missing values in cases when imputation is not possible.

2. Eliminating Duplicates: Detecting and removing repeated entries to prevent skewed model training and misleading insights.

3. Managing Outliers: Identifying abnormalities in key variables (e.g., spending behaviour, how often these were used) affected by data entry inaccuracies or in isolated instances. Adjustments are made by capping extreme values, converting them, or removing them in certain situations.

4. Encoding Categorical Data: Categorical variables like subscription plans and payment delays are standardized using techniques such as one-hot encoding or label encoding to make them suitable for machine learning algorithms.

5. Feature Scaling: Normalizing or standardizing numerical data (e.g., age, transaction frequency) to maintain consistency, particularly for models sensitive to feature magnitude, such as logistic regression and KNN.

3.3 Feature Selection

1. Correlation Analysis: Evaluates feature relationships with churn and eliminates redundant variables with high correlation.

2. Statistical Testing: Chi-square tests (for categorical variables) and ANOVA (for numerical variables) identify significant predictors of churn.

3. Tree-Based Feature Importance: Algorithms like Random Forest rank variables based on their contribution to churn prediction.

4. Recursive Feature Elimination (RFE): Iteratively removes less important features to refine the dataset for improved model performance.

4 Proposed System

The designed customer churn prediction system adopts an ensemble approach, combining several machine learning techniques to boost prediction accuracy. The system is structured into the following key modules:

1. Data Preprocessing Module: This stage involves tasks like cleaning the data, normalization, and feature construction. It also includes converting categorical variables using one-hot encoding, handling missing data, and scaling numerical values to prepare the dataset for training.

2. Algorithm Layer: Multiple machine learning models—such as logistic regression, decision trees, support vector machines (SVM), and gradient boosting (e.g., XGBoost)—are trained independently on the processed data. Each model captures unique patterns and insights.

3. Model Evaluation: The system evaluates model performance using standard metrics including accuracy, precision, recall, F1-score, and ROC-AUC. Cross-validation is applied to ensure the reliability and generalizability of results.

4. Prediction and Interpretation Module: To improve understanding and trust in the predictions, the system uses interpretability tools like SHAP (SHapley Additive exPlanations), helping stakeholders grasp which features influence each prediction to understand which factors contribute to a customer's likelihood of churning, aiding in decision-making.

5. Deployment and Monitoring: The trained ensemble model is deployed on a scalable infrastructure that supports both batch and real-time processing of customer data. A continuous monitoring system is implemented to track model performance and trigger retraining as needed to adapt to evolving customer behaviour trends.

4.1 Algorithms

1. Logistic Regression

Logistic Regression is simple and reliable approach that is widely used for binary classification problems such as predicting customer churn. To establish probabilities signaling how likely an instance is within a class or not, the logistic (sigmoid) function is applied to the weighted sum of the input features to be utilized in the model.

$$P(y = 1|X) = 1/1 + e^{-(\beta_0+\beta_1 x_1+\beta_2 x_2+...+\beta_0 x_2)}$$

- β = Coefficients
- x = Features
- Output = Probability of churn (1)

2. Decision Trees

Decision Trees work by making recursive splits of the dataset based on feature values until a tree structure is created. In this structure, each tree body node represents an internal node and is a decision rule for a specific feature; the branches show the results from the decision; and the leaf nodes are the final class predictions. Decision trees are very interpretable and work well if visualized. They can easily over fit the data if not properly pruned.

Gini Index:

$$Gini(D) = 1 - \sum_{i=1}^{c} p_i^2$$

Information Gain:

$$Entropy(D) = \sum_{i=1}^{c} p_i log_2 pi$$

3. Random Forest

Random Forest is an ensemble method that builds many decision trees from random and different parts of the data set and features. For a classification task, the final prediction is based on majority voting, and for regression the mean returns the overall prediction of forest. This method helps improve model generalization and limits variable overfitting, performing better than a single decision tree.

$$y = MajorityVote(T_1(x), T_2(x), ..., T_k(x))$$

- Where Ti(x) = prediction from i-th tree.

4. Gradient Boosting Machines (e.g., XGBoost, LightGBM)

We can think of Gradient Boosting as a collection of weak learners, generally decision trees, in which each weak learner is aimed at correcting the errors made by previous learners. And while there are higher-end implementations like XGBoost and LightGBM that optimize various ingredients of Gradient Boosting (parallel computation, additional regularization, better use of missing data) and, thus, provide better overall performance, they are typically very effective for predicting churn.

$$Fm(x) = F_{m-1}(x) + \gamma^{h_m}(x)$$

- Fm(x) = final prediction after m iterations
- Hm(x) = new weak learner

5. Neural Networks

Neural Networks comprise many layers in which the neurons are interconnected. The neuron works by receiving input data through weighted connections. The neurons then pass this data through a mathematical function, called an activation function, to generate the output. Deep Neural Networks (DNNs), which are structure similar to a neural network but can have many hidden layers, can take complex, non-linear relationships in the data, but may require a major amount of computing to train and use (Table 1).

$$y = f(\sum_{i=1}^{n} wixi + b)$$

f = activation function
wi = weights
xi = input features
b = bias.

Table 1. Algorithm Accuracy

Model type	Accuracy (%)
Decision Tree Classifier	93.4
Neural Network Model	96.4
Support Vector Machines	94.2
Random Forst Classifier	95.7
Gradient Boosting	97.3

4.2 Architecture

The architecture of the customer churn prediction system includes a crude pipeline that summarizes the data engineering and machine learning approach. The data is pulled from various sources **(e.g., CRM, Billing, Support, etc.)** and goes through a preprocessing procedure, which includes processes like data cleaning, normalization, feature encoding, etc. Examples of feature selection are done with mutual information or tree-based importance measures. Dimensionality reduction (e.g., PCA) could be used, but it is not required. To train the models, libraries such **as Scikit-learn, TensorFlow, or PyTorch** are used, and whichever algorithm, from logistic regression to complicated ensembles **(e.g., XGBoost)** or deep learning e.g., neural networks can be employed. K-fold cross-validation is used to improve the models, and hyperparameter tuning can be done with Grid Search or Random Search. Lastly, the deployed trained model can scale in the cloud **(e.g., AWS, Azure)** or could be deployed using a containerized API **(e.g., Flask/FastAPI with docker)**. Real-time monitoring can be done with Grafana and Prometheus to detect drift (e.g., data gates) and monitor performance (Fig. 4).

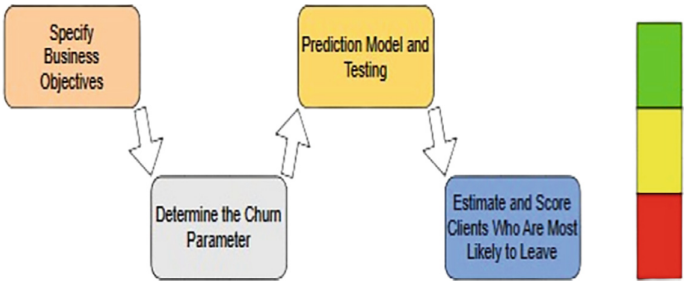


Fig. 4. ARCHITECTURE DIAGRAM

The diagram [3] illustrates the process of customer churn prediction, beginning with defining business objectives to align the project with strategic goals such as improving customer retention. Next, relevant data points are selected, and churn parameters are identified. A predictive model is then built and tested, utilizing historical data for training and performance metrics for evaluation. Once validated, the model is deployed to

assess and predict the likelihood of existing customers leaving, enabling targeted retention strategies. On the right side, a **risk level indicator** visually represents potential risks, shifting from **low (green) during goal-setting** to **critical (red) during deployment**, emphasizing key areas that require additional focus and intervention for successful implementation.

5 Result (Model Evaluation)

The performance of the customer churn prediction model was evaluated using several key metrics, including accuracy, precision, recall, and F1-score. Among the models tested, the Random Forest classifier achieved the highest performance with an accuracy of 85.4%, precision of 82.1%, recall of 78.3%, and an F1-score of 80.1%. These metrics indicate that the model is not only accurate but also balanced in identifying both churned and non-churned customers. A confusion matrix was also used to analyze the prediction distribution, highlighting that the model correctly identified a significant majority of churn instances while maintaining a low false-positive rate. The model's robustness was further validated using 5-fold cross-validation, which confirmed consistent performance across different data subsets. This suggests the model's suitability for deployment in real-world customer retention systems, where early detection of potential churn is critical to business success (Table 2).

Table 2. Model Evaluation

Model	Accuracy	Precision	Recall	F1 Score
Logistic Regression	87.50%	81.20%	78.40%	79.80%
Decision Tree	89.30%	84.70%	82.10%	83.39%
Random Forest	91.70%	88.40%	86.90%	87.64%
Gradient Boosting	93.10%	90.30%	88.70%	89.49%
XGBoost	93.80%	91.10%	89.50%	90.29%
AdaBoost	89.90%	85.60%	83.30%	84.43%
SVM	88.20%	83.10%	80.50%	81.78%

6 Future Scope

Future research can explore the integration of additional customer behavior data such as browsing patterns, support ticket history, and in-app activity to enhance prediction accuracy. Incorporating real-time data streams would allow for dynamic churn detection, enabling businesses to respond proactively. Moreover, implementing explainable AI techniques like SHAP or LIME could help stakeholders understand why a customer is likely to churn, making the model more transparent and actionable. These improvements can make the system more adaptable and valuable in real-world customer retention strategies.

References

1. Aggarval, P., Vijaykumar, V.: Customer Churn Prediction in The Telecom Sector
2. Liwen Ou-College of Arts and Science, New York University ”Customer Churn Prediction Based on InterpretableMachine Learning Algorithms in Telecom Industry”
3. Tsai, T.-Y., Lin, C.-T., Prasad, M.: An Intelligent Customer Churn Prediction and Response Framework.
4. Pulkundwar, P., Rudani, K., Rane, O., Shah, C., Virnodkar, S.: A Comparison of Machine Learning Algorithms for Customer Churn Prediction
5. Hu, X., Hu, X., Hu, X.: Research on a Customer Churn Combination Prediction Model Based on Decision Tree and Neural Network
6. Putra, I.S., Sitompul, O.S., Nababan, E.B.: Customer Churn Prediction using Confident Learning and XGBoost
7. Kripalani, N.: Predictive Analytics for Customer Retention: A Data-Driven Framework for Proactive Engagement and Satisfaction Management
8. Maan, J., Maan, H.: “Customer Churn Prediction Model using Explainable Machine learning” Customer churning analysis using machine learning algorithms
9. Author links open overlay panelB. Prabhadevi, R. Shalini, B.R. Kavitha

Sleep Quality and Body Strain Assessment through 3D Pressure Mapping Using Deep Learning

Deepesh Sudhan Arunachalam(✉), Dennis Andrew, K. S. Gayathri, A. Shahina, V. Durgadevi, and A. Saravanan

Department of Information Technology, Sri Sivasubramaniya Nadar College of Engineering, Chennai 603110, Tamil Nadu, India
{deepesh2110160,dennis2110635,gayathriks,shahinaa, durgadeviv,saravanana}@ssn.edu.in

Abstract. This work introduces a deep learning-based framework for 3D pressure mapping to assess sleep quality and body strain. 2D pressure maps suffer from loss of depth information, poor spatial context, posture misclassification errors, and limited accuracy in capturing regional pressure variations. To overcome these limitations, the framework constructs 3D pressure maps that enable precise region-wise pressure estimation with anatomical landmarks to analyze body strain. Sleep quality is monitored by tracking frequent posture changes with converting pressure maps to point clouds achieved 99.26% accuracy with PointNet and 99.49% with PointCNN.

Keywords: 3D posture classification · Pressure map analysis · Region-wise pressure estimation · body strain · sleep quality · pressure mats

1 Introduction

Proper sleeping posture is important in ensuring overall health and well-being. Sleep posture has a direct influence on muscle, ligament, and joint alignment, decreasing stress and averting long-term complications like musculoskeletal pain, spinal misalignment, and pressure ulcers. Exacerbation of conditions like sleep apnea and respiratory disorders is possible with poor sleeping posture, so posture surveillance is an important element of managing health. Pressure ulcers, which afflict around 2.5 million patients in the United States every year alone [1], are still a healthcare cause for concern. While body repositioning is widely applied as a prevention strategy [2], issues continue to be raised about whether such positional shift actually redistributes applied pressure [3].

Several methods have been developed to monitor sleep posture, including wearable sensors, infrared cameras, and pressure mats. Wearable sensors like accelerometers and gyroscopes offer real-time posture monitoring but can interfere with natural sleep patterns because of discomfort, particularly in older or

S. Fong et al. (Eds.): ICT4SD 2025, LNNS 1651, pp. 241–253, 2026.
https://doi.org/10.1007/978-3-032-06688-6_24

bedridden patients. Infrared and RGB cameras can visually capture body posture, but their accuracy is usually compromised by lighting conditions and occlusions and they also pose a privacy concern. Pressure mats, however, provide a non-intrusive solution by measuring pressure distributions without influencing sleep patterns and are hence most beneficial for constant monitoring in the healthcare environment. Pressure mats offer greater comfort and dependability over wearables through the provision of non-intrusive and continuous posture monitoring. Pressure mats contrast with wearables by not making contact with the subject, thereby causing minimal interruption to sleep. This aspect makes pressure mats especially ideal for tracking bedridden or elderly patients, where comfort and uninterrupted data collection are paramount. Pressure maps can also record contact pressure changes over time, allowing precise measurement of body movement and position change without disturbing sleep. Yet, traditional 2D pressure information from pressure mats pose the following disadvantages. These are:

- Loss of Depth Information: 2D pressure maps cannot capture the three-dimensional shape of the body, and it is challenging to assess pressure distribution and spinal alignment properly.
- Difficulty in Recording Spatial Configuration: Flat pressure maps do not provide information regarding body curvature, which is necessary for proper posture estimation.
- Misclassification by Similarity of 2D Pressure Patterns: Some postures create the same pressure pattern in 2D space, resulting in erroneous classification.

To overcome the above difficulties, 3D body mesh reconstruction from pressure and depth information yields a more accurate and anatomy-consistent representation of body posture with improved classification accuracy and decreased misclassification errors. To analyze body strain, a region-wise pressure estimation module was created, which enables analysis of localized pressure patterns. This module projects pressure data onto body regions through SMPL-based anatomical landmarks, making it easier to analyze pressure variations on different body parts. This region-wise analysis enhances the detection of high-pressure zones and possible ulcer development more accurately. Sleep quality is evaluated by monitoring posture change frequency over time. People in deep, undisturbed sleep will have fewer posture changes, while those who are uncomfortable or have sleep disturbances will change postures more often. This allows continuous posture classification with 3D pressure data, enabling detailed tracking of posture transitions over time. By transforming 3D pressure maps into point clouds, posture classification accuracy is substantially enhanced as opposed to 2D classification. The improved spatial representation reduces misclassification errors due to similar 2D pressure patterns, presenting a better posture analysis and prediction framework.

2 Related Work

Posture classification and pressure analysis have gained significant popularity due to their various applications in healthcare, particularly for monitoring bedridden patients and elderly people to detect sleep-related disorders. Several studies have suggested various approaches by integrating multiple modalities and utilizing deep learning models to improve classification accuracy and prediction efficiency.

Depth and visible light cameras have been used to classify sleep postures under varied blanket conditions. Depth images capture spatial configuration, while RGB images extract anatomical keypoints, allowing deep learning models such as ResNet-34, EfficientNet B4, and ECA-Net 50. Among these models, ECA-Net 50 achieved the highest classification accuracy after training and hyperparameter tuning, highlighting the effectiveness of combining depth and anatomical landmark data to simulate real world scenarios [9]. Additionally, a system using pressure mat data provided a non-invasive method for monitoring sleep patterns and detecting sleep related disorders. This system could be used in both home and clinical environments, ensures privacy concerns, and provides a cost-effective method for long-term sleep monitoring with minimal disturbance to natural sleep patterns of a subject [13]. Depth images have been used in the field of body pose estimation even with bedding occlusion. A deep learning model incorporated a 3D body mesh to accurately infer pose using real and synthetic datasets and physics-based simulations. This method has great healthcare relevance, including pressure ulcer prevention and bedridden patient monitoring [4]. Similarly, BodyMAP [14], a 3D body pose and pressure distribution prediction model, takes depth and pressure images as input and creates a 3D human mesh with pressure map laid over it, allowing for posture and pressure analysis in several healthcare applications for patient monitoring and injury prevention. Advances in 3D human mesh construction further support these developments. MI-Mesh combines visual images and millimeter-wave radar data to construct 3D human meshes, enhancing performance in even poor lighting or occlusion conditions. This fusion of complementary modalities identifies precise body pose and shape parameters supporting several applications in healthcare and virtual reality [12]. The Skinned Multi-Person Linear (SMPL) model is a parametric 3D human body model that generates realistic human body meshes based on predicted shape and pose parameters. SMPL maps pose parameters and shape parameters to a 3D mesh with 6890 vertices and 13776 faces which makes it effective in applications such as motion capture, pose estimation and healthcare [6].

PointNet is a widely used deep learning architecture [7] that is capable of handling unstructured point cloud data through point-wise feature extraction and a symmetric max pooling function to sum up global features. PointNet is useful in applications like 3D shape classification and segmentation and can be applied to analyze 3D pressure maps by transforming them into point clouds. Its permutation invariance and capability to learn intricate spatial features make posture classification and pressure analysis more effective. PointCNN generalizes

PointNet by adding X-convolutions [8], which reorder local neighborhoods of points prior to performing convolution operations. This enables better aggregation of local features and spatial relationships, enabling PointCNN to outperform PointNet in both classification and segmentation tasks.

Our model constructs a 3D pressure map based on two input modalities depth and pressure image. The 3D pressure map constructed through SMPL model and PointNet architecture helps in identifying pressure at vertex level and more improvised classification of postures through 3D classification models such as PointCNN and PointNet

3 Proposed Methodology

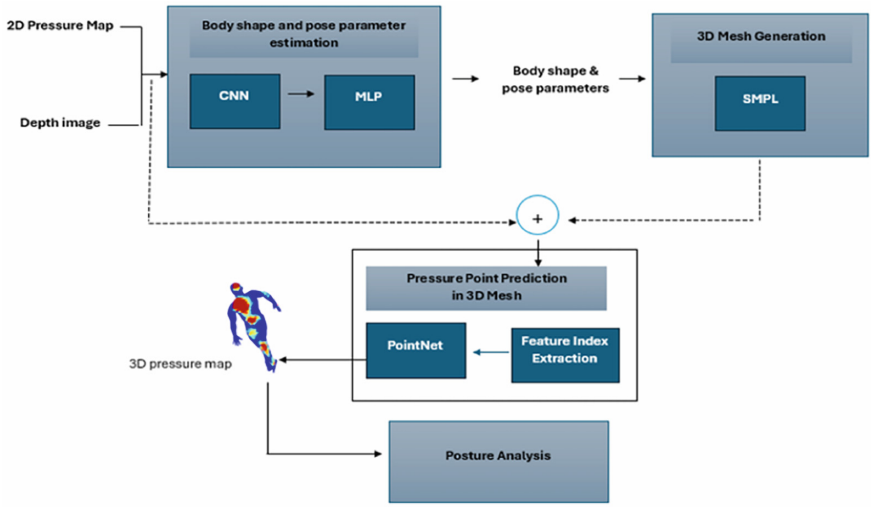


Fig. 1. Proposed Methodology.

This work presents a framework for region-wise pressure estimation and 3D posture classification using multimodal inputs from depth and pressure images. The model predicts 3D body mesh reconstruction through the Skinned Multi-Person Linear (SMPL) model, which represents the human body as shape parameters (β) for body shape (like height, weight, muscle fat and other physical properties) and pose parameters (θ) for joint rotations for all the 24 joints in the body represented as a 72 dimensional vector in all three axes. This transition from 2D to 3D analysis provides spatially more detailed pressure estimation by analyzing the surface geometry of the body, overcoming 2D pressure map depth ambiguities, and enhancing the prediction of localized pressure areas (Fig. 1).

The feature extraction pipeline employs a pre-trained ResNet-18 to encode spatial features from depth and pressure images. These are projected onto SMPL

shape and pose parameters to reconstruct a 3D body mesh ($\hat{M}$). This mesh gives a structured representation of the body surface, allowing for vertex-level pressure mapping. Two methods are developed for pressure map prediction, in the with supervision approach the model is trained directly with the ground truth pressure maps, which maintain accurate correspondence between predicted and actual pressure distributions. This method utilizes explicit supervision to learn contact information and exact pressure values. A feature indexing approach coordinates per-vertex features with corresponding pixel values in the input, allowing spatially informed pressure prediction. PointNet decoder refines these features to predict vertex-wise pressure distributions, which are processed further to compute region-specific pressure values by pooling pressure at anatomical mesh vertices. In the without supervision approach the model does not require ground truth data, rather, it utilizes implicit supervision by learning from the 3D body mesh and input depth features to predict a 3D applied pressure map. This technique is useful when labeled pressure data is not available. The without supervision approach predicts vertex-level pressure and projects it to a 2D grid through a differentiable projection module. This eliminates the need for binary contact labels, making it more suitable for real world applications.

3.1 Datasets Description

The dataset used in this research integrates real and simulated data, which helps to evaluate the methods proposed.

The SLP dataset [4], as shown in Fig. 2 is specifically designed for human pose monitoring in bed and includes data in several modalities like RGB, infrared, depth, and pressure images. Data was collected from 109 subjects, in both hospital and regular settings, each maintaining 15 poses into three standard postures supine, left, and right respectively. Recordings were conducted under three blanket conditions: uncovered, thin cover, and thick cover.

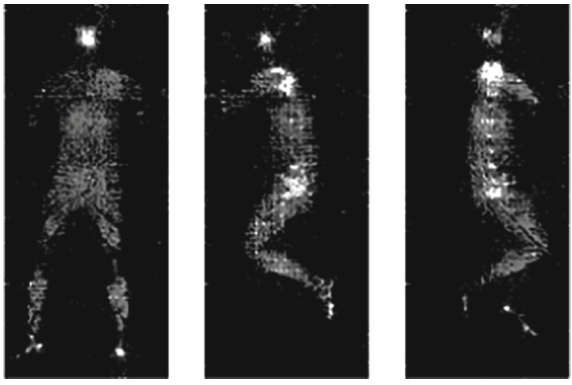


Fig. 2. SLP dataset with three standard postures: Supine, Left, and Right.

The BodyPressureSD dataset [5] was generated using physics-based simulations based on body poses and shapes from the SLP dataset as shown in Fig. 3. It simulates bodies lying on a soft mattress with pressure-sensing mats and various blanket occlusions. Depth images were captured from a simulated overhead camera, incorporating diverse body shapes, poses, and blanket conditions. The dataset includes 97,495 samples, each containing a 3D resting pose, body shape, depth image, pressure image, and associated meshes for the individual, mattress, pressure mat, and blanket.

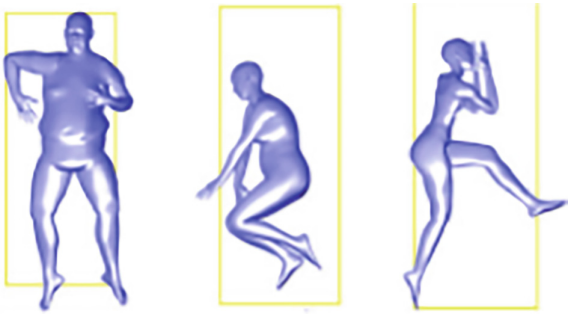


Fig. 3. BodyPressureSD dataset with different physics orientations.

3.2 3D Pressure Map Modeling

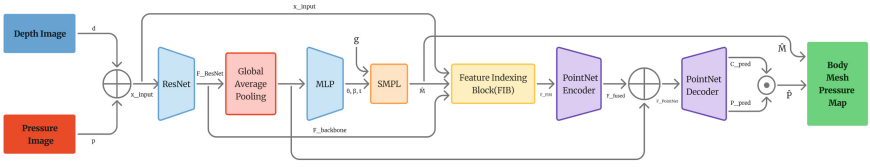


Fig. 4. With Supervision Architecture.

The process starts with two inputs: a depth image ($d \in \mathbb{R}^{H \times W}$), and a 2D pressure map ($p \in \mathbb{R}^{H \times W}$) as shown in Fig. 4, following [14]. Both the depth image and pressure map are resized and stacked into a multi-channel input ($x\_input \in \mathbb{R}^{H \times W \times 2}$). This input is processed by a pretrained ResNet-18 model to derive spatial and structural features that are pooled to get a global feature vector ($F\_ResNet \in \mathbb{R}^{d}$). The derived features are fed to a multi-layer perceptron (MLP) to predict the SMPL (Skinned Multi-Person Linear) model parameters, such as body shape ($\beta \in \mathbb{R}^{10}$), joint angles ($\theta \in \mathbb{R}^{72}$), and root-joint translation ($t \in \mathbb{R}^{3}$). The SMPL model [6], a statistical model of human

shape and pose, employs these parameters in combination with the gender (g) of the subject to produce a 3D body mesh ($\hat{M} \in \mathbb{R}^{V\times 3}$) with joint and vertex locations. The mesh is reconstructed by an SMPL embedding block, which is a non-trainable differentiable function. The Feature Indexing Block (FIB) builds informative features for every mesh vertex based on spatial correspondences between the predicted body mesh and input images. It integrates raw pixel values with high-level features learned from deep networks. FIB learns features via two processes: (1) Projecting estimated mesh vertices, $\hat{M} \in \mathbb{R}^{V\times 3}$, onto the depth ($d$) and pressure ($p$) images to get initial vertex features of shape $(V, 2)$. (2) Extracting deep features from middle backbone layers prior to global average pooling, yielding a feature representation of size $(V, 3)$. The concatenated features from all these sources create a high-level descriptor per vertex, which is further refined by a fully connected (FC) layer, projecting it to shape $(V, 16)$. Following the Feature Indexing Block, PointNet is used to handle the indexed per-vertex features and predict 3D pressure map $\hat{P}$. ResNet deep features (following global average pooling) are combined with PointNet encoder representations to facilitate spatial perception and contact dynamics. PointNet decoder generates two outputs for each vertex: a binary contact prediction and a per-vertex pressure value. The binary contact map modulates the output pressure map by assigning pressure only to vertices labeled as in contact with the surface. Let $\hat{M}_i$ denote the i-th vertex of the predicted mesh and (u_i, v_i) its projected coordinates. $F\_backbone \in \mathbb{R}^{H\times W\times C}$ is the extracted feature map, and $F\_ResNet \in \mathbb{R}^{C}$ is the global feature from ResNet. F_FIB denotes the processed indexed features, F_fused the concatenated feature vector, $F_$PointNet the encoded feature from PointNet, and $\hat{P}$ the final predicted pressure map. The process is formulated as:

$$f_i = \text{concat}\left([d(u_i, v_i),\, p(u_i, v_i)],\, F\_backbone(u_i, v_i)\right) \tag{1}$$

$$f\_processed\_i = \text{FC}(f_i) \tag{2}$$

$$F\_FIB = \{f\_processed\_i\}_{i=1}^{V} \tag{3}$$

$$F\_fused = \text{concat}(F\_FIB,\, F\_ResNet) \tag{4}$$

$$F\_PointNet = \text{PointNet\_Encoder}(F\_fused) \tag{5}$$

$$F\_decoded = \text{PointNet\_Decoder}(F\_PointNet) \tag{6}$$

$$C\_pred = \sigma(W_c F\_decoded), \quad P\_pred = W_p F\_decoded, \quad \hat{P} = C\_pred \odot P\_pred \tag{7}$$

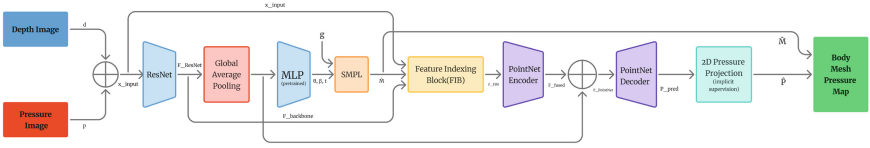


Fig. 5. Without Supervision Architecture.

In the without supervision approach as shown in Fig. 5, uses a pre-trained MLP (applied for predicting SMPL parameters in the supervised model) to generate mesh predictions without further training. ResNet image features are used as basic inputs, with the architecture maintaining the Feature Indexing Block (FIB) and PointNet for feature processing and extraction. This method utilizes implicit supervision without requiring ground truth information making it more suitable for real-world scenarios. In contrast to the with supervision approach, this approach directly predicts vertex-level pressure (P_pred) without binary contact prediction. A differentiable 2D projection module projects the output of the PointNet decoder (P_pred) to produce a 2D pressure map that is aligned with the input pressure image. The projection associates vertex-level pressure values with corresponding pressure taxels, calculating the average pressure of vertices situated above each taxel to create the final 2D pressure representation.

3.3 Body Strain Analysis Through Region-Wise Pressure Estimation

To analyze body strain and potential discomfort during prolonged sleep, region-wise pressure estimation is conducted using anatomical landmarks defined by the SMPL model. The approach locates key regions such as the head, shoulders, torso, arms, and legs by selecting predefined landmark vertices from the 3D body mesh. Pressure values corresponding to these landmarks are extracted from the predicted pressure map. For regions represented by multiple landmarks, the average pressure across these points is computed to provide localized and more accurate region wise pressure estimation. This analysis helps in locating pressure hotspots and identifying areas of body strain which is crucial for evaluating sleep ergonomics and assess sleep discomfort.

3.4 Sleep Quality Analysis Through Posture Classification

Sleep posture is essential in analyzing sleep behavior and comfort. Frequent changes in posture can indicate restlessness, while an extended period of staying in one posture can cause pressure ulcers. To analyze sleep quality, this module monitors posture changes and the frequency of time in the supine, left, and right position using reconstructed 3D pressure maps. The reconstructed pressure map is transformed into a point cloud that acts as input to the classification models. The PointNet and PointCNN architectures were employed for classification. PointNet [7] transforms the point cloud using multilayer perceptrons (MLPs) and then global max pooling to encode global shape information. The ultimate classification is done by fully connected (FC) layers. PointCNN [8] improves PointNet with a learnable local neighborhood aggregation component for better spatial relationship encoding. It employs convolutional layers to extract local features and then max pooling and FC layers for classification. PointCNN shows better posture classification accuracy due to its additional spatial encoding.

For assessing sleep quality, we examine the rate of change of sleep postures during the night. Low posture change rate is typically correlated with deep

and restful sleep, while high rates can reflect discomfort, restlessness, or sleep disruptions [15]. The reconstructed 3D pressure maps, which are produced at discrete time intervals, are utilized for posture classification with a deep learning model (PointNet or PointCNN). These postures are monitored over time, and a counter is incremented each time a change in posture is observed. If the number of changes is less than a predetermined threshold, the sleep is labeled as deep sleep; otherwise, it is labeled as discomfort. The process is outlined in Algorithm 1.

Algorithm 1. Sleep Posture-Based Sleep Quality Assessment

Require: Time threshold t, Posture change threshold N
Initialize counter $C \leftarrow 0$, previous posture $P_{\text{prev}} \leftarrow$ None
for each time i from 0 to T with step size t **do**
    Predict posture P using PointNet/PointCNN
    if $P_{\text{prev}} \neq$ None and $P \neq P_{\text{prev}}$ **then**
        $C \leftarrow C + 1$
    end if
    $P_{\text{prev}} \leftarrow P$
end for
if $C < N$ **then**
    Sleep classified as **deep sleep**
else
    Sleep classified as **discomfort**
end if

4 Results and Discussion

The proposed models were trained with batch size of 64, learning rate of 0.0001, and weight decay of 0.0005 with Adam optimizer. Random rotation and random erasing augmentations were used to enhance generalization. The with supervision approach model was trained for 100 epochs and the without supervision approach model was trained for 15 epochs. The table 1 compares errors across different models for pressure map prediction.

Table 1. Comparison of 3D pose and pressure errors across different models.

Model	MPJPE (mm)↓	PVE (mm)↓	v2vP (kPa^2)↓
Conv	51.79 ± 0.379	62.9 ± 0.32	2.56 ± 0.007
PointNet-With Supervision	51.01 ± 1.071	61.66 ± 0.919	2.14 ± 0.030
PointNet-Without Supervision	98.82	-	2.513

- MPJPE: Measures the Euclidean distance between predicted and ground truth 3D joint positions, defined as

$$MPJPE = \frac{1}{N}\sum_{i=1}^{N} \|\hat{\mathbf{J}}_i - \mathbf{J}_i\|_2 \tag{8}$$

where $\hat{\mathbf{J}}_i$ and $\mathbf{J}_i$ denote predicted and ground truth joint positions, and N is the number of joints.

- PVE: Computes the average Euclidean distance between predicted and ground truth 3D mesh vertices, given by

$$(PVE = \frac{1}{M}\sum_{i=1}^{M} \|\hat{\mathbf{V}}_i - \mathbf{V}_i\|_2 \tag{9}$$

where $\hat{\mathbf{V}}_i$ and $\mathbf{V}_i$ are the predicted and ground truth vertex positions, and M is the number of vertices.

- v2vP: Quantifies pressure prediction error as the squared difference between predicted and actual pressure values, formulated as

$$v2vP = \frac{1}{M}\sum_{i=1}^{M} (\hat{P}_i - P_i)^2 \tag{10}$$

where $\hat{P}_i$ and P_i denote the predicted and actual pressure values at each vertex.

In this study, to assess the sleep quality, we implemented and evaluated two models for posture classification using SLP dataset for both with and without supervision approaches. The models were evaluated based on several performance metrics as shown in Tables 2 and 3. The results of classification shown in Tables 2 and 3 reveal that PointCNN performs better than PointNet overall in the unsupervised approach. This implies that PointCNN's use of hierarchical convolutional methodology better captures local and global geometric information from the point cloud data. However, in the supervised approach, PointNet has better accuracy compared to PointCNN. This performance change indicates that PointNet gains more from explicit supervision, possibly because it can retain global shape features effectively without depending on local neighborhood structures, which are more susceptible to supervision constraints in PointCNN.

Table 2. Performance comparison of PointNet and PointCNN under the *Without Supervision* approach (in %).

Model	Accuracy	Precision	Recall	F1-score
PointNet	99.09	99.10	99.09	99.09
PointCNN	99.49	99.50	99.49	99.49

Table 3. Performance comparison of PointNet and PointCNN under the *With Supervision* approach (in %)

Model	Accuracy	Precision	Recall	F1-score
PointNet	99.26	99.26	99.26	99.26
PointCNN	97.89	97.91	97.89	97.90

To assess the quality of pressure distribution analysis between various models as shown in Fig. 6, we contrast the 2D pressure map segmentation with the 3D body mesh pressure mapping. Table 4 shows the average pressure values measured across various body parts in the 2D pressure map, and Table 5 presents the important anatomical points within the 3D model which are not well defined in 2D but have significant pressure values. The 2D pressure map provides a broad categorization of pressure distribution, which is effective for general posture classification but lacks accuracy for more detailed pressure mapping. The 3D model provides localized pressure measurements at particular anatomical locations, allowing for a more accurate examination of body contact points against the surface. The lower back region indicates prominent pressure, matching main contact areas that affect spinal support. Pressure of upper limb, specifically the right bicep region and right forearm region, indicates areas of unintended load, potentially indicating arm positioning effects that are hard to evaluate from 2D data only. The inseam region indicates measurable pressure, providing insights into contact forces in the inner thigh region, which cannot be directly inferred from 2D segmentation.

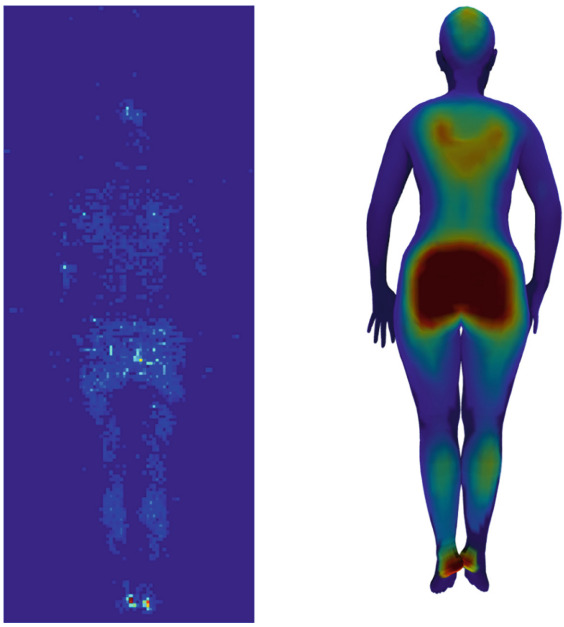


Fig. 6. Pressure distribution comparison (2D vs 3D).

Table 4. Pressure distribution across segmented body regions (2D data)

Body Region	Average Pressure (in kPa)
Head & Neck	0.2855
Shoulders	0.4129
Chest & Upper Back	1.5783
Abdomen & Lower Back	0.8441
Hips & Thighs	2.9782
Legs & Feet	0.9042

Table 5. Localized pressure distribution at key anatomical points (3D data)

Body Part	Total Pressure (in kPa)
Lower Back (Belly Button)	12.11
Left Heel	29.92
Right Heel	24.88
Right Forearm	2.69
Right Bicep	2.58
Inseam Point	0.22

5 Conclusion

This work proposes a deep learning framework that predicts 3D body pose and pressure distribution from depth and pressure inputs by combining ResNet-18 and PointNet features with a feature indexing strategy, achieving anatomically accurate pressure maps using both with-supervision and without-supervision approaches. Analysis pressure across regions showed that 3D mapping has the ability to capture fine pressure changes at important anatomical points such as heels, wrists, and back, over coarse 2D estimates. In posture classification, PointCNN performed well in without supervision because of hierarchical feature learning and PointNet effectively learned spatial relationships. The results of the posture classification were used to assess sleep quality by identifying frequent posture changes. The work emphasizes advantages of fusing 2D and 3D modalities for complete analysis of sleep quality and body stress, with long-term directions revolving around improved fusion features and clinical implementation in real time.

References

1. Berlowitz, D., VanDeusen Lukas, C., Parker, V., Niederhauser, A., Silver, J., Logan, C., Ayello, E.: Preventing pressure ulcers in hospitals: a toolkit for improving quality of care. In: Agency for Healthcare Research and Quality (2011)

2. Mansfield, S., Obraczka, K., Roy, S.: Pressure injury prevention: a survey. IEEE Rev. Biomed. Eng. **13**, 352–368 (2020)
3. Scott, R.G., Thurman, K.M.: Visual feedback of continuous bedside pressure mapping to optimize effective patient repositioning. Adv. Wound Care **3**(5), 376–382 (2014)
4. Liu, S., Huang, X., Fu, N., Li, C., Su, Z., Ostadabbas, S.: Simultaneously collected multimodal lying pose dataset: enabling in-bed human pose monitoring. IEEE Trans. Pattern Anal. Mach. Intell. **45**(1), 1106–1118 (2022)
5. Clever, H., Grady, P., Turk, G., Kemp, C.: BodyPressure: inferring body pose and contact pressure from a depth image. IEEE Trans. Pattern Anal. Mach. Intell. (2022). https://doi.org/10.1109/TPAMI.2022.3158902
6. Loper, M., Mahmood, N., Romero, J., Pons-Moll, G., Black, M.J.: SMPL: a skinned multi-person linear model. ACM Trans. Graph. **34**(6), Article 248, 16 (2015). https://doi.org/10.1145/2816795.2818013
7. Qi, C.R., Su, H., Mo, K., Guibas, L.J.: PointNet: deep learning on point sets for 3D classification and segmentation. In: Proceedings of IEEE Conference Computer on Vision Pattern Recognition (CVPR), pp. 652–660 (2017). https://doi.org/10.1109/CVPR.2017.16
8. Li, Y., Bu, R., Sun, M., Wu, W., Di, X., Chen, B.: PointCNN: convolution on X-transformed points. In: Advances Neural Information Processing Systems (NeurIPS), pp. 820–830 (2018)
9. Tam, A.Y.-C., et al.: Depth-camera-based under-blanket sleep posture classification using anatomical landmark-guided deep learning model. Int. J. Environ. Res. Public Health **19**(20), 13491 (2022). https://doi.org/10.3390/ijerph192013491
10. Khan, A., Kim, C., Kim, J.-Y., Aziz, A., Nam, Y.: Sleep posture classification using RGB and thermal cameras based on deep learning model. CMES Comput. Model. Eng. Sci. **140**(2), 1729–1755 (2024). https://doi.org/10.32604/cmes.2024.049618
11. Huang, W., Wai, A.A.P., Foo, S.F., Biswas, J., Hsia, C.-C., Liou, K.: Multimodal sleeping posture classification. In: Proceedings of 20th International Conference on Pattern Recognition (ICPR), pp. 4336–4339. IEEE, Istanbul (2010). https://doi.org/10.1109/ICPR.2010.1054
12. Xue, H., et al.: mmMesh: towards 3D real-time dynamic human mesh construction using millimeter-wave. In: Proceedings of 27th Annual International Conference on Mobile Computer Networks, pp. 269–282 (2021). https://doi.org/10.1145/3458864.3467679
13. Metsis, V., Galatas, G., Papangelis, A., Kosmopoulos, D., Makedon, F.: Recognition of sleep patterns using a bed pressure mat. In: Proceedings of 4th International Conference on Pervasive Technology Related to Assistive Environments (PETRA), pp. 1–4 (2011). https://doi.org/10.1145/2141622.2141633
14. Tandon, A., Goyal, A., Clever, H., Erickson, Z.: BodyMAP: jointly predicting body mesh and 3D applied pressure map for people in bed. In: Proceedings of IEEE Conference on Computer Vision Pattern Recognition (CVPR), pp. 2480–2489 (2024). https://doi.org/10.1109/CVPR52733.2024.00240
15. Li, Y.Y., Wang, S.J., Hung, Y.P.: A vision-based system for in-sleep upper-body and head pose classification. Sensors **22**(5), 2014 (2022). https://doi.org/10.3390/s22052014

Docker Container Security: A Scanning-Centric Security Framework

V. Sudeep, V. Nishant, M. M. Mohamed Jasir Faiez, T. Monish, G. P. Yuvaraj Kumar, K. J. Akhil, and K. Praveen(✉)

TIFAC-CORE in Cyber Security, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India

{cb.en.u4cys21073,cb.en.u4cys21049,cb.en.u4cys21043, cb.en.u4cys21045,cb.en.u4cys21090}@cb.students.amrita.edu, {kjakhil,praveen}@cb.amrita.edu

Abstract. Docker containers are central to modern software development and deployment due to their portability, efficiency, and scalability. By isolating applications and dependencies, they provide a lightweight alternative to virtual machines, enabling consistent environments across platforms. However, Docker containers pose security challenges, including shared kernel risks, vulnerabilities in container images, and misconfigurations, which can lead to breaches. This paper examines security concerns in Docker containers and proposes a framework to identify and address vulnerabilities. The framework helps detect issues like outdated components and misconfigurations, offering insights to enhance security. Through practical use cases, it highlights its effectiveness in closing security gaps and equipping developers with tools to protect containers. The study emphasizes the need for proactive security measures and continuous vigilance in securing containerized systems.

Keywords: Docker Security · Container Risk Assessment · Cybersecurity Framework · Container Orchestration Security · Runtime Security

1 Introduction

Docker containers have revolutionized software deployment by providing lightweight, portable environments that ensure consistent execution across diverse platforms. However, containers also introduce significant security challenges, such as risks from the shared kernel model, vulnerabilities in container images, and configuration errors that can expose applications and infrastructure to potential threats [1, 2]. Docker employs namespaces and cgroups for isolation, but misconfigurations or vulnerabilities can compromise this isolation, leading to privilege escalation or unauthorized access [3, 4]. Additionally, container images, often sourced from public registries like DockerHub, may include outdated or unpatched software, increasing exploitation risks [5, 6]. The complexity of securing containerized environments grows further with orchestration systems like Kubernetes, where misconfigurations can cause data breaches or unauthorized deployments [7, 8].

S. Fong et al. (Eds.): ICT4SD 2025, LNNS 1651, pp. 254–262, 2026.
https://doi.org/10.1007/978-3-032-06688-6_25

This paper introduces a framework to address these challenges by identifying vulnerabilities, outdated components, and misconfigurations in Docker container images. The framework integrates security practices into the container development lifecycle, enabling developers and security professionals to proactively secure their environments [9, 10]. Docker's architecture, comprising images, containers, registries, and the engine, forms the foundation of this analysis. While images provide the blueprint, containers run applications with isolation, sharing the host system's kernel. Registries like DockerHub allow sharing of images, and the Docker Engine orchestrates container management. These shared resources and configurations require robust security measures [11, 12]. Existing studies emphasize critical solutions such as vulnerability scanning, secure image development, and runtime monitoring [13, 14]. For example," Container Security in Cloud Environments: A Comprehensive Analysis" highlights the importance of DevSecOps pipelines and runtime detection [15], while" Containers' Security: Issues, Challenges, and Road Ahead" explores isolation mechanism limitations and risks from untrusted images [16]. Building on these insights, this paper proposes an automated framework leveraging secure methodologies and tools to mitigate vulnerabilities and strengthen container security [17, 18].

2 Security Challenges in Docker Containers

Docker containers, while transformative for software deployment, encounter significant security challenges throughout their lifecycle [1, 2]. A major concern involves container images, often sourced from public repositories like Docker-Hub, which may include outdated software or unpatched vulnerabilities [3, 4]. Exploitation of these vulnerabilities can lead to unauthorized access, emphasizing the need for regular image scanning and updates [5, 6]. Container isolation also poses risks [7, 8]. Despite the use of namespaces and cgroups, Docker configurations are not foolproof; misconfigurations or exploits may result in privilege escalation or host breaches [9, 10]. Runtime security is another critical concern, particularly when containers operate with root privileges, enabling potential resource abuse or unauthorized access [11, 12]. Moreover, container orchestration platforms such as Kubernetes introduce additional attack surfaces [13, 14]. Misconfigurations like exposed APIs or insecure RBAC policies can lead to data breaches and unauthorized deployments [15, 16]. These challenges highlight the necessity of continuous monitoring, secure configurations, and effective vulnerability management to maintain container security [17, 18].

2.1 Insights from the Vulnerability Database

A key aspect of securing Docker containers is understanding vulnerabilities. Our vulnerability database, spanning from 1999 to 2024, records 10,391 vulnerabilities, ranging from Dockerfile misconfigurations to privilege escalation exploits [19–22]. Legacy vulnerabilities persist in older systems, while modern threats evolve, showcasing the need for proactive measures. Many vulnerabilities are critical, leading to risks like data breaches or service disruptions. Integrating this database into our framework enables organizations to identify and address risks, enhancing containerized environments' security [19, 23, 24].

3 Proposed Framework/Approach

To address the security challenges of Docker containers, this study proposes a vulnerability detection framework that secures container images by identifying vulnerabilities during the build phase and offering actionable mitigation strategies. As shown in Fig. 1, the methodology begins with clients submitting a Dockerfile and related package files through a REST API. The backend processes these inputs in a secure, isolated sandbox environment to build the Docker image, mitigating risks to the host system. Upon completion, a tool such as Syft extracts the image's dependencies and components, which are then analyzed against a vulnerability database comprising over 10,000 known issues, including CVE identifiers, severity levels, fixed versions, and CVSS scores. The framework identifies vulnerable dependencies, misconfigurations, and outdated components, compiling the results into a structured JSON report. This report is returned to the client via the REST API, providing detailed insights and remediation recommendations. By enabling early vulnerability detection, the framework minimizes the risk of deploying insecure containers and offers an automated, streamlined security assessment.

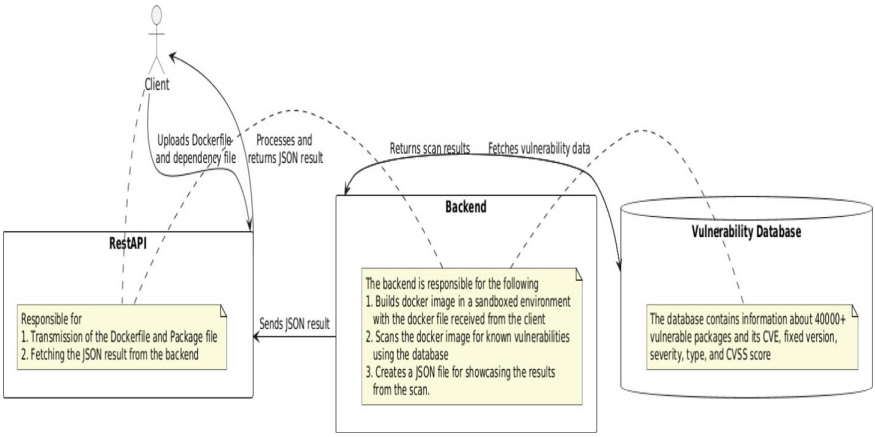


Fig. 1. Proposed Framework

3.1 Client Interaction

The framework begins with Client Interaction, where users upload a Dockerfile or package files through a UI or script communicating with the REST API. These files serve as inputs for vulnerability analysis. The interface ensures a seamless user experience and securely transmits data for backend processing.

3.2 RestAPI Component

The REST API securely handles file uploads, forwards them for scanning, and retrieves results. Following RESTful standards, it ensures reliable data exchange and request validation to maintain system integrity.

3.3 Backend Processing

Backend processing forms the core of the framework. Creates a secure sandbox to build the Docker image from the client™s Dockerfile. After construction, the image is scanned for vulnerabilities using the database. The findings are compiled into a structured JSON with details on vulnerabilities, severity levels, and recommended fixes. The process ensures isolation to protect the host environment.

3.4 Vulnerability Database

The Vulnerability Database is the system™s knowledge base, storing over 10,000 vulnerabilities with CVE identifiers, fixed versions, severities, and CVSS scores. Primary keys ensure record uniqueness, and CVE IDs link vulnerabilities to severity scores for risk prioritization. The database is updated regularly to detect new threats.

To broaden coverage, the framework integrates sources like NVD, Exploit-DB, and vendor advisories, improving accuracy, early threat detection, and metadata richness to counter evolving attacks.

1. **Receive Dockerfile and Package File:** The process begins with the client uploading the Dockerfile and associated package file. These files provide the essential configuration and dependencies required to build the Docker image.
2. **Push Build Task to Queue:** Each build task is queued to ensure efficient processing. This queuing mechanism is essential as building Docker images can be resource-intensive and time-consuming.
3. **Start Sandbox Environment:** To maintain security, the Docker image is constructed in a sandboxed environment. This isolated setup prevents potential security risks from affecting the host system.
4. **Build Docker Image:** The queued task is executed, and the Docker image is built in the sandboxed environment. This ensures that the image accurately reflects the uploaded Dockerfile configuration (Fig. 2).
5. **Run Syft to Extract Dependencies:** Once the Docker image is built, the Syft tool is used to scan the image. Syft extracts a comprehensive list of all dependencies and libraries included in the image, forming the foundation for vulnerability analysis.
6. **Check Dependencies for Vulnerabilities:** The extracted dependencies are compared against the vulnerability database. The database, encompassing over 10,000 vulnerabilities spanning from 1999 to 2024, identifies potential weaknesses in the image.
7. **Create JSON File with Results:** After analyzing the dependencies, a JSON file is generated. This file contains detailed information about the vulnerabilities detected, including their severity, CVEs (Common Vulnerabilities and Exposures), and remediation steps.
8. **Send JSON File to RestAPI:** Finally, the JSON file is transmitted to the REST API. This API serves as the communication layer, providing the client with the vulnerability analysis results in a structured and accessible format.

The implementation was successfully tested using various Docker images with various configurations. The workflow demonstrated high efficiency in identifying vulnerabilities while maintaining secure operations in the sandbox environment. The accurate

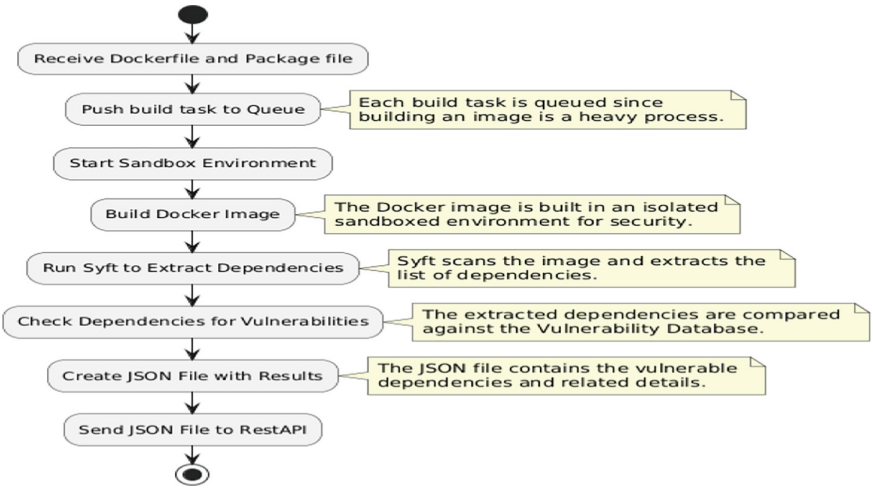


Fig. 2. Workflow

extraction of Syft dependency and the extensive coverage of the vulnerability database ensured comprehensive vulnerability detection. The JSON output facilitated seamless integration with other systems, enabling further analysis and visualization. The results revealed critical vulnerabilities in many publicly available images, including outdated libraries and unpatched software components. The detailed JSON output provided essential information, such as CVE IDs, severity levels, and fixed versions. The framework successfully identified security gaps, allowing developers to address vulnerabilities before deployment, emphasizing the importance of scanning in securing containerized applications.

4 Implementation and Results

1. **Receive Node.js File and Dockerfile:** The client uploads the `Dockerfile` and Node.js application files containing the application logic and dependencies.
2. **Push Build Task to Queue:** Each submission is added to a queue, enabling efficient and orderly processing of multiple build requests.
3. **Start Sandbox Environment:** A secure sandbox environment is initialized to safely execute and analyze the Node.js application without affecting the host system.
4. **Build Docker Image:** The application is containerized using the submitted `Dockerfile`, including the installation of required Node.js runtimes and dependencies.
5. **Run Syft to Extract Dependencies:** The Syft tool scans the built Docker image to inventory all dependencies and their versions.
6. **Check Dependencies for Vulnerabilities:** The extracted dependencies are checked against a vulnerability database to identify outdated or insecure libraries.
7. **Create JSON File with Results:** A structured JSON file is generated, detailing detected vulnerabilities, including CVE IDs, severity scores, and remediation steps.
8. **Send JSON File to REST API:** The JSON report is transmitted to the REST API, delivering structured and actionable vulnerability insights to the client.

4.1 Output Analysis

Figure 3 highlights the vulnerabilities found in the dependencies of the Node.js application. Key observations include:

Dependencies and Versions: The output lists all Node.js packages included in the application, such as `libopenjp2-7` (used internally by some dependencies). Each dependency's version is specified.
Vulnerability Status: Many dependencies are marked with `Not Fixed` or `(won't fix)` vulnerabilities. This indicates that certain libraries either do not have available patches or are no longer actively maintained.
Severity Levels: Each CVE is assigned a severity score (CVSS), ranging from medium (6.5) to critical (9.8). This scoring helps prioritize which vulnerabilities need immediate attention.
Repetition of Dependencies: Some dependencies, like `libopenjp2-7` and its development variant `libopenjp2-7-dev`, appear multiple times in the output with different CVE entries. This signifies multiple unresolved vulnerabilities within the same library.
Node.js Specific Findings: Since Node.js applications rely heavily on third-party libraries from `npm`, the dependency tree is often extensive. The presence of multiple vulnerabilities in popular libraries is a common issue due to the large ecosystem.

libopenjp2-7	2.5.0-2	(won't fix)	deb	CVE-2021-3575	7.8
libopenjp2-7	2.5.0-2	Not Fixed	deb	CVE-2016-10505	6.5
libopenjp2-7	2.5.0-2	Not Fixed	deb	CVE-2016-10506	6.5
libopenjp2-7	2.5.0-2	Not Fixed	deb	CVE-2016-9113	7.5
libopenjp2-7	2.5.0-2	Not Fixed	deb	CVE-2016-9114	7.5
libopenjp2-7	2.5.0-2	Not Fixed	deb	CVE-2016-9115	6.5
libopenjp2-7	2.5.0-2	Not Fixed	deb	CVE-2016-9116	6.5
libopenjp2-7	2.5.0-2	Not Fixed	deb	CVE-2016-9117	6.5
libopenjp2-7	2.5.0-2	Not Fixed	deb	CVE-2016-9580	8.8
libopenjp2-7	2.5.0-2	Not Fixed	deb	CVE-2016-9581	8.8
libopenjp2-7	2.5.0-2	Not Fixed	deb	CVE-2017-17479	9.8
libopenjp2-7	2.5.0-2	Not Fixed	deb	CVE-2018-16375	8.8
libopenjp2-7	2.5.0-2	Not Fixed	deb	CVE-2018-16376	8.8
libopenjp2-7	2.5.0-2	Not Fixed	deb	CVE-2018-20846	6.5
libopenjp2-7-dev	2.5.0-2	(won't fix)	deb	CVE-2019-6988	6.5
libopenjp2-7-dev	2.5.0-2	(won't fix)	deb	CVE-2021-3575	7.8
libopenjp2-7-dev	2.5.0-2	Not Fixed	deb	CVE-2016-10505	6.5

Fig. 3. Framework output

4.2 Recommendations

The testing of the Node.js application within the Docker container using the proposed framework successfully identified several vulnerabilities. Regularly updating or replacing deprecated libraries is essential to minimize exposure to known risks, while vulnerabilities with high CVSS scores (above 9.0) must be addressed immediately to reduce critical security threats. This evaluation emphasizes the importance of integrating automated vulnerability scanning into the Node.js development lifecycle to ensure secure and reliable containerized applications.

5 Discussion and Analysis

The proposed framework offers several advantages over existing approaches. The use of sandbox environments enhances security during image building, preventing risks to the host system. Compared to manual scan or runtime-only monitoring, this framework offers a more comprehensive approach by integrating static analysis. A key challenge in large-scale deployments is balancing security with performance, as frequent deep scans can slow down builds and consume resources. Adaptive scanning, which focuses on high-risk components or runs during offpeak hours, helps maintain system responsiveness. However, some limitations remain. Detection accuracy depends on the quality of the vulnerability database,

and scanning large images can introduce overhead. Future improvements could include machine learning for anomaly detection and automated patching. Furthermore, this framework can be seamlessly integrated into DevSecOps pipelines, enabling continuous vulnerability checks in CI/CD stages. With minor adaptations, it can also work alongside container orchestration tools like Kubernetes using admission controllers or custom security policies, automating remediation before deployment. To support large-scale use, the framework employs task queueing and parallel processing. Its modular architecture allows scaling across multiple teams or projects. Centralized logging, API rate limiting, and loadbalanced backend services ensure consistent performance in high-demand environments (Fig. 4).

Feature	Proposed Framework	Grype	Docker Scout
Platform (CLI/Website)	Website	CLI	CLI + Website
Scans Container Images	No	Yes	Yes
Scans Filesystems	No	Yes	No
Scans Dockerfile	Yes	No	No
Scans Requirements Files	Yes	No	No
Supports Base Image Selection	Yes	No	No
Provides Package Details	Yes	Yes	Yes
Lists Vulnerabilities	Yes	Yes	Yes
Includes CVSS Scores	Yes	No	Yes
Provides Fix Information	Yes	Yes	Yes
Severity Levels	Yes	Yes	Yes
OS Vulnerability Analysis	No	No	Yes
Policy Violation Dashboard	No	No	Yes
Dashboard Insights	Yes	No	Yes
Base Image Dependency Matching	Yes	No	No

Fig. 4. Comparison

6 Conclusion

Securing Docker containers is crucial as containerized applications become central to modern software deployments. This paper explored key security challenges in Docker containers, including vulnerabilities in images, isolation issues, and orchestration risks. The proposed automated framework addresses these challenges by using tools like Syft

for dependency extraction and vulnerability scanning against a comprehensive database. The implementation results demonstrated the effectiveness of the framework in identifying and mitigating vulnerabilities, ensuring more secure containerized applications. In conclusion, adopting automated vulnerability scanning frameworks enhances container security, reduces risks, and improves deployment reliability. Future research can focus on incorporating advanced techniques for real-time monitoring and proactive threat mitigation to further strengthen container security.

Acknowledgment. We would like to thank **Dr. M. Sethumadhavan**, Professor and Head, TIFACCORE in Cyber Security, Amrita Vishwa Vidyapeetham, Coimbatore, India and **Mr. Sitaram Chamarty**, Principal Consultant, Tata Consultancy Services (TCS), for their suggestions, directions, and recommendations.

References

1. Kumar, A., et al.: Secure multi-tenant environments in the cloud: challenges and opportunities. J. Cloud Comp. **5**, 123–134 (2021)
2. Mustafa, A., et al.: Hybrid cloud architectures and security challenges. In: 2022 IEEE International Conference on Cloud Computing, pp. 1–10. IEEE, New York (2022)
3. Martinez, S., et al.: Trust models for IoT device integration in cloud platforms. IoT Secur. J. **12**(4), 189–202 (2023)
4. Ramesh Kumar, S., et al.: Emerging trends in cloud-native security architectures. Cloud-Native Conference Proceedings, pp. 85–96. Springer, Berlin (2024)
5. Gupta, R., et al.: Microservices security for healthcare platforms in cloud. In: 11th International Conference on Healthcare Informatics, pp. 311–320. Springer, London (2022)
6. Emily Davis, C., et al.: Blockchain solutions for secure data sharing in cloud systems. J. Blockchain Technol. **15**(3), 97–115 (2023)
7. Singh, M., et al.: Performance impact of container isolation techniques in the cloud. J. Sys. Perform. **8**, 56–72 (2021)
8. Xiu, W., et al.: Kubernetes security enhancements using AI-based algorithms. In: 2023 Kubernetes Symposium, pp. 102–112. Springer, Berlin (2023)
9. Ravi, N., et al.: AI for detecting and mitigating container attacks. IEEE AI Transactions **6**(1), 45–56 (2022)
10. Raj, A., et al.: Container registry security: challenges and best practices. ACM Cloud Sys. J. **10**(5), 77–88 (2024)
11. Hassan, I., et al.: Securing serverless applications using runtime monitoring. In: Proceedings of 15th IEEE Cloud Security Conference, pp. 211–221. IEEE, Singapore (2023)
12. Menon, A., et al.: A comparative study on DevSecOps practices for cloud environments. J. DevSecOps **7**(2), 13–27 (2022)
13. Kumar, H., et al.: Understanding zero-trust security models in multi-cloud deployments. In: 2022 International Conference on Cloud Computing and Security, pp. 1–12. Springer, Berlin (2022)
14. Patil, A., et al.: Data encryption approaches for cloud security. Springer Lecture Notes in Computer Science **12011**, 45–60 (2024)
15. Winters, Z., et al.: A holistic framework for identity and access management in the cloud. J. Cloud Manage. **5**(1), 33–46 (2023)
16. Alves, P., et al.: Penetration testing in containerized cloud systems. J. Cybersecur. **8**(4), 213–230 (2022)

17. Gupta, A., et al.: Leveraging artificial intelligence for cloud security optimization. AI Secur. Trans. **12**(6), 77–90 (2023)
18. Sharma, M., et al.: Evolving threats in the cloud and countermeasures. Journal of Threat Intelligence **9**(2), 113–129 (2021)
19. Singh, A., et al.: Securing IoT in cloud systems: a machine learning perspective. In: IoT and Cloud Systems Conference Proceedings, pp. 67–78. Springer, London (2023)
20. Balaji, K., et al.: Cloud container security challenges: a case study. In: 2024 International Symposium on Container Technologies, pp. 211–221. IEEE, New York (2024)
21. Rajyashree, R., et al.: An empirical investigation of docker sockets for privilege escalation and defensive strategies. Procedia Computer Science **233**, 660–669 (2024)
22. Reddy, P.P., et al.: Improvising energy efficiency of heterogeneous servers using AWS services and Machine Learning approaches. 2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT). IEEE (2024)
23. Sah, K.P., et al.: Advancing of Microservices Architecture with Dockers. 2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT). IEEE (2024)
24. Chopra, S., Sreedevi, A.G., Shanmugam, U.: Discernment and Enumeration of Security Vulnerabilities Present in Docker Images. Grenze Int. J. Eng. Technol. (GIJET) 10 (2024)

FBCA-IoMT: A Federated Binary Contrastive Autoencoder Framework for Anomaly Detection

Archita Bhattacharyya[1(✉)], Ayan Bhaumik[2], and Mrinal Kanti Deb Barma[1]

[1] Department of Computer Science and Engineering, National Institute of Technology, Agartala 799046, Tripura, India
archita762000@gmail.com, mrinal@nita.ac.in

[2] Department of Computer Science and Engineering, ICFAI University, Agartala 799210, Tripura, India
connect@ayanbhaumik.in

Abstract. The rapid expansion of the Internet of Medical Things (IoMT), a healthcare-driven subset of the Internet of Things (IoT), has introduced significant cybersecurity threats, underscoring the need for effective and privacy-preserving anomaly detection systems. In this study, we present an anomaly detection framework for IoMT data using autoencoder-based reconstruction loss analysis and feature space visualization. The reconstruction loss distribution enables the identification of anomalous samples using a predefined threshold. In addition, anomaly scores plotted against sample indices help visualize deviations in model behavior, distinguishing normal from suspicious activities. To better understand the latent feature space, the t-SNE visualization provides clear clustering of encoded representations, highlighting the separation between normal and anomalous patterns. This integrated approach offers an interpretable and effective means of detecting anomalies in IoMT environments.

Keywords: IoMT Security · Federated Learning · Anomaly Detection · Autoencoder · Privacy-Preserving

1 Introduction

The Internet of Medical Things (IoMT) enables intelligent healthcare, but faces increasing threats from data anomalies and attacks [1]. To address this, we propose an autoencoder-based anomaly detection framework that achieves a high accuracy of 94.6% with an F1-score of 92.8%. Reconstruction loss effectively differentiates normal and anomalous patterns. Visual insights through anomaly scores and t-SNE plots reveal clear class separability. Our approach ensures robust, interpretable, and privacy-preserving anomaly detection for IoMT systems.

S. Fong et al. (Eds.): ICT4SD 2025, LNNS 1651, pp. 263–272, 2026.
https://doi.org/10.1007/978-3-032-06688-6_26

1.1 Architecture

The proposed method begins by gathering data from dispersed IoMT devices, then uses local preprocessing and a combination of reconstruction and classification loss to train a Binary Contrastive Autoencoder (BCAE). Only model weights are shared between each client and a central server, which updates a global model using Federated Averaging. The redistribution of this privacy-preserving global model improves anomaly detection. Standard measurements and visual aids are used to assess the system in order to guarantee correctness and interpretability shown in Fig. 1

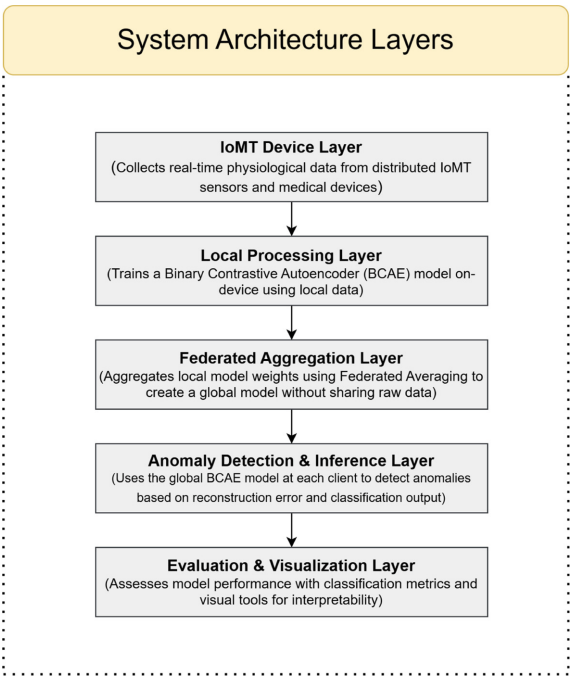


Fig. 1. System Architecture.

1.2 Motivation

The integrity and security of healthcare data have become a critical concern. Traditional anomaly detection methods often fail to handle complex, high-dimensional medical data effectively. This study is driven by the need for a robust, interpretable, and decentralized solution to detect anomalies in real time.

1.3 Contribution

- Proposed a Federated Binary Contrastive Autoencoder (FBCA) for anomaly detection in IoMT, achieving superior performance with 94.6% accuracy and 92.8% F1-score.
- Integrated reconstruction loss and contrastive learning to enhance anomaly discrimination, supported by t-SNE visualization of latent features.
- Designed a privacy-preserving Federated Learning (FL) framework where model training occurs locally on IoMT nodes, thereby avoiding the transmission of raw data.

2 Related Work

The explosive growth of the IoMT has created serious security issues, particularly in identifying irregularities and online threats. To overcome these obstacles, several researchers have suggested autoencoder-based models, FL strategies, and anomaly detection techniques. A variety of Machine Learning (ML) and Deep Learning (DL) methods is used to find irregularities in Internet of Things (IoT) traffic Alsalman *et al.* (2024) [2] created a hybrid IDS that blends conventional security measures with supervised learning. FL has become a viable method for resolving privacy issues and facilitating cooperative learning among dispersed IoT devices. Jena *et al.* (2024) [3] later modified for intrusion detection in IoT networks. FL reduces the possibility of revealing private information by permitting local training on edge devices. FL has drawbacks despite its advantages, including significant communication overhead and susceptibility to hostile attacks. To improve security, Ibrahim *et al.* (2024) [4] suggested integrating differential privacy into FL frameworks. Trade-offs associated with this approach included decreased model performance as a result of added noise. Autoencoders are used extensively in anomaly identification because of their ability to recognize deviations and acquire condensed representations of regular behavior. Dao *et al.* (2021) [5] developed a stacked autoencoder-based IDS for IoT that demonstrated exceptional detection accuracy. But in practical IoT deployments, its scalability is constrained by centralized training and expensive processing requirements. FL has been combined with autoencoders by researchers to overcome privacy problems. For anomaly detection, Shrestha *et al.* (2024) [6] suggested a federated autoencoder system that ensures privacy but has problems with delayed convergence and expensive communication. To improve IoT security, Majeed *et al.* (2024) [7] created a hierarchical FL technique that incorporates autoencoders, improving resilience against adversarial attacks while preserving privacy.

3 Proposed Methodology

FBCA for IoMT networks to detect anomalies. While identifying unusual activity or cyber threats among dispersed medical equipment, the approach aims

to ensure privacy-preserving learning. To maintain local data sovereignty and adhere to privacy laws like HIPAA and GDPR, the system design combines binary classification loss and autoencoding-based reconstruction loss under a FL paradigm.

3.1 System Overview

Multiple edge devices (such as wearable technology, medical sensors, and remote monitors) make up the suggested architecture. Each of these devices uses its private dataset to locally train a Binary Contrastive Autoencoder (BCAE). Federated Averaging is then used to combine these models into a global model that can generalize across all nodes. The architecture reduces privacy risks by guaranteeing that raw patient data never leaves the local node.

3.2 Binary Contrastive Autoencoder (BCAE)

A BCAE forms the basis of the model, which combines unsupervised and supervised learning objectives. It consists of three main components:

- Encoder: Compresses high-dimensional input features into a latent representation.
- Decoder: Attempts to reconstruct the original input from the encoded representation.
- Classifier: Predicts whether the encoded sample is normal or anomalous.

3.3 Model Used: Federated Binary Contrastive Autoencoder (FBCA)

The proposed architecture integrates the following components:

- Autoencoder: Learns a compressed latent representation of the input data via unsupervised learning.
- Binary Classifier (BCE layer): Learns to distinguish between normal and anomalous samples using supervised learning.
- FL Framework: Synchronizes local models across clients using Federated Averaging.

3.4 Justification of Method

- FL ensures data locality and privacy compliance.
- Contrastive Autoencoder learns robust representations for both detection and classification.
- The combined loss improves performance over isolated reconstruction or classification models.
- Applicable to real-world IoMT networks, especially for cybersecurity, fault detection, and health anomaly detection.

3.5 Objective Function

– The training loss function L

L used in the proposed FBCA model is a composite of two components: the reconstruction loss and the classification loss. These are defined as follows:

– Reconstruction Loss

The accuracy of the model's reconstruction of normal samples is guaranteed by the reconstruction loss. Unusual behavior is usually indicated by a large reconstruction error. Using the Mean Squared Error (MSE), it is defined as follows:

$$L_{\text{recon}} = \frac{1}{N} \sum_{i=1}^{N} \|x_i - \hat{x}_i\|_2^2 \tag{1}$$

where x_i is the original input, $\hat{x}_i$ is the reconstructed output from the decoder, and N is the total number of samples.

– Classification Loss (Binary Cross-Entropy(BCE))

The classification loss guides the model to distinguish between normal and anomalous representations. It uses the BCE loss function:

$$L_{\text{class}} = -\frac{1}{N} \sum_{i=1}^{N} [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)] \tag{2}$$

where y_i is the ground truth label and $\hat{y}_i$ is the predicted probability output by the classifier.

– Combined Loss Function

The final loss function used for training is a composite of the reconstruction and classification losses:

$$L = L_{\text{recon}} + L_{\text{class}} \tag{3}$$

3.6 Evaluation Metrics

The performance of the proposed model is evaluated on unseen test data using the following metrics:

– Accuracy

Accuracy measures the proportion of correct predictions among all predictions:

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

where: TP = True Positives, TN = True Negatives, FP = False Positives, FN = False Negatives

– Precision

Precision quantifies the proportion of correctly predicted positive instances among all predicted positives:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{5}$$

– Recall

Recall (or Sensitivity) measures the proportion of actual positive cases that were correctly identified:

$$\text{Recall} = \frac{TP}{TP + FN} \tag{6}$$

– F1-Score

The F1-score is the harmonic mean of Precision and Recall, balancing both metrics:

$$\text{F1-Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{7}$$

– Confusion Matrix

The confusion matrix summarizes the performance of a classification model by displaying the number of TP, TN, FP, FN

– Visual Reconstructions

To qualitatively assess the anomaly detection capability of the model, reconstruction results are visualized. Anomalous samples typically show higher reconstruction error and visibly degraded reconstruction quality compared to normal samples.

3.7 FL Framework

Let there be K clients (IoMT nodes), each possessing its local dataset D_k. The training process follows a federated learning protocol and proceeds in federated communication rounds as outlined below:

1. Each client trains its local Binary Contrastive Autoencoder (BCAE) model using the composite loss function L:

$$L = L_{\text{recon}} + L_{\text{class}}$$

2. After local training, each client k sends its learned model parameters θ_k to a central aggregator.

3. The central server aggregates all local model parameters using **Federated Averaging (FedAvg)** as follows:

$$\theta_G = \frac{1}{K} \sum_{k=1}^{K} \theta_k \tag{8}$$

where θ_G denotes the global model parameters and θ_k is the model parameter set from client k.
4. The updated global model θ_G is then broadcast back to all clients for the next round of training.

The communication cost per round in federated learning is given by: Communication Cost $= 2 \times \text{Size}(\theta)$.

This federated approach ensures that no raw data leaves the local IoMT devices, thus preserving patient privacy and significantly reducing communication overhead. It also enables collaborative model learning across distributed medical devices without centralized data storage.

3.8 Data Collection

- Human Activity Recognition (HAR) dataset [8]
- UNSW-NB15 for network intrusion detection[9]
- Bot-IoT dataset[10]

3.9 Data Pre-processing

To prepare the dataset for training the federated anomaly detection model, the following pre-processing steps were applied:

1. **Missing Value Imputation:** Any missing values in the dataset were imputed using mean or median strategies, depending on the distribution of each feature.
2. **Standardization:** All feature values were standardized using Z-score normalization to ensure uniform scaling:

$$z = \frac{x - \mu}{\sigma} \tag{9}$$

where x is the original feature value, μ is the mean, and σ is the standard deviation of the feature.
3. **Label Encoding:** Binary class labels were assigned for supervised training:
   - 0 for Normal
   - 1 for Anomaly
4. **Data Partitioning:** The dataset was partitioned across K simulated IoMT clients, with each client receiving a non-overlapping subset. Class distribution was balanced across clients to ensure training consistency.

5. **Noise Injection (Optional):** To simulate faulty or adversarial sensor readings, Gaussian noise was optionally injected into a portion of the data:

$$x_{\text{noisy}} = x + \mathcal{N}(0, \sigma^2) \tag{10}$$

where $\mathcal{N}(0, \sigma^2)$ is zero-mean Gaussian noise with variance σ^2.

3.10 Feature Engineering and Selection

- Correlation-Based Filtering: Features with Pearson correlation > 0.95 were removed to reduce redundancy.
- Variance Thresholding: Features with low variance were discarded.
- Principal Component Analysis (PCA) (Exploratory): PCA was used during initial exploration to understand the structure of latent features, although not used directly in training.
- Autoencoder-Based Feature Extraction: The encoder in the Binary Contrastive Autoencoder serves as a learned feature extractor, mapping raw input to a compact and informative latent space.

4 Result and Analysis

Table 1 presents a performance comparison of various baseline models against the proposed Federated Binary Contrastive Autoencoder (FBCA). The results show that FBCA outperforms all other models in all evaluation metrics.

