Impact Factor – 6.625 | Special Issue – 389 | March 2026 | ISSN – 2348-7143

INTERNATIONAL RESEARCH FELLOWS ASSOCIATION'S

RESEARCH JOURNEY

Multidisciplinary International E-Research Journal

PEER REFEREED AND INDEXED JOURNAL

One Day National Level Conference On

ADVANCED COMPUTING AND ARTIFICIAL INTELLIGENCE (ACAI)

- CHIEF EDITOR -

Dr. Dhanraj T. Dhangar

- EDITORS -

Dr. Asha Patil
Dr. Amit P. Patil

For Details Visit To : www.researchjourney.net

Printed By : ACADEMIC BOOK PUBLICATIONS

Impact Factor – 6.625 ▪ Special Issue - 389 ▪ March 2026 ▪ ISSN – 2348-7143

INTERNATIONAL RESEARCH FELLOWS ASSOCIATION'S

RESEARCH JOURNEY

Multidisciplinary International E-research Journal

PEER REFREED AND INDEXED JOURNAL

One Day National Level Conference On

ADVANCED COMPUTING AND ARTIFICIAL INTELLIGENCE (ACAI)

-: Chief Editor :-

Dr.Dhanraj T. Dhangar

-: Editor :-

Dr. Asha Patil

Dr. Amit P. Patil

Printed by : ACADEMIC BOOK PUBLICATIONS

Impact Factor – 6.625 ▪ Special Issue - 389 ▪ March 2026 ▪ ISSN – 2348-7143

INTERNATIONAL RESEARCH FELLOWS ASSOCIATION'S

RESEARCH JOURNEY

UGC Approved Journal
Multidisciplinary International E-research Journal

One Day National Conference On

ADVANCED COMPUTING AND ARTIFICIAL INTELLIGENCE (ACAI)

-: Editors :-
Dr. Asha Patil
Dr. Amit P. Patil

Printed by

ACADEMIC BOOK PUBLICATIONS

Dyandeep Apartment, Plot No.2, Chaitanya Nagar, Opp.Progressive English Medium School, Jalgaon.
Email: academicbooksjalgoanl@gmail.com Ph: 0257-2253274,

ISSN : 2348-7143
March 2026

: Convener's Message :

It gives me immense pleasure to present the proceedings of the National Level Conference on **Advanced Computing and Artificial Intelligence** (ACAI), held on 7th March 2026 and organized by R.C. Patel Educational Trust's Institute of Management Research and Development (IMRD), Shirpur, Dist. Dhule.

The successful execution of this conference stands as a witness to well planning, coordinated teamwork, and the unwavering commitment of the organizing committee. Every aspect of the event from call for papers and peer review processes to the seamless management of hybrid technical sessions online and offline was carefully designed to ensure a productive and enriching academic experience for all participants.

The conference served as a dynamic platform for academicians, researchers, industry professionals, and students to exchange ideas, present innovative research findings, and deliberate on emerging technological trends. The participation of research scholars from Maharashtra, Gujarat, and Madhya Pradesh reflects the growing relevance and impact of Artificial Intelligence and advanced computing in today's rapidly evolving technological landscape.

I am delighted to share that the conference received 125 registrations, with 56 high-quality research papers presented across multiple technical sessions. The contributions covered diverse and current areas such as Artificial Intelligence, Machine Learning, Natural Language Processing, Cybersecurity, and Indian Sign Language Processing, showcasing the interdisciplinary scope and research depth of the domain.

The grand success of the conference was made possible through the visionary leadership of our Conference Chair, Dr. Vaishali B. Patil, Director, IMRD Shirpur. We were honored by the presence of distinguished dignitaries, including Prof. Dr. Dharmendra Bhatti, who delivered an enlightening keynote address, and Dr. Maya Ingle, who presented an insightful plenary session. Their inspiring inputs greatly enriched the intellectual environment of the conference and inspired participants to explore innovative and responsible technological solutions.

I extend my heartfelt gratitude to all authors, reviewers, session chairs, speakers, delegates, and members of the organizing committee for their dedication, cooperation, and tireless efforts in ensuring the smooth conduct and success of this event. I also sincerely thank the management, faculty, and staff of IMRD Shirpur for their constant support and encouragement.

I firmly believe that the research contributions compiled in these proceedings will significantly advance knowledge in the fields of Artificial Intelligence and advanced computing. May this conference continue to foster innovation, collaboration, and excellence, and serve as a promoter for future technological advancements.

Dr. Asha Patil
Conference Convenor,
IMRD Shirpur

: Director's Message :

Dear Esteemed Guests, Participants, and Scholars,

It gives me immense pleasure to welcome you to the National Level Conference on **Advanced Computing and Artificial Intelligence** (ACAI-2026). This conference provides a valuable platform for academicians, researchers, industry experts, and students to engage in meaningful dialogue, share innovative ideas, and contribute to the advancement of emerging technologies, reflecting our institute's vision of delivering quality higher education and fostering a global perspective in Information Technology and computer science.

In alignment with the principles of the National Education Policy 2020, ACAI-2026 emphasizes multidisciplinary education, research and innovation, skill development, digital literacy, and critical thinking. This focus vibrates with our mission to impart high-quality, value-based education that nurtures responsible, proactive, and competent professionals. By encouraging experiential learning, creativity, and problem-solving, the conference supports our goal of molding students into socially responsible individuals capable of successful in a global and technology-driven environment.

The rapid growth of fields such as Artificial Intelligence, Machine Learning, Natural Language Processing, Cyber Security, and Data Analytics highlights the importance of a knowledge-driven and technology-enabled society. ACAI-2026 provides an opportunity to explore these advancements while promoting interdisciplinary collaboration and industry-academia partnerships, preparing students to become self-reliant, globally competent, and future-ready core objectives of our institutional mission.

I extend my sincere gratitude to all participants, speakers, reviewers, and organizing members for their contributions in making this conference a success. I am confident that ACAI-2026 will inspire continued research, innovation, and excellence, fulfilling our vision of shaping capable, socially responsible, and innovative professionals.

Thank you and I wish you all a productive and enriching experience.

Dr. Vaishali B. Patil
Director – RCPET's IMRD, Shirpur

ISSN : 2348-7143
March 2026

: Editorial Message :

It is with great pride and enthusiasm that we present the proceedings of the National Level Conference on **Advanced Computing and Artificial Intelligence** (ACAI), held on 7th March 2026, and organized by R.C. Patel Educational Trust's Institute of Management Research and Development (IMRD), Shirpur, Dist. Dhule. This conference reflects our continued commitment to promoting research excellence, innovation, and knowledge dissemination in the rapidly advancing fields of computing and artificial intelligence.

The conference received an overwhelming response, with a total of 125 registrations and 56 research papers selected for presentation across multiple technical sessions conducted in hybrid mode. The accepted papers represent diverse and contemporary research areas, including Artificial Intelligence, Machine Learning, Natural Language Processing, Cyber Security, Indian Sign Language Processing, and other emerging technologies, demonstrating the interdisciplinary scope and relevance of the conference.

A key highlight of this publication is the rigorous quality assurance process adopted for paper selection. All submissions underwent a peer-review process, evaluated by experts based on originality, technical quality, clarity, and relevance. Authors were required to incorporate reviewers' suggestions to meet the prescribed academic standards.

To ensure academic integrity and originality, all manuscripts were screened using plagiarism detection tools, and only papers with a similarity index of less than 10% (excluding references) were accepted for publication. This strict faithfulness to ethical publication practices enhances the credibility and reliability of the research presented in this volume.

These proceedings are published with ISSN–2348-7143 under Research Journey Multidisciplinary International E-research Journal Peer Refreed and Index Journal, ensuring formal academic recognition. We extend our sincere gratitude to all authors, reviewers, session chairs, keynote and plenary speakers, and members of the organizing committee for their dedication and valuable contributions. Their collective efforts have ensured not only the successful conduct of the conference but also the high academic quality of this publication.

We firmly believe that the research compiled in these proceedings will serve as a valuable reference for academicians, researchers, and industry practitioners, and will contribute significantly to the advancement of Artificial Intelligence and advanced computing technologies. We hope this initiative will continue to inspire high-quality research, innovation, and interdisciplinary collaboration in the years ahead.

Warm regards.

Dr. Amit P. Patil
Editor,
Assistant Professor,
RCPET's IMRD, Shirpur

:CONTENTS:

English

RESEARCH JOURNEY International Multidisciplinary E-Research Journal
Impact Factor (SJIF) - 6.625 | ***Special Issue 389 :*** *Advanced Computing and Artificial Intelligence (ACAI)*
ISSN : 2348-7143
March 2026

Automated Solid Waste Classification Using Supervised Machine Learning Algorithms

Giteshwari Patil,
Rajnandini Patil
R. C. Patel Educational Trust's Institute of Management Research and Development, Shirpur.

Abstract:

Waste management is a major issue in today's world, which is a highly pressing concern, 50 % in rural and semi-urban areas where qualified personnel and sophisticated infrastructure are not readily available. This paper presents a machine learning-based system for the automatic classification of waste into three major categories: recyclable. The proposed system utilizes a variety of sensor-driven parameters such as weight, moisture content, pH, conductivity, chemical composition, intensity of smell, and physical properties such as colour, texture, and flammability. These parameters are pre-processed using normalization and categorical encoding techniques and then used in four different classification algorithms: K-Nearest Neighbours, Support Vector Machine, Decision Tree, and Naïve Bayes. The performance of each algorithm is measured using accuracy, precision, recall, F1-score, and confusion matrices to determine which algorithm performs best. The proposed system has the potential to decrease manual effort, facilitate the safe disposal of hazardous materials, and encourage the practice of efficient recycling.

Keywords: Waste segregation, Classifier algorithms, K-Nearest Neighbours, Support Vector Machine, Decision Tree.

Introduction:

India generates approx. 62-65 million of municipal solid waste annually. Biodegradable waste forms about 55 % of total municipal waste while on other hand recyclable waste contributes nearly 30-35 % of total municipal waste. One of the most important steps in waste management is the process of classification, which helps in determining whether the material is recyclable, organic, or hazardous [1]. Classification is essential as it helps in reusing recyclable materials, composting or naturally treating organic waste, and properly handling hazardous waste to avoid any kind of contamination or accidents. Unfortunately is many rural and semi-urban regions, waste classification is still a manual process, which is performed by untrained personal with minimal equipment. Manual classification not only makes the process slower but also poses serious health threats to the workers and results inaccuracies that make recycling and disposal ineffective [1].

In recent years, the emergence of smart technology has paved the way for automation in waste management. Today, sensors can measure physical and chemical properties of waste, such as wight, moisture content pH value, odor intensity, texture, combustibility, and chemical composition. These properties possess sufficient information to identify different types of waste but interpreting them in an efficient manner requires sophisticated computational algorithms. This is where machine learning (ML) technology becomes extremely useful [2].

Waste generation is an inevitable byproduct of human activity, from domestic trash to industrial waste. With the increasing pace of urbanization, population, and consumption patterns, the volume and complexity of waste are steadily rising. Estimates suggest that billions of tons of solid waste are produced annually, and a substantial amount of it is either inadequately managed of improperly disposed of. This has serious repercussions on human health, soil quality, water purity, and the entire ecological system[3].

Machine learning is set of tools and techniques that can learn from data patterns and make prediction without being explicitly programmed. By training models on a dataset of waste samples with labelled categories (recyclable, organic, hazardous), an ML system can learn to automatically classify new waste samples in real-time. This makes the process less prone to human error, more efficient, and more systematic about waste management[4].

The Purpose of this research is to create and evaluate a machine learning classification system for waste based on sensors data. The research compares various classification algorithms such as K-Nearest Neighbours (KNN), support Vector Machine (SVM), Decision Tree, and Native bayes to see which one provides the highest accuracy and reliability. The system is intended to be applied in real-world settings, particularly in rural recycling centres, where the best resources and infrastructure are not readily available[4].

This research tackles the problem of automated waste classification, which can help make the process of handling hazardous waste safer and more sustainable. It can also help increase recycling rates and overall sustainability. The results of this research are expected to lay the groundwork for creating cost-effective, real-world waste management systems using artificial intelligence[5].

Problem Definition:

Inadequate waste classification can result in heath issues, pollution and inefficient recycling rates. Traditional approaches are time-consuming, prone to errors, and lack scalability for large population.

Objectives:

1. To develop a system that can automatically distinguish between different kinds of waste.
2. To reduce the amount of money and time that goes into waste sorting and collection by implementing an automated system.
3. To develop a system that can be applied in villages, small towns, or large cities.
4. To help to achieve a clean and green tomorrow by reducing waste, reusing, and recycling.