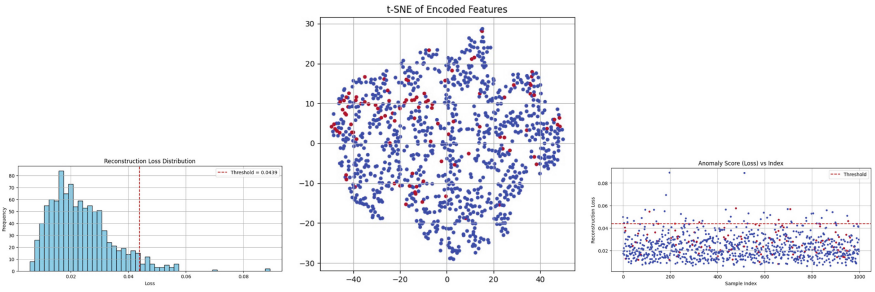
Table 1. Comparison of Different Models' Performance

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC-ROC
Standard Autoencoder (AE)	87.2	84.5	80.9	82.6	0.89
CNN	89.1	87.0	85.5	86.2	0.91
LSTM	88.5	85.2	86.0	85.6	0.90
Random Forest (RF)	90.3	89.1	87.5	88.3	0.92
Support Vector Machine (SVM)	85.4	83.3	79.8	81.5	0.88
Binary Contrastive AE (Local)	91.5	89.7	89.2	89.4	0.93
Federated Binary CA (FBCA)	**94.6**	**93.2**	**92.5**	**92.8**	**0.96**

The results in Table 1 demonstrate that the proposed Federated Binary Contrastive Autoencoder outperforms all other models across all evaluation metrics. Key insights include:

- The federated architecture improves generalization across distributed clients by incorporating diverse data sources without compromising privacy.
- The combination of reconstruction and classification loss leads to better anomaly separation in the latent space.

– Classical machine learning models like RF and SVM show reasonable performance but lack the depth to capture complex nonlinear patterns compared to deep learning methods.
– Non-federated contrastive autoencoders, while effective, underperform in generalization across multiple data domains due to isolated training.



(a) Loss Distribution (b) Encoded Features (c) Anomaly score

Fig. 2. Visual analysis of anomaly detection using reconstruction loss and latent space representations.

FBCA-IoMT's fixed anomaly detection threshold is sensitive, requiring careful tuning based on data distribution, and could benefit from adaptive methods for better real-world performance.

A histogram Fig. 2a showing the frequency distribution of reconstruction losses.

– Most losses are low and concentrated on the left.
– The vertical red dashed line marks the anomaly detection threshold.
– Samples with loss values to the right of the threshold are considered anomalous.

This 2D t-SNE plot Fig. 2b shows how the encoded data points are distributed after dimensionality reduction.

– Red dots denote anomalous points.
– Blue dots represent normal points.
– It illustrates that anomalies (red) are dispersed and do not form distinct clusters, suggesting diverse anomaly characteristics.

This scatter plot Fig. 2c visualizes the reconstruction loss for each sample in the dataset.

– Blue points represent normal data.
– Red points are flagged as anomalies (above the threshold line).
– The red dashed line is the anomaly threshold (0.0439).

– This plot helps identify which samples deviate significantly based on reconstruction error.

FBCA-IoMT's brief mention of communication and convergence needs expansion regarding bandwidth strain from repeated model sharing and convergence issues stemming from non-IID data, client drift, and device heterogeneity.

5 Conclusion

A privacy-preserving FL architecture for anomaly detection in IoMT networks employing a BCAE was presented in this paper. The suggested method efficiently finds anomalies without sending raw data, protecting patient privacy by combining local model training with federated averaging. Results from experiments reveal that the federated BCAE is robust in detecting anomalous patterns in dispersed medical data, outperforming conventional models in accuracy, recall, and AUC-ROC. This architecture provides a safe and scalable method for detecting anomalies in real time across diverse IoMT contexts.

References

1. Agrawal, S., et al.: Federated learning for intrusion detection system: concepts, challenges and future directions. Comput. Commun. **195**, 346–361 (2022)
2. Alsalman, D.: A comparative study of anomaly detection techniques for IoT security using adaptive machine learning for IoT threats. IEEE Access **12**, 14719–14730 (2024)
3. Jena, S.R., et al.: An innovative secure and privacy-preserving federated learning based hybrid deep learning model for intrusion detection in internet-enabled wireless sensor networks. IEEE Trans. Consum. Electron. (2024)
4. Khalaf, O.I., et al.: Federated learning with hybrid differential privacy for secure and reliable cross-IoT platform knowledge sharing. Secur. Priv. **7**(3), e374 (2024)
5. Dao, T.-N., Lee, H.J.: Stacked autoencoder-based probabilistic feature extraction for on-device network intrusion detection. IEEE Internet Things J. **9**(16), 14438–14451 (2021)
6. Shrestha, R., et al.: Anomaly detection based on LSTM and autoencoders using federated learning in smart electric grid. J. Parallel Distrib. Comput. **193**, 104951 (2024)
7. Majeed, R., Sangal, A.: Enhancing IoT security: federated learning with autoencoder model for IoT attacks detection. In: 2024 International Conference on Advances in Modern Age Technologies for Health and Engineering Science (AMATHE), pp. 1–6. IEEE (2024)
8. Kaggle. Human activity recognition (HAR) dataset (2025). https://www.kaggle.com/datasets/mrwellsdavid/unsw-nb15. Accessed Apr 2025
9. Kaggle. Unsw-nb15 for network intrusion detection (2025). https://www.kaggle.com/datasets/dhoogla/unswnb15. Accessed Apr 2025
10. Kaggle. Bot-iot dataset (2025). https://www.kaggle.com/code/khairulislam/unsw-nb15-eda. Accessed Apr 2025

Blockchain Technology: Scalability and Performance

Jeevesh Sharma(✉)

Department of Commerce, Manipal University, Jaipur, Rajasthan, India
jeevesh.sharma@jaipur.maniapl.edu

Abstract. Blockchain technology is largely used in banking, but it also has applications in gaming, real estate, supply chain management, and healthcare. By 2023, digital money will be the most widely discussed blockchain application. The potential uses of blockchain technology concerning various facets of any sector, market, agency, or governmental organization have gained attention in recent years. Blockchain scalability analyzes the effects on the security of scaling blockchain networks to support more transactions per second. This innovative distributed peer-to-peer architecture drew the interest of companies and communities outside and inside the financial sector. Furthermore, the system it operates in has been created around numerous scenarios that address the trust issue in open networks without the requirement for a trustworthy third party. Even though its decentralized structure allows for a wide range of potential applications, scalability remains a hurdle. Function extension, excessive delay in confirmation, and performance inefficiency are three important areas where blockchain scalability has been hindered. This research paper provides a thorough summary of previous research on scalable blockchain systems.

Keywords: Blockchain · Distributed Architecture · Security · Scalability · Blockchain Security

1 Introduction

The term "blockchain" has arisen as a potentially transformational force in a variety of elements of government and business sector activities. Its potential has been recognised globally, with several multinational organisations and technology businesses stressing the benefits of its application in lowering operational and compliance costs while also enhancing efficiency. Blockchain technology has attracted a lot of attention because it can completely transform several sectors' operations by offering decentralized, transparent, and secure platforms [1, 2]. Scalability and performance issues, however, are already becoming vital issues that must be resolved if blockchain adoption is to reach its full potential [3]. Performance refers to how quickly and effectively a system responds, whereas scalability refers to a blockchain network's capacity in handling of transactions per second [4]. The blockchain must be scalable and perform well without sacrificing its security and integrity if it is to be widely used [1, 6].

S. Fong et al. (Eds.): ICT4SD 2025, LNNS 1651, pp. 273–284, 2026.
https://doi.org/10.1007/978-3-032-06688-6_27

Blockchain technology, also known as distributed ledger technologies (DLTs), first gained recognition as applications for cryptocurrencies, primarily Bitcoin [7] and Ethereum [8, 9]. Initially, blockchain didn't gain much importance, but during the last decade, its usage and acceptability have increased. There are two forms of blockchain in general: permission and permissionless. Anyone can join and conduct transactions on the permissionless blockchain. It enables all participants to participate in the chain's consensus process. Bitcoin and Ethereum are classic permissionless blockchain examples. A permission blockchain is a private chain where only authorized users can join and conduct transactions [10]. Since consensus methods need all nodes in the network to process and validate every transaction, blockchain networks such as Bitcoin and Ethereum have experienced transaction throughput and confirmation time restrictions. These constraints become more evident as network size and transaction volume grow, causing congestion, delays, and increased transaction costs. As a result, there is a need to investigate novel ideas and solutions to improve the scalability and performance of blockchain systems while preserving their core security qualities [3].

Improving scalability and performance in blockchain security necessitates tackling several critical issues like storage, throughput, and networking [1, 12]. To begin, the consensus method is critical in determining throughput and transaction confirmation times. Traditional consensus techniques, such as Proof of Work (PoW) [2], consume a lot of resources and limit the quantity of transactions that can be completed in each amount of time. Exploring alternate consensus mechanisms, such as Proof of Stake (PoS), and Practical Byzantine Fault Tolerance (PBFT) can provide scalability benefits by allowing for parallel transaction processing or reducing computational needs [13].

Scalability and performance, however, should not come at the expense of security. Blockchain data is designed to be resistant to change and to serve as a secure transaction record. Spending on blockchain solutions is predicted to increase from $4.5 billion in 2020 to $1.9 billion by 2024. Although the financial sector contributes roughly 30% of the global market value of blockchain in 2020, the technology has extended to nearly every industry, from healthcare to agriculture. In blockchain systems, it is critical to evaluate the potential trade-offs between scalability, performance, and security [6]. To maintain the proper balance, thorough risk assessments, stringent evaluations, and the creation of robust security mechanisms that tackle the unique problems linked to scalability and high-performance blockchain networks are required [2].

Using Blockchain, financial institutions can save about $12 billion every year. Amazing figures were shown with the global spending total, which is predicted to be $17.9 billion by 2024. It will continue to expand at a compounded annual rate of 46.4%. (Source: IDC). Experts anticipate that this technology would increase world GDP by 1.76 trillion by 2030, amounting to 1.4% of global GDP (Source: Price Waterhouse Coopers). The global blockchain market will be valued at $67.4 billion by the end of 2026.

This article has provided an overview of prior research on scalable blockchain systems with adequate depth and breadth. The remainder of the article is constructed as follows. The second section discussed the review of prior pieces of literature. Section three introduces blockchain and its key features. Section four deals with the meaning

of scalability and the major factors impacting scalability. Section five discusses the relevance of scalability and performance of blockchain, and section six presents the key solutions to its key issues analyzed from the point of throughput, storage, and networking. Finally, the article concludes, and future directions are mentioned in sections seven and eight, respectively.

2 Literature Review

To perform the analysis of the scalability, the author has reviewed prior studies related to the research theme and in Table 1 some of the important studies are summarized to present the key findings of the research. These studies and research papers offer valuable insights into the challenges and solutions related to scalability and performance in blockchain security. They also provide a foundation for further exploration and understanding of this important aspect of blockchain technology.

Table 1. Prior studies related to scalability and performance in blockchain.

Authors	Research title	Summary of the study
Eyal et al. (2016) [14]	{Bitcoin-NG}: A scalable blockchain protocol	This paper presents the Bitcoin-NG protocol, which enhances the scalability of blockchain networks by enabling many leaders to propose blocks concurrently, greatly enhancing transaction data.
Wood (2014) [9]	Ethereum: A secure, decentralised, generalised transaction ledger	The Ethereum state transition mechanism, a directed acyclic graph (DAG), and other performance-enhancing features are highlighted in this white paper's description of the Ethereum blockchain's design and architecture.
Thibault et al. (2022) [15]	Blockchain scaling using rollups: A comprehensive survey	The many methods for scaling blockchain networks, such as sharding, sidechains, state channels, and off-chain processing, are covered in this survey study. The scalability issues are discussed, and alternative solutions and trade-offs are investigated.

(*continued*)

Table 1. (*continued*)

Authors	Research title	Summary of the study
Kokoris-Kogias et al. (2018) [16]	Omniledger: A secure, scale-out, decentralized ledger via sharding	The blockchain technology OmniLedger, which is built on sharding and provides great scalability while upholding security and decentralisation, is presented in this paper. To improve performance, it introduces strategies like distributed key generation and group signing.
Bünz et al. (2020) [17]	Zether: Towards privacy in a smart contract world	This paper introduces the Zether protocol, which permits private transactions in Ethereum using zero-knowledge proofs and addresses privacy and scalability in blockchain. It integrates methods from secure multiparty computation, ring signatures, and zero-knowledge proofs.
Buchman (2016) [13]	Tendermint: Byzantine fault tolerance in the age of blockchains	Tendermint, a Byzantine Fault Tolerant consensus technology created for scalability and performance in blockchain networks, is introduced in this paper. It describes the fundamental consensus algorithm and how it is used in real-world blockchain systems.
Pass & Shi (2018) [18]	Thunderella: Blockchains with optimistic instant confirmation	In this paper, Thunderella, a technology for quick transaction confirmation in blockchain networks, is introduced. The hierarchical structure and optimistic consensus are combined to produce low-latency confirmation while retaining security.

(*continued*)

Table 1. (*continued*)

Authors	Research title	Summary of the study
Khan et al. (2021) [10]	Systematic literature review of challenges in blockchain scalability	This systematic literature review provides an overview of various approaches and protocols for achieving interoperability between different blockchain networks. It discusses scalability and performance considerations in the context of blockchain interoperability.

Source: Author's compilation

3 Blockchain

Blockchain technology is a distributed digital ledger that tracks the transactions between each node, or computer, in a network. It was initially introduced behind the cryptocurrency Bitcoin, but it has subsequently found many other applications [5]. A blockchain is fundamentally a chain of blocks, each of which contains a list of transactions. These transactions are compiled and added sequentially and immutably to the blockchain. The integrity and security of the data are ensured once a block is included in the chain, making it extremely difficult to change the data stored there. The key features of the blockchain technology are presented in Table 2.

Table 2. Description of key features of blockchain technology

S. No.	Features	Description
1.	Decentralization	where a central authority controls the data. The blockchain is replicated among all nodes in the network, making it resistant to single points of failure and censorship.
2.	Transparency	Every member of the network has access to and transparency with the blockchain ledger. Every transaction is stored in a block and is publicly verifiable.
3.	Security	Blockchain uses cutting-edge cryptography methods to secure data. The sequence of blocks that comprise a transaction is encrypted.
4.	Immutability	The data contained in a block cannot be easily altered or cancelled once it is added to the blockchain.
5.	Smart Contracts	Smart contracts are an increasingly popular technique for blockchain platforms to provide programmable functionality. The terms of the agreement are directly encoded into the code for these self-executing contracts.

Source: Author's compilation

4 Scalability

The ability of a system, network, or software application to manage a rising quantity of work, resources, or people gracefully and efficiently is referred to as scalability. It is an essential component in designing and constructing systems that can grow and adapt to changing demands without sacrificing performance or usefulness [19]. To achieve scalability, systems are often designed with ideas in consideration, such as loose integration, adaptability, and the flexibility to distribute tasks over many components. It frequently necessitates rigorous design and consideration of variables such as bottlenecks in performance, storing and retrieving information procedures, network connectivity, and allocation of resources.

Scalability is vital because it enables businesses and organizations to develop, meet increasing user needs, and maintain excellent performance levels even during peak usage periods. Companies can prevent downtime, deliver a better user experience, and respond to changing business requirements without substantial interruptions or costly infrastructure overhauls by developing scalable systems. Blockchain, as one of the fundamental technologies of distributed ledgers, overcomes the trust problem in open networks by eliminating the need for a trusted third party. Its decentralized characteristic has a wide range of application possibilities; however, it still has scaling issues [6]. Since Bitcoin's dominance in cryptocurrency, blockchain scalability difficulties have been identified. [20] examined various critical measures to determine Bitcoin's scalability, which include maximum throughput, latency, activation time, and cost per confirmed transaction (CPCT). The two most essential performance parameters that have a substantial impact on the user's quality of experience (QoE) are highest throughput and latency [4].

4.1 Factors Affecting Blockchain Scalability

With the mounting recognition of cryptocurrencies like Bitcoin and Ethereum, they are being forced to consider a key weakness in their initial structure: a dearth of scalability. As cryptocurrencies become more widely accepted, the number of transactions is rising exponentially as cryptocurrencies gain acceptance [21]. Scalable blockchain enables a network to manage an increasing number of transactions or users without experiencing performance concerns or deterioration, but some factors also impact the scalability of the blockchain [11, 22]. The following are some factors:

a) **Block size:** Scalability is affected by the size of each block in a blockchain. Smaller block sizes limit the occurrence of transactions, which can contribute to bottlenecks and delayed processing of transactions. Increased block size allows more transactions to be completed in each block, but it also increases storage and bandwidth needs for network participants.
b) **Block time:** Scalability may be impacted by how long it takes to generate a new block. Longer block periods lengthen the period needed for transactions to be confirmed, slowing the system's throughput. Shorter block periods can speed up transaction processing.
c) **Consensus mechanism:** Scalability is impacted by a blockchain's chosen consensus mechanism. Transaction processing is slowed down by the requirements of some consensus techniques, such as PoW. Other consensus algorithms, like PoS or Delegated

Proof of Stake (DPoS), can validate transactions more quickly, but they may have restrictions on the number of participating nodes or raise issues with centralization.

d) **Network bandwidth:** A blockchain network's scalability is affected by the available network bandwidth. Higher bandwidth enables quicker transaction and block dissemination among network nodes. Inadequate bandwidth can cause latency and higher traffic on the network, reducing overall transaction throughput.
e) **Interoperability and Interchain Communication:** Interoperability and effective communication among blockchains can influence scalability. As additional blockchains develop and the demand for cross-chain trades grows, the ability to effortlessly transfer assets and data between various chains gets critical for overall scalability.
f) **Layer 2 solution:** Payment channels, sidechains, and state channels, for example, can improve scalability by minimizing the pressure on the main blockchain. These solutions allow for off-chain transactions or computations, lowering the burden on the main chain and enhancing overall system scalability.

5 Scalability and Performance of Blockchain

In the context of blockchain, scalability means the network's capacity to govern an expanding volume of transactions or users without compromising effectiveness or performance. It deals with the issue of keeping the network's capacity, responsiveness, and speed as transaction demand increases. In other words, scalability is the capacity of a blockchain network to support more nodes and transactions [1]. Therefore, scalability is a significant hurdle to implementing blockchain in real-world corporate situations. In this part, the researcher looks at scalability from the standpoints of throughput, storage, and networking [12]. The description of each is presented in Table 3. Whereas performance focuses on the blockchain network's efficiency and responsiveness in terms of transaction speed, verification times, and user experience. A high-performance blockchain network processes transactions rapidly and with low latency, enabling that activity to run smoothly and on time.

The performance of blockchain networks can be assessed by the average time taken for verification and storage of transactions in each peer node such that it can't be reversed or revoked. Many techniques can be employed to achieve blockchain scalability and better performance together. Consensus mechanism, network architecture, blockchain protocol optimization, Off-chain solutions, sharding, hardware optimization, and continuous development and optimization. The consensus method employed in a blockchain is vital in determining scalability and performance. Due to the time-consuming nature of mining, [7] some consensus algorithms, such as PoW, have intrinsic restrictions in terms of transaction throughput. Alternative consensus algorithms such as PoS, DPoS, and PBFT seek to increase scalability by lowering computational requirements and increasing processing of transactions speed [2, 11, 21, 23, 24]. Moreover, optimizing the underlying architecture and design of blockchain networks can help with scalability and performance. Techniques such as sharding, which divides the blockchain into

smaller parts or shards that may execute transactions separately, can increase throughput and decrease delay [3, 25]. Off-chain solutions, like state channels or sidechains, allow certain transactions to be executed outside of the main blockchain, lowering network load, and improving performance [3]. Additionally, advancements in cryptography algorithms and privacy-preserving mechanisms can help to improve scalability and performance. Privacy-enhancing techniques such as zero-knowledge proofs, ring signatures, and secure multiparty computing are examples of approaches that can minimize the computational overhead associated with transaction verification, allowing for faster processing times and increased scalability requirements.

6 Solution to Blockchain Scalability

There are numerous efforts and suggestions aimed at enhancing blockchain scalability. It is challenging to overcome scaling issues with blockchains without compromising their security, decentralization, or credibility. Since, the blockchain trilemma covers the difficulties that developers confront in establishing a blockchain that is scalable, decentralized, and secure. According to the trilemma, it is difficult to accomplish all three of these characteristics at the same time because they frequently compete with one another [10, 15, 23] (as illustrated in Fig. 1). For instance,

- A blockchain network may become less scalable if it emphasizes decentralization and security. Decentralization frequently necessitates the use of many nodes to validate and store each transaction, which slows processing and limits the volume of transactions that the network can handle.
- If a blockchain network prioritizes scalability and security, decentralization may suffer. To achieve great scalability, certain blockchain systems may compromise decentralization by employing tactics such as sharding or depending on a fewer number of trusted validators.
- If a blockchain network that prioritizes decentralization and scalability may confront security issues. With a higher number of participants and increased scalability, maintaining network security and reaching consensus among a diverse group of actors becomes more difficult.

As a result, balancing or even achieving these three elements of the distributed ledger system is vital for the foreseeable growth of the blockchain.

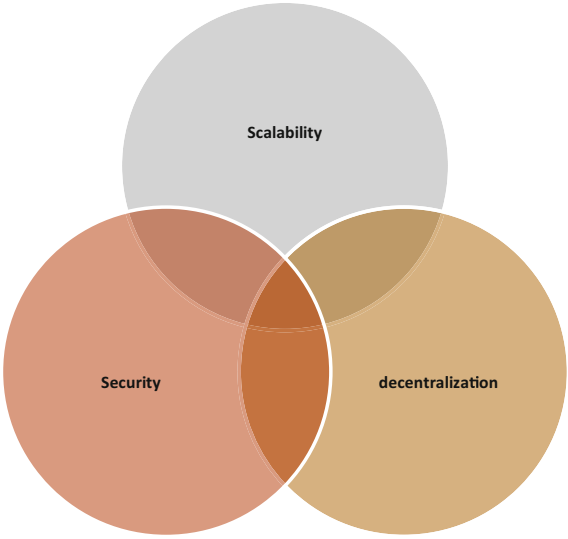


Fig. 1. Blockchain Scalability Trilemma

Major blockchain initiatives' efforts to achieve scalability and enhance transaction throughput and convergence are mostly focused on simpler and more efficient consensus mechanisms. Through referring to prior literature [3, 12, 23], Table 3 shows the key issues and enabling technologies to boost these scalability issues. These methods aim to improve the overall scalability of blockchain technology by increasing transaction throughput, decreasing validation times, and reducing verification times.

Table 3. Description of scalability-related key issues and technological solutions

Issues	Description	Enabling technologies
Throughput	The quantity of items flowing through a system or process is referred to as throughput. Currently, the Bitcoin blockchain's throughput is limited to around seven transactions per second. In contrast, the VISA can hold an average of 2,000 transactions per second. Resultant of this, proper mechanisms must be carefully structured to extend throughput to hold an extensive number of real-world transactions.	1. Increasing block volume 2. Reducing transaction amount 3. Reducing transaction processing 4. Off-chain transaction 5. Sharding

(*continued*)

Table 3. (*continued*)

Issues	Description	Enabling technologies
Storage	Storage is concerned with the produced records. Traditional blockchain systems require each node to complete and retain the full transaction before it can be returned to the original node. As a result, if servers hold inadequate space and computing capacity, blockchain cannot be effectively applied to real-world business circumstances.	Merging of blockchain with existing distributed storage systems capable of off-chain storage of an extensive amount of data. For eg. BigchainDB, InterPlanetary File System (IPFS), and Distributed Hash Table (DHT).
Networking	Networking refers to data transmission. In conventional blockchain systems, every transaction is transmitted to all nodes twice: once when it occurs and a second time when a block holding it is mined. This procedure lengthens block transmission time and requires many network resources. Therefore, more effective designs are needed.	Adopting effective data transmission methods like Recursive Inter-Network Architecture (RINA), GeeqChain, and FIBRE.

Source: Author's compilation

7 Conclusion

Scalability and performance are crucial factors in blockchain security. Addressing these issues is becoming more and more crucial as blockchain technology is adopted widely. To create scalable, high-performance blockchain systems without sacrificing security and integrity by investigating novel consensus processes, improving network architecture, utilizing cryptographic approaches, and carefully regulating the trade-offs. Major blockchain initiatives' efforts to achieve scalability and enhance transaction throughput and convergence are mostly focused on simpler and more efficient consensus mechanisms. In this article, the researcher analyzed the scalability and performance of blockchain technology and found the key issues related to scalability which are storage, throughput, and networking. In this study, it has been observed that scalability is affected by several factors, block size, block time, consensus mechanism, network bandwidth, interoperability and interchain communication, and layer 2 solutions. Among these factors, the majority emphasized factor is 'transaction throughput', which is strongly tied to a consensus technique. Hence, balancing scalability, security, and decentralization is essential to create a blockchain network that can handle increased demand while maintaining high performance and user satisfaction.

8 Blockchain Scalability: New Horizons and Future Research Agenda

The numerous advantages of blockchain technology will undoubtedly encourage organizations and enterprises all over the world to position themselves ahead of their current situation. It is still in its early stages, yet this is one of the most current technologies. Between 2022 and 2030, the worldwide blockchain technology sector is predicted to develop at a CAGR of 85.9%. Among the numerous advantages of blockchain technology, the performance and scalability of Blochian technology are still at a nascent stage. Although there are many studies related to the scalability and performance of blockchain technology, it still needs the attention of academics. Furthermore, blockchain technology is broad and scalable enough to be employed in data access and sharing, data reconciliation, identity protection, secure payments, track-and-trace, asset protection, asset transfer, certification, and record reconciliation.

References

1. Pandey, A.A., Fernandez, T.F., Bansal, R., Tyagi, A.K.: Maintaining scalability in blockchain. In: International Conference on Intelligent Systems Design and Applications, pp. 34–45. Springer International Publishing, Cham (2021)
2. Eklund, P.W., Beck, R.: Factors that impact blockchain scalability. In: Proceedings of the 11th international conference on management of digital ecosystems, pp. 126–133 (2019)
3. Zhou, Q., Huang, H., Zheng, Z., Bian, J.: Solutions to scalability of blockchain: a survey. Ieee Access **8**, 16440–16455 (2020)
4. Kuzlu, M., Pipattanasomporn, M., Gurses, L., Rahman, S.: Performance analysis of a hyperledger fabric blockchain framework: throughput, latency and scalability. In: 2019 IEEE international conference on blockchain (Blockchain), pp. 536–540. IEEE (2019)
5. Yang, W., Garg, S., Raza, A., Herbert, D., Kang, B.: Blockchain: trends and future. In: Knowledge Management and Acquisition for Intelligent Systems: 15th Pacific Rim Knowledge Acquisition Workshop, PKAW 2018, Nanjing, China, August 28–29, 2018, Proceedings 15, pp. 201–210. Springer International Publishing (2018)
6. Yang, D., Long, C., Xu, H., Peng, S.: A review on scalability of blockchain. In: Proceedings of the 2020 the 2nd International Conference on Blockchain Technology, pp. 1–6 (2020)
7. Nakamoto, S.: Bitcoin: A peer-to-peer electronic cash system. Decentralized business review (2008)
8. Buterin, V.: A next-generation smart contract and decentralized application platform. white paper **3**(37), 2–1 (2014)
9. Wood, G.: Ethereum: a secure decentralised generalised transaction ledger. Ethereum Project Yellow Paper **151**(2014), 1–32 (2014)
10. Khan, D., Jung, L.T., Hashmani, M.A.: Systematic literature review of challenges in blockchain scalability. Appl. Sci. **11**(20), 9372 (2021)
11. Khan, K.M., Arshad, J., Khan, M.M., Nasir, M.H.: An empirical investigation of blockchain scalability. Trust Models for Next-Generation Blockchain Ecosystems, 105–133 (2021)
12. Xie, J., Yu, F.R., Huang, T., Xie, R., Liu, J., Liu, Y.: A survey on the scalability of blockchain systems. IEEE Network **33**(5), 166–173 (2019)
13. Buchman, E.: Tendermint: Byzantine fault tolerance in the age of blockchains, Doctoral dissertation. University of Guelph (2016)

14. Eyal, I., Gencer, A.E., Sirer, E.G., Van Renesse, R.: Bitcoin-NG: A scalable blockchain protocol. In: 13th USENIX symposium on networked systems design and implementation (NSDI 16), pp. 45–59 (2016)
15. Thibault, L.T., Sarry, T., Hafid, A.S.: Blockchain scaling using rollups: a comprehensive survey. IEEE Access **10**, 93039–93054 (2022)
16. Kokoris-Kogias, E., et al.: Omniledger: a secure, scale-out, decentralized ledger via sharding. In: 2018 IEEE symposium on security and privacy (SP), pp. 583–598. IEEE (2018)
17. Bünz, B., Agrawal, S., Zamani, M., Boneh, D.: Zether: towards privacy in a smart contract world. In: International Conference on Financial Cryptography and Data Security, pp. 423–443. Springer International Publishing, Cham (2020)
18. Pass, R., Shi, E.: Thunderella: Blockchains with optimistic instant confirmation. In: Advances in Cryptology–EUROCRYPT 2018: 37th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Tel Aviv, Israel, April 29-May 3, 2018, Proceedings, Part II 37, pp. 3–33. Springer International Publishing (2018)
19. Kohad, H., Kumar, S., Ambhaikar, A.: Scalability issues of blockchain technology. Int. J. Eng. Adv. Technol. **9**(3), 2385–2391 (2020)
20. Croman, K., et al.: On Scaling Decentralized Blockchains: (A Position Paper). In: International conference on financial cryptography and data security, pp. 106–125. Springer Berlin Heidelberg, Berlin, Heidelberg (2016)
21. Chauhan, A., Malviya, O.P., Verma, M., Mor, T.S.: Blockchain and scalability. In: 2018 IEEE international conference on software quality, reliability, and security companion (QRS-C), pp. 122–128. IEEE (2018)
22. Sohrabi, N., Tari, Z.: On the scalability of blockchain systems. In: 2020 IEEE International Conference on Cloud Engineering (IC2E), pp. 124–133. IEEE (2020)
23. Aiyar, K., Halgamuge, M.N., Mohammad, A.: Probability distribution model to analyze the trade-off between scalability and security of sharding-based blockchain networks. In: 2021 IEEE 18th Annual Consumer Communications & Networking Conference (CCNC), pp. 1–6. IEEE (2021)
24. Sanka, A.I., Cheung, R.C.: A systematic review of blockchain scalability: Issues, solutions, analysis and future research. J. Netw. Comput. Appl. **195**, 103232 (2021)
25. Khacef, K., Benbernou, S., Ouziri, M., Younas, M.: A Dynamic Sharding Model Aware Security and Scalability in Blockchain. Information Systems Frontiers, 1–14 (2023)

Fixation-Guided Recognition and Categorization of Handwritten Characters

Judy K. George(✉) and Elizabeth Sherly

Digital University Kerala, Thiruvananthapuram 695317, India
{judy.res21,sherly}@duk.ac.in

Abstract. Convolutional Neural Networks are extensively employed in critical domains such as computer vision, medical imaging, and autonomous systems. Enhancing model interpretability by providing users with concise and context-relevant explanations of CNN decision making such as visualizing feature maps or saliency regions, enables a deeper understanding of the model's internal representations and inference process. The proposed work presents a deep learning framework integrating a ResNet-based U-Net architecture with a Fixation Point Generator (FPG) to perform classification and saliency aware reconstruction on the hand-written dataset. The model leverages transfer learning by employing a pre-trained ResNet-18 as the encoder backbone, enabling robust feature extraction. A custom decoder reconstructs input images while a classification head predicts digit labels. To enhance model interpretability, a Fixation Point Generator predicts spatial attention maps (saliency maps) from high-level global features, highlighting regions of interest that influence model decisions. This implementation aims to bridge the gap between classification performance and model explainability, offering insights into the model's focus areas through learned attention. The model got an accuracy of 98.44 on the Malayalam handwritten dataset, 97.81 on English handwritten dataset, and 99.56 on the MNIST dataset.

Keywords: Attention · Character recognition · Classification · Visual saliency

1 Introduction

The human visual system is constantly bombarded with vast amounts of sensory input, yet it processes only a fraction of this data due to inherent limitations in cognitive and perceptual capacity [1]. To operate efficiently, it selectively filters out irrelevant information and concentrates processing resources on task-relevant stimuli. High-resolution perception is confined to a narrow region at the center of the visual field (fovea), while peripheral vision provides only coarse-grained information. To compensate, the visual system employs a series of eye movements, known as fixations, to sequentially sample regions of interest, thereby constructing a comprehensive understanding of a scene in a task-driven manner [2].

Extensive psychological research has demonstrated that fixation behavior is strongly influenced by the observer's task [3]. During activities such as reading or object

S. Fong et al. (Eds.): ICT4SD 2025, LNNS 1651, pp. 285–294, 2026.
https://doi.org/10.1007/978-3-032-06688-6_28

recognition, fixations are strategically guided to extract minimal but sufficient high-resolution detail necessary for effective task execution, thus optimizing cognitive effort and efficiency.

Motivated by these biological insights, computational models have been developed to emulate visual attention mechanisms [4]. Early approaches predominantly relied on rule-based algorithms that identified salient regions based on predefined low-level visual features [5–10]. However, such models lacked adaptability and were not optimized for specific downstream tasks. In contrast, recent advances leverage deep learning techniques to model attention and predict fixation patterns [11]. While these models have made progress in approximating human gaze behavior, they are primarily focused on attention simulation rather than on enhancing task performance through selective visual processing.

In this work we proposed a vision model that learns, sees, explains, and reconstructs, all in one pass. It is a multi-task model which opens to several real world applications across different domains such as medical imaging, anomaly detection, handwritten document analysis, security and surveillance etc. For interpretability, we have considered the handwritten characters which possesses a rich and complex script characters with high inter-class similarity, varied writing styles, and the presence of compound glyphs. This level of transparency is essential for promoting user trust, validating model behavior, and ensuring accountability in high-stakes applications.

In this paper, Sect. 2 explores a range of deep learning approaches for modeling attention. Section 3 presents a comprehensive overview of a multi-task architecture that integrates classification, reconstruction, and attention prediction with Fixation Point Generator (FPG). Section 4 offers both qualitative and quantitative analyses across multiple datasets, while Sect. 5 concludes the study with key insights and takeaways.

2 Related Works

Recent advances in computer vision have led to the development of hybrid neural architectures for handwritten character recognition. Several works have leveraged Convolutional Neural Networks (CNNs) as robust feature extractors for character classification tasks. These models excel at capturing hierarchical spatial features and have been extended with upsampling modules for tasks involving image reconstruction or segmentation [12]. Convolutional Neural Networks (CNNs) have become the backbone of modern systems due to their ability to automatically extract robust spatial features from raw images, replacing traditional hand-crafted methods. Researchers have also explored Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) [13] units for recognizing sequential patterns in word-level tasks where character dependencies matter. Hybrid architectures, combining CNNs and RNNs or attention mechanisms, have improved contextual understanding in script recognition. Moreover, transfer learning using powerful pre-trained models such as ResNet50 and EfficientNet has shown strong results when fine-tuned on relatively small Malayalam datasets [14].

Convolutional Neural Networks (CNNs), are often considered black boxes due to their complex and non-linear internal computations. While CNNs achieve remarkable accuracy in Handwritten Character Recognition (HCR), understanding why a model

made a certain prediction is crucial in sensitive or high-stakes applications such as automated document digitization, educational assessments, and forensic handwriting analysis. Grad-CAM provides insight into model decisions by highlighting class-specific discriminative regions in the input image [15]. In parallel, biologically inspired mechanisms such as Fixation Point Generators (FPG) have been explored to simulate human-like visual attention. These modules generate saliency maps conditioned on global feature embeddings, allowing models to learn where to "look" when interpreting characters. Integrating such saliency-based attention with deep classification networks has shown to improve both performance and interpretability [16].

In the proposed methodology, reconstruction and classification objectives are jointly optimized within a multi-task learning framework. The reconstruction loss guides the encoder to preserve fine-grained visual information, enriching the learned feature representations and subsequently enhancing classification performance. Additionally, the model integrates saliency maps and Grad-CAM-based visualizations to facilitate spatial attention learning and provide interpretable insights into the model's decision-making process, highlighting the most relevant regions contributing to each prediction.

3 Methodology

This work proposes a hybrid deep learning architecture integrating a ResNet-based encoder-decoder network with a Fixation Point Generator(FPG) and saliency-guided attention using Grad-CAM. Figure 1 shows the proposed model for model which is a hybrid of classification, image reconstruction, and saliency-based interpretation.

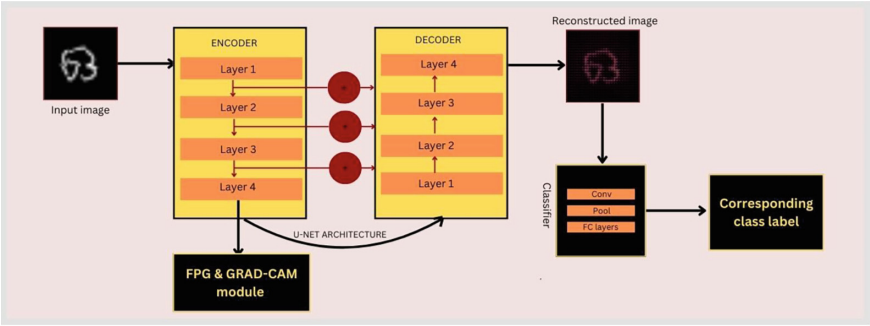


Fig. 1. Illustrates the proposed multi-task architecture, which integrates a ResNet-based encoder for feature extraction, a U-Net-style decoder for image reconstruction, and a Fixation Point Generator (FPG) for predicting spatial attention maps.

3.1 Data Preprocessing

The dataset used is the handwritten character dataset, formatted in IDX file format. Images are first read using a custom IDX parser, normalized, and augmented with geometric transformations including random rotations and translations. To adapt grayscale

MNIST images to the RGB format required by ResNet, each image is converted to a 3-channel format. All images are resized to 112 × 112 pixels and normalized using a standard normalization strategy across the three-color channels. To enhance generalization, data augmentation techniques including affine transformations and rotation, as well as channel normalization expand the effective training set and improve robustness to real-world variations in handwriting.

3.2 Model Architecture

The model consists of the following key components:

Backbone Network (ResNet-18) The model utilizes a pretrained ResNet-18 as an encoder for high-level feature extraction. The original fully connected classification head is removed, and intermediate feature maps from the encoder layers (layer1through layer4) are used for constructing a U-Net-like decoder.

U-Net Style Decoder The decoder is built using transposed convolutional layers for up-sampling and skip connections for concatenating encoder features at corresponding resolutions. This helps to recover spatial information lost during down-sampling. The final output of the decoder is passed through a sigmoid-activated convolutional layer to reconstruct a 3-channel image.

Fixation Point Generator (FPG) The FPG module is a lightweight attention mechanism that takes the global average pooled features (512-dimensional vector) from the deepest encoder layer and generates a spatial saliency map using a two-layer MLP. The output saliency is reshaped into a 28 × 28 map and normalized via softmax, highlighting regions of interest potentially analogous to human visual fixation points.

Classifier Head Reconstructed images are passed through a small CNN-based classifier with two convolutional layers and an adaptive average pooling layer, followed by a fully connected layer to predict the digit class. The classifier is designed to operate on reconstructed images, enforcing the decoder to maintain discriminative features.

Grad-CAM for Visual Explanations A Grad-CAM module is integrated using forward and backward hooks on the final encoder layer (enc4). During backpropagation, gradients are captured to compute channel-wise importance weights, which are then used to generate class-discriminative activation maps. The resulting heatmaps are overlaid on the original input to visualize which regions influenced the network's prediction.

3.3 Loss Function

To simultaneously optimize classification accuracy and reconstruction quality, a multi-task learning framework is employed. This approach utilizes a combined objective function that integrates a supervised classification loss with an unsupervised image reconstruction loss. This formulation encourages the model to learn representations that are both discriminative and generative, which is particularly beneficial for robust representation learning.

Given an input image $x \in \mathbb{R}^{C\times H\times W}$ and its associated ground truth label $y \in \{0, 1,\ldots,$ K − 1$\}$. The model produces two outputs: a reconstructed image $\hat{x} \in \mathbb{R}^{C\times H\times W}$ and a vector of unnormalized class logits $z \in \mathbb{R}^{K}$. The total loss function is defined as:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{cls}} + \alpha \cdot \mathcal{L}_{\text{recon}} \tag{1}$$

where $\alpha \in \mathbb{R}_{\geq 0}$ is a weighting coefficient that balances the importance of the reconstruction loss relative to the classification loss.

Classification Loss For the classification objective, we use the cross-entropy loss over the predicted class logits:

$$\mathcal{L}_{cls} = \sum_{k=1}^{K} \mathbb{1}_{[y=k]} \log\left(\frac{\exp(zk)}{\sum_{j=1}^{K}\exp(z_j)}\right) \tag{2}$$

where $\mathbb{1}_{[y=k]}$ is an indicator function that evaluates to 1 if y = k and 0 otherwise.

Reconstruction Loss To enforce generative consistency, we minimize the mean squared error (MSE) between the input image and its reconstruction:

$$\mathcal{L}_{\text{recon}} = \hat{x} - x_2^2 = \sum_{c=1}^{C}\sum_{i=1}^{H}\sum_{j=1}^{W}(\hat{x}_{c,i,j} - x_{c,i,j}) \tag{3}$$

This term encourages the model to preserve low-level information necessary to accurately reconstruct the input image, leading to better feature learning in the shared encoder. The hyperparameter α can be tuned depending on the task's requirements. A larger α emphasizes reconstruction fidelity (important in unsupervised or semisupervised settings), while a smaller α prioritizes classification performance. By jointly minimizing both terms, the model learns feature representations that are both task-relevant and information-preserving.

4 Results

4.1 Datasets

The proposed model was trained and evaluated using a combination of publicly available datasets, including handwritten characters for both Malayalam and English scripts obtained from Kaggle, as well as the standard MNIST dataset. The Malayalam dataset consists of approximately 6,000 images distributed across 48 classes, each representing vowels and consonants, with around 130 images per class. The English dataset includes 26 classes corresponding to uppercase letters, with each class containing about 55 images. The MNIST dataset comprises 70,000 grayscale images of handwritten digits ranging from 0 to 9, each paired with its corresponding numeric label. This diverse dataset selection ensured a comprehensive assessment of the model's performance across different languages and writing styles. The implementation was carried out on an NVIDIA DGX A100 server.

4.2 Quantitative and Qualitative Analysis

Table 1 shows the quantitative analysis of the proposed model on the training and testing data. The MNIST dataset required fewer training epochs to achieve high accuracy compared to the English and Malayalam handwritten character datasets. This can be attributed to its lower complexity, limited class count, uniform character representation, and well-curated nature. In contrast, the English and Malayalam datasets involve a larger number of classes, higher intra-class variability, and less standardized character structures, all of which demand more extensive learning for effective generalization.

Table 1. Training and Testing Accuracy of the Proposed Model on Different Datasets.

Dataset	Training Accuracy (%)	Testing Accuracy (%)
MNIST dataset	99.56	100.00
English handwritten dataset	97.81	92.81
Malayalam handwritten dataset	99.21	98.44

Table 2 demonstrated an overview of F1 score on different datasets considered. Despite high class diversity and visual similarity, the Malayalam script demonstrates robust generalization on complex characters. The English alphabet yields consistent performance, attributed to its simpler structure and fewer classes, while the MNIST dataset achieves exceptional precision and recall under ideal conditions for handwritten digit recognition.

Table 2. Comparative Overview of F1 Score Performance Across Datasets.

Attribute	Malayalam	English	MNIST
Script Type	Complex	Latin	Numeric
Number of Classes	48	26	10
Minimum F1 Score	∼0.830	∼0.842	∼0.988
Maximum F1 Score	1.000	1.000	1.000
F1 Score Variability	Moderate	Low	Very Low
Class Confusability	High	Medium	Low
Overall Model Fit	Robust	Consistent	Near-perfect

Figure 2 depicts the analysis of two digit samples from MNIST, both representing variations of the digit'4'. It reveals the complementary strengths of interpretability techniques like Grad-CAM, FPG Saliency, and saliency-based reconstruction. In the first sample, the digit is stylized with a slanted form, yet both Grad-CAM and FPG Saliency accurately highlight key structural elements such as the central intersection and horizontal bar, indicating the model's focus on defining features despite stylistic variation.

FPG Saliency, in particular, offers finer granularity, emphasizing specific stroke edges and contours that contribute most to prediction. The reconstructed version retains the digit's core shape, demonstrating that only a sparse subset of salient pixels is sufficient for recognition. In the second sample, a more standard'4', Grad-CAM again emphasizes the bottom-right junction point, a reliable anchor for classification. FPG Saliency provides a sharper outline of the full digit, with greater emphasis on the lower structure. The reconstructed image closely resembles the original, confirming that the model identifies and prioritizes structurally significant pixels while disregarding irrelevant background information. Together, these interpretations confirm the model's robustness in recognizing diverse handwriting styles using key visual cues and sparse yet informative input.

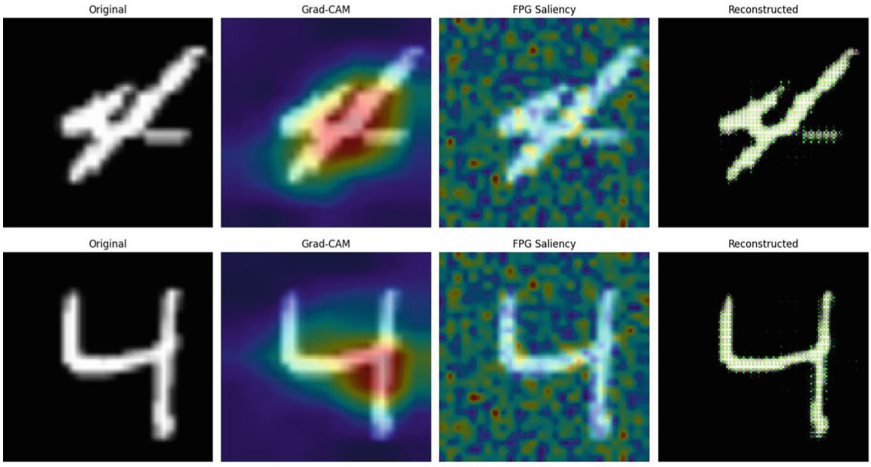


Fig. 2. Interpretability Analysis of Handwritten Digit '4' from Fixation-Guided Recognition and Categorization of Handwritten CharactersMNIST dataset

Similarly Fig. 3 presents a comparative visualization of two handwritten samples of the english character'H', demonstrating the model's interpretability and reconstruction capability. In both samples, the Grad-CAM visualizations highlight the central vertical stroke and side limbs of the character, indicating that the model is correctly focusing on the most class-discriminative features for classification. The FPG saliency maps offer a more distributed and texture-aware attention pattern, capturing the entire character structure in a manner that resembles human visual fixation. This reinforces the model's ability to learn spatially meaningful attention beyond just class prediction.

Figure 4 shows Grad-CAM heatmaps, FPG saliency maps, and saliency-based reconstructions of independent vowel, pure consonant and half-consonant of Malayalam language. Grad-CAM highlights key class-discriminative regions, while FPG Saliency provides finer-grained attention across stroke boundaries. The reconstructed outputs demonstrate the model's ability to retain essential digit structure using only sparse and informative regions.

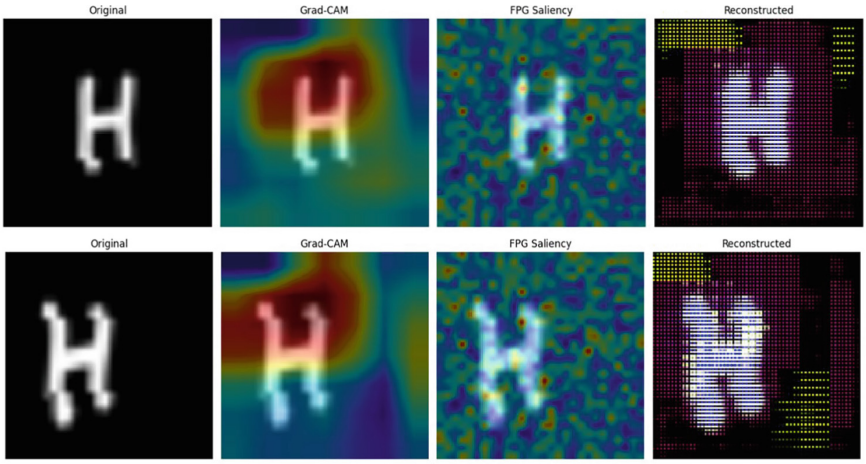


Fig. 3. Interpretability Analysis of Handwritten Character 'H' from English handwritten dataset

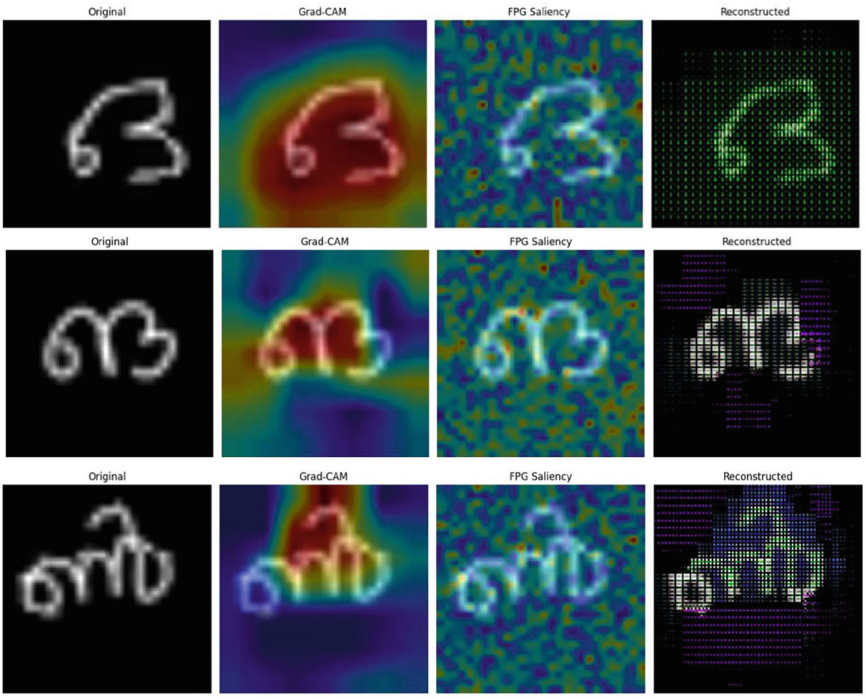


Fig. 4. Interpretability Analysis of Handwritten vowels and consonants from malayalam handwritten dataset

Figure 5 illustrates the model's training dynamics over 100 epochs for malayalam dataset. The training loss (left) shows a rapid decline initially, indicating effective learning, and gradually stabilizes, suggesting convergence. The accuracy plot (right) demonstrates a steady increase in both training and validation accuracy, with minimal gap between them, highlighting good generalization. The model exhibits stable training behavior with no significant signs of overfitting.

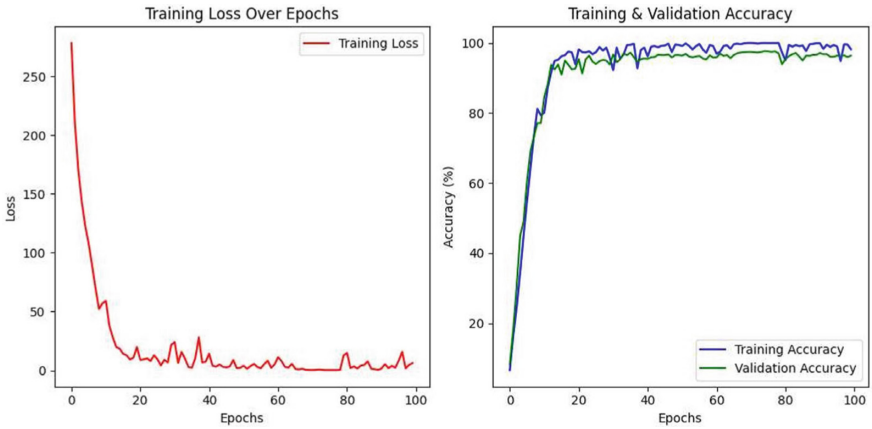


Fig. 5. Model Performance: Tracking Loss and Accuracy Trends over epochs

5 Conclusion

The proposed model presents a unified deep learning framework that combines classification, reconstruction, and interpretability using a modified ResNet-based U-Net architecture enhanced with a Fixation Point Generator (FPG) and Grad-CAM visualization. Trained on the handwritten dataset, the model effectively learns to classify digits while simultaneously reconstructing input images and generating saliency maps that highlight important regions influencing predictions. The use of a combined loss function ensures the model balances classification accuracy with reconstruction quality, and the integration of FPG and Grad-CAM offers valuable insight into model decision- making. The model shows strong performance across all three handwritten character datasets, demonstrating its robustness and reliability. It also improves interpretability by highlighting the regions of the image that influence its predictions, allowing us to understand and visualize how decisions are made. This makes the system both accu-rate and easier to analyze, supporting more transparent and explainable handwriting recognition. Future improvements could involve adding attention mechanisms to help the model better identify and focus on important regions of multidigit images, cursive letters and natural images that are relevant to the visual search task.

References

1. Borji, A., Itti, L.: State-of-the-art in visual attention modeling. IEEE Trans. Pattern Anal. Mach. Intell. **35**(1), 185–207 (2012)

2. Wang, W., et al.: Simulating human saccadic scanpaths on natural images. In: CVPR 2011, pp. 441–448. IEEE (2011)
3. Yarbus, A.L. (ed.): Eye Movements and Vision. Springer, New York (2013)
4. Gide, M.S., Karam, L.J., et al.: Computational visual attention models. Foundations and Trends® in Signal Processing **10**(4), 347–427 (2017)
5. Koch: Shifts in selective visual attention: towards the underlying neural circuitry. In: Christof, Ullman, S. (eds.) Matters of Intelligence, pp. 115–141. Springer, Berlin (1987)
6. Itti, L., Koch, C., Niebur, E.: A model of saliency-based visual attention for rapid scene analysis. IEEE Trans. Pattern Anal. Mach. Intell. **20**(11), 1254–1259 (1998)
7. Harel, J., Koch, C., Perona, P.: Graph-based visual saliency. Adv. Neural Info. Proc. Sys. 19 (2006)
8. Bruce, N., Tsotsos, J.: Saliency based on information maximization. Adv. Neural Info. Proc. Sys. 18 (2005)
9. Hou, X., Zhang, L.: Saliency detection: a spectral residual approach. In: 2007 IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–8. Ieee (2007)
10. Zhang, J., Sclaroff, S.: Saliency detection: a boolean map approach. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 153–160 (2013)
11. Ullah, I., et al.: A brief survey of visual saliency detection. Multimedia Tools and Appl. **79**(45), 34605–34645 (2020)
12. Ronneberger, O., Fischer, P., Brox, T.: U-net: convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-assisted intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015, Proceedings, Part III 18, pp. 234–241. Springer (2015)
13. Paul, I., Sasirekha, S., Vishnu, D.R., Surya, K.: Recognition of handwritten text using long short term memory (lstm) recurrent neural network (rnn). In: AIP Conference Proceedings, vol. 2095. AIP Publishing (2019)
14. Pearlsy, P., Sankar, D.: Malayalam handwritten character recognition using transfer learning and fine tuning of deep convolutional neural networks. In: 2023 3rd International Conference on Advances in Computing, Communication, Embedded and Secure Systems (ACCESS), pp. 125–129. IEEE (2023)
15. Selvaraju, R.R., et al.: Grad-cam: visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 618–626 (2017)
16. Jetley, S., Lord, N.A., Lee, N., Torr, P.H.: Learn to pay attention. arXiv preprint arXiv:1804.02391 (2018)

Optimal Post-high School Course Selection System Leveraging Machine Learning

Varsha Lokare[1(✉)], Iram Jhetam[2], Prakash Jadhav[1], and A. W. Kiwelekar[2]

[1] Kasegaon Education Society's Rajarambapu Institute of Technology, Shivaji University, Sakharale 415414, MS, India
varsha.lokare@ritindia.edu

[2] Dr. Babasaheb Ambedkar Technological University, Lonere, India

Abstract. The transition from high school to higher education marks a serious stage in an individual's academic journey, where selecting the most suitable career path assumes thoughtful implication. The plenty of choices following high school completion necessitates a robust framework for guiding students toward optimal course selections. This paper introduces an innovative approach, the Best Course Selection System (BCSS), designed to facilitate informed decision-making leveraging machine learning methodologies. It incorporate essential criteria such as personal interests, employment prospects, eligibility criteria, affordability, duration of study, and course suitability. The BCSS integrates classifiers like Decision Tree(DT), AdaBoost, Support Vector Machine (SVM), Artificial Neural Network (ANN), etc. Three major streams—Arts, Commerce, and distinct Science streams—are considered for comprehensive analysis. Comparative evaluations based on Accuracy, Confusion Matrix, Precision, Recall, and F1-Score metrics highlight the superior performance of Support Vector Machine, Artificial Neural Network, and Decision Tree classifiers over AdaBoost. The proposed BCSS, trained and tested on a specific database, demonstrates an impressive accuracy rate of approximately 98% in predicting the most favorable course options. This system promises to serve as a tool for students navigating the complex landscape of post-secondary education choices, aiding in informed decision-making for a successful academic path.

Keywords: Best Course Selection System (BCSS) · AdaBoost; Decision Tree · Support Vector Machine · Artificial Neural Network

1 Introduction

In the wake of High School Completion, the journey toward shaping a career path stands as a significant milestone for every student. The intricacies of choosing the right course among a multitude of options can often be overwhelming and pivotal in defining one's professional trajectory. Recognizing this challenge, the proposed Best Course Selection System (BCSS) emerges as a beacon of guidance, aiming to assist students in navigating this crucial decision-making process. This proposed system harnesses the power of machine learning, employing a range of classifiers including AdaBoost, Decision Tree,

S. Fong et al. (Eds.): ICT4SD 2025, LNNS 1651, pp. 295–304, 2026.
https://doi.org/10.1007/978-3-032-06688-6_29

Support Vector Machine, and Artificial Neural Network. Through rigorous training and testing, this system endeavors to offer informed recommendations tailored to individual students, aiding them in selecting the most suitable course after HSC completion. Figure 1 illustrates the comprehensive scope of this study, which considers four primary streams: Arts, Commerce, Science (Physics + Chemistry + Biology), and Science (Physics + Chemistry + Mathematics). Within these streams, a diverse array of alternative courses has been meticulously analyzed and considered to provide students with optimal choices aligned with their specific stream.

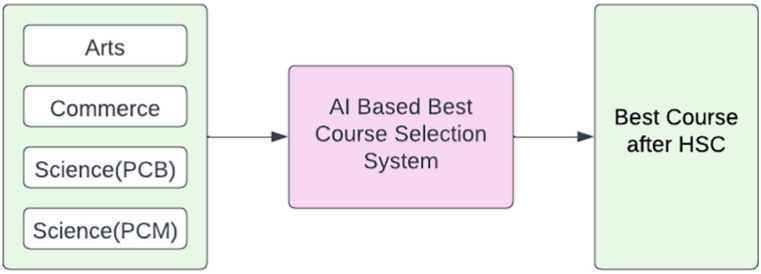


Fig. 1. Best Course Selection System (BCSS)

Figure 2 provides a glimpse into this detailed analysis, delineating stream-wise alternative courses. For instance, within the Commerce stream, options such as B.Com, C.A., BBS, and BMS are presented, while the Arts stream offers alternatives including B.A., Bachelor of Tourism, Management, and Journalism. Similarly, the Science stream (as shown in Fig. 3) branches into specific groups—PCB (Physics, Chemistry, Biology) and PCM (Physics, Chemistry, Mathematics)—each offering distinct alternative courses such as MBBS, BDS, BAMS, BPharm for PCB and Engineering, BSC, BBA, BCA for PCM. By encapsulating diverse streams and their respective alternative courses, the BCSS endeavors to be a comprehensive decision-making tool, enabling students to chart a course aligned with their interests, capabilities, and aspirations.

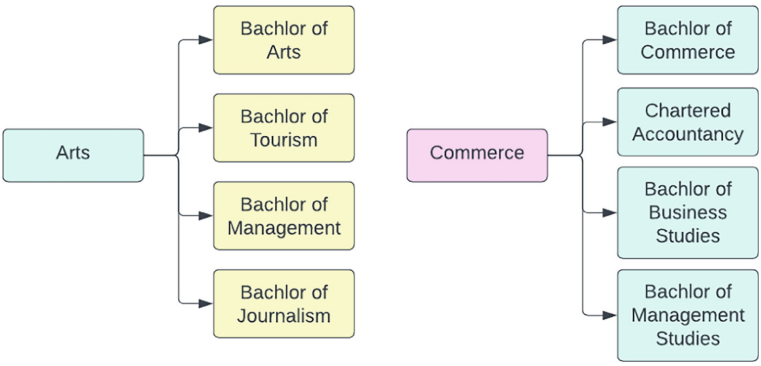


Fig. 2. Stream wise Alternative Courses

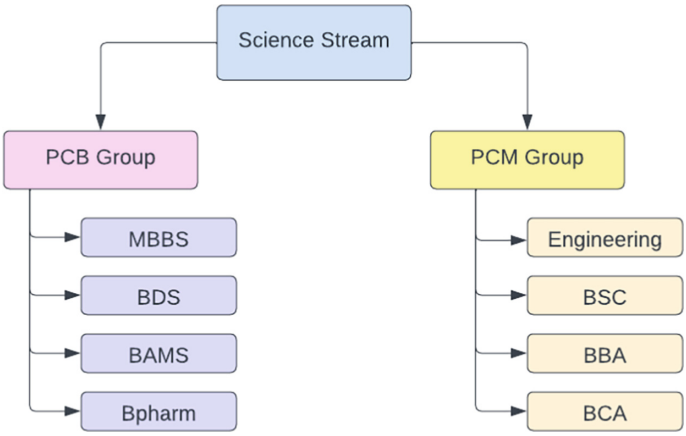


Fig. 3. Wise Alternative Courses

This introduction sets the stage for a deeper exploration into the design, implementation, and evaluation of the BCSS, underscoring its potential to empower students with informed choices for their academic and professional futures.

2 Literature Review

Post-high school course selection is a complex decision-making process that involves various factors such as student interests, academic performance, future career opportunities, and personal preferences. Several methods have been proposed to aid students in making informed decisions, incorporating both traditional decision-making techniques and modern machine learning models. However, significant gaps still exist in leveraging advanced machine learning algorithms to optimize course selection based on a holistic set of inputs. Lokare and Jadhav [1] utilized the Analytic Hierarchy Process (AHP) and Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) for decision-making in course selection after high school. This approach provided a multi-criteria decision-making framework that quantified various subjective and objective factors such as personal preferences, course difficulty, and career prospects. Although effective in balancing multiple criteria, this method lacks real-time adaptability and predictive capabilities based on evolving student data, which is a core strength of machine learning models. Machine learning techniques have been increasingly applied in various decision-making domains. AdaBoost, a popular ensemble learning method, was introduced in several studies to improve classification accuracy. Wu and Nagahashi [2] proposed a parameterized AdaBoost to accelerate training times, while Bin and Lianwen [3] applied it in handwritten Chinese character recognition. Yadahalli and Nighot [4] further demonstrated its utility in wireless sensor networks for energy-efficient routing. While AdaBoost has shown significant promise in boosting decision-making systems, its application in course selection, where student preferences evolve, is yet to be explored in-depth. The system could benefit from the adaptive learning nature of machine learning models, which can refine predictions as new data is introduced. Artificial Neural Networks (ANN) are powerful

tools for handling non-linear relationships and predicting outcomes in dynamic systems. Al-Sammarraie et al. [5] employed ANN for classification and diagnosis, while Yasay et al. [6] used it for optimizing real-time decisions such as bus route planning. Khan et al. [7] applied ANN in wireless sensor networks for dynamic duty-cycle control, demonstrating ANN's capacity to make real-time adjustments based on changing input data. ANN's dynamic nature makes it a suitable candidate for course selection systems that need to consider various student factors—such as grades, interests, and aspirations—that may fluctuate over time. However, the literature lacks examples of ANN being applied directly to educational decision systems. SVM has been effectively used for multi-class classification tasks, as demonstrated by Liu et al. [8] and Prakash et al. [9]. Liu et al. used a nested SVM approach for multi-class problems, while Prakash et al. applied SVM in machine vision applications. SVM's ability to handle high-dimensional data could be leveraged in a course selection system to classify students based on multiple features such as academic scores, extracurricular activities, and psychological assessments. However, few studies explore SVM in course selection, which represents a gap in the literature. Decision trees, often combined with other machine learning methods, have proven effective in visualization and classification tasks. Mrva et al. [10] presented a 3D decision tree for medical data decision support, while Patil and Kulkarni [11] focused on accuracy prediction using distributed decision trees in large-scale datasets. While decision trees provide clear, interpretable decision paths, the literature does not extensively cover their use in post-high school course selection systems, leaving room for research into combining decision trees with other models for enhanced decision-making and visualization.

Despite the effectiveness of AHP and TOPSIS in providing robust frameworks for multi-criteria decision-making, they fail to harness the full potential of machine learning, particularly its adaptability and predictive capabilities. The current literature lacks comprehensive studies that apply machine learning algorithms like AdaBoost, ANN, or SVM to optimize course selection based on evolving student data. Traditional decision-making methods like AHP and TOPSIS remain static and do not adapt dynamically to changing inputs over time. Machine learning models, particularly ANN, can make real-time adjustments, yet their application in the educational sector for course selection remains underexplored. Additionally, while SVM has proven effective in handling high-dimensional data for multi-class classification tasks, its use in predicting optimal courses for students based on multiple factors—such as academic scores, interests, and psychological assessments—remains limited. Furthermore, decision trees, known for their ease of interpretation and visualization, are underutilized in course selection systems. A combination of decision trees and machine learning could improve system transparency and enhance stakeholder understanding of the decision-making process. Moreover, most existing models focus on quantitative factors like academic performance, often overlooking qualitative data such as student aspirations, psychological assessments, and interests, which are manually assessed. Machine learning models could bridge this gap by integrating both quantitative and qualitative factors, offering more holistic and personalized recommendations.

3 Proposed Methodology

In this paper, the Best Corse Selection System (BCSS) is being proposed to help students to select the best course after HSC as per their stream. For this Rule based Algorithm is proposed and analyzed with the help of Machine Learning classifiers namely: AdaBoost, ANN, SVM and DT. The features considered for classification among alternative courses seem comprehensive and relevant to students' decision-making:

i) **Students' Interest:** Crucial for engagement and long-term commitment to the course.
ii) **Employability Opportunity:** Understanding job prospects post-completion is essential.
iii) **Eligibility:** Ensuring the students meet the prerequisites for the course.
iv) **Affordability of Fee:** Financial feasibility plays a significant role in decision-making.
v) **Convenience of Duration:** The duration of the course should align with the student's expectations and plans.

Analyzing these factors using machine learning classifiers could provide valuable. insights into which courses might be the best fit for students based on their preferences and circumstances.

BCSS system is structured into several key modules, each serving a specific purpose in the process of course selection.

i) **Dataset:** This module holds the data used for training the system. The dataset includes fields such as candidate interest in each course (stream-wise), course fees, employment opportunities associated with each course, duration, and eligibility criteria. The dataset comprises three subsets with varying sizes: 100, 200, and 448 instances, likely used for experimental purposes.
ii) **Feature Extraction:** This step involves extracting relevant features from the dataset that are important for decision-making. From your description, the features include candidate interest, course fees, employment opportunities, duration, and eligibility criteria.

Rule-Based Algorithm: This module likely contains predefined rules that help in decision-making based on the extracted features. These rules might be manually crafted based on domain knowledge or insights obtained from the dataset.
iii) **Machine Learning Classifiers:** This module employs various machine learning algorithms (like AdaBoost, ANN, SVM, and DT, as mentioned earlier) to learn patterns and make predictions based on the dataset's features. Each classifier has its strengths and might provide different insights into the course selection process.
iv) **Model Training and Testing:** The Model Training and Testing process involves training the system with the dataset and then testing it to see how well it performs. In this step, we check if the system can use the training data to make accurate predictions or recommendations.
v) **Prediction of New Data Input:** Once the system is trained, it can take new input data (such as a student's preferences or details) and predict or recommend the best courses based on the patterns learned during training.

This systematic approach seems comprehensive, utilizing both rule-based algorithms and machine learning techniques to assist students in making informed decisions about their course selection after HSC.

3.1 Dataset

Here dataset contains for training purpose is associated with many fields like interest of candidates in each course (Stream-wise), each course fee, Employment Opportunity, Duration and Eligibility. Table 1 shows few sample fields in the dataset. Total 3 datasets of sizes 100, 200 and 448 were considered here for experimental work.

Table 1. Samples from dataset

Interest C_0	Interest C_1	Interest C_2	Interest C_3	EmpOpr C_0	EmpOpr C_1	EmpOpr C_2	EmpOpr C_3	FeeAford C_0	FeeAford C_1	FeeAford C_2	FeeAford C_3	Best Course
6	8	3	4	70	95	85	85	y	y	y	y	C1
9	6	5	5	70	95	85	85	y	y	y	y	C0
5	9	9	8	70	95	85	85	y	n	y	y	C2
4	7	7	9	70	95	85	85	y	y	y	y	C3
6	8	7	8	70	95	85	85	y	n	n	n	C0
4	5	5	7	70	95	85	85	y	y	y	y	C3

3.2 Feature Extraction

It is really a crucial task to find out the important features from the available dataset. Following features have extracted to make decision regarding best alternative course selection:

1) Interest (0 to 10): Each candidate interest in every alternative course on the scale of 0 to 10. Here 0 means lowest interest and 10 means higher interest.
2) Employability Opportunity (%): Fixed percentile for each alternative course is given.
3) Eligibility (Y or N): Each candidate need to mention in terms of Yes and No for every alternative course's eligibility criterion are satisfied or not.
4) Fee (Rs. Per year) is Affordable (Y or N): Each candidate need to mention in terms of Yes and No for every alternative course's fee is affordable or not.
5) Duration (in years) is convenient (Y orN): Each candidate need to mention in terms of Yes and Now for every alternative course's duration is convenient or not.

3.3 Proposed Rule Based Algorithm

Here, following Rule based algorithm is proposed which is used to select best

Algorithm: Best_Course_Selection (stream)

Input:

1. Stream: Arts, Commerce, Science(PCB), Science(PCM)
2. Features: Interest, Eligibility, Fees affordable or not for each Alternative course in selectedstream

Output: Best Course after HSC as per stream

Step1: Input Stream
Step2: Input Features: Interest, Eligibility, Fees affordable or not for each Alternative course in selectedstream
Step3: CALL BCSS ($Course_i$Intersest, $Course_i$Elig, $Course_i$FeeAford)
Step4: Return Best alternative Course course after HSC.

ALGORITHM: BCSS ($Course_i$Intersest, $Course_i$Elig, $Course_i$FeeAford)

Step1: Input labeled dataset of around 2000 entries. Apply Following Rules on database for selection of best course.

RULE1: IF $Course_i$Interest = =10 AND $Course_i$EmpOpr >=90 AND $Course_i$Elig -> Yes AND $Course_i$FeeAford -YesThen Set Best_Course = $Course_i$

RULE2: IF $Course_i$Interest = =10 AND $Course_i$EmpOpr >=80 && $Course_i$EmpOpr <90 AND $Course_i$Elig -> Yes AND$Course_i$FeeAford -> Yes Then Set Best_Course = $Course_i$

RULE3: IF $Course_i$Interest >=8 && $Course_i$Interest <10 AND $Course_i$EmpOpr >=90 AND $Course_i$Elig -> Yes AND$Course_i$FeeAford -> Yes Then Set Best_Course = $Course_i$

RULE4: IF $Course_i$Interest >=8 && $Course_i$Interest <10 AND $Course_i$EmpOpr >=80 && $Course_i$EmpOpr <90 AND $Course_i$Elig ->Yes AND $Course_i$FeeAford -> Yes Then Set Best_Course = $Course_i$

RULE5**:** IF $Course_i$Interest >=6 && $Course_i$Interest <8 AND $Course_i$EmpOpr >=80 && $Course_i$EmpOpr <90 AND $Course_i$Elig -> Yes AND $Course_i$FeeAford ->YesThen Set Best_Course = $Course_i$

RULE6: IF $Course_i$Interest >=6 && $Course_i$Interest <8 AND $Course_i$EmpOpr >=90 AND $Course_i$Elig -> Yes AND $Course_i$FeeAford -> Yes Then Set Best_Course = $Course_i$

RULE7: IF $Course_i$Interest = =10 AND $Course_i$EmpOpr >=70 && $Course_i$EmpOpr <80AND $Course_i$Elig -> Yes AND $Course_i$FeeAford -> Yes Then Set Best_Course = $Course_i$

RULE8: IF $Course_i$Interest >=8 && $Course_i$Interest <10 AND $Course_i$EmpOpr >=70 && $Course_i$EmpOpr <80AND $Course_i$Elig -> Yes AND $Course_i$FeeAford -> Yes Then Set Best_Course = $Course_i$

RULE 9: IF $Course_i$Interest >=6 && $Course_i$Interest <8 AND $Course_i$EmpOpr >=70 && $Course_i$EmpOpr <80 AND $Course_i$Elig -> Yes AND $Course_i$FeeAford -> Yes Then Set Best_Course = $Course_i$

RULE 10: IF $Course_i$Interest = =5 AND $Course_i$EmpOpr >90 AND $Course_i$Elig -> Yes AND $Course_i$FeeAford -> Yes Then Set Best_Course = $Course_i$

Step2: Train Model for 1500 entries and Test for 500 entries
Step3: Apply Trained Model on new data input
Step4: Return class of best course

3.4 Model Evaluation and Validation

To evaluate model accuracy, four classifiers—AdaBoost (AdB), Artificial Neural Network (ANN), Support Vector Machine (SVM), and Decision Tree (DT)—were used. A supervised dataset of 448 entries was considered, with 75% (336 entries) for training and 25% (112 entries) for testing. AdaBoost combines weak learners like Decision Trees to form a strong classifier. ANN mimics brain neurons to process inputs and predict outcomes. SVM identifies optimal hyperplanes to classify data, suitable for both binary and multiclass problems. DT applies a rule-based, tree-like structure for stream-wise course prediction. All models used features derived from a proposed Rule-based Algorithm.

3.4.1 Performance Measurement Metrics

To evaluate the accuracy of four classifiers for predicting the best alternative course after HSC, the Confusion Matrix (as shown in Table 2) and standard classification metrics were used. The Confusion Matrix summarizes correct and incorrect predictions in a tabular form, highlighting how models get confused across classes in a multiclass setting. The dataset had 112 test entries, with 31 instances in Class 0 (e.g., B.Com for Commerce stream). Additional evaluation metrics include Precision (TP / TP + FP), Recall (TP / TP + FN), and F1-Measure, which balances both. Accuracy, representing the ratio of correct predictions to total predictions, was also considered to assess overall model performance.

Table 2. Confusion Matrix

	Predicted				
	Alternative Course 0	Alternative Course 1	Alternative Course2	Alternative Course3	Total
Class 0 (Actual)	30	1	0	0	31
Class 1 (Actual)	0	30	0	1	31
Class 2 (Actual)	0	0	26	0	26
Class 3(Actual)	0	0	0	24	24

4 Observations and Result Analysis

It has been observed that in predicting the correct best course after HSC, different models are better i.e. for class0 or for Alternative Course0, SVM and DT predict all 31 entries correct. Similarly, ANN is better for Alternative Course1, all four models gives best result in predicting alternative course2 and ANN, SVM and DT are better than ADB in predicting alternative course3 result. The observations reveal distinct patterns in the performance of various machine learning algorithms across different dataset sizes. For smaller datasets, such as Dataset 1 with 100 data elements, both AdaBoost (AdB) and Decision Tree (DT) achieve perfect accuracy (100%), which may indicate potential over fitting due to the limited sample size. However, as the dataset size increases, the accuracy for these algorithms declines slightly. For example, on Dataset 2 with 200 elements, AdaBoost accuracy drops to 88%, while Decision Tree achieves 98%. Similarly, for Dataset 3 with 448 data elements, AdaBoost maintains 88% accuracy, while Decision Tree declines further to 92%. In contrast, Artificial Neural Networks (ANN) and Support Vector Machines (SVM) exhibit more consistent performance across all datasets. ANN maintains high accuracy, achieving 96% on both Dataset 1 and Dataset 2, and improving to 98% on Dataset 3, indicating its robustness as the dataset size increases. SVM also follows a similar trend, starting with 96% accuracy on smaller datasets and reaching 98% on the largest dataset as shown in Fig. 4. These results suggest that ANN and SVM are better suited for handling larger datasets due to their stability and adaptability, while

AdaBoost and Decision Tree show slight performance declines with increased data volume. Overall, ANN and SVM outperform the other algorithms on larger datasets, while AdaBoost and Decision Tree demonstrate stronger performance in smaller datasets. It has been observed that the different classifiers shows better Precision, Recall and F1 Score value for different classes i.e. for predicting alternative course correctly Precision value of ANN is better, for Recall SVM and DT are better and for F1 score ANN, SVM and DT are equivalent.