Challenges:

1. Obtaining sufficient data about waste to train machine learning algorithms may be challenging.
2. Training machine learning algorithms may require a substantial amount of data and processing power.
3. The system will require periodic maintenance to ensure accuracy and efficiency.

Review of Literature:-

The given research literature focus on modern waste management and the use of advanced technologies to improver waste classification. Due to rapid urban growth and increasing waste, manual sorting is no longer effective. Technologies like AI , ML , IOT , and Deep Learning help in accurately identifying waste, reducing human effort, and supporting environmental sustainability.

1. Abdullah Alourani and M. U. (2025): developed and intelligent smart solid waste management system (ISSWMS) that integrates IOT sensors for monitoring bin fill levels and AI-driven VGG-19 classification for segregating plastic, glass, mantal, and trash on a conveyor belt.
2. Bingbing Fang and Jianying Yuan Shi (2023): reviewed artificial intelligence application is waste management for smart cities, covering area like waste-to-energy conversion, smart bins, sorting robots, and predictive modelling.
3. Siedlecka, J. S. (2025): explored circular economy strategies to tansform fine glass, ceramic and plastic residues from glass recycling plants into valuable resources, emphasizing form glass production and other valorisation techniques.
4. Pieters, Luntungan Stephen (2025): developed an automatic waste classification system using CNN-based deep learning to support smart waste management (SWM), targeting plastic, paper, organic, and non-organic waste.
5. Mohammed A. AI Doghan and V. P. (2023): AI-enabled reverse logistic and big data analytics to improve waste and resource management in Saudi Arbias manufacturing sector, focusing on circular economy supply chains.
6. Mona Jabor AI-Thani et al. (2025): conducted the bibliometric and systematics review of smart food supply chain management (SFSCM), analysing trends, technologies, and research gaps to enhanced food security and sustainability.
7. Muhammad Aminullah Abd Rahim and S, A. (2025): evaluated 3R (Reduce, Reuse, Recycle) waste management practices for construction and demolition (C&D) waste among Malaysian contractors, addressing daily generation of ~25,600 tonnes.
8. Sameh Fuqha and N. N. (2025): review AI and IOT application and smart waste management, outlining challenges like high cost and infrastructure gaps alongside opportunities for automatic sorting and predictive collection.

The literature review points out the growing use of AI, machine learning, IoT and deep learning concepts use in automatic waste segregation to boost circular economy. Most of the studies have proposed automatic waste classification techniques using IoT sensors and CNN based classification models for automated waste evaluation. While review based models has applications such as smart bins, robotic sorting, predictive collection, waste -to-waste energy systems, and reverse logistics enhanced by data analytics. This research works focused on circular economy for classification of glasses, plastic, and ceramic residues, 3R practices in construction and sustainability in smart food supply chains. Despite of these techniques and advancements, most of research address the gap in the classification of recyclable and biodegradable waste using machine learning techniques. To bridge this gap, the present research aims to built an ML based classification approach specifically targeting recyclable and biodegradable waste, leads to sustainable waste classification.

Methodology:

The research paper will analyse the dataset of 500 images, which were resized and normalized, and processed with the CNN algorithm to automatically extract input features from the images. The extracted features represented were stored in CSV (Comma-Separated Values) format to create a structured dataset. This CSV file was used as input

for training machine learning classifier algorithms such as Support Vector Machine, K-Nearest Neighbors, Naïve Bayes, and Decision Tree. Neighbours with a preference for predicting waste as recyclable or organic.

Machine Learning Algorithms:

KNN: (K-Nearest Neighbours) is a supervised learning algorithm that classifies a new data point based on the nearest neighbours. To find the distance between two points mainly Euclidean distance are used. The value of k determines the number of neighbours.

This table includes the dataset trained on KNN classifier algorithm along with tunning parameters.

Table 1: KNN PREDICTION TABLE:

No of features	K Value	Train%	Test%	Accuracy%
8	3	0.8	0.2	85.00
		0.7	0.3	85.33
		0.6	0.4	87.25
	5	0.8	0.2	86.50
		0.7	0.3	85.33
		0.6	0.4	86.67

This table showing the model accuracy for different settings of K value, train-test split, and number of features, accuracy ranges 85.00% to 87.25%. Best accuracy: 87.25% (K=3, Train/Test=0.6/0.4, 8 features).

SVM: Support Vector Machine (SVM) is a supervised machine learning algorithm used for classification as well as regression purposes by determining an optimal hyperplane to separate classes by maximizing the margin between data points. Work better with high-dimensionality dataset.

This table includes the dataset trained on SVC classifier algorithm along with tunning parameters.

Table 2: SVC PREDICTION TABLE:

No of features	Kernel Value	Train %	Test %	Accuracy%
8	linear	0.8	0.2	88.00
		0.7	0.3	89.00
		0.6	0.4	88.75
	rbf	0.8	0.2	90.00
		0.7	0.3	88.67
		0.6	0.4	93.00
	poly	0.8	0.2	89.00
		0.7	0.3	88.00
		0.6	0.4	92.00

This table shows the Support Vector Machine (SVM) model performance with different kernel types (linear, polynomial, RBF), train-test splits, and 8 selected features. Accuracy ranges 88.00% to 93.00%. Best accuracy: 93.00% (RBF kernel, train/test=0.6/0.4, 8 features).

Naive Bayes: Naive Bayes is a supervised machine learning classifier that uses Bayes' theorem and supposes independence between features and chooses the class with maximum probability. This data table includes the dataset trained on the Naïve Bayes regressor algorithm with tuning parameters.

Table 3: NAIVE BAYES PREDICTION TABLE:

No of features	NB	Alpha	Binarize	Accuracy%
8	Gaussian NB	-	-	71.00
	Bernoulli NB	1.0	0.5	69.00
	Complement Naive Bayes	1.0	-	69.00

This table shows the performance of Naïve Bayes on three different models (Gaussian NB, Bernoulli NB, Complement NB) on 8 features.Best accuracy: 71.00% (Gaussian NB). Other model (Bernoulli NB, Complement) accuracy is approximately 69%.

Decision Tree: Decision Tree is a machine learning algorithm that splits data into branches based on feature decisions to predict outcomes. It creates a tree where each node represents a condition and leaves represent final predictions.

Table 4: Decision Tree Prediction Table:

This table includes the dataset trained on Decision Tree classifier algorithm along with tunning parameters.

No of features	Tune Parameters	max_depth	Accuracy %
8	criterion="gini"	None	79.50
	criterion="entropy"	None	78.50
	criterion="gini"	5	84.00
	criterion="entropy"	5	81.00
	min_samples_split=2	None	79.50
	min_samples_split=2	5	84.00

This table shows the Decision Tree model performance with different splitting criteria (gini, entropy) and max_depth settings. Accuracy range: 78.50% to 84.00%. Best accuracy 84.00% (criterion=gini, max_depth=5,8 features).

Result:

This table shows the accuracy comparison of all machine learning classifier algorithms.

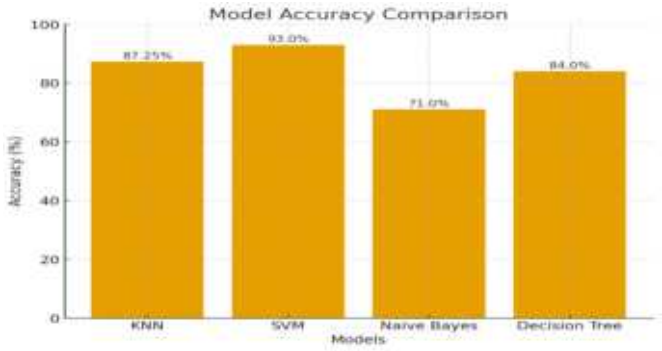
Table 5: Performance Comparison of Classification Algorithms

Sr no	Classifiers	Accuracy%
1	KNN	87.25%
2	SVM	93.00%
3	Naïve Bayes	71.00%
4	Decision tree	84%

This table shows that after comparing the accuracy score of all machine learning classifier algorithms, SVM gives the highest accuracy of 93.00 % among all classifiers.

This graph shows the comparative analysis of classification accuracy among KNN, SVM, Naïve Bayes, and Decision Tree algorithms. The X-axis denotes the machine learning models, whereas the Y-axis denotes the accuracy score obtain during testing.

Figure 1: Comparative analysis of all machine learning classifier algorithm accuracy



This graph shows the comparative analysis of classification accuracy among KNN, SVM, Naïve Bayes and Decision Tree. This represents that SVM gives highest accuracy to classify waste as organic or recyclable upto 93.00%.

Conclusion:

This project proves that it is possible to sort the waste into recyclable and organic waste categories using machine learning techniques. The accuracy of the classification of the waste into the different categories is good. The features of the waste, such as weight, moisture content, pH, odor, and chemical content, improve the accuracy of the classification of the waste compare to guessing the type of waste. The accuracy of the classification of the waste into different categories by the models (KNN, SVM, Decision Tree, and Naïve Bayes) Varies, but it proves the possibility of waste sorting using machine learning techniques. The SVM model has the highest accuracy of up to 93.00% in the classification of the waste. This system can be useful in the classification of the waste in the rural areas.

References :

1. M. U. Abdullah Alourani, "Smart waste management and classification system using advanced IoT and AI technologies," PeerJ Computer Science, vol. 24, 2025.
2. J. Y. S. Bingbing Fang, "Artificial intelligence for waste management in smart cities: a review," Environmental Chemistry Letters, Apr. 24, 2023.
3. J. S. Siedlecka, "From Waste to Resource: Circular Ewa Economy Approaches to Valorize Fine Glass, Ceramic, and Plastic Residues in a Glass Recycling Plant," Sustainability, p. 20, Sep. 4, 2025.
4. L. S. Pieters, "Development of automatic waste classification system using CNN based deep learning to support smart waste management," Jurnal Inovtek Polbeng - Seri Informatika, 2025, pp. 2527–9866.
5. V. P. Mohammed A. Al Doghan, "AI-enabled reverse logistics and big data for enhanced waste and resource management," Operational Research in Engineering Sciences: Theory and Applications, pp. 15–335, 2023.
6. M. H. Mona Jabor Al-Thani, "Smart food supply chain management: A bibliometric and systematic review," Food and Humanity, p. 100736, 2025.
7. S. A. Muhammad Aminullah Abd Rahim, "Evaluating 3R waste management practice for construction and demolition waste among contractors in Malaysia," Journal of Sustainable Civil Engineering and Technology, pp. 78–89, 2025.
8. N. N. Sameh Fuqaha, "Artificial Intelligence and IoT for Smart Waste Management: Challenges, Opportunities, and Future Directions," Future Techno Science, pp. 3048–3719, Jun. 4, 2025.
9. Y. Zhou, "AI-driven digital circular economy with material and energy sustainability for Industry 4.0," Energy and AI, vol. 20, p. 13, 2025.
10. S. M. L. A. S. A. B. A. G. P. Kiran S. Pillai, "Municipal Solid Waste Management: A Review of Machine Learning Applications," EDP Sciences, Vols. 455, 2023, no. E35 Web conf, p. 12, 5 December 2023.
11. F. Fotovvatikhah, I. Ahmedy, R. Md Noor, and M. U. Munir, "A systematic review of AI-based techniques for automated waste classification," Sensors, vol. 25, no. 10, p. 3181, May 2025.
12. M. Chhabra, B. Sharan, M. Elbarachi, and M. Kumar, "Intelligent waste classification approach

based on improved multi-layered convolutional neural network," Multimedia Tools and Applications, vol. 83, pp. 84095–84120, 2024.

13. A. Arishi, "Real-time household waste detection and classification for sustainable recycling: A deep learning approach," Sustainability, vol. 17, no. 5, p. 1902, Feb. 2025.
14. P. A. Rajeev et al., "Advancing e-waste classification with customizable YOLO based deep learning models," Sci. Rep., vol. 15, Art. no. 18151, 2025.
15. D. Ghosh and A. Goswami, "HybridSOMSpikeNet: A deep model with differentiable soft self-organizing maps and spiking dynamics for waste classification," arXiv, Oct. 2025.
16. W. Qiu, C. Xie, and J. Huang, "An improved EfficientNetV2 for garbage classification," arXiv, Mar. 2025.
17. I. Dawar, A. Srivastava, M. Singal, N. Dhyani, and S. Rastogi, "A systematic literature review on municipal solid waste management using machine learning and deep learning," Artif. Intell. Rev., vol. 58, p. 183, 2025.
18. "Intelligent waste sorting for sustainable environment: A hybrid deep learning and transfer learning model," Gondwana Research, vol. 146, pp. 252–266, Oct. 2025, doi:10.1016/j.gr.2024.07.014.
19. "Intelligent waste sorting for urban sustainability using deep learning," Sci. Rep., 2025, doi:10.1038/s41598-025-08461-w.
20. J. C. Vesga Ferreira, H. E. Perez Waltero, and J. A. Vesga Barrera, "Design of a waste classification system using a low experimental cost capacitive sensor and machine learning algorithms," Appl. Sci., vol. 15, no. 3, p. 1565, Feb. 2025.
21. A. Wadhwanwala and T. Kumar, "Revolutionizing waste management: Leveraging YOLOv8 for enhanced waste categorization," Preprints, vol. 2025, Jul. 2025.