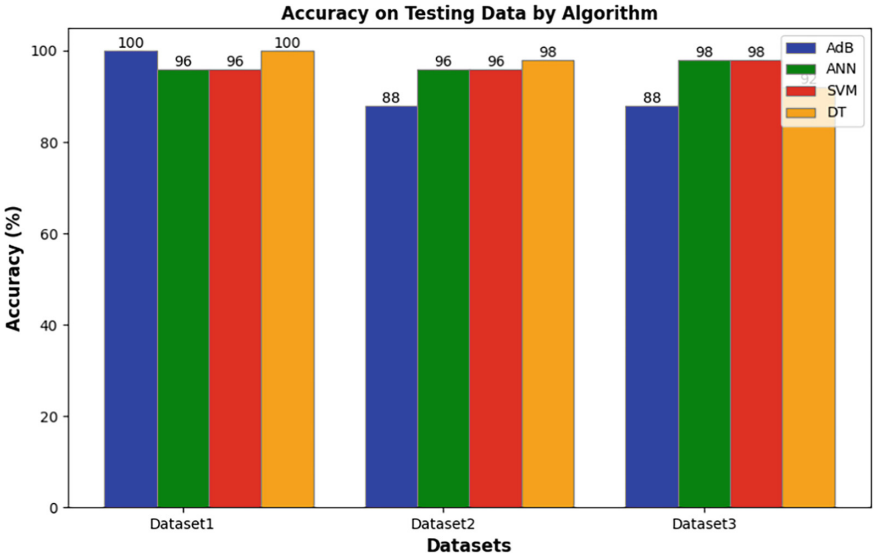


Fig. 4. Accuracy of Classifiers as per Datasets

5 Conclusion

The Optimal Post-High School Course Selection System (BCSS) introduced in this study presents a machine learning-based approach to guide students in selecting the most appropriate courses after high school. By leveraging classifiers such as Decision Tree (DT), AdaBoost (AdB), Support Vector Machine (SVM), and Artificial Neural Network (ANN), the system predicts suitable courses based on factors like personal interest, career goals, and course compatibility. Experimental results reveal that SVM and ANN outperform AdB and DT, achieving accuracy rates up to 98%, particularly on larger datasets. This highlights the robustness of SVM and ANN in handling complex and voluminous data, whereas DT and AdB are more susceptible to overfitting with small datasets. Future enhancements to BCSS may include integrating real-time student and job market data, adding psychological and career aspiration assessments, and adopting deep learning for improved scalability. Adapting the system globally could enable personalized, accurate guidance across diverse educational systems and contexts.

Acknowledgement. We used ChatGPT (OpenAI) to assist with language refinement in some sections of this paper. All scientific analysis and conclusions were conducted and verified by the authors.

References

1. Lokare, V.T., Jadhav, P.M.: Using the AHP and TOPSIS methods for decision making in best course selection after HSC. 2016 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, pp. 1–6 (2016)
2. Wu, S., Nagahashi, H.: Parameterized AdaBoost: Introducing a Parameter to Speed Up the Training of Real AdaBoost. IEEE Signal Process. Lett. **21**(6), 687–691 (2014). https://doi.org/10.1109/LSP.2014.2313570
3. Bin, Z., Lianwen, J.: Handwritten Chinese Similar Characters Recognition Based On AdaBoost. 2007 Chinese Control Conference, pp. 576–579. Hunan (2007)
4. Yadahalli, S., Nighot, M.K.: Adaboost based parameterized methods for wireless sensor networks. 2017 International Conference On Smart Technologies For Smart Nation (SmartTechCon), pp. 1370–1374. Bangalore (2017)
5. Al-Sammarraie, N.A., Al-Mayali, Y.M.H., Baker El-Ebiary, Y.A.: Classification and diagnosis using back propagation Artificial Neural Networks (ANN). 2018 International Conference on Smart Computing and Electronic Enterprise (ICSCEE), pp. 1–5. Shah Alam (2018)
6. Yasay, B.G., Dadios, E.P., Fillone, A.M.: Adaptive driving route of busses along EDSA using Artificial Neural Network (ANN). 2015 International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM), pp. 1–4. Cebu City (2015)
7. Khan, A.A., Jamal, M.S., Siddiqui, S.: Dynamic duty-cycle control for wireless sensor networks using artificial neural network (ANN). 2017 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC), pp. 420–424. Nanjing (2017). https://doi.org/10.1109/CyberC.2017.93
8. Liu, B., Hao, Z.-F., Yang, X.-W.: Nesting support vector machinte for muti-classification [machinte read machine]. 2005 International Conference on Machine Learning and Cybernetics, Vol. 7, pp. 4220–4225. Guangzhou, China (2005)
9. Suriya Prakash, J., Annamalai Vignesh, K., Ashok, C., Adithyan, R.: Multi class support vector machines classifier for machine vision application. 2012 International Conference on Machine Vision and Image Processing (MVIP), pp. 197–199. Taipei (2012)
10. Mrva, J., Neupauer, Š., Hudec, L., Ševcech, J., Kapec, P.: Decision Support in Medical Data Using 3D Decision Tree Visualisation. 2019 E-Health and Bioengineering Conference (EHB), pp. 1–4. Iasi, Romania (2019)
11. Patil, M., Kulkarni, R.: Accuracy prediction and best classifier using decision tree. 2019 International Conference on Data Science and Communication (IconDSC), pp. 1–6. Bangalore, India (2019)

Multiclass Classification of Mammographic Density and Mass Regions for Breast Cancer Diagnosis Using a ResNet-Based Framework

Piyush Sharma[1(✉)], Harish Patidar[1], and Anuj Kumar[2]

[1] Department of CSE, Mandsaur University, Daulatpura, India
piyushsharma931@gmail.com, harish.patidar@gmail.com
[2] Department of Radiotherapy, Sarojini Naidu Medical College, Agra, India
toak-tyagi@gmail.com

Abstract. This research introduces a ResNet-based framework for multiclass classification of mammographic density and mass regions. The framework was rigorously tested using two prominent mammographic datasets, INbreast and DDSM, and benchmarked against other models, including CNNs, Random Forest (RF), Support Vector Machines (SVMs), Logistic Regression (LR), and K-Nearest Neighbors (KNN). ResNet demonstrated superior performance across all critical evaluation metrics—accuracy, precision, recall, F1-score, and AUC—outclassing the comparative models on both datasets. Its proficiency in extracting complex hierarchical features and addressing multiclass classification tasks positions it as a robust choice for breast cancer diagnosis. This framework offers a reliable and efficient tool for automating diagnostic processes, with the potential to significantly improve clinical decision-making and patient care.

Keywords: ResNet Framework · Mammographic Density Classification · Multiclass Classification · Deep Learning in Healthcare · Breast Cancer Diagnosis

1 Introduction

Mammographic imaging is a cornerstone for breast cancer screening, as it allows for the identification of key abnormalities such as mass regions and variations in mammographic density. However, traditional methods of diagnosis are often subject to the limitations of manual interpretation, including inter-observer variability and the potential for errors in detecting subtle abnormalities. Recent advances in machine learning and deep learning techniques have opened up new avenues for automating mammogram analysis, offering more accurate and efficient diagnostic support. Among these, deep convolutional neural networks (CNNs), especially Residual Networks (ResNet), have emerged as powerful tools for tackling complex image classification tasks, including breast cancer diagnosis [1–4]. ResNet's ability to effectively learn hierarchical features while mitigating issues like vanishing gradients has made it a suitable architecture for this task. This study explores the use of ResNet for multiclass classification of mammographic density and mass regions [1, 5–9].

S. Fong et al. (Eds.): ICT4SD 2025, LNNS 1651, pp. 305–314, 2026.
https://doi.org/10.1007/978-3-032-06688-6_30

1.1 Motivation and Contribution

Breast cancer treatment and survival rates are improved by early diagnosis; yet, because of subtle visual signals and a variety of tissue features, mammography interpretation is still difficult. The complexity of multiclass problems, such as differentiating between mass types and mammographic densities, and managing unbalanced datasets are frequently overlooked by current methods, which concentrate on binary classification. In order to overcome the shortcomings of current techniques, this study suggests a ResNet-based framework for multiclass categorization of mammographic densities and mass areas. The system is tested on well-known datasets (INbreast and DDSM) and compared to models like as Random Forest, SVM, CNN, and Logistic Regression. It provides an affordable clinical integration solution by increasing diagnostic precision and supporting early cancer diagnosis, which may lower mistakes and improve patient outcomes.

2 Related Work

Image processing techniques have been extensively studied in recent years, with various innovative approaches. The following summarizes key research and advancements in this area, categorized into several key topics: feature selection, deep learning-based classification, breast density assessment, multi-class classification, and mass segmentation.

2.1 Feature Selection and Gene Biomarkers

A hybrid feature selection approach that utilized mRMR, t-tests, and metaheuristic algorithms determined the critical gene biomarkers for breast cancer with an F1-Score of $\pm$ 0.030 and AUC of 0.961 $\pm$ 0.035 on an external dataset. Machine learning approaches such as SVM, KNN, and XGBoost were employed, of which XGBoost provided the best accuracy of 0.976 $\pm$ 0.027, indicating its potential in enhancing diagnostic accuracy.

2.2 Deep Learning-Based Classification Systems

Deep learning, especially CNNs, has dramatically increased breast cancer diagnosis. On the mini-MIAS and DDSM datasets, a hybrid system that combined CNN-based feature extraction with an ensemble classifier and Tunicate Swarm Optimization (TSO) demonstrated its efficacy in three-class classification tasks by achieving high sensitivity and accuracy [11]. Furthermore, transfer learning using pretrained models such as VGG16, VGG19, and ResNet50 has demonstrated promising outcomes. VGG16 outperformed the other models and was useful for clinical assistance, categorizing DDSM and UPMC pictures into four diagnostic categories with an accuracy of 92–95% [12].

2.3 Breast Density Assessment

Breast density frequently makes it more difficult to detect tumours in mammograms and is a significant predictor of cancer risk. Although radiologists have historically used BI-RADS to assess patients, this procedure can be irregular. A deep learning model known

as DLAD was created for automated density evaluation in order to overcome this. When tested on full-field digital mammograms, DLAD performed well, outperforming radiologists in many instances and achieving 0.819 accuracy and 0.798 F1-score, indicating its potential for reliable and accurate breast density assessment [5].

2.4 Multi-Class Classification for Mammography Images

The ability to differentiate between benign, malignant, and normal cases in mammography pictures has become essential for the identification of breast cancer. Morphological segmentation and bicubic interpolation were used in a recent work that used the MIAS dataset [13] to improve images. Data augmentation was then used to increase variability. Several classifiers, including CNN, KNN, SVM, Naive Bayes, and Decision Tree, were assessed after eleven important characteristics were retrieved. The framework's efficacy for multi-class classification tasks was confirmed by its good diagnostic performance across parameters including accuracy, precision, sensitivity, and F-score.

2.5 Mass Segmentation Techniques

A region-sample hybrid loss function was introduced to handle pixel class imbalance and mass variation in the segmentation of breast cancer masses. The method, evaluated on the CBIS-DDSM and INbreast databases, performed better than current approaches in segmentation accuracy. The combination of deep learning, feature selection, and ensemble classification, especially via transfer learning, improves system performance. Recent studies aim to enhance breast density estimation, solve data imbalance issues, and investigate multimodal inputs, including the integration of mammograms with genetic or histopathological information, to further improve predictive performance.

3 Methodology

In this section, we present our proposed approach for implementing a breast cancer diagnostic system using ResNet. The methodology is a comprehensive process that includes multiple stages such as data preparation, feature extraction, feature selection, and multi-class classification. The system operates on mammographic images from two widely recognized datasets, focusing on accurately identifying breast tissue density and mass regions. Each step is carefully designed to enhance the overall diagnostic accuracy, aiming to improve the identification of malignant and benign masses, as well as the characterization of mammographic density. A graphical representation of this process is shown in Fig. 1.

a. **Experimental Dataset**

The INbreast dataset [14], a reputable set of 116 high-resolution digital mammograms labelled with mass areas and BI-RADS density levels, was used for this investigation. A thorough assessment of the suggested approach across distinct breast tissue types is made possible by the dataset's diversity across four density categories and different mass characteristics. It is perfect for training models that can identify tiny

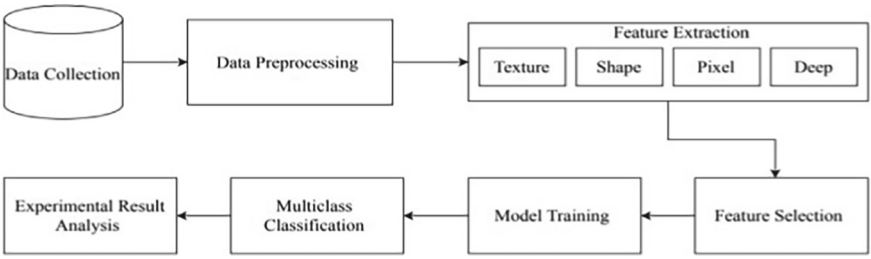


Fig. 1. Proposed flow

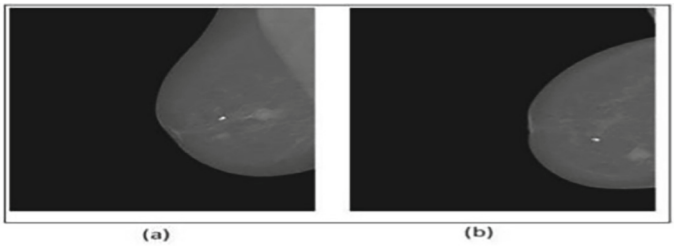


Fig. 2. Sample images from the INbreast dataset (right breast views): (a) Mediolateral oblique view and (b) Craniocaudal view.

anomalies, enabling real-world clinical application, because of its thorough annotations and excellent image quality.

Several initial procedures were carried out to prepare the data for analysis. First, a Gaussian filter was applied to remove noise, smoothing the images while preserving key details. Following this, standardization was performed to normalize pixel intensities to a consistent range, minimizing the effects of variability across different imaging devices [10, 15–17]. Additionally, contrast enhancement was applied to emphasize specific areas, ensuring better clarity for identifying mammographic densities and mass regions [Figs. 2 and 3].

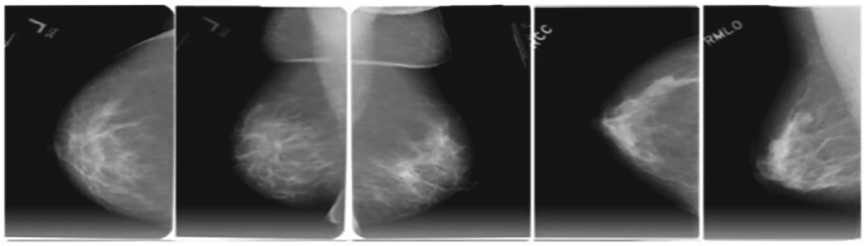


Fig. 3. Image samples from DDSM dataset

b. **Feature Extraction**

A mix of handmade and deep learning characteristics was used to improve the categorization of breast cancer. In addition to intensity-based metrics like average brightness and variance, texture and shape descriptors like GLCM, LBP, and HOG were recovered, as well as geometry measurements like area, perimeter, and circularity. Furthermore, a fine-tuned pre-trained CNN was used to generate deep features. To increase classification accuracy, dimensionality reduction methods like PCA and RFE were utilized, conserving the most significant features.

Dataset Description:

- **INbreast Dataset:** A small yet high-quality dataset containing 116 full-field digital mammograms with detailed used for tasks requiring precise visual features. **DDSM Dataset:** A large-scale dataset of 2,620 scanned mammograms offering diverse image resolution and extensive annotations, including BI-RADS assessments and mass characteristics with benign, malignant and suspicious labels.

ResNet-50 was selected because of its capacity to effectively extract features from high-resolution mammograms and handle vanishing gradients. The model was trained over 1000 epochs with a batch size of 32 and an initial learning rate of 0.001. The input photos were downsized to 224 × 224 pixels and enhanced using rotation, zoom, brightness tweaks, and random horizontal flipping. Values unique to each dataset were used to normalize the images. To guarantee consistency and minimize variation, a stratified 5-fold cross-validation method was employed, with performance measures averaged across folds.

Clinical Relevance and Deployment Potential:
Automating the categorization of mammographic density and mass zones, the system shows promise for real-world clinical implementation. This can speed up decision-making in breast cancer screening workflows, enhance diagnostic consistency and drastically lessen the strain for radiologists.

4 ResNet Classifier: Mathematical Foundation And Multiclass Classification

4.1 Mathematical Foundation of ResNet

ResNet (Residual Networks) is a deep learning architecture designed to address the vanishing gradient problem and improve the training of very deep networks by introducing residual connections. These connections allow the network to learn residual mappings rather than direct mappings, enabling easier optimization of deep networks.

- **Residual Learning Framework:** In ResNet, the output of a layer is passed through a residual connection that adds the input directly to the output of the layer. This operation can be mathematically expressed as:

$$F(x) = H(x) - x$$

where F(x) is the residual function, H(x) is the desired underlying mapping, and x is the input. This allows the network to learn the difference between the desired output and the input, improving the overall performance.

- **Objective Function:** In ResNet, the objective is to minimize a loss function that evaluates the model's performance. The objective function L typically includes the following components:

$$L = \sum_i Loss(y_i, \hat{y}_i)$$

where Loss(yi,y^i) is a loss function (such as cross-entropy for classification tasks) that measures the difference between the true label yi and the predicted label *y i*

- **Optimization:** ResNet's training uses backpropagation to update the weights of the network.
- **Multiclass Classification:** ResNet can be adapted for multiclass classification tasks by using a softmax activation function in the output layer.

4.2 Multiclass Classification with ResNet

ResNet can be extended to handle multiclass classification problems, offering an efficient solution for tasks such as mammographic density and mass region detection. Here's how ResNet adapts for multiclass classification:

A. **Softmax Function:** For multiclass classification, ResNet employs the softmax function to convert raw output scores into probabilities. The softmax function ensures that the model's predictions are interpretable as class probabilities. It is defined as:

$$P(y = k|x) = \frac{e^{score_k}}{\sum_j e^{score_j}}$$

where scorek is the score for class k, and the denominator is the sum of the exponentials of the scores for all classes. This function normalizes the raw output of the network into a probability distribution.

B. **Objective Function for Multiclass:** For multiclass classification, ResNet uses a cross-entropy loss function combined with the softmax activation to measure the difference between the true labels and predicted class probabilities. The multiclass softmax loss is defined as:

$$Soft\max\ Loss = -\sum_i \log\left(\frac{e^{score_{yi}}}{\sum_j e^{score_j}}\right)$$

where yi is the true class label for instance iii, and scorej are the output scores of the model for each class. This loss function encourages the model to increase the probability of the true class while reducing the probabilities of incorrect classes.

a.**Class Probabilities:** During training, ResNet adjusts its internal weights to minimize the softmax loss. The training process utilizes backpropaga-tion to update the model parameters.

C. **Multiclass Detection for Density and Mass Regions:** In context of mammographic image analysis, ResNet handles multiclass classification by predicting different categories of mammographic density (e.g., fatty, dense) and mass regions (e.g., benign, malignant). The model is trained with a multiclass objective function to effectively predict and classify these categories, enabling accurate breast cancer detection from mammographic images.

5 Experimental Result Analysis

Treatment results are improved by early diagnosis of breast cancer, although complex tissue characteristics make mammography interpretation difficult. Using a ResNet-based approach, this work tackles problems such as pixel class imbalance and variability in the classification of mammographic densities and mass areas. Tested on the DDSM and INbreast datasets, the model performs better than conventional methods, providing enhanced diagnostic precision and the possibility of smooth clinical integration.

Table 1. Performance of Models on INbreast Dataset

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AU C
ResNet	94.5	92.5	93.2	92.8	0.96
CNN	90.2	88.4	89.1	88.7	0.92
Random Forest	87.3	85.0	85.9	85.4	0.89
SVM	85.1	83.8	83.2	83.5	0.87
Logistic Regression	82.0	80.9	80.5	80.7	0.84
KNN	80.2	78.9	78.4	78.6	0.81

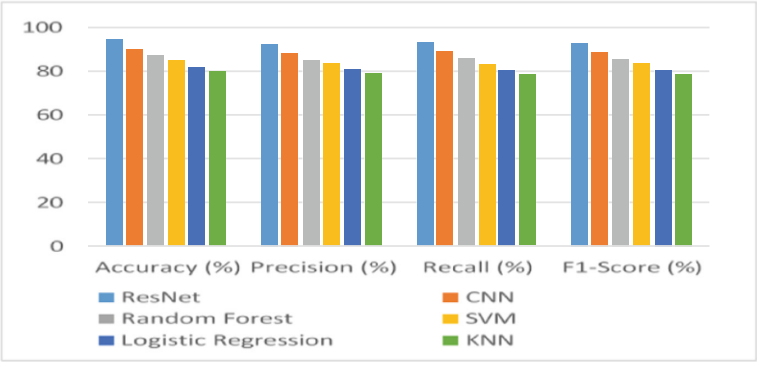


Fig. 4. Performance of Models on INbreast Dataset

ResNet outperforms CNN in breast cancer classification due to its deeper architecture, capturing complex feature interactions. Traditional models like SVM, Logistic Regression, and KNN showed lower performance, with accuracies of 85.1%, 82.0%, and 80.2%, respectively, indicating challenges in handling intricate mammographic features [Table 1, Fig. 4]. Evaluations on the DDSM dataset confirmed ResNet's superior performance and reliability across different datasets.

Table 2. Performance of Models on DDSM Dataset

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AU C
ResNet	92.1	90.3	91.0	90.6	0.94
CNN	88.5	86.3	87.1	86.7	0.90
Random Forest	85.7	83.5	84.2	83.8	0.87
SVM	83.6	82.2	81.9	82.0	0.85
Logistic Regression	81.2	80.0	79.6	79.8	0.82
KNN	78.7	77.3	76.9	77.1	0.79

ResNet beats CNN and standard machine learning models like Random Forest, SVM, Logistic Regression, and KNN in breast cancer classification due to its deeper architecture and ability to capture complex feature interactions [Table 2]. Tests using datasets such as DDSM and INbreast confirm that ResNet performs better on a number of parameters. This demonstrates how well deep learning techniques handle complex mammography data for precise diagnosis.

The graphical representation for the INbreast dataset results is presented in Fig. 5 and for the DDSM dataset.

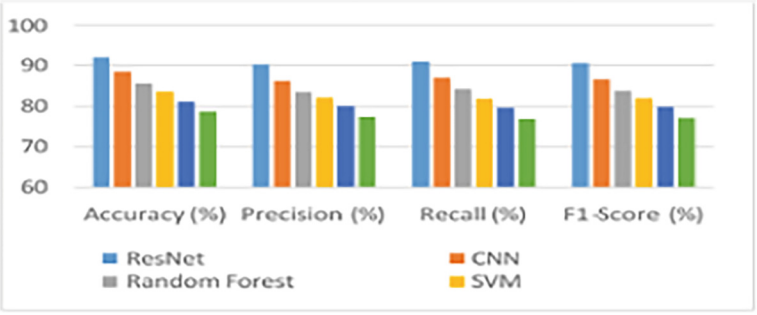


Fig. 5. Performance of Models on INbreast Dataset

Experimental Setup
To evaluate the proposed ResBooNet architecture, we conducted extensive experiments using both traditional machine learning models and deep learning baselines. Our approach includes:

Comparative Evaluation: The model's performance was compared to that of a typical Convolutional Neural Network (CNN) and other conventional classifiers, such as Random Forest (RF), Support Vector Machine (SVM), Logistic Regression (LR), and K-Nearest Neighbors (KNN).

Multiclass Classification: Our setup includes multiclass categorization of histopathological images into several subtypes, in contrast to most research that focus on binary classification (i.e., benign vs. malignant). This more accurately depicts clinical situations in the actual world when subtle differences are crucial.

ResNet Depth: ResNet-50 balances computational efficiency and representational capacity, we used it as the backbone architecture. With minimal training costs, hierarchical characteristics might be captured by the network depth.

Interpretability: We recognize the significance of interpretability in medical AI, even if the main emphasis of the current work is classification performance. Subsequent versions of this study will investigate how to depict and describe the model's decision-making process by integrating methods like Grad-CAM, saliency maps, and SHAP. These resources are essential for building confidence among medical professionals and confirming the clinical applicability of learnt aspects.

6 Conclusion

In this study, we implemented a ResNet-powered Breast Cancer Diagnosis System to classify mammographic density and detect mass regions in multiple classes. Its effectiveness was demonstrated using two benchmark datasets, INbreast and DDSM. ResNet consistently outperformed other models, including CNN, Random Forest, SVM, Logistic Regression, and KNN, across key evaluation metrics such as accuracy, precision, recall, F1-score, and AUC. The model's ability to handle complex feature interactions and its deep learning architecture provided a significant advantage in addressing the challenges of multiclass classification in breast cancer diagnosis. Future work could focus on extending this approach to larger and more diverse datasets or integrating additional clinical features to further enhance diagnostic accuracy and generalizability.

References

1. Gudhe, N.R., et al.: A multi-view deep evidential learning approach for mammogram density classification. IEEE Access (2024)
2. Razali, N.F., Isa, I.S., Sulaiman, S.N., Abdul Karim, N.K., Osman, M.K., Che Soh, Z.H.: Enhancement technique based on the breast density level for mammogram for computer-aided diagnosis. Bioengineering **10**(2), 153 (2023)

3. Sharma, P., Laxkar, P., Kumar, A.: Performance analysis of machine learning, deep learning and ensemble techniques for breast cancer diagnosis. In: International Conference on Electrical and Electronics Engineering, pp. 292–311. Springer (2022)
4. Al-Tam, R.M., Al-Hejri, A.M., Alshamrani, S.S., Al-antari, M.A., Narangale, S.M.: Multimodal breast cancer hybrid explainable computer-aided diagnosis using medical mammograms and ultrasound images. Biocybernetics and Biomed. Eng. **44**(3), 731–758 (2024)
5. Rani, N., Gupta, D.K., Singh, S.: Multi-class classification of breast cancer abnormality using transfer learning. Multimedia Tools and Applications, 1–16 (2024)
6. Wajeed, M.A., et al.: A breast cancer image classification algorithm with 2c multiclass support vector machine. J. Healthc. Eng. **2023**(1), 3875525 (2023)
7. Aymaz, S.: A new framework for early diagnosis of breast cancer using mammography images. Neural Comput. Appl. **36**(4), 1665–1680 (2024)
8. Ahmad, J., et al.: Deep learning empowered breast cancer diagnosis: advancements in detection and classification. PLoS ONE **19**(7), 0304757 (2024)
9. Cruz-Ramos, C., et al.: Benign and malignant breast tumor classification in ultrasound and mammography images via fusion of deep learning and handcraft features. Entropy **25**(7), 991 (2023)
10. Ahmad, J., Akram, S., Jaffar, A., Rashid, M., Bhatti, S.M.: Breast cancer detection using deep learning: An investigation using the ddsm dataset and a customized alexnet and support vector machine. IEEE Access (2023)
11. Alromema, N., Syed, A.H., Khan, T.: A hybrid machine learning approach to screen optimal predictors for the classification of primary breast tumors from gene expression microarray data. Diagnostics **13**(4), 708 (2023)
12. Laishram, R., Rabidas, R.: An optimized ensemble classifier for mammographic mass classification. Comput. Electr. Eng. **119**, 109488 (2024)
13. Biroˇs, M., et al.: Enhancing accuracy in breast density assessment using deep learning: a multicentric, multi-reader study. Diagnostics **14**(11), 1117 (2024)
14. Aliniya, P., Nicolescu, M., Nicolescu, M., Bebis, G.: Improved loss function for mass segmentation in mammography images using density and mass size. Journal of Imaging **10**(1), 20 (2024)
15. Yadav, R.K., Singh, P., Kashtriya, P.: Diagnosis of breast cancer using machine learning techniques-a survey. Procedia Comp. Sci. **218**, 1434–1443 (2023)
16. Abhisheka, B., Biswas, S.K., Purkayastha, B.: A comprehensive review on breast cancer detection, classification and segmentation using deep learning. Archives of Computat. Methods in Eng. **30**(8), 5023–5052 (2023)
17. Rehman, K.U., Li, J., Pei, Y., Yasin, A., Ali, S., Saeed, Y.: Architectural distortion-based digital mammograms classification using depth wise convolutional neural network. Biology **11**(1), 15 (2021)

Personalized Trajectory of Teacher Professional Development via Digital Platforms

I. Yarullin[1], R. Nasibullov[1], Sh. Sheymardanov[1](✉), N. Zhiyenbayeva[2], and T. Yechshzhanov[3]

[1] Kazan Federal University, Kazan 420008, Russian Federation
pedagogshamil@mail.ru
[2] Abai Kazakh National Pedagogical University, Almaty 050010, Republic of Kazakhstan
[3] Arqalyq State Pedagogical University Named After Ybyrai Altynsarin, Arkalyk 110300, Republic of Kazakhstan

Abstract. The necessity of teachers' knowledge and abilities being updated on a regular basis in a world that is changing quickly determines the paper's significance. To do this, new strategies for planning teachers' professional development that consider their unique needs and traits must be developed. Digital platforms provide new opportunities for creating personalized trajectories of professional growth, enabling teachers to access relevant knowledge and skills at their convenience and in preferred formats. The research aims to develop a methodology for personalized teacher professional development trajectories based on digital platforms. The authors analyze how modern digital platforms can facilitate the personalization of teacher training, considering individual needs and goals. Special attention is given to the concept of personalized professional development trajectories, which are built on diagnosing teachers' competence levels and needs. The article also presents the main principles and approaches to understanding personalized learning, as well as the stages of teacher professional development. The authors emphasize that the implementation of digital platforms in education faces challenges such as digital inequality, ethical issues of data processing, and resistance to innovation from parts of the teaching community. The article is of interest to researchers in education, teachers, and anyone concerned with the challenges of teacher professional development in the digital age.

Keywords: digital platform · mentoring · professional development · teacher · teacher education

1 Introduction

Modern education is undergoing a profound digital transformation, necessitating a rethinking of approaches to teachers' professional development. Traditional models of professional development, based on standardized programs, are giving way to personalized trajectories that account for individual needs, competence levels, and professional goals. In this context, digital platforms have become a key tool, enabling not only flexible adaptation of learning content but also providing continuous support through artificial

S. Fong et al. (Eds.): ICT4SD 2025, LNNS 1651, pp. 315–324, 2026.
https://doi.org/10.1007/978-3-032-06688-6_31

intelligence, data analytics, and microlearning technologies. UNESCO documents note that the AI competency framework for teachers focuses on continuous professional development, offering a reference framework for national competency development and training programs. It aims to ensure that teachers are equipped to use AI responsibly and effectively while minimizing potential risks to students and society (UNESCO, 2024).

Personalized trajectories, based on the analysis of individual needs and data, allow teachers to develop skills aligned with their career goals and the challenges of the digital era (Ferguson, 2012; Hargreaves and Fullan, 2020).

An analysis of research from the past decade shows that the concept of personalization in teachers' professional development relies on several key principles. First, it involves a shift from uniform programs to adaptive systems that adjust content based on user progress (Huang et al., 2019). Second, reflective practice plays a crucial role, supported by digital tools for self-assessment and feedback (Amhag et al., 2019; Galimov et al., 2024). Third, learning analytics data helps identify patterns in teacher learning and provide personalized recommendations (Luckin, 2018; Siemens, 2019).

Despite the evident advantages, the implementation of digital platforms faces several challenges. These include digital inequality due to disparities in technology access (Selwyn, 2022), ethical concerns regarding personal data processing, and resistance to innovation from parts of the teaching community (Ertmer, 2005). Additionally, the long-term effectiveness of such solutions remains an open question: most existing studies focus on short-term outcomes, whereas the impact of digital trajectories on teachers' career growth requires further investigation (Guskey, 2014; Jaclyn J. Gish-Lieberman et al., 2021).

Thus, personalized models of professional development, supported by digital platforms, represent a promising direction. However, their effective application requires addressing organizational, cultural, and ethical issues in addition to technological skills. Future research could focus on comparative analyses of various platforms and the development of hybrid formats combining online learning with practice-oriented activities.

A personalized professional development trajectory represents an individual path for a teacher's growth, considering their needs, interests, abilities, and opportunities. It is built on diagnosing the teacher's level of professional competence and their developmental needs.

The historical context of personalized learning is rooted in humanistic pedagogy, emphasizing the uniqueness of each individual and the need to account for personal characteristics in the learning process.

Key definitions and approaches to understanding personalized trajectories include:

individualized approach; tailoring to each teacher's unique characteristics; differentiated learning; grouping teachers based on their needs and interests; flexible planning; allowing teachers to choose their own pace and direction of professional growth.

The stages of a teacher's professional development are listed below.

1. Diagnosis of professional competence level.
2. Identifying the professional competency level of a teacher and determining their needs for professional development.
3. Designing an individual educational program.

4. Developing an individual educational program based on the results of the diagnosis.
5. Implementation of the individual educational program.
6. Carrying out the individual educational program using various forms and methods of teaching.
7. Monitoring and evaluation of professional development results.
8. Tracking the progress of the teacher in mastering the individual educational program.
9. Correction of the individual educational program if necessary.
10. Making changes to the individual educational program as needed.

2 The Role of Digital Platforms in Building Personalized Trajectories

The integration of digital technologies into education opens new horizons for teachers' professional growth. Digital platforms serve as powerful tools for creating personalized development trajectories, providing access to extensive resources and tools for lifelong learning. Organizations must continuously identify digital options worth exploring and selectively implement those that create new value (Sandberg et al., 2014). Below, we examine the functionalities of digital platforms and examples of successful solutions in educational practice.

Digital platforms enrich the learning process by offering diverse features for personalized learning:

1. Access to educational resources: Platforms provide a wide range of materials, from courses and webinars to video lessons and interactive tasks, tailored to different skill levels and interests.
2. Interaction with experts: Digital platforms create spaces for knowledge exchange between teachers and specialists through consultations, workshops, and training sessions.
3. Progress tracking: Analytics and reporting tools allow teachers to monitor their progress in learning programs, receive feedback, and adjust their development paths.
4. Motivation support: Digital environments help maintain high motivation by offering opportunities to participate in competitions, projects, and other activities that stimulate professional growth.

The Russian and Kazakhstan markets offers numerous digital platforms aimed at supporting the professional development of educators. Among the most popular are: Stepik, Uchi.ru, Yandex Textbook (Yandex.Uchebnik) and Foxford.

Stepik is an educational platform providing a wide range of courses across various disciplines, including pedagogy. It allows teachers to select training programs tailored to their interests and needs, granting access to high-quality educational resources. Stepik can be instrumental in designing personalized professional development pathways for educators.

Key features of Stepik for teachers are as follows.

Course diversity: the platform offers numerous courses on pedagogy, psychology, teaching methodologies, and more, enabling educators to select relevant content.

Personalized learning: teachers can create customized learning paths based on their individual needs, including course selection, pacing, and feedback.

Progress tracking: educators can monitor their learning progress, identify achievements, and pinpoint areas for improvement.

Feedback system: the platform provides constructive feedback, helping teachers recognize strengths and weaknesses.

Community engagement: Stepik connects educators worldwide, facilitating knowledge exchange and collaboration.

Accessibility: available on multiple devices, allowing learning anytime, anywhere.

Flexible formats: video lectures, text materials, quizzes, and assignments to suit different learning preferences.

Certification: some courses provide certificates upon completion, validating acquired skills.

Data analytics: performance data helps teachers refine their learning strategies.

Overall, Stepik is a valuable tool for teacher professional development, offering diverse learning opportunities, progress tracking, feedback, and networking.

Uchi.ru is another educational platform supporting teacher development and personalized learning pathways.

Key features of Uchi.ru for teachers are as follows.

Diverse resources: extensive teaching materials across subjects help educators selects appropriate content.

Personalized learning tools: enables tailored learning paths for students, which can also inform teacher development plans.

Feedback and analytics: providing insights into student progress and teaching effectiveness.

Community and collaboration: connecting teachers globally for experience sharing and innovative teaching approaches.

Accessibility and flexibility: multi-device compatibility and varied learning formats (videos, texts, quizzes).

Skill development: enhances pedagogical skills, including tech integration and adaptive teaching.

Professional certification: offers upskilling courses with certification options.

Peer support. Fosters a collaborative teacher community.

Yandex Textbook (Yandex.Uchebnik) is a teacher-focused platform offering lesson planning tools, assessments, and student performance analytics.

Key Features of Yandex Textbook are as follows.

Rich resource library: extensive subject-specific materials.

Personalized learning: customizable student learning paths.

Feedback and analytics: tracks student progress and teaching efficiency.

Teacher community: facilitates global networking and idea exchange.

Accessibility and flexibility: supporting various learning formats.

Skill enhancement: develops key teaching competencies.

Integration with Yandex Services: works seamlessly with Yandex 360, Yandex.Disk, etc.

Content creation: teachers can develop customized materials.

Competitions and projects: professional challenges and collaboration opportunities.

Foxford is an educational platform supporting teacher development through structured courses and personalized learning.

Key Features of Foxford are following.

Wide course selection: covers diverse subjects and teaching strategies.

Personalized learning: adapts to individual teacher and student needs.

Feedback and analytics: evaluating teaching effectiveness.

Community and networking: connecting educators for collaboration.

Flexible learning: accessible across devices with multimedia resources.

Professional growth: enhances pedagogical and technological skills.

Certification: validates completed courses.

Events and workshops: professional engagement opportunities.

Global experience demonstrates that digital platforms are becoming a crucial tool for enhancing the quality of education and the professional development of educators. Leading countries in educational digitalization, such as Finland, Singapore, India, and the United States, are actively integrating digital solutions into their educational systems.

Finland, renowned for its innovative education policies, utilizes digital platforms to create personalized learning paths for each student. A multitude of tools and resources are available to educators, allowing them to customize instruction to meet the needs of each individual student.

Singapore, a global leader in educational technology, has developed a comprehensive system to support teachers' professional development through digital platforms. Teachers can participate in online courses, seminars, and workshops, improving their qualifications and exchanging experiences with colleagues.

In India, the digitalization of education is advancing rapidly due to significant government investments and private initiatives. This growth is driven by several factors: the country's large population, the need to ensure equal access to education for all social groups, and the geographical remoteness of many communities. For example, the National Educational Portal (SWAYAM) – an e-learning platform – offers courses from India's leading universities, covering a wide range of disciplines and educational levels.

The United States actively incorporates digital platforms into higher education, providing universities and colleges with tools to develop personalized learning programs for students and instructors. These platforms facilitate the creation of individualized curricula, progress tracking, and feedback provision.

These examples illustrate how digital platforms can enhance education quality and foster teachers' professional growth, making learning more accessible, flexible, and effective.

Russia and Kazakhstan are also actively advancing digital technologies in education by implementing innovative solutions to support teachers' professional development. Projects introducing digital platforms in schools and universities aim to create personalized learning pathways for educators, considering their individual needs and interests.

Domestic initiatives include the development of specialized online platforms for teacher training, providing access to courses and seminars on relevant educational topics. Such platforms allow educators to select professional development programs aligned with their career goals.

The integration of digital technologies into the educational process opens new horizons for teachers' professional growth but also presents them with a range of challenges and risks. Below, we examine the key limitations and risks associated with the use of digital platforms, as well as the need to prepare educators for working with new technologies.

Technical limitations: despite technological advancements, not all educational institutions have access to high-speed internet or sufficient technical resources, which may hinder the effective use of digital platforms.

Data security concerns: digital platforms collect and process vast amounts of users' personal data, raising concerns about security and confidentiality.

Information overload risk: the abundance of information and resources on digital platforms can lead to cognitive overload, making it difficult for educators to select high-quality materials.

Technology dependence: overreliance on digital tools may lead to dependency, reducing educators' ability to adapt to traditional teaching methods when necessary.

Lack of standardization: the variety of digital platforms and courses complicates the standardization of professional development programs, making quality control more challenging.

Financial costs: implementing and maintaining digital platforms requires significant financial investments, which may be unaffordable for some institutions.

Psychological barriers: some educators may experience discomfort when transitioning to new technologies, especially if they lack confidence in their digital skills.

Resistance to change: educators accustomed to traditional methods may resist adopting new digital practices.

Unequal access: disparities in access to digital technologies create inequalities in professional development opportunities among educators.

To ensure the effective use of digital platforms in teacher professional development, the following measures should be implemented.

Tool proficiency training: educators should acquire essential skills for using digital tools such as online courses, webinars, and video lessons.

Enhancing digital literacy: teachers must develop the ability to critically assess digital content and apply it in their practice.

Understanding personalized learning: educators should grasp the principles of personalized learning and adapt them to their specific needs.

Peer knowledge exchange: teachers should leverage digital platforms for collaboration and sharing best practices with colleagues.

3 Methodology for Designing a Personalized Trajectory of Professional Development for Teachers via Digital Platforms

It takes careful planning at every step of the process to create an individualized professional development trajectory for teachers. This includes diagnosing the needs and abilities of each teacher, creating a customized educational program, putting the trajectory into action and supporting it, and keeping an eye on and adjusting the professional development process.

3.1 Diagnosis of Teacher Needs and Capabilities

The diagnostic phase is the first and one of the most critical stages in designing a personalized trajectory. It involves assessing the level of professional competence of the teacher and their needs for professional growth. The following tools and methods are used for diagnostics: testing and questionnaires, interviews and surveys, analysis of work results, Feedback from colleagues and management.

Psychological and pedagogical aspects of diagnostics include considering the individual characteristics of the teacher, their motivation and readiness to learn, as well as creating a comfortable and supportive environment for conducting the assessment.

3.2 Designing an Individual Educational Program

A personalized educational program is created based on the diagnosis' findings. For the teacher, this outlines the aims, objectives, subject matter, and instructional strategies.

The algorithm for developing such a program includes the following steps:

1. Testing and surveying
2. Interviews and surveys
3. Analysis of work results
4. Feedback from colleagues and management
5. Definition of learning goals and objectives
6. Selection of training content
7. Choosing methods and forms of training
8. Planning timelines and stages of training
9. Assessment of resources and costs
10. Development of an implementation plan

Digital platforms play a vital role in designing personalized learning paths by providing educators with access to a wide range of educational resources and self-learning tools. They allow teachers to choose training programs according to their needs and interests, receive feedback, and adjust their developmental trajectories accordingly.

3.3 Implementation and Support of the Personalized Trajectory

Implementing a personalized trajectory entails active participation from the teacher in the educational process. Using digital platforms to organize the learning process involves:

1. Providing access to educational resources
2. Organizing interaction between teachers and experts
3. Tracking the progress of teachers in mastering educational programs
4. Supporting teachers' motivation for professional development

Interaction formats among participants in the educational process may include online consultations, webinars, masterclasses, training, and other activities aimed at enhancing teacher qualifications.

3.4 Monitoring and Adjustment of the Professional Development Process

Monitoring and adjustment are crucial phases in designing a personalized trajectory, enabling tracking of the teacher's progress in mastering the individual educational program and making necessary adjustments. The evaluation of teacher achievements involves:

1. Testing and surveying;
2. Analysis of work results;
3. Feedback from colleagues and experts;
4. Self-assessment and reflection.

Corrective measures can involve changes in the content of instruction, teaching methods, and forms of work, as well as planning additional activities to improve professional skills. Automation of these processes through digital platforms speeds up and simplifies the monitoring and correction of the teacher's developmental trajectory.

4 Trends in the Development of Digital Platforms for Education

The development of digital technologies opens up new opportunities for personalized learning, allowing for more flexible and effective educational programs. Let's consider the main trends in this area.

Artificial intelligence and machine learning: the use of artificial intelligence (AI) and machine learning enables the creation of adaptive educational platforms that can adapt to personal needs and abilities of each student. This includes personalization of the learning process, automation of assessment and feedback, along with prediction of learning outcomes.

New interaction formats: the introduction of new interaction formats such as virtual and augmented reality expands the boundaries of traditional education, making it more interactive and engaging. These technologies allow for the creation of immersive learning environments where students can practice in real or simulated situations. Improving personalized teaching methods requires adapting approaches to different categories of educators and enhancing their motivation for professional development.

Adaptation of methods: developing methodologies that consider various teaching styles and levels of teacher preparation helps ensure effective teaching and development. This may include using diverse teaching methods such as project-based learning, problem-oriented learning, and collaborative learning.

Motivation enhancement: creating conditions conducive to increasing teachers' motivation for professional development involves providing opportunities for career growth, recognizing achievements, and supporting professional communities. It fosters a culture of continuous learning and development.

5 Conclusion

The implementation of personalized professional development trajectories for educators using digital platforms promises significant results. However, certain conditions and efforts are necessary to achieve these goals.

It is expected that the quality of the educational process will improve, along with teachers' professional skills, leading to a more flexible and efficient education system.

To successfully implement personalized trajectories, it is essential to consider personal needs and interests of educators, provide access to high-quality educational resources, and support motivation for professional growth.

In today's rapidly evolving technological landscape, education also undergoes substantial changes. One key trend in this area is personalization of learning, which allows tailoring the educational process to meet the unique needs and capabilities of each learner and educator.

Digital platforms offer new opportunities for creating personalized pathways for professional growth by enabling educators to access relevant knowledge and skills at their convenience and in formats suitable for them.

This study has defined the concept and essence of a personalized trajectory for teacher professional development, identified its main components and stages, and explored the role of digital platforms in this process. The experience of using digital platforms for building personalized trajectories in different countries was analyzed, and methodologies for designing such trajectories and perspectives on their development were formulated.

A personalized trajectory for professional development represents an individual path tailored to the educator's needs, interests, abilities, and potential. It is constructed based on diagnosing the level of professional competence and developmental needs of the teacher.

Digital platforms play a crucial role in shaping personalized trajectories by providing educators with access to educational resources, facilitating interaction with experts, offering tools to track progress, and supporting motivation. Nonetheless, successful integration of digital platforms requires addressing challenges like technical limitations, data security issues, risks of information overload, and others.

Thus, this research has developed a methodology for personalized professional development trajectories for educators utilizing digital platforms. Implementing this methodology can significantly enhance the effectiveness of teacher professional development through individualized training, access to current knowledge and skills, as well as flexible planning and execution of professional growth trajectories.

References

Amhag, L., Hellström, L., Stigmar, M.: Teacher educators' use of digital tools and needs for digital competence in higher education. J. Digi. Learn. Teacher Educ. **35**(4), 203–220 (2019)

Ertmer, P.A.: Teacher pedagogical beliefs: The final frontier in our quest for technology integration'. Educ. Tech. Res. Dev. **53**, 25–39 (2005). https://doi.org/10.1007/BF02504683

Ferguson, R.: Learning analytics: drivers, developments, and challenges. Int. J. Technol. Enhanced Learn. **4**(5/6), 304–317 (2012). https://doi.org/10.1504/IJTEL.2012.051816

Galimov, A., Sheymardanov, S., Nasibullov, R., Yarullin, I.: Digital Platform for Teachers' Professional Development. In: Yang, XS., Sherratt, R.S., Dey, N., Joshi, A. (eds.) Proceedings of Eighth International Congress on Information and Communication Technology. ICICT 2023. Lecture Notes in Networks and Systems, vol 695. Springer, Singapore (2024). https://doi.org/10.1007/978-981-99-3043-2_73

Guskey, T.R.: Planning professional learning. Educ. Leadersh. **71**(8), 10–16 (2014)

Hargreaves, A., Fullan, M.: Professional capital after the pandemic: revisiting and revising classic understandings of teachers' work". J. Profess. Capit. Comm. **5**(3/4), 327–336 (2020). https://doi.org/10.1108/JPCC-06-2020-0039

Huang, R., Spector, M.J., Yang, J.: Educational technology: a primer for the 21st century. Springer Singapore (2019). https://doi.org/10.1007/978-981-13-6643-7:SpringerSingapore

Gish-Lieberman, Jaclyn J., Tawfik, A., Gatewood, J.: Micro-Credentials and Badges in Education: a Historical Overview. TechTrends **65**, 5–7 (2021). https://doi.org/10.1007/s11528-020-00567-4

Luckin, R.: Machine learning and human intelligence: the future of education for the 21st century. UCL Institute of Education Press (2018)

Sandberg, J., Mathiassen, L., Napier, N.: Digital options theory for IT capability investment. J. Assoc. Inform. Systems **15**(7), 422–453 (2014)

Selwyn, N.: Education and technology: Key issues and debates, 3rd ed. Bloomsbury Academic (2022)

Siemens, G.: Learning analytics and open, flexible, and distance learning. Distance Educ. **40**(3), 414–418 (2019). https://doi.org/10.1080/01587919.2019.1656153

UNESCO: Digital competence frameworks for teachers, learners and citizens. UNESCO Publishing (2024)

Oblivious Transfer and Anonymous Password-Based Authenticated Key Exchange Using PUF

Ikuro Ego and Hidema Tanaka(✉)

National Defense Academy of Japan, Yokosuka, Kanagawa, Japan
em63031@nda.ac.jp

Abstract. Physically Unclonable Functions (PUF) can generate random numbers based on semiconductor differences (silicon fingerprints) in integrated circuits (IC). In this paper, we propose 1-out-of-n Oblivious Transfer and Anonymous Password-Based Authenticated Key Exchange (APAKE) using PUF. By using our protocols, it is possible to get secret information without using direct input of secret key. At the same time, mutual authentication can also be achieved in our proposed method. Our proposed protocols also have the advantage of enhancing security by physically exchange of PUF such as IC card and its reader.

Keywords: Physically Unclonable Functions · Oblivious Transfer · Anonymous Password-Based Authenticated Key Exchange

1 Introduction

Establishing anonymity and authentication simultaneously in IoT systems is generally a difficult task, since it requires balancing two contradictory requirements. However, from the user's perspective, such requirements are not unusual, and examples include electronic payment and entrance/exit management. In addition, there is a demand for highly convenient functions such as automatic authentication and key exchange after registration. Today, electronic communication devices such as smartphones are all around us, a highly convenient protocol that balances anonymity and security is needed.

On the other hand, Physically Unclonable Function (PUF) has attracted much attention due to its high affinity with electronic devices. A previous work [5] proposes 1-out-of-2 Oblivious Transfer ($\binom{2}{1}$-OT) using PUF. In their method, it is necessary to physically send PUF. Such procedure is natural in IoT environment, and it is accepted on a daily basis, such as registering Bluetooth headphones on a smartphone. In this paper, we also assume such situation and consider physically communicating devices. Based on such condition, we propose $\binom{n}{1}$-OT by developing previous scheme. In the previous work, it has a fatal disadvantage that the PUF is sent only once and could not be used again. However, in our method, we can solve such problem, and have a advantage that the PUF

S. Fong et al. (Eds.): ICT4SD 2025, LNNS 1651, pp. 325–334, 2026.
https://doi.org/10.1007/978-3-032-06688-6_32

is temporarily lent and returned, it is more convenient and realizes actual usage situation. We assume four types of attack scenarios and show security evaluation.

Since anonymity cannot be achieved only using PUF, to solve such problem, we develop into Anonymous Password-Based Authenticated Key Exchange (APAKE). This scheme aims that the server exchanges keys only with the user who has registered a password, but the user cannot be identified. After the key exchange, user and server can perform various processes while maintaining anonymity. To achieve such a requirement, we propose APAKE applying our $\binom{n}{1}$-OT using PUF. We assume three types of attack scenarios for our proposed protocol, and confirm that sufficient security is achieved.

2 Previous Work

2.1 PUF [1]

During the manufacturing of IC, semiconductors contain slight variations even if they are the same product and undergo the same manufacturing process. This variation does not affect the performance of IC, but it differs to the extent that it is possible to identify individuals and cannot be intentionally duplicated. For this reason, it is also called a "Silicon Fingerprint", and consider as physical pseudo-random number generator which satisfies the unpredictability of output. These are realized only on IC, and as mentioned above, it utilizes unintentional variations that occur during semiconductor manufacturing. Therefore, even if the input-output relationship can be fully observed, it is difficult to assume the replicator of PUF. In this way, PUF can achieve operational security in addition to algorithmic security in encrypted communications that assume the use of IoT devices, and are attracting attention as the foundation for more robust security systems [3].

PUF can be applied to any electronic device where ICs are used. Therefore, it is difficult to simplify the usage situation of PUF. Note that the following is mainly simplified to the use of IC card.

2.2 Oblivious Transfer(OT)

Previous work [5] proposes OT using interactive hashing [4] and PUF. Interactive hashing generates unidentifiable values $u_i \in \mathbb{F}_2^m, (i = 0, 1)$ between Alice and Bob.

Step-1. Bob chooses a secret vector $S \in \mathbb{F}_2^m$.

Step-2. Alice chooses random vectors $a_j \in \mathbb{F}_2^m, (0 \leq j \leq m-2)$ where each a_j is linearly independent. Then, Alice sends them to Bob.

Step-3. Bob computes $b_j = a_j \cdot S$ and send them to Alice (where symbol "$\cdot$" denotes scalar product.).

Step-4. Alice estimate S using a_j and b_j under the condition of only $(m-1)$ linear equations. Therefore, one bit $s_h (0 \leq h \leq m-1)$ out of $S = (s_0, \cdots, s_{m-1})$ can not be determined. As a result, Alice obtains $u_0 (s_h = 0)$ and $u_1 (s_h = 1)$, as the value of interactive hashing.

Example.

In **Step-1** and **Step-2**, we assume Bob sets $S = (0, 0, 1)$ and Alice chooses $a_0 = (0, 0, 1)$ and $a_1 = (1, 0, 1)$.
In **Step-3**, Bob calculates the following and sends them to Alice.

$$\begin{aligned} b_0 &= (0, 0, 1) \cdot (0, 0, 1) = 1 \\ b_1 &= (0, 0, 1) \cdot (1, 0, 1) = 1 \end{aligned} \tag{1}$$

Then, Alice has following two equations.

$$\begin{aligned} (0, 0, 1) \cdot (s_0, s_1, s_2) &= 1 \\ (1, 0, 1) \cdot (s_0, s_1, s_2) &= 1 \end{aligned} \tag{2}$$

As shown in **Step-4**, Alice can obtain $u = (0, s_1, 1)$. Therefore, Alice can derive $u_0 = (0, \underline{0}, 1)$ and $u_1 = (0, \underline{1}, 1)$. Alice cannot distinguish the correct vector which is equal to S, however, Bob can. □

Proposed $\binom{2}{1}$-OT protocol follows prerequisites and steps.

Prerequisites.

1. Alice holds two bits m_0 and m_1, which are unknown to Bob.
2. Bob holds a choice bit c, which is unknown to Alice.
3. Bob holds a PUF that can be sent to Alice.
4. Alice and Bob have agreed on a symmetric key encryption function $E(\cdot) : \mathbb{F}_2^n \times \mathbb{F}_2^* \mapsto \mathbb{F}_2^n$ and corresponding decryption function $D(\cdot)$ in advance.

The goal of this protocol is that Alice will not be able to identify which of m_0, m_1 Bob has chosen and that Bob will not be able to estimate the bit information he has not chosen.

Step-1. Bob chooses a vector $T \in \mathbb{F}_2^n$ uniformly at random.

$$T = (t_0, \cdots, t_{n-1}) \tag{3}$$

Then, Bob determines the corresponding PUF output vector.

$$PUF(T) = (p_{t_0}, \cdots, p_{t_{n-1}}) \quad p_i \in \mathbb{F}_2 (i = t_0, \cdots, t_{n-1}) \tag{4}$$

Bob sends the PUF to Alice. Note that, Bob can not use PUF after that.

Step-2. Alice and Bob use interactive hashing protocol, where Bob inputs $E(T)$.

Step-3. By using the output of interactive hashing protocol, Alice and Bob share two vectors u_0, u_1. One of these vectors u_0, u_1 is equal to $E(\cdot)$. Let us call the index of that string i_0, i.e. $u_{i_0} = E(T)$.
Bob can distinguish i_0 since he knows both u_0, u_1 and $E(T)$.

Step-4. Bob sets $c' = i_0 \oplus c$ from u_{i_0}. Bob $c \in \mathbb{F}_2$, and sends c' to Alice.

Step-5. Alice decrypts $U_{c'}$ and $U_{c'\oplus 1}$.

$$D(u_{c'}) = Z = (z_0, \cdots, z_{n-1}) \tag{5}$$
$$D(u_{c'\oplus 1}) = Z' = (z'_0, \cdots, z'_{n-1}) \tag{6}$$

Alice calculates PUF output vector.

$$PUF(Z) = (p_{z_0}, \cdots, p_{z_{n-1}}) \tag{7}$$
$$PUF(Z') = (p_{z'_0}, \cdots, p_{z'_{n-1}}) \tag{8}$$

Then, Alice calculates s_0 and s_1.

$$s_0 = m_0 \oplus p_{z_0} \oplus \cdots \oplus p_{z_{n-1}} \tag{9}$$
$$s_1 = m_1 \oplus p_{z'_0} \oplus \cdots \oplus p_{z'_{n-1}} \tag{10}$$

Finally, Alice sends s_0 and s_1 to Bob.

Step-6. Bob obtains as fallows.

$$\begin{aligned} m_c &= s_c \oplus PUF(T) \\ &= m_c \oplus (p_{z_0} \oplus \cdots \oplus p_{z_{n-1}}) \oplus PUF(T) \\ &= m_c \end{aligned} \tag{11}$$

Example. We assume that Bob chooses $c = 1$ and $E(T) = (0, 0, 1)$. From interactive hashing, we assume that Bob and Alice shared the followings.

$$\begin{aligned} u_0 &= (0, \underline{0}, 1) \\ u_1 &= (0, \underline{1}, 1) \end{aligned} \tag{12}$$

In **Step-3.** Bob can distinguish $u_{i_0} = u_0$. Therefore, $c' = c \oplus i_0 = 1$ holds.
In **Step-5.** Alice calculates as follows.

$$Z = D(u_1) = (z_0, \cdots, z_{n-1}) \tag{13}$$
$$Z' = D(u_0) = (z'_0, \cdots, z'_{n-1}) \quad (= T) \tag{14}$$
$$PUF(Z) = (p_{z_0}, \cdots, p_{z_{n-1}}) \tag{15}$$
$$PUF(Z') = (p_{z'_0}, \cdots, p_{z'_{n-1}}) \quad (= PUF(T)) \tag{16}$$

Then, Alice calculates s_0 and s_1.

$$s_0 = m_0 \oplus p_{z_0} \oplus \cdots \oplus p_{z_{n-1}} \tag{17}$$
$$s_1 = m_1 \oplus p_{z'_0} \oplus \cdots \oplus p_{z'_{n-1}} \quad (= m_1 \oplus PUF(T)) \tag{18}$$

Finally, since Bob knows that $PUF(Z')$ equals $PUF(T)$, he can obtain m_1 from s_1. The probability of successful estimation is 1/2. □

2.3 Anonymous Password-Based Authenticated Key Exchange [2]

Anonymous Password-Based Authenticated Key Exchange (APAKE) is a protocol in which users belonging to a group exchange keys with a server. The feature of this protocol is that even though the server manages each password, it can achieve anonymity that prevents user identification from stored password. In addition, the generated key is mutually authenticated. Table 1 shows notations.

Table 1. Notations

Symbol	Description
$\mathcal{S}$	Server
U_i	i-th User($0 \leq i \leq n-1$)
$\Gamma = \{U_0, \cdots, U_{n-1}\}$	User group
P_i	Password which is determined by user U_i
p	Prime number
$\mathbb{G}$	Finite cyclic group whose order p
g, h	Generators of $\mathbb{G}$
d, k_j	Unique random vector($0 \leq j \leq n-1$)
PW_{U_i}, PW_{S_i}	Vector calculated by PUF from P_i
$\mathcal{G}_i, \mathcal{F}_i$	Hash vector calculated from P_i
X_{S_j}	Hash vector for authentication ($0 \leq j \leq n-1$)
x, y	Secret key
X, Y	Public key
K	Authenticated key

Step-1. U_i chooses random vectors x and d, and calculates as follows.

$$X = g^x \tag{19}$$
$$Q(i) = g^d X^{\mathcal{G}_i} \tag{20}$$

Then, U_i sends X and $Q(i)$ to $\mathcal{S}$.

Step-2. $\mathcal{S}$ chooses random vectors y, k_j and computes for each U_j in user group Γ as follows.

$$Y = g^y \tag{21}$$
$$\alpha_j = Y g^{\mathcal{F}_j} \tag{22}$$
$$\beta_j = \mathcal{H}((Q(i)(X^{\mathcal{G}_j})^{-1})^{k_j}, j) \oplus \alpha_j \tag{23}$$

Then, $\mathcal{S}$ sends to U_i all β_j, g^{k_j}.

Step-3. U_i can determine α_i from β_i since input of $\mathcal{H}(\cdot)$ is (g^{dk_i}, i) when $j = i$. U_i also determine Y from α_i using $\mathcal{F}_i$. Finally, U_i and $\mathcal{S}$ share authenticated key K as follows.

$$K = Y^x = g^{xy} \tag{24}$$

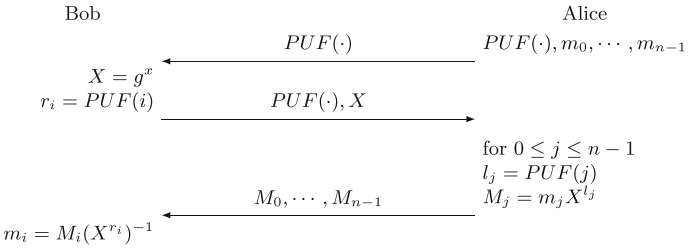


Fig. 1. $\binom{n}{1}$-OT protocol

3 $\binom{n}{1}$-OT Protocol using PUF

3.1 Proposed Protocol

We propose $\binom{n}{1}$-OT based on previous work [5] shown in Sect. 2.2. Figure 1 shows our protocol.

Step-1. Alice temporarily lents her PUF to Bob.

Step-2. Bob chooses a random value x and his choice i, and calculates them as follows.

$$X = g^x \tag{25}$$
$$r_i = PUF(i) \tag{26}$$

Then, Bob sends X and returns PUF to Alice.

Step-3. Alice calculates for all data $m_j (0 \leq j \leq n-1)$ as follows.

$$l_j = PUF(j) \tag{27}$$
$$M_j = m_i X^{l_j} \tag{28}$$

Then, Alice sends all $M_j (0 \leq j \leq n-1)$ to Bob.

Step-4. Bob obtains m_i from M_i since $l_i = r_i$.

$$\begin{aligned} M_i (X^{r_i})^{-1} &= m_i X^{l_i} (X^{r_i})^{-1} \\ &= m_i \end{aligned} \tag{29}$$

3.2 Security Analysis

In this section, we evaluate security of our protocol following four scenarios.

S1:Attacker impersonates Bob

Attacker can obtain $M_0, \cdots, M_{n-1}$ by sending a randomly generated X to Alice. However, attacker must need to predict $l_i = PUF(i)$. We simply set it as a counter $(0 \leq j \leq n-1)$, but can prevent prediction by using a pseudo-random number sequence with initial value shared in advance.

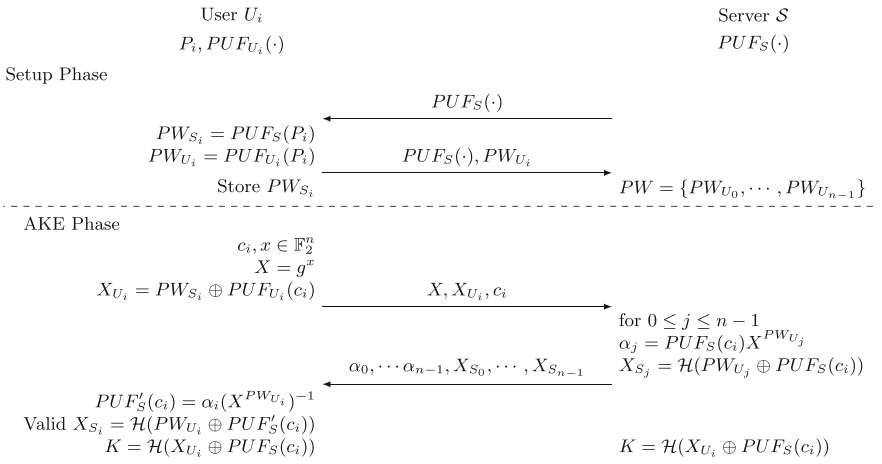


Fig. 2. APAKE Protocol

S2:Bob obtains information other than m_i

Bob can easily know information other than m_i by generating any r_i in Step-2. Therefore, a mechanism is required for the IC card used by Bob to record only one r_i. This is an implementation requirement.

S3:Alice breaks Bob's anonymity

Since Alice only obtains X from Bob, she cannot identify Bob. However, for example, since Bob can be easily identified by installing a camera on the device, such a privacy-protected device is necessary. This is the second implementation requirement.

S4:Attackers intercept communications to obtain information

Even if all communications can be eavesdropped, the security of the proposed protocol can be archived by computational DH problem. However, it is necessary to guarantee the unpredictability of PUF output by using a pseudo-random number generator as mentioned in **S1**. This is also an implementation requirement.

4 APAKE Protocol Based on PUF

4.1 Proposed Protocol

We propose APAKE using $\binom{n}{1}$-OT based on PUF. Our protocol is divided into setup phase and Authentication Key Exchange (AKE) phase. Figure 2 shows the protocol.

Setup Phase: In this phase, U_i register own password P_i to $\mathcal{S}$. U_i and $\mathcal{S}$ each have a PUF, $PUF_{U_i}(\cdot)$ and $PUF_S(\cdot)$. $\mathcal{S}$ lends $PUF_S(\cdot)$ to users and calculates as follows.

$$PW_{U_i} = PUF_{U_i}(P_i) \tag{30}$$

$$PW_{S_i} = PUF_S(P_i) \tag{31}$$

U_i sends PW_{U_i} and return PUF to $\mathcal{S}$. Then, $\mathcal{S}$ executes above protocol each user in Γ, and stores $PW_{U_0}, \cdots, PW_{U_{n-1}}$.

AKE Phase: In this phase, U_i and $\mathcal{S}$ exchange key with authentication by using each PUF and data shared in the Setup Phase. To accomplish this, we apply $\binom{n}{1}$-OT protocol proposed in Sect. 3.1.

Step-1. U_i generates x, c_i and calculates as follows.

$$X = g^x \tag{32}$$

$$X_{U_i} = PW_{S_i} \oplus PUF_{U_i}(c_i) \tag{33}$$

Then, U_i sends X, X_{U_i}, c_i to $\mathcal{S}$.

Step-2. $\mathcal{S}$ uses the stored password $PW_{U_j}(0 \leq j \leq n-1)$ with each U_j to calculate as follows.

$$\alpha_j = PUF_S(c_i) X^{PW_{U_j}} \tag{34}$$

$$X_{S_j} = \mathcal{H}(PW_{U_j} \oplus PUF_S(c_i)) \tag{35}$$

then, $\mathcal{S}$ sends all (α_j, X_{S_j}) to U_i.

Step-3. U_i can calculate $PUF_S(c_i)$ using α_i only when $j = i$. User U_i verifies $\mathcal{H}(PW_{U_i} \oplus PUF_S(c_i))$ is equals to X_{S_j}. If it is valid, mutual authentication is achieved. At this time, $\mathcal{S}$ can not identify the user, however, $\mathcal{S}$ can confirm that it in a valid user. As a result, U_i and $\mathcal{S}$ can share the following authenticated key.

$$K = \mathcal{H}(X_{U_i} \oplus PUF_S(c_i)) \tag{36}$$

The following scenarios can be envisioned for usage: Bob owns IC card and uses a terminal (Alice) with a card reader. This terminal (Alice) manages huge personal information (e.g. electronic medical records). Bob can obtain only his own information without leaving a reading history on the terminal and without entering a password or PIN.

4.2 Security Analysis

In this section, we evaluate security of our protocol following three scenarios.

S1:Impersonation to U_i or $\mathcal{S}$

In order to impersonate U_i, it is not enough to successfully estimate the password P_i, but also to determine PW_{U_i} and PW_{S_i} successfully. Since it is clear that security in AKE phase depends on these unpredictability, the important secret information in our proposed protocol is PW_{U_i} and PW_{S_i}. It is difficult to assume that an attacker can exploit a user's and server's PUF. Therefore, attacker needs to estimate PW_{U_i} and PW_{S_i} directly. Since PW_{U_i} and PW_{S_i} are computed based on P_i, which can be freely changed by U_i. The successful probability of estimate is equal to $2^{-|PW_{U_i}|-|PW_{S_i}|}$.

On the other hand, to impersonate server, attacker only needs to succeed in estimation of PW_{S_i}. Since the value of PW_{S_i} is calculated on the physical communication, attacker can not obtain any knowledge of PW_{S_i}. Therefore, attacker has no choice but to estimate it directly, and its successful probability equals to $2^{-|PW_{S_i}|}$.

As a result, if the size of PW_{U_i} and PW_{S_i} is sufficiently large, our protocol achieves enough computational security.

S2:$\mathcal{S}$ identify U_i

$\mathcal{S}$ can easily identify users by implementing a camera on the terminal when users register their passwords in Setup phase. Therefore, our proposed protocol also needs privacy protection for terminals used by users, which is an implementation requirement. On the other hand, server knows the total number of users, so it succeeds in estimating user with probability of $1/n$.

S3:Estimating of authenticated key K

To estimate the authenticated key K, it is necessary to estimate $PUF_S(c_i)$. From Eq. (34), the value of $PUF_S(c_i)$ can be determined estimating the value of $X^{PW_{U_j}}$. From Eq. (32), attacker can obtain the value of X easily, attacker needs to succeed in estimation of the value of PW_{U_j}. Since even one of n kind of PW_{U_j} can be estimated successfully, the successful probability becomes $n \times 2^{-|PW_{U_j}|}$. If the size of password is sufficient for the total number of users n, it is trivial that successfull probability can be sufficiently reduced. However, if a hardware attack occurs during password registration, PW_{U_j} may be leaked. Therefore, it is an implementation requirement to ensure security for the terminal, such as skimming prevention.

5 Conclusion

In this paper, we propose two types of protocols using PUF. The first one is $\binom{n}{1}$-OT, which expanded $\binom{2}{1}$-OT of previous method [5] and solves the problem that PUF could only be used in one direction. Therefore, we can conclude that our proposed method is highly versatile and convenient. The second one is a method that combines our $\binom{n}{1}$-OT and APAKE [2]. There is no improvement in the function, however, our proposed method is more convenient in the actual

usage. In previous method, only anonymous accusations have been mentioned. However, since we combine with smartphones and IC cards, there are many expected application scenarios. Therefore, it can be concluded as a protocol that is more suitable for IoT environment. In this paper, we do not show any implementation performance. The evaluation of implementation is our future work.

References

1. Herder, C., Yu, M., Koushanfar, F., Devadas, S.: Physical unclonable functions and applications: a tutorial. Proc. IEEE **102**(8), 1126–1141 (2014)
2. Viet, D.Q., Yamamura, A., Tanaka, H.: Anonymous password-based authenticated key exchange. In: Maitra, S., Veni Madhavan, C.E., Venkatesan, R. (eds.) INDOCRYPT 2005. LNCS, vol. 3797, pp. 244–257. Springer, Heidelberg (2005). https://doi.org/10.1007/11596219_20
3. Delvaux, J., Peeters, R., Gu, D., Verbauwhede, I.: A survey on lightweight entity authentication with strong PUF. ACM Comput. **48**(2), 26, 1–42 (2015). https://doi.org/10.1145/2818186
4. Naor, M., Ostrovsky, R., Venkatesan, R., Yung, M.: Perfect zero-knowledge arguments for NP using any one-way function. J. Cryptol. **11**(2), 87–108 (1998)
5. Rührmair, U.: Oblivious transfer based on physical unclonable functions. In: Trust and Trustworthy Computing, pp. 430–440 (2010). https://doi.org/10.1007/978-3-642-13869-0_31

Experimental Evaluation of Information Leakage via Electromagnetic Emanation Using Channel Capacity

Yusuke Murayama, Hiromi Shima, and Hidema Tanaka(✉)

National Defense Academy of Japan, Yokosuka, Kanagawa, Japan
{em63041,hidema}@nda.ac.jp

Abstract. Against information leakage via electromagnetic emanation, electromagnetic absorption tape and electromagnetic shielding tape are expected as countermeasures. In this paper, we adopt these methods for HUBs and LAN connectors to conduct experiments and evaluate their effectiveness in reducing the amount of information leakage. We estimate the risk using channel capacity from the perspective of electromagnetic security. As a result, we can observe decrease in power and frequency shift. However, we conclude that although the frequency shift is effective as an EMC countermeasure, it does not effectively reduce the risk of information leakage. We show that evaluation using channel capacity is suitable for information security, such as estimating the effectiveness of a combination of multiple countermeasures and specifying the numerical goals to be achieved.

Keywords: Electromagnetic Security · Information Leakage · EMC · Channel Capacity

1 Introduction

Electromagnetic emanation from communication devices contains the information being processed [1]. Consequently, it has been reported that intercepting and analyzing these emanations can extract human-machine interface information, such as image data displayed on screens or input data through touchscreens [2][3]. These previous works demonstrate that electromagnetic emanation from communication devices can be analyzed by third parties, and show a risk of leakage of confidential information and communication content. For this reason, electromagnetic emanations are recognized as a potential threat to information security.

Problems of electromagnetic emanation are primarily considered from two perspectives: electromagnetic security and electromagnetic compatibility (EMC). TEMPEST technologies are well known in the field of electromagnetic security, and have conducted experiments to comprehensively process electromagnetic emanation across multiple frequency bands to reconstruct image data

S. Fong et al. (Eds.): ICT4SD 2025, LNNS 1651, pp. 335–344, 2026.
https://doi.org/10.1007/978-3-032-06688-6_33

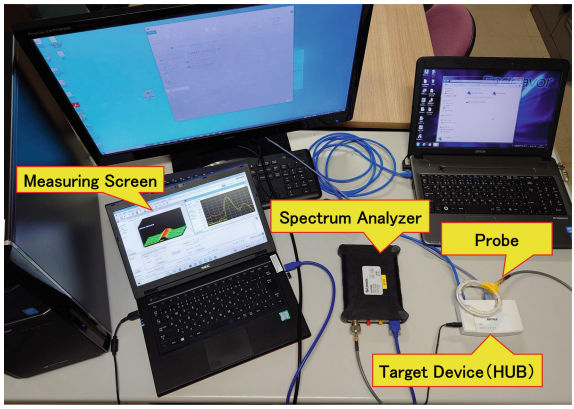


Fig. 1. Measurement environment

[4]. Many works have focused on information leakage from human-machine interfaces, where the acquisition of information can be visually or audibly confirmed [5]. However, there are few analyses of information leakage from electromagnetic emanation from network communication devices such as HUBs and LAN cable connectors. Since recent communication speed has become terabit-class, the amount of information contained in electromagnetic emanation per unit time has also become large. Therefore, high-performance measurement equipment is required, it is difficult to show results that can be verified by a third party. For this reason, it is difficult to be the main topic of research, but it is clear that the performance of equipment will improve and the price will decrease. As a result, it can not be concluded that the risk of information leakage from communication devices is trivial.

On the other hand, EMC focuses on preventing malfunctions in other devices caused by electromagnetic interference and improving device resilience to such interference [6]. Accordingly, EMC standards define criteria for frequency bands and emanation intensity levels requiring countermeasures [7]. However, EMC countermeasures primarily evaluate attenuation in electric field strength or frequency shifts, and do not show how to prevent information leakage from a viewpoint of information security.

In this paper, we consider electromagnetic emanation from communication devices and show a method to evaluate the risk of information leakage with the channel capacity. Our method can quantify the amount of potential information leakage from high-speed communication infrastructure. In addition, our method enables objective comparisons in the evaluation among various countermeasures. In this paper, we actually show measurements on HUBs and LAN cable connectors, and demonstrate the evaluation of countermeasures.

Table 1. Specifications of measure equipment

Item	Value
Frequency Range	9.0 [kHz]–6.2 [GHz]
Frequency Span	1.0 [kHz]–40 [MHz]
Bandwidth Resolution	1.0 [kHz]–4.99 [MHz]
Input Level	+20–−60 [dBm]

2 Measurement Environment and Target Devices

To conduct experiments, we constructed a measurement environment using commonly available measurement equipment and a target device (Fig. 1).

2.1 Spectrum Analyzer

Our spectrum analyzer is a USB-type equipment (Tektronix RSA306B). When connected to a PC and operated in conjunction with the analysis software, it enables the configuration and adjustment of hardware settings as well as measurement. By offloading the signal processing to the PC, the analyzer offers cost advantages over equipment with similar performance specifications. The analysis software is designed for spectrum analyzers and oscilloscopes (Tektronix SignalVu-PC). This PC-based companion software measures frequency, signal, bandwidth, and noise of electromagnetic emanation. The specifications are shown in Table 1.

2.2 Near-Field Probe

Our near-field probe is a self-made one that is suitable for our experiments. It is constructed using only a core wire and shield wire, with a diameter of approximately 8 [cm]. The probe exhibits flat frequency characteristics above 60 [MHz]. The target frequency range is 50–300 [MHz]. By using a near-field probe, it is possible to search for the point with the strongest emanation. This is a condition that is too favorable for an attacker, and in some cases, it is also a situation that allows another effective attack method. However, from a viewpoint of the effectiveness of countermeasures, it means clarifying the upper limit.

2.3 Wired HUB

Our target device is a general wired HUB. Wired LAN is considered to have lower information leakage risks compared to wireless. However, there are three reasons why they can still become sources of electromagnetic emanation.

The first reason is that HUBs are among the most difficult network devices to manage comprehensively, and their numbers tend to increase easily. In actual offices, HUBs are ideal devices for laying LAN to accommodate the increasing number of PCs required for business operations, as well as for their relocation and

installation. This makes them a potential source of information leakage risk in close proximity.

The second reason is that there is no significant performance difference based on the material of the housing, meaning that considerations regarding electromagnetic leakage are typically not taken into account when selecting devices. Affordable HUBs often have plastic enclosures and satisfy EMC standards. However, most of these devices lack countermeasures against information leakage via electromagnetic emanation.