ISSN : 2348-7143
March 2026
RESEARCH JOURNEY International Multidisciplinary E-Research Journal
Impact Factor (SJIF) - 6.625 | ***Special Issue 389 :*** *Advanced Computing and Artificial Intelligence (ACAI)*

Hybrid Machine Learning Based Recommendation System

Karishma Pardeshi,
Janvi Patil
Student of RCP IMRD, Shirpur

Abstract

Nowadays, people get a lot of options on online platforms, such as online shopping, watching movies or web series, and using different apps. While having more options is good, it can be confusing to choose the right one. People feel confused when deciding which product or information they would like the most. This can reduce their satisfaction and engagement.

Therefore, a recommendation system is used. This system suggests appropriate things to users based on their previous usage information. But some traditional methods, such as Collaborative Filtering and Content-Based Filtering, do not work well if the users are new or have little information. Sometimes they give the same recommendations again.

To address this problem, this research uses a hybrid machine learning approach. It combines two different recommendation methods. The system provides convenient and diverse recommendations to users by seeing what they like and known about items. The model is tested using Precision, Recall, F1-Score, and RMSE metrics.

Experiments have shown that using multiple techniques together yields better results. Therefore, the hybrid approach improves personalized recommendations and improves the user experience.

Keywords

Hybrid Recommendation, Machine Learning, Personalisation, Collaborative Filtering, Content-Based Filtering.

Introduction

In today's digital age, online platforms such as shopping websites, streaming apps and social media collect and manage large amounts of user data. Users are faced with many options – like products, movies or various services – but the sheer number of options makes it difficult to make the make right choice. This creates confusion and can reduce user satisfaction.

Recommender systems are used on most digital platforms to solve this problem. These systems analyze user's behavior, past transactions and interests and suggests appropriate content accordingly. For example, platforms like Amazon and Netflix use recommendation system to improve user experience and increase their engagement.

Conventional recommendation methods mainly use Collaborative Filtering and Content-Based Filtering techniques. Although these methods are effective, they have some limitations. Such as cold-start problem (lack of information about a new user), limited data, difficulties encountered as the system grows, and lack of diversity in recommendations. Many researches focus only on accuracy, but not enough attention is paid to system efficiency and balanced evaluation.

To overcome these limitations, a Hybrid Machine Learning based recommendation system has been proposed in this research. This system uses a combination of both Collaborative and Content-Based methods. Its main objective is to provide more accurate, personalized and reliable recommendations as well as develop more efficient systems for digital platforms.

Problem Statement:

Digital platforms are currently growing at a very fast pace. Online shopping, streaming services and social media are generating large amounts of user data. Recommender systems are used to suggest suitable options to users. But traditional methods like Collaborative Filtering and Content-Based Filtering have some limitations.

These methods suffer from cold-start problems (lack of information about a new user), insufficient data, performance constraints as the system grows, and lack of diversity in recommendations. This can reduce the accuracy of recommendations and the quality of personalization. Also, many researches focus only on accuracy, but not enough attention is paid to the overall performance, reliability and proper evaluation of the system.

Hence there is a need to develop more effective Hybrid Machine Learning based recommendation system to overcome these limitations. This system will help in providing more accurate, personalized and reliable recommendations using a combination of various techniques. Therefore, a more efficient and useful system can be created for digital platforms.

Objectives:

1. Combining two methods to create a system to get better recommendations.
2. Using machine learning to suggest items based on what users like.
3. To check if the system works properly.
4. Comparing the new system with the old once to see which one is better.

5. Making recommendations more accurate and user experience better.

Literature Review:

Recommendation System have improved a lot over time. Now they provide users with more appropriate and personalized suggestions as per their interests. Initially Collaborative Filtering was used in most of systems. In this method suggestions are made based on people with similar interest. For example, if two people like the same movies, the movies that one has seen are recommended to the other.

After this came the method of Matrix Factorisation. This technology discovers the hidden relationship between the user and the object. So, one can more accurately predict what someone will like.

Also, the method of Content-Based Filtering was developed. In this method, suggestions are given based on what the user has seen or selected before and what are the characteristics of those objects. For example, if a person likes action movies, then action movies are more recommended to him.

But these traditional methods have some problem. These methods do not work well if the information is sparse (Data Sparsity), there is a new user or new object (Cold-Start Problem), large amount of information has to be handled (Scalability Issues), or repeated requested of the same type (Limited Diversity). Hybrid Recommendation System were created as a solution to these problems. The systems combine information about what users like and item to make recommendations. Hence recommendations are more accurate, diverse and reliable.

Table 1. Presents a summary of major research contributions in hybrid machine learning-based recommendation systems.

Research	Year	Technique Used	Key Contribution	Limitations
Koren et al.	2009	Matrix Factorization (SVD)	Improved prediction accuracy using talent factor models	Cold-start problem, requires large dataset
Burke	2002	Hybrid Recommendation	Introduced hybrid strategies combining CF and CBF	Increased system complexity
Sarwar et al.	2001	Item-Based Collaborative Filtering	Scalable and efficient recommendation method	Limited diversity, cold-start issue
Adomavicius & Tuzhilin	2005	Context-Aware Recommender Systems	Expanded evaluation beyond accuracy metrics	Complex implementation
Amatrial (Netflix)	2013	Ensemble & Hybrid Systems	Real-world large-scale hybrid implementation	High computational
Mobasher	2007	Web usage Mining + Hybrid	Personalized web recommendation	Data preprocessing challenges
Kalideen & Yagli	2025	ML-Based Hybrid Systems	Addressed sparsity and scalability	Mostly theoretical analysis
Sami et al.	2024	CF + NCF + RNN + Content Hybrid	Improved RMSE and precision in hybrid model	Increased model complexity

Methodology:

In this section we have explained in simple terms how we have built our Hybrid Machine Learning based recommendation system. First, we created the dataset. After that the data was cleaned. Next, we created two different models – Collaborative Filtering and Content-Based Filtering. After that, these two models were combined to form a Hybrid system. Finally, the system was tested to check its performance. Our main goal is to provide more accurate recommendations and at the same time ensure that the system works faster.