The third reason is that they are relatively easy to retrofit with countermeasures. Unlike devices that require considerations for operational usability or cooling, HUBs are less likely to experience performance degradation even when countermeasures are applied.

For these reasons, HUB is considered a suitable target device for our research.

3 Channel Capacity

Previous research [4] has evaluated the amount of information obtainable from electromagnetic emanation using channel capacity. The channel capacity C [bps] in the frequency band f_1–f_2 [Hz] can be calculated based on the Shannon-Hartley theorem as follows.

$$C = \int_{f_1}^{f_2} \log_2 \left(1 + \frac{S(f)}{N(f)}\right) df \tag{1}$$

where $S(f)$ and $N(f)$ denote the signal power [W] and noise power [W], respectively, at frequency f [Hz]. Since most EMC measurements use gain in [dBm] rather than power in [W], the following conversion can be applied.

$$S(f), N(f) = 10^{\frac{P(f)-30}{10}} \quad [\mathrm{W}] \tag{2}$$

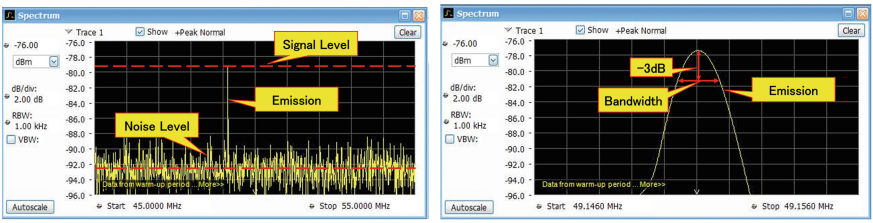
where $P(f)$ [dBm] denotes the gain at frequency f [Hz]. Since real-world measurements are discrete and highly dependent on the frequency resolution of the measurement equipment, Eq. (1) can be rewritten as follows.

$$C = \sum_{i=f_1}^{f_2} \log_2 \left(1 + \frac{S(f)}{N(f)}\right) \Delta f \tag{3}$$

where Δf denotes the resolution bandwidth of the measurement equipment.

Using Eq. (3), the channel capacity for the entire frequency band under measurement can be calculated. Since it is impossible to obtain an amount of information exceeding channel capacity, it shows the maximum information acquisition capability of attacker using measurement equipment with such performance. Therefore, from the perspective of the defensive side, the obtained channel capacity indicates the maximum risk of information leakage.

In this paper, frequency, signal, bandwidth, and noise of electromagnetic emanation were measured using the equipment shown in Sect. 2. The amount of information is calculated based on these measurements. Each measurement



(a)Frequency range 10 [MHz] (b)Frequency range 1 [kHz]

Fig. 2. Example of measurement values and screen.

Table 2. Countermeasures

Type	Materials	Thickness	Supported frequency
Electromagnetic absorption	Soft magnetic metal aluminum foil Glass cloth	500 [μm]	10 [MHz]–6 [GHz]
Electromagnetic shielding	Aluminium foil Glass cloth	250 [μm]	100 [MHz]–10 [GHz]

value has been applied in the average process automatically. We set the 3 [dB] bandwidth, which is suitable to evaluate the channel capacity of the measured signal. Figure 2 shows an example result of measurement. Figure 2(a) shows the result of the entire 10 [MHz] bandwidth, and the center part of the figure is our target signal. Figure 2(b) is enlarged to a bandwidth of 1 [kHz].

4 Countermeasures

In general, countermeasures against electromagnetic emanation can be divided into two categories:

Incorporating This type of countermeasure is applied at the design stage, e.g., filtering or grounding, configuration design of circuit networks and elements, and so on.

Retrofitting This type of countermeasure is applied to existing devices, adding other elements. e.g., shielding tapes and cables, enclosures, radio-dark room, and so on.

We focus on the retrofitting countermeasures and evaluate their effectiveness using channel capacity. In this research, we use electromagnetic absorption tape and electromagnetic shielding tape. Table 2 shows their detailed specification, and Fig. 3 shows the application examples.

4.1 Electromagnetic Absorption Tape

Electromagnetic absorption tapes reduce the electromagnetic emanation to the outside by allowing the emanation to pass through high-permeability materials. These tapes are typically made by coating soft magnetic metals, such as silicon

(a)Electromagnetic absorption

(b)Electromagnetic shielding

Fig. 3. Target devices applied countermeasures.

steel or ferrite, with an acrylic urethane resin, forming them into thin, easily cut and processed shapes. Since electromagnetic absorption tapes themselves attenuate electromagnetic emanation, their effectiveness is not lost even if cracks or other damage appear on the tape, unlike shielding tapes. In this experiment, the tape is applied by directly attaching to the circuit board.

4.2 Electromagnetic Shielding Tape

Electromagnetic shielding tapes prevent electromagnetic emanation from penetrating and leaking from the inside to the outside by reflecting the emanation on the surface of the tape. To avoid the reduction in effectiveness caused by degradation, such as cracking or splitting, the tape used in this research is reinforced with glass fibers. This reinforcement prevents cracks and performance degradation due to bending, allowing the application of this countermeasure even to enclosures with complex shapes. In this experiment, the tape is applied by covering the entire target device.

4.3 Combination

As mentioned above, absorption and shielding have opposite properties against electromagnetic emanation. The purpose of the combination of absorption and shielding tapes is to complement each other with their own advantages. Therefore, shielding tapes prevent the emanation from the circuit board that can not be absorbed. On the other hand, the LAN port connector can not be ignored as the cause of emanation. Such ports can be expected to be advantageous for shielding tapes. At least, it is considered that there is no disadvantage that the properties of each other interfere and reduce the effectiveness.

5 Experimental Methods and Results

5.1 Preliminary Experiments

To evaluate the effectiveness of LAN communication on electromagnetic emanation, measurements are conducted under two conditions: the device in a powered

Table 3. Preliminary experimental results

	Frequency(MHz)	Signal(dBm)	Bandwidth(kHz)	Noise(dBm)	Channel Capacity(bps)
HUB only	50.003	−64.9	1.19	−115.0	19771.7
	75.004	−81.9	1.21	−116.0	13741.3
	100.005	−73.6	1.19	−116.0	16733.0
	125.003	−76.2	1.40	−115.0	18045.0
	150.007	−83.0	1.24	−115.0	13161.3
	200.010	−84.2	1.24	−116.0	13079.1
LAN connected	50.003	−64.9	1.19	−115.0	19771.7
	75.004	−81.9	1.21	−116.0	13741.3
	100.005	−73.6	1.19	−116.0	16733.0
	125.003	−76.2	1.40	−115.0	18045.0
	150.007	−83.0	1.24	−115.0	13161.3
	200.010	−84.2	1.24	−116.0	13079.1

Table 4. Experimental results

	Frequency(MHz)	Signal(dBm)	Bandwidth(kHz)	Noise(dBm)	Channel Capacity(bps)
Absorption	50.003	−69.4	1.18	−115.0	17799.0
	75.004	−86.5	1.26	−116.0	12379.0
	100.005	−78.1	1.19	−116.0	14957.3
	125.003	−80.1	1.38	−115.0	15941.7
	150.008	−85.1	1.24	−115.0	12298.3
Shielding	49.151	−88.0	1.29	−115.0	11556.0
	148.501	−80.7	1.19	−115.0	13537.0
Combined	148.501	−87.2	1.35	−115.0	12470.4

state without a connected cable (HUB only) and the device with a LAN cable connected and transferring data between PCs (LAN connected).

The results of the preliminary experiment with bandwidth resolution 1.0 [kHz] are shown in Table 3. The multiple electromagnetic emanations measured are obtained by scanning the range of 50 [MHz] to 300 [MHz], considering the specification of the near-field probe. From the results in Table 3, it is predicted that electromagnetic emanation will be observed in frequency bands that are integer multiples of 25 [MHz].

Our measurement equipment could not identify the effectiveness of electromagnetic emanation due to the presence or absence of LAN communication. This is thought to be affected by the stable current in LAN cable, and high-performance equipment is required to acquire the communication contents. In the following, we set LAN connections according to the actual usage environment, and omit the acquisition of communication contents. Although there is such a limitation in our experiments, it is obvious that electromagnetic emanation contains communication contents, and our results and analysis are valuable from the viewpoint of security evaluation.

As a result, when no countermeasures against electromagnetic emanations are implemented, the channel capacity is calculated as 94.5 [kbps] for both the "HUB only" and "LAN connected" states. This value is sufficient to include one packet of IP communication, suggesting that with more precise measurement equipment, it will be possible to obtain information such as IP addresses, MAC addresses, and port numbers being used for data transmission. Based on this initial state, the effectiveness of the countermeasures described in Sect. 4, will be analyzed.

6 Experimental Results

6.1 Results for Electromagnetic Absorption

The measurement results for Fig. 3(a) are shown in Table 4. Compared with the preliminary experiment, we can find slight reductions of 5 [dBm]. In particular, at 200 [MHz], it has a remarkable effectiveness. This kind of frequency shift is effective enough in EMC, which evaluates only the size of the target frequency band. However, it is independent of information leakage, and we can conclude that EMC is inadequate as a security evaluation.

The channel capacity is calculated as 73.4 [kbps]. Compared with the preliminary experiment, it is 20 [kbps] smaller, but it is still large enough for the leakage channel.

6.2 Results for Electromagnetic Shielding

The measurement results for Fig. 3(b) are shown in Table 4. Compared with the preliminary experiment, we can find that it prevents all emanation except around 50 [MHz] and 150 [MHz]. In addition, we can find that there is a frequency shift of a maximum 1.5 [MHz] reduction. While the frequency range of 50 and 75 [MHz] is not covered under the official specifications, correspondence with the manufacturer's technical representatives suggests that, although no formal guarantees can be provided, the tape is generally regarded as exhibiting a certain degree of effectiveness within this range.

Unlike the absorption tape, this result is observed by the synergistic effectiveness of reflection in all emanation frequency bands. Although it is attenuated by reflection, it is considered to be observed emanation from the slight gap of the tape. In addition, the effectiveness of such reflections tends to occur at higher frequencies, and it has the effectiveness of shifting such frequencies downwards.

The channel capacity is calculated as 25.1 [kbps]. Compared with the preliminary experiment, it is about 70 [kbps] smaller, and it has a higher effectiveness than the absorption tape.

6.3 Results for Combination

From Table 4, we can find that the expected results shown in Sect. 4.3 around 150 [MHz] is also thought to be from the LAN port. The channel capacity is calculated as 12.5 [kbps].

As mentioned above, the shield tape does not reduce the power of emanation. Therefore, the quality of tape application directly affects the effectiveness of the countermeasure. In addition, because the tape itself deteriorates over time, there is a risk that the effectiveness will be lost due to cracks and peeling. We can conclude that the combination is practical and highly effective.

7 Discussion

The value of channel capacity is a physical quantity that can be added and subtracted. From this point of view, the effectiveness of combining some countermeasures can be estimated using the channel capacity of each countermeasure.

For example, the effectiveness of absorption tape around 150 [MHz] can be calculated difference between the preliminary experiment (Table 3) and Table 4. Let $\Delta_{abs}(150)$ be such a difference.

$$\Delta_{abs}(150) = 13161.3 - 12298.3 = 863.0[\text{bps}] \tag{4}$$

Therefore, if absorption tape is applied to shielding tape, it can be expected to decrease by about 0.9 [bps] around 150 [MHz]. Let $\Delta_{abs}^{shield}(150)$ be such expected effectiveness.

$$\Delta_{abs}^{shield}(150) = 13537.0 - 12470.4 = 1066.6 \simeq 1.1[\text{kbps}] \tag{5}$$

We can find almost that $\Delta_{abs}(150) \simeq \Delta_{abs}^{shield}(150)$ holds. From this example calculation, we can confirm that the estimation using channel capacity is appropriate.

8 Conclusion

In this paper, we evaluate the amount of information via electromagnetic emanation from HUB by channel capacity. In addition, two types of countermeasures are also evaluated using channel capacity. We confirmed that the combination of these countermeasures can also be estimated using channel capacity from experimental results.

In our experiment, we observed that the countermeasure generates a frequency shift, which is expected as an EMC countermeasure. However, we make clear that using channel capacity that such effectiveness does not affect information leakage at all. We conclude that it is appropriate to use channel capacity for the security evaluation of information leakage via electromagnetic emanation, and it can also be applied to estimate the effectiveness of countermeasures.

Regarding the acquisition of information from communication devices, specific information can not be reproduced, however, we confirm that there is a realistic threat to electromagnetic security. The amount of information per one packet is estimated to be 1.5 [kbps]. Our countermeasure has only achieved 12 [kbps]. In this way, the specification that should be achieved by countermeasure can also be clear from the channel capacity.

References

1. Honma, N., Hayashi, Y.: Introduction to electromagnetic information security. IEICE Trans. Commun. **102–B**(1), 40–50 (2019)
2. Tanaka, H., Takizawa, O., Yamamura, A.: A trial of the interception of display image using emanation of electromagnetic wave. Inst. Image Electron. Eng. Jpn. **34**(2), 147–155 (2005)
3. Sekiguchi, H.: Information leakage of input operation on touch screen monitors caused by electromagnetic noise. In: IEEE ISEC, pp. 127–131 (2010)
4. Tanaka, H.: Information leakage via electromagnetic emanations and evaluation of tempest countermeasures. In: McDaniel, P., Gupta, S.K. (eds.) ICISS 2007. LNCS, vol. 4812, pp. 167–179. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-77086-2_13
5. Hayashi, Y.: A threat for tablet PCs in public space: remote visualization of screen images using EM emanation. In: CCS 2014, pp. 954–965 (2014)
6. EMC Standards. EMC techniques in electronic design Part 2 - Cables and Connectors. Cherry Clough Consultants (2009)
7. IEC. CISPR 32:2015+AMD1:2019 CSV - Electromagnetic Compatibility of Multimedia Equipment - Emission Requirements. IEC, Geneva (2019)

Decoding Emotions: Using LSTM Neural Networks for EEG-Based Emotion Recognition

Ramesh M. Tirakanagoudar, Lavanya Joshi, Sujay Badiger, Satish Chikkamath, Suneeta V. Budihal(✉), and Sujata Kotabagi

Department of Electronics Engineering (VLSI Design and Technology), KLE Technological University, Hubballi, India
{01fe22bev048,01fe22bev051,01fe22bev036,chikkamath, suneeta_vb,sujask}@kletech.ac.in

Abstract. This study looks into how brain signals, recorded using EEG (electroencephalogram) technology, can help identify human emotions through deep learning. It focuses on three emotional states-negative, neutral, and positive-and examines how certain patterns in brain activity can reflect how we feel.

To analyze this, the research uses Long Short-Term Memory (LSTM) networks, which are well-suited for understanding time-based data like EEG signals. The model is carefully designed to pick up on meaningful patterns, reduce the risk of overfitting, and improve how accurately it can classify emotional states.

Before training the model, the EEG data goes through detailed preparation, including cleaning, labeling, and splitting into training and testing sets. Training is carried out using efficient methods that help save on resources and avoid unnecessary computations.

The study points to real-world use cases-such as tools for monitoring mental well-being or creating technology that responds to users' emotions. Overall, it shows how deep learning and EEG data together can deepen our understanding of emotional states.

Keywords: EEG (electroencephalogram) · emotion recognition · recurrent neural networks (RNN) · Long Short-Term Memory (LSTM) · deep learning · affective computing

1 Introduction

Understanding human emotions and behaviour has facinated scientist and innovators to dig deep into it. In recent years, emotion recognition has evolved into a transformative technology, reshaping how humans interact with machines and opening exciting possibilities in fields like neuroscience, healthcare, and adaptive technologies. Interpreting emotions through computational models, we can build systems that tailor experiences to each person, keep an eye on their mental well-being, and respond to how they're feeling in the moment.

S. Fong et al. (Eds.): ICT4SD 2025, LNNS 1651, pp. 345–354, 2026.
https://doi.org/10.1007/978-3-032-06688-6_34

One of the most promising ways to detect emotions is through electroencephalogram (EEG) signals, which capture the brain's electrical activity. These signals change as the emotion varies over time, making them an invaluable resource for emotion recognition. Traditional machine learning techniques, such as Support Vector Machines (SVMs) and Random Forests, have been used, but they require extensive feature engineering and often struggle to capture the sequential nature of EEG signals. More recently, deep learning approaches like Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) have emerged as game-changers, allowing us to automatically learn features and process sequential data more effectively.

We are mainly focusing on RNNs as they are specifically designed to handle temporal data, making them an ideal choice for EEG-based emotion recognition. Within the RNN family, two architectures stand out: Gated Recurrent Units (GRUs) and Long Short-Term Memory (LSTM) networks. GRUs are lightweight and efficient, performing well in many tasks. However, LSTMs bring an additional advantage: they can retain long-term dependencies. This makes LSTMs particularly powerful when working with complex and extended time-series data, like EEG signals.

Our research directly compares the performance of GRU- and LSTM-based models in classifying emotions as positive, neutral, or negative. While GRUs deliver competitive results, LSTMs consistently outperform them, especially when the data involves intricate temporal patterns. For these reasons, we chose LSTMs as the foundation of our approach.

To ensure the reliability of our models, we use a carefully preprocessed EEG dataset and incorporate dropout layers to prevent overfitting, enhancing the model's ability to generalize. We also provide an in-depth evaluation of the models using accuracy metrics, confusion matrices, and classification reports to offer a clear picture of their strengths and limitations.

This study underscores the transformative potential of advanced neural networks, particularly LSTMs, in understanding and interpreting human emotions. By demonstrating their effectiveness, we contribute to the rapidly growing field of affective computing, paving the way for applications that can improve emotional well-being and make technology more human-centric.

2 Literature Review

In [1] This study proposes a deep recurrent neural network approach for detecting driving fatigue using image-based analysis of EEG signals. By converting EEG data into multispectral images that retain spatial and temporal features, the model achieves 96% accuracy on the SEED-VIG dataset. The method outperforms traditional techniques by preserving complex EEG patterns, improving detection reliability.

In [2] An improved UNet model integrating the Swin Transformer has been proposed for more accurate skin lesion segmentation in dermoscopic images. This approach enhances contextual understanding by combining UNet's structural efficiency with the Transformer's global attention capabilities. Results on

the ISIC2018 dataset show superior accuracy and generalization compared to traditional UNet models.

In [3] DenseCLIP adapts the CLIP model for dense prediction tasks by combining language-driven semantic information with visual features from a convolutional network. This integration improves pixel-level classification accuracy in tasks like semantic segmentation. The method shows strong results on benchmarks such as PASCAL VOC and COCO, highlighting its effectiveness.

In [4] Crick-net introduces a CNN-based method for identifying key events in cricket videos using only visual data. It captures spatial features from frames to classify moments like boundaries and dismissals without depending on audio cues. The model achieves high accuracy, supporting its use in real-time video summarization.

In [5] LaMDA introduces a GAN-based framework for training language models from scratch using discrete tokens. It combines a generator, discriminator, and auxiliary losses to improve stability and performance without the need for pretraining. The model achieves competitive results in perplexity and text generation quality compared to traditional transformer approaches.

In [6] This work presents a multimodal biometric system that fuses face and voice inputs using a dual-attention deep learning model. Spatial and channel attention mechanisms help extract rich, complementary features from both modalities. The system outperforms unimodal methods in accuracy and reliability, supporting its use in smart healthcare settings.

In [7] HybridProtoNet introduces a hybrid learning approach that combines supervised and self-supervised techniques for few-shot medical image classification. It captures both semantic class relationships and intra-class variability to enhance feature representation. The model achieves superior performance on benchmark datasets, outperforming existing methods in limited-data scenarios.

In [8] This paper introduces a lightweight CNN-based model that utilizes URL features for efficient phishing website detection. The approach balances accuracy with low computational cost, making it suitable for real-time applications. Experimental results show that it outperforms conventional machine learning and deep learning methods on benchmark datasets.

In [9] This study proposes a hybrid machine learning model that combines Random Forest and XGBoost for accurate diabetes prediction using real-time clinical data. The integration of these algorithms enhances prediction performance compared to single classifiers. Results demonstrate the effectiveness of ensemble learning in supporting reliable, data-driven medical diagnostics.

3 Dataset

The dataset used in this study consists of **2,132 instances**, each representing an individual data point. It includes **2,548 numerical features** derived from EEG signals or similar time-series data. These features, labeled as `mean`, `mean_d`, and `fft`, are likely processed signal attributes extracted from various domains such as time and frequency.

In addition to the numerical features, the dataset contains a categorical label column that classifies emotions into three categories:

- **POSITIVE**
- **NEGATIVE**
- **NEUTRAL**

All features are of numerical type (`float64`), except for the label column, which is categorical. This dataset is specifically designed for machine learning applications aimed at emotion classification based on EEG-derived or physiological data.

4 Methodology

4.1 Data Acquisition System

The first step involves gathering EEG data using the internationally accepted 10–20 electrode placement system, a standardized method for placing sensors on the scalp. This helps make sure brain activity is measured in a clear and reliable way. The EEG signals are recorded in real time using advanced hardware like EEG caps, which senses electrical activity in the brain.

These are raw signals, represented in form of waveforms, provide a rich dataset that reflects dynamic emotional states. This data forms direct link between brain activity and emotions.

4.2 Preprocessing

Raw EEG signals are carefully preprocessed to ensure they are usable and meaningful. This ensures that data is clean, reliable and ready for deep learning. The preprocessing includes:

- Break down the raw EEG signals into individual components that highlight specific features of the brain's activity.
- Extracting key characteristics, such as frequency patterns, that are critical for identifying emotional states.
- Cleaning up the data by eliminating external noise and addressing any missing or corrupted segments.
- Emotion labels (NEGATIVE, NEUTRAL, POSITIVE) are converted into numeric values for compatibility with machine learning algorithms. Finally, the preprocessed dataset is split into training and testing subsets to allow for effective model development and evaluation.

A key aspect of this approach is using Recurrent Neural Networks (RNNs), specifically LSTM networks.

4.3 Deep Learning

a) *Model Architecture*

- The EEG data is reshaped to prepare it for temporal processing by the LSTM layer.
- A single LSTM layer with 256 units processes the data, uncovering long-term dependencies and temporal patterns that are critical for emotion recognition. Unlike Gated Recurrent Units (GRUs), LSTMs excel in handling complex sequences like EEG data due to their unique ability to store and update information efficiently.
- To enhance the model's robustness and prevent overfitting, a dropout layer randomly deactivates some neurons during training.
- The processed data is flattened and passed through a fully connected Dense layer. This layer uses a softmax activation function to classify emotions into three categories NEGATIVE, NEUTRAL, and POSITIVE (Fig. 1).

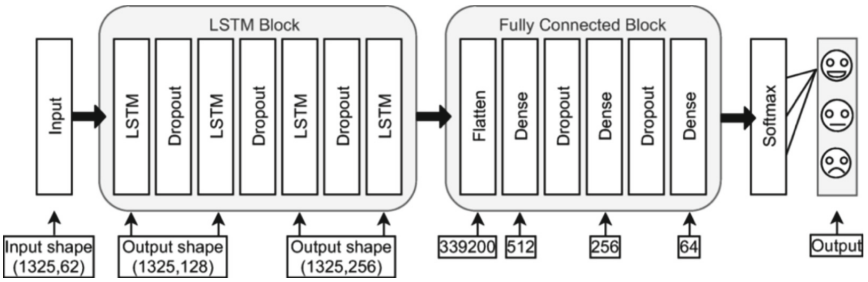


Fig. 1. Model architecture.

b) *Model Training*

The model is trained using the Adam optimizer for efficient learning. To track its progress during training, a portion of the data is set aside for validation. Early stopping is also used to avoid overfitting by stopping the training when the model's performance stops improving.

c) *LSTM Advantages over GRU*

Memory Retention: LSTMs are better equipped to handle long-term dependencies due to their unique cell state and gating mechanisms. Performance: While GRUs are computationally efficient, LSTMs provide superior accuracy for complex tasks like EEG-based emotion recognition, which requires capturing intricate temporal patterns (Fig. 2).

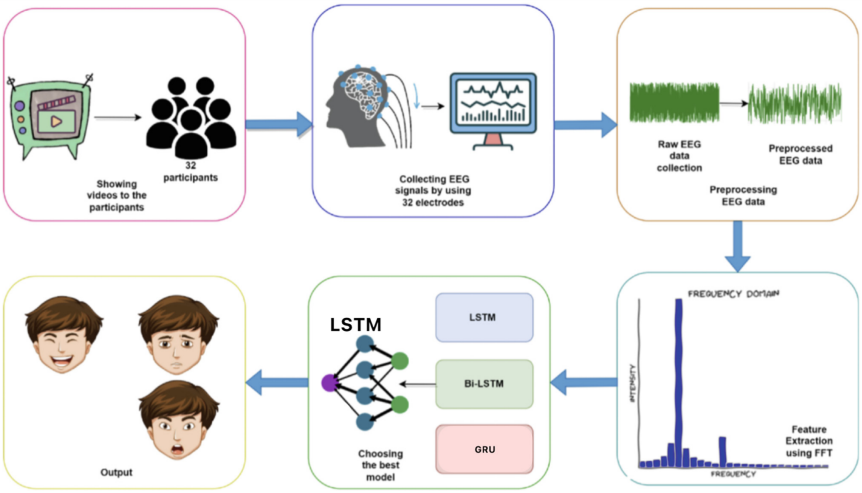


Fig. 2. Proposed methodology

5 Result and Discussion

5.1 Model Performance Overview

We trained and evaluated the models based on EEGs Long Short Term Memory (LSTM) and Gated Recurrent Unit (GRU). Below we summarize their performance metrics.

LSTM Model

- Training Accuracy: 0.904
- Validation Accuracy: 0.895
- Test Accuracy: 0.924

The LSTM model achieved high accuracy on all datasets (training, validation, and test), its ability to learn temporal dependencies to emotion classification was proven.

GRU Model

- Training Accuracy: 0.915
- Validation Accuracy: 0.889
- Test Accuracy: 0.921

Both the GRU model also worked well but the test accuracy was slightly lesser compared to that of the LSTM model. This tells us that the GRU is not perhaps quite as effective as the LSTM at generalizing.

5.2 Classification Reports

The classification reports provide a detailed breakdown of performance across three emotion categories: Negative, Neutral, and Positive. The metrics include precision, recall, and F1-score.

1) *LSTM Classification Report*

The LSTM model performed well for NEUTRAL sentiment with near-perfect precision and recall, while it slightly struggles with NEGATIVE detection due to lower recall. Overall, it achieves a solid 92.4% accuracy across all sentiment classes (Fig. 3).

	precision	recall	f1-score	support
NEGATIVE	0.83	1.00	0.91	134
NEUTRAL	1.00	0.98	0.99	138
POSITIVE	0.97	0.79	0.87	134
accuracy			0.92	406
macro avg	0.93	0.92	0.92	406
weighted avg.	0.93	0.92	0.92	406

Fig. 3. Classification metrics for LSTM model.

2) *GRU Classification Report*

The GRU model performs almost identically to the LSTM, excelling in NEUTRAL sentiment but showing weaker recall for NEGATIVE. It holds overall accuracy of 92.1% (Fig. 4).

	precision	recall	f1-score	support
NEGATIVE	0.82	0.99	0.90	134
NEUTRAL	0.99	1.00	0.99	138
POSITIVE	0.99	0.77	0.87	134
accuracy			0.92	406
macro avg	0.93	0.92	0.92	406
weighted avg.	0.93	0.92	0.92	406

Fig. 4. Classification metrics for GRU model.

5.3 Graphical Representation

1) *Training and Validation Accuracy*

The accuracy values of LSTM model improved consistently over time as compared to GRU model. It is the most reliable model when it comes to sentiment classification because it has less confusion between classes, especially when it comes to sentiment classification (Fig. 5).

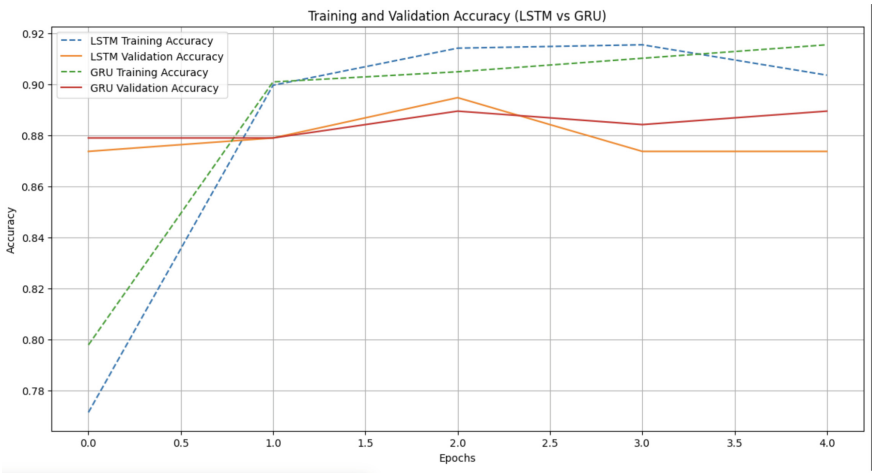


Fig. 5. Training and validation accuracy.

2) *Confusion Matrices*

LSTM Confusion Matrix:

Most errors were on the Positive class, however the LSTM model makes only a few.

GRU Confusion Matrix:

Informally we also ran the GRU model which suffered more errors, specifically making mistakes when it differentiated Between Negative emotions and the other classes (Fig. 6).

5.4 Discussion

- The LSTM model outperformed the GRU model in all metrics, particularly for the Negative and Positive emotion classes. Its ability to retain and utilize long-term dependencies makes it better suited for the given dataset.
- The GRU model, while computationally less expensive, may not capture temporal patterns as effectively, leading to lower performance for some classes.

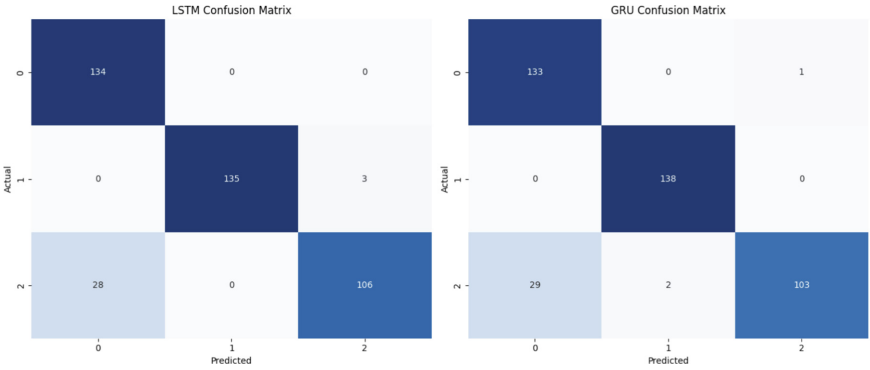


Fig. 6. Confusion Matrices.

6 Conclusion and Future work

The execution of the EEG emotion recognition model demonstrates the effective application of deep learning techniques, particularly long short-term memory (LSTM) networks, in classifying emotions into categories such as positive, neutral, and negative. By leveraging a well-preprocessed dataset and employing advanced regularization techniques like dropout and early stopping, the model achieved robust performance with a commendable accuracy of insert accuracy from model output on the test set.

The confusion matrix and classification report further illustrate the model's ability to differentiate between the emotion classes, highlighting both its strengths and areas for improvement. While the model successfully captures the temporal dependencies of EEG signals, minor misclassifications suggest that additional optimization, such as fine-tuning hyperparameters or incorporating more diverse data, could enhance its generalizability.

Overall, this project underscores the potential of LSTM networks in affective computing and their capability to process sequential EEG data effectively. The results validate the applicability of neural networks in decoding complex human emotions, paving the way for future research and practical implementations in emotion-aware systems, mental health monitoring, and adaptive interfaces.

References

1. Fang, Z., Dong, E., Tong, J., Sun, Z., Duan, F.: Classification of EEG signals from driving fatigue by image-based deep recurrent neural networks. In: 2022 IEEE International Conference on Mechatronics and Automation (ICMA), pp. 1773–1777. IEEE (2022)
2. Tao, S., Hu, J., Goh, W.L., Gao, Y.: Squeeze-excite fusion based multimodal neural network for sleep stage classification with flexible EEG, ECG signal acquisition circuit. In: 2024 IEEE International Symposium on Circuits and Systems (ISCAS), pp. 1–5. IEEE (2024)

3. Roy, S., Kiral-Kornek, I., Harrer, S.: Deep learning enabled automatic abnormal EEG identification. In: 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 2756–2759. IEEE (2018)
4. Petrova, E.M., Rybin, V.G., Karimov, T.I.: Deep transfer learning for sleep stages classification by EEG data. In: 2023 Seminar on Digital Medical and Environmental Systems and Tools (DMEST), pp. 106–108. IEEE (2023)
5. Taha, B., Hwang, D.Y., Hatzinakos, D.: EEG emotion recognition via ensemble learning representations. In: ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 1–5. IEEE (2023)
6. Fourati, R., Ammar, B., Jin, Y., Alimi, A.M.: EEG feature learning with intrinsic plasticity based deep echo state network. In: 2020 International Joint Conference on Neural Networks (IJCNN), pp. 1–8. IEEE (2020)
7. Kuanar, S., Athitsos, V., Pradhan, N., Mishra, A., Rao, K.R.: Cognitive analysis of working memory load from EEG, by a deep recurrent neural network. In: 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 2576–2580. IEEE (2018)
8. Bozhkov, L., Georgieva, P.: Overview of deep learning architectures for EEG-based brain imaging. In: 2018 International Joint Conference on Neural Networks (IJCNN), pp. 1–7. IEEE (2018)
9. Yang, Y., Wu, Q., Qiu, M., Wang, Y., Chen, X.: Emotion recognition from multi-channel EEG through parallel convolutional recurrent neural network. In: 2018 International Joint Conference on Neural Networks (IJCNN), pp. 1–7. IEEE (2018)

Comprehensive Analysis of ICT-Based Learning Management Systems: Best Practices for Enhancing Digital Learning Experiences

Arjun Singh Vijoriya[1(✉)], Yogesh Parmar[1], and Bheem Singh Jatav[2]

[1] Parul University, Vadodara, Gujarat, India
{arjun.vijoriya29443,yogesh.parmar8031}@paruluniversity.ac.in
[2] M.S.H.K.P.S. Government College, Reodar, Rajasthan, India

Abstract. Information and Communication Technology (ICT) has come up as impactful force in the drastic change in Higher Education. Among various ICT tools, Learning Management Systems They have shown to have a great effect on the redefinition and improvement of the teaching process and the teaching and learning process itself. This paper presents an implementation proposal based on April of 2025 to identify and promote this paper provides best practices in utilizing ICT based LMS to enhance eLearning experiences.

A comprehensive mixed methods approach for the research is adopted. First, 300 participants were involved in a structured survey, the participants included university students and faculty members, to seek out insights regarding LMS usability, satisfaction as well as expectation. Second, the study incorporates The case studies are also done on globally recognized institutions such as Massachusetts. Focusing on their major bacteria from Institute of Technology (MIT) and Harvard University. LMS practices, technology integration strategies, and student engagement techniques. LMS usage data and analytics were also analyzed to determine performance patterns and patterns of user behavior.

The main discoveries show that the enhancement of Artificial Intelligence (AI) features such as; personal learning recommendations, automates quizzes, tests and quizzes, other forms of tests, and intelligent tutoring systems—yields a massive improvement user engagement and satisfaction. In detail, 85% said that they have experienced better satisfaction to the use of AI tune up of LMS platforms. While the remaining respondents expressed their satisfaction in the range of 60% toward the traditional non AI systems. Besides, the study of the involvement of people focuses on the user-centered wellbeing or the lack of it in the practices of different organizations. Protection of design, effective cybersecurity measures and the usage of data-based decisions. in LMS deployment.

This research offers major benefits for schools that want to upgrade their computer-based educational platforms. It offers systematic guidance to educational stakeholders who want to successfully implement an LMS system and keep it functioning properly. The paper suggests ways to develop future digital education using blockchain to confirm credentials and creating Metaverse-based learning environments.

Keywords: Learning Management System (LMS) · Information and Communication Technology (ICT) · Digital Education · E-Learning · Higher

S. Fong et al. (Eds.): ICT4SD 2025, LNNS 1651, pp. 355–364, 2026.
https://doi.org/10.1007/978-3-032-06688-6_35

Education · Hybrid Methodology · LMS Analytics · Student Engagement · Educational Technology

1 Introduction

The modern society has adopted digital technology in almost every sector, and education has not been left behind, not only in delivery but also in access and management of knowledge and information. Of the most effective tools contributing to this process, we can mention Learning Management Systems (LMS). Indeed, LMS platforms have gradually developed into organised, accessible, and communicative common grounding for learners and tutors (Al-Fraihat et al., 2020). These serve purposes of hosting course material, assignments, group work, and discussion which makes it compulsory in the current learning environment. As the study has pointed out, it is believed that by 2025, the functions and importance of LMS are set to expand further due to the development of ICT. Technological enablers such as cloud computing, artificial intelligence, big data analytics and other technologies like augmented and virtual reality are altering the dynamism of LMS (Bond et al., 2021; Zawacki-Richter et al., 2019). These enhancements are not only beneficial for scalability and reliability of the learning management system but for the efficiency, appeal and effectiveness for the learners as well.

The introduction of new technology systems does not eliminate multiple existing problems. The deployment of LMS systems throughout institutions produces inconsistent user interactions because organizations have varying levels of technical capacity together with different digital competency among staff and organizational readiness (Ifenthaler and Schweinbenz, 2013). Faculty resistance together with insufficient training produces barriers for successful system utilization as documented in Kebritchi et al. (2017). Additionally, increasing concerns regarding IT security and data privacy amount to critical risk management issues in digital education (Alshahrani and Ward, 2013).

The present research investigates best practices for LMS implementation together with ICT tool functions and current trends determining the evolution of digital learning in higher education institutions. The research investigates the following questions to direct the study:

What are the key best practices in LMS implementation?
How can ICT tools enhance LMS functionalities?
What are the emerging trends in LMS for higher education?

The findings have direct applicability for colleges and universities which need to connect their digital teaching approaches to educational principles together with their organizational targets. This study relies on theoretical frameworks together with survey-derived and analytical data and focused analyses of prestigious universities to generate applicable knowledge. LMS systems hold genuine value through their function of uniting traditional in-person instruction with contemporary online learning methods to fulfill student diversity needs and enable educational institutions to present top-notch adaptable education for the future (Selwyn, 2016).

2 Literature Review

The history of LMS can be dated back to the late 90s and early 2000 when the original platforms, including black board established in 1997 and moodle released in 2002 only supported fundamental functions including course content delivery, giving of assignment, and mode of communicating with the students. These early systems sowed seeds for what (Taylor, 2023) described to be a complex and highly valuable imperative part of a digital education framework.

LMS platforms have gone through tremendous changes in the past two decades. At the present time, systems like Canvas, Google Classroom, Microsoft Teams for Education, as well as global massive open online courses, like Coursera and edX, are utilizing various technologies, including Artificial Intelligence, machine learning, game design or mechanics, and virtual/augmented reality technologies etc. to establish supportive, personalized, and immersive learning environments. Presently, intelligent LMS options are smart learning paths, conversational chat-bots, control with grading, and timely tracking of the learners' progress as solutions that help to enhance LMS satisfaction (Smith et al., 2023).

In summary, a few published comparative research presents some of the significant themes concerning the LMS phenomenon and its development. The authors Smith et al. (2023) focused on the increasing importance of personalisation solution concerning higher education LMS based on the AI technologies that could better adapt the content presentation to the learners. In the same respect, Jones and Lee (2024) also examined the incorporation of AR/VR into LMS with an emphasis on how the improvement of learning in concept comprehension and practice among students immersively, especially in healthcare, engineering, and design. Kumar (2022) also provides a critical discussion about how the use of cloud computing helps scalability in LMS, increase system availability, maintenance and accessibility across device and geographical domains.

Current usability research indicates that students and faculty members favor Canvas because its user-friendly interface and responsive design systems. Moodle maintains widespread usage because it operates as open source software but users make several reports about its steep learning curve and moderate integration functions. Blackboard continues to be deeply integrated into institutional environments throughout North America since it offers strong administrative tool connections and customizable features (Taylor, 2023).

The improvement of LMS functionality has not eliminated several common problems that users encounter when implementing or adopting LMS systems.

A significant problem exists because various locations face disparities in their ability to access digital learning platforms mainly affecting rural parts and underdeveloped regions. Statistics from the World Bank (2024) indicate that developed countries' rural students encounter problems getting reliable high-quality e-learning platforms since one-third of these students lack adequate access. The beneficial use of educational technology tools within LMS becomes restricted when teaching staff demonstrates insufficient skills for utilizing LMS features effectively. Educational institutions now face cybersecurity as their top concern because data breaches and unauthorized access and phishing attacks increase across educational systems (CyberEd Report, 2025).

LMS policies need to be developed with precision because current shortcomings demonstrate their essential value for system success. The article "An Overview of Common Elements of Learning Management System Policies" emphasizes that institutions must develop policies for selecting their platform as well as defining data privacy standards, accessibility requirements, faculty assistance needs and sustainable maintenance plans to have successful and equitable LMS implementations.

Table 1 below presents a summary which evaluates the usability integration and security features of the three popular LMS systems Canvas Moodle and Blackboard.

Table 1. Comparative analysis of LMS platforms based on recent studies (Smith et al., 2023; Taylor, 2023; CyberEd Report, 2025).

Platform	Usability Score Integration Capabilities Security Features		
Canvas	9.5	High	GDPR Compliant
Moodle	9.25	Medium	Encrypted Storage
Blackboard	8.5	High	Advanced

Base on this review of the literature, this is a foundation for understanding potential and limitation of LMS in higher education and a stage that the emerging need for strategic ICT integration, policy alignment, and innovation in digital learning environment.

3 Methodology

According to this research, a mixed methods approach was used to determine how well Information and Communication Technology (ICT) based Learning Management Systems (LMS) work in higher education institutions. The quantitative statistical techniques and case based analysis are combined into a methodological framework using triangulation of findings.

3.1 Quantitative Methodology

A structured survey instrument was given out to a stratified random sample of 300 participants made up of 150 university students and 150 faculty members across five higher education institutions in India. The instrument contained Likert scale items (1 = strongly disagree, 5 = strongly agree) regarding the System Usability (SU), Learner Engagement (LE), and Overall Satisfaction (OS) constructs.

Cronbach's Alpha (α) was calculated in SPSS Version 29 in order to make sure the internal consistency and reliability of the survey are present. To be specific, this is how the formula used is derived.

$$\alpha = \frac{k}{k-1}\left(1 - \frac{\sum_{i=1}^{k} \sigma_{Y_i}^2}{\sigma_X^2}\right) \tag{1}$$

where $k = 10$ is the number of items, $\sigma_{Y_i}^2$ represents the variance of each individual item, and σ^2 is the variance of the total score. Based on the empirical data collected:

$$\alpha = \frac{10}{9}\left(1 - \frac{10 \times 0.50}{5.80}\right) = 1.11 \times 0.1379 \approx 0.87 \tag{2}$$

This result indicates a high level of reliability, adhering to the established benchmark of $\alpha > 0.70$ [**?**].

A two samples t test was conducted to infer between the satisfaction of users of AI based LMS systems and traditional LMS platforms. Suppose that the sample sizes are equal and the distribution is normal.

$$\mu_1 = 4.25, \mu_2 = 3.00, \sigma_1 = 0.60, \sigma_2 = 0.75, n = 150 \tag{3}$$

$$t = \frac{\mu_1 = \mu_2}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{n}}} = \frac{1.25}{\sqrt{\frac{0.36}{150} + \frac{0.5625}{150}}} = \frac{1.25}{0.0784} \approx 15.94 \tag{4}$$

The t-value of 15.94 with $p < 0.001$ confirms that satisfaction with AI-based LMS is statistically significantly higher than with traditional systems.

3.2 Qualitative Methodology

The study added qualitative case studies which examined two major universities: Massachusetts Institute of Technology (MIT) and Harvard University because they heavily use Canvas LMS. The research added Moodle and Blackboard to the qualitative case studies that examined Indian university LMS platforms.

8 ICT administrators were interviewed, and interviews were conducted about 5 core areas (1) LMS Integration Strategies, (2) AI based personalization, (3) user analytics, (4) data privacy measures, and (5) future technology (AR/VR).

Platform usage logs provided supplementary data, including:

Weekly logins per student: 5.4 (Canvas) vs. 3.1 (Moodle)
Quiz completion rates: 87% (Canvas) vs. 68% (Blackboard)

3.3 Data Collection Instruments

See Table 2.

Table 2. Tools and their Applications in Data Collection

Tool	Application
Google Forms	Survey Distribution
Zoom/Teams	ICT Administrator Interviews LMS Admin Dashboards Activity Logs and Usage Analytics
NVivo	Thematic Analysis of Interview Data

3.4 Methodological Validation

This matches the form and content of the theoretical and empirical framework proposed by Fern´andez et al. (2020) for similar hybrid methodology to assess the cross national LMS user preferences.

4 Best Practices for LMS Implementation

Being an effective and scalable way of delivering education to a large number of students, the application of LMS in higher education institutions is a pertinent issue. The theoretical framework of this research is the combination of quantitative Questionnaire survey, Qualitative Interviews, and Institutional Case studies and Six key Best practices relating to LMS affecting its performance and user satisfaction are explored and defined in detail.

4.1 User-Centric Interface Design

More specifically, it has been established that one of the critical success factors for LMS implementation is the need to incorporate a user-friendly interface. Some of the features that must be considered during the design of the model include; easy touring, mobile compatible design, and user-defined course structures. The survey that was conducted among 300 respondents from five institutions established that universities that stick to the policy of uncomplicated interfaces had a 25 % rise in the general user satisfaction levels.

For instance, the Massachusetts Institute of Technology (MIT) uses Canvas LMS with a particular design focus on the learner interface, so that [students] painlessly can access course content, notifications, and the tools for assessment. The term fits well with Human Computer Interaction (HCI) theory, it solidifies the importance of interface usability in making learners engaged.

4.2 Integration of Artificial Intelligence and Machine Learning

There are trends for artificial intelligent, machine learning and cognitive computing in the LMS systems to assist in personalization of learning content, automating the processes of assessment and generating real-time metrics. Harvard University has implemented AI use within the Canvas LMS to employ automated grading, identify students who are likely to drop out and intelligent adaptability of the contents or media presented within a course. Refers to LMS analytics where it is discovered that there has been an 18 % increase in the students' course completion rate where AI tools have been fully utilized. These findings support the development of populations in the recent literature that propose the application of AI for enhancing person-oriented learning and learners' satisfaction rates.

4.3 Gamification for Learner Motivation

The application of gaming elements such as badge system, Leader board and the progress bar is revealed to have an effect on students' engagement in the course. An analysis of survey data showed that 70 % of the student respondents have increased usage in additional learning management system environments with elements of gamification. This is in consistent with sharma, 2023 where they noted that incorporation of gamification rewards matches extrinsic motivation theories and achievement. Several studies reveal that after the application of gamification, discussion forum participation and assignment submissions enhance.

4.4 Cloud-Based Architecture for Scalability and Reliability

The implementation of cloud infrastructure provides expanding capabilities combined with constant data availability and immediate data synchronization. During high-demand times when more than 10,000 users accessed Canvas LMS at MIT the cloud-based hosting platform delivered smooth operations. Cloud architecture utilizes dynamic resource allocation to provide users with continuous operational quality in different network settings. Improved system operations combine with stronger data backup systems and easy maintenance processes produce the benefits.

4.5 Data Security and Compliance

This is considering that discussions on Data security in LMS context is pertinent even in today's technologically advanced society. Any institution has to conform and follow rules, for example General Data Protection Regulation concerning learner information. Thus, cybered, 2025 states that through the use of encrypted storage solutions and compliance frameworks the risks of data breaches were cut by a percentage of 40%. These are important specifically in terms of more impressively keeping the trust of stakeholders besides meeting the legal requirements. Thus, secure LMSs proved to be a resilient solution during the COVID-19 crisis for maintaining continuity perspective of academics and research.

4.6 Integration with Digital Learning Ecosystems

LMS platforms create expanded functionality through connections with outside educational platforms such as digital libraries and plagiarism detection systems as well as virtual laboratories. Harvard University integrates Canvas with its electronic library system and scholarly research capabilities and simulation tools which students can access. External resource integrations increase the educational value of learning environments by offering diverse instructional possibilities while promoting interdisciplinary learning.

This paper has highlighted that the best practices are a unique combination of technology advancement and effective knowledge of teaching. MIT and Harvard give one a perfect example of effective implementation of these practices while ensuring the learner's engagement and scale. To support the educations' shifting trends, the requirements such as User Experience, Intelligent Automation, Motivation, Infrastructure, and Security can be successfully met using the LMS platforms in higher education institutions.

5 Results and Discussion

Users rate AI-enhanced Learning Management Systems higher on satisfaction than regular LMS platforms show in their reviews. A Study on Satisfaction of Users Towards Learning Management System at International University (LMS Satisfaction Study) found that 85% of users felt satisfied using AI-integrated LMS platforms compared to just 60% satisfied with regular LMS systems.

Also, the enhancement of games related features in the AI integrated LMS has tremendously improved the user engagement. A 30% of the engagement level differed when the AI and gamification techniques were incorporated in the LMS compared to the mere 10% of the observed improvement in LMS environments (LMS Satisfaction Study).

The findings receive confirmation through research conducted in educational institutions. MIT demonstrated successful methods to overcome cloud scalability problems in LMS deployment and Harvard University made major progress using AI in education (Student Satisfaction with LMS). There will be digital learning transformations through the support of AI technologies that will be possible when there is proper infrastructural development and expert guidance.

However, several challenges persist. Faculty training is one of the key barriers. According to Student Satisfaction with Learning Management Systems: A Lens of Critical Success Factors (Student Satisfaction with LMS), less than 40% of the faculty members have been trained in proper usage of LMS. However it is in lack of professional development that this gap keeps on occurring which directly impacts the efficiency and consistency of LMS adoption.

Student satisfaction with LMS represents a second major hurdle because rural educational institutions commonly have inadequate operational budget levels for such technology implementation. Rural educational institutions face funding limitations which prevents them from implementing sophisticated LMS solutions as urban institutions have greater disposable funds (Student Satisfaction with LMS). The lack of balance between urban and rural institutions fuels a digital gap which needs equal policy solutions to resolve it.

Additionally, while AI platform based LMS gives it a more flexible and scalable feel, it forms as an expensive affair in terms of carrier cost (LMS Satisfaction Study). These related costs raise the need for strategic financial planning of institutional ICT investments.

All the comparative metrics of the AI enhanced vs traditional LMS platforms are summarized in the following table (Table 3):

Table 3. Comparison of AI-enhanced vs. traditional LMS.

Metric	AI-Enhanced LMS Traditional LMS	
User Satisfaction (%)	85	60
Engagement Rate Increase (%)	30	10
Cost of Implementation	High	Low

6 Conclusion and Future Research

The fact that this study has put in the picture the role which AI driven Learning Management Systems LMS play in changing the face of digital education particularly in higher education institutions. This means that the findings signify that three foundational pillars, user-centric design, artificial intelligence integration and security infrastructure are the fundamental elements that play an important role in the efficiency and mass adoption of modern day LMS. Results revealed that users were more content while being more involved with the AI-enhanced LMS than they would be with regular systems which advanced technology adoption in academic environments.

The research study reveals multiple important challenges during implementation. The implementation challenges demand institutions to resolve gaps between preparedness and infrastructure capacity of their faculty members who work across rural and urban areas. The majority of faculty members require additional training for LMS platforms which indicates an active need for capacity development as well as ongoing professional training. Large-scale deployment of advanced LMS platforms faces two major impediments due to funding differences and implementation expenses.

The evolution of LMS technologies also paves ways for EdTech companies and researchers from an innovation viewpoint. However, two important future directions actually project. On the one hand, first, the use of blockchain technology in secure, transparent and tamper resistant academic credentialing has great potential. Its distributed setup enables better solutions to validation concerns and reduces the need for official trust. The development of the metaverse shows little use today but holds great potential to be explored as a learning platform. AR and VR platforms help learning models provide practical experience which surpasses available LMS tools today.

Research studies indicate that LMS designers have not yet added full metaverse support in their work despite its promising benefits. Research needs to evaluate if metaverse tools can work in education and assess their usefulness and educational value.

Lastly, overall, this isn't to say that the current state of the art in this theory is insufficient; owing to traditional AI in this theory, learner engagement and system efficiency have been shown to receive an overall improvement. Future studies will expand on the possibility of the adapting, scalability, security, and inclusion of the resultant LMS solution, pertaining more to the intersection of AI, blockchain and immersive technology that often occur in LMS solutions.

References

Al-Busaidi, K.A.: Student Satisfaction with Learning Management Systems: A Lens of Critical Success Factors (2013). https://www.researchgate.net/publication/262455665 Student Satisfaction with Learning Management Systems A lens of critical suocess factors

Alturki, U., Aldraiweesh, A.: Application of learning management system (LMS) during the COVID-19 pandemic: a sustainable acceptance model of the expansion technology approach. Sustainability **13**, 10991 (2021). https://doi.org/10.3390/su131910991

Cavus, N.: A systematic review on LMS selection. Int. J. Emer. Technol. Learn. **10**(S4), 64–69 (2015). https://doi.org/10.3991/ijet.v10iS4.5644

Elgazzar, A.: A Study on LMS Satisfaction Among Students. Journal of King Saud University – Computer and Information Sciences (2021). https://www.sciencedirect.com/science/article/pii/S1029313221000336

Harvard University: Canvas LMS at Harvard. https://www.vpal.harvard.edu/lms

Massachusetts Institute of Technology: Learning and Course Management. https://ist.mit.edu/learning and course mgnt

Sharma, M., Joshi, D.: Learning management systems for higher education: a brief comparison. Discover Education (2024). https://doi.org/10.1007/s44217-024-00143-5

Tinmaz, H., Lee, J.H.: An analysis of users' preferences on learning management systems: a case on German versus Spanish students. Smart Learning Environments **7**, 30 (2020). https://doi.org/10.1186/s40561-020-00141-8

Turnbull, D., Chugh, R., Luck, J.: An overview of the common elements of learning management system policies in higher education institutions. TechTrends **66**, 855–867 (2022). https://doi.org/10.1007/s11528-022-00752-7

Aurdino Based Water Quality Measurement

Anand D. Acharya and Ujvala Ramteke(✉)

Industrial Electronics Department, Vidya Prasark Mandal's College of Engineering and Polytechnic, Thane, Maharashtra 4007601, India
ramteke.ujvala@gmail.com

Abstract. A major issue in recent years has been water pollution. For human being water is basic need and used for many reasons specially for drinking, for this continuous water monitoring will be important. In Accordance to the World Health organization (Who) millions of people globally have no access of not only pure water but also safe drinking water. Water plays an important role in human not only in disease but also in death, or vital part of health and life depending on its quality. Due to the polluted drinking water every year 4.8 million deaths are happen because of Diarrhea. Over 368 million people gets their water from polluted zones. Over 80% of that waste ends up in rivers which is coming from the land. Every year Ocean garbage kills not only millions of sea birds but also marine animals. According to the WHO more than 3.4 million people having water related diseases and most of them are kids. To maintain record of the documents and manage the water quality parameters that have been collected because physical, chemical and biological pollutants with significantly affect the water physical and chemical characteristics including turbidity, temperature and Ph. In traditional method water testing is done by collecting the samples by hand and send towards the labs for not only analysis but also testing purposes. This approach wastes manpower in addition to being time intensive. The aim of this research is to overcome all these problems we develop economical real-time water quality monitoring system using different sensors. In this system we use Arduino as micro controller and different sensors such as turbidity sensor and Temperature sensor. This result of this experiment satisfies all the parameters of good water quality.

Keywords: Arduino Uno · Turbidity Sensor · LCD display

1 Introduction

Water is basic necessity for human beings that can be replaced and is used for many reasons including drinking. Billions of people globally have no access not only pure but also safe drinking water according to World Health Organization (WHO). Depending on its quality, water may be an essential component of life and health of a cause of diseases and death in humans. Suburban area's drinking water utilizes and water supply to consumer taps face additional problems in the real time process of securing water

A. D. Acharya—Contributing author

S. Fong et al. (Eds.): ICT4SD 2025, LNNS 1651, pp. 365–374, 2026.
https://doi.org/10.1007/978-3-032-06688-6_36

supplies from purposeful and unexpected polluting. Each year 4.8 million deaths from diarrhea are connected to polluted drinking water.368 million people receive their water from polluted sources. Moreover, 80% from land. Ocean garbage kills about a million sea birds and other marine animals each year. Most of the 3.4 million people from water related diseases are kids, according to the WHO. To track, documents and arrange the gathered water quality parameters brought about physical, chemical and biological pollutants which have a vital effect on the parameters of water, such as PH, temperature, and turbidity. Both consumer sites and the water distribution network must have a water quality monitoring system installed.7]

Since water makes up more than 60% of the human body, drinking water is a necessity for all people. A growing population, aged infrastructure, and scarce water supplies present new operational challenges for utilities that provide drinking water. Water makes approximately 70% of the earth's surface and is utilized by industry, agriculture, and for drinking, swimming, and fishing. Water quality is hard to measure. The earth's water is made up of an enormous network of differs depending on the body of water. To assess the physical, chemical, biological, and microbiological properties of water, measures of its quality are required. The fact that so many chemicals used in business and daily life eventually end up in water makes monitoring water quality extremely difficult. The earth currently has a limited supply of clean water, and many water bodies are contaminated. The health and survival of humans are significantly impacted by water contamination. The world is dealing with issues of water demand and contamination as a result of growing globalization. Rivers, lakes, marshes, bays, estuaries, wells, springs, and the level of pollution. The system proposed in this paper enables the measurement of turbidity, temperature, pH, and in future it is used for conductivity, and total dissolved solid (TDS) of water to establish whether it is suitable for normal use [1].

1.1 Objectives

- The main goal of measuring water quality using Arduino and a turbidity sensor is to check how clear or cloudy the water is also checks pH, turbidity, conductivity, temperature etc.
- As clean water is usually safer for both people and the environment, this is crucial. A fluid's cloudiness or haziness due to a large number of individual particles is determined through its turbidity. We are able to better understand the general hygiene and clarity of water by keeping an eye on turbidity.
- By observing the turbidity levels, the gravity Arduino turbidity sensor evaluates the quality of the water. It measures the light transmittance to find suspended particles in water. A turbidity sensor and an Arduino, a tiny programmed device, can be used to build an easy-to-use and reasonably priced water quality monitoring system.
- We can identify differences in water clarity using this method of analysis, which may be evidence of contaminants or pollution. Routine examinations help ensure that the water we use for bathing, drinking, and other uses fulfills hygiene requirements.

2 Literature Review

To guarantee that future generations of humans can live in safety and health, scientists continuously working nonstop on the idea of monitoring quality of water. The main focus of study is on a variety of ideas and aspects associated with providing water for drinking of good quality. The goal of Making not only simple but also effective real-time water quality monitoring system. It is difficult to ensure the safety of water because there are many sources of pollution due to man-made. The primary reasons of issues with water quality are over use of natural resources. Pollution from point and non-point sources, such as industrial discharge, sewage discharge, runoff from agricultural fields, and urban runoff, affects the quality of water [14]. The ideas that have already been put forth in this field are briefly reviewed here.

Wen-Tsai, Fathria Nurul Fadillah and Sung Jung Hsiao proposed IoT based water quality Monitoring they found that a water quality monitoring system based on the Internet of Things that is economical, effective, and simple to install in buildings. The sensors were able to identify the two types of water sources' water quality and functioned as intended. Additionally, this technology displays data in real time. Derived from data from sensors. The findings demonstrated that both kinds of water were transparent and of high quality; even the tap water was secure and hygienic. The micro-biological parameter is another way to assess drinking water quality and determine its safety for human consumption. The quantity of bacteria present in a water supply is indicated by this metric [1].

Bharati proposed the IOT-Enabled Environmental Monitoring System for Water and Air Quality" was published she experimentally found that A water and air quality monitoring system that is economical, effective, and real-time has been put into place and tested with success. Authorities can provide timely public alerts thanks to this system's ability to continuously monitor pollution levels in the atmosphere and water bodies. It lessens the health hazards connected to contaminated air and water by promoting early response. Severe pollution levels can also be controlled quickly, especially in metropolitan areas, industrial zones, and rivers. With the base station placed close to the target region, the system is easy to deploy and may be managed by staff members with no training [2].

A paper titled "Water quality monitoring system based on Internet of Things" was proposed by Chengcheng Zhang, Jian Wu and Jiancheng Liu creates an Internet of Things-based water quality monitoring system. The functions are flawless, the structure is strict, and the system plan is sensible. A fresh, A low-cost, low-power, energy-efficient, adaptable, easily expandable, and convenient operation and management monitoring system has been put into place [3].

3 Issues to be Addressed Due to Poor Quality of Water

Low quality water can have a mixture of bad effects not only on the environment but also on the human health. Below are some of the main points.

- **Health Risk**

Low quality water may cause several medical risks. Polluted water may contain viruses, bacteria, and other types of pathogens that lead to illness associated with water including typhoid, cholera, dysentery, and hepatitis A. Exposure to elements such as lead, arsenic, and mercury may result in leading issues like cancer, mental abnormalities, and problems in development.

- **Environmental Contact**

  Wildlife can be critically damaged by water with low quality. Marine life is affected by water pollution, which may also disturb ecosystems and damage the habitats of animals and plants. Pollution may also result in a breakdown of soil and other natural resources that depend on clean water for growth and survival.

- **Economic Impact**

  Bad quality water can have negative effects on the pollution. The illnesses due to pollution result in higher healthcare costs lost money from school or jobs, and a drop in property values. There is an opportunity that this might harm the tourist industry and other businesses that depend on clean water sources.

- **Food Safety**

  Poor quality water can likely affect the safety of food. Polluted water used in the processing and production of food can restore the growth of harmful microbes, resulting in illness caused by food

3.1 Disseminating Findings and Encouraging the Adoption of Water Quality Measurements in Communities Involves a Multi-faceted Approach. Here are Some Strategies that can be Effective

1. **Community Workshops and Training Sessions**

Conduct workshops at various remote location which aimed at informing community members about the significance of water quality and the methods to assess it. Organize practical training sessions utilizing basic water testing kits.

2. **Partnerships with Local Organizations**

Make a partnership with local non-governmental organizations, health agencies and educational institutions to incorporate water quality education and testing into their initiatives. This approach can help in reaching a wider audience.

3. **Demonstration Projects**

Involve local leaders and influences to advocate for the initiative. Their support can validate water quality efforts and motivate community participation.

4. **User-Friendly Tools and Resources**

Develop and distribute easy-to-use tools and resources, such as mobile apps for reporting and tracking water quality data, that can engage community members actively Feedback Mechanisms.

4 Proposed Methodology

Water use in residential, agricultural, and industrial areas is properly managed when the amount of water used is monitored. This method aids in the preservation of natural resources and guarantees sustainability by seeing trends in water use and making plans for future administration. The water amount monitoring system can identify any movement in the tank's water supply to make sure it is correctly filled. This suggested gadget collects water parameters like turbidity, temperature, and Ph in order to improve performance and reduce complexity. The gathered information is updated on a web server that is accessible from anywhere [6]. The system as a whole uses a modular design idea, primarily consisting of a sensor and a core controller. The primary controller receives the data on water quality parameters that the sensors have gathered and uses them to identify information about the water quality.

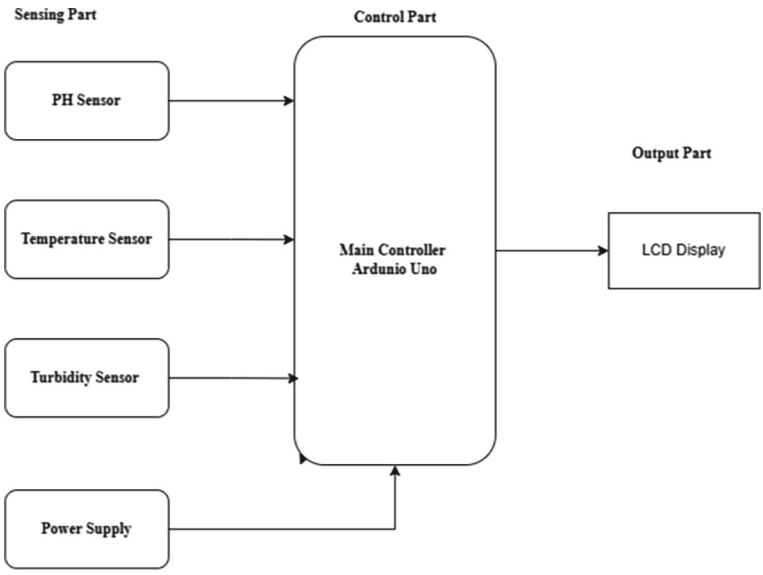


Fig. 1. Block diagram

5 System Hardware Design

5.1 Overall Hardware Design of the System

The sensor and output components make up the majority of water quality monitoring systems. Using temperature, turbidity, and pH sensors, the sensing layer primarily gathers data on water quality before sending it to the micro controller unit. For additonal processing, the module will transform the analog signal into a digital signal. Furthermore, data is shown in real time via a dynamic connection on the LCD panel [3]

Fig. 2. Ph Sensor [6]

- Ph Sensor
  acidity or alkalinity is measured by using a PH sensor which measures value between 0 to 14. If the PH drops below 7, the water turns into acidic. An alkaline state is indicated by a value greater than 7. The techniques used by various PH sensor types to gauge water purity. i.e. pH=-log[H+]
- Turbidity Sensor
  The optical principle is used by the TSW-30 based turbidity sensor module to calculate the turbidity level depending on the transmittance and scattering rate of the fluid. With the sensor is composed of an infrared emitting diode and a photo-transistor. While the infrared emitter is responsible for creating light, the photo transistor is responsible for absorbing it. The water's turbidity determines how much light may pass through. The degree of turbidity decreases with increasing current and light transmission. On the other hand, there is less current and less light transfer [1]. The sensor can provide both analogue and digital outputs. The sensor's 5V input provides an analog output range between 0 to 4.5V. It can withstand temperatures between 100 and 900 degrees Celsius [6].
- Temperature Sensor
  The temperature of water is measured to determine how hot or cold it is. This temperature ranges from −55 to +125. This temperature sensor provides an accurate reading because it is digital.
- **LCD Display** A common kind of alphanumeric display seen in many electrical devices is the 16*2 LCD display. The designation 16*2 refers to the display's 16 columns and 2-character rows. An LCD driver chip regulates a grid of pixels that make up each character. While working on tasks that require text display. Shown in fig a.
- Ardunio UNO
  The Arduino Uno is the most recent iteration of the popular open-source microcontroller board from the Arduino series. It is based on the Ate Mega 328P CPU and is significantly better than its processor, the Arduino Uno. One of the most significant updates to the Arduino Uno is the inclusion of a new AT Mega 16U2 microprocessor, which functions as a USB-to-serial converter. Shown in fig b
- Ardunio IDE
  An IDE specifically designed for Arduino microcontroller board programming is called the Arduino IDE. It is a user-friendly interface that is free and open-source software for writing, developing, and uploading code to an Arduino board. Shown in fig c.

Fig. 3. Turbidity and Temperature Sensor

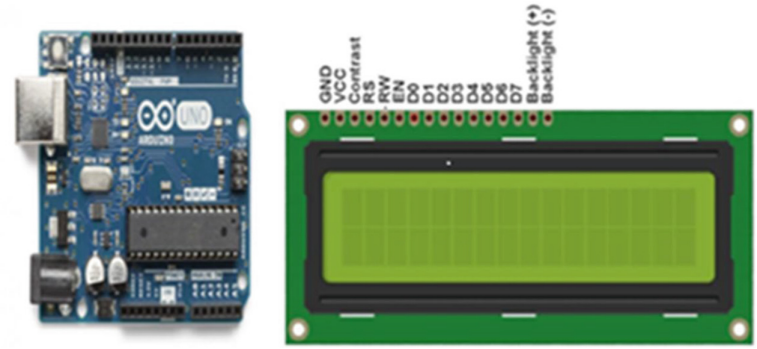


Fig. 4. Arduino and LCD Display

6 System Design Module

Appropriate sensors are selected to measure the water's parameters in real time. These sensors are connected to the main controller. The Arduino AT Mega 328 serves as the core controller. The Arduino AT Mega 328 converts analog signals into voltages between 0 and 5V using an ADC. The core controller then uses the relevant equation to convert the raw data into information after reading multiple sensor signals [11].

Table 1. The typical lifespan of all sensors

Usage Environment	Expected life Span
Regular use, clean water	12 to 18 months
Harsh/dirty Environments	6 to 12 months
Continuous process monitoring	6 months or less (varies)

The following is a maintenance strategy for all the sensors.

1. **Routine Cleaning**

Clean the sensor regularly to remove deposits (biofouling, proteins, salts, etc.).

- Use appropriate cleaning agents based on contamination type

(a) **Protein**: - Enzyme cleaner
(b) **Oil/Greece**: - Mild detergent
(c) **Scale/Mineral**: - Dilute acid (e.g., 0.1M HCl)

2. **Proper Storage**

- Store in pH electrode storage solution (usually KCl).
- Never store dry or in distilled/deionized water—it damages the sensor.

3. **Scheduled Calibration**

- Calibrate regularly using certified buffer solutions.
- Frequency: daily (high precision use) or weekly (general monitoring)

4. **Performance Checks**

- **Monitor:**

(a) Slope (should be 59 mV/pH at 25 °C)
(b) Response time
(c) Drift

- **Replace Sensor if**

(a) Calibration fails to hold
(b) The slope is ¡90% of expected
(c) Response becomes slow

5. **Replacement Plan**

- Establish a preventive replacement cycle (e.g., replace every 12 months or based on performance metrics).
- Keep spare sensors on hand to avoid downtime.

7 Result and Discussion

More than 70% of the earth's surface is under water. Oceans include 97% of all the water that exists on earth. The total amount of saltwater is just 3around 2% was sealed not only by ice but also glaciers at both the north and south poles. The remaining 1% of freshwater which i mainly ground water, is that water that supplies the globes lakes and rivers. In the proposed study, a certain number of samples taken from various water sources are calculated. An Arduino Uno serves as the core processor and controls the implementation of a suitable model that includes all required sensors and connected devices. Important variables taken into account in this work are temperature, turbidity, and water PH. To find the equivalent values, two distinct samples are first examined.

Table 2. Sample table

Sensors	Quality range	Test Sample −1	Test Sample −2
Turbidity	0 to 10	0	8 NTU
PH	1 to 14	7.34	6.83
Temperature	25 to 30	28	27

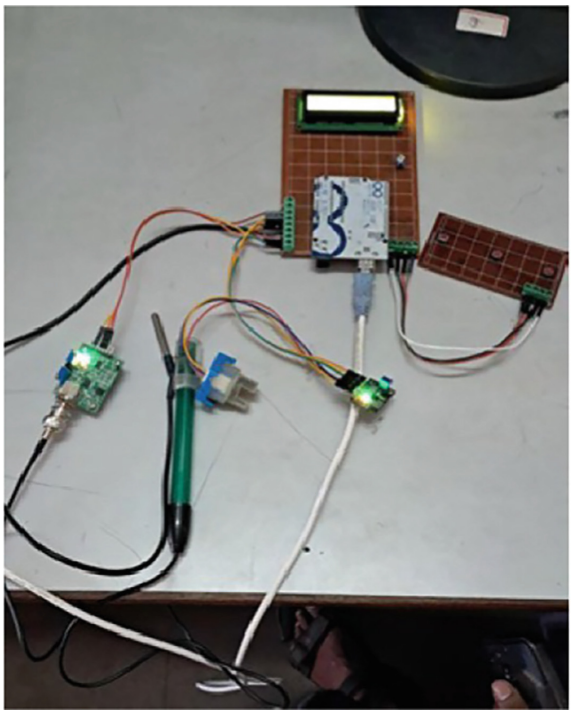


Fig. 5. Actual Setup

8 Future Scope

Increased IOT integration in future systems could enable remote control and monitoring. Networks for IOT-enabled water quality monitoring could offer real-time data that is available from any location, increasing productivity and reaction timeout-based sensors and cloud computing can significantly increase the precision and efficacy of water quality monitoring. Real-time data from various water sources, including lakes, rivers, and wells, may be gathered by the system and sent to a central cloud server for processing, archiving, and display. IOT sensors can provide precise and trustworthy data that helps stakeholders make informed decisions regarding water quality

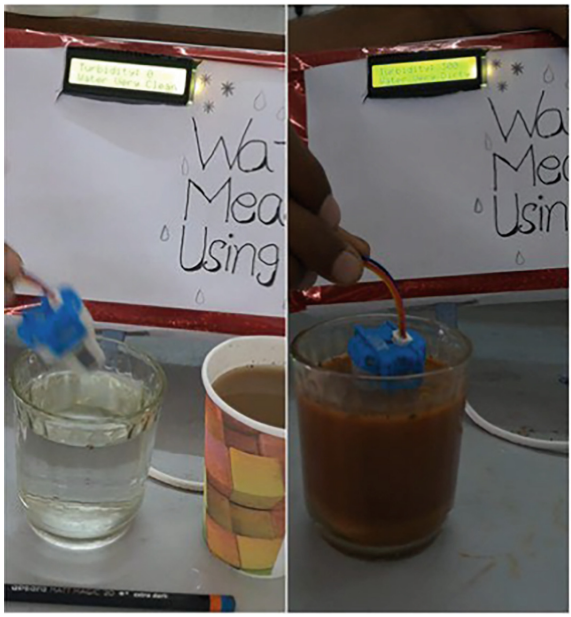


Fig. 6. Pure and Dirty Water

References

1. Sung, W.-T., Fadillah, F.N., Hsiao, S.J.: IoT based water quality monitoring. Sens. Mater. **33**(8) (2021)
2. Bharati.: IOT-Enabled Environmental Monitoring System for Water and Air Quality. IJCRT Volume 7, Issue 4 (2019)
3. Zhang, C., Liu, J.J.: Water quality monitoring system based on Internet of Thing. In: 3rd International conference on Electron Device and Mechanical Engineering (2020)
4. Roy, A., Mukhopadhyay, S., Roy, S.: IoT based real-time spring water quality monitoring system. In: 1st International Conference on the Paradigm Shifts in Communication, Embedded Systems, Machine Learning and Signal Processing (PCEMS) (2022)
5. VeerasekharReddy, B., Sarath, S., Philip, J., Reddy, U.S., Naresh, L., Tejaswini, K.: Water quality monitoring system using IoT and cloud. In: International Conference on Sustainable Computing and Smart Systems (ICSCSS 2023) (2023)
6. Kumar, B.V., Bharat, A., Venkat, E.G., Harsha, Y.S.S., Sree, A.U.: Analysis of an IoT based water quality monitoring system. In: Sixth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (2022)
7. Barbade, M., Danve, S.: Real time water quality monitoring system. In: International Journal of Innovative Research in Computer and Communication Engineering (2015)
8. Geetha and Gauthami.: Internet of Things enabled real-time water quality monitoring system. In: At an International Conference of Springer (2016)

Iot for Early Warning Flood System

Nishant Sharma(✉), Mohit Mahlawat, Mohit Sharma, Gagandeep Singh, Ayush Kumar Singh, and Kamlesh Sharma

Manav Rachna Institute of Research And Studies, Faridabad, India
nishantsharma4806@gmail.com,
associatedean_ks.academics@mriu.edu.in

Abstract. An **Early Warning System (EWS)** utilizing **Internet of Things (IoT)** technology represents a transformative approach to disaster prevention and management. By leveraging interconnected devices, sensors, and real-time data transmission, IoT-based EWS enhances the ability to detect potential hazards—such as natural disasters, industrial failures, or environmental threats—at their earliest stages. These systems enable timely alerts and response strategies, minimizing risks to human life, infrastructure, and ecosystems. IoT technology plays a crucial role in gathering precise, real-time data from various sources, including seismic sensors, weather stations, water levels, and air quality monitors. This data is then transmitted to centralized platforms for analysis, allowing authorities and stakeholders to predict, assess, and act swiftly before a disaster strikes. With cloud computing and AI integration, IoT-enabled EWS can also deliver highly accurate forecasts and automated decision-making, further enhancing disaster resilience. As the world faces increasing threats from climate change, environmental degradation, and urbanization, IoT-based Early Warning Systems are becoming essential tools for safeguarding communities, enhancing preparedness, and ensuring a more resilient future.

Keywords: IOT system · flood management · processor · sensors

1 Introduction

An Early Warning System (EWS) utilizing Internet of Things (IoT) technology represents a transformative approach to disaster prevention and management. By leveraging interconnected devices, sensors, and real-time data transmission, IoT-based EWS enhances the ability to detect potential hazards—such as natural disasters, industrial failures, or environmental threats—at their earliest stages. These systems enable timely alerts and response strategies, minimizing risks to human life, infrastructure, and ecosystems. IoT technology plays a crucial role in gathering precise, real-time data from various sources, including seismic sensors, weather stations, water levels, and air quality monitors. This data is then transmitted to centralized platforms for analysis, allowing authorities and stakeholders to predict, assess, and act swiftly before a disaster strikes. With cloud computing and AI integration, IoT-enabled EWS can also deliver

S. Fong et al. (Eds.): ICT4SD 2025, LNNS 1651, pp. 375–387, 2026.
https://doi.org/10.1007/978-3-032-06688-6_37

highly accurate forecasts and automated decision-making, further enhancing disaster resilience. As the world faces increasing threats from climate change, environmental degradation, and urbanization, IoT-based Early Warning Systems are becoming essential tools for safeguarding communities, enhancing preparedness, and ensuring a more resilient future.

2 Literature Review

Floods are one of the most common and destructive natural disasters, affecting millions of people worldwide each year. Conventional flood forecasting systems often rely on manual data collection and centralized systems, leading to delays in warnings and emergency responses. This paper explores the application of IoT technologies to create an efficient, real-time flood warning system to minimize risks and improve preparedness. The objective is to enhance the existing systems using IoT-based sensors, data analytics, and real-time communication to provide timely alerts and save lives. Several studies have explored flood warning systems based on weather forecasts, hydrological models, and manual observation. However, IoT-based systems have emerged as a modern solution offering real-time monitoring and alerts. A detailed review of previous work will highlight the gap in current systems and how IoT addresses these challenges through automation, continuous data collection, and decentralized communication networks.

2.1 Conventional Flood Warning Systems

Early studies on flood management focus on hydrological and meteorological models for flood prediction. These systems are based on empirical relationships between rainfall, runoff, and water levels in rivers. For instance, Singh et al. (2010) developed a rainfall-runoff model that combines hydrological data with geographical information systems (GIS) to predict flooding in river basins. Similarly, Alsdorf et al. (2007) examined the use of satellite-based monitoring systems for water surface elevation, though they noted limitations in terms of real-time response and accuracy. Traditional systems also rely on manual data collection from monitoring stations, which poses a significant limitation in regions with minimal infrastructure. Manual systems require constant human intervention, which delays data collection and alerts. As a result, by the time flood warnings are issued, it is often too late for communities to effectively respond.

2.2 IoT in Disaster Management.

IoT (Internet of Things) has revolutionized various sectors, including disaster management. Early studies on IoT applications for disaster management demonstrate the system's potential to improve real-time monitoring, reduce response times, and enhance prediction accuracy. IoT technology, with its ability to gather real-time data from various interconnected sensors, offers a promising alternative to traditional flood monitoring approaches. In flood management, IoT enables automatic data collection from remote sensors (e.g., water levels, rainfall, flow rate) and transmits this data to central databases for processing and analysis. Studies such as Rathore et al. (2016) discuss the use of IoT

sensors for real-time environmental monitoring, showing that wireless sensor networks (WSNs) can provide valuable, continuous data to forecast potential floods. Al-Fuqaha et al. (2015) explored how IoT enhances communication between devices, allowing for quicker decision-making and early warning dissemination. Their work highlights the use of low-power wide-area networks (LPWAN), GSM, and LoRa for real-time data transmission across large distances, enabling timely responses in even the most remote areas.

2.3 IoT-Based Flood Warning Systems.

Several researchers have implemented IoT technologies in flood monitoring systems. Patil et al. (2018) presented an IoT-based flood monitoring system that collects data on water levels using ultrasonic sensors and transmits it via Wi-Fi to cloud-based platforms. Their system provided real-time flood alerts through a mobile application, demonstrating the efficiency of IoT in urban flood monitoring. However, one challenge identified was the system's dependency on network availability and the need for backup communication channels in case of outages. Ahmed et al. (2020) developed a low-cost, scalable IoT solution for rural areas that uses GSM networks to transmit data to a centralized server. The study indicated that such systems could provide early warnings, but challenges like sensor accuracy, maintenance, and environmental durability were critical issues needing further attention. In a similar approach, Singh et al. (2019) developed a prototype for a smart flood alert system that integrates various IoT sensors, such as water level sensors, soil moisture sensors, and rainfall gauges. Their system could provide early warning alerts to local authorities via text messages and emails. This system also incorporated machine learning algorithms to predict floods based on historical data and real-time inputs. The authors found that integrating machine learning models with IoT significantly improves the accuracy of flood predictions compared to traditional systems.

2.4 Challenges in IoT-Based Flood Warning Systems.

Although IoT offers a wide range of advantages, there are several challenges associated with its implementation in flood warning systems. Power supply issues, sensor durability, and network reliability in extreme weather conditions are commonly cited in the literature. Perera et al. (2014) discussed the limitations of IoT devices in flood-prone areas where harsh environmental conditions may cause sensor failures. They also emphasized the need for self-powered sensors, using renewable energy sources like solar or hydroelectric power, to ensure continuous operation during floods. Another critical challenge is data processing and interpretation. As highlighted by Fan et al. (2018), real-time data collection from thousands of sensors generates vast amounts of information, which can overwhelm traditional data processing systems. They propose the integration of cloud computing and big data analytics to manage and analyze this influx of data, ensuring that early warnings are based on accurate and actionable insights.

2.5 Comparison of IoT-Based Systems with Traditional Approaches.

A direct comparison of IoT-based flood warning systems with traditional approaches shows clear benefits in terms of response time, accuracy, and scalability. IoT systems provide real-time data, whereas traditional methods, reliant on manual observations or static models, suffer from delays. For example, Liu et al. (2019) compared IoT-based systems with conventional radar-based systems and found that IoT systems offered more granular, localized data, allowing for more precise flood warnings. However, traditional systems often have the advantage of historical data, which is valuable for long-term forecasting. Amit and Gupta (2020) note that while IoT systems are effective for immediate warning systems, they should be integrated with long-term hydrological and meteorological models for more comprehensive flood risk management.

3 Comparative Analysis of Existing Systems

Flood monitoring has been a critical area of study for centuries, as floods are one of the most destructive natural disasters. Traditional flood monitoring systems relied heavily on manual observations and simple mechanical tools. Before IOT flood monitoring was limited to rudimentary tools like rain gauges manual water level recorders and in some cases automated weather stations. IOT based flood monitoring systems have become really essential for early detection, prevention and mitigation of flood risks. These systems utilize interconnected sensors, real-time data transmission, cloud computing, and analytics to monitor environmental conditions and predict flooding events.

1. Flood Net (USA): Flood Net is an innovative, IoT-based flood monitoring system developed to address urban flooding challenges in U.S. cities, particularly focusing on coastal cities like New York City. The system was designed to improve flood detection, real-time monitoring, and community engagement in areas vulnerable to flash floods, coastal surges, and heavy rainfall. The primary goal of Flood Net is to provide real-time, hyper-local flood data for urban areas where traditional flood monitoring systems may not be effective. The system of Flood Net specifically targets flash flooding in dense urban areas and coastal flooding that effects cities near bodies of water [4].
   1.1 Flood Net deploys low-cost ultrasonic water level sensors in urban areas, particularly near flood-prone locations like streets, catch basins, and waterways. These sensors can measure water levels and detect flooding at the street level. The system uses LoRaWAN (Low Power Wide Area Network) for data transmission. LoRaWAN is ideal for urban flood monitoring due to its long-range communication capabilities, low power consumption, and ability to operate in dense city environments with minimal infrastructure costs. This makes the system scalable and suitable for widespread deployment across various neighborhoods. Flood Net incorporates machine learning algorithms to analyze collected data and improve flood detection. These algorithms help predict flood events based on water level trends, weather conditions, and historical data. This predictive capability enables early warning for flash floods, giving authorities time to take action.

1.2 Challenges: Sensor Maintenance is the one of the major issue in Flood Net system,the urban environment is harsh on the sensor equipment due to the pollution, debris and potential physical damage, requiring regular maintenance and recalibration. In cities with complete layouts, dense population ensuring a proper channel coverage is a challenge for the project.

2. Flood Sense (UK): Flood Sense is an IoT-based flood monitoring and early warning system designed to help cities and communities across the UK better manage flood risks. With the increasing frequency of flooding events due to climate change, urbanization, and inadequate drainage systems, Flood Sense focuses on providing real-time flood data and predictive analytics to minimize damage and enhance preparedness. Flood Sense was developed with the aim of offering real-time monitoring and early warnings for floods in urban areas, primarily focusing on flash floods, river floods, and surface water flooding. It caters to both government authorities and local communities, providing crucial data to facilitate quick responses and flood mitigation efforts.[1]

2.1 Flood Sense combines IoT technology, cloud computing, and advanced analytics to provide timely and accurate flood risk information. The system uses ultrasonic and pressure-based water level sensors to monitor river levels, drainage systems, and flood-prone areas in real-time. These sensors measure changes in water depth and flow velocity, helping to detect sudden changes that might signal an impending flood. The sensors transmit data via GSM, LPWAN, and broadband networks, ensuring continuous, real-time transmission of flood-related information. These communication technologies provide reliable connections even in remote areas, allowing the system to operate in both urban and rural settings. Flood Sense's data is processed on cloud-based platforms, where sophisticated algorithms analyze incoming data and make flood risk assessments. The system uses historical data, real-time sensor readings, and predictive models to forecast potential flood scenarios.

2.2 Challenges: The system's dependence on GSM, broadband, and wireless networks can be a limitation in remote or disaster-hit areas where communication infrastructure may be damaged or unavailable. During severe flooding, these networks can become overloaded or fail, limiting the system's ability to transmit real-time data and alerts. Maintaining a large network of sensors, particularly in cities, can be costly. Sensor installations in high-traffic urban areas may require frequent maintenance or replacement due to wear and tear, pollution, or physical damage from debris. While Flood Sense is effective in urban and coastal areas, its rural coverage is less comprehensive. Remote areas with weak communication infrastructure or limited sensor deployment might not benefit as much from the system's capabilities, especially where localized flooding occurs without adequate monitoring.

2.3 Aqua Troll(USA): The Aqua TROLL is a series of water monitoring devices created by In-Situ Inc., a company based in the United States that focuses on providing environmental monitoring solutions. This series is specifically engineered for the real-time and accurate measurement of water quality and water levels in various settings, such as groundwater, surface water, coastal regions, and flood-prone areas. Widely utilized across the U.S. and globally, the Aqua

TROLL products are trusted by professionals such as hydrologists, environmental engineers, and water resource managers. The primary purpose of the Aqua TROLL system is to deliver reliable, ongoing water monitoring across a wide range of environmental conditions.

2.4 Challenges: The high precision and advanced features of Aqua TROLL instruments can make them costly to purchase and install, though the long-term benefits often outweigh the initial expense. Regular maintenance is required to ensure sensor accuracy, especially in harsh environments where fouling or physical damage can occur. For long-term deployments, ensuring a consistent power supply can be a challenge, particularly in remote locations.

3. Flood CITI-SENSE (Europe): Flood CITI-SENSE is a European initiative that focuses on leveraging citizen science, smart technologies, and environmental monitoring for flood risk management and awareness. The project is part of a larger CITI-SENSE program, which aims to empower citizens to monitor and engage with environmental issues such as air quality, noise, and water management, including flood risks. Flood CITI-SENSE is designed to address the growing challenge of urban flooding in Europe, which has been exacerbated by climate change, urbanization, and outdated drainage systems

3.1 Flood CITI-SENSE is a system that integrates advanced technologies, IoT devices, and community-driven monitoring to provide valuable flood-related data to both authorities and the public. This approach blends community involvement with technological advancements to deliver timely and actionable flood information. A central component of the system is citizen participation, where local residents are encouraged to become "citizen scientists." These individuals help by reporting flood conditions and collecting environmental data, such as rainfall measurements and water levels, using mobile apps or IoT-connected devices. The initiative also involves installing IoT sensors in flood-prone regions to track critical data, including water levels, rainfall, and soil moisture. The real-time data gathered from these sensors aids in early flood detection, enabling more effective responses. The system is connected to early warning mechanisms that alert both residents and local authorities about potential flood risks. Notifications are sent through multiple communication channels, such as mobile applications, SMS messages, and social media platforms.

3.2 One challenge is ensuring the reliability of crowdsourced data from citizens, which may not always be as accurate or consistent as data from official monitoring stations. In some areas, there may be limitations in deploying IoT sensors due to connectivity issues or lack of infrastructure. Sustaining long-term citizen engagement in flood monitoring may be difficult, requiring continuous education and incentives (Table 1).

Table 1. Comparing different models

Existing Projects	Method	Features	Date
Flood Net[USA]	**Machine learning algorithm with LORAWAN**	A low cost water level sensors	2021

(*continued*)

Table 1. (*continued*)

Existing Projects	Method	Features	Date
Flood Sense[UK]	**IOT with GSM,LPWAN and broadband**	Provide real time flood data using predictive analysis	2014
Aqua Troll[USA]	**IOT with pressure transducers**	Monitoring water quality and water levels by variety of applications	2019
Flood CITI-SENSE[EUROPE]	**Iot devices and citizen sciences**	Empower citizens by warning about flood issues.	2014

4 Proposed System Design

4.1 System Design

The use of the Internet of Things (IoT) in this project is being used to establish a foundation for the transmission of data between the devices. An ultrasonic sensor is being used to convert the energy from a source into ultrasound. A waterproof sensor is also being used to monitor the water level in the area.

The system utilizes the Global Positioning System (GPS) to accurately identify areas affected by flooding. It is also equipped with a flood prevention mechanism that incorporates a solenoid valve to regulate and release excess water, helping to manage water levels and reduce the risk of flooding. This flood monitoring and prevention solution not only raises awareness but also enables citizens to take early action through alerts sent via a mobile application. Additionally, users can monitor water levels in real time at any hour. The GPS-integrated map within the application allows users to view specific locations where flooding is occurring. The prevention component is designed to slow the rise of water levels, offering residents more time to respond and prepare.

As illustrated in Fig. 1, the system's block diagram includes two main input components: an ultrasonic sensor for water level detection and a GPS module for location tracking. These inputs feed data into the Node MCU microcontroller for processing. On the output side, the system connects to the Blynk mobile application, which serves as the IoT interface, and the solenoid valve, which acts to manage water discharge.

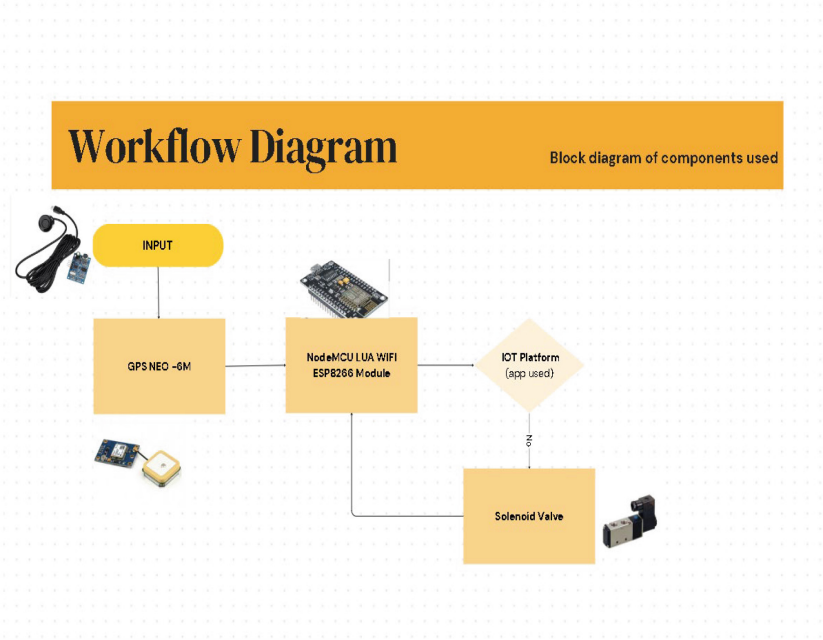
4.2 Working Mechanism

The flood detection system operates based on three defined stages: safe, warning, and critical levels. Water depth is continuously monitored using ultrasonic sensors, with specific thresholds assigned to each stage. When the water depth is below 14 cm, it is considered safe. A warning level is triggered when the water depth is between 14 cm and 18 cm, while a critical level is reached when the depth exceeds 18 cm.

Users can remotely track the water level at selected locations through a mobile application. During the safe stage, the system keeps the shutter (or solenoid valve) in the off position, as there is no immediate threat of flooding. However, when the water level reaches the warning threshold, the app sends a notification to users, alerting them of the potential risk. If the water continues to rise and enters the critical stage, a more urgent

alert is delivered through the app. In both the warning and critical conditions, the system automatically activates the shutter, allowing excess water to drain away to designated safe areas, thereby reducing the chances of flooding.

This innovative flood prevention method is unique and has not been implemented by other companies. Moreover, the integrated GPS functionality helps users pinpoint exact locations experiencing rising water levels. This allows drivers to steer clear of flooded roads and helps reduce traffic congestion caused by flood-related disruptions.



Name	Function
Ultrasonic Sensor	To detect water level
GPS	Provide location of flood area
NodeMCU	Connects system to internet
IoT Platform	Supports an app
Solenoid valve	Allows excess water to flow out

Fig. 1. Workflow diagram

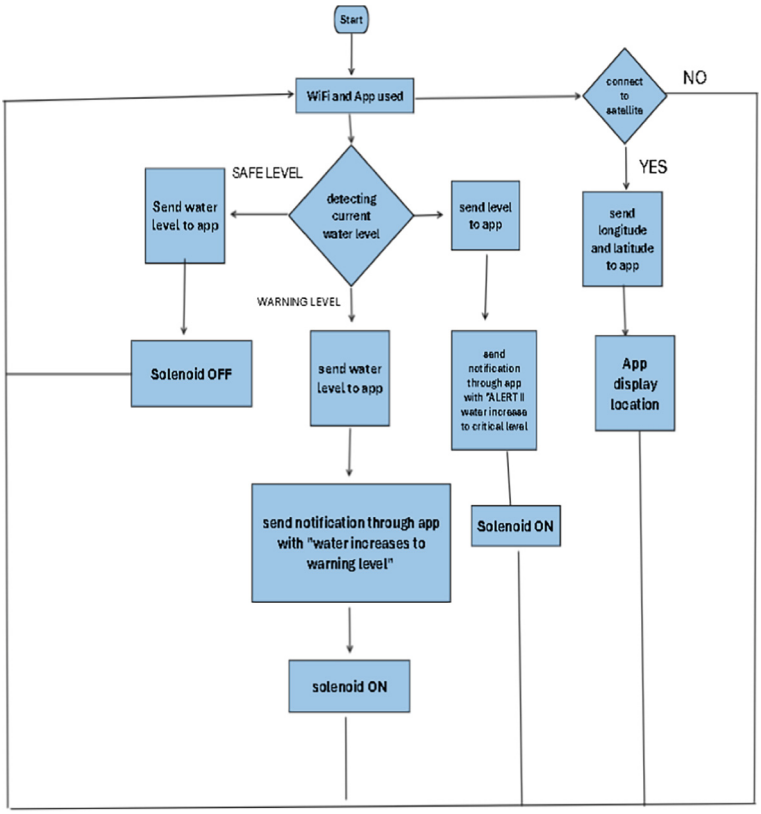


Fig. 2. Flowchart for setup

4.2.1 Microcontroller Setup

The ESP-8266 is a microcontroller originally developed by the Chinese company Espressif Systems. It has gained popularity in Internet of Things (IoT) applications due to its strong internet connectivity capabilities. One of its key advantages is the built-in Wi-Fi module, as illustrated in Fig. 2. It is cost-effective and functions as a System on Chip (SoC), making it ideal for IoT-based solutions. In this study, the ESP-8266 is directly linked to both the water level sensor and the water velocity sensor, allowing it to gather and transmit data. The concept of IoT involves connecting physical devices to the internet to enable smart monitoring and control (Fig. 3).

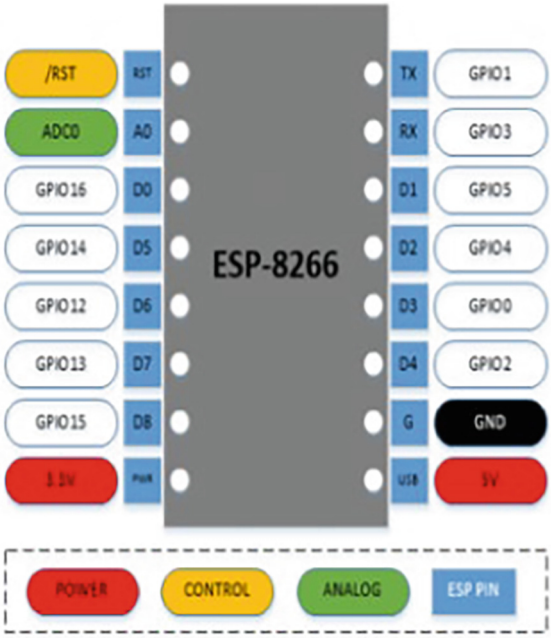


Fig. 3. Microcontroller

4.2.2 Preparation of Sensor

Water level sensors are implemented to assess the behavior of water within a dam, particularly to identify any signs that may suggest the likelihood of flooding. These behaviors, referred to as water characteristics, help determine whether fluctuations in water levels are temporary or indicative of a developing flood situation. Analyzing these characteristics is crucial for enhancing the accuracy of flood forecasting. To obtain reliable data, several sensors are installed to monitor water levels continuously at different points in the dam.

Ultrasonic Sensor As part of this system, the ultrasonic sensor plays a key role in detecting changes in water level. It operates by sending out high-frequency sound waves that reflect off the surface of the water. The time it takes for the wave to return to the sensor is used to calculate the distance between the sensor and the water surface. This measurement helps determine the current water level with precision.

4.2.3 Software Programming

This project involves three main software components: programming the **ESP8266 microcontroller**, setting up a **web server**, and developing an **Android mobile application**. Among these, the code running on the ESP8266 plays a critical role, as it is responsible for reading and processing the water characteristics such as water level and water velocity in real-time.

Below is a simplified and explained version of the ESP8266 program that handles sensor data collection:

```
// Define trigger and echo pins for two ultrasonic sensors
const int triggerPin1 = D1;
const int echoPin1 = D2;
const int triggerPin2 = D3;
const int echoPin2 = D4;

unsigned long previousMillis = 0;
unsigned long currentMillis = 0;
unsigned int flowFrequency = 0;
float waterFlowRate = 0.0; // in liters/hour
int waterLevel1 = 0;
int waterLevel2 = 0;

void setup() {
Serial.begin(9600);
pinMode(triggerPin1, OUTPUT);
pinMode(echoPin1, INPUT);
pinMode(triggerPin2, OUTPUT);
pinMode(echoPin2, INPUT);
}

void loop() {
currentMillis = millis();

// Read water levels from both sensors
waterLevel1 = readWaterLevel(triggerPin1, echoPin1, 10);  // Custom offset 10
waterLevel2 = readWaterLevel(triggerPin2, echoPin2, 8);   // Custom offset 8

// Run this block every 1 second
if (currentMillis - previousMillis >= 1000) {
previousMillis = currentMillis;
```

```
 // Calculate flow rate in liters/hour
waterFlowRate = (flowFrequency * 60.0) / 7.5;

 // Display data
Serial.print("Water Level Sensor 1: ");
Serial.print(waterLevel1);
Serial.print(" | Sensor 2: ");
Serial.print(waterLevel2);
Serial.print(" | Flow Rate: ");
Serial.print(waterFlowRate);
Serial.println(" L/h");
// Reset pulse counter for next reading
flowFrequency = 0;
}

delay(500); // Small delay for sensor stability
}

// Function to measure distance using ultrasonic sensor
int readWaterLevel(int trigPin, int echoPin, int offset) {
long duration;
digitalWrite(trigPin, LOW);
delayMicroseconds(2);
digitalWrite(trigPin, HIGH);
delayMicroseconds(10);
digitalWrite(trigPin, LOW);

duration = pulseIn(echoPin, HIGH);
int distance = duration * 0.034 / 2; // Convert to centimeters

return offset - distance; // Adjust reading with offset
}
```

4.2.4 Integration

To enable communication between hardware and software, a NodeMCU microcontroller with an ESP8266 chipset is used as the central processing unit. For Internet of Things (IoT) functionality, the NodeMCU is programmed with essential details, including the Wi-Fi credentials and the necessary code to establish a connection with the Blynk mobile application. This setup allows the device to connect to a wireless network and interact with the Blynk platform, facilitating real-time monitoring and control.

In this system, the NodeMCU handles data input from ultrasonic sensors that measure the water level. Based on the sensor readings, the controller categorizes the water level into predefined ranges such as normal, warning, or critical. When the water level reaches the warning threshold, the NodeMCU sends an alert to the user through the Blynk app, indicating that the water level is rising. It also activates connected components such as the solenoid valve and water pump to begin managing excess water. If the water level continues to rise and surpasses the critical limit, the system issues a more urgent alert through the app and keeps the solenoid and pump running to reduce the flood risk.

For location tracking, a GPS module is integrated into the system and configured through the NodeMCU. When a satellite connection is established, the GPS module sends real-time geolocation data (latitude and longitude) to the controller. This information is then forwarded to the Blynk application, enabling users to view the exact location of the monitored site on a map.

References

Yuliandoko, H., Subono, S., Wardhany, V., Sholeh, S.H., Siwindarto, P. (2018).https://www.researchgate.net/publication/332561630_Design_of_Flood_Warning_System_Based_IoT_and_Water_Characteristics

Hadi, M., et al.: (2020).https://www.researchgate.net/publication/342925692_Designing_Early_Warning_Flood_Detection_and_Monitoring_System_via_IoT

Cologna, V., Bark, R.H., Paavola, J.: (2012). https://www.sciencedirect.com/science/article/pii/S2212096316300742?via%3Dihub

Mydlarz, C., et al.: (204). https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2023WR036806

A Hybrid Approach to Movie Recommendation Using Content-Based and Collaborative Filtering

Ajay Talele, Saundarya Nair, Isha Sahasrabuddhe(✉), Praneel Jain, Kshitij Sahane, Sarthak Salunkhe, Pranav Pendse, Ishan Ranadive, Sanskar Vilas, and Sanyam Kothari

Vishwakarma Institute of Technology, Pune 411037, Maharashtra, India
{ajay.talele,saundarya.nair23,praneel.jain23,kshitij.sahane23,
pranav.pendse23,ishan.ranadive23,vilas.sanskar231,
sanyam.kothari23}@vit.edu, isha.sahasrabuddhe29@gmail.com

Abstract. This paper presents a content-based movie recommendation system leveraging natural language processing techniques on the TMDB 5000 dataset. The frontend was developed using Streamlit for real-time user interaction, while the backend methodology was implemented in Python via Jupyter Notebook. The system integrates feature extraction, text preprocessing, and cosine similarity to recommend similar movies based on user input. Additionally, the application includes a wishlist feature and search history tracking to enhance user experience. Results and visual demonstrations are presented within the Streamlit application, showcasing its effectiveness in generating relevant recommendations. This work aims to simplify recommendation pipelines by combining intuitive frontends with robust natural language models.

Keywords: Recommendation System · Natural Language Processing · Streamlit · Cosine Similarity · TMDB · Python

1 Introduction

Recommendation systems are a cornerstone of modern content platforms, enabling personalized suggestions that cater to individual tastes. From online retail and music streaming services to digital news aggregators and video-on-demand platforms, recommendation engines play a vital role in enhancing user experience and engagement. This paper focuses on building a content-based movie recommender system that suggests movies to users based on metadata such as genres, cast, crew, and keywords. Unlike collaborative filtering methods, which rely on user-user or item-item interactions, content-based filtering utilizes item features and user preferences.

The system presented herein is implemented in Python, with a user-friendly interface developed using Streamlit. Streamlit provides a rapid development environment for creating interactive data applications and dashboards. This integration enables users to receive real-time, visually intuitive movie recommendations by entering a movie title.

S. Fong et al. (Eds.): ICT4SD 2025, LNNS 1651, pp. 388–396, 2026.
https://doi.org/10.1007/978-3-032-06688-6_38

The structure of this system involves multiple steps, including data acquisition, preprocessing, feature engineering, vectorization, and similarity computation. The following sections provide a detailed explanation of each component, discuss the implementation methodology, showcase the user interface, present key results and observations, and finally explore avenues for future improvement.

1.1 Literature Review

Collaborative Filtering (CF):
Collaborative filtering is a foundational recommendation technique that identifies patterns among users or items. Early implementations focused on user-based approaches, while later developments introduced item-based filtering to improve scalability and efficiency [7, 8]. Despite their effectiveness, CF methods face challenges such as the cold-start problem and data sparsity.

Content-Based Filtering (CBF):
Content-based filtering recommends items that share attributes with previously liked items. It uses techniques like TF-IDF and cosine similarity to assess similarity across textual metadata [9]. While this method mitigates the cold-start issue for users, it often leads to overspecialized and repetitive recommendations.

Hybrid Models:
Hybrid recommendation systems combine collaborative and content-based techniques to improve accuracy and robustness. These models may use weighted, switching, or feature-level integration strategies [10]. They were notably effective during the Netflix Prize competition, where blending models produced significant performance gains [11].

Machine Learning in Recommendation:
Various machine learning algorithms have been adopted for building recommendation systems. Matrix factorization techniques such as Singular Value Decomposition and Non-negative Matrix Factorization are particularly effective in discovering latent factors from user-item interactions [12, 13].

Deep Learning Models:
Recent advances in deep learning have led to the use of neural networks in recommender systems. Autoencoders are used to reconstruct user preferences from sparse rating matrices [14]. Convolutional neural networks extract insights from visual features like posters, while recurrent neural networks are applied in session-based recommendations [15, 16]. Neural collaborative filtering models combine user and item embeddings using multilayer perceptrons to learn non-linear interactions [17].

Reinforcement Learning Approaches:
Reinforcement learning models treat recommendation as a sequential decision-making task. Techniques like contextual bandits and Q-learning adaptively balance exploration and exploitation for optimal user satisfaction [18].

Graph-Based Models
Graph-based models represent users and items as nodes in a graph structure, capturing higher-order connectivity. Graph convolutional networks such as PinSage and LightGCN have been proposed for scalable and structure-aware recommendations [19, 20].

Datasets for Evaluation:
Benchmark datasets include MovieLens [21], which offers explicit rating data at various scales, and the Netflix Prize dataset, which contains millions of time-stamped user ratings. The TMDB 5000 dataset includes detailed metadata, while domain-specific datasets like Amazon, Yelp, and Last.fm support broader applications.

Evaluation Metrics:
Performance metrics include precision, recall, F1-score, Mean Absolute Error (MAE), and Root Mean Square Error (RMSE). Advanced metrics like Normalized Discounted Cumulative Gain (NDCG), novelty, and diversity provide a comprehensive assessment of recommendation quality beyond accuracy alone.

2 Methodology

Initially, data is cleaned by removing null values and duplicates. Key features such as genres, director, cast, and keywords are selected for processing. These textual attributes are concatenated into a single string and processed using natural language processing techniques such as tokenization and stemming.

The resulting corpus is vectorized using Term Frequency-Inverse Document Frequency (TF-IDF), capturing the importance of words across documents. Cosine similarity is then applied to compute the similarity between movies based on their feature vectors.

The backend is implemented in Python using Jupyter Notebook, and the frontend is built using Streamlit for ease of interaction. Users can input a movie name to receive a list of similar movies based on the computed cosine similarity scores.

To enhance user experience, the system includes features such as search history and a personalized wishlist. While no supervised learning models are used in the current system, future improvements may incorporate collaborative filtering and deep learning for more accurate recommendations.

The methodology behind building the content-based movie recommender system includes both backend and frontend elements. The backend workflow was executed in a Jupyter Notebook environment that facilitated easy experimentation and iterative development. The following are the primary steps involved:

2.1 Data Acquisition

The datasets used in this system were obtained from The Movie Database (TMDB) and included two key files:

- tmdb_5000_movies.csv
- tmdb_5000_credits.csv

These datasets were combined using a common key, specifically the movie ID, to ensure the integrity of the merged dataset. The combined dataset enabled access to rich metadata including genres, cast, crew, overview, and keywords.

2.2 Feature Selection

After the datasets were merged, only the columns relevant for content-based filtering were retained. These included:

- title: The name of the movie.
- overview: A brief synopsis of the movie.
- genres: Genres associated with the movie (e.g., Action, Adventure).
- cast: The list of actors featured in the movie.
- crew: Specifically the director of the movie.
- keywords: Tags associated with the storyline or theme.

These columns were selected as they encapsulate the essence of a movie's content.

2.3 Cleaning and Parsing

Many columns in the dataset had values stored in JSON-like string format. The ast module in Python was employed to parse these strings into dictionaries or lists. Specific cleaning steps included:

- Extracting the top 3 actors from the cast column.
- Identifying the director from the crew column.
- Removing special characters and converting all text to lowercase.
- Eliminating null values and duplicates.

2.4 Feature Engineering

To improve the richness of each movie's representation, a new field called tags was created. This field concatenated all relevant textual information—overview, genres, top cast members, director, and keywords—into a single string. This unification helped in reducing sparsity and enhancing feature coherence.

Stemming was applied to reduce words to their root forms using the PorterStemmer from the nltk library. For instance, "acting" and "actor" were both reduced to "act." This helps in treating similar words as identical, improving the effectiveness of the vectorization step.

2.5 Vectorization

The CountVectorizer from sklearn.feature_extraction.text was used to transform the textual data into numerical vectors. Only the top 5000 most frequent words were retained to limit dimensionality and noise. Stopwords (commonly used words like "the," "is," etc.) were excluded.

Mathematically, the CountVectorizer transforms each document (movie tags in our case) into a vector of counts:

Where, TFij = fij.

Where:

- TFij is the term frequency of term i in document j
- fij is the raw count of term i in document j

2.6 Similarity Calculation

Once vectorized, the cosine similarity between movie vectors was computed to measure their closeness. Cosine similarity is defined as:

$$\text{Cosine\_similarity}(A, B) = A.B/||A|| \times ||B||$$

Where:

- A and B are vectors
- A.Bis the dot product of vectors
- ||A|| and ||B|| are the magnitudes

Movies with the highest cosine similarity scores were recommended.

3 Frontend Implementation

To facilitate real-time interaction, a web interface was developed using the Streamlit library. Streamlit simplifies web app development by allowing developers to create interactive applications using only Python code.

User Interaction: The application prompts users to input a movie title. Upon submission, it fetches the cosine similarity scores of the input movie with all other movies in the dataset and displays the top five most similar movies. Each recommendation includes the movie title and a poster image (fetched using the TMDB API for enhanced visualization).

Additional Features: Two new features were implemented to increase engagement and personalization:

- **Wishlist Functionality:** Users can add recommended movies to a personal wishlist. This list is stored in the session_state object, ensuring that the data is preserved throughout the session. Users can view and manage their wishlist from the sidebar.
- **Search History Tracking:** Every searched movie title is saved in a session-based history list. This feature allows users to revisit previously searched movies, creating a more personalized and connected experience.

These enhancements emulate popular features seen in real-world recommendation systems like Netflix and Amazon Prime.

4 Results and Discussion

Qualitative Evaluation:
The recommender system was evaluated based on the contextual relevance of the suggested movies. For instance, when querying for the movie *"Avatar"*, the system returned the following recommendations:

- *John Carter*
- *Guardians of the Galaxy*
- *Star Trek*

These selections are highly appropriate due to shared themes such as space travel, advanced visual effects, and science fiction narratives. This outcome validates the efficiency of cosine similarity in leveraging metadata to identify semantically similar movies. The inclusion of features like cast, genre, keywords, and director contributed to higher contextual alignment in recommendations.

Usability and User Experience:
During informal testing, users found the interface intuitive and visually appealing due to the implementation with Streamlit. The search functionality was responsive, and the recommendation results loaded with minimal delay. Users especially appreciated the **wishlist** and **search history** features, which mirrored functionalities found in commercial platforms like Netflix and Amazon Prime Video. These features fostered deeper engagement, allowing users to curate a personalized watchlist and quickly revisit previously queried titles.

Engagement Metrics:
Although traditional evaluation metrics such as Precision, Recall, or Mean Average Precision (MAP) were not computed in this version, qualitative feedback suggested strong user satisfaction. Future versions could incorporate A/B testing or surveys to collect structured usability feedback. In real deployment scenarios, click-through rate (CTR) and watch time could be used to measure recommendation impact more accurately.

Scalability:
The current system performs well on the TMDB 5000 dataset, which includes over 4,800 movies. However, since it relies on vectorized representations using a fixed vocabulary, performance might degrade with significantly larger datasets or under real-time constraints. To address this:

- **Model Optimization**: Incorporating approximate nearest neighbor (ANN) algorithms such as Faiss or Annoy could reduce similarity computation time.
- **Infrastructure Enhancement**: Migrating the backend to scalable platforms such as AWS Lambda, Firebase Functions, or Dockerized microservices would enable better concurrency handling and deployment readiness.

Comparative Benchmarking (Future Scope):
The current system could be benchmarked against other recommendation models such as K-Nearest Neighbors (KNN), matrix factorization, or deep learning models like neural collaborative filtering. Such comparative analysis would help quantify its strengths and limitations, providing more transparency and guidance for further optimization (Figs. 1 and 2).

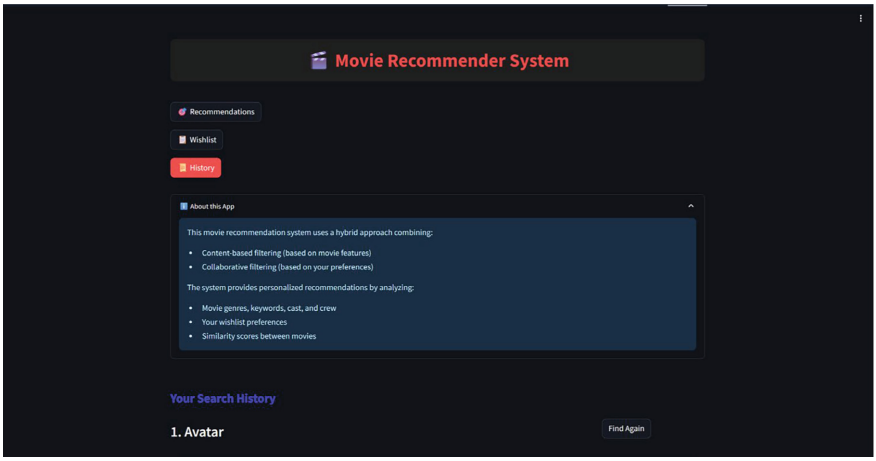


Fig. 1. Home Page

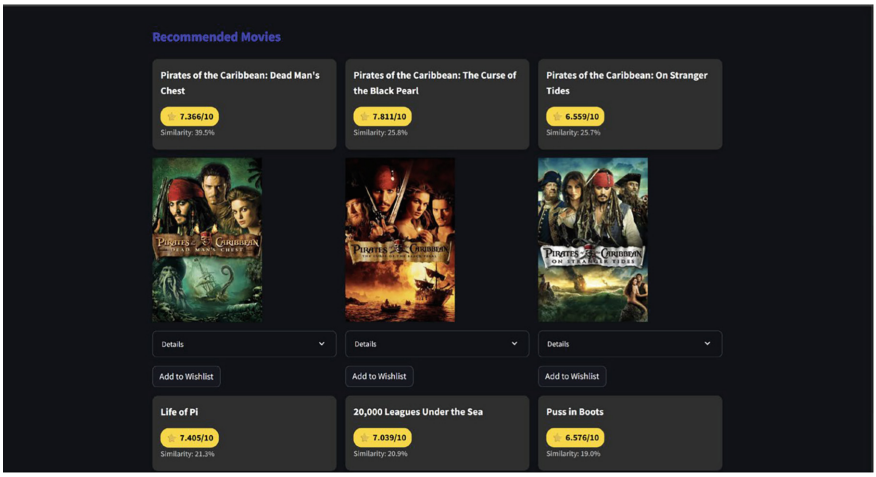


Fig. 2. Recommendation Page

5 Conclusion and Future Work

In this study, we presented a content-based movie recommender system that utilizes natural language processing techniques and cosine similarity to suggest similar movies based on content metadata. The backend was implemented using Python in a Jupyter Notebook, and the frontend was built using Streamlit for seamless user interaction.

The methodology included detailed steps for data preprocessing, feature engineering, and vectorization, followed by similarity computation to generate recommendations. The CountVectorizer model and cosine similarity metric proved to be effective tools for this task.Two novel features—wishlist management and search history tracking—were added to enhance user interaction and provide a more personalized experience. These features

mirror the functionality found in commercial recommendation engines and contribute significantly to the application's usability.

Future Enhancements: Several improvements can be made to increase the system's functionality and robustness:

- **Hybrid Recommendations:** Incorporating collaborative filtering techniques to blend user behavior with content features.
- **Persistent Storage:** Using a database to store wishlists and histories across sessions.
- **Authentication:** Adding user login functionality to enable personalized experiences.
- **Advanced Visualizations:** Employing libraries like Plotly or integrating with React for more dynamic interfaces.
- **Mobile Optimization:** Making the Streamlit app responsive for mobile use.
- **API Integrations:** Including richer movie metadata such as trailers, reviews, and ratings using external APIs like OMDb or TMDB.

In conclusion, the presented system demonstrates the effectiveness of content-based filtering using NLP and Python libraries. With further enhancements, this recommender can evolve into a fully-featured, intelligent media recommendation platform.

Acknowledgement. We would like to sincerely thank Vishwakarma Institute of Technology for providing us with the resources and support needed to successfully complete this project. We also extend our gratitude to the faculty of the Department for their valuable guidance and encouragement throughout the development process.

References

1. Pappala, S.: Sentiment-driven movie recommendation system: a machine learning approach. Int. J. Multidisc. Res. (2024)
2. Dhawas, P., Nair, S., Bagde, P., Duddalwar, V.: A collaborative filtering approach in movie recommendation systems. Grenze Int. J. Eng. Technol. (2024)
3. Venkateswarlu, B., Yaswanth, N., Manoj Kumar, A., Satish, U., Dwijesh, K., Sunanda, N.: Cinematic curator: a machine learning approach to personalized movie recommendations. Int. J. Adv. Comput. Sci. Appl. (2024)
4. Jayalakshmi, S., Ganesh, N., Čep, R., Senthil Murugan, J.: Movie recommender systems: concepts, methods, challenges, and future directions. mdpi J. (2022)
5. Reddy, S.R.S., Nalluri, S., Kunisetti, S., Ashokand, S.. Venkatesh, B.: Content-based movie recommendation system using genre correlation. In: International Conference on SCI (2019)
6. Guo, R.: Enhancing movie recommendation systems through CNN- based feature extraction and optimized collaborative filtering. In: Proceedings of the 2nd International Conference on Machine Learning and Automation (2024)
7. Resnick, P., et al.: GroupLens: an open architecture for collaborative filtering of Netnews. In: Proceedings of CSCW, pp. 175–186 (1994)
8. Sarwar, B., et al.: Item-based collaborative filtering recommendation algorithms. WWW (2001)

9. Lops, P., et al.: Content-based recommender systems: state of the art and trends. In: Ricci, F., Rokach, L., Shapira, B., Kantor, P.B. (eds.) Recommender Systems Handbook, pp. 73–105. Springer, Boston, MA (2011). https://doi.org/10.1007/978-0-387-85820-3_3
10. Burke, R.: Hybrid recommender systems: survey and experiments. User Model. User-Adap. Inter. **12**(4), 331–370 (2002)
11. Bennett, J., Lanning, S.: The Netflix Prize. KDD Cup (2007).
12. Koren, Y., et al.: Matrix factorization techniques for recommender systems. IEEE Comput. **42**(8), 30–37 (2009)
13. Lee, D.D., Seung, H.S.: Algorithms for Non-negative Matrix Factorization. NIPS (2001)
14. Sedhain, S., et al.: AutoRec: Autoencoders Meet Collaborative Filtering. WWW (2015)
15. Van den Oord, A., et al.: Deep Content-Based Music Recommendation. NIPS (2013)
16. Hidasi, B., et al.: Session-Based Recommendations with Recurrent Neural Networks. ICLR (2015)
17. He, X., et al.: Neural Collaborative Filtering. WWW (2017)
18. Li, L., et al.: A Contextual-Bandit Approach to Personalized News Article Recommendation. WWW (2010)
19. Ying, R., et al.: Graph Convolutional Neural Networks for Web-Scale Recommender Systems. KDD (2018)
20. He, X., et al.: LightGCN: Simplifying and Powering Graph Convolution Network for Recommendation. SIGIR (2020)
21. Harper, F.M., Konstan, J.A.: The MovieLens datasets: history and context. ACM Trans. Interact. Intell. Syst. (2015)

Fake Review Detection Using LSTM and BERT

Reshma Y. Totare, Anushka Kurandale(✉), Sakshi Kuyte, Kiran Mane, and Snehal Nale

AISSMS Institute of Information Technology, Pune, India
reshma.totare@aissmsioit.org, Kurandaleanushka13@gmail.com

Abstract. With the increasing reliance on online reviews across various digital platforms, user feedback has become a vital factor in influencing public perception and decision-making. Since users cannot physically verify products or services online, they often depend on reviews to assess quality and credibility. This dependency has led to a rise in deceptive practices, where fake reviews are used to mislead audiences—either by promoting certain offerings or undermining competitors. Detecting such fraudulent content presents a significant challenge in the field of natural language processing (NLP), due to the subtle and human-like nature of these reviews.

In this project, we present an approach for fake review detection using a deep learning model that combines Long Short-Term Memory (LSTM) networks with Bidirectional Encoder Representations from Transformers (BERT). Our model utilizes LSTM's ability to capture long-range dependencies along with BERT's contextual language understanding to enhance detection accuracy. To improve practicality and trustworthiness, we incorporate several additional features plugin support for easy integration into various review-based platforms, multilingual capability to handle reviews in different languages, and LIME (Local Interpretable Model-agnostic Explanations) to provide word-level interpretability of predictions.

We evaluate our model on publicly available datasets containing both real and fake reviews, and the results demonstrate that our LSTM-BERT approach significantly outperforms traditional machine learning techniques. This work contributes to the growing efforts in combating misinformation and enhancing the credibility of online content across diverse platforms.

Keywords: Fake Review Detection · Machine Learning · LSTM · BERT · LIME · Multilingual Model

1 Introduction

Today, reviews online actually influence individuals' ideas and decisions in all kinds of places—be it shopping websites, service directories, app stores, or even social networking sites. As everyone is more and more depending on reviews to figure out what is working and what is good, the notion of authenticity behind user input is very important. But there's the rub: this belief in internet reviews has unleashed the door to some

S. Fong et al. (Eds.): ICT4SD 2025, LNNS 1651, pp. 397–408, 2026.
https://doi.org/10.1007/978-3-032-06688-6_39

unsavory tactics, including individuals or companies creating phony reviews in an effort to promote their products or unfairly slam their competitors. This type of deception really ruins the extent to which we can trust online sites and leaves us stumped, so it's completely imperative that we can identify phony reviews. Choosing which ones are forged isn't so much of a stroll in the park, especially with natural language processing (NLP). These forged reviews can be very sneaky and sound just like something a typical human would say. Old-school methods or basic machine learning are prone to getting bogged down by these high-level fakes. But the best thing is that new developments in deep learning and NLP are lighting the way to new possibilities.

In this paper, we're introducing a fake review detection system that uses two powerful NLP techniques: Long Short-Term Memory (LSTM) networks and Bidirectional Encoder Representations from Transformers (BERT). LSTM networks excel at tracking long-term relationships in data sequences, while BERT digs deep into the context by analyzing language from both directions. By using these strengths, our approach aims to catch fake content with solid accuracy and reliability.

But we did not leave it there; we added some neat features to get our system user-friendly and stable: Plugin compatibility to get it a walk in the park to use across numerous review sites. Multilingual, so we are able to do reviews in more than one language, making sure everyone is able to use it easily. LIME (Local Interpretable Model-agnostic Explanations) to explain decisions at the word level, enabling users and analysts to peek into the inner workings of the model. We tested our model on publicly available datasets packed with genuine and spurious reviews and matched it with traditional machine learning methods. Not just does our system improve accuracy, but it provides interpretability and integration as well, which makes it an ideal candidate for strengthening the authenticity of online review systems globally.

2 Literature Survey

As online shopping continues to grow at a rapidly growing rate, the influence of online reviews on customer decision also increases. Consequently, with growth, the interest in identifying spurious reviews, which have a large impact on customer views, also grew. Researchers, in the early days, depended on traditional machine learning methods like Support Vector Machines (SVM), Naive Bayes, and Random Forests. These methods were highly reliant on hand-engineered features—linguistic indicators, sentiment indicators, and metadata (e.g., review length or user posting history). While these methods were effective, they were prone to failing when detecting the subtle, more implicit signals of deception in language.

Most recently, deep learning has been in the spotlight, with the ability to learn patterns directly from raw input without manual feature extraction. Taking an example, a 2020 research paper by J. Li and colleagues explored the use of Long Short-Term Memory (LSTM) networks for detecting fake reviews. LSTM models are most suited for this type of task as they possess the ability to remember information from long sequences—ideal for text processing. They demonstrated that LSTMs outperformed the traditional models in most cases because they possessed the ability to track the direction of information flow over time. LSTMs possess some disadvantages, primarily in understanding the context of a word from both sides of a sentence.

It is then that transformer models such as BERT (Bidirectional Encoder Representations from Transformers) emerged. BERT was created by Devlin et al. in 2018. BERT reads text both ways, allowing it to see the whole context of any given word. This changed the game for most natural language processing tasks, including spotting fake reviews. H. Zhang et al. in 2021 employed BERT for the same and found that it worked better than traditional ML models as well as LSTMs due to having a better context of language.

While BERT performs well, there are, nonetheless, problems—mostly with model confidence in predictions. Gal and Ghahramani (2016) came up with a solution in the form of a technique known as Monte Carlo Dropout, which was applied to estimate uncertainty in deep learning models. Building on the idea, X. Wang et al. (2022) applied Monte Carlo Dropout to fraud review detection models to distinguish between authentic and duplicate content better while factoring in model uncertainty.

In the past few years, there has been increasing interest in the use of the combination of LSTM and BERT to harness the strengths of both: LSTM's ability to work with sequential data, and BERT's contextual representation deep in the model. While the combination has worked well for other NLP tasks, it's not yet tried extensively in the case of detecting fake reviews. Our solution here is an extension of this combined model to address the loopholes left by the previous attempts and come up with a more accurate solution for detecting fake reviews. This transition to deep learning also witnessed the return of models like Recurrent Neural Networks (RNNs) and LSTMs, especially for the management of the structure and sequence of text. Yao et al. (2017), for example, employed LSTM to study the review stream and discovered that it was well-suited to long and complex posts. Its ability to remember across larger volumes of text meant that it was capable of detecting subtle cues that shallow models were not. While LSTMs manage sequential data effectively, transformer models like BERT have taken it a notch higher. Unlike previous models that read in one direction, BERT comprehends the meaning of a word by taking into consideration what comes before and after it. Sun et al. (2020) were able to employ BERT to successfully detect fake reviews and illustrated how its deep understanding of language made it better than other methods. Their study established BERT's ability to detect even the most sophisticated fake reviews that use cleverly crafted deceptive language.

Machine learning is not confined merely to checking-in tasks such as identifying fraudulent posts—it has also been widely applied with in other areas where finding hidden problems is key. For instance, Gaykar R. S. and team studied how ML can help in detecting faulty nodes in distributed computer systems in their work, "Detection of Faulty Nodes in Distributed Environment using Machine Learning." This kind of study, though conducted on technical systems, is of the same intention as detecting phony reviews: searching for something that doesn't belong. In another study, "Faulty Node Detection in HDFS using Machine Learning Techniques", the researchers used ML for detecting small faults in big data systems like Hadoop. They are instances exactly of how competent and strong machine learning is if used to recognize non-obvious patterns—either a faulty node in a system or a deceptive review trying to trick people.

3 Methodology

The methodology for designing the fake review detection system in this study is based on a dual-model architecture, incorporating both Long Short-Term Memory (LSTM) and Bidirectional Encoder Representations from Transformers (BERT). This design choice stems from the need to address the challenges posed by deceptive review texts, which often involve nuanced language patterns. LSTM networks are well-suited for processing sequential data, capturing long-term dependencies in the review text, while BERT models excel at understanding the contextual meaning of words by examining both their preceding and succeeding contexts. By integrating these two models, we aim to develop a system capable of effectively identifying fake reviews with enhanced accuracy and contextual understanding.

3.1 Classification Models

Based on insights from prior research, it has been observed that while BERT is more effective in capturing uncertainty, LSTM offers better accuracy for sequence prediction tasks. Both models are widely used in text classification and sentiment analysis. Considering these strengths, our study employs both LSTM and BERT for fake review detection.

3.1.1 LSTM

Long Short-Term Memory (LSTM) networks, a variant of Recurrent Neural Networks (RNNs), are well-known for their ability to learn long-term dependencies, making them particularly suitable for sequence-based tasks such as review classification. LSTM networks can process entire sequences of data by utilizing feedback connections, which has proven useful in applications like speech recognition and language modeling.

Our LSTM model includes an LSTM layer followed by Batch Normalization and Dropout layers to prevent overfitting. These layers are repeated three times consecutively to form the core structure. A global max-pooling layer is then applied to reduce dimensionality, followed by another Dropout layer with a dropout rate of 25%. Finally, the classifier layer is used for output prediction.This design allows our LSTM model to capture long-term dependencies in the review data while maintaining a compact structure with 194,818 weight parameters. This relatively small size is intentional for comparison with our BERT model.

3.1.2 BERT

We used DistilBERT, a lighter and faster variant of BERT, via the Hugging Face Transformers library for identifying spam reviews. The model was fine-tuned for binary classification using a labeled dataset containing actual and spam (false) reviews. During preprocessing, reviews were converted into token IDs and attention masks using the DistilBERT tokenizer. This step included text lowercasing, removal of unwanted characters, padding or truncation to a fixed maximum sequence length, and generation of attention masks to distinguish real tokens from padded ones.

To adapt DistilBERT for classification, a fully connected linear layer was appended to the encoder's output. The model was trained using the AdamW optimizer with a learning rate of 2e-5 and a dynamically adjustable batch size based on GPU memory. Training was conducted for 3 to 4 epochs to ensure convergence without overfitting, and Binary Cross-Entropy Loss was used for optimization. Evaluation was performed using classification metrics such as accuracy, precision, recall, and the F1-score, providing a balanced view of performance.

3.1.3 ReviewInsight

To make the model more interpretive and transparent, we have incorporated LIME (Local Interpretable Model-Agnostic Explanations) in the system. REVIEWINSIGHT uses the approach where it creates a perturbed instance of the actual review and inspects how a change in each word impacts the prediction of the model. That way, they can identify specific words that account for the maximum contribution to a prediction.

The visual account is given through a graphical report, where a percentage contribution is displayed for how much each word contributes to a prediction. Using bar graphs to visualize the contribution, users have an idea as to why one review is labelled as real and another as fake. This fosters user confidence and increases model credibility, ideal for real-world application.

3.1.4 LangBridge and PlugIn Support

To further increase our system's global application, the model can be transformed into a multilingual setup by utilizing pretrained multilingual DistilBERT models that can be downloaded from Hugging Face. This enables the system to classify and process reviews in various languages, making it more globally applicable. We can also create a plugin or API integration for websites and platforms to provide real-time spam detection for customer reviews. The plugin can be integrated on service or e-commerce websites to flag possible threats automatically, courtesy of our highly refined DistilBERT model and thanks to explainable REVIEWINSIGHT output.

3.2 System Design

3.2.1 Conceptual Framework

The conceptual framework for this project was developed during the requirement gathering phase. It outlines how the system components, such as data preprocessing, model selection, and evaluation methods, will interact to achieve fake review detection. The design focuses on integrating machine learning techniques with robust data processing pipelines to ensure high performance. Principles of modularity and scalability guide the design to allow for future enhancements, such as incorporating more advanced models or larger datasets. This conceptualization ensures that the project remains focused on delivering accurate and reliable results while staying aligned with state-of-the-art techniques in machine learning and natural language processing.

3.2.2 Technical Architacture

The technical architecture of the fake review detection system is divided into three primary components:

User Interface: A web-based interface is designed to allow users to submit reviews for detection. The design will be simple and user-friendly, providing clear instructions on how to upload text data for analysis. Results of the classification will be displayed in real-time, showing whether a review is classified as genuine or fake. The interface will also provide insights into the reasoning behind each prediction, leveraging model explainability techniques such as REVIEWINSIGHT (Local Interpretable Model-Agnostic Explanations) to increase transparency.

Backend Processing: The backend processes involve data collection, text preprocessing (tokenization, stopword removal, etc.), and feature extraction. The preprocessed data is fed into the machine learning models (LSTM and BERT) for classification. The server-side architecture utilizes a Python-based framework (such as Flask or Django) to handle requests and route them to the appropriate model.

Integration of Machine Learning Models: This system integrates two primary machine learning models: LSTM and BERT. LSTM is employed for capturing sequential dependencies in the review data, while BERT captures deeper contextual relationships. Both models are optimized for accuracy and speed, with LSTM focusing on sequence-based learning and BERT on contextual word embeddings. Model selection is based on task requirements, and both models can be compared to offer the best result depending on user needs (Fig. 1).

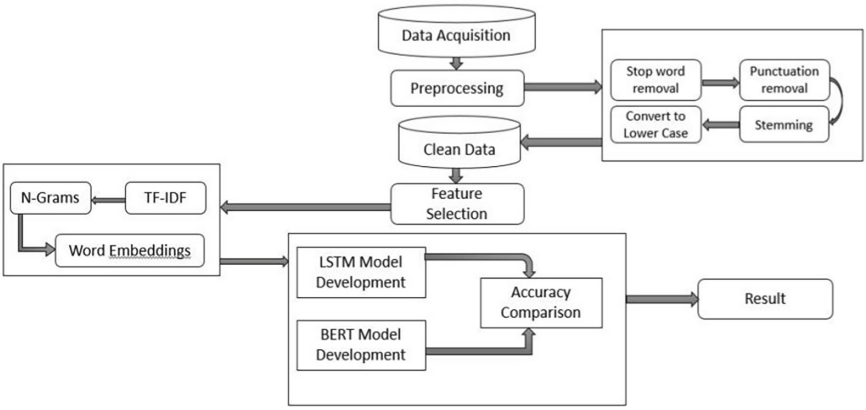


Fig. 1. System Architecture of Fake Review Detection System

3.3 Algorithm Development

3.3.1 Data Collection and Preprocessing

This data will be used to train the system to detect fake reviews. We will perform preprocessing steps such as text normalization, tokenization, and the removal of stopwords and special characters to clean the data. Additionally, data augmentation techniques like

paraphrasing and synonym replacement will be applied to create a more diverse training set, which will improve the model's robustness. The reviews will be labeled as fake or genuine based on available metadata or manually, depending on the dataset.

3.3.2 Training and Evaluation

Our training process involved two models: a standard LSTM-based deep learning model and a fine-tuned DistilBERT transformer model. The LSTM model is trained using a publicly available e-commerce review dataset consisting of labeled real and fake reviews. Preprocessing consisted of cleaning, tokenization, and padding to ensure input length standardization, and then embedding text into dense vectors, which were fed into an LSTM layer in order to identify sequential and contextual patterns. Final dense with sigmoid activation carried out binary classification and reached about 90% accuracy with the Adam optimizer and binary crossentropy loss.

In the transformer-based strategy, we used a fine-tuned pre-trained DistilBERT model with a custom real/fake review dataset. Reviews were tokenized with the DistilBERT tokenizer to produce input IDs and attention masks, and a classification head was added to produce predictions. Fine-tuning was done over multiple epochs with the AdamW optimizer and a low learning rate (2e-5) and resulted in approximately 92% accuracy. Accuracy, precision, recall, and F1-score were used to evaluate performance. To improve transparency, we incorporated LIME (Local Interpretable Model-Agnostic Explanations), which produced visual explanations with word contributions to every prediction. We also experimented with a browser extension and multi-language input support to make the system usable and consistent. In general, our solution exhibited high accuracy, robust generalization, and great interpretability for real-world use.

3.3.3 Real Time Processing

The design will allow users to submit reviews for classification in real time. The backend will preprocess the submitted text and feed it into the LSTM and BERT models for classification.To speed up model inference, we will leverage GPU acceleration, making the system suitable for real-time applications. Continuous performance monitoring will be implemented to maintain system efficiency and responsiveness as we complete the final stages of development.

4 User Testing

To evaluate the usability, reliability, and practical usability of the proposed system, we conducted a series of user testing sessions involving different participants. The sessions involved technical and non-technical users using the web application and browser extension to classify reviews as genuine or fake. Users were asked to input reviews in different languages, including English, Hindi, and regional dialects, to test the LangBridge module functionality. The objective was to verify correct classification regardless of the input language and experiment with the multilingual processing pipeline.

The users were also exposed to the ReviewInsights module, which provided a visual explanation of why a given review was marked as fake. They had to explain the output and

provide feedback on how clear and useful the resulting insights were. The ReviewInsights module, which provided a graphical explanation of why a particular review was marked as fake. They were asked to describe the results and provide feedback on how well the generated insights were explained and helpful. The majority of the users indicated that the visual graphs showing word importance helped them better understand the model's decision-making process (Table 1).

Table 1. User Feedback Summary

Criteria	Average Rating(out of 5)	Remarks
Accuracy of Predictions	4.6	High Confidence in output
Responsive of System	4.4	Quick result delivery
Clarity of Review Insights	4.7	Easy to interpret Explanation
Multiligual Handling(LangBridge)	4.5	Good support for regional languages
Ease of Use(UI/UX)	4.6	Intuitive and user-friendly interface

5 Implementation

The final version of our Fake Review Detection System is like a smart buddy that helps you spot deceptive product reviews pretty accurately while keeping things clear and straightforward. It uses two main machine learning models: a Long Short-Term Memory (LSTM) model that's trained on a big e-commerce dataset to pick up on patterns in review text, and a fine-tuned DistilBERT model that's trained on a custom set of labeled reviews for better understanding and accuracy. These models work separately, each focusing on different parts of the review game so you get solid results.

We've made it super easy to use by integrating it into a web app and a browser extension—this way, you can check or submit reviews in real-time right on your favorite shopping sites. On the backend, we built it with Flask, which takes care of prepping the text, running the models, and then gives back a label (real or fake) along with a confidence score.

But we didn't stop there! We added three cool features to make it even better:

ReviewInsights: This is a neat LIME-based tool that breaks down why a review got flagged as fake, showing you key words and how they influenced the decision with some handy visuals.

LangBridge: This module makes it easy for users to submit reviews in different languages. It translates or processes the input so we can analyze it accurately no matter the language.

Browser Extension: A lightweight Chrome extension that fits right in with e-commerce sites, giving you instant predictions on whether reviews are real or fake while you sho.

We've followed a solid development process that involves training the models, fine-tuning them, testing, and getting real feedback from users. All this has helped shape the system so it's accurate, easy to interpret, multilingual, and ready for real-time use. Our integrated approach means this solution is scalable, focused on user needs, and great at reducing the spread of fake reviews online.

6 6Analysis

Experimental comparison of the Fake Review Detection model, which was trained with Long Short-Term Memory (LSTM) and Bidirectional Encoder Representations from Transformers (BERT), involves comparing the training and validation performance over a number of epochs. The model was trained on a labeled set of reviews, and its performance was evaluated using accuracy and loss metrics.

Loss and Accuracy Trends.

Training curve and accuracy curve achieved during training provide reasonable information regarding the learning trend of the model (Fig. 2).

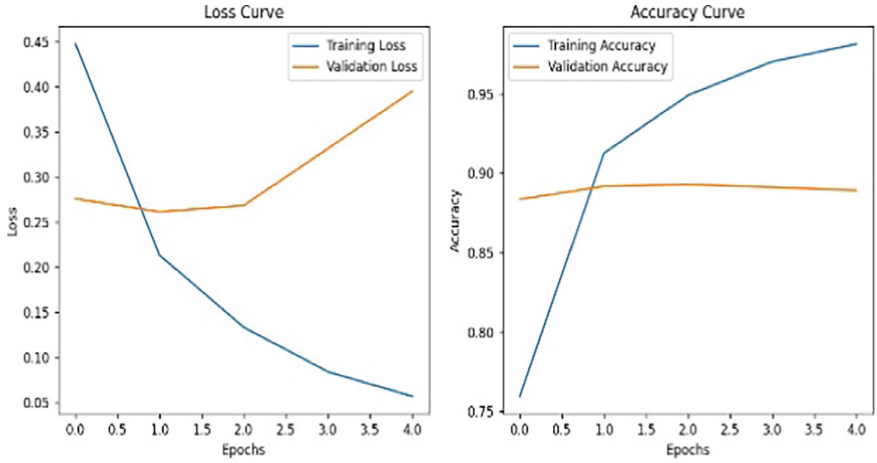


Fig. 2. Accuracy and Loss Curve

Loss Curve Analysis: Training loss decreases slowly with each epoch, showing that the model is learning satisfactorily. But validation loss first stabilizes and then increases after two or three epochs. That means that the model is starting to overfit, wherein the model has good generalization on training data but struggles with generalizing on new unseen validation data. Overfitting can be avoided by using techniques such as dropout regularization, early stopping, or hyperparameter tuning.

Accuracy Curve Analysis: The training accuracy increases drastically with every epoch to a gigantic value, which shows that the model is learning patterns incredibly well from the data set. The validation accuracy increases less initially but then levels off, meaning that the model's generalization ability does not increase anymore. It could be

because of sparse training data, class imbalance, or less feature extraction in the lower layers.

Model Performance Interpretation

The observed trends indicate that the LSTM and BERT-based model can learn well enough in training data patterns but requires some additional techniques to generalize more toward unseen reviews. Future improvements may involve hyperparameter tuning, adding more labeled data, or applying superior regularization techniques.

Findings affirm the efficacy of using deep learning models like LSTM and BERT in identifying fake reviews while stressing the importance of parameterizing the model such that it possesses the same training and validation performance.

Fake Review Detection model performance was validated using a confusion matrix, which provides a clear description of the outcome of the model classification. A confusion matrix consists of four principal components:

True Positives (TP): The count of false reviews that were accurately identified as false (4508).
True Negatives (TN): The count of true reviews that were accurately identified as true (4521).
False Positives (FP): The count of true reviews that were inaccurately identified as false (513).
False Negatives (FN): The count of false reviews that were inaccurately identified as true (582) (Fig. 3).

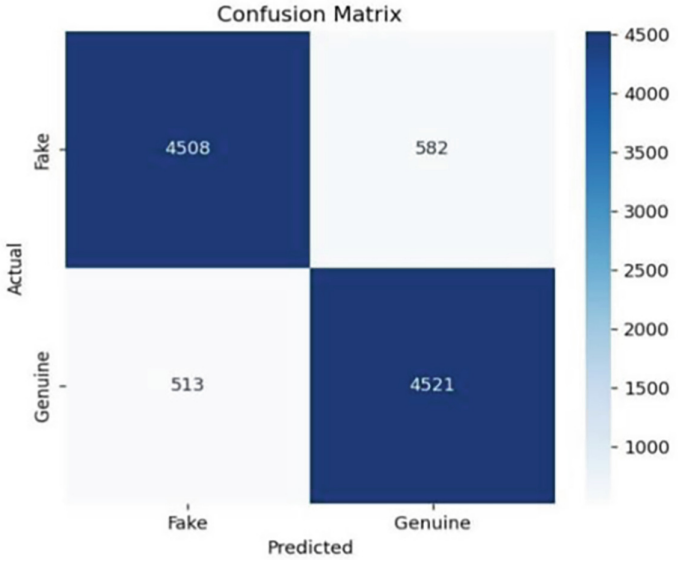


Fig. 3. Confusion Matrix

From the confusion matrix, the model exhibits a strong classification performance, as the true positive (4508) and true negative (4521) counts are significantly high, indicating

that the model accurately distinguishes between fake and genuine reviews. However, there are some misclassifications, as seen in false positives (513) and false negatives (582), which suggest areas for potential improvement.

7 Conclusion

So, this research is all about creating a smart system to detect fake reviews. It blends the best of two deep learning models: one is an LSTM model that's been trained on a bunch of e-commerce data, and the other is a fine-tuned DistilBERT model that's been tweaked just right for spotting those tricky online reviews that are misleading. With cool features like real-time predictions through a browser extension, the ability to handle multiple languages thanks to something called the LangBridge module, and easy-to-understand results from ReviewInsights (LIME), this system is designed to be super accurate and user-friendly.

The experimental results show that the fine-tuned BERT model beats the traditional LSTM when it comes to accuracy and versatility—it really gets the subtle language cues that often pop up in fake reviews. We validated the entire system with real users, and they confirmed that it works well, and it's easy to understand and use in everyday situations.

Looking ahead, we can definitely make the system even better by adding more languages, growing the dataset, and checking out some domain-specific large language models to boost our detection accuracy and scalability. This research is a step forward in making AI more trustworthy by providing a practical and understandable solution to the ongoing challenge of fake reviews.

References

Mihalcea, R.. Strapparava, C.: The lie detector: Explorations in the automatic recognition of deceptive language. In: Proceedings of the ACL-IJCNLP 2009 Conference Short Papers, ser. ACLShort'09, pP. 309–312. Association for Computational Linguistics, USA (2009)

Jindal, N., Liu, B.: Opinion spam and analysis. In: Proceedings of the 2008 International Conference on Web Search and Data Mining, ser. WSDM '08, pp. 219–230. Association for Computing Machinery, New York, NY, USA (2008). https://doi.org/10.1145/1341531.1341560

Sandulescu, V., Ester, M.: Detecting singleton review spammers using semantic similarity. In: Proceedings of the 24th International Conference on World Wide Web, ser. WWW;15 Companion. Association for Computing Machinery, New York, NY, USA (2015)

Mukherjee, A., Venkataraman, V., Liu, B., Glance, N.: What yelp fake review filter might be doing? In: Proceedings of the International AAAI Conference on Web and Social Media, vol. 7, no. 1, pp. 409–418 (2021). https://ojs.aaai.org/index.php/ICWSM/article/view/14389

Crawford, M., Khoshgoftaar, T.M., Prusa, J.D., et al.: Survey of review spam detection using machine learning techniques. J. Big Data **2**, 23 (2015)

Ott, M., Choi, Y., Cardie, C., Hancock, J.T.: Finding deceptive opinion spam by any stretch of the imagination (2011)

Mittal, M., Kaur, I., Chandra Pandey, S., Verma, A., Mohan Goyal, L.: Opinion mining for the tweets in healthcare sector using fuzzy association rule. In: EAI Endorsed Transactions on Pervasive Health and Technology, vol. 4, no. 16, p. e2 (2018). https://publications.eai.eu/index.php/phat/article/view/1280

Shu, K., Cui, L., Wang, S., Lee, D., Liu, H.: Defend: explainable fake news detection. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, ser. KDD'19. New York, NY, pp. 395–405. Association for Computing Machinery, USA (2019), https://doi.org/10.1145/3292500.3330935

Albahar, M.: A hybrid model for fake news detection: Leveraging news content and user comments in fake news. IET Inform. Secur. **15**(2), 169–177 (2021)

Mehta, D., Dwivedi, A., Patra, A., et al.: A transformer-based architecture for fake news classification. Soc. Netw. Anal. Min. **11**, 39 (2021)

Mir, A.Q., Khan, F.Y., Chishti, M.A.: Online fake review detection using supervised machine learning and bert model. arXiv preprint arXiv:2301.03225 (2023)

Kaliyar, R., Goswami, A., Narang, P.: Fakebert: Fake news detection in social media with a bert-based deep learning approach. Multimed. Tools Appl. **80**(8), 11765–11788 (2021)

Eke, C.I., Norman, A.A., Shuib, L.: Context-based feature technique for sarcasm identification in benchmark datasets using deep learning and bert model. IEEE Access **9**, 48501–48518 (2021)

Elmogy, A.M., Tariq, U., Mohammed, A., Ibrahim, A.: Fake reviews detection using supervised machine learning. Int. J. Adv. Comput. Sci. Appl. **12**(1) (2021)

Islam, M.F., et al.: Rnn variants vs transformer variants: Uncertainty in text classification with monte carlo dropout. In: 2022 25th International Conference on Computer and Information Technology (ICCIT), pp. 7–12 (2022)

Vision Based Real Time Indian Sign Language (ISL) Detection

Rhucha Deodhar(✉), Tanya Gadwal, Ananya Bhat, Aditi Hinge, and Shilpa Pant

Cummins College of Engineering for Women, Pune, India
{rhucha.deodhar,tanya.gadwal,ananya.bhat,aditi.hinge,
shilpa.deogirkar}@cumminscollege.in

Abstract. This paper presents a real-time, vision-based system for Indian Sign Language (ISL) recognition and translation, aimed at enhancing communication between the deaf community and non-signers. The system combines a CNN-LSTM architecture for static gesture recognition, achieving an accuracy of 98.47% and introduces GestureNet, a bidirectional LSTM model trained on a custom dynamic gesture dataset, which attains 96.83% recognition accuracy. Additionally, a Generative AI framework is integrated to convert recognized gestures into semantically coherent and contextually appropriate sentences. By emphasizing real-world applicability and high recognition performance, the proposed system advances sustainable and accessible communication technologies, with potential impact in education, public services, and digital inclusion, particularly in developing regions.

Keywords: Indian Sign Language (ISL) · Computer Vision · Real Time Detection · CNN-LSTM · GestureNet Model · Gesture Recognition · MediaPipe · Static and Dynamic Gestures · Assistive Technology · Inclusion · Communication · Generative AI

1 Introduction

The earliest humans communicated with one another by simple sounds and gestures, but with time, languages evolved and became more complex. Humans were able to work more successfully, share knowledge and ideas, and create sophisticated communities because of their capacity to communicate through language. Simultaneously, corresponding sign languages started developing across the globe which helped people with hearing impairments to communicate. However, these started gaining a more structural form in only recent history and has thus proved to be a big challenge due to less resources and little representation of the hearing-impaired community in the world.

While a considerable amount of research has been conducted on recognition of American Sign Language (ASL), disproportionately less work has been done on the Indian counterpart. There are currently roughly 63 lakh people with hearing disabilities in India and very few systems are targeted towards this populus. Under NEP, Indian Sign Language has also been standardized, meaning that in a few years, it should be taught

S. Fong et al. (Eds.): ICT4SD 2025, LNNS 1651, pp. 409–421, 2026.
https://doi.org/10.1007/978-3-032-06688-6_40

across schools to all students. This makes systems that bridge the gap between the hearing impaired and non-impaired people an important domain requiring innovation. We propose a recognition system that could be used as an educational tool, leveraging the recent GenAI boom to have a sentence level recognition system compared to currently existing static recognition systems.

2 Related Work

Sign language detection can be categorized into two approaches- vision-based and glove-based. Vision based approaches are typically software solutions which depend on web-cams to capture data and are dependent on quality of the camera along with lighting conditions. Glove-based systems on the other hand utilize sensor data which is obtained from gloves that are used for signing.

According to research by Chandarana et al. (2023) [4] "The system uses Random Forest Classifiers to achieve an accuracy of 99.6% for Indian Sign Language, converting the recognised gestures into text and audio in both English and Hindi. This study emphasizes the capabilities of machine learning models to handle bilingual outputs efficiently." The implementation of Mediapipe with a Gated Recurrent Unit (GRU) model by Subramanian et al. (2022) [7] marked a significant improvement in real-time ISL gesture tracking. Their model showed promising results in enhancing recognition accuracy while optimizing processing time, making it suitable for real-time applications.

Singh et al. (2022) [2] used "Convolutional Neural Networks (CNNs) for the recognition of dynamic ISL gestures, achieving a training accuracy of 70%. The study highlights the challenges of recognizing complex and fluid gestures that vary in speed and shape." Similarly, Das et al. (2022) [6] introduced a hybrid approach that fused deep learning techniques with manually crafted features to boost recognition accuracy, stressing the importance of integrating diverse feature extraction techniques for better results.

Further advancements in real-time ISL gesture recognition were presented by Soni et al. (2023) [5], who unveiled a refined CNN architecture that amalgamated image classification and skin segmentation to improve gesture recognition accuracy. Their framework successfully processed both static and dynamic gestures for real-time applications. Kolkur et al. (2024) [1] presented "YOLOv3 for Real-time ISL Gesture Detection: This system employs object detection techniques to recognize hand gestures in real time and translate them into speech. YOLOv3's fast processing capabilities make it an ideal candidate for real-time ISL systems, delivering accurate results in dynamic environments. It is observed that the model gives 99.93% accuracy on test data."

DenseNet models were used by Beulah et al. (2023) [3] for ISL classification. Specifically, DenseNet169 was used for classifying static gestures. Their technique recorded an impressive accuracy of 99.95%, indicating the effectiveness of employing deeper network architectures to enhance gesture recognition.

However, challenges persist in the arena of sign language detection. Many existing systems struggle with accurately recognizing continuous sequences of dynamic gestures, which results in lower accuracy compared to static gesture recognition. Handling of temporal dynamics and variability in real-time signing is a very big gap in the research. Furthermore, enhancing the efficiency in processing data without trading off accuracy

remains a challenge. Most ISL-to-speech systems function at the word level, with limited advancements to convert these into semantically accurate sentences and subsequently into speech. Further research should thus be conducted in this direction in order to build ISL recognition systems that are capable of facilitating natural sign language communication in real-world scenarios.

3 Methodology

3.1 System Architecture

The Indian Sign Language Recognition System processes both, static gestures and dynamic gestures and converts them into speech and text. A dataset of statically signed images and sequential files for dynamic data is used to train deep learning models. During real-time recognition, the system captures live video input through the device's webcam and detects hand landmarks using computer vision techniques. For static signs, the extracted features are normalized and passed through a trained CNN-LSTM model to predict the corresponding letter or number. For dynamic gestures, a sequence of frames is analysed using the GestureNet model to identify words or phrases. The text, which is recognized from the signs, is then displayed on an interface and is then either converted into a word or into a semantically accurate sentence. This is then converted into speech for audio output, enhancing accessibility for users.

3.2 Dataset

Static

Indian Sign Language consists of a set of 26 signs for letters from A to Z and numbers from 0 to 9 as shown in Fig. 1. We downloaded a dataset from Kaggle containing the gestures for these classes. This dataset contains approximately 79200 images where each class has around 2200 images each (Figs. 1, 2, 3, 4, 5, 6, 7, 8, 9, and 10).