Description of Dataset:

This dataset includes information about users, the items they have rated, and the scores for those ratings. Also, the features of each item are also given. To test how well the system works, we split the data into two parts – 80% training and 20% testing. Data for training is used to build the model, while data for testing is used to measure how accurately the model predicts.

Data Preprocessing:

Before building the model, we cleaned the data to ensure it was accurate and consistent. This included removing duplicate (recurring) entries, filling in empty values, normalizing ratings, creating a user-item matrix, and creating feature vectors of item features.

Collaborative Filtering Model:

In this method we used we used user-based collaborative filtering. This method provides recommendations by looking at the interests of users who have similar interests.

Similarity between users is measured by cosine similarity method.

$$sim(u,v) = \frac{u \cdot v}{\| u \| \| v \|}$$

Then an estimated rating of an item is calculated based on the ratings of similar users.

$$\hat{r}_{u,i} = \frac{\sum v \in N(u) sim(u,v) \cdot r_{v,i}}{\sum v \in N(u) \mid sim(u,v) \mid}$$

Users whose interests match more, are given more importance.

This method understands user's interests based on their past rating history and accordingly recommends items preferred by users with similar interests.

Content-Based Filtering Model:

This method suggests products to the user based on his interest.

Each product contains some information, such as its type, category or description. We convert this information into numbers and create a list of features for each product. After this, product features and user preferences are compared using the cosine similarity method. This method measures how similar two things are to each other. If a new product enters the market and does not yet have any ratings, this method can still provide a valid recommendation. So ths method is very useful.

Hybrid Integration Strategy

In this system we use two methods together:

Collaborative Filtering (CF)

Content-Based Filtering (CF)

The final recommendation is given by combining the scores of both these methods.

The following formula is used to calculate the final marks:

Final Score = a(CF) + (1 – a) (CB)

Here:

A is a number between 0 and 1. It decides how much importance to give to which method. CF is the score obtained from Collaborative Filtering.

CB is the score obtained from Content-Based Filtering.

We try different values of a and a choose the value that gives the best result.

Thus, this system provides more accurate and better recommendations by balancing user interests and product information.

Input:

User-Item Rating Matrix R

Item Feature Matrix F

Weight Parameter a

Target user u

Output:

Top – N recommended items

Step 1: Preprocess dataset

Step 2: Compute user similarity using cosine similarity

Step 3: Predict ratings using collaborative filtering

Step 4: Compute content-based similarity scores

Step 5: Combine scores using a weighted hybrid formula

Step 6: Rank and recommend Top-N

Computational Complexity

The costs of the approach are:

- User similarity computation: O (n2)
- Calculation of similarity of items: O(m2)
- Hybrid Integration: o(m)

Overall complexity:

Where:

n = number of users

m = number of items

k = number of features

Experimental Setup and Evaluation Metrics

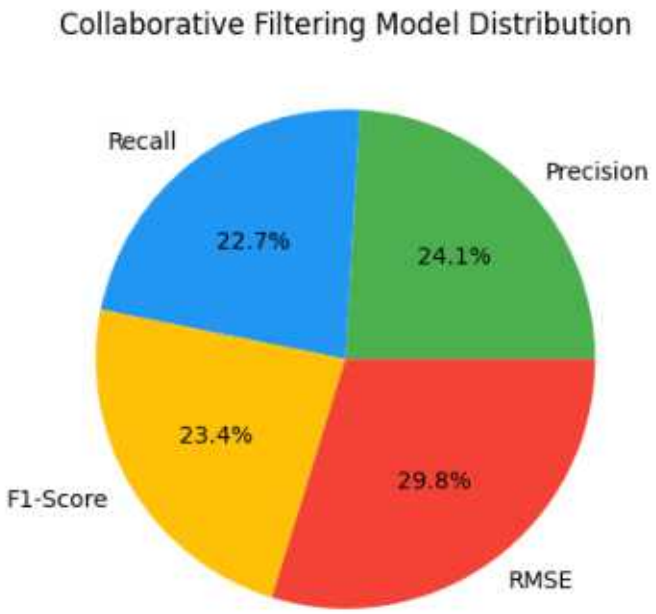
This section presents the performance evaluation of the proposed Hybrid Machine Learning-Based Recommendation System. The dataset was divided into 80% training data and 20% testing data to evaluate model performance. The hybrid model was compared with standalone collaborative filtering and content-based filtering approaches.

The system performance was measured using Precision, Recall, F1-Score, and Root Mean Squared Error (RMSE).

Table 2. Performance Comparison of Recommendation Models.

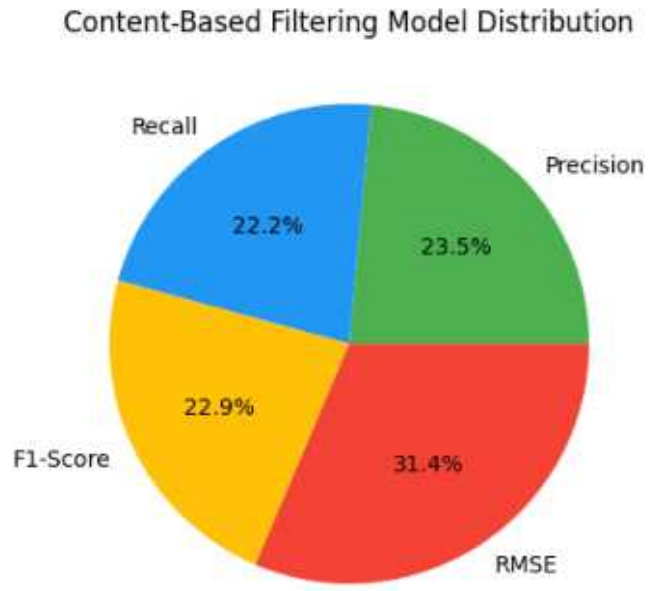
Model	Precision	Recall	F1_Score	RMSE
Collaborative Filtering	0.72	0.68	0.72	0.89
Content-Based Filtering	0.69	0.65	0.67	0.92
Hybrid Model	0.84	0.80	0.82	0.74