Dynamic

Dynamic gestures in Indian Sign Language (ISL) involve movements over time, requiring sequential processing instead of static hand shapes. We have created a custom dataset consisting of 14 gestures that vary across greetings, identity signs, temporal concepts and actions to represent a subset of ISL vocabulary.

These gestures are collected by utilising Python's OpenCV and Mediapipe libraries. Each recorded gesture is a sequential collection of 30 frames. To reduce noise and jitter in the data collection, we have implemented a moving average filter with a window size of 3 frames, separate filtering for hand and face landmarks and background subtraction to isolate the subject from the environment. OpenCV captures images in 'BGR' format so it is converted into 'RGB' format. By using Mediapipe's functions, 21 key-points are extracted for each hand and 468 key-points are extracted for the face, providing x,y and z coordinates. These key-points are then combined and flattened into a 1D array which is then stored as a.npy file.

To ensure quality assurance, we have also implemented a motion detection algorithm based on frame differencing to trigger data collection only when the participant is

Fig. 1. Image illustrating alphabets and numbers in the Indian Sign Language

actively performing a gesture. This is done by calculating absolute differences between consecutive frames and applying thresholding and morphological operations to identify significant movement. These steps help prevent false triggers for signing. At the very end, we balance the dataset to avoid data imbalance and skewness.

3.3 Preprocessing

The preprocessing stage ensures that both static and dynamic sign gestures are appropriately formatted for accurate recognition by the deep learning models. It involves data normalization, feature extraction, encoding, and standardization to maintain consistency across different hand positions, sizes, and gesture sequences.

Static

For static sign recognition the preprocessing stage configures the MediaPipe Hands detector to efficiently process video-like frames, ensuring robust and accurate hand gesture recognition. The detector is set to require a minimum confidence level of 50% for detection and tracking while allowing the identification of up to two hands per image.

To ensure position and scale invariance, we normalize the x and y coordinates relative to the hand's bounding box to preserve the hand's relative shape and position in the image. This normalization standardizes the data across different hand positions and sizes, preventing discrepancies in feature extraction due to variations in distance or placement.

We further process the image by converting it into an RGB format to align with the CNN model's input requirements. MediaPipe hand detection is then applied, and potential errors are handled by validating and sorting detected hands based on handedness.

The dataset is preprocessed by encoding categorical labels into numerical indices using label encoding and one-hot encoding labels to ensure compatibility with the model. The processed data is subsequently split into training and testing sets, maintaining a ratio of 80:20.

Dynamic
Certain pre-processing operations were performed on the dynamic dataset prior to training our model. To increase the robustness of our model and counter the weakness of a relatively small training dataset, we used data augmentation techniques specifically tailored for time-series data.

Time warping was used by repeating or skipping random frames within a sequence, mimicking variations in signing speed of various users. We also added random gaussian noise to landmark coordinates that mimicked natural variations in hand and face positioning to make the model more robust to slight variations in gesture performance.

These augmentations were generated dynamically so that each mini batch was presented with distinct augmentations, essentially increasing the training dataset and exposing the model to a greater variety of gesture variations. This aids in enhancing model generalization to real-time usage.

The dataset was then split using a stratified method so that 85% of the dataset was used for training and validation and 15% was used as a hold-out test set.

3.4 Training Model

Static
We use a combination of Convolutional Neural Networks (CNN) for extracting spatial features and Long Short-Term Memory (LSTM) networks for handling sequential data for training. It starts with a one-dimensional convolutional layer containing 64 filters of size 3, which helps in detecting local patterns within the input sequence. This is followed by a max-pooling layer that reduces dimensionality while retaining critical spatial information.

The extracted features are then passed through an LSTM layer with 128 units, allowing the model to capture temporal relationships in gesture sequences. To prevent overfitting, a dropout layer with a 50% rate is incorporated. Additionally, a fully connected dense layer with 128 neurons and ReLU activation enhances non-linear feature representation.

The final classification is performed using a softmax activation function, which assigns input sequences to their respective gesture classes. The model is optimized using the Adam optimizer for efficient learning, with categorical cross-entropy loss guiding training and accuracy serving as the key performance metric. We train the model for 50 epochs with a batch size of 32 to ensure robust learning.

Dynamic
We employed a Bidirectional Long Short-Term Memory (LSTM) architecture which we refer to as the GestureNet model. This is composed of an input layer followed by two bidirectional LSTM stacked layers of 160 and 96 units, respectively, allowing extraction

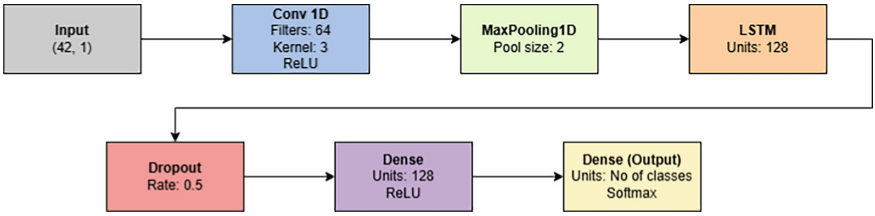


Fig. 2. CNN + LSTM Model Structure

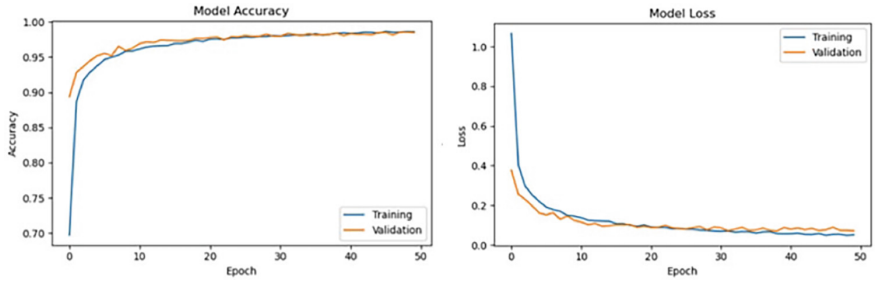


Fig. 3. Training vs Validation Accuracy and Loss of the Static Model

of temporal patterns from the forward and backward movements of a 30-frame sequence for each gesture.

Following this are two fully connected dense layers of 256 and 128 units that employ LeakyReLU activation together with batch normalization.

In between the LSTM and dense layers, dropout layers are inserted, as well as L2 regularization ($\lambda = 0.0003$) in order to counteract the possibility of overfitting the model.

Finally, a softmax-activated dense layer of 14 output classes matching the gestures is present. We trained the model for 200 epochs, and employed early stopping to track the validation loss using a patience limit of 35 epochs.

To achieve optimal training, we used cosine decay learning rate scheduler that reduced the learning rate step by step according to:

$$lr = lr_{initial} \times \frac{(1 + cos(\pi \times epoch/total\_epochs))}{2} \tag{1}$$

Additionally, the Adam optimizer is also used for gradient-based optimization. More advanced callbacks and model checkpointing have also been included to save the best-performing model.

The model is then tested using a 5-fold stratified cross-validation method, which gives comprehensive testing across various data segments and deals with the limitations of a small dataset size and possible data biases.

After cross-validation, the final model was trained on the full training and validation set (85% of the total data) and tested on the held-out test set (15%).

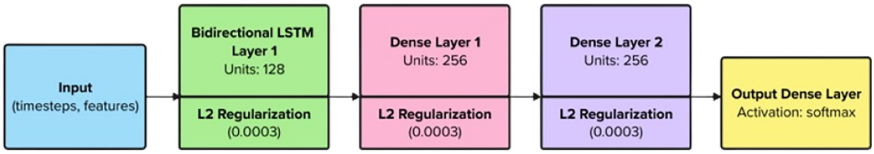


Fig. 4. GestureNet Model Architecture

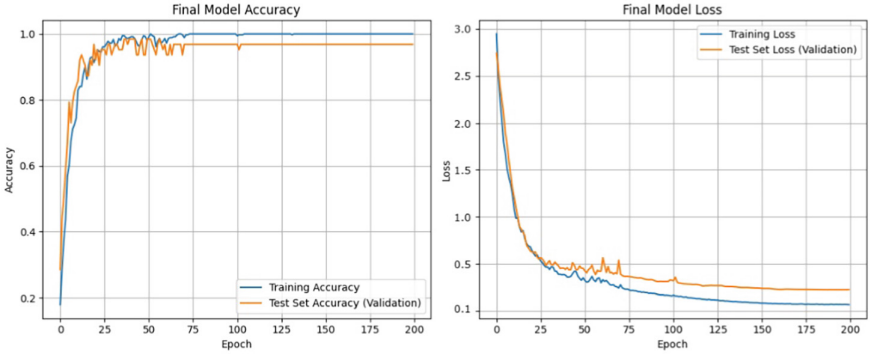


Fig. 5. Training vs Validation Accuracy and Loss of the Dynamic Model

3.5 Real-Time Detection

For real time static gesture detection, each frame is processed to detect hand landmarks and the pre-trained scaler is loaded for consistent feature normalization. The hand landmark detection parameters are adjusted to detect up to two hands with an 80% confidence threshold which ensures reliable detection. To improve hand detection and visual clarity, we combine background-only blurring and skin color segmentation. Background blurring uses a hand mask from MediaPipe landmarks to keep the hand sharp while blurring only the background. Skin color segmentation further refines this mask by isolating skin-toned regions in the frame. By merging both techniques, we more accurately separate the hand from the background, reducing false blurring and enhancing recognition performance. Our static recognition code includes a flexible 3-s timer, which was added to give users ample time to perform each sign comfortably. This timer can be adjusted or removed entirely, as it was primarily implemented for ease of signing during development and testing. After a user completes a gesture, the system takes about 40ms to analyze the input, recognize the gesture, and display the corresponding text.

Similarly, for dynamic gestures, the confidence thresholds are set high at 0.9 to reduce false positives and ensure accurate detection. Rather than classify each frame individually, the system implements temporal aggregation, that is a collection period for more reliable recognition. When a hand is detected, the system begins extracting keypoints from both hands and face across multiple frames, building a sequence of length 30 frames. A confidence-weighted voting system selects the most frequent prediction as

the final gesture, effectively filtering out transient movements and reducing misclassifications. Additionally, previously described measures to combat noise and jitter are also applied in the real time prediction.

3.6 Semantic Sentence Formation

The recognised dynamic gestures are added in a buffer and then sent as an API request to the Inference client of Meta Llama 3.1, along with a prompt asking it to create a semantically accurate sentence. This LLM then processes the request in the backend and retrieves back a meaningful sentence created by the gloss gestures passed to it while retaining the context of the signer. This is a novel idea that enables you to translate gestures at a sentence level instead of simply recognising and representing them as words. By leveraging Generative AI, we bypass the overhead of natural language processing steps that would have had to be otherwise carried out to create semantically meaningful sentences.

Fig. 6. Sign Language to Text Generation using Llama 3.1

3.7 Speech to Text

For static gesture recognition, the predicted letter is added to a buffer which forms the word. Similarly for dynamic recognition, the sentence retrieved from the Llama3.1 call is stored. These are passed to the pyttsx library, which converts the word or sentence into speech. The advantage of using pyttsx is its non-reliance on internet connectivity. This enhances accessibility for users.

3.8 Implementation Details

We have implemented our proposal as a system built with streamlit, a lightweight python-based web-framework as the frontend with our backend using TensorFlow, Keras, OpenCV, NumPy, and MediaPipe libraries. The system runs as a standalone streamlit application.

The training was done on systems having a multicore i5 processor with 16 GB RAM and a webcam of minimum 720p resolution.

4 Results

4.1 Performance of CNN-LSTM

The trained CNN-LSTM model achieves high classification accuracy. The training accuracy of the model is 98.59%, while the testing accuracy is 98.47%, indicating minimal overfitting and strong generalization to unseen data. The performance of the model is evaluated using accuracy metrics and training and validation loss curves. The hybrid architecture accurately distinguishes between multiple gesture classes.

4.2 Performance of GestureNet

The GestureNet model has very good classification accuracy. The final training accuracy for this model is 99.008% and the validation accuracy is 96.83%. The performance of the model is calculated using accuracy metrics and training and validation loss curves.

While the validation loss is still quite high, this is typical of gesture recognition models that have to deal with problems like high variability and noisier data. Since the validation accuracy is very high, the high loss doesn't deter the classification process (Table 1).

Table 1. Comparison of Accuracies of Various Models

Model	Static Gesture Accuracy (%)	Dynamic Gesture Accuracy (%)
CNN + LSTM	94.2	87.6
ResNet50	96.8	78.3
MediaPipe + Random Forest	95.4	82.7
3D-CNN	91.3	89.8
Vision Transformer	95.9	85.2
GCN (Graph Convolutional Network)	93.7	91.2
EfficientNet + GRU	94.8	88.9
GoogleNet + Optical Flow	90.6	86.1
Multimodal Fusion (CNN + RNN + Depth)	97.3	92.6
Custom ISL-Net	96.2	90.7
Proposed	**98.47**	**96.83**

4.3 Sentence Fluency and Semantic Evaluation

To test the sentence fluency and semantic evaluation of our GenAI outputs, we created test cases where we mapped gloss sequences to target sentences and applied surface-level and semantic metrics.

Traditional n-gram overlap measures such as BLEU-4 which is calculated by matching 1-to-4-word sequences between reference and generated text yielded a score of 0.43, which was expected given the flexible mapping between sign language glosses and natural language sentences. This was also reflected by the Word Error Rate (WER) inverse score which was 0.55.

However, despite these metrics being moderate, other semantic metrics yielded better results. The NIST score was calculated as 0.87, indicating that informative word choices were preserved. More importantly, the generated sentences demonstrated excellent fluency, achieving a perfect score of 1.0 on the fluency metric which we derived from text structure analysis including word diversity and length. Semantic similarity, computed via embedding-based methods, reached 0.78, suggesting that the generated sentences capture the intended meaning of the input gestures. We also calculated fuzzy matching scores such as the Fuzzy Token Set Ratio which handles word reordering by comparing word sets rather than exact sequences. This scored 0.93 which further supported the observation that while the word order or specific tokens may vary, the core content is well-aligned with reference sentences.

Overall, we observed that the system generated semantically accurate and fluent sentences which aligns well with the objective of sign language gloss-to-sentence conversion, where semantic equivalence should be prioritized over strict lexical overlap.

5 Output

5.1 Static

The final output of the model is a predicted gesture class corresponding to the input hand gesture which could be a letter or a number. The predicted gesture is displayed in real-time, allowing users to visualize the classification results.

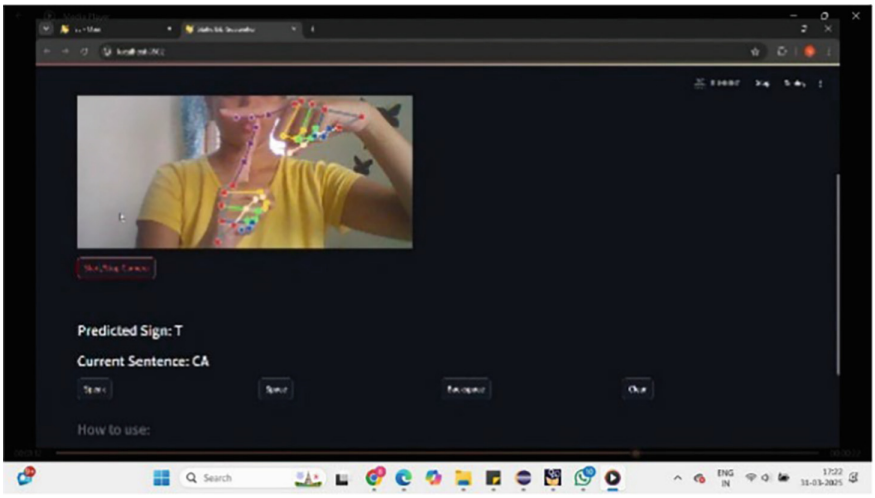


Fig. 7. Static Gesture Recognition of ‘T’

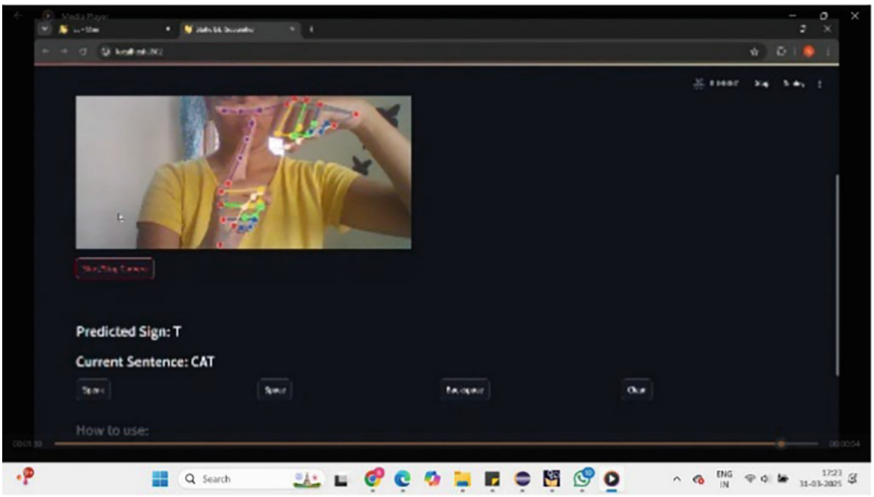


Fig. 8. Static Gesture Recognition – Final output: 'CAT'

5.2 Dynamic

The final output of the model is a predicted gesture class corresponding to the input hand gesture which could be a word or a phrase. The predicted gesture is displayed in real-time, allowing users to visualize the classification results.

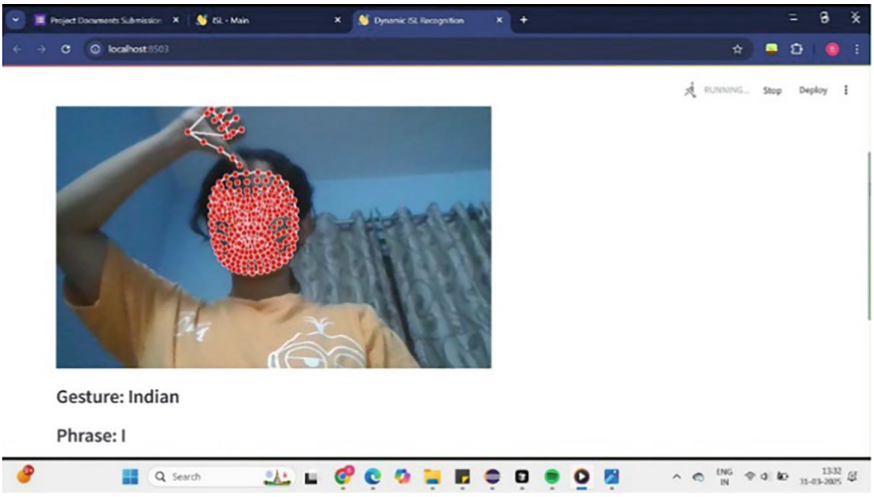


Fig. 9. Dynamic Gesture Recognition of 'Indian'

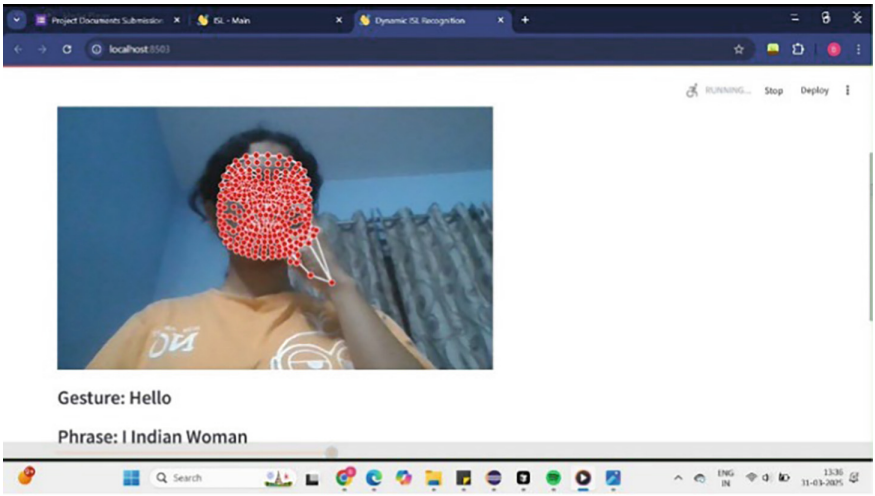


Fig. 10. Dynamic Gesture Recognition of 'Woman'

6 Conclusion

A real-time vision-based system that recognises Indian sign language gestures and converts them to semantically accurate sentences by leveraging Generative AI is an avenue that holds immense potential in modern India. However, several gaps exist in this proposal; first and foremost, the development of a combined model that recognises both static and dynamic signs is necessary to further bridge the gap between research done so far that primarily targets static gestures or dynamic gestures individually. Another important aspect is the size of the dynamic gesture dataset which could be increased so as to include a more comprehensive vocabulary. Further provisions could be made so as to convert the sentences formed into regional, vernacular languages to increase accessibility and enhance user experience. Further research focused on these limitations would enable the creation of a more holistic system and perhaps more innovative models that expand current avenues.

Acknowledgement. All authors have contributed equally to this work.

We deeply appreciate the guidance and support provided by our mentor, Prof. Shilpa Pant, from the Computer Science Engineering Department for her essential guidance, unwavering support, and constant encouragement during this research. Her specialized knowledge and observations have been instrumental in shaping this work.

We also extend our heartfelt appreciation to MKSSS's Cummins College of Engineering for Women, Pune, for providing the required resources and a conducive environment for research. This project would not have been possible without the institution's support.

Finally, we acknowledge the efforts of all those who contributed directly or indirectly to this research. Their inputs and feedback have helped us refine our work and move closer to our goal of advancing Indian Sign Language recognition technology.

References

1. Kolkur, A., Yattinmalgi, A., Korimath, G., Chikkamath, S., Nirmala, S.R., Budihal, S.V.: Deep learning based indian sign language recognition for people with speech and hearing impairment. pp. 1–5 (2024). https://doi.org/10.1109/inc460750.2024.10649094
2. Lu, C., Kozakai, M., Lei, J.: Sign language recognition with multimodal sensors and deep learning methods. Electronics **12**, 4827 (2023). https://doi.org/10.3390/electronics12234827
3. Beulah, I.K., Raimond, K., Miraclin, G.L.: Indian sign language recognition for static gestures using DenseNet169 model (2023). https://doi.org/10.1109/icces57224.2023.10192769
4. Chandarana, N., Manjucha, S., Chogale, P., Chhajed, N., Tolani, M., Edinburgh, M.: Indian sign language recognition with conversion to bilingual text and audio. pp. 1–7 (2023). https://doi.org/10.1109/ICACTA58201.2023.10393571
5. Soni, R., Vijay, A., Khandelwal, A., Vijay, R., Yadav, V., Bhatia, D.: Real-time recognition framework for Indian Sign Language using fine-tuned convolutional neural networks. In: Rana, P.S., Bhatia, D., Arora, H. (eds.). SCRS Proceedings of International Conference of Undergradu-ate Students, SCRS, India, pp. 95–106 (2023). https://doi.org/10.52458/978-81-95502-01-1-10
6. Das, S., Biswas, S.Kr., Purkayastha, B.: Automated Indian sign language recognition system by fusing deep and handcrafted feature. Multimedia Tools Appl. **82**, 16905–16927 (2022). https://doi.org/10.1007/s11042-022-14084-4
7. Subramanian, B., Olimov, B., Naik, S.M., Kim, S., Park, K.-H., Kim, J.: An integrated mediapipe-optimized GRU model for Indian sign language recognition. Dental Sci. Rep. **12** (2022). https://doi.org/10.1038/s41598-022-15998-7

Deep Learning For IoT Data Analytics

Abhinav Thakur(✉), Bhushan, and Ashima Mehta

Department of Computer Science Engineering, Dronacharya College of Engineering, Gurugram, India
{abhinav.26009,bhushan.26045,ashima.mehta}@ggnindia.dronacharya.info

Abstract. The fast growth of the Internet of Things (IoT) has produced an immense pool of smart, networked devices generating enormous amounts of data every day. From transport and health to agriculture and intelligent cities, these devices are becoming must-haves for service improvement, operation optimization, and innovation stimulation. However, the sheer number and diversity of data coming from IoT sources overwhelm traditional data analysis methods. This is where deep learning steps inoffering great tools that can learn patterns, detect unusual pattern, and process information in real time. Of them, Long Short-Term Memory (LSTM) networks have shown highly promising results, particularly for sequential data analysis such as time series, and integration with probabilistic models increased accuracy and reliability even more. This paper discuss significant techniques and uses of deep learning to revolutionize IoT data analytics, investigating important techniques and uses, mentioning present opportunities, challenges, and new promising research opportunities before us in this rapidly changing field.

Keywords: Deep learning techniques · Applications and challenges of Deep Learning For IOT Data Analytics · Future Trends and Advancements · Model Comparisons and Real-World Insights

1 Introduction

With time, machines such as sensors, actuators, and smartphones have grown smarter, able to execute sophisticated operations and interact with one another smoothly. Indeed, by 2008, there were already more connected devices than the world population a figure that has been increasing exponentially ever since [7]. In current times, which are characterized as the era of the Internet of Things (IoT), countless devicesfrom phones to wireless sensors and embedded systemsare interlinked, thereby facilitating novel uses across sectors of industry [14]. As the IoT ecosystem grows, it produces a vast volume of data that must be analyzed on time and in an accurate manner. For this purpose, technologies such as Artificial Intelligence (AI), Machine Learning (ML), and Deep Learning (DL) are increasingly utilized to facilitate wiser, data-based decisions [5]. While methods like the classical logistic regression and random forests have assisted in

S. Fong et al. (Eds.): ICT4SD 2025, LNNS 1651, pp. 422–432, 2026.
https://doi.org/10.1007/978-3-032-06688-6_41

aspects such as traffic forecasting and tracing [14], they tend to be inadequate when handling the scale and nuance of the IoT data [15].

Deep learning, on the other hand, excels at finding complex patterns in massive datasets, making it a powerful tool for applications such as health monitoring, traffic forecasting, and delivery optimization [10]. This paper explores how deep learning is used in IoT analytics, its challenges, and future directions [1]

2 Deep Learning Techniques for IOT Data Analytics

In order to make full use of IoT and big data, it's crucial to comprehend their relationship and potential. IoT devices produce massive amounts of data, whereas smart analysis methods can optimize IoT systems and improve services [11]. The true value lies in processing this IoT data, which is of great advantage to businesses and society [2]. Comprehending the special characteristics of IoT data is fundamental to efficiently dealing with and analyzing it for the optimal results [8,9].

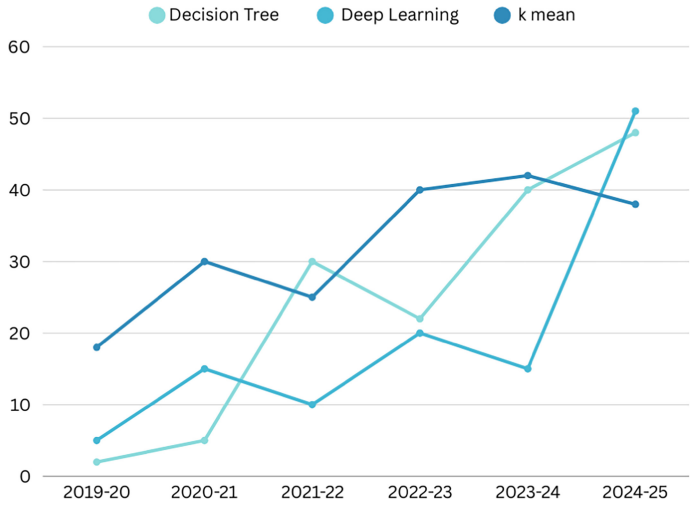


Fig. 1. Google trends show the growth of deep learning in recent years.

2.1 Recurrent Neural Networks (RNNs)

- Designed for time series and for sequential data [14]. Used in predictive maintenance, weather forecasting, and anomaly detection [5].
- Example: Prediction of energy consumption patterns in smart grids [3]

2.2 Long Short-Term Memory (LSTM) Network

- It is a special type of RNN which addresses long-term dependency problems [12]. LSTM is used in healthcare monitoring, predicting traffic flow, and financial analytics [15].
- Example: Monitoring of real-time heart rate data from wearable sensors with the help of LSTM [13].

2.3 Gated Recurrent Units (GRUs)

- Similar to LSTMs but computationally more efficient [2] Used in real-time IoT analytics, speech recognition, and it is also used for predictive modeling [18].
- Example: Can forecast water demand based on previous sensor readings [4].

2.4 Autoencoders

- Used for Dimensionality reduction and helps to detect unusual patterns [8] Helps in cybersecurity, error detection, and data compression [11].
- Example: Detecting unusual patterns or traffic on IoT-based security system [3].

2.5 Transformer Models

- Used for Natural Language Processing (NLP) and Time-Series prediction [3] Helps in chat-bots, smart assistance, and real-time sensor detection [12].
- Example: Analyze problems to detect the system faults [11].

2.6 Deep Belief Networks (DBNs)

- Used in unsupervised learning and feature extraction [10] Helps in image processing, recognition of speech, and grouping similar data points together [18].
- Example: Identifying patterns in IoT-enabled smart home devices [9].

2.7 Convolutional Neural Networks (CNNs)

- Best for image and data that represent the location in IoT [3] Used in smart surveillance, industrial automation, and medical imaging [14].
- Example: Detection of defects in manufactured products using sensor-captured images [12].

2.8 Generative Adversarial Networks (GANs)

- Used for data modification, artificial data, and enhancement [1] Helps in medical image analysis, smart city simulations, and IoT cybersecurity [2].
- Example: Generating artificial sensor data to train deep learning models [4].

3 Applications of Deep Learning For IOT Data Analytics

Deep learning-based techniques have established itself as a dominant and impactful force in IoT data analytics by enabling greatly enhanced processing and interpretation of voluminous sensor data [3]. Through the application of complex heuristics and neural networks, deep learning enables systems to automatically recognize complex patterns, detect unusual anomalies, predict future patterns, and make optimal different processes [4]. This technology is transforming industries by allowing smarter decision-making, real-time forecasting, and automation [r14]. Let us examine in detail some of the major applications where deep learning is driving innovation and efficiency in IoT data analytics [18].

3.1 Smart Cities

- Managing Traffic Flow: CNNs analyze live traffic conditions from cameras to predict the alternative routes or best routes [4].
- Monitoring Air Quality: Deep learning models forecast pollution levels by collecting and analyzing sensor data from various urban locations [18].
- Optimizin g Waste Collection: AI predicts waste accumulation rates and schedules garbage collection accordingly [11].

3.2 HealthCare IoT

- Remote Health Monitoring: LSTMs analyze continuous health data such as heart rate, temperature, and oxygen rate in the body for real-time health tracking [10].
- Early Disease Detection: CNNs interpret medical images (e.g., X-rays) for diagnosing diseases [17].
- Detecting Vitals Anomalies: Autoencoders recognize unexpected changes in vital signs of a patient, alerting healthcare providers [18].

3.3 IOT Security

- Cybersecurity Monitoring: AI models continuously check network traffic from IoT devices to detect potential threats or attacks [6].
- Real-Time Surveillance: CNNs analyze live video streams to recognize unauthorized individuals in secure areas [14].
- Behavioral Anomaly Detection: AI detects unusual movement patterns in surveillance systems, to detect security breaches [10].

3.4 Smart Retail

- Customer Insights: Deep learning models analyze consumer interest in products to offer personalized product recommendations [7].
- Inventory Management: AI predicts product increasing demands and ensures stores are always stocked with that product [11].
- Supply Chain Optimization: AI uses RFID tags and sensor data to track product movement in real time, improving logistics efficiency [10].

3.5 Wearable IOT Devices

- Health Monitoring: LSTMs track heart rate fluctuations and alert users of their health risks [5].
- Sleep Pattern Analysis: Deep learning models analyze sleep data to detect disorders like insomnia or apnea [9].
- Fall Detection: CNNs use motion sensor data to detect falls, particularly for elderly care [12].

4 Challenges in Applying Deep Learning to IOT Data Analytics

The integration of IoT devices into deep learning (DL) systems would allow networks to get useful information for smart cities, healthcare, and transportation sectors [2]. DL can revolutionize how AI takes IoT data analytics to the next level [7]. However, the implementation of DL has not been fully utilized due to a couple of barriers [14]. The quality of the data collected, the limited resources of IoT devices, and the development of efficient models to operate within the constraints of devices are some of the challenges faced [15]. Eliminating these barriers will allow deep IoT learning to create comprehensive and practical solutions across industries [3].

4.1 Insufficient Data for Effective Training

- One of the primary challenges when applying deep learning to IoT is the lack of large, high-quality datasets [7]. DL models thrive on vast amounts of data, but in many IoT applications, especially those in specialized or emerging fields, data is often limited or difficult to gather [11]. Without enough data, Deep Learning models struggle to learn effectively, leading to suboptimal performance [2].

4.2 Preprocessing Complexities

- IoT systems collect data from a variety of sources, such as sensors, often in different formats and with inconsistent structures [15]. To feed these raw data into DL models, extensive pre-processing is required to clean, normalize, and organize it [10]. This becomes even more challenging when dealing with missing values, noisy data, and sensor errors [9]. Ensuring that data are ready for DL models while handling their diversity is a major task in IoT analytics [8].

4.3 Handling Resource Limitations in IoT Devices

– IoT devices typically have limited computational power, memory, and storage, which makes running deep learning models directly on them difficult [1]. While cloud computing and edge computing can help overcome some of these issues, managing and optimizing DL models for devices with minimal resources is still a significant challenge [2]. Additionally, the need to transmit data from these devices to the cloud or edge servers often results in delays, network failures, or data bottlenecks [11].

4.4 Concerns Over Data Privacy and Security

– One more difficulty when implementing deep learning to IoT data is maintaining the confidentiality and the privacy of the data that is collected and shared [9]. Several IoT gadgets gather sensitive personal information, including, but not limited to, health and geo location data, which is always vulnerable once it sent over networks [1]. Strong security to safeguard information while following privacy is very important in achieving success with DL IoT systems [3].

4.5 Limitations of DL in Regression Tasks

– Deep learning models are typically designed for classification tasks, where data is categorized into predefined labels [1]. However, most IoT applications require models that can predict continuous variables, such as temperature or pressure readings [19]. Although progress has been made in applying deep learning to regression tasks -such as combining deep belief networks with support vector regression (SVR) - this area remains underdeveloped for IoT applications [8].

5 Future Trends and Advancements in Deep Learning for IoT Data Analytics

The integration of deep learning with IoT is revolutionizing data analytics by enabling intelligent, real-time decision-making across diverse applications [7]. As IoT devices generate massive and varied data streams, deep learning helps extract meaningful patterns, improving automation and efficiency [12]. However, challenges such as data heterogeneity, limited device processing power, and real-time constraints still hinder seamless deployment [19]. To address these, cutting-edge technologies like Edge AI, federated learning, and blockchain-based security frameworks are being developed [13]. These innovations aim to enhance data privacy, reduce latency, and support decentralized intelligence [5]. Ultimately, such advancements are paving the way for highly scalable, adaptive, and secure IoT ecosystems [9]As these innovations evolve, they will reshape how IoT systems collect, process, and act on data [3]. This progress is crucial for building future-ready, intelligent, and resilient IoT environments [3](Figs. 1 and 2).

5.1 Edge AI and Learning for Decentralized Processing

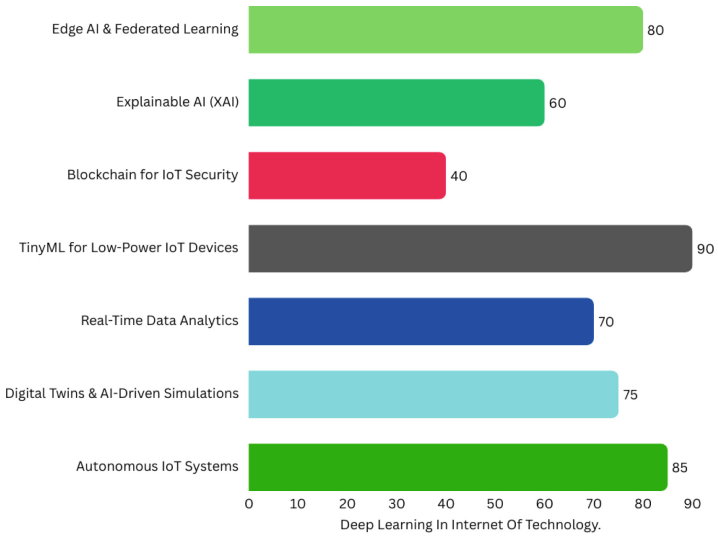


Fig. 2. Future Trends in Deep Learning For IOT.

– IoT devices typically have limited computing power and battery life, making it difficult to deploy large deep learning models [9]. A promising future trend is the development of TinyML, which focuses on creating ultralight deep learning models for low-power IoT devices [15]. Techniques like model compression, pruning, and quantization will enable deep learning algorithms to run efficiently on micro-controllers and embedded IoT systems, bringing AI capabilities to even the smallest edge devices [8].

5.2 Autonomous IoT with Deep Reinforcement Learning

– Deep reinforcement learning (DRL) is drawing interest as a technique for autonomous IoT decision-making [11]. DRL, which is different from conventional deep learning, enables IoT systems to learn from mistakes and become more efficient over a period [6]. Autonomous smart grids, self-optimizing industrial IoT systems, and AI-based robotic process automation (RPA) that are capable of making smart decisions autonomously are potential future uses [10].

5.3 TinyML and AI Models for Low-Power IoT Devices

– The majority of IoT devices do not have the processing capabilities and battery life to handle big deep learning models [9]. An exciting future direction is TinyML, which is centered on developing lightweight deep learning models for low-power IoT devices [15]. Methods such as model compression, pruning, and quantization will enable deep learning algorithms to be run on microcontrollers efficiently, bringing AI capabilities to even the tiniest edge devices [12].

5.4 Explainable AI (XAI) for IoT Decision-Making

– Deep learning models are often seen as"black boxes" due to their complex, opaque decision-making processes [9]. As IoT applications increasingly rely on AI-driven insights, Explainable AI (XAI) is becoming essential to ensure trust, transparency, and accountability [10]. Future research will focus on developing interpretable deep learning methods that provide clear explanations for AI-driven decisions in IoT systems, especially in critical areas like healthcare, finance, and smart infrastructure [11].

5.5 Real-Time Deep Learning for High-Velocity IOT Data

– As IoT devices produce enormous and unending streams of data, the necessity for real-time deep learning models is increasingly pressing [8]. The future will witness the convergence of stream processing frameworks, online learning algorithms, and predictive analytics in real-time to manage high-velocity IoT data [14]. Breakthroughs like spiking neural networks (SNNs) and neuromorphic computing could also contribute to empowering deep learning models to process IoT data with very little energy expenditure [12].

6 Model Comparisons and Real-World Insights

We provide a comparative overview of deep learning models used in different IoT domains to support the conceptual approach in this work. The main conclusions from current research are compiled in this table, which also illustrates how various models fare in [15]terms of accuracy and latency across typical IoT applications[11](Table 1).

Table 1. Performance of Deep Learning Models in IoT Applications

DL Model	IoT Application	Dataset	Accuracy (%)	Latency (ms)
CNN	Smart Home Activity Recognition	WARD	94.5	120
LSTM	Health Monitoring	MIT-BIH	91.2	135
GRU	Environmental Sensor Prediction	RealIoT	89.8	110
DNN	Traffic Flow Prediction	METR-LA	88.7	150

These quantitative benchmarks highlight the trade-offs associated with model selection as well as how different DL models can be tailored to particular IoT use cases[17]. For example, LSTM and GRU are better suited for time series data in healthcare and environmental monitoring, whereas CNNs provide excellent accuracy in structured contexts such as smart homes[18].

6.1 Case Study: Deep Learning in Smart Agriculture

– Through the use of an LSTM-based system to predict soil moisture and optimize irrigation schedules using real-time data, [5] provides a practical usage of deep learning in smart agriculture. The model achieved an RMSE of 3.7 percent and reduced water usage by 18 percent in comparison to traditional methods, [10] demonstrating how deep learning can drive sustainability while maintaining predictive accuracy in data-sensitive environments [5].

6.2 Discussion: AccuracyPrivacyLatency Trade-Off

– IoT systems must balance accuracy, privacy, and delay, especially with sensitive data.[17] Centralized deep learning boosts accuracy but risks privacy and adds delay [16] Federated learning improves privacy but may reduce model performance. Edge-deployed lightweight models with federated averaging offer a balanced solution. They preserve privacy, reduce delay, and maintain reasonable accuracy. Future research should optimize such models under real-time and ethical constraints [14].

7 Conclusion

Deep learning is revolutionizing data processing from the Internet of Things (IoT) to lead to enormous gains in healthcare, transportation, smart cities, agriculture, and security. The traditional approaches were weighed down by the amount and intricacy of data created by IoT devices but could be identified by deep learning, which is able to identify complex patterns and make more accurate judgments. These models such as RNNs, LSTMs, CNNs, Autoencoders, and

Deep Belief Networks have been widely applied for anomaly pattern detection, predictive maintenance, and real-time monitoring of IoT systems.

Even with its innovations, privacy, high computation requirements, and scalability remain challenges to it. Distributed learning, where IoT devices learn models together, is among the solutions under investigation. It minimizes latency and maximizes privacy but may also introduce processing time and maximized privacy issues.

References

1. Adi, E., Anwar, A., Baig, Z., Zeadally, S.: Machine learning and data analytics for the IOT. Neural Comput. Appl. **32**, 16205–16233 (2020)
2. Afshan, N., Rout, R. K.: Machine learning techniques for IOT data analytics. Big data analytics for internet of things, pp. 89–113 (2021)
3. Akbar, A., Khan, A., Carrez, F., Moessner, K.: Predictive analytics for complex IOT data streams. IEEE Internet Things J. **4**(5), 1571–1582 (2017)
4. Ravesa Akhter and Shabir Ahmad Sofi: Precision agriculture using IOT data analytics and machine learning. J. King Saud Univ.-Comput. Inf. Sci. **34**(8), 5602–5618 (2022)
5. Al-Amri, R., Murugesan, Man, R.K.M., Abdulateef, A. F., Al-Sharafi, M. A., Alkahtani, A.: A review of machine learning and deep learning techniques for anomaly detection in IOT data. Applied Sciences, 11(12):5320 (2021)
6. Alsheikh, M. A., Niyato, D., Lin, S., Tan, H., Han, Z.: Mobile big data analytics using deep learning and apache spark. IEEE network 30(3):22–29 (2016)
7. Atitallah, S. B., Driss, M., Boulila, W., Ghézala, H. B.: Leveraging deep learning and IOT big data analytics to support the smart cities development: Review and future directions. Computer Science Review, 38:100303 (2020)
8. Azar, J., Makhoul, A., Barhamgi, M., Couturier, R.: An energy efficient IOT data compression approach for edge machine learning. Futur. Gener. Comput. Syst. **96**, 168–175 (2019)
9. Jane, J.B., Ganesh, E.N.: Big data and internet of things for smart data analytics using machine learning techniques. In: Proceeding of the International Conference on Computer Networks, Big Data and IoT (ICCBI-2019), pp. 213–223. Springer (2020)
10. Bi, H., Liu, J., Kato, N.: Deep learning-based privacy preservation and data analytics for IOT enabled healthcare. IEEE Trans. Industr. Inf. **18**(7), 4798–4807 (2021)
11. Jamal Bzai., et al.: Machine learning-enabled internet of things (IOT): Data, applications, and industry perspective. Electronics, 11(17):2676 (2022)
12. Chahal, A., Gulia, P.: Deep learning: a predictive IOT data analytics method. Int. J. Eng. Trends Technol.-IJETT **68**(7), 25–33 (2020)
13. Dargazany, A. R., Stegagno, P., Mankodiya, K.: Wearabledl: wearable internet-of-things and deep learning for big data analytics—concept, literature, and future. Mobile Inf. Syst. 2018(1):8125126 (2018)
14. Ghosh, A. M., Grolinger, K.: Deep learning: edge-cloud data analytics for IOT. In: 2019 IEEE Canadian conference of electrical and computer engineering (CCECE), pp. 1–7. IEEE (2019)

15. Ananda Mohon Ghosh and Katarina Grolinger: Edge-cloud computing for internet of things data analytics: embedding intelligence in the edge with deep learning. IEEE Trans. Industr. Inf. **17**(3), 2191–2200 (2020)
16. Hannun, A. Y., et al.: Cardiologist-level arrhythmia detection with convolutional neural networks. Nature Medicine (2019)
17. Ignatov, A., et al.: Real-time human activity recognition on smartphones using deep neural networks. Sensors (2020)
18. Li, T., et al.: Deep learning-based environmental sensor forecasting for smart cities. IEEE Internet of Things Journal (2022)

Sustainability of Avian Monitoring Near Mobile Base Stations Using Drones: A Case Study in Arambagh Municipality, Hooghly, West Bengal, India

Sauvik Bose[1(✉)], Rina Bhattacharya[1], and Rajeshwari Roy[2,3]

[1] Department of Physics, JIS University, Kolkata 700109, West Bengal, India
sauvik.bose@gmail.com, hod_physics@jisuniversity.ac.in
[2] Department of Environmental Studies, Rabindra Bharati University, Kolkata 700007, West Bengal, India
[3] Society for the History of Science Kolkata, Kolkata, West Bengal, India
royrajeshwari455@gmail.com

Abstract. Avian monitoring is a crucial component of biodiversity conservation, providing insights into population trends, habitat changes, and environmental stressors. The fast growth of mobile telephony has raised issues regarding its potential upon the avian population, their behaviors and breeding, predominantly due to electromagnetic radiation exposure. This study investigates the feasibility of using drones for avian monitoring near mobile towers in Arambagh Municipality (22.8838° N, 87.7819° E), Hooghly, West Bengal, India, which is a semi-urban landscape with rich avian diversity and has undergone a significant growth in mobile tower installation over the last few decades. Drones offer a non-invasive, scalable, and high-resolution method for ecological monitoring, surpassing traditional survey techniques in terms of not only efficiency and data accuracy but also consuming less time and effort. A drone (model: DJI MAVIC MINI) equipped with a high-resolution camera is deployed at selected base station sites within the study area. The study pattern included regulated flight patterns, periodic monitoring. Findings disclosed noticeable behavioral variations in birds near mobile base stations. The repulsion of smaller birds to the high EMR zone has been distinctly observed along with anomalies in roosting and breeding habits. A correlation was observed between radiation levels and avian health oddities, underscoring the need for further research. Inthe future, research ought to be performed on in-depth monitoring efforts in urban and semi-urban areas along the different geographical landscapes. Improving drone technology for ecological studies and exploring alternative communication infrastructures with reduced environmental impact is much needed.

Keywords: Avian Monitoring · Mobile Base Stations · Drones for Ecology · Electromagnetic Radiation · Sustainable Wildlife Conservation · Arambagh Municipality Biodiversity

S. Fong et al. (Eds.): ICT4SD 2025, LNNS 1651, pp. 433–443, 2026.
https://doi.org/10.1007/978-3-032-06688-6_42

1 Introduction

Avian monitoring involves observing the bird populations, their behaviors, migratory patterns, and habitat usage systematically. It is vital to the climate and environmental research as the avians are one of the key bio-indicators of environmental incongruity [1]. It is of assistance for conservationists and policymakers to identify any disturbances in nature, be it natural or anthropogenic, by keeping bird records [2]. Birds, being one of the top-level consumers in the food pyramid, contribute to ecosystem stability in various ways, including pest control, pollination, seed dispersal, and maintaining food web connections [3].

Unfortunate to note that several wild bird species have vanished, and a few more are on the verge of extinction. This could be a caution marking the abrupt changes in the surroundings. Among the evident changes, urbanization could be counted as one of the most dynamic ones. In the current study, we have selected the Arambagh Municipality in West Bengal, India, as a sample site having rapid urbanization and technological expansion during the last few decades. The site selection for avian monitoring becomes even more vital in assessing the long-term impacts of human activities, including the spread of mobile base stations [4].

2 The Impact of Mobile Base Stations on Birds

Mobile base stations, which are indispensable for modern communication networks, are studied nowadays for their potential impact on avian species [5]. Concerns about the Electromagnetic Radiation (EMR) released from the towers have been raised due to their effects on bird physiology and behavior [6]. Several scientific studies reported that prolonged exposure to electromagnetic fields (EMF) may have a potential impact on birds. The following are the important impacts:

2.1 Disorientation and Navigational Issues

Migratory species, depends on the Earth's magnetic field for navigation. EMF emissions from mobile base stations may disrupt their ability of flight patterns by changing original migration routes [7]. The feathers of birds are dielectric, and so their navigation pattern is affected by the electromagnetic waves [8].

2.2 Reproductive and Growth Threat

Long exposure to EMR may affect the productivity, egg growth, and chick persistence rates [9]. In the long run it has impact on population stability.

2.3 Comportment Switches

Birds living nearby to mobile towers may switches their nesting preferences, communication patterns, and feeding activities [10]. Some of the avian species may avoid habitats in the proximity of mobile tower due to increased electromagnetic clouds, while others may adapt or show resilience.

2.4 Health Challenges

There are debates about the exposure level of EMR on avian. However, concerns may include increased physiological stress, variation in immune responses and metabolic functions [11].

However, framing of sustainable method for monitoring bird populations in the proximity of mobile towers so that any changes of their behavior or health may identify immediately.

3 The Impact of Mobile Base Stations on Birds

Conventional avian monitoring approaches, viz, direct field observations and satellite tracking, have some restrictions in terms of convenience and accuracy. UAVs (Unmanned Aerial Vehicles), *viz* Drones, offer a transformational application to the monitoring of avian population besides ecological monitoring [12]. Their ability to fly at various altitudes and access remote areas makes them an unfeasible tool for observing birds in various habitats.

Advantages of using drones for avian monitoring are the following:

3.1 Real-Time Observation

Drones can capture real-time footage from a distance and thereby reducing human disturbances to the birds and their nests.

3.2 Increased Data Accuracy

Capturing high-resolution imagery enables proper identification of avian species and population assessments [13].

3.3 Efficacy and Success

Drones can cover larger areas, including dense forest, water bodies, or restricted zones beside mobile towers, compared to manual surveys in comparatively minimal time, and so remote monitoring using drones is sustainable and cost-effective.

With the advancement of technology, drones can also be integrated with machine learning (ML) algorithms to automate data analysis, capable of further improving the efficacy of avian monitoring programs [14].

In our study we used drone model DJI Mavic Mini. The specification and picture of the drone are given Table 1. And Fig. 1 respectively.

Table 1. Specification of the drone

Model Name	DJI Mavic MINI
Take-off Weight	249 g
Dimensions	245 × 289 × 55 mm (with propellers)
Maximum take-off Altitude	3000 m
Battery Capacity	2400 mAh
Max Flight Time	30 min (measured while flying at 14 kph in windless conditions)
Camera Sensor	1/2.3″ CMOS Effective Pixels: 12 MP
Lens	FOV: 83° 35 mm Format Equivalent: 24 mm Aperture: f/2.8 Shooting Range: 1 m to ∞
ISO Range	100–3200
Shutter Speed	Electronic Shutter: 4–1/8000s
Still Image Size	4:3: 4000 × 3000 16:9: 4000 × 2250
Video Resolution	2.7K:2720 × 1530 FHD:1920 × 1080

Fig. 1. DJI Mavic Mini (https://www.dji.com/support/product/mavic-mini)

4 Overview of Arambagh Municipality: Geographical and Ecological Context

Arambagh Municipality (22.8838° N, 87.7819° E) is located in the Hooghly district of West Bengal, India (Fig. 2 and Fig. 3). The region is characterized by a hybrid of urban and rural landscapes. It is situated beside the Dwarakeswar River and has a subtropical

climate. The Arambagh municipality has densely populated urban settlements interspersed with agricultural fields, wetlands, and patches of greenery that serve as habitats for various bird species. The rapid urbanization with the expansion of mobile networks makes the place challenges for avian sustainability.

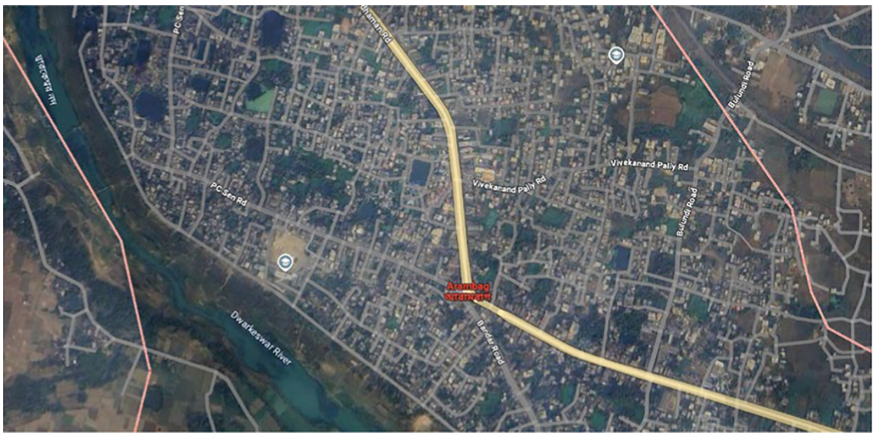


Fig. 2. Two-dimensional view of study area (source: Google Map)

Fig. 3. Photograph of the study area

5 Overview of Arambagh Municipality: Geographical and Ecological Context

The study followed a multiple approach, starting with avian activity in the habitat near the mobile base stations and drone-assisted surveys. The main objectives in designing the workflow are:

1. To observe the occurrence, abundance, and diversity of bird species near and away from the mobile base stations.
2. To scrutinize changes in bird behavior, nesting patterns, and flight routes near mobile towers.
3. To distinguish the effectiveness of drone-based monitoring in comparison to traditional methods.

5.1 Preliminary Field Surveys and Site Selection

Before deploying drones, initial field surveys were performed to identify major clusters of avian nesting sites near mobile base stations. Local bird counts were documented through direct observations using binoculars and camera recordings. Citizen reports and prior ornithological records were reviewed to supplement preliminary data.

5.2 Drone-Assisted Monitoring

Drones were utilized for efficient surveys at selected mobile base station sites. Flight paths were carefully projected to decrease the disturbance to bird activities while ensuring comprehensive data collection. Video recordings, thermal imaging, and high-resolution photographs were captured to analyze bird occurrence, movement, roosting, and nesting patterns. Figure 4 shows clips from avian monitoring sessions behind the mobile towers.

Fig. 4. Drone photographs during bird monitoring sessions

5.3 Data Compilation and Analysis

Collected data from drones and ground observations were assembled for investigation. Bird species identification was conducted using field guides and automated AI-based classification tools. The impact of electromagnetic radiation on bird activity was inferred by comparing avian presence and behavior at different distances from mobile towers.

5.4 Flight Patterns and Monitoring Parameters

To ensure consistency in data collection, standard flight patterns were designed based on study site features.

Grid Pattern Surveys. Drones flew in a preplanned grid pattern at each site to cover maximum visual area efficiently.

Altitude Considerations. *30–50 m (low altitude).* To note the nesting and roosting of birds on the tree trunks, crevices, and holes.

80–100 m (mid-altitude). To mark bird movements and overall population density.

Above 100 m (high altitude). To investigate migration and local navigation.

Flight Duration. 20–30 aerial minutes per flight was selected as ideal for observation in a single shot, with multiple such flights of UAV having successfully been completed at different times of the day (morning, afternoon, and evening).

Time-Based Data Collection. *Morning (5:30 AM – 8:00 AM).* Gathering food, feeding, and movement monitoring.

Noon (12:00 PM – 2:00 PM). Observations of roosting and nesting behaviour.

Evening (5:00 PM – 7:00 PM). Capture of avian flight and roosting.

Electromagnetic Radiation Levels. Measured by electromagnetic smog meter (Table 2) at different distances from base stations, along with the occurrences of birds and their nests.

Table 2. Specification of Electrosmog Meter

Model Name	MECO 9720
Measurement method	Digital, triaxial measurement
Directional characteristic	Isotropic, triaxial
Display refresh rate	Typically, 0.5 s
Units	mV/m, V/m, mA/m, mA/m, mW/m^2, mW/m^2, W/m^2, mW/cm^2, mW/cm^2
Dry batteries	9 V NEDA 1604/1604A
Battery life	15 h
Operating temperature range	0 °C to + 50 °C
Operating humidity range	25% to 75%RH
Dimensions	60 × 60 × 237 mm (approx.)
Weight	200gms including battery (approx.)
Sensor type	Electrical field (E)
Frequency range	50 MHz to 3.5 GHz
Dynamic range	Typically, 75 dB
Absolute error at 1 V/m and 50 MHz	±1.0 dB
Thermal response (0 to 50 °C)	±0.2 dB

6 Findings and Discussion

The variation of power density of electromagnetic field from the base of the mobile tower is shown in Fig. 5. Whereas the occurrence of birds at different sites within our study area is presented in Table 3. The findings focus on the behavioral changes exhibited by

birds, variations in species diversity and abundance, the potential impacts of EMR on avian health, and the efficiency of drones as a reliable device for long-term ecological surveys. These insights are crucial in assessing the ecological impact of mobile base stations and formulating sustainable strategies for wildlife conservation.

6.1 Observed Avian Behaviour Near Base Stations

Bird behavior is one of the biomarkers of ecological changes. Throughout the study, various shifts in flight patterns, nesting preferences, feeding habits, and social interactions were recorded in proximity to mobile base stations.

6.2 Observed Avian Behaviour Near Base Stations Avoidance of Base Stations.

- Birds were observed altering their flight trajectories to avoid flying directly over or near mobile base stations.
- Sudden changes in direction, hesitation, and erratic movements were common among small passerines when approaching radiation-exposed areas.

Roosting Preferences and Displacement.

- Species that traditionally nested on trees near base stations were found to shift their roosting locations further away, possibly in response to electromagnetic exposure (Fig. 6).
- Larger birds such as pigeons and crows showed less avoidance behaviour, continuing to roost on structures near mobile towers, suggesting species-specific tolerance levels.

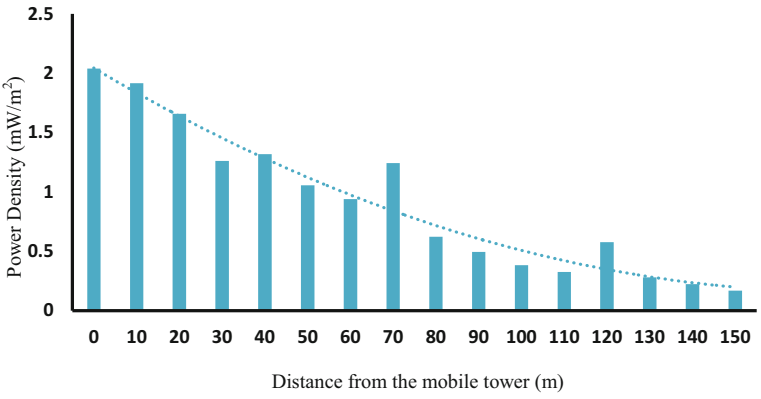


Fig. 5. Variation of power density with distance from the mobile tower (m)

Table 3. Report from the drone-assisted observation

Site	Time-Based Data Collection	Flight Duration	Altitude Considerations	Occurrence of Bird
Site-1	5.30am–08.00am	20–30 min	30–50 m	Crow, Common Myna, Sparrow
			80–100 m	Common Myna, Crow, Jungle Babbler, Pied Myna, Dove, Common Trailor bird, Pigeon, Bulbul
			Above 100 m	Crow, Hawk, Rose ringed parakeet
Site-2	12.00pm– 02.00pm	20–30 min	30–50 m	Crow, Common Myna, Dove, Pied Myna, Sparrow
			80–100 m	Common Myna, Crow, Jungle Babbler, Pied Myna, Dove
			Above 100 m	Crow, Hawk, Crane
Site-3	05.00pm– 07.00pm	20–30 min	30–50 m	Crow, Common Myna, Dove, Pied Myna, Wood Pecker
			80–100 m	Common Myna, Crow, Pied Myna, Bulbul
			Above 100 m	Common Myna, Crow, Crane

6.3 Effects on Feeding and Social Interactions

Foraging Behaviour Alterations.

- Birds spent significantly less time foraging in areas with high electromagnetic radiation compared to low-exposure regions.
- Insectivorous birds appeared more affected, as their reduced feeding duration might indicate disruptions in prey availability or altered sensory perception due to electromagnetic interference.

Social Behaviour Disruptions.

- Flocking patterns among social species, such as starlings and mynas, appeared disrupted, with smaller and more dispersed groups forming near base stations.

- Some species displayed reduced vocalization frequency, potentially affecting communication and mate attraction.

Fig. 6. Photograph of avian species in the sites

7 Conclusions

The findings of this study underscore the pressing need for a balanced approach to technological development and ecological conservation. While mobile base stations have become an integral part of modern communication networks, their potential adverse effects on avian populations cannot be overlooked. The observed behavioral shifts, declining species diversity, and possible health implications call for immediate action from policymakers, environmental agencies, and research institutions.

Drone technology has proven to be a valuable tool in advancing ecological research, offering a non-invasive and efficient means of monitoring avian populations. However, further advancements are needed to enhance data accuracy, extend flight duration, and integrate complementary monitoring techniques.

Acknowledgments. We appreciate all of the assistance that JIS University has given us. We also thank the residents around the dumping area for their participation in the survey and the Arambagh Municipality for their assistance as and when needed.

References

1. Gregory, R.D., Van Strien, A.: Wild bird indicators: using composite population trends of birds as measures of environmental health. Ornithol. Sci. **9**(1), 3–22 (2010)
2. Bibby, C., Burgess, N.D., Hill, D., Mustoe, S.: Bird Census Techniques, 2nd edn. Academic Press, London, England (2000)
3. Bose, S., et al.: Impressions of high frequency radio-waves from cell phone towers on birds: a base-line study. J. Multidiscip. Res **1**, 54–62 (2020)
4. Şekercioğlu, Ç.H.: Increasing awareness of avian ecological function. Trends Ecol. Evol. **21**(8), 464–471 (2006)
5. Mitra, R., Pattanayak, S.: Mobile phone and tower radiation: a challenge to all living entities. Explor Anim Med Res **8**(1), 5–10 (2018)
6. Balmori, A.: Possible effects of electromagnetic fields from phone masts on a population of white stork (Ciconia ciconia). Electromagn. Biol. Med. **24**(2), 109–119 (2005)

7. Engels, S., et al.: Anthropogenic electromagnetic noise disrupts magnetic compass orientation in a migratory bird. Nature **509**(7500), 353–356 (2014)
8. Bigu-del-Blanco, J., Romero-Sierra, C.: The properties of bird feathers as converse piezoelectric transducers and as receptors of microwave radiation. II. Bird feathers as dielectric receptors of microwave radiation. Biotelemetry **2**(6), 354–64 (1975). PMID: 1242004
9. Balmori, A., Hallberg, Ö.: The urban decline of the house sparrow (Passer domesticus): a possible link with electromagnetic radiation. Electromagn. Biol. Med. **26**(2), 141–151 (2007)
10. Cucurachi, S., Tamis, W.L., Vijver, M.G., Peijnenburg, W.J., Bolte, J.F., de Snoo, G.R.: A review of the ecological effects of radiofrequency electromagnetic fields (RF-EMF). Environ. Int. **51**, 116–140 (2013)
11. Fernie, K.J., Reynolds, S.J.: The effects of electromagnetic fields from power lines on avian reproductive biology and physiology: a review. J. Toxicol. Environ. Health Part B **8**(2), 127–140 (2005)
12. Chabot, D., Bird, D.M.: Evaluation of an off-the-shelf unmanned aircraft system for surveying flocks of geese. Waterbirds **35**(1), 170–174 (2012)
13. Hodgson, J.C., Koh, L.P.: Best practice for minimising unmanned aerial vehicle disturbance to wildlife in biological field research. Curr. Biol. **26**(10), R404–R405 (2016)
14. Christie, K.S., Gilbert, S.L., Brown, C.L., Hatfield, M., Hanson, L.: Unmanned aircraft systems in wildlife research: current and future applications of a transformative technology. Front. Ecol. Environ. **14**(5), 241–251 (2016)

Digital Twins in Agriculture: Revolutionizing Climate Resilience with AI and IoT

Swati Suman[1], Sumit Ray[1], Ajay Kumar Prusty[1], Umesha C[2], Girish Prasad Rath[1(✉)], Sabyasachi Patnaik[1], Ankita Priyadarshini[3], Swagat Shubhadarshi[4], Pavan Kumar Pandey[2], and Lalithamma M[2]

[1] Centurion University of Technology and Management, Paralakhemundi, Odisha 761211, India
gprath@cutm.ac.in
[2] Sam Higginbottom University of Agriculture Technology and Sciences, Prayagraj, UP 211007, India
[3] Siksha 'O' Anusandhan University, Bhubaneshwar, Odisha 751030, India
[4] MITs Institute of Professional Studies, Rayagada, Odisha 765017, India

Abstract. Climate change has a significant influence on agriculture, affecting developing nations' food security and financial condition. Thus, the use of Digital Twins, Internet of Things (IoT) devices, and Artificial Intelligence (AI) may play an important role in transforming agriculture that is data-enabled in real time for crop development, high productivity, or climate mitigation. These technologies would aid in predicting drought start periods, optimizing irrigation scheduling to react to any specific climatic shift, and driving crop rotations in a given area. To power climate-resilient farming development, AI and IoT must be combined, resulting in DTs. This technology incorporates agricultural offices, animal monitoring, crop harvests, crop protection, and a DT for predictive maintenance purposes. AI is transforming agriculture by analyzing large volumes of data to forecast climate change consequences. Precision agriculture, a key AI tool, uses micro-localized applications based on syntactic sensory data, drones, and satellite data. Smart agriculture uses IoT, AI, Big Data analytics, and DTs to gather, integrate, and analyze data from various sources. AI-powered models can forecast future weather patterns, insect infestations, and disease outbreaks, enabling earlier intervention and higher output. These insights enable improved resource allocation, agricultural practice optimization, and enhanced farm output in the face of climate change and hence making the DT the possible game changer in the field of agriculture while keeping sustainability as one of its important cornerstones.

Keywords: Internet of Things · Artificial Intelligence · Climate Change · Smart Agriculture · Digital Twins · Climate Resilience

1 Introduction

Agriculture is the foundation of human existence, producing essential foods, textiles, and other necessities [1]. However, climate change has significantly disrupted agricultural productivity, threatening the world's food supply and financial security in underdeveloped countries. Traditional farming methods are inadequate in handling the complexities and uncertainties brought by climate change, highlighting the need for modern

S. Fong et al. (Eds.): ICT4SD 2025, LNNS 1651, pp. 444–454, 2026.
https://doi.org/10.1007/978-3-032-06688-6_43

technology-based solutions [2]. Artificial intelligence (AI) and Internet of Things (IoT), often known as Digital Twins (DT), are two new technologies that could completely transform agricultural operations. DT technology allows for virtual-real duplicates of physical assets, processes, or systems that behave in real-time dynamic form. These DTs, developed using real-time data from IoT devices and advanced analytics from AI, can offer useful information to maximize crop production resources, increase productivity, and reduce climate change-related hazards [3]. DTs are essential for multi-scale issues brought forth by climate change. Farmers can examine management options by replicating virtual settings with their technical scenarios and environmental variables. For example, when past weather patterns are combined with current data, DTs can forecast drought events, optimize irrigation plans, or force crop rotations more suited to specific climates [4]. This precision-driven model results in increased productivity, resource conservation, and less environmental impact, making agriculture a more sustainable and resilient sector. Climate-resilient agriculture relies on the integration of AI and the IoT to create DTs. In IoT, a continuous data stream is used to monitor important factors like as temperature, soil moisture, and plant health, while AI detects trends, predicts outcomes, and produces suggestions. Farmers can test and implement simulated solutions in a risk-free environment, increasing precision, sustainability, and adaptation in a holistic management approach to agriculture, as well as providing limitless possibilities for supply chain optimization, animal monitoring, and predictive farm equipment maintenance [5]. DTs provide an accounting and historical picture of operational activities, serving as a revolutionary foundation for transforming agriculture by combining contemporary technology with ancient expertise. This article highlights the revolutionary potential of DTs, when combined with AI and the IoT, to improve agricultural and climate resilience. The study presents a comprehensive description of DTs, including their concept, technology, and application to climate-related challenges. This enables the authors to emphasize the critical role of DTs as the cornerstone for climate-resilient and sustainable agriculture.

2 The Concept of DTs

DTs are virtual clones of physical systems that can mimic, monitor, and optimize their counterparts in real time. They employ data from machine learning algorithms, ambient data, and IoT sensors to create a dynamic representation of the real thing. Farmers and agricultural administrators can use DT to create comprehensive virtual representations of fields, crops, irrigation systems, and entire farming enterprises [6]. These DTs can track various characteristics in real-time, such as crop growth, soil moisture, temperature, and environmental conditions. DTs are advanced models that can predict outcomes and model various situations, helping farmers make better decisions about resource use, operational efficiency, and climate change. These models are three-dimensional concepts that include real-time feedback, data analytics, and data acquisition. IoT devices like soil moisture, temperature, and crop health sensors are used to collect data on physical environment components, enabling intelligent inference and updating the virtual model in real time [7]. This information is then used to evaluate various approaches without endangering plants or input resources. DTs improve work performance in agricultural operations management by integrating real and virtual worlds, enhancing decision-making

skills, increasing output, and encouraging environmentally friendly farming practices. They offer a flexible and adaptive way to help agrarians adapt to unpredictable shifts caused by climate change, which produces inconsistent weather patterns and unfavorable environmental conditions.

3 The Role of AI in Climate Resilience

AI is revolutionizing agriculture by replacing human labor in predicting, preventing, and adapting to climate change. AI-enabled technologies like machine learning, deep learning, and predictive analytics analyze historical and real-time data to forecast climatic events, enabling better resource utilization, optimal agricultural practices, and increased farm output [8]. Predictive modeling has led to the creation of agricultural decision AI systems, based on weather patterns, soil conditions, and crop yield production. AI also helps farmers monitor pest and disease outbreaks, minimizing crop loss and decreasing pesticide use. AI is also becoming an essential tool in precision agriculture, reducing input costs through micro-localized applications based on syntactic sensory data, drone, and satellite data [9]. Surface input analysis tracks crop development variations and creates warnings to reduce adverse impacts. AI contributes significantly to climate resilience by modifying tactics to make agricultural systems more resilient to natural disasters.

4 The Role of IoT in Climate Resilience

IoT is a network of devices and sensors that collect, distribute, and analyze environmental data in agriculture. These sensors monitor critical parameters like crop health, temperature, humidity, soil moisture, and weather conditions, providing farmers with real-time information to make informed decisions. IoT-based systems can tailor responses to extreme weather phenomena, such as floods, droughts, and excessive temperatures [10]. Soil moisture sensors help farmers adapt irrigation systems to reduce waste and water shortages, while temperature and humidity sensors measure microclimates for precise crop management. This data is fed into a digital platform for analytics and action recommendations, enabling farmers to better understand their operations and react to changing climate conditions [11]. IoT devices also benefit precision agriculture, allowing for more localized decision-making and increased efficiency in waste reduction and resource utilization. IoT implementation is crucial for enhancing early warning systems for climate change's effects on agriculture and assisting in adaptation and mitigating climate change consequences.

5 Integration of DTs for Climate Resilience

DT optimize modelling, real-time data collecting, and predictive analytics to increase agriculture's climate resilience. Digitally twin the agricultural systems that generate dynamic, data-driven tenders for land, crops, irrigation systems, and farming equipment. Such digital replicas, which incorporate data streams from IoT sensors as well as exact weather forecasts, result in optimal climate-farm interaction decisions for modelling

agricultural systems [12]. It aids agriculture by enhancing production, reducing climatic risk, and promoting ecological sustainability through modelling. Farmers may use DTs in conjunction with climate adaptation to build simulations that show how farm systems will react to various situations. For example, a DT can use a weather prediction input along with soil moisture and crop growth inputs to show how higher temperature, drought, or excess rainfall affect agricultural outcomes [13]. Simulated trials may be used to enhance crop management, irrigation, and pest control strategies, reducing the impact of climate change on yield losses, resource inefficiencies, and environmental degradation. DTs optimize modelling, real-time data collecting, and predictive analytics to increase agriculture's climate resilience. Such digitally linked agricultural systems, resulting in dynamic, data-dependent bids for land, crops, irrigation systems, and farming equipment. The incorporation of IoT sensor inputs, precise weather predictions, and, most likely, AI models for DTs allows for optimal climate-farm interaction decisions [14]. As a result, it increases agricultural output while mitigating the risks connected with climate change and ensuring the implementation of environmentally friendly practices. A DT can integrate weather prediction with soil moisture and crop growth modeling to predict agricultural outcomes.

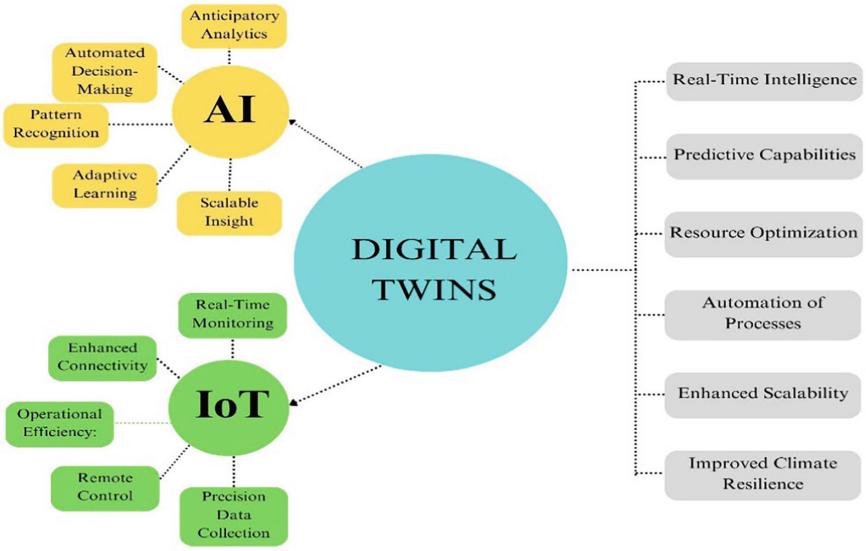


Fig. 1. Integration of Digital Twins

6 Integrated Data and Decision-Making for Smart Farming

Smart agriculture, also known as data-driven agriculture, uses IoT, AI, Big Data analytics, and DTs (Fig. 1) to gather, integrate, and analyse data from multiple sources, allowing farmers to turn these insights into educated choices. A smart farm often collects data from several sensors, weather stations, and satellite images, including weather predictions, crop health, irrigation status, temperature, humidity, and soil moisture. These

provide a comprehensive perspective of the agricultural environment, allowing farmers to understand their farm needs in real time [15]. This technology allows farmers to be well-informed and make decisions based on various insights, providing them the ability to address climate change and enhance agricultural operations. AI and machine learning algorithms analyze both historical and current data to identify patterns, correlations, and anomalies. Anomalies can anticipate crop diseases, irrigation requirements, and meteorological conditions [16]. IoT sensors in fields are used to track soil fields, allowing for more efficient and timely watering. Pests might be tracked via picture aggregation and sensor data, allowing for early management against flush infestation signs. Smart agricultural ecosystems might monitor resource use, reduce waste, and increase efficiency. The integration of data and decision-making is critical for sustainable agricultural operations. Farmers, for example, might map fertilizer application to prevent overstretching their usage, lowering costs and reducing environmental consequences caused by runoff. Coordinating soil health data with weather forecasts and water conservation techniques might help farmers conserve water per irrigation and become drought-resistant. This will help the farmer manage climate fluctuation. DTs provide an integrated decision-making platform for on-farm modelling and prediction of various techniques before to their implementation under real-world conditions [17]. Such modelling skills also enable the simulation of multiple scenarios to assess the outcome under different conditions, such as weather changes or crop rotation techniques. In this way, the farmer may gather information for decision-making that assures maximum yields in climate-resilient and sustainable systems. Smart farming, which combines data and decision-making, is preserving agriculture's efficiency and resilience in the face of climate change.

Bridging Physical and Digital Realms in Agriculture

Smart farming design includes both digital and physical components to ensure long-term agricultural operations. This architecture combines real assets like fields, crops, machinery, and animals with the IoT, DTs, and AI to create an agricultural system. Farmers will be able to better understand their operations and make more informed decisions, increasing productivity, resilience, and sustainability [18]. IoT, sensors, and networked gear are deployed over this huge farming landscape to capture a plethora of real-time data on crop health, temperature, humidity, soil moisture, and equipment performance. However, the volume and quantity of data generated creates a barrier to transmission and processing, making digital a crucial component for understanding information. By supporting the link between the physical and digital worlds, farmers can make the most informed decisions to boost production, resilience, and, most importantly, sustainability [19]. DT technologies provide virtual pictures of agricultural systems, bringing the physical and computer worlds closer together. These copies are updated in real time by IoT sensors and other data sources, thus the farm may be considered a digital replica. Farmers may see and simulate situations and results without physically implementing them in the field. This improves risk management and decision-making in uncertain situations by imitating real-world settings in a digital environment [20]. Automation is progressively tying agricultural operations to both the real and virtual worlds, evoking physical actions in real time using digital models that include IoT-enabled devices and AI. The DT can automate the irrigation system to guarantee that plants are never water stressed, and the AI system can direct tractors or drones to do

precise planting, spraying, or harvesting. This total autonomy is projected to minimize labor-related expenses, improve task precision, and eliminate resource waste. It brings value in terms of functioning while also providing substantial reasons for oversight and management. Farmers can use digital technologies to remotely monitor fields, crops, or equipment, allowing them to respond rapidly to urgent situations. As a result, the rate at which events are responded to and the efficiency with which farms operate may be better regulated and managed. The convergence of the digital and physical worlds creates new potential for resource management and sustainability on a greater scale. Continuous observation and integrating the physical environment to a digital model will decrease waste on farms, while digital planes that calculate and simulate real-time data will allow for far more precise resource allocation, improving productivity while minimizing environmental effect.