Figure 1. Collaborative Filtering Model



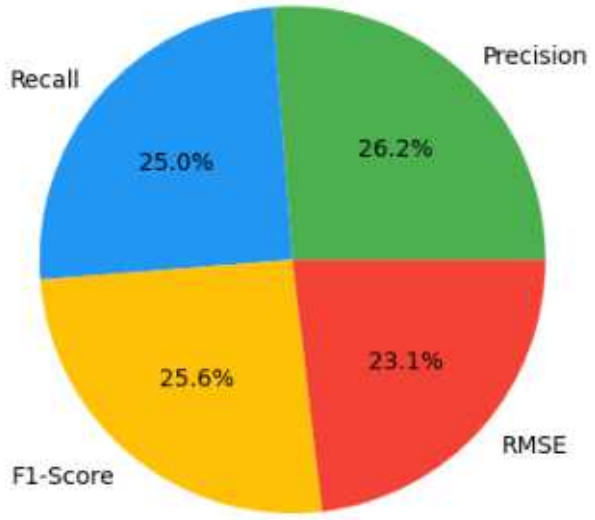
Collaborative filtering provides medium accuracy/ recall and high RMSE.

Figure 2. Content-Based Filtering Model



Content-Based model is balanced and has a greater RMSE compared to the hybrid.

Figure 3. Hybrid Recommendation Model



The hybrid model is more precise, and has a higher recall, F1-Score, and lower RMSE.

As indicated in the graph, collaborative and content-based methods have better performance on the overall recommendation when combined.

Discussion:

The results show that using the two methods together gives better results than using a single method.

Collaborative Filtering is a method that finds people with similar interests and suggests items accordingly. But if the information is less or the new user has not given any rating yet, this method does not work well. Content-Based Filtering this method provides recommendations based on the characteristics of objects. So, it works well for new items. But it may recommend the same type of item over and over, so the recommendations may seem a bit similar. The Hybrid model combines the best of both methods. It takes into account both the interests of users and information about objects.

This makes the system provide more accurate, balanced and personalized recommendations. It also reduces miscalculations and improves overall system performance.

It is true that the Hybrid model requires a bit more computation and processing. But it is worth using it as the recommendations are better and more accurate.

Hybrid system is also a good option when creating new digital platforms, as they provide more accurate and reliable recommendation.

Conclusion:

In this study, a recommendation system based on hybrid machine learning is developed. In this system two methods are used together:

Collaborative Filtering, Content-Based Filtering.

The main objective of this system is to overcome the problems of the old (traditional) recommendation system. Those problems are:

Cold-start problem, gradually increase the data, less variety in recommendations, repeatedly recommending the same type of item.

A weighted hybrid method was developed and used to solve these problems.

This system combines both user interests (Collaborative Filtering) and object characteristics (Content-Based Filtering). This makes recommendations more accurate and personalized.

The results showed that the hybrid model performed better than the single method. In this model:

Precision, Recall and F1 Score were, higher Reduced RMSE (i.e. reduced prediction error) This shows that using the two methods together gives more accurate results and reduced errors.

The hybrid method is slightly more complex and requires more computation. But it is worth it because of the good and quality recommendations it gets. This method is especially useful for e-commerce websites and streaming services.

In the future, this system can be made even better by adding deep learning, real-time recommendations, and contextual recommendations, and contextual recommendations system. This will make the system more personalized and widely usable. The main objective of this research is that multiple machine learning methods can be used together to create a more effective and powerful recommendation system.

References :

1. B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, “Item-based collaborative filtering recommendation algorithms.” In Proc. 10th Int. World Wide Web Conf. (WWW), 2001, pp.285-295.
2. R. Burke, “Hybrid recommender systems: Survey and experiments,” User Modeling and User-Adapted Interaction, vol. 12, no. 4.pp. 331-370, 2002.
3. G. Adomavicius and A. Tuzhilin, “Toward the next generation of recommender systems: A survey of the state-of-the-art and possible, extensions,” IEEE Trans. Knowl. Data Eng., vol. 17, no. 6, pp.734-749, Jun. 2005.
4. Y. Koren, R, Bell, and C. Volinsky, “Matrix factorization techniques for recommender system,” IEEE Computer, vol. 42, no. 8, pp. 30-37, Aug. 2009.
5. X. Su and T. M. Khoshgoftaar, “A survey of collaborative filtering techniques,’ Advances in Artificial Intelligence, vol. 2009, Article ID 421425, 2009.
6. C. Desrosiers and G. Karypis, “A comprehensive survey of neighborhood-based recommendation methods,” In Recommender System Handbook, Springer, 2011, pp. 107-144.
7. M. J. Pazzani and D. Billsus, “Content-based recommendation systems,” In the Adaptive Web, Springer, 2007, pp. 325-341.
8. C. Amatriain and J. Basilico, “Netflix recommendations: Beyond the 5 stars,” Netflix Tech Blog, 2013.
9. X. He, L. Liao, H. Zhang, L. Nie, X. Hu, and T. S. Chua, “Neural collaborative filtering.” In Proc. 26th Int. World Wide Web Conf. (WWW), 2017, pp. 173-182.

RESEARCH JOURNEY International Multidisciplinary E-Research Journal
Impact Factor (SJIF) - 6.625 | ***Special Issue 389 :*** *Advanced Computing and Artificial Intelligence (ACAI)*
ISSN : 2348-7143
March 2026

Emotion Recognition from the audio using Machine Learning techniques: A comparative study

Aelis Kothiya, Hetvi Vamja, Divya Ghori, Rakesh R. Savant
Babu Madhav Institute of Information Technology, Uka Tarsadia University Bardoli, India

Abstract

Speech can be defined as one of the most expressive and natural types of human communication that not only covers the linguistic meaning, but also reflects the emotional state. Automatic Speech Emotion Recognition (SER) is a field that seeks to recognize human emotions based on speech signals through the analysis of the acoustic features and has been gaining growing significance because of its uses in human-computer interaction, virtual assistants, customer service systems, and mental health monitoring. However, recognition of emotions is not always accurate because of the differences in the speaking styles, tone, intonation, and personal distinctions. In this study, the Speech Emotion Recognition system is classical Machine Learning based system. A hybrid collection of acoustic characteristics, i.e. Mel-Frequency Cepstral Coefficients (MFCC), delta and delta-delta coefficients, energy-based characteristics, spectral representations, zero-crossing rate and pitch is derived on speech signals. The standardization of features is done before classification and an SVM classifier with Radial Basis Function (RBF) kernel is used to classify emotions. The experiments are conducted on a publicly available Kaggle Voice Emotion Dataset with 6 emotional classes, which are anger, happiness, sadness, fear, disgust, and neutral. Cross-validation- Five cross-validation is done to determine strength and contribution to features. The effectiveness of the proposed SVM-based system coupled with well-contracted acoustic features in performing the speech emotion recognition tasks is proved by experimental results of reliability and consistent performance.