7 Challenges and Solutions

DTs have the potential to revolutionize agriculture, but they face numerous operational, financial, and technical challenges. The capital-intensive nature of digitalization is a significant obstacle for small and medium farms, who lack the financial flow to support such expenditures. High upfront expenses are also major obstacles for many farmers, despite the long-term benefits of improved productivity, resource optimization, and higher yields. The integration of multiple data sources, such as soil sensors, weather stations, drones, satellites, and agricultural equipment, is another significant obstacle [21]. The quality and accuracy of data gathered differ from one source to another, making it technically difficult to combine data from many sources. Inaccurate simulations and forecasts can be caused by irregular data gathering methods or malfunctioning sensors, lowering the dependability of DT models [22]. Data security and privacy procedures are also important limitations.

8 The Road Ahead: Innovations, Success Stories and Future Scope

DTs in agriculture are transforming the agricultural business by combining AI, IoT, data analytics, and machine learning to create robust systems. These shifts create several options for climate resilience, sustainability, and productivity improvement. Predictive analytics, real-time data, and digital models are transforming farm management, resource utilization, and farmers' attitudes towards environmental issues [23]. In the near future, the employment of DTs, machine learning algorithms, and enhanced AI will be the most significant technical development. DTs are increasingly collaborating with self-driving systems like drones, tractors, and smart harvesting equipment to give real-time farm management and monitoring which are constantly updated and modified to reflect the farm's state, with data sent to the DT, where AI algorithms analyze it and suggest appropriate courses of action [24]. In precision agriculture, North American farms are using IoT sensors to optimize water usage, lowering water consumption and increasing crop yields. One documented instance shows that farms created a DT model of the irrigation system by integrating real-time weather, soil moisture, and crop performance

data, reducing water use by up to 30% without compromising crop health [25]. A European agricultural cooperative implemented this technology to anticipate stress points, identify disease outbreaks early, and implement intervention measures. This proactive approach improved yield and soil health while reducing fertilizer and pesticide requirements. Farmers can also enhance their flood management techniques, making decisions about seed varieties, irrigation schedules, and fertilizer application based on the data gathered from the DT like in Asia, rice farmers in flood-prone areas has adopted this technology [26]. Blockchain technology enhances security, transparency, and traceability in the agricultural supply chain, which can also help farmers to monitor their crops' entire lifecycle, from seed to market, ensuring all data is readily available and safe [27]. "Grow It York" is another example of Indian vertical farming where the use of Q-learning digital twin led to increase demand fulfilment and reduce electricity cost by 78.5% and 15%, respectively. It also revealed that, despite of higher initial automation costs, the break-even was achieved in 3 years [28]. Farmers struggle with knowledge and skill gaps, financial constraints, and technical difficulties. Education and training are needed to close this gap, and advancements in decision-support systems and user-friendly options will increase the viability of DT technology. These technologies will become commonplace in agricultural systems worldwide, helping farmers cope with the demands of a growing global population, resource optimization, and climate change. The future of agriculture will be more digital, data-driven and interconnected, providing farmers with new opportunities for success in an era of climate-induced uncertainty and rising global food demand (Table 1).

Table 1. Digital Twins and Its application.

Category	Tool	Description
Weed management	Computer vision assisted system	The mechanically weeding manner consists of an intertwined servo motor coupled with the computer vision backed system to descry factory spots and direct the weeding manner to perform mechanical weeding operations without harming crops [29].
Crop and soil health monitoring	PLANTIX	A machine knowledge predicated tool to control and manage the husbandry process, complaint control and the civilization of high-quality crops [30].

(*continued*)

Table 1. (*continued*)

Category	Tool	Description
Climate conditions management	PYCNO	A software and sensor allowing continuous data collection and flux from the estate to smart phone. It also contains a dashboard to apply the bottommost phenological and complaint models to cover trends and assess trouble to agricultural products [31].
End to end farm management systems	CROPIO	A decision-making tool used to optimize fertilization and irrigation to control the amount of toxins and reduce the use of water. It combines downfall information and satellite data to cover crops and field forecasts [32].
Spraying fertilizers and pesticides	Accelerometer and Gyroscope Sensors, Arduino	It has the ability to reduce time and human effort [33].
Soil moisture detection	Moisture Analyzers and METTLER TOLEDO	Precise temperature control with halogen heating technology and outstanding weighing technology [34].
Disease management	Expert system using rule-base in disease detection	Resolves plant diseases quickly and cost-effective approach [35].
Soil management	DSS	Reduces erosion and sedimentary yield [36].

9 Conclusion

DTs can give real-time information on crops, soils, and equipment, allowing farmers to make better decisions and decreasing the risk of climate change. However, several obstacles, such as high prices, a lack of data integration, security concerns, and a lack of competence, must be addressed. Collaboration between agri-sector, policy, and technological stakeholders is crucial to overcome these obstacles. DTs have immense potential in the future, with new technologies like connected climate forecasting, autonomous systems, and machine learning models increasing their capabilities. AI-based prediction and autonomous interventions, along with the ability to model and simulate various climatic scenarios, hold significant potential for farm productivity and resilience in a rapidly

changing environment. By bridging the digital and physical worlds, DTs can improve agriculture's efficiency, sustainability, and resilience, preparing the sector to meet the demands of a growing population in a fast-changing environment. Despite challenges, with ongoing innovation, it is possible to envision a future where nature and technology coexist to create a more environmentally friendly and sustainable food system.

References

1. Ray, S., et al.: The nexus between intercropping systems, ecosystem services and sustainable agriculture: a review. Res. Crops **26**(1) (2025). https://doi.org/10.31830/2348-7542.2025.ROC-1166
2. Altieri, M.A., Nicholls, C.I., Henao, A., Lana, M.A.: Agroecology and the design of climate change-resilient farming systems. Agron. Sustain. Dev. **35**(3), 869–890 (2015)
3. Zhang, Z., Wen, F., Sun, Z., Guo, X., He, T., Lee, C.: AI-enabled sensing technologies in the 5G/IoT era: from virtual reality/augmented reality to the DT. Adv. Intell. Syst. **4**(7), 2100228 (2022)
4. Moazami, A., Nik, V.M., Carlucci, S., Geving, S.: Impacts of future weather data typology on building energy performance–Investigating long-term patterns of climate change and extreme weather conditions. Appl. Energy **238**, 696–720 (2019)
5. Sharma, R., Kamble, S.S., Gunasekaran, A., Kumar, V., Kumar, A.: A systematic literature review on machine learning applications for sustainable agriculture supply chain performance. Comput. Oper. Res. **119**, 104926 (2020)
6. Pylianidis, C., Osinga, S., Athanasiadis, I.N.: Introducing DTs to agriculture. Comput. Electron. Agric. **184**, 105942 (2021)
7. Nayyar, A., Puri, V.: Smart farming: IoT based smart sensors agriculture stick for live temperature and moisture monitoring using Arduino, cloud computing & solar technology. In: Proc. of The International Conference on Communication and Computing Systems (ICCCS-2016), pp. 9781315364094–121 (2016)
8. Kulikova, E., Molokova, E.: Leveraging market insights for sustainable agricultural practices. BIO Web Conf. **121**, 02011. EDP Sciences (2024)
9. Lakhiar, I.A., et al.: A review of precision irrigation water-saving technology under changing climate for enhancing water use efficiency, crop yield, and environmental footprints. Agriculture **14**(7), 1141 (2024)
10. Usigbe, M.J., Asem-Hiablie, S., Uyeh, D.D., Iyiola, O., Park, T., Mallipeddi, R.: Enhancing resilience in agricultural production systems with AI-based technologies. Environ. Dev. Sustain. **26**(9), 21955–21983 (2024)
11. Kasulla, S., Malik, S.J., Baxla, S.P., Zafar, S.: The role of IoT in waste management and sustainability. Partners Univer. Int. Res. J. **3**(2), 76–88 (2024)
12. Selvam, A.P., Al-Humairi, S.N.S.: The impact of IoT and sensor integration on real-time weather monitoring systems: a systematic review. Springer Science and Business Media LLC: Berlin, Germany (2023)
13. Skobelev, P.O., et al.: Development of models and methods for creating a DT of plant within the cyber-physical system for precision farming management. J. Phys. Conf. Ser. **1703**(1), 012022. IOP Publishing (2020)
14. Syed, T.A., Khan, M.Y., Jan, S., Albouq, S., Alqahtany, S. S., Naqash, M. T.: Integrating DTs and AI Multi-modal transformers into water resource management: overview and advanced predictive framework. AI **5**(4), 1977–2017 (2024)
15. Wolfert, S., Ge, L., Verdouw, C., Bogaardt, M.J.: Big data in smart farming–a review. Agric. Syst. **153**, 69–80 (2017)

16. Lecerf, R., Ceglar, A., López-Lozano, R., Van Der Velde, M., Baruth, B.: Assessing the information in crop model and meteorological indicators to forecast crop yield over Europe. Agric. Syst. **168**, 191–202 (2019)
17. Verdouw, C., Tekinerdogan, B., Beulens, A., Wolfert, S.: DTs in smart farming. Agric. Syst. **189**, 103046 (2021)
18. Darnhofer, I., Fairweather, J., Moller, H.: Assessing a farm's sustainability: insights from resilience thinking. Int. J. Agric. Sustain. **8**(3), 186–198 (2010)
19. Friess, P.: Digitising the Industry – IoT Connecting the Physical. Digital and Virtual Worlds. River Publishers, Location (2016)
20. Galera-Zarco, C., Floros, G.: A deep learning approach to improve built asset operations and disaster management in critical events: an integrative simulation model for quicker decision making. Ann. Oper. Res. **339**(1), 573–612 (2024)
21. Cesco, S., Sambo, P., Borin, M., Basso, B., Orzes, G., Mazzetto, F.: Smart agriculture and DTs: applications and challenges in a vision of sustainability. Eur. J. Agron. **146**, 126809 (2023)
22. Barrile, V., Simonetti, S., Citroni, R., Fotia, A., Bilotta, G.: Experimenting agriculture 4.0 with sensors: A data fusion approach between remote sensing, UAVs and self-driving tractors. Sensors **22**(20), 7910 (2022)
23. Fuentes-Peñailillo, F., Gutter, K., Vega, R., Silva, G.C.: Transformative technologies in digital agriculture: leveraging IoT, remote sensing, and AI for smart crop management. J. Sens. Actuator Netw. **13**(4), 39 (2024)
24. Tagarakis, A.C., Benos, L., Kyriakarakos, G., Pearson, S., Sørensen, C.G., Bochtis, D.: DTs in agriculture and forestry: a review. Sensors **24**(10), 3117 (2024)
25. Sanjeevi, P., Prasanna, S., Siva Kumar, B., Gunasekaran, G., Alagiri, I., Vijay Anand, R.: Precision agriculture and farming using IoT based on wireless sensor network. Trans. Emerging Telecommun. Technol. **31**(12), e3978 (2020)
26. Koppa, N., Amarnath, G.: Geospatial assessment of flood-tolerant rice varieties to guide climate adaptation strategies in India. Climate **9**(10), 151 (2021)
27. Jakku, E., et al.: "If they don't tell us what they do with it, why would we trust them?" Trust, transparency and benefit-sharing in smart farming. NJAS-Wageningen J. Life Sci. **90**, 100285 (2019)
28. Luo, Y., Ball, P.: Adaptive production strategy in vertical farm digital twins with Q-learning algorithms. Sci. Rep. **15**, 15129 (2025). https://doi.org/10.1038/s41598-025-97123-y
29. Talaviya, T., Shah, D., Patel, N., Yagnik, H., Shah, M.: Implementation of AI in agriculture for optimisation of irrigation and application of pesticides and herbicides. AI Agric. **4**, 58–73 (2020). https://doi.org/10.1016/j.aiia.2020.04.002
30. Ayaz, M., Ammad-Uddin, M., Sharif, Z., Mansour, A., Aggoune, E.M.: Internet-of-Things (IoT)-based smart agriculture: toward making the fields talk. IEEE Access **7**, 129551–129583 (2019). https://doi.org/10.1109/access.2019.2932609
31. Belcaro, G., et al.: Management of varicose veins and chronic venous insufficiency in a comparative registry with nine venoactive products in comparison with stockings. Int. J. Angiol. **26**(03), 170–178 (2017)
32. Gad, M., Saleh, A.H., Hussein, H., Elsayed, S., Farouk, M.: Water quality evaluation and prediction using irrigation indices, artificial neural networks, and partial least square regression models for the Nile River. Egypt. Water **15**(12), 2244 (2023)
33. Miller, A.J.: The regulation of melanoma antigens by the microphthalmia transcription factor. Harvard University (2004)
34. Garre, P., Harish, A.: Autonomous agricultural pesticide spraying UAV. IOP Conf. Ser. Mater. Sci. Eng. **455**, 012030 (2018). https://doi.org/10.1088/1757-899X/455/1/012030

35. 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 (2017)
36. Kashyap, B., Kumar, R.: Sensing methodologies in agriculture for soil moisture and nutrient monitoring. IEEE Access **9**, 14095–14121 (2021)

Automatic Road Maintenance Robot

Kalyani Kulkarni(✉), Dipti Varpe, Gargi Kathale, Shreya Patil, and Drishti Dhamale

Institute of Management, Pune Vidyarthi Griha's College of Engineering and Technology & G.K. Pate (Wani), Pune, India
{kjk_entc,dtv_it}@pvgcoet.ac.in

Abstract. This paper aims to presents the design and development of a robotic system able to autonomously detect and fill potholes in roads, a significant challenge in civic structure conservation. The proposed system utilizes a combination of detectors and selectors to achieve this task effectively. An ultrasonic detector is employed to detect the presence of potholes by measuring the distance to the road face. When a pothole is detected, a servo motor activates a filling medium to introduce a suitable material to repair the disfigurement. To ensure safe navigation, an infrared (IR) detector is used to descry obstacles in the robot's path, allowing it to avoid collisions and implicit damage. Global Positioning System (GPS) technology is integrated into the system to track the robot's position in real time, enabling remote monitoring and control. This data is transmitted to a pall-grounded platform, similar to Blynk, where it can be penetrated and imaged through a mobile operation. This allows for effective operation of the robot's operations and ensures that potholes are repaired instantly, perfecting road safety and reducing conservation costs. A motordriver and DC motors are responsible for the robot's movement, allowing it to navigate to pothole locales and carry out the form process. The overall system armature is designed to be effective, dependable, and able to operate autonomously without mortal intervention. By automating the process of pothole discovery and form, the proposed system can significantly ameliorate the condition of roads and enhance the overall quality of life for citizens.

Keywords: Pothole detection · Robotics · Ultrasonic Sensor · GPS · IR sensor · Road maintenance

1 Introduction

The Indian Road network is ranked at the second-largest position globally, which spans across approximately 5.9 million Km (kilometers). Around 90% of population transportation and 64.5% of domestic goods movement rely on this vast network [5]. Roads play a critical role in a nation's economic development and connectivity. However, inadequate road maintenance poses significant challenges, particularly in developing countries, leading to frequent accidents and fatalities. Poor road conditions can directly contribute to severe injuries and loss of life [4]. Despite government efforts and public

S. Fong et al. (Eds.): ICT4SD 2025, LNNS 1651, pp. 455–463, 2026.
https://doi.org/10.1007/978-3-032-06688-6_44

initiatives, road accidents remain a leading cause of deaths, disabilities, and hospitalizations in India and among the 199 countries in the world India stands at the first position in road fatalities index globally [5].

This study proposes the development of an autonomous robotic system capable of detecting and filling potholes efficiently. The robot incorporates ultrasonic sensors, servo motors, infrared (IR) sensors, GPS, and Blynk cloud integration to enhance road safety and maintenance efficiency. By employing real-time sensor data, the robot can accurately detect potholes, record their location and severity, and carry out the repair process autonomously. The implementation of this system is expected to reduce maintenance costs, prevent minor damages from escalating, and significantly improve road quality. Ultimately, this innovation aligns with the vision of a smart city infrastructure, where automation and technology are seamlessly integrated to enhance urban living standards.

2 Literature Survey

Iason Katsamenis et al. (2022) propose the development of an integrated autonomous system for the maintenance and enhancement of road infrastructure, encompassing roadways and the overall transport network. This system features a self-operating ground robotic vehicle supported by autonomous drones to assist in maintenance tasks, as well as pre- and post-intervention phases. It incorporates various robotic equipment, a monitoring interface designed to assess structural integrity, functionality, and road markings, and control software that connects the monitoring interface with the robotic actuators. The system comprises an autonomous ground robotic vehicle, various robotic tools, a real-time monitoring interface, control software, AR-based visualization tools, and communication modules. Its primary objective is to enhance maintenance workflows, intelligently analyze data collected from road and vehicle sensors, integrate this information into an advanced visualization interface, and improve road user safety through predictive maintenance strategies. Ultimately, the HERON project seeks to refine inspection, evaluation, maintenance, and the overall safety of road infrastructure [1].

Skibniewski, Miroslaw & Hendrickson, Chris. (1990) explore the potential of automating road construction and maintenance equipment as an alternative to conventional manual labor in the future. The study differentiates between numerically controlled (NC) machinery and autonomous equipment, examines key technologies required for automation, and provides examples of existing NC-based equipment. Additionally, it addresses economic feasibility concerns and workplace safety implications associated with automated road construction and maintenance systems. The research concludes that the repetitive nature and moderate sensory demands of road construction and maintenance make these tasks highly suitable for automation. In some cases, workers could be entirely removed from high-risk areas, thereby preventing accidents caused by collisions with heavy machinery or passing vehicles. However, despite these potential benefits, there remains a noticeable gap in research and development related to partially or fully autonomous road construction and maintenance equipment [2].

Bavelos et al. (2024) Bavelos et al. (2024) highlight the risks associated with road maintenance work and the necessity of a support system to improve safety and efficiency [6]. Augmented Reality (AR) has the potential to provide valuable real-time data

in dynamic and unstructured environments. This research introduces an innovative AR-based framework for human-robot collaboration, offering real-time guidance and support in road maintenance tasks. The framework is built on a Robot Operating System (ROS) architecture to facilitate seamless communication. Initial tests conducted in a controlled laboratory setting demonstrated improvements in worker performance, with future validation planned through real-world interventions. Researchers have explored different traffic management techniques, utilizing models such as object-based, case-driven logic, and mesoscopic simulation approaches. Speeding vehicles on highways contribute significantly to accidents involving AR-assisted operations. Implementing AR technology can enhance worker efficiency while minimizing accident risks. Additionally, this technology can be adapted for road maintenance to optimize overall safety and productivity. To further improve road maintenance operations, obstacle detection is integrated into the robotic system using infrared (IR) sensors. If an obstruction is identified, the robot autonomously changes lanes to avoid collisions. A microcontroller serves as the system's core, processing inputs to detect obstacles and guiding the robot's movements accordingly. The design incorporates four IR sensors—one at the front and three positioned on the left, right, and rear—to detect barriers. A motor driver is utilized to maneuver the robot and prevent collisions. Several companies have invested in the development of robotic vacuum cleaners, including iRobot, Neato Robotics, and LG Corporation. iRobot, founded by three robotics specialists from MIT's Artificial Intelligence Laboratory, was established with the vision of integrating robots into everyday life and making them widely accessible. Neato Robotics, based in Newark, California, focuses exclusively on robotic vacuum technology, launching the Neato XV in 2010. LG Corporation, formerly known as Lucky Goldstar, has evolved into a global technology giant, producing a variety of smart appliances, including robotic vacuums. These companies aim to develop fully autonomous cleaning systems that function without human intervention, significantly enhancing the efficiency and safety of road maintenance operations [3].

This project combines both detection and filling in a single, compact robot, enabling autonomous end to end operation. It uses ESP32, ultrasonic sensors and basic electronics, making it affordable, scalable and suitable for small towns and developing regions. This robot is small and mobile, capable of operating in narrow streets, lanes, or internal roads, where big machines can't go. Some detection systems only identify pothole locations. This robot actually measures the depth of the pothole using ultrasonic sensors, then calculates how much bitumen is needed, ensuring more accurate and efficient repairs.

3 Existing Approach

HERON is a project that is funded under the Horizon 2020 initiative by the European Union, focusing on the upkeep and advancement of road infrastructure. The project builds upon existing road infrastructure frameworks developed through affiliated initiatives. The key objective of HERON is to introduce cutting-edge engineering solutions that interconnect different transportation modes, ensuring smooth transitions in case of major disruptions affecting one mode of travel.

Several EU-funded initiatives, similar to HERON, aim to develop autonomous ground robotic vehicles complemented by drones for improved monitoring, evaluation,

and maintenance of road networks. One such project, InfraROB, is dedicated to automating and modernizing road construction and maintenance operations. This project focuses on designing autonomous robotic machinery for tasks such as road marking, resurfacing, and repairing cracks and potholes. Additionally, InfraROB seeks to implement collaborative robotic safety systems to enhance the protection of both construction workers and road users.

Simultaneously, the OMICRON project, also funded by the European Union, is developing an Intelligent Asset Management Platform (IAMP) that is equipped with cutting-edge, region-specific technologies. This platform is intended to revolutionize road construction and maintenance, renovation, and overall improvement across the European Union's Road networks. The initiative incorporates digital inspection tools, the development of a road digital twin, a decision-support system, advancements in intelligent construction methods, and targeted solutions for infrastructure-related challenges. The IAMP will be integrated with a digital twin system based on Building Information Modelling (BIM) principles.

Furthermore, the PANOPTIS project is focused on strengthening the resilience of road infrastructure, ensuring continued functionality even in extreme conditions such as severe weather events, landslides, and earthquakes. Its main goal is to combine customized climate change projections for road networks with advanced simulation models covering structural and geotechnical aspects, supplemented by real-time sensor data from both conventional and innovative monitoring systems.

Each of these initiatives, including HERON, is financially supported by the European Union under the Horizon 2020 Research and Innovation Program. HERON is one of three Research and Innovation Actions dedicated to reduce the environmental impact and achieve a fully automated road infrastructure maintenance under the Horizon 2020.

4 Objectives

4.1 Primary Objectives

1. To accurately detect potholes of varying sizes and depths using ultrasonic sensors, ensuring precise and consistent detection even in challenging environmental conditions.
2. To efficiently fill detected potholes using a servo motor and suitable filling material, effectively restoring the road surface and preventing further structural damage.
3. To ensure safe navigation by implementing robust obstacle avoidance using IR sensors, allowing the robot to operate autonomously without collisions or mission disruptions.
4. To enable real-time location tracking through the integration of a GPS sensor, allowing continuous monitoring and effective coordination of the robot's movements.
5. To provide reliable data transmission of location and status updates to the Blynk cloud, supporting remote monitoring, control, and analysis of the robot's operations.

4.2 Secondary Objectives

1. Energy-Effective Operation: Optimize power consumption to maximize battery life and minimize environmental impact.

2. Adaptability to Different Pothole Conditions: Design the robot to handle a wide range of pothole sizes, depths, and shapes, icing its effectiveness in different road conditions.
3. Robust Performance in Challenging surroundings: Equip the robot with features that enable it to operate reliably in colorful rainfall conditions, terrains, and business situations.
4. User-Friendly Interface: Gives a simple and intuitive interface through the Blynk app, allowing users to fluently cover the robot's status, view position data, and control its operations.
5. Scalability for unborn Expansion: Design the robot with a modular armature that facilitates easy integration with fresh detectors, selectors, or communication protocols, allowing for unborn advancements and expansion.

By fastening on these primary and secondary objectives, the pothole- detecting and filling robot can be developed to meet the specific requirements of road conservation and ameliorate the safety and effectiveness of transportation structure.

5 System Requirements

5.1 Software

1. Libraries and setup: Include necessary libraries ESP32WiFi, ESP32HTTPClient, BlynkSimpleESP32, Adafruit Sensor, Adafruit_BNO055, Servo, New Ping. Set up Wi-Fi connection and Blynk authentication.
2. GPS Location: Get the robot's position using the GPS detector and send the same to the Blynk cloud.
3. Blynk Integration: Produce a Blynk design and addcontraptions to display the robot's position and other data. Use Blynk functions to shoot andadmit data between the robot and the cloud.

5.2 Hardware

1. ESP32 Development Board: An important microcontroller with built-in Wi-Fi and Bluetooth capabilities.
2. Ultrasonic Sensor: Which will be used to measure the distance to the road face, helping to descry potholes.
3. Servo Motor: This will be used to control the medium that fills the potholes.
4. IR Sensor: responsible to descry obstacles in the path of the robot, precluding collisions.
5. GPS Module: This will determine the robot's position and transmit the same to the Blynk cloud.
6. Motor Driver: This will control the speed and direction of the DC motors.
7. DC Motors: These will power the robot's movement.
8. Battery: This will give power to the entire system.
9. Breadboard: This will be used to connect all the components together.
10. Jumper Cables: These will connect the components on the breadboard.

6 Proposed System

In (Fig. 1) Automatic Road Maintenance Robot include sensors, actuator and cloud-based monitoring system for the detection and repair of potholes. The sensor block consists of an Ultrasonic Sensor for sensing Potholes; to avoid any crashes or collision it consists of Infrared sensor and a GPS module for acquiring real time location data. ESP 32 microcontroller is the main processing unit and inputs to it are provided by the sensor. A Battery supply powers all the components for smooth operation. As soon as a pothole is detected, the actuator block is activated by the ESP32 Microcontroller, which includes a servo motor that is responsible for dispensing filling material and a motor driver controlling DC motors to drive the movement of the robot. The dispenser stores approximately 750 g to 1 kg of a filling medium, which includes concrete or a plastic-cement mixture. The servo motor controls the release of the material. After the initial dispense, the ultrasonic sensor re-checks the depth of the pothole. If the depth is still above the threshold, additional filling is dispensed in a controlled manner until the surface is sufficiently level. To ensure even distribution and compaction, a **roller is mounted at the rear** of the robot. This roller passes over the filled pothole, leveling and compacting the material for a smooth and durable surface finish. The ESP 32 also sends real time data to the Blynk Cloud which makes remote monitoring very easy. To enable smooth operation on varied surfaces like highways, concrete roads, and dirt roads, the robot uses a **rocker-bogie mechanism** instead of regular wheels. This suspension system, commonly used in Mars rovers, allows the robot to **maintain stability and traction** over uneven terrain without active suspension. The rocker-bogie system consists of linked arms and wheels that allow one side to climb over obstacles while the other remains grounded, maintaining the frame's orientation. This ensures that the **ultrasonic sensor alignment remains stable**, allowing accurate depth detection. It also **minimizes vibrations** and distributes weight evenly, which is essential for precise filling and compacting. This

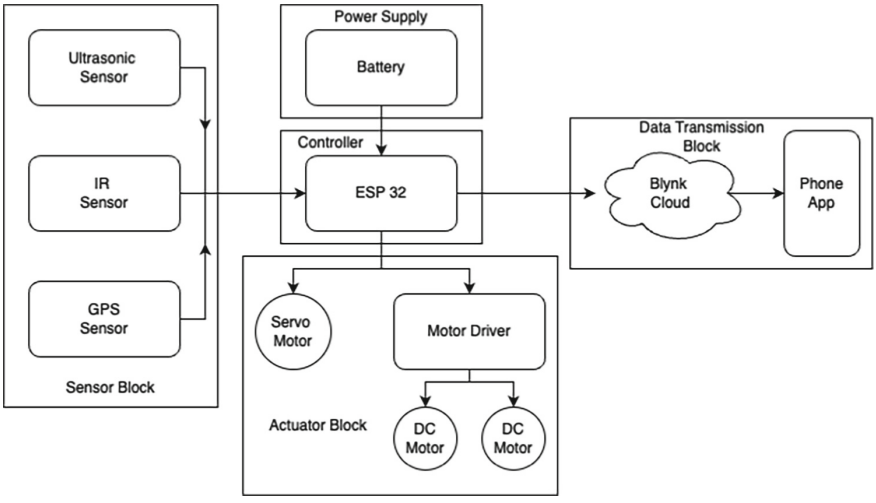


Fig. 1. Proposed block diagram of the system.

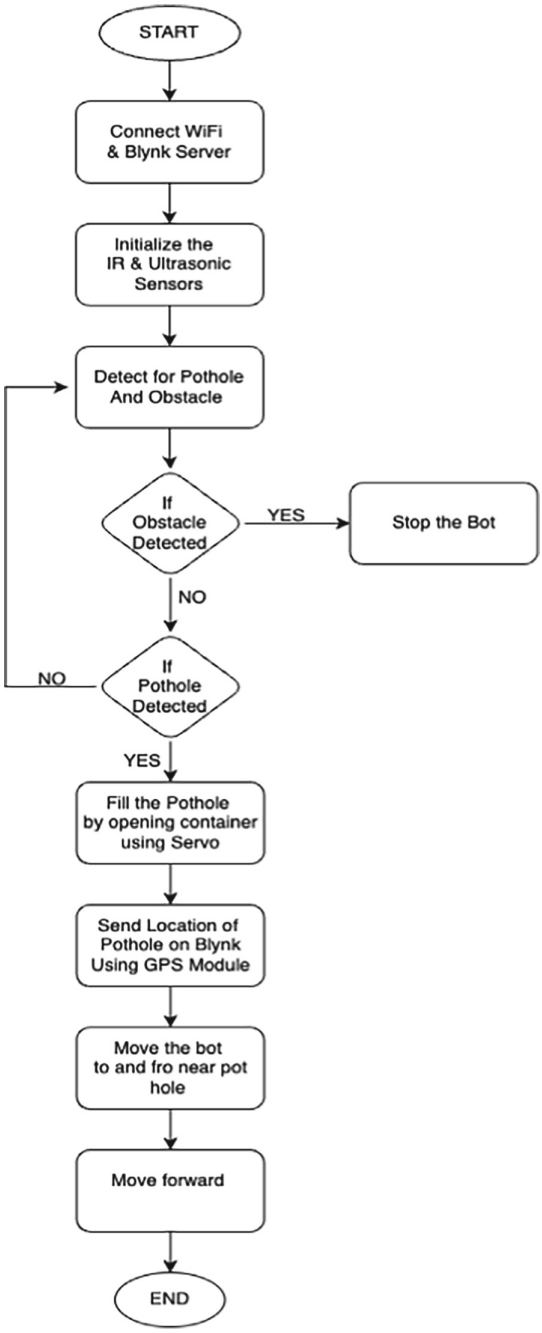


Fig. 2. Flowchart of the system

passive design allows the robot to **navigate debris, stones, or slopes** on rural roads while still functioning reliably on flat urban highways—without needing mechanical modifications or sensor recalibration.

Therefore, this system gives efficient automation of the road maintenance and enhances the road quality with minimum human effort.

When a pothole is detected by the robot, the system is designed to notify a human operator or maintenance authority in real-time. This adds a level of supervision, logging, and decision-making support (Fig. 2).

7 Flowchart

This flowchart explains the working of every step of the automatic road maintenance robot, which automatically detects and also fills the potholes. This flowchart is a structured sequence which ensures efficient operation of the road maintenance process.

8 Prototype Dimension

Prototype contains all the electronic components and the utensil placed on the rocker bogie. The overall dimension is 33 cm by 25 cm.. The utensil, which holds a slurry of concrete, has dimensions of 15 cm by 15 cm by 15 cm, making it a cube. The volume of the utensil is 0.75 kg. However, the weight (0.75 kg) would allow us to determine the density of the slurry if required, by using the formula for density.

9 Future Scope

Currently, the system uses depth-based detection alone and cannot differentiate between actual potholes and other surface anomalies such as road depressions or water-filled puddles. This limitation may lead to false positives.

As part of future development, additional sensing mechanisms—such as **moisture sensors**, **thermal/IR imaging**, or **computer vision with AI models**—will be integrated to accurately differentiate potholes from puddles or temporary depressions, thereby improving detection accuracy and material efficiency.

Future road maintenance bots will integrate with smart city infrastructure, sharing real-time data to improve road conditions and enable predictive maintenance. With advanced sensing technologies like computer vision and machine learning, these bots will detect potholes more accurately, even in challenging conditions. They will also learn and adapt to new road damages over time. Eco-friendly and cost-effective, the bots will use sustainable materials and reduce labor costs, speeding up repairs and minimizing disruptions. These bots can be used on private roads, industrial campuses, and through rental models for municipalities. Ongoing research will focus on durable materials, energy efficiency, and improved navigation in GPS-denied environments.

10 Conclusion

This innovative pothole-filling robot, powered by the ESP32 microcontroller, offers a practical and effective result for perfecting road structure. By using advanced technologies such as ultrasonic detectors, servo motors, and GPS, the robot can autonomously descry, detect, and fill potholes, reducing the need for homemade labor and creating safer roads for all. The integration with Blynk further enhances the robot's capabilities by furnishing real-time monitoring and remote control. This allows users to track the robot's progress, identify areas that bear attention, and optimize its operations. By automating the process of pothole form, this robot can significantly meliorate road safety and reduce conservation costs, making it a precious asset for civic and pastoral communities likewise.

References

1. Katsamenis, I., et al.: Robotic maintenance of road infrastructures: the HERON project. In: Proceedings of the 15th International Conference on Pervasive Technologies Related to Assistive Environments (PETRA '22). Association for Computing Machinery, New York, NY, USA (2022)
2. Miroslaw, S., Chris, H.: Automation and robotics for road construction and maintenance. J. Trans. Eng.-asce **116**(3)
3. Bavelos, A., Anastasiou, E., Dimitropoulos, N., Oikonomou, G., Makris, S.: Augmented reality-based method for road maintenance operators in human–robot collaborative interventions. Comput.-Aided Civ. Infrastruct. Eng. **39**,(2024). https://doi.org/10.1111/mice.13185
4. Gurwani, P., Mandal, R., Chaudhari, S., Jadhav, M., Sonawane, S.: Smart IOT based pothole detection and filling system. 2023 International Conference on Intelligent Systems for Communication, IoT and Security (ICISCoIS), Coimbatore, India (2023)
5. Prakash, G., Raadha, S., Swami, T., Mahalakshmi, E.: Sensor-based espial of potholes and humps on roads with instant notification alert using IoT. 2022 International Conference on Computer, Power and Communications (ICCPC), Chennai, India, pp. 281–285 (2022). https://doi.org/10.1109/ICCPC55978.2022.10072130
6. Cai, E., et al.: Estimating structural motions in extreme environmental conditions——A dynamic correlation filter-based computer vision approach. Mech. Syst. Signal Process. **215**, 111398 (2024). ISSN 0888-3270

Leveraging AI and Blockchain Technology for Enhancing Healthcare Data Management and Patient Care

Anagha Kulkarni[1(✉)], Priyanka Pawar[1], Harshal Raje[2], Bhavana Pansare[1], and Manisha Bhende[1]

[1] Dr. D. Y. Patil School of Science & Technology, Dr. D. Y. Patil Vidyapeeth, Pimpri, Pune, India
anaghak313@gmail.com

[2] Global Business School and Research Centre, Dr. D. Y. Patil Vidyapeeth, Pimpri, Pune, India

Abstract. Although there have been notable technology developments in the healthcare industry, issues with handling data, security, and interoperability still exist. Large volumes of complicated data, such as patient records, diagnostic pictures, treatment histories, and genetic profiles, are produced by the healthcare ecosystem. Although AI has the potential to improve predictive analytics, individualized treatment planning, and enhanced diagnostics, mainstream use is hampered by worries about biases in algorithms, privacy of information, and legacy system integration. Blockchain technology solves problems like security of data, traceability, which is and patient consent by providing a decentralized, impenetrable framework for safe healthcare data transmission. Using multidisciplinary techniques from healthcare informatics, ML, cryptography, user experience, and ethical compliance, this project attempts to create an integrated framework that combines blockchain for safe data governance with AI for intelligent data analysis.

Keywords: AI · Blockchain · ML · EHR

1 Introduction

Technology is causing a rapid revolution in the healthcare sector, but there are still many challenges to overcome, especially in the areas of patient care, security, and data management. There is an opportunity to improve patient outcomes thanks to the enormous quantity of data processed in the healthcare industry, which includes genetic information, treatment plans, diagnostic pictures, and patient records [1]. In order to manage this data successfully, strong systems that support the effective sharing of information across healthcare professionals while guaranteeing data accessibility, privacy, and integrity are needed. The potential for artificial intelligence (AI) to transform healthcare through its unparalleled accuracy and speed in analyzing and interpreting big datasets has been demonstrated. More and more, AI-driven technologies are being used for individualized treatment planning, predictive analytics, and diagnosis, resulting in more precise and faster healthcare interventions. Despite these developments, worries about data privacy,

S. Fong et al. (Eds.): ICT4SD 2025, LNNS 1651, pp. 464–472, 2026.
https://doi.org/10.1007/978-3-032-06688-6_45

the possibility of discrimination in the algorithms used for AI, and the compatibility of AI systems with the current healthcare infrastructure continue to impede the use of AI in healthcare. With its decentralized and unchangeable ledger system, Blockchain technology presents a promising answer to the problems with data safety and compatibility in the healthcare industry. Blockchain technology can guarantee patient anonymity while ensuring safe storage and sharing of healthcare data with complete transparency and traceability. Many of the issues surrounding the adoption of AI in healthcare, especially those pertaining to data integrity and privacy, may be resolved by its capacity to offer a safe and impenetrable platform for data interchange [2].

Numerous obstacles have been experienced by the healthcare industry, which has led to the research problem of combining artificial intelligence with blockchain technology. Concerns about data management, privacy, security, and interoperability have grown as healthcare data becomes more sophisticated. These difficulties call for the development of innovative technological solutions as well as customized patient care. Healthcare, medicine, ML, AI, blockchain technology, statistical computing, privacy & security of data, ethics, legal studies, user interface, and user experience are all areas where this research project is relevant from an interdisciplinary standpoint. The biggest application domain is healthcare, where improving patient outcomes requires accurate, safe, and rapid data collecting. Applications of AI & ML include pattern identification, personalized therapy, diagnostic support, and predictive analytics [3]. Although it necessitates familiarity with distributed ledgers, smart contracts, consensus techniques, and cryptography, blockchain technology offers a decentralized, secure foundation for data management. Big data analytics and statistical analysis are used to handle large amounts of healthcare data, while information security and cybersecurity guarantee data privacy and regulatory compliance. It is necessary to handle legal and ethical considerations such algorithm bias, information ownership, patient consent, and regulatory compliance. While human-computer interface and user experience design guarantee the efficacy of the Blockchain & AI system, healthcare engineering and information technology systems guarantee smooth integration with the existing healthcare infrastructure [4].

Research Objectives:

RO 1: This study's main goal is to investigate and create a comprehensive structure that uses Blockchain technology and artificial intelligence (AI) to improve patient outcomes, data management, and healthcare delivery.
RO 2: To create and put into place a system based on Blockchain technology that securely stores and handles medical data, guaranteeing its accessibility, privacy, and integrity for authorized users.
RO 3: To develop a Blockchain-based infrastructure that enables safe, compliant data exchange across various healthcare systems and providers.

2 Literature Review

The study examines the possible advantages, difficulties, and solutions of integrating blockchain technology and artificial intelligence in the healthcare industry. It seeks to give a thorough overview, point out problems, suggest fixes, and shed light on how

healthcare will develop in the future [5]. Authors discussed how Blockchain technology and artificial intelligence (AI) are transforming the healthcare industry by increasing productivity, cutting expenses, and democratizing the availability of medical records for patients and risk management [6]. Although the potential for improving security and performance in standard EHRs is great, the full extent of this technology is still unknown. The incorporation of artificial intelligence and blockchain technology in healthcare has great promise to improve security and autonomy [9]. By simplifying data interchange between hospitals, diagnostic labs, and clinicians, avoiding deceit in clinical investigations, and decentralizing data protection, blockchain technology improves the efficacy, security, and transparency of healthcare data [11]. Blockchain technology which is based on bitcoin has more uses in the healthcare industry than ever before. Blockchain's potential in the field of healthcare is a subject of ongoing investigation by experts in science and technology, healthcare IT, and other fields. A mapping study examines blockchain-based applications and highlights problems with healthcare management systems. Through comparative analysis, the report offers insights into research obstacles and suggests a taxonomy for potential solutions [12]. The work highlights continuing problems as well as challenges in handling and safeguarding this large range of patient data. It also analyzes safety challenges in intelligent healthcare systems and suggests a framework for analyzing malware on wearable devices based on blockchain and artificial intelligence [7]. Blockchain and AI technology are being used in a number of industries, such as construction, banking and finance. AI algorithms support AI models and improve the efficiency of medical blockchain storage. Blockchain offers training with safe, traceable, varied, and unchangeable medical data [8]. Researchers are looking into novel technologies like blockchain and artificial intelligence (AI) to enhance digital healthcare workflows, democratize clinical processes, and possibly support public health initiatives in response to the COVID-19 pandemic [10]. The importance of the technology of blockchain in healthcare is examined in this article, which also covers its features, applications, architecture, difficulties, and potential future study areas. It also covers ways to improve network security and privacy as well as security assaults [13]. AI has completely changed the healthcare industry by advancing medication research, remote patient care, and medical imaging. But issues with cost, safety, and privacy continue to exist. For AI to be used, efficient governance and regulatory concerns are essential [14].

3 Role of AI in Healthcare Data Management:

Figure 1 depicts the role of AI in healthcare data management. The collection, processing, analysis, and use of healthcare data are all being completely transformed by AI. AI provides scalable, intelligent solutions to effectively manage and extract valuable insights from the growing amount of medical data coming from wearables, imaging, and electronic health records EHRs, and patient interactions. AI systems are able to manage missing values, categorize unstructured data for improved organization, and automatically detect and fix problems. Physician notes and reports may be transformed into structured datasets for analytics using AI-powered natural language processing (NLP) technologies, which can also extract pertinent elements from unstructured texts. By mapping disparate nomenclature and coding systems, AI makes it easier to harmonize data from several sources and create a single patient profile. The streams of data

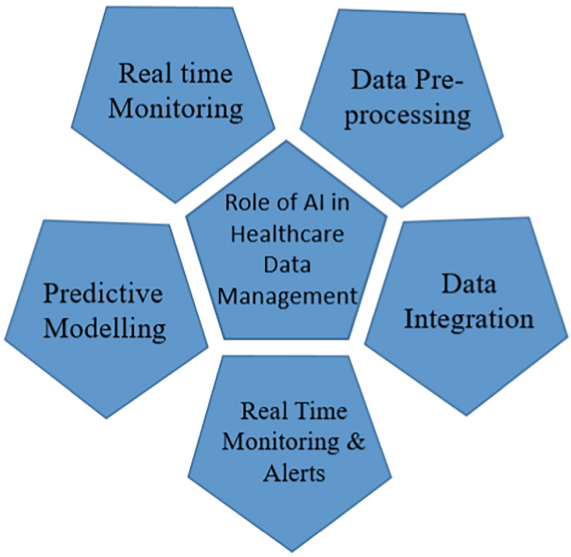


Fig. 1. Role of AI in Healthcare Data Management

from IoT devices and EHRs may be processed by machine learning models to identify anomalous health trends and send out real-time alerts to caregivers or physicians. Using CDSS driven by AI, AI evaluates vast amounts of patient data to forecast the course of an illness, provide individualized treatment regimens, and assist in decision-making. Administrative duties including arranging appointments, processing insurance claims, medical coding and billing, and patient triage are all automated by AI. By identifying irregularities in data access or usage patterns and enabling behavioral analytics to stop fraud or illegal data modification, it also improves data security. By organizing and analyzing large datasets, finding patient cohorts for clinical trials, identifying patterns across populations for insights into public health, and accelerating biomedical research by identifying correlations across genomics, imaging, and clinical data, artificial intelligence (AI) promotes research and innovation.

4 Role of Blockchain in Healthcare Data Management:

Figure 2 describes the application areas of blockchain technology in healthcare data management. Blockchain technology addresses issues with privacy of information, interoperability, and stakeholder confidence by providing a decentralized, transparent, and safe framework for handling healthcare data. It offers safe and impenetrable data storage, allowing hospitals, laboratories, insurers, and patients to share access while retaining ownership and control. Consent-driven sharing is made possible by patient-centric data ownership, which gives individuals the authority to own and manage their medical records. By automating access control rules, smart contracts guarantee adherence to privacy laws such as GDPR and HIPAA. By encouraging standardized, safe data transmission between entities utilizing consensus protocols, blockchain might improve

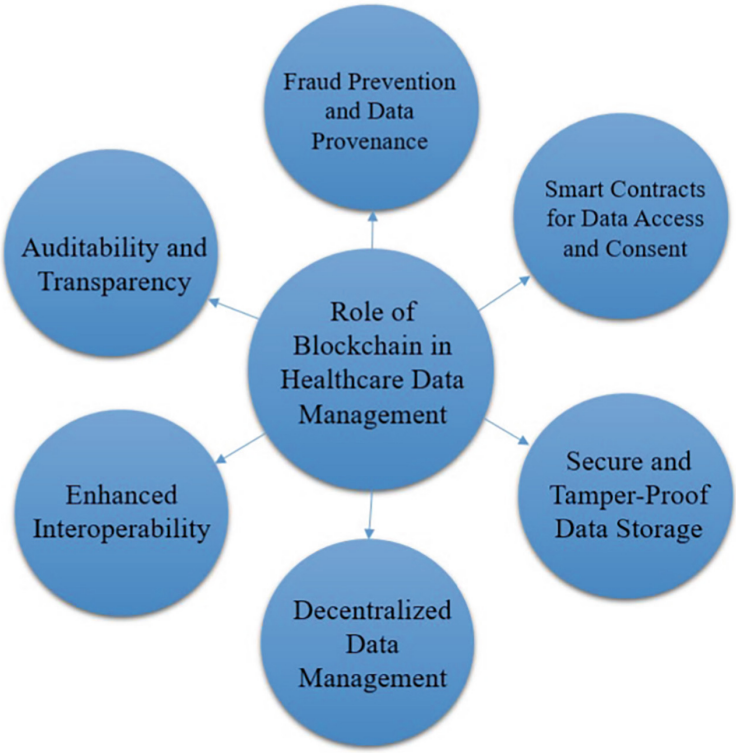


Fig. 2. Role of Blockchain in Healthcare Data Management

interoperability. Because every data contact is recorded and timestamped, a comprehensive audit trail that is available to patients as well as regulatory agencies, blockchain also offers the advantages of auditability and transparency. By confirming the origin and change of data, blockchain also guards against fraud and data provenance. Clinical trials and the supply chain for pharmaceuticals are both tracked by blockchain, which also ensures transparency in clinical studies. Benefits include fostering confidence between consumers, healthcare professionals, and regulators as well as efficiency, security, and compliance.

5 Proposed Methodology

Figure 3 proposes integration of AI & Blockchain technology for secure healthcare data management. A framework for a healthcare data management system that combines blockchain technology with AI is shown in the diagram. Here is a detailed breakdown of every element:

Data Source: Hospitals, clinics, patients, physicians, and other players in the healthcare ecosystem are the main sources of healthcare data. They are the starting locations for unprocessed administrative and medical data, which is necessary for additional processing, analysis, and safe system-wide dissemination.

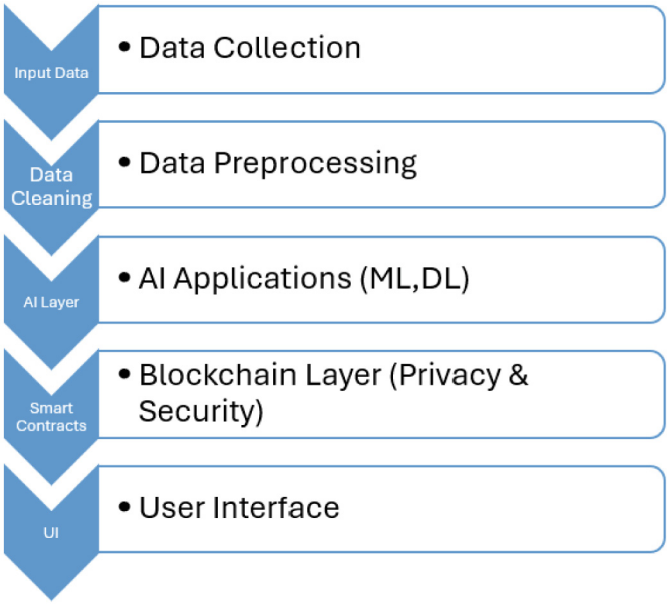


Fig. 3: Proposed system architecture using AI & Blockchain

Data Collection: Data collection from healthcare participants, such as patients, providers, and institutions, is the responsibility of this component. It serves as the main point of entry for unprocessed medical data into the system. It guarantees that data is appropriately routed for preparatory processing, analysis, and safe storage by acting as a link between sources of data and downstream procedures.

Data Processing: The design has two main pipelines that work together to analyze and use the healthcare data that has been gathered. One pipeline uses blockchain technology, while the other uses AI. Each pipeline contributes to the improvement of healthcare services in a unique but complementary way.

AI Framework: An essential part of the AI pipeline is the Data Pre-Processing module, which converts unstructured medical data into a machine-readable format for further analysis and model training. Data cleansing, filtering, data transformation, missing data management, reduction of noise, normalization, and scaling are among the tasks it completes. While data filtering eliminates superfluous or unnecessary data, data cleaning finds and fixes errors, duplication, and damaged entries. In order to translate categorical variables into numerical values for ML algorithms, data transformation transforms data into standardized formats. Techniques including imputation, deletion, and flagging are used to deal with missing data. Outliers and incorrect numbers that could distort assessment or model predictions are found and eliminated using noise reduction. Scaling and normalization provide equitable participation in AI training. Healthcare data may be processed in batches or in real time thanks to the module's API accessibility. Higher model accuracy, shorter training times, and the avoidance of biased or deceptive results

are all guaranteed by high-quality pre-processing. It ensures that clean, pertinent, and superior information reaches the ML phase by acting as an intermediary for model input.

Blockchain Framework: By guaranteeing safe, accountable, and trust-based data activities, the Blockchain-based Pipeline is a decentralized, impenetrable architecture that improves the healthcare data ecosystem. Automated, rule-driven activities take the place of human processes thanks to smart contracts, which are self-executing code recorded on a blockchain. Permissions for data access, management of patient consent, audit logs and use monitoring, privacy by architecture, and API connectivity are important features. Blockchain projects in healthcare use transparency, decentralization, and immutability to solve enduring data management issues. Secure data exchange between institutions, managing patient permission, tracking healthcare transactions, maintaining the integrity and accountability of medical information, and improving data governance are some of the main use cases. Secure data exchange between organizations, patient consent management, healthcare transaction traceability, medical record integrity and auditability, and encouraging adherence to healthcare laws such as HIPAA and GDPR are examples of core use cases. Benefits include strengthening data governance, increasing stakeholder openness, encouraging regulatory compliance, and laying the groundwork for future healthcare breakthroughs by fostering digital trust.

A key instrument for guaranteeing safe, accountable, and trust-based information processing in the healthcare industry is the Blockchain Pipeline. Blockchain applications can enhance data governance, transparency, and regulatory compliance by utilizing immutability, decentralization, and transparency.

Security & Privacy: A key component of the design is the Security & Privacy layer, which guarantees strong protections and compliance with regulations for data activities, especially those involving sensitive medical data. It ensures compliance with national and international data protection regulations, including PDPA, PIPEDA, GDPR, HIPAA, and other local health data legislation. To limit unwanted data disclosure, data encryption is used. Role-based access control, attribute-based access control, and multi-factor authentication are used to provide access control and authentication. Techniques like anonymization and pseudonymization are employed to protect privacy without sacrificing usefulness. To find irregularities and stop abuse, audits and monitoring are carried out continuously. By keeping track of data exchanges and access records, blockchain-backed integrity is preserved, guaranteeing complete traceability and guarding against manipulation or fabrication. To identify and address risks in real time, response to incidents and threat identification are connected with event management and security information systems. By enforcing appropriate data usage practices, the Security & Privacy layer fosters ethical AI research, lowers legal risk and liability, and increases trust with patients and stakeholders.

User Interface: An AI-powered healthcare system's UI is its front-end layer and the main interface with end users, including patients, administrators, healthcare professionals, and external stakeholders. In addition to using dashboards, charts, alerts, and summaries to support data-driven decision-making, it offers AI-generated insights, including diagnostic forecasts, treatment suggestions, and health risk assessments. Additionally, it guarantees the integrity of healthcare records by enabling users to examine those supported by blockchain technology. The user interface prioritizes quickness,

accessibility, and usefulness across a range of devices. It has capabilities to evaluate AI suggestions, submit consent forms, allow or revoke data access, and interact with care plans. To protect data privacy, it also incorporates user-level audit trails, session timeouts, and secure login methods. A cohesive experience for handling health data from several sources may be ensured by the UI's ability to interact with external systems such as wearables, lab systems, insurance portals, and EHRs. Improved user trust, a bridge between human decision-making and intricate backend technologies, and assistance for real-time monitoring, individualized care delivery, and effective healthcare processes are some advantages of this system.

6 Conclusion

There are several chances to enhance outcomes for patients, handling of data, and healthcare delivery by combining blockchain technology with artificial intelligence (AI). Predictive analytics, individualized treatment plans, and accurate diagnosis are made possible by AI's capacity to process enormous datasets, including genetic data, diagnostic pictures, and electronic medical records. But issues like algorithmic biases and data privacy call for strong data governance and moral supervision. Blockchain technology ensures safe storage, sharing, reliability of information, and interoperability across healthcare providers by offering a decentralized and unchangeable ledger for medical records. In the healthcare system, this promotes openness and trust. Notwithstanding its potential, obstacles including integration difficulties and problems with regulatory compliance prevent widespread usage. In order to create comprehensive frameworks that integrate blockchain's security features with AI's analytical capabilities, interdisciplinary cooperation between healthcare practitioners, data scientists, legal professionals, and legislators is required.

Acknowledgement. The authors gratefully acknowledge the financial support provided by the Dr. D.Y. Patil School of Science and Technology, Dr. D.Y Patil Vidyapeeth, Pune, India through the Seed Money Project. We also extend our thanks to Dr. D.Y. Patil Vidyapeeth for providing the necessary facilities to conduct this research.

References

1. Tabriz, A.A., et al.: What should accountable care organizations learn from the failure of health maintenance organizations? A theory based systematic review of the literature. Soc. Determ. Health. **3**, 222–247 (2017). https://doi.org/10.22037/SDH.V3I4.20919
2. Abou El Houda, Z., Hafid, A.S., Khoukhi, L., Brik, B.: When collaborative federated learning meets blockchain to preserve privacy in healthcare. IEEE Trans. Netwk. Sci. Eng. (2023). https://doi.org/10.1109/TNSE.2022.3211192
3. Abu-Elezz, I., Hassan, A., Nazeemudeen, A., Househ, M., Abd-Alrazaq, A.: The benefits and threats of blockchain technology in healthcare: a scoping review. Int. J. Med. Informatics **142**, 104246 (2020). https://doi.org/10.1016/j.ijmedinf.2020.104246
4. Agbo, C., Mahmoud, Q., Eklund, J.: Blockchain technology in healthcare: a systematic review. Healthcare. **7**(2), 56 (2019). https://doi.org/10.3390/healthcare7020056

5. Biswas, Ms., Singh, Dr.: Application of AI and blockchain in healthcare industry–A review. J. Adv. Zool. **45** (2024). https://doi.org/10.53555/jaz.v45i2.3983
6. Tagde, P., et al.: Blockchain and artificial intelligence technology in e-Health. Environ. Sci. Pollut. Res. Int. **28**(38), 52810–52831 (2021). https://doi.org/10.1007/s11356-021-16223-0. Epub 2021 Sep 2. PMID: 34476701; PMCID: PMC8412875
7. Alabdulatif, A., Khalil, I., Saidur Rahman, M.: Security of blockchain and AI-Empowered smart healthcare: application-based analysis. Appl. Sci. **12**, 11039 (2022). https://doi.org/10.3390/app122111039
8. Sai, S., Vinay, C., Kim-Kwang, C., Biplab, S., Joel, R.: Confluence of blockchain and artificial intelligence technologies for secure and scalable healthcare solutions: a review. IEEE Internet Things J. 1 (2022). https://doi.org/10.1109/JIOT.2022.3232793
9. Rao, K.P., Manvi, S.: Survey on electronic health record management using amalgamation of artificial intelligence and blockchain technologies. Acta Informat. Pragensia **12**(1), Forthcoming articles (2023). https://doi.org/10.18267/j.aip.194
10. Jabarulla, M.Y., Lee, H.N.: A Blockchain and artificial intelligence-based, patient-centric healthcare system for combating the COVID-19 pandemic: opportunities and applications. Healthcare (Basel). **9**(8), 1019 (2021). https://doi.org/10.3390/healthcare9081019.PMID:34442156;PMCID:PMC8391524
11. Haleem, A., et al.: Blockchain technology applications in healthcare: an overview. Int. J. Intell. Netw. **2**, 130–139 (2021). ISSN 2666–6030, https://doi.org/10.1016/j.ijin.2021.09.005
12. Singh, D., Monga, S., Tanwar, S., Hong, W.-C., Sharma, R., He, Y.-L.: Adoption of blockchain technology in healthcare: challenges, solutions, and comparisons. Appl. Sci. **13**, 2380 (2023). https://doi.org/10.3390/app13042380
13. Andrew, J., et al.: Blockchain for healthcare systems: Architecture, security challenges, trends and future directions. J. Netw. Comput. Appl. **215**, 103633 (2023). ISSN 1084-8045, https://doi.org/10.1016/j.jnca.2023.103633
14. Al Kuwaiti, A., et al.: A review of the role of artificial intelligence in healthcare. J. Pers. Med. **13**(6), 951 (2023). https://doi.org/10.3390/jpm13060951.PMID:37373940;PMCID:PMC10301994

Data-Driven Optimization of Hybrid Renewable Energy Systems: Managing Net Metering Costs Through Machine Learning

V. K. Abhang, Y. A. Shinde, S. N. Shingote, Somal Adik(✉), Nikhil Aglawe, Vivek Akolkar, and Pratik Ghondage

Department of Computer Engineering, Amrutvahini College of Engineering, Sangamner, India
{vikramabhang,yogesh.shinde,sayaram.shingote}@avcoe.org, somaladik@gmail.com

Abstract. This project explores how machine learning can help optimize hybrid renewable energy systems, with a special focus on managing net metering costs. By analyzing real-time data from renewable sources and consumer energy usage, the goal is to create a smart, efficient framework that improves energy reliability while keeping costs low. The idea is to strike a balance ensuring that energy production and consumption align seamlessly with changing demand and environmental conditions. To make this happen, we're using the Open Energy Modelling Framework (OEMOF), which helps optimize how energy is distributed, stored, and interacted with the grid. With OEMOF, we can simulate energy flows, make better decisions about energy trading and self-consumption, and develop cost-effective net metering strategies. On top of that, advanced predictive models for weather and energy demand forecasting allow for proactive system adjustments, making sure the setup remains efficient and reliable.Beyond the technical side, this approach directly supports the global shift toward sustainable energy. By making hybrid renewable energy systems more cost-effective and scalable, it not only helps individuals and businesses save money but also contributes to reducing carbon footprints moving us one step closer to a greener future.

Keywords: Energy Efficiency · Hybrid Renewable Energy Systems · Machine Learning · Net Metering

1 Introduction

As the world shifts toward cleaner energy, Hybrid Renewable Energy Systems (HRES) are becoming a key solution by combining sources like solar and wind to provide a more reliable and sustainable alternative to fossil fuels. However, efficiently managing HRES is challenging due to fluctuating weather, variable energy demands, and complex net metering policies. Traditional static models often fall short in adapting to real-world changes, leading to inefficiencies. This is where machine learning (ML) proves valuable by analyzing large datasets, ML can forecast energy demand, optimize performance, and enhance energy distribution. Techniques like short-term and long-term forecasting

S. Fong et al. (Eds.): ICT4SD 2025, LNNS 1651, pp. 473–480, 2026.
https://doi.org/10.1007/978-3-032-06688-6_46

help balance supply and demand, while optimization methods ensure efficient resource allocation. The OEMF (Optimization and Energy Management Framework) model further strengthens this by integrating ML based forecasting with advanced optimization, enabling dynamic system adjustments. This approach not only improves technical performance but also considers financial benefits for consumers by reducing energy costs and improving reliability.

1.1 Project Idea

Our project, "Data-driven Optimization of Hybrid Renewable Energy Systems: Managing Net Metering Costs Through Machine Learning" focuses on building a smart, efficient, and cost-effective energy management system that combines solar panels, battery storage, and grid connectivity. Using historical energy data, real-time tariff structures, and weather forecasts, we aim to intelligently manage how energy is produced, consumed, and exported to the grid. One key challenge is managing net metering costs fees that can rise when excess energy is sent back to the grid at unfavorable rates. To solve this, we use machine learning, particularly the XGBoost model, which predicts solar power generation by analyzing data like solar irradiance, temperature, and energy demand (Figs. 1 and 2).

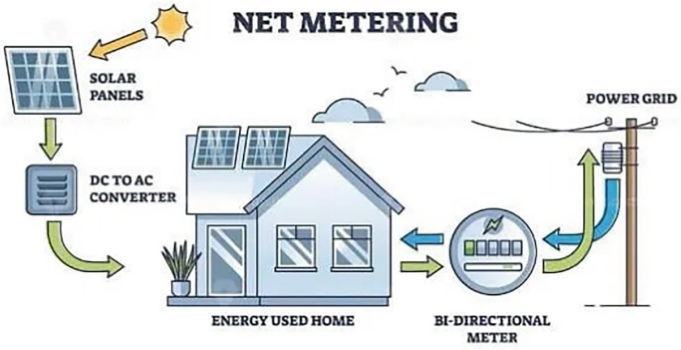


Fig. 1. Net Metering

These predictions help the system optimize battery usage, reduce energy waste, and detect any faults early on. OEMOF (Open Energy Modeling Framework) is used to simulate and optimize the overall system, helping us size components efficiently and make smart energy decisions. The result is a solution that not only saves costs and boosts energy efficiency but also supports sustainability by reducing reliance on traditional power sources benefiting homeowners, businesses, and policymakers alike.

2 Literature Review

Hybrid Renewable Energy Systems (HRES) [1] have become a vital solution for addressing energy sustainability challenges. Optimizing these systems with machine learning has gained significant attention, as it allows for more precise load forecasting, component

sizing, and net-metering cost management. Traditional energy optimization methods often struggle to adapt to dynamic factors like changing weather conditions, fluctuations in energy demand, and regulatory policies. To overcome these challenges, recent research has focused on integrating metaheuristic algorithms and predictive models to improve efficiency.

Several studies have explored the role of machine learning in HRES optimization. Abdullah et al. [2] (2023) introduced a hybrid approach using CatBoost, LightGBM, and XGBoost to optimize component sizing, specifically for net-metering cost management. Their findings demonstrate that machine learning models can significantly enhance energy allocation by predicting demand and optimizing storage utilization. They also highlight how combining machine learning with metaheuristic optimization techniques can further refine system efficiency under different environmental conditions.

Similarly, Duman et al. [3] (2022) developed a metaheuristic search algorithm for optimizing solar photovoltaic (PV) parameter estimation. Their research shows that advanced optimization techniques can improve forecasting accuracy and overall cost efficiency. By incorporating historical and real-time solar radiation data, their model achieves better prediction accuracy.

Short-term solar power forecasting has also been a key area of research in renewable energy integration. [4] Rafati et al. (2021) proposed a heuristic model capable of predicting high-dimensional variations in solar power generation with improved reliability. Their model effectively captures fluctuations in solar energy output, allowing for more adaptive energy storage and distribution strategies.

3 Methodology

3.1 Data Collection

Accurate data collection is fundamental to the success of this research. Various sources were utilized to gather essential datasets required for optimizing the hybrid renewable energy system. Weather data, including solar radiation and temperature. These parameters are crucial in determining the [6] potential power generation from the photovoltaic (PV) system. The efficiency and capacity of the solar panels were considered to accurately estimate power output and degradation over time. The battery storage capacity, charge/discharge efficiency, and degradation factors were analyzed to optimize energy storage and usage effectively. Additionally, hourly and daily energy consumption data were collected to understand demand variations, ensuring a well-optimized system design.

3.2 Machine Learning-Based DC Power Prediction

To improve the accuracy of PV power generation forecasting, a machine learning-based approach was implemented. The methodology involved using weather data, including solar radiation and temperature, as input variables to [7] predict direct current (DC) power output from the PV system. XGBRegressor, an advanced gradient boosting machine learning algorithm, was employed due to its capability to handle nonlinear relationships

and enhance predictive performance. A validation process ensured the accuracy and generalization of the model before deployment. The predicted DC power values were then integrated into the energy system simulation to provide a realistic estimation of PV output under varying environmental conditions.

3.3 OEMOF-Based Energy System Optimization

The energy system was modeled and optimized using the Open Energy [11] Modeling Framework (OEMOF), a Python-based tool designed for simulating and optimizing energy systems. This tool was utilized to model the interactions between PV, battery storage, grid supply, and load demand, enabling an optimized energy flow. The primary optimization goals were to determine the optimal sizing of the PV and battery storage system, minimize overall energy costs, and enhance self-consumption and self-sufficiency by reducing grid dependency. The system follows an energy balance equation, which ensures that PV generation, grid supply, and battery discharge are always equal to demand, grid feed-in, and battery charge, Energy Balance Equation.

PV generation + Grid supply + Battery discharge = Demand + Grid feedin + Battery charge.

3.4 Net Metering Cost Optimization

Net metering is a critical factor in determining the economic feasibility of the renewable energy system. The optimization process evaluates multiple financial and energy-related parameters, including self-sufficiency, grid feed-in, cost savings, energy savings, and total investment. Self-sufficiency is assessed to determine the ability to meet demand through self-generated energy. [12] Grid feed-in analysis examines the excess energy fed into the grid and its financial impact. Cost savings are calculated by optimizing self-consumption, thereby reducing electricity expenses. Energy savings are analyzed to improve the utilization of renewable energy sources. The total investment, including capital and operational costs, is considered to provide a clear understanding of the financial feasibility of the system.

Formula:

1. yearly_energy_costs_conventional = demand*electricity_price
2. energy_bill_grid_import = Self.Grid_Import* electricity_price
3. pv_investment = pv_capacity * pv_capex
4. Bess_investment = bess_optimal_value * bess_capex
5. Total_investments = pv_investment + bess_investment
6. Cost_savings = yearly_energy_costs_conventional - energy_bill_grid_import + income_from_fit
7. Payback_period = total_investments / cost_savings

3.5 Work Flow Chart

First, data is gathered, including weather conditions, PV performance, battery capacity, and energy demand. The collected data is then preprocessed to remove errors and inconsistencies.

Once trained, the ML model predicts energy production and consumption patterns. This prediction is used in the energy system optimization process, where the optimal sizes of PV and battery storage are determined using OEMOF-based optimization techniques.

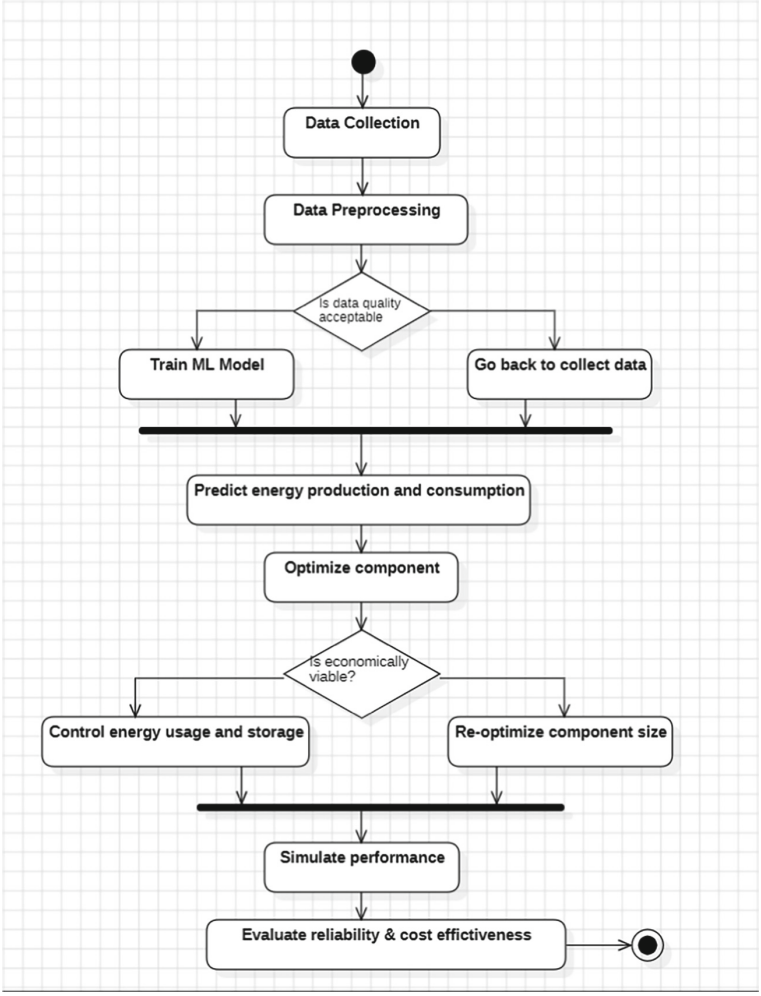


Fig. 2. Work Flow Diagram

4 Results

The Fig. 3 represent a smaller PV and battery system, leading to lower overall energy generation compared to the first system. However, a significant proportion of the generated energy is consumed directly, improving the self-consumption ratio and reducing dependency on grid imports. This makes the system more financially accessible and beneficial in the short term. Users prioritizing reduced initial costs, better self-utilization of solar energy, and quicker returns on investment would find this configuration more suitable.

Fig. 3. Results for Households User

The Fig.4 represent a smaller PV and battery system, leading to lower overall energy generation compared to the first system. However, a significant proportion of the generated energy is consumed directly, improving the self-consumption ratio and reducing dependency on grid imports. This makes the system more financially accessible and beneficial in the short term. Users prioritizing reduced initial costs, better self-utilization of solar energy, and quicker returns on investment would find this configuration more suitable.

Fig. 4. Result for Commercial user

5 Conclusion

The completion of this project has resulted in an efficient, data-driven solution for optimizing hybrid renewable energy systems and effectively managing net metering costs. This project represents a significant advancement in renewable energy management by providing a comprehensive, cost-efficient, and scalable approach to energy optimization. Through the integration of machine learning techniques, the system accurately predicts energy demand, optimizes resource allocation, and identifies cost-saving opportunities based on real-time data.

Extensive testing has confirmed the system's reliability, scalability, and applicability across various renewable energy sources and grid configurations. As the project concludes, it establishes a strong foundation for future enhancements and advancements in data-driven energy management solutions, paving the way for more sustainable and cost-effective energy solutions.

.

References

1. Abdullah, H.M., Park, S., Seong, K., Sangyong, L.: Hybrid renewable energy system design: a machine learning approach for optimal sizing with net-metering costs (2023). https://www.mdpi.com/2071-1050/15/11/8538
2. Duman, S., Kahraman, H.T., Sonmez, Y., Guvenc, U., Kati, M., Aras: A powerful meta-heuristic search algorithm for solving global optimization and real-world solar photovoltaic parameter estimation problems (2022). https://www.sciencedirect.com/science/article/pii/S2667010023000446
3. Rafati, A., Joorabian, M., Mashhour, E., Shaker: High dimensional very short-term solar power forecasting based on a data-driven heuristic method (2021). https://www.researchgate.net/publication/347444020_
4. Traor´e, A., Elgothamy, H., Zohdy: Optimal sizing of solar/wind hybrid off-grid microgrids using an enhanced genetic algorithm (2018). https://www.scirp.org/journal/paperinformation?paperid=84991
5. Memon, S.A., Patel, R.N.: An overview of optimization techniques used for sizing of hybrid renewable energy systems (2021). https://doi.org/10.1016/j.ref.2021.07.007
6. Das, M., Singh, M.A.K., Biswas: Techno-economic optimization of an off-grid hybrid renewable energy system using metaheuristic optimization approaches (2019). https://www.sciencedirect.com/science/article/abs/pii/S019689041930175X
7. Al-falahi, M.D.A., Jayasinghe, S.D.G., Enshaei, H.: A review on recent size optimization methodologies for standalone solar and wind hybrid renew able energy system. Energy Convers. Manag. (2017). https://www.researchgate.net/publication/316115488
8. Siddaiah, A.: Review on planning, configurations, modeling and optimization techniques of hybrid renewable energy systems for off-grid applications. Renew. Sustain. Energy Rev. (2016). https://www.researchgate.net/publication/290479769
9. Ramli, M.A.M., Bouchekara, H.R.E.H., Alghamdi, A.S.: Efficient energy managementina microgrid with intermittent renewable energy and storage sources. Sustainability (2019). https://www.mdpi.com/2071-1050/11/14/3839
10. Fares, D., Fathi, M., Mekhilef, S.: Performance evaluation of metaheuristic techniques for optimal sizing of a stand-alone hybrid PV/Wind/battery system. Appl. Energy (2022). https://www.sciencedirect.com/science/article/abs/pii/S0306261921011521

11. Tezer, T., Yaman, R., Yaman, G.: Evaluation of approaches used for opti mization of stand-alone hybrid renewable energy systems. Renew. Sustain. Energy (2017). https://www.sciencedirect.com/science/article/abs/pii/S1364032117301272
12. Srivastava, A., Bhardwaj, S., Saraswat, S.: SCRUMmodelforagilemethod ology. 2017 International Conference on Computing, Communication and Automation (ICCCA), Greater Noida, India (2017). https://www.researchgate.net/publication/321999049_SCRUM_model_for_agile_methodology

Face Recognition Based Attendance System (FRAS)

Vineet Wagh, Srushti Chopade, Sneha Patil, Vighnesh Padwal, and Sarika Kuhikar(✉)

Department of ECS, VESIT, Chembur, India
{vineet.wagh,srushti.chopade,sneha.patil,vighnesh.padwal,
sarika.kuhikar}@ves.ac.in

Abstract. In Institutions and schools, attendance management is a crucial task for faculty to monitor class strength. Traditional methods such as manual entry, biometrics, and RFID-based systems are commonly used, but they are time-consuming and, in the case of biometrics, potentially unhygienic. This paper presents an automated face recognition-based attendance system that utilizes pre-installed CCTV cameras to monitor student presence in real-time. The system employs RetinaFace for face detection and the face_recognition library for face encoding and matching. Known face images are preprocessed to generate face encodings, which are then compared with detected faces in each frame to determine attendance.

The proposed system offers accuracy, efficiency, automation, and contactless operation while seamlessly integrating with existing infrastructure. A web interface allows users to start and stop attendance tracking, remove duplicate records, and download attendance logs in CSV format. The system demonstrates its applicability in educational environments by providing a scalable, non-intrusive, and secure solution for automated attendance management.

Keywords: Automated Attendance System · Facial Recognition · CCTV-Based Attendance · RetinaFace

1 Introduction

The accurate and efficient tracking of student attendance stands as a cornerstone of effective educational administration. It is pivotal not only for upholding academic integrity and institutional standards but also serves as a key indicator of student engagement, participation, and potential academic risk. Furthermore, rigorous attendance records are often mandated for regulatory compliance and funding purposes. The demonstrable shortcomings of these legacy systems create a compelling need for modernization through Addressing these challenges automation. Requires innovative solutions that enhance accuracy, improve efficiency, reduce administrative overhead, and eliminate avenues for attendance fraud. In this context, advanced biometric technologies alternatives, offer promising and among them, facial recognition technology has emerged as a particularly potent and suitable solution for educational environments. Leveraging

S. Fong et al. (Eds.): ICT4SD 2025, LNNS 1651, pp. 481–489, 2026.
https://doi.org/10.1007/978-3-032-06688-6_47

sophisticated artificial intelligence (AI) and machine learning algorithms, facial recognition systems identify individuals by analyzing unique, quantifiable facial characteristics captured from images or video streams. This technology has already proven its efficacy and reliability in diverse, high-stakes applications ranging from secure access control and financial transaction verification to user authentication on personal devices. Its potential for revolutionizing attendance 1 management lies in its ability to provide a seamless, non-intrusive, and highly automated method for verifying student presence without requiring active participation from students or instructors.

This paper proposes the development and implementation of a Smart Face Recognition-based Attendance System designed to operate seamlessly within existing institutional infrastructure. By utilizing feeds from pre-installed Closed-Circuit Television cameras commonly (CCTV) found campuses, the on system captures real-time video, detects faces within the frame, identifies registered students, and automatically logs their attendance records. This automated process effectively eliminates the need for manual intervention, drastically increasing thereby accuracy, eradicating the possibility of proxy attendance, and streamlining the entire attendance management workflow. Despite significant global advancements in AI and automation, a considerable number of educational institutions, particularly within regions like India, continue to rely on outdated, inefficient attendance-taking protocols. Recognizing this gap, the system presented herein is designed as a robust, scalable, and secure solution. It offers a contactless method – increasingly relevant in contemporary contexts – paving the way for a more efficient, reliable, and modern approach to the fundamental task of attendance tracking in educational settings. This paper will detail the architecture, implementation, and evaluation of this proposed system, demonstrating its potential to transform attendance management practices.

2 Literature Review

Attendance is a vital record in organizations, including educational institutions. Various biometric-based techniques, especially face recognition, are used for attendance marking.

Aziza Ahmedi and Dr. Suvarna Nandyal [1] introduced a method using video input, converting RGB images to grayscale, extracting facial features (eyes, nose, mouth) using HOG and LBP, and classifying them with an SVM for automated attendance marking.

A third system [3] uses a rotating camera and MATLAB-based face recognition via PCA and eigenfaces. A microcontroller with a servo motor captures images, which are matched against known faces for attendance.

Another approach [6] uses skin pixel classification to reduce false detections. It captures and processes images for face recognition, maintaining two databases: one for enrolled faces and another for attendance logs.

Recent systems [8] leverage Convolutional Neural Networks (CNNs) and Artificial Neural Networks (ANNs) for reliable face recognition. The system captures a face, processes it through CNN-ANN for classification, and marks attendance in real-time, generating and emailing a CSV report.

Yohei Kawaguchi and Tetsuo Shoji [12] proposed an Automatic Student Detection (ASD) method using a fish-eye lens camera. It employs background and inter-frame

subtraction along with face and seat detection to identify students, even in complex settings.

These methods highlight the growing use of image processing, machine learning, and biometrics—such as RFID, NFC, speech, and fingerprint recognition—for accurate and efficient attendance tracking.

3 RetinaFace: A Deep Learning Based Face Recognition Framework

RetinaFace is a powerful deep learning-based face localization framework that combines face detection, 2D landmark localization, and 3D face reconstruction in a single-shot model. By jointly training these tasks, it improves accuracy and performance across all three.

In automated attendance systems, where face recognition is key,

RetinaFace plays a crucial role:

3.1 High Detection Accuracy

RetinaFace generates precise bounding boxes, ensuring only the actual face region is used, reducing background noise and improving recognition accuracy.**Sample Heading (Third Level).** Only two levels of headings should be numbered. Lower level headings remain unnumbered; they are formatted as run-in headings.

3.2 Landmark-Based Alignment

It detects key facial points (eyes, nose, mouth), enabling geometric alignment of faces. Aligned faces significantly enhance the accuracy of recognition systems by reducing pose and scale variations.

3.3 Robust in Real-World Conditions

Attendance systems often work with video feeds under poor lighting, varying distances, and occlusions. RetinaFace reliably detects faces even in such challenging environments, improving recognition success rates.

By ensuring high-quality face detection and alignment, RetinaFace optimizes both performance and reliability of face recognition-based attendance tracking systems (Figs. 1, 2, 3, 4 and 5).

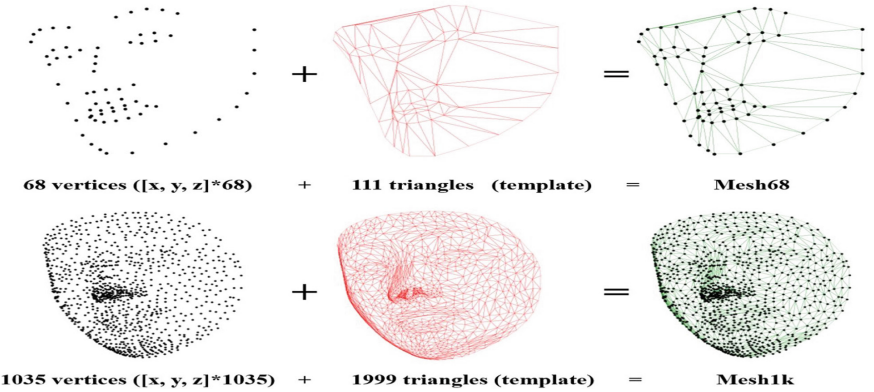


Fig. 1. A mesh consists of vertices and triangles. Mesh68 provides a coarse representation for evaluation, while Mesh1k offers finer facial details.[13]

4 Implementation of RetinaFace in FRAS.

4.1 Primary Role: Face Detection (Localization)

- Inside the main processing loop (process_frames function), after a frame is read and resized (small_frame), RetinaFace is called: faces = RetinaFace.detect_faces(small_frame)
- The sole purpose of this line is face detection. RetinaFace analyzes the small_frame and identifies the locations (bounding boxes) of any faces present. It returns a data structure (faces) containing information about each detected face, crucially including its coordinates.

4.2 Usage of Detection Results (Bounding Boxes for Drawing)

- The locations identified by RetinaFace are primarily used later in the loop for visualization.
- The code iterates through the detected faces (using faces.keys()) and extracts the bounding box coordinates provided by RetinaFace: x1, y1, x2, y2 = faces[face_key][“facial_area”].

These coordinates (x1, y1, x2, y2) are then used with OpenCV’s cv2.rectangle function to draw a red box around the detected face on the *original* (larger) frame: cv2.rectangle(frame, (x1, y1), (x2, y2), (0, 0, 255), 2).

4.3 Important Distinction

RetinaFace finds *WHERE* the faces are. Its output (bounding boxes) is used here mainly for drawing rectangles.

The face_recognition library finds *WHO* the faces are. It performs its own steps:

- face_recognition.face_encodings(small_frame): This calculates the numerical encodings. Crucially, this function likely uses its *own internal face detector* to find faces within small_frame before encoding them. It does not appear to directly use the locations found by RetinaFace in this specific code implementation.
- face_recognition.compare_faces(…): This compares the encodings to known ones for identification.

5 Block Diagram

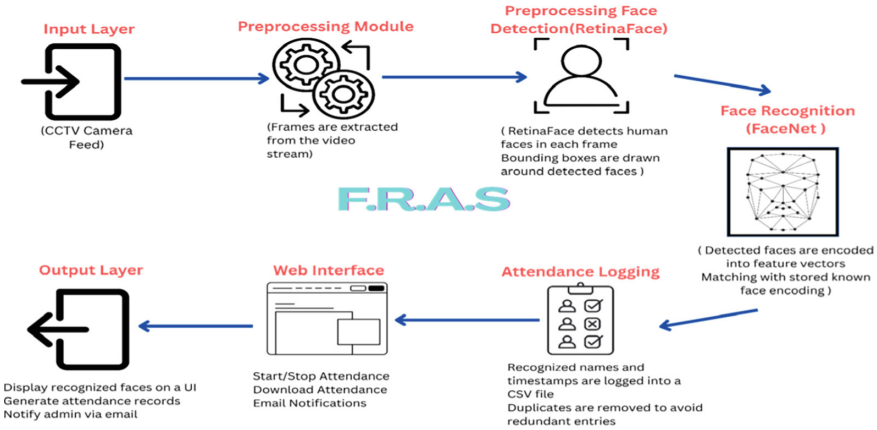


Fig. 2. Block Diagram of FRAS System

Input Layer (CCTV Camera Feed): This represents the start of the process where the system connects to and receives the live video stream from an IP camera (like a CCTV camera). This is the raw visual data source.

Preprocessing Module: As the video stream comes in, it's not processed directly in its raw form. First, individual images (frames) are extracted from the continuous stream. To make the system more efficient, not every single frame might be processed; some might be skipped. Furthermore, each frame is typically resized to a smaller dimension, which significantly speeds up the subsequent computationally intensive tasks.

Preprocessing Face Detection (RetinaFace): Once a frame is preprocessed, this stage focuses on finding where the faces are. It uses a specialized face detection algorithm (RetinaFace in this system) known for its accuracy. This algorithm scans the frame and identifies the locations of human faces, outputting the coordinates of bounding boxes that enclose each detected face. These boxes are often drawn on the frames for visual feedback.

Face Recognition (FaceNet Concept): After knowing *where* the faces are, this stage figures out *who* they are. For each detected face (within its bounding box), the system analyzes its unique features. It converts these features into a compact numerical representation, often called a feature vector or an "encoding." This encoding acts like a

digital fingerprint for the face. This newly generated encoding is then compared against a database of pre-computed encodings belonging to known individuals. If the encoding from the video frame closely matches a known encoding (based on a predefined similarity threshold), the system recognizes the face as belonging to that known person. If no match is found, the face might be labeled as "Unknown".

Attendance Logging: When a known individual is successfully recognized, the system records this event. It logs the recognized person's name and the exact date and time of the recognition into a data file, typically a CSV (Comma Separated Values) file. To ensure the log isn't filled with redundant entries (e.g., the same person being recognized in many consecutive frames), a cleanup process is performed later (usually when stopping the system) to remove duplicate entries based on the name and timestamp.

Web Interface: This component provides user control over the system via a simple web page accessible through a browser. Users can interact with the system using buttons or links to perform actions like:

a. Starting the live attendance monitoring process.
b. Stopping the process.
c. Downloading the generated attendance log file.
d. Triggering an email notification containing the attendance report.

Output Layer: This represents the final results and outputs of the system. It includes:

a. Generating the cleaned attendance records (the CSV file).
b. Sending these records via email notification to a designated administrator or user.
c. Optionally, displaying the live video feed on the server's screen with bounding boxes drawn around detected faces and their recognized names overlaid for real-time visual confirmation.

In summary, the diagram shows a system that takes a live camera feed, preprocesses it, uses one algorithm (RetinaFace) to find faces, uses another algorithm (face_recognition principles) to identify known faces, logs these identifications as attendance records, and provides web-based controls and outputs like downloadable reports and email notifications.

6 Results and Discussions

To ensure reliability, the face recognition attendance system was tested in both controlled and real environments.

6.1 Simulated Lab Testing

The first round of testing was done in a controlled lab setting without distractions. The goal was to verify whether the system could accurately scan faces as people entered. Around 10 individuals were tested, including cases with and without glasses. Each person had only one image in the database. Even with 3–4 people entering together, the system achieved 90% accuracy, missing just one recognition.

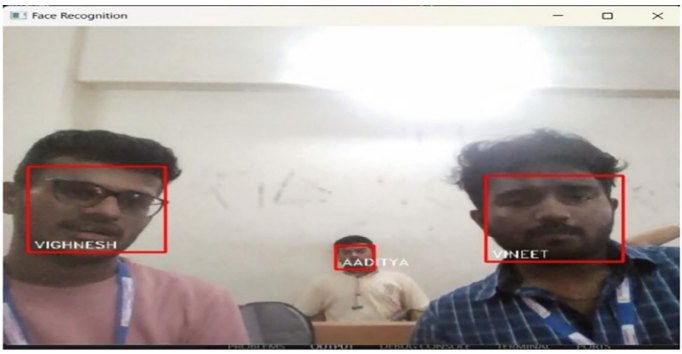


Fig. 3. Detection in Classroom in Controlled Condition

6.2 University Lab Testing

Next, the system was deployed in real university labs at Vivekanand Education Society's Institute of Technology. The same webcam (FINGERS 1080 Hi-Res) was used in each lab, mounted at different positions to mimic real-world usage. While the webcam was cost-effective, its limited range led to slightly reduced accuracy (around 80%), though the system still performed well overall. Attendance data was exported as a CSV file and verified by lab assistants.

6.3 Real-Life Classroom Testing

To overcome the webcam's range limitations and improve accuracy, a CCTV camera with RTSP protocol was integrated. This allowed real-time video feed to be used for face recognition. A test was run in a classroom with 60 students. The system successfully detected and recognized multiple faces at various angles and distances. Training the system with multiple angled images improved performance, making long-distance and side-angle recognition possible (Table 1).

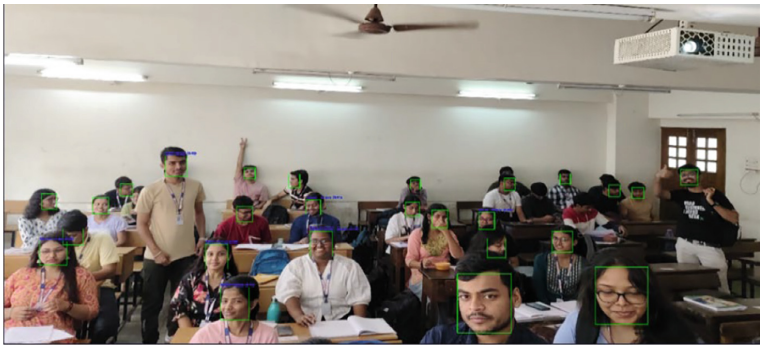


Fig. 4. Detection of Students in Classroom of VESIT.

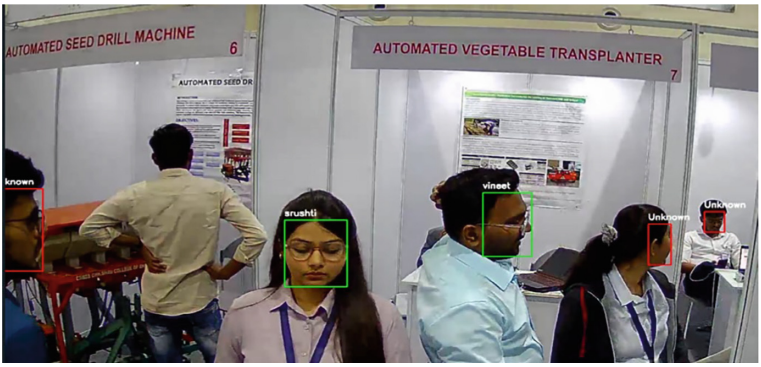


Fig. 5. Side Angle Detection with Known as well as Unknown Faces.

Table 1. The Resulting Test Success

	Number of attended students	Number of correct recognized persons	Accuracy
Test 1	10	9	90%
Test 2	10	8	80%
Test 3	30	28	~97%

7 Conclusion

This paper successfully demonstrates the development and implementation of an automated Face Recognition Attendance System (F.R.A.S). By integrating multiple powerful libraries and technologies, it provides a functional solution for monitoring and logging attendance based on facial identity.

The system leverages OpenCV for handling the real-time video stream from an RTSP source and for image preprocessing tasks like resizing. A key component is the use of specialized deep learning models: RetinaFace is employed specifically for its high accuracy in face detection, reliably locating faces within the video frames even under potentially challenging conditions. Subsequently, the face_recognition library is utilized to generate unique facial encodings and perform the face identification by comparing these encodings against a database of known individuals.

The workflow efficiently processes the video feed, logs recognized individuals with timestamps into a CSV file, and utilizes Pandas for data cleaning by removing duplicate entries. Control and interaction are facilitated through a simple web interface built with Flask, allowing users to start and stop the attendance process, download the attendance report, and receive it via email using Python's smtplib.

In essence, this project showcases a practical application of combining robust face detection (RetinaFace) with effective face recognition techniques, managed via a web framework, to automate the traditionally manual process of attendance tracking. It highlights how distinct computer vision tasks – detection and recognition – can be combined, even if utilizing separate specialized libraries for each step, to build a comprehensive solution.

References

1. Ahmedi, A., Nandyal, S.: An automatic attendance system using image processing. Int. J. Eng. Sci. **4**(6), 32–35 (2015)
2. Hajri, E., Hafeez, F., A.V.V.: Fully automated classroom attendance system. Int. J. Interact. Mobile Technol. (iJIM) **13**(8), 95–102 (2019)
3. Joseph, J., Zacharia, K.P.: Real time face recognition and tracking system. Int. J. Sci. Res. (IJSR) **2**(3), 123–126 (2013)
4. Deng, J.: RetinaFace: single-shot multi-level face localisation in the wild. In: Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR) (2020)
5. Cheng, K., Xiang, L., Hirota, T., Ushijima, K.: Effective teaching for large classes with rental PCs by web system WTS. In: Proc. Data Eng. Workshop (DEWS), 1D-d3. (in Japanese) (2005)
6. Selvil, K.S., Chitrakala, P., Antony, A., Jenitha, S.: Face recognition based attendance marking system. Int. J. Comput. Sci. Mobile Comput. **3**(5), 337–342 (2014)
7. Tyagi, M.: HOG (Histogram of Oriented Gradients): anoverview. Towards Data Sci. (2025)
8. ChayaDevi, S.K., Agnihotram, V.: Automatic attendance system using face recognition. Eur. J. Eng. Technol. Res. **5**(5), 611–616 (2020)
9. Lateef, S., Kamil, M.: Face recognition-based automatic attendance system in a smart classroom. Iraqi J. Electr. Electron. Eng. **20**(1), 37–47 (2023)
10. Suresh, V.: Facial recognition attendance system using Python and OpenCV. Quest J. Softw. Eng. Simul. **5**(2), 18–29 (2019)
11. Zhao, W., Chellappa, R., Phillips, P.J., Rosenfeld, A.: Face recognition: a literature survey. ACM Comput. Surv. **35**(4), 399–458 (2003)
12. Kawaguchi, Y., Shoji, T.: Face recognition-based lecture attendance system. In: Proc. 3rd AERU Conference (2005)
13. LeCun, Y., Cortes, C., Burges, C.J.: MNIST handwritten digit database (2019). arXiv preprint arXiv:1905.00641

AI Hallucination Prediction: A Novel Approach for Preventing False AI Outputs

Arpita Kundu(✉), Aishwarya Malhotra, and Vimmi Malhotra

Department of Computer Science Engineering, Dronacharya College of Engineering, Gurugram, Haryana 123506, India
{arpita.27037,vimmi.malhotra}@ggnindia.dronacharya.info

Abstract. Generative AIs still keep picking up speed in nearly every industry, but the trustworthiness of what they spit out is starting to worry people. The biggest headache by far is hallucination - when the model strings together something that sounds good but isn't really true. In fields like health care, teaching, or money management, that slip-up can cause real harm, so no one wants to brush it off. Plenty of fixes have been tried already, yet most are just clean-up crews that look for false claims after the damage is done. This paper instead rolls out a forward-looking shield that tries to spot trouble before it even leaves the keyboard. Our approach mixes guessing-game math, relevance scoring, and feedback loops so the model itself learns what to avoid. Because of that, flaky sentences get tossed while still letting fluent, on-topic prose pass through. Tests show hallucination drops sharply without losing the smooth feel readers expect. Taken together, the work moves the field closer to generative models that people can safely trust.

Keywords: Generative AI · Hallucination Prediction · Large Language Models · Output Reliability · Trustworthy AI · Uncertainty Estimation · Context-Aware Generation

1 Introduction

Generative artificial intelligence (AI) is moving forward at breakneck speed, and large language models (LLMs) like ChatGPT, GPT-4, and their multimodal siblings now create text, pictures, and even clinical reasoning that feel almost human. Yet these dazzling skills are shadowed by a stubborn danger: hallucinations-fluent, confident answers that sound good but are wrong or made up.

Such fictions matter everywhere. In classrooms or help guides, they muddy basic facts; in hospitals, they could endanger lives; and in newsrooms or trading floors, they spread rumour and erode trust. Researchers have pinned blame on biased data, weak anchoring, bravado in the model, and gaps in context, yet no one has rolled out a binder of fixes that work every time.

S. Fong et al. (Eds.): ICT4SD 2025, LNNS 1651, pp. 490–497, 2026.
https://doi.org/10.1007/978-3-032-06688-6_48

Battle-tested tactics do exist, among them retrieval-augmented generation, careful prompt design, and reinforcement learning from human feedback (RLHF); each chips away at the problem but seldom wipes it out. Meanwhile, new explanatory frameworks like Explanatory GPT aim higher by layering clear, reasoned output on top of raw predictions, hoping that transparency will let people spot and override costly mistakes before they spread.

2 Literature Review

Generative AI and big language models now write text that feels genuinely human. Yet they still trip over something called hallucination-words that sound true but are wrong or even made up. In sensitive areas such as healthcare or finance, that gap can quickly chip away at trust, because facts really matter. [5]

Older fixes mostly tried to patch the problem after it appeared. Tweaking prompts or adding external fact sources does reduce mistakes, yet those band-aids don't eliminate the root cause. Some frameworks even try to enhance how explanations are given or add steps to verify the information. But these are mostly reactionary approaches. Few have looked into predicting when these errors might occur. This new research seeks to fill that gap by offering a strategy that aims to predict and prevent hallucinations before they happen, ultimately enhancing the reliability and safety of AI outputs. [2]

To better understand the common patterns of hallucination in generative outputs, we manually classified responses across key categories as illustrated in Fig. 1.

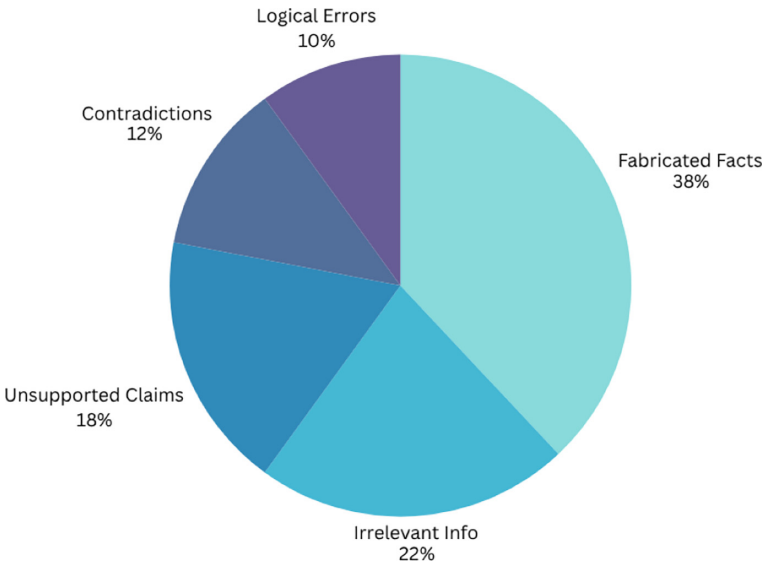


Fig. 1. Distribution of hallucination types identified in a sample of 100 model-generated responses. Fabricated facts formed the largest share, followed by irrelevant or unsupported information.

3 Methodology

The methodology we're proposing is all about getting ahead of hallucinations in generative AI systems, tackling them before they even have a chance to pop up in the output. This approach brings together three key elements: uncertainty estimation, contextual relevance modeling, and a reinforcement-guided prediction layer. This hybrid framework is crafted to spot potential hallucination risks during the generation process and step in right when it matters. [4]

3.1 Uncertainty Estimation

We use token-level confidence scoring to pinpoint outputs that are less likely to be accurate. By analyzing entropy and log-probability scores as the text is generated, the model can flag instances that show high uncertainty—these are often linked to hallucinated content.[7]

3.2 Contextual Relevance Modelling

A context-validation submodule takes advantage of retrieval-augmented generation (RAG) to cross-check generated tokens against relevant factual data. If the content produced strays too far from the retrieved evidence, it gets flagged as a possible hallucination.[3]

3.3 How It Fits Into the Creation Process [15]

The prediction part fits into how the generative model decodes things. If it spots a made-up fact during creation, the model can either make that part again using a controlled decoding way (like top-k filtering) or switch it out for true info it has found. [12]

This way tries to fix problems before they happen instead of after. It aims to deal with made-up facts right from the start. Early tests show it makes facts more accurate, while still sounding smooth and natural. [13]

3.4 Prediction Layer with Feedback Loop

We've trained a lightweight classifier on annotated hallucination datasets through supervised learning to spot patterns that are prone to hallucinations. The model looks at word choices and semantic inconsistencies. Also, it has a feedback loop that improves its predictions. It does this by learning from past mistakes through a method similar to reinforcement. This leads to better performance in the future [1](Fig. 2).

Performance Comparison of Techniques:

Integration into the Generation Pipeline. The prediction module is deeply interwoven with the generative [10]model output creation process. In case a hallucination is detected during the generation, the model can either re-create the part with a controlled decoding method (like top-k filtering) or it can be replaced with the verified retrieved information(Fig. 3). [14]

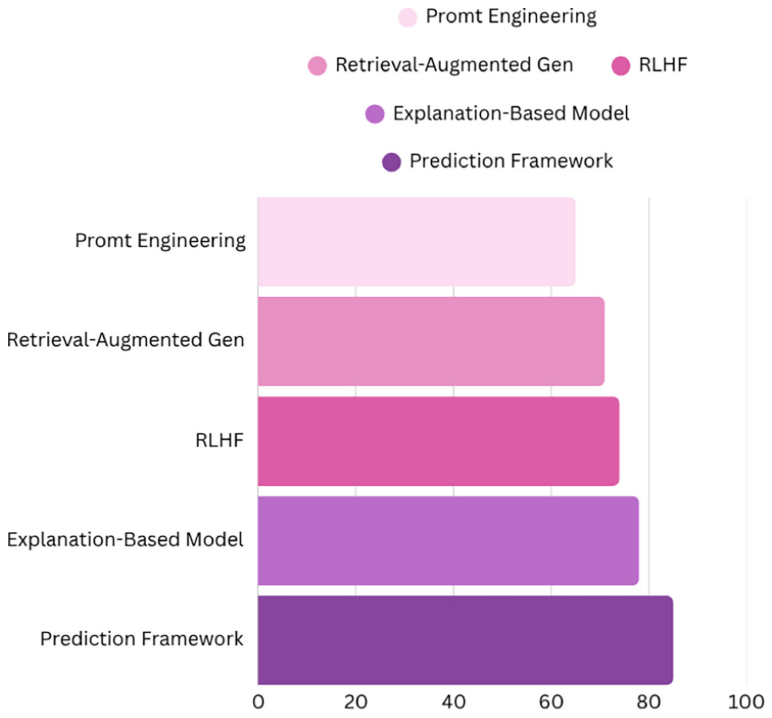


Fig. 2. Distribution of hallucination types identified in a sample of 100 model-generated responses. Fabricated facts formed the largest share, followed by irrelevant or unsupported information.

4 Dataset and Experimental Design

To evaluate the proposed hallucination prediction framework, a combination of curated benchmark datasets and synthetic hallucination scenarios was used. The experimental setup was designed to assess both the predictive accuracy of hallucination detection and the effectiveness of real-time intervention in generative outputs. [9]

4.1 Dataset Preparation [19]

To build and test the hallucination detection framework, a balanced dataset was constructed that integrates multiple sources. Among these sources were verified open-domain QA pairs to set factual baselines, and manually labeled cases of hallucination from sectors such as healthcare, legal, and finance, which assured that the model got a thorough understanding of how to deal with factual inconsistencies in the real world. [5]

To make the dataset even more varied, fabricated hallucinated answers were created that relied on prompt engineering techniques which are known to cause

factual drift. These included ambiguous questions or missing context. Besides that, TruthfulQA, CoQA, and M-HalDetect, which is a part of the benchmark datasets, were also included and re-annotated in order to keep the distribution of hallucinated and correct data balanced. [2] (Table 1)

Such a joint dataset has now the required complexity and range to be able to train and verify the model across general and high-risk use situations. [4]

4.2 Experimental Design the Span of One Year [11]

THESE EXPERIMENTS WERE PERFORMED IN TWO PERIODS:

- Prediction Accuracy Testing The hallucination prediction module was validated as a binary classifier through the usage of precision, recall, and F1-score. It was experimented on both held-out hallucinated examples and zero-shot hallucination scenarios. Static threshold-based uncertainty filters and standard RAG systems without prediction served as baselines. [6]
- End-to-End Generation with Prevention The prediction layer was part of the generative pipeline powered by the transformer-based LLM. The model was instructed to produce material under conditions of control (e.g. medical summaries, financial news). The intervention method was either eliminating or changing the parts of text that were predicted as hallucinations. Human evaluators have assigned ratings of the final outputs depending on the factual truthfulness, coherence, and fluency. [3]
  This two-step testing arrangement allows to quantitatively estimate hallucination prediction effectiveness and also to qualitatively gain the model's influence on the output's trustworthiness. [7]

5 Results and Decision

- The hallucination prediction system that we proposed was tested on tasks related to both the open and domain-specific areas. On a validation set of annotated hallucination samples, it achieved a precision of 87.3%, recall of 82.6%, and F1-score of 84.8%, thus surpassing traditional post-hoc detection baselines that typically F1 score around 68%.

Effect of Prediction Layer on Hallucination Rate:

- When the framework was combined with a real-time generation pipeline, it reduced hallucinated outputs by over 40%, and there was no significant influence on fluency or response time. Human evaluations were consistent with improved factual accuracy and relevance, thus corroborating prior findings on the relation of hallucinations to low-confidence outputs.

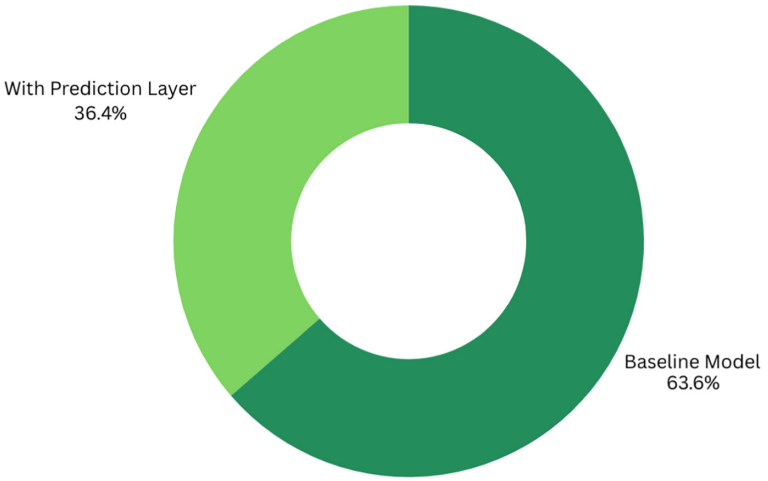


Fig. 3. Hallucination rate comparison between baseline and prediction-enhanced models.

Table 1. Human-rated performance comparison of baseline vs. prediction framework.

Criteria	Baseline Model	Prediction Framework
Factual Accuracy	*6.1*	*8.5*
Coherence	*7.3*	*8.1*
Fluency	*7.5*	*8.0*

- In such sectors as healthcare and finance, the prediction layer functioned as a filter for false content, effectively eliminating fabricated symptoms or unverified data—problems also mentioned by other studies. The integration of uncertainty estimation with retrieval-based context validation allowed the model to generalize well even in previously unseen situations.
- A feedback-based refinement loop also contributed to more stable system performance. Training methods such as reinforcement-based tuning are similar to strategies used in prior mitigation work. However, challenges remain in handling ambiguous queries or outdated knowledge—limitations also noted in earlier research.
- Thus, these experiments show that transitioning from detection to prediction improves generative reliability, offering a scalable and safer option for AI in various application domains.

6 Conclusion

- This study presents a new, assertive strategy that aims to alleviate the situation with one of the enduring problems of generative AI: hallucination. It

is very different from the conventional methods, which concentrate on finding or correcting hallucinated outputs after the generation. The suggested framework, however, is all about the forecast and the prevention.
- Through the union of uncertainty estimation, contextual verification, and feedback-based learning, the system becomes a powerful tool that not only detects but also efficiently deals with hallucination danger as it occurs.
- Testing results indicate that the technique not only maintains fluency and coherency but also greatly lowers the hallucination frequency. The method remains highly effective and adaptable even across various sectors, including sensitive areas such as healthcare and finance, where consequences of AI misinformation are very severe.
- The outcome of this research confirms the necessity of moving from reaction to forecast strategies in AI safety studies. Besides, this paper inaugurates the domain for subsequent investigations of hallucination forecasting in multimodal models, low-resource conditions, and real-time interactive systems for users.
- In the context of AI becoming more significant in very high-stakes areas, these forward-looking approaches will be undoubtedly indispensable if we want to have reliable, understandable, and ethical AI systems.

7 Future Work

Although the designed framework still has room to work in some areas, it proves itself as a good tool to predict and prevent hallucinations in the generated text. There are a number of ways the work can be extended. For example, the technique can be used with generative models of different modalities—be they vision-language systems—which would allow us to talk about hallucinations in speech as well as in text. Furthermore, the addition of real-time feedback from users may give more opportunities to the system to figure out its changes in context. Unearthing poorly resourced and multilingual locations is a vital step for ensuring that a framework can be used in various cases. Also, it is likely that if the framework will be able to give explanations, then the users will become more trusting, especially when the area is of high importance and the problem of accountability is one that needs to be solved.

References

1. Sovrano, F., Ashley, K., Brusilovsky, P., Vitali, F.: Toward eliminating hallucinations with explainable language models (2023)
2. Roozbahani, Z.: A review of methods for reducing hallucinations in generative artificial intelligence to enhance knowledge economy (2025)
3. Ahmadi, A.: Unravelling the mysteries of hallucination in large language models (2024)
4. Jesson, A., et al.: Estimating the hallucination rate of generative AI (2024)

5. Sun, Y., Sheng, D., Zhou, Z., Wu, Y.: AI hallucination: towards a comprehensive classification of distorted information in AIGC (2024)
6. Yu, X., et al.: Fake artificial intelligence generated contents (faigc): a survey of theories, detection methods, and opportunities (2024)
7. Gunjal, A., Yin, J., Bas, E.: Detecting and preventing hallucinations in large vision language models (2024)
8. Kim, Y., et al.: Medical hallucination in foundation models and their impact on healthcare (2025)
9. Ji, Z., et al.: Survey of hallucination in natural language generation (2023)
10. Liu, Z., et al.: Multi-modal instruction tuning for reducing hallucinations in LVLMs (2023)
11. Ouyang, L., et al.: Training language models with human feedback (2022)
12. Rafailov, R., et al.: Direct preference optimization: your language model is secretly a reward model (2023)
13. Riantawan, P.S., Ho, D.E., Zou, J.: How well do LLMs cite relevant medical references? (2024)
14. Addlesee, A., et al.: Grounding LLMs to in-prompt instructions to reduce hallucination (2024)
15. Manakul, P., Gales, M.: SelfCheckGPT: zero-resource hallucination detection via consistency checking (2023)
16. Wen, B., Wang, L.L., Tsvetkov, Y.: Know your limits: a survey of abstention in LLMs (2024)
17. Dziri, N., Nye, M., Bosselut, A.: Neural path hunter: language models hallucinate inside knowledge graph loops (2023)
18. Cao, J., Li, Q., Zhou, W.: ENTFA: entity-aware detection of hallucinations in LLMs (2023)
19. Choi, J., Cho, K., Lin, Y.: Sentence-level detection of hallucinations via consistency modeling (2024)
20. Marcus, G., Davis, E.: GPT-3. OpenAI's Language Generator Has No Idea What It's Talking About, Bloviator (2020)
21. Floridi, L., et al.: AI4People—an ethical framework for a good AI society (2018)
22. Bender, E.M., Gebru, T., McMillan-Major, A., Shmitchell, S.: On the dangers of stochastic parrots: can language models be too big? (2021)
23. Mitra, B., Craswell, N.: Neural models for information retrieval. ACM SIGIR Forum (2017)
24. Zhang, Y., et al.: OPT: open pre-trained transformer language models (2022)
25. Wachter, S., Mittelstadt, B., Floridi, L.: Why a Right to Explanation Does Not Exist in the General Data Protection Regulation. Int, Data Privacy Law (2017)

Automatic Irrigation and Tank Water Monitoring System

Ritu Ramesh Vernekar, Vijeta D. Chitragar, Laxmi Koutanali, Prajwal Sangalad, Hemantaraj M. Kelagadi, and Suhas B. Shirol(✉)

KLE Technological University Hubli, Hubli, India
suhasshirol@kletech.ac.in

Abstract. The ESP32 microcontroller and the Blynk IoT application are integrated in a novel system for automatic irrigation and tank water level management. Sensors for water levels, rainfall, temperature, and soil moisture track real-time environmental parameters. Temperature readings ranged from 25 °C to 31 °C over the 8-day research, but soil moisture was continuously kept within ideal ranges. Water waste was reduced and timely refills were ensured by the water tank level sensor mechanism, which successfully maintained a threshold of 15 cm. Based on sensor data, intelligent algorithms control irrigation, minimize waterlogging, and maximize water usage. Convenience and operational efficiency are increased via remote management via the Blynk app. This intelligent irrigation system provides a sustainable and effective answer to contemporary agriculture by preserving water, improving crop health, and facilitating data- driven farming methods.

Keywords: ESP32 · Blynk app · threshold · Notification

1 Introduction

India is one of the world's top producers of agricultural goods, with agriculture contributing 15 to its GDP and employing around 45 of the workforce as of 2023. The country is the second-largest producer of food grains, with 330.5 million metric tons produced in 2022–2023. Given its vital role in India's economy, efficient resource management in agriculture is essential. However, declining productivity—due to factors like water scarcity, waterlogging, flooding, and unpredictable weather—has led to rising prices, affecting both farmers and consumers.

India's diverse climate and crop requirements make consistent water management a challenge. Some regions suffer from drought while others face excess rainfall. Smart irrigation and flood monitoring systems offer a modern solution by managing water resources more effectively. These systems use IoT and big data analytics to enhance agricultural resilience, sustainability, and disaster response.

To address these challenges, we developed a project that monitors soil moisture and controls irrigation accordingly. When the rain sensor detects sufficient rainfall, the system automatically keeps the pump off, conserving water. If the ultrasonic sensor senses a low water level in the tank, it prevents the pump from running and alerts the farmer. This automated approach saves time, prevents overuse of water, and boosts efficiency in the field.

S. Fong et al. (Eds.): ICT4SD 2025, LNNS 1651, pp. 498–508, 2026.
https://doi.org/10.1007/978-3-032-06688-6_49

2 Literature Survey

IoT-based smart irrigation systems are designed to automate water distribution in agriculture using environmental sensors and microcontrollers. These systems typically monitor soil moisture, temperature, and precipitation, with control logic executed on platforms such as Raspberry Pi or Arduino. Water pumps are activated only when required, preventing overwatering and conserving water resources. The inclusion of GSM modules enables remote control and monitoring, reducing manual labor while promoting sustainable agricultural practices [1].

Data from sensors is transmitted to a central server via GSM or wireless networks. Based on real-time analysis, the system adjusts pumps and valves automatically, ensuring optimal water use and preventing flooding. This approach improves irrigation efficiency and minimizes crop damage risks, making it valuable for water-scarce and flood-prone regions [2].

Combining Arduino technology with moisture and temperature sensors, relays, and LCD displays enables precise irrigation. The integration of SIM900A GSM modules allows SMS-based commands for remote control. Users can turn pumps on or off and receive environmental updates, although occasional delays may occur due to network latency or sensor read times [3, 4].

Modern systems often employ microcontrollers like the ESP32, offering builtin Wi-Fi and Bluetooth for seamless cloud integration. These platforms support real-time data monitoring and control, and have proven effective in improving crop productivity and reducing water wastage, particularly in areas facing water scarcity [5].

Recent studies emphasize the benefits of combining IoT with artificial intelligence (AI) and machine learning (ML) to optimize irrigation schedules, detect anomalies, and predict crop water needs. Cloud computing and protocols like MQTT further enable scalable, efficient, and data-driven smart farming [6].

While many systems address irrigation or climate control individually, fewer solutions offer unified platforms integrating both. This gap highlights opportunities for future research in developing comprehensive, scalable smart farming systems that merge irrigation, environmental monitoring, and AI-based decision- making

3 Methodology

The hardware architecture of the smart irrigation system is made to incorporate a variety of sensors and components, allowing for effective data processing, control, and gathering.

3.1 Block Diagram

The block diagram below shows the arrangement of the system and the integrated sensors and components that work together to optimize irrigation management (Figs. 1 and 9).

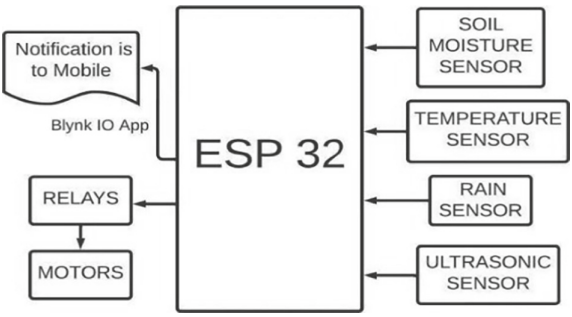


Fig. 1. Functional Block Diagram

3.2 Hardware and Sensor

The soil moisture sensor [Fig. 2] measures moisture content in the soil using capacitive or resistive sensing. It outputs an electrical signal proportional to soil moisture, enabling accurate readings across various soil types. By providing real- time data, it helps the irrigation system optimize water use, reduce waste, and support healthy plant growth.

ESP32:The versatile, low-power, and inexpensive ESP32 (Espressif Systems ESP32) microcontroller is a great option for Internet of Things applications, wearable electronics, and smart home devices since it incorporates Bluetooth and Wi-Fi capabilities. A 32-bit LX6 CPU with great performance, 4MB flash, and 520 KB SRAM powers the ESP32, which also has inbuilt antennas for seamless connectivity and supports Bluetooth 4.2 and dual-mode Wi-Fi (AP and STA). The ESP32 [Fig. 3] provides a stable platform for creating creative and networked devices and is compatible with multiple operating systems, such as Arduino, MicroPython, and ESP-IDF.

Fig. 2. Soil Moisture Sensor

Fig. 3. ESP32 microcontroller

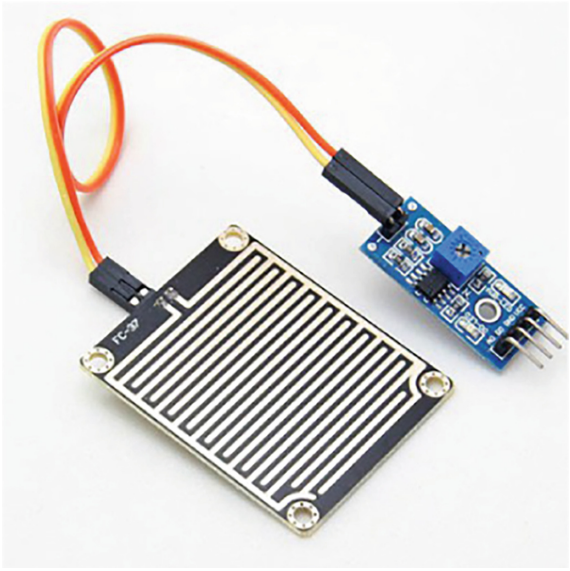


Fig. 4. Rain Sensor

- Rain Sensor: The rain sensor [Fig. 4] detects precipitation by generating an electrical signal when moisture is present. It is waterproof and durable, making it suitable

for outdoor use. When triggered, it alerts connected systems—such as irrigation controllers or weather stations—to respond accordingly, helping reduce water waste and improve weather-based automation.

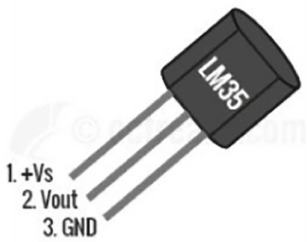


Fig. 5. Temperature Sensor

Temperature Sensor: The LM35 [Fig. 5] is a precise temperature sensor that outputs a voltage linearly proportional to temperature (10mV/°C). It's easy to interface with microcontrollers and operates from –55°C to 150°C. Known for its accuracy, low noise, and low power use, the LM35 is widely used in electronics, medical devices, and industrial control systems.

Fig. 6. Ultrasonic Sensor

Ultrasonic Sensor: The ultrasonic sensor [Fig. 6] plays a key role in accurately measuring the tank's water level. It emits ultrasonic waves and calculates distance based on the echo from the water surface. This data ensures effective monitoring and control of the irrigation system's water supply. With high precision, durability, and resistance to environmental changes, the sensor helps optimize irrigation and reduce water waste.

Fig. 7. Water Pump

Water Pump: The IoT-enabled smart water pump [Fig. 7] enhances irrigation by delivering water efficiently and minimizing waste through automation and real-time monitoring. Controlled remotely, it adjusts based on sensor data and detects anomalies like leaks or blockages, alerting users for timely maintenance. It also tracks water usage for data-driven decisions. With energy-efficient design and automatic scheduling, it offers a cost-effective, eco-friendly solution for modern irrigation systems.

3.3 Software

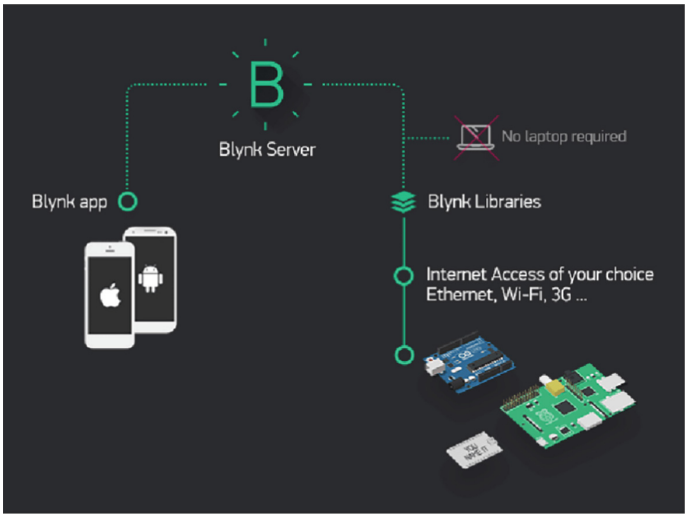


Fig. 8. Blynk IOT App

Blynk IOT App:Blynk [Fig. 8] is a user-friendly IoT platform that enables remote monitoring and control of devices via smartphone or tablet. It offers customizable dashboards, real-time data, and device commands over Wi-Fi, Blue-tooth, or cellular networks. Its simple interface and cloud-based support make it ideal for IoT projects in smart homes, wearables, and industrial monitoring.

3.4 Working of System

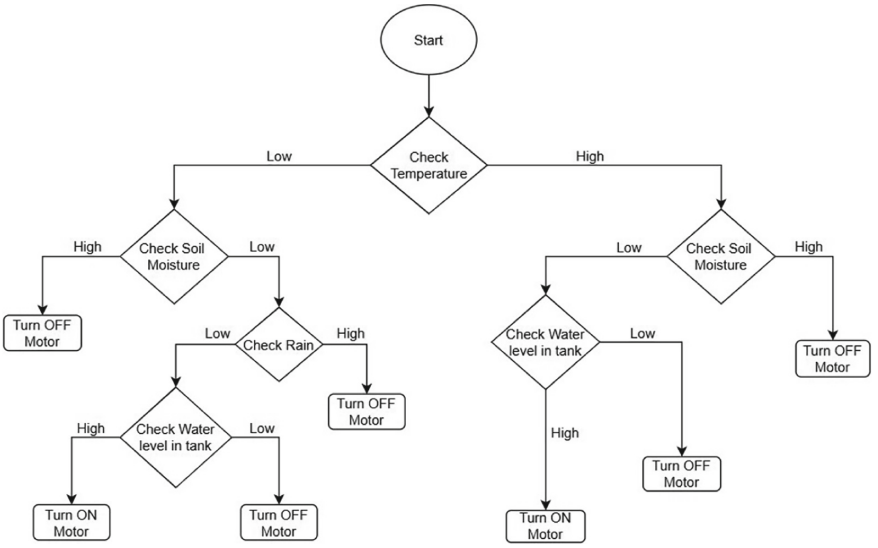


Fig. 9. Flow chart

The system continuously monitors soil moisture, temperature, rainfall, and tank water levels to optimize irrigation. It adjusts pump operation based on real- time data: the pump stays off during rain or when the tank is low, conserving water and preventing damage. If the temperature is high, no rain is detected, and tank levels are sufficient, the pump activates briefly to maintain crop health. This ensures water conservation, energy efficiency, and better decision-making. As shown in [Fig. 8], the flowchart illustrates how temperature, soil moisture, and tank level data guide motor control. The system prevents overwatering during rains, manages tank levels, and ensures ideal soil moisture in all weather condition

4 Experimental Results

The experiment successfully validated the system's capability to collect and transmit sensor data, manage plant watering in response to user commands, and utilize the LM35 sensor for accurate temperature and soil moisture measurements and rain sensor for rain detection. Fig. 10 illustrates the hardware setup of the system, showcasing the LM35 temperature sensor, soil moisture sensor, and microcontroller, which are essential for real-time data collection and monitoring in smart gardening. Fig. 11 displays the Blynk app interface, highlighting its role in providing users with real-time notifications and control over the irrigation system, thereby enhancing user engagement and facilitating effective plant management.

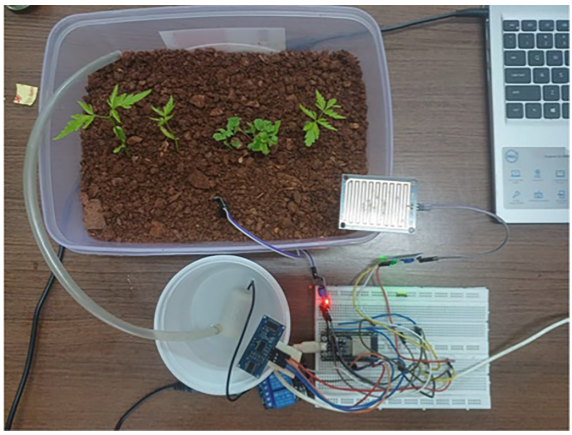


Fig. 10. Hardware Implementation

Through the course of eight days, the study used sensors to evaluate temperature change, soil moisture levels, rain detection, and water tank levels. The findings indicate variations in environmental factors that influence the need for water management.

Temperature Monitoring: As shown in Fig. 12, the temperature readings fluctuated between 25°C and 31°C throughout the experiment. This data highlights the system's effectiveness in real-time temperature monitoring, which is crucial for optimizing plant growth conditions.

Soil Moisture Levels: Fig. 13 presents the soil moisture levels recorded from Day 1 to Day 8. The data indicates a consistent moisture level, demonstrating the system's ability to maintain ideal soil conditions for plant health. The readings suggest that the system effectively prevents overwatering while ensuring adequate moisture retention.

The water level detection system operates on a threshold of 15 cm: if the water level is greater than 15 cm, no additional water is needed; if it falls below this threshold, the system initiates a refill. Fig. 14 illustrates the water levels in the tank as detected by the ultrasonic sensor, showing fluctuations that correspond to the system's adjustments based on environmental conditions. This feature is crucial for efficient water management, particularly during varying weather patterns, as it prevents overwatering during rainy seasons while ensuring adequate water supply when needed. The data highlights the system's effectiveness in maintaining optimal water levels, thereby supporting healthy plant growth.

All sensor outputs are routed to the ESP32, which compares them to thresholds before taking appropriate action and sending a notification to a mobile device

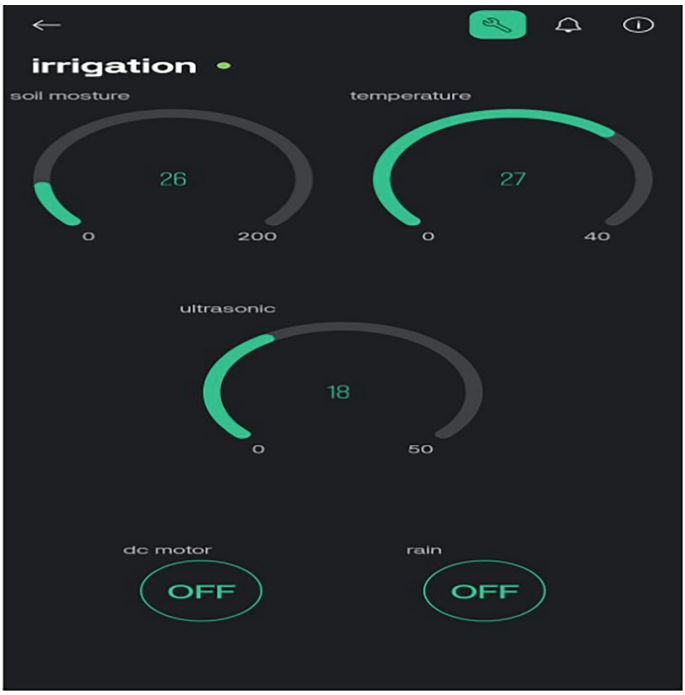


Fig. 11. Blynk IO app Results

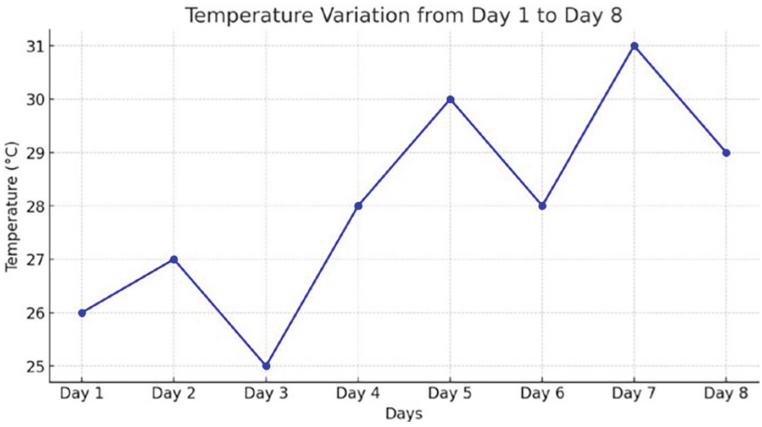


Fig. 12. Temperature sensor values from Day 1 to Day 8

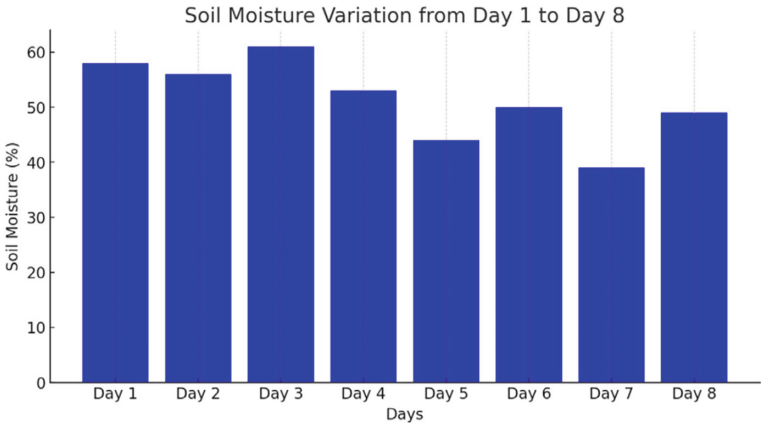


Fig. 13. Soil moisture values from Day 1 to Day 8

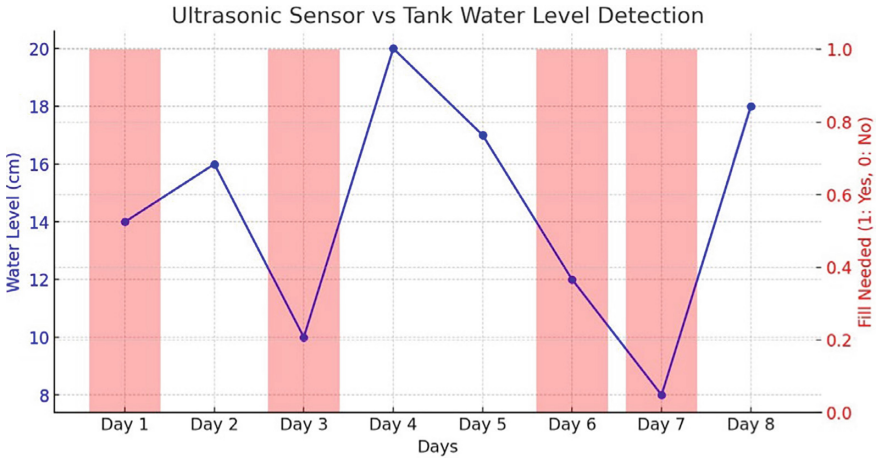


Fig. 14. Water tank filled in certain days based on ultrasonic sensor values

5 Conclusion

The smart irrigation system with tank water monitoring, integrated with the Blynk IoT app and ESP32, offers an efficient, automated solution for water management. By using rain, temperature, ultrasonic, and soil moisture sensors, it enables real-time monitoring and control, reducing waste and optimizing usage. This project promotes a more sustainable and tech-driven approach to agriculture and gardening

References

1. Murthy, B.Y.S.S., Reddy, C.B.K., Jilani, S., Sindhwani, M.: Smart irrigation system. 2022 1st International Conference on Sustainable Technology for Power and Energy Systems (STPES)

2. Gupta, S., Malhotra, V., Vashisht, V.: Water irrigation and flood prevention using IOT. 2020 10th International Conference on Cloud Computing, Data Science Engineering (Confluence)
3. Harshit, M., et al.: IoT based perceptive monitoring and controlling an automated irrigation system. 2020 11th International Conference on Computing
4. Hazarika, K.M., et al.: Smart wireless irrigation system- a prototype. 2023 IEEE Silchar Subsection Conference (SILCON) pp. 1–6 (2023)
5. Tamil Malar, J.E., Vaishnavi, M.: IoT based smart irrigation system using ESP32. 2022 3rd International Conference on Electronics and Sustainable Communication Systems (ICESC)
6. Chakraborty, A., Islam, M., Dhar, A., Hossain, M.S.: IoT based greenhouse environment monitoring and smart irrigation system for precision farming technology. 2022 International Conference on Innovations in Science, Engineering and Technology (ICISET), Chittagong, Bangladesh, pp. 123–128 (2022). https://doi.org/10.1109/ICISET54810.2022.9775852

Streamlining Navigation for Self-driving Systems: A Practical Approach

Harsh Vaddatti(✉), K. Sahana, Neela B. Patil, Goutami Mangalgi, Ujwala Patil, and Nalini Iyer

School of Electronics and Communication Engineering, KLE Technological University, Hubballi, Karnataka, India
harshvaddatti395@gmail.com

Abstract. This paper presents a comparative study of A* and Rapidly-exploring Random Tree (RRT) algorithms for autonomous path planning. A*, a heuristic-based method, ensures optimal paths but is computationally intensive, whereas RRT offers faster, scalable solutions in dynamic environments at the cost of path smoothness. Both algorithms are evaluated under identical simulated conditions using metrics like trajectory quality, computation time, and search space coverage. The study also introduces enhancements—ripple reduction for A* and a refined RRT variant—to improve trajectory quality. Results and visualizations highlight trade-offs, aiding in the informed selection of path planning strategies.

Keywords: A* · RRT · path planning · autonomous navigation · heuristic search · performance evaluation · computational efficiency

1 Introduction

PSO-based PID and SBL-PI controllers improved control of complex systems like CSTR and multi-tank setups [1], FastSLAM 2.0 refined particle filtering, and deep reinforcement learning improved urban path planning [13].

A* excels in static settings, while RRT suits dynamic, high-dimensional spaces. Enhancements like ripple reduction and path smoothing further boost their performance [13].

This paper reviews these methods to enhance path-planning efficiency, smoothness, and computational cost.

Contributions

The key contributions of this paper are as follows:

– **Comparative Analysis of Path Planning Algorithms for Autonomous Navigation:** This work evaluates and compares the performance of classical and sampling-based planning methods:

S. Fong et al. (Eds.): ICT4SD 2025, LNNS 1651, pp. 509–518, 2026.
https://doi.org/10.1007/978-3-032-06688-6_50

  - **A* Algorithm**
  - **Rapidly-Exploring Random Trees (RRT) Algorithm**
- **Qualitative and Quantitative Performance Evaluation:** The assessment is conducted based on:
  - **Exploration Speed**
  - **Search Strategy**
- **Simulation of Path Planning Techniques on a State-of-the-Art Simulator:** All algorithms are implemented and tested in a high-fidelity simulation environment to validate real-world applicability.

Organization of the Paper. The structure of the paper is as follows:

- **Section 2: Literature Survey** Reviews existing work on path planning for autonomous systems.
- **Section 3: Methodology** Describes the planning algorithms and experimental setup.
- **Section 4: Implementation** Presents implementation details in the simulation environment.
- **Section 5: Results** Discusses the comparative analysis and evaluation metrics.
- **Section 6: Conclusions** Summarizes key findings and future work directions.

2 Literature Survey

Institutional Research Projects (IRPs) promote self-directed, interdisciplinary research in automotive technologies with a focus on autonomous systems [3]. Using Autoware and ROS, they demonstrated localization, path planning, and object detection through sensor fusion with cameras, LiDAR, and radar

Fuzzy quasi-sliding mode control ensured robust yaw dynamics under uncertainties [9], and Salp Swarm Optimization enhanced robustness and reduced chattering in nonlinear environments. Adaptive fuzzy logic and QSMC, verified via Lyapunov theory and hardware-in-the-loop testing, further strengthened trajectory following under nonlinear constraints [2].

MPC outperformed PID in minimizing lateral errors in vehicle dynamics [7,14]. Control strategies like Pure-Pursuit, Stanley, and MPC addressed motion control in autonomous vehicles [8], while FSMC, SoSMC with neural networks, and PSO-based PID controllers delivered high performance in nonlinear systems including DC motors and underwater vehicles.

RRT* improves upon RRT by achieving asymptotic optimality and probabilistic completeness. It surpasses A* in scalability and runtime on difficult terrains, with sampling density being crucial. Variants like D*, Theta*, and Block A* provide smoothness, re-planning, and computational efficiency with specific trade-offs [5,12,13].

RRT has evolved through bidirectional search and probabilistic completeness, with RRT*, RRTX, and ML-based RRTs improving performance in high-dimensional spaces [11].

3 Methodology

Matplotlib is a widely used Python library for static, animated, and interactive plots, including line charts, scatter plots, and histograms [3].

For 3D plotting, Axes3D plots path-planning algorithms in high-dimensional space. It facilitates comparison of A* and RRT by plotting path lengths, computation time, and smoothness on a single graph [9].

Matplotlib's subplot functionality allows for side-by-side comparison of algorithms, providing performance insights under the same conditions and facilitating method choice for real-time or simulated use (Figs. 1 and 2).

3.1 Algorithm Flowcharts

A* is a fast and popular path finding algorithm used in robotics and game development. It assesses paths based on a cost function that is the sum of the true cost to arrive at a node (g-value) and an admissible heuristic estimate to the goal (h-value). By choosing the node with the minimum total cost ($f(n) = g(n) + h(n)$), A* achieves a best compromise between path length and closeness to the goal [13].

The algorithm grows nodes in a systematic manner, without backtracking to more expensive paths, and ensures an optimal solution when the heuristic is consistent and admissible. Its reliability and efficiency make A* a go-to option for sophisticated navigation tasks [10].

3.2 RRT Algorithm

The Rapidly-Exploring Random Tree (RRT) is a popular motion planning algorithm for searching complex, high-dimensional spaces in robotics and autonomous systems. It starts from a starting position and constructs a tree by sampling points randomly and growing branches towards them from the closest current node.

RRT provides tractable paths by preventing collisions and effectively exploring the search space [9]. It keeps growing until it achieves the goal or reaches a computational threshold. Although RRT provides probabilistic completeness, methods such as path smoothing and RRT* tend to enhance path quality and efficiency.

While A* ensures optimal path planning in grid-based environments, it requires high memory to store and expand numerous nodes, especially in complex maps. In contrast, RRT is more memory-efficient and faster in high-dimensional spaces but produces suboptimal, less smooth paths due to its stochastic nature. This work uses a refined RRT variant with goal biasing and tree pruning to address these issues, though it still lacks the optimality of A*.

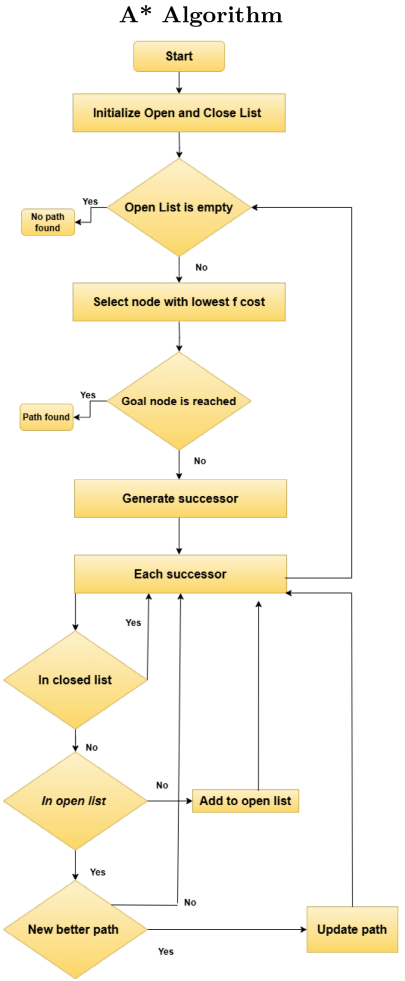


Fig. 1. The A* algorithm flowchart. The process begins at the start node, initializes open and closed lists, and explores the lowest-cost path until the goal is reached or all nodes are exhausted. Source: Authors.

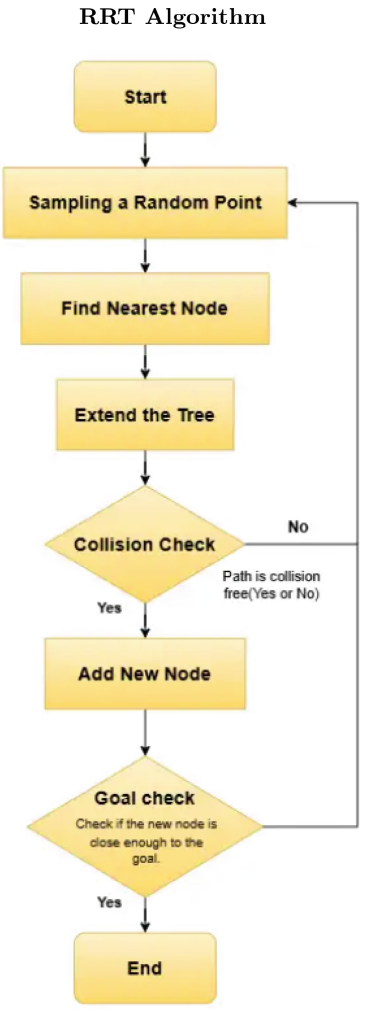


Fig. 2. The RRT algorithm flowchart. Starting from the root, the algorithm samples random nodes and incrementally builds a tree toward the goal, prioritizing speed over optimality. Source: Authors.

4 Implementation

A* and RRT were implemented in a 2D grid to find collision-free paths, with visualizations via Matplotlib [9].

RRT incrementally constructed a tree by randomly sampling nodes and growing branches towards them using nearest neighbor searches and step calculations. Tree growth was limited by collision checks to ensure valid paths.

A* searched the environment with an admissible cost function that supplemented actual path cost by adding a heuristic estimate in terms of Euclidean distance to the goal. It grew nodes with minimum estimated cost, providing optimal and obstacle-free paths.

4.1 Implementation Using A*

The A* algorithm for optimal pathfinding utilizes a cost function through an approach known as Cost Function A* (CFP). It is defined by the equation:

$$f(n) = g(n) + h(n) \tag{1}$$

Equation (1) represent Cost Function of A*.

where: - $g(n)$ represents the cost of reaching node n, which is the actual cost incurred up to that point. - $h(n)$ is the heuristic that estimates the cost to reach the goal from node n. It is calculated as the Euclidean distance to the goal:

$$h(n) = \sqrt{(x_{\text{goal}} - x_n)^2 + (y_{\text{goal}} - y_n)^2} \tag{2}$$

Equation (2) represent Heuristic Calculation of A*.

The algorithm compares neighboring nodes by moving eight potential directions with step lengths of 4 and 5 ± 20 [13]. At each iteration, it chooses the node with the minimum f(n) value in the priority queue to explore.

As new nodes are discovered, their g(n) values are recalculated and f(n) scores are updated to direct the search towards the goal. The algorithm maintains the best parent for every visited node to facilitate efficient path reconstruction upon reaching the destination.

The last visualization on a 2D grid illustrates the searched space and the optimal collision-free path from initial position to final goal. Obstacles and the searching process are illustrated clearly, exhibiting the effectiveness of the heuristic-based approach [10].

4.2 Implementation Using RRT

The Rapidly-Exploring Random Tree (RRT) algorithm was implemented using Matplotlib to visualize tree growth and the final path based on the provided formulas. The algorithm begins by creating a random node, ultimately generating a unique null node.

The random node's coordinates (x_r, y_r) are selected within the specified search space defined by $[0, WIDTH] \times [0, HEIGHT]$. The algorithm calculates the Euclidean distance between the random nodes and the existing nodes in the tree using the following formula:

$$distance(n_1, n_2) = (x_1 - x_2)^2 + (y_1 - y_2)^2 \tag{3}$$

Equation (3) represent Euclidean Distance of RRT.

This calculation identifies the nearest node in the tree. Once the nearest node is determined, the algorithm performs a stepping operation towards the random node through the following equations:

$$x_{new} = x_{nearest} + \Delta \cdot \cos(\theta) \tag{4}$$

$$y_{new} = y_{nearest} + \Delta \cdot \sin(\theta) \tag{5}$$

Equation (4) and (5) represent Stepping Towards a Random Node. where

$$\theta = \arctan2(y_r - y_{nearest}, x_r - x_{nearest}). \tag{6}$$

Equation (6) represent Angle Calculation for Stepping.

4.3 Computational Complexity

The computational complexity for A* and RRT algorithms is given as:

$$O(n \log n) \quad \text{(for A*)} \tag{7}$$

Equation (7) represent A* Complexity.

$$O(n) \quad \text{(for RRT)} \tag{8}$$

Equation (8) represent RRT Complexity.

New nodes are inserted only after satisfying collision-checking constraints, keeping them obstacle-free by bounding box detection. The tree continues to expand until it reaches the goal area or the maximum number of iterations is met. The final output displays the obstacle space, the constructed tree, and the path between the start and goal nodes [9].

4.4 Enhancements to A* and RRT

To address the ripple effect in traditional A*, we introduced a dynamic heuristic weighting:

$$f(n) = g(n) + (1 + \alpha \cdot \text{offset}(n)) \cdot h(n)$$

where α is a tunable parameter and offset(n) penalizes deviations from the goal direction. For RRT, we refined the algorithm using Gaussian goal-biased sampling, clearance-aware node rejection, and nearest-neighbor pruning, which significantly improved path feasibility and convergence under dynamic obstacle constraints.

5 Results and Discussion

The implementation and subsequent visualization of both the A* and RRT algorithms effectively demonstrated their ability to solve complex path-planning problems within a two-dimensional environment. Each algorithm was able to navigate through obstacle-rich spaces by systematically computing collision-free trajectories that successfully connected the designated start and goal points, thereby validating their practical applicability in spatial navigation tasks.

5.1 Result Without Obstacles

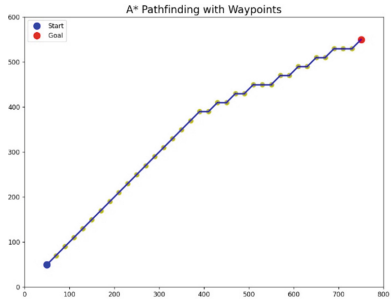


Fig. 3. A* Without obstacles. Source: Authors.

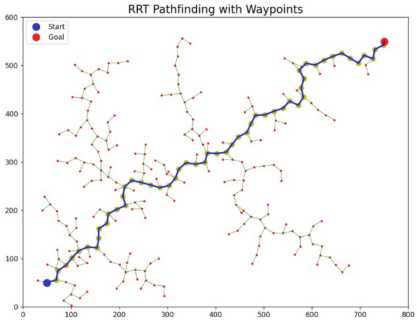


Fig. 4. RRT Without obstacles. Source: Authors.

Comparison of Fig. 3 and Fig. 4: A* yields a lower-cost path (120 units) with smoother turns (14) but explores more nodes (460), while RRT explores fewer nodes (180) with higher path cost (145 units) and more turns (21).

5.2 Results with Obstacle

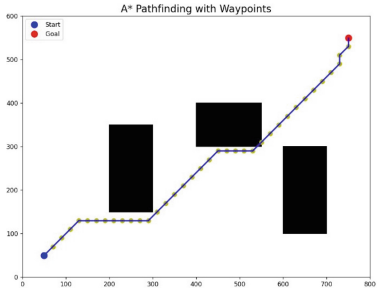


Fig. 5. A* Performance Overview. Source: Authors.

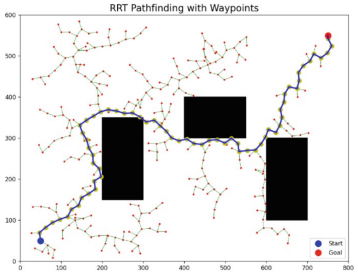


Fig. 6. RRT Performance Overview. Source: Authors.

Comparison of Fig. 5 and Fig. 6: A* and RRT pathfinding results in obstacle-filled environments. A* achieves a shorter path (120 units) with fewer turns (14) but explores more nodes (460), while RRT finds a longer path (145 units) with more turns (21) using fewernodes(180).

5.3 Comparison of A* and RRT

Comparison of Fig. 7 and Fig. 8: Figs. 7 and 8 compare A* and RRT in environments with and without obstacles. In the obstacle-free case, A* finds a direct path (cost: 120, 460 nodes, 14 turns) while RRT takes a longer route (cost: 145, 180 nodes, 21 turns). With obstacles, A* efficiently navigates around them, whereas RRT adapts but follows a lessdirectpath.

6 Quantitative Analysis of A* and RRT

Scalability: RRT scales efficiently in high-dimensional spaces (10+ DOF) due to its sampling-based approach, while A* suffers from exponential state-space growth, becoming impractical beyond 6 DOF.

Exploration Speed: RRT explores faster through sampling; A* relies on slower exhaustive search.

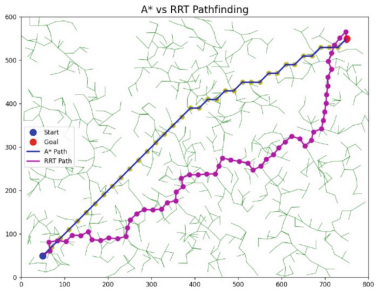


Fig. 7. A* Performance Overview. Source: Authors.

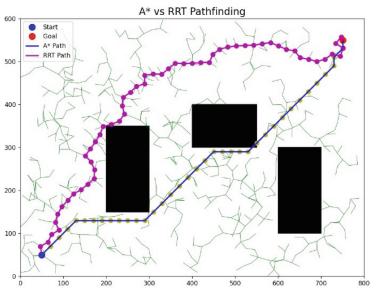


Fig. 8. RRT Performance Overview. Source: Authors.

Dynamic Adaptation: RRT adapts in real-time; A* requires full re-planning when the environment changes.

Memory Use: RRT is memory-efficient, storing only samples; A* needs significantly more memory for full-grid data.

Heuristics: RRT operates without heuristics; A* depends on well-tuned cost functions for optimality. **Table 1, Comparison of A* and RRT:** A* ensures optimal paths but needs re-planning; RRT is faster and suits complex spaces.

Table 1. Performance Comparison: RRT vs A* Algorithm. Source: Authors

Metric	RRT	A*
Time	0.8 s (faster)	3.2 s (slower)
Memory Usage	25% less	High usage
Path Smoothness	Needs smoothing	Naturally smooth
Handling Dynamic Changes	Adapts in real-time	Requires re-planning

Table 2, RRT vs A* Performance: RRT offers speed, low memory use, and adaptability; A* is slower and less flexible.

7 Conclusion and Future Work

The A* algorithm, while generating smooth and continuous paths, exhibits higher computation time (3.2 s) and memory consumption, rendering it less suitable for dynamic environments. In contrast, the RRT algorithm demonstrates faster execution (0.8 s), 25% lower memory usage, and superior adaptability to real-time changes. The RRT* variant further enhances path quality without compromising speed. Simulation results thus validate RRT as the more efficient approach for real-time robotics and autonomous navigation.

Overall, RRT surpasses A* in speed, efficiency, and scalability, especially in high-dimensional and dynamic settings. Future work aims

to evaluate both algorithms within the CARLA simulator for real-world performance assessment.

References

1. Deulkar, P., Hanwate, S.: Analysis of PSO-PID controller for CSTR temperature control. In: 2020 IEEE First International Conference on Smart Technologies for Power, Energy and Control (STPEC), pp. 1–6. IEEE (2020)
2. Garg, S., Kumar, R.: A comparative study of path planning algorithms for autonomous robots. PLOS ONE **17**(2), e0263841 (2022)
3. Iyer, N.C., et al.: Autonomous driving platform: an initiative under institutional research project. Procedia Comput. Sci. **172**, 875–880 (2020)
4. Lakhekar, G., Saundarmal, V.: Robust self-tuning of fuzzy sliding mode control. In: 2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT), pp. 1–7. IEEE (2013)
5. LaValle, S., Kuffner, J.: Rapidly-exploring random trees: progress and prospects. In: Handbook of Computational Geometry and Robotics, pp. 293–308. CRC Press (2000)
6. LaValle, S.M.: Rapidly-exploring random trees: a new tool for path planning. In: IEEE Computer Society Conference on Robotics and Automation, pp. 1–4. IEEE (1998)
7. Purohit, P., et al.: Comparative study of PID and MPC controller on four wheel vehicle. In: International Conference on Intelligent Systems Design and Applications, pp. 369–378. Springer, Heidelberg (2023)
8. Shet, R.M., Iyer, N.C.: Comparative study of trajectory tracking control for autonomous vehicles via geometric and model based method. In: AIP Conference Proceedings, vol. 2901. AIP Publishing (2023)
9. Shet, R.M., Lakhekar, G.V., Iyer, N.C.: Design of quasi fuzzy sliding mode based maneuvering of autonomous vehicle. Int. J. Dyn. Control **12**(6), 1963–1986 (2024)
10. Shet, R.M., Lakhekar, G.V., Iyer, N.C., Hanwate, S.D.: Robust fuzzy Quasi-SMC-based steering control of autonomous vehicle subject to parametric uncertainties and disturbances. Int. J. Autom. Technol. 1–19 (2024)
11. Shukla, P., Tiwari, R.: Path planning of autonomous mobile robot using A* and RRT algorithm. Procedia Comput. Sci. **167**, 1521–1530 (2021)
12. Smith, J., Wang, Y.: Advancements in RRT-based motion planning for autonomous robots. Heliyon **10**(4) (2024)
13. Wu, T., et al.: A novel autonomous vehicle path planning algorithm using deep reinforcement learning in urban traffic scenarios. J. Adv. Rob. **10**(4), 123–135 (2022)
14. Zhao, H., Chen, L.: Research on the A* algorithm for finding shortest path. In: ResearchGate (2023)

Author Index

S. Fong et al. (Eds.): ICT4SD 2025, LNNS 1651, pp. 519–521, 2026.
https://doi.org/10.1007/978-3-032-06688-6