Keywords: Speech Emotion Recognition, Support Vector Machine, MFCC, Acoustic Features, Machine Learning, Audio Classification.

1. Introduction

Speech is regarded as one of the most expressive and natural types of human communication that expresses not only linguistic information but also expresses emotional states [5]. The human emotions of happiness, sadness, anger, fear, disgust, and neutrality are channeled through the change in the pitch, tone, energy, and speaking rate [3,6,19]. Speech emotion recognition, also known as Speech Emotion Recognition (SER), has received a lot of attention over the past years because of its extensive applications in human-computer interaction, intelligent virtual assistants, customer service analytics, call-center monitoring, e- learning platforms and mental health assessment systems [3,11,14]. Speech Emotion Recognition is the process that involves acoustic signal character of sounds in speech to label emotional states [3]. Even with the intensive research on the topic, it is not easy to determine precisely if a speech shows emotion or not. The expression of emotions differs greatly among speakers because of the tonal differences, style of speaking, accent, gender, cultural background and recording situations [6,10,11]. Also, the noises in the background and the restrained access to labelled emotional speech data add to the complications linked with the emotion recognition process [11,17]. As such the creation of dependable SER systems will necessitate strong feature extraction methods and useful classification algorithms that are able to generalize sufficiently across different speech conditions [3,10]. The classical machine learning algorithms are vital in the development of SER systems, especially when dealing with moderate-size datasets [11,14]. Support Vector Machines (SVM), the Random Forest, K-Nearest Neighbors (KNN), Logistic Regression and Naive Bayes are some of the techniques that have been used to perform the emotion classification tasks [6,7,9,13]. Such methods are based on the well-designed acoustic characteristics such as Mel-Frequency Cepstral Coefficients (MFCC), energy-based characteristics, pitch and spectral characteristics to extract emotional patterns within the speech cues [3,8,10,18,19]. SVM is one of such methods, which have been repeatedly cited as effective classifier because of its high generalization 9 and the capacity to operate in high dimension feature space [1, 2]. The present research is a comparative study of traditional machine learning algorithms to Speech Emotion Recognition on the basis of complete set of acoustic features derived out of the speech signals [6,11,14]. Various classifiers are run and analyzed based on common performance metrics, such as accuracy, precision, recall, F1-score, and confusion matrices, which include Support Vector Machines (SVM), Random Forest, K-Nearest Neighbors (KNN), Logistic Regression, and Naive Bayes [6,9,13]. According to the comparative analysis, SVM shows its better performance and is thus chosen to be analyzed in more details and experimented further [1,7,14]. The results of this work are useful in the design of the effective and understandable SER systems that are able to be implemented into the real-life practice

efficiently [11,17].

2. Related Work

Old and recent researches in Speech Emotion Recognition (SER) are mainly aimed at acoustic features of speech signals, the most frequently used feature being Mel-Frequency Cepstral Coefficients (MFCC) that are more effective in the modelling of human auditory perception [3,10,16]. A number of researchers have developed MFCC-based representations with more prosodic and spectral information including energy, zero-crossing rate, spectral centroid, and pitch to better emotion discrimination [8,9,18,19].

Classical machine learning algorithms that have been substantially studied within the context of emotion recognition include K-Nearest Neighbours (KNN), Naive Bayes, Logistic Regression, Random Forest, and Support Vector Machine (SVM) as far as classification methods are concerned [6,9,13,14]. Support Vector Machine is one of such techniques and has always performed well in SER since it is capable of dealing with high-dimensional feature-spaces, as well as nonlinear decision boundaries [1,2,14]. Experiments that used Radial Basis Function (RBF) kernel are said to achieve higher classification rates than linear classifiers especially when the acoustical features are nonlinearly separable [1,7].

Some comparative works have noted that SVM tends to perform better than other conventional classifiers in the case they are trained on well-engineered sets of acoustic features [6,13,14]. These results are able to prove that SVM is a strong foundation model among the speech emotion classification when working with classical machine learning-based SER.

Although successful, current SER strategies are characterized by a number of shortcomings. Several researches use a small number of acoustic characteristics that might not be able to get the whole picture of the emotional speech patterns [3,11]. Also, variability of speakers, background noise, and variability in recording conditions are known to have a major influence on model robustness and capacity to generalize [10,17]. The significance of feature scaling is also not mentioned in some existing works and this hinders the performance of the margin-based classifiers e.g., SVM to use heterogeneous acoustic features [1,2]. Moreover, bias in classification and poor performance on minority emotion classes is common because of the use of unbalanced datasets without proper weighting of the classes [6,14].

Unlike the other works, the given work will entail comparative study of the traditional machine learning classifiers within the area of Speech Emotion Recognition and then move on to the explanation of a superior Support Vector Machine (SVM)-based model [6,14]. Some classifiers are applied and tested on a set of typical acoustic features and SVM still remains superior to other traditional approaches. The analysis has reached the combination of the complete set of acoustic features, which are MFCC and delta and delta-delta MFCC, RMS energy, zero-crossing rate, spectral centroid, spectral bandwidth, spectral rolloff and pitch unlike the current methods which only use a few features [3,8,10,18]. Such more feature space enables one to model the spectral and prosodic variation more effectively than in the case of emotional speech. In addition, the process of standardizing features is done explicitly before the process of classification is done to balance the contribution of features, and to ensure stability of learning, especially with margin-based classifiers such as SVM [1,2]. The selected SVM model is grounded on the RBF kernel with equal weight of classes that is effective not only to handle nonlinearity but also an imbalance between classes. All this enhances the proposed Speech Emotion Recognition system and has a greater potential to generalize to other scenarios than what was previously reported by the traditional machine learning techniques.

3. Methodology

This section outlines the entire methodology which has been embraced in developing the proposed Speech Emotion Recognition (SER) system. The methodology includes dataset preparation, audio preprocessing, acoustic feature extraction, normalization of features, comparison of the results with the classical machine learning classifiers, comparative evaluation with a Support Vector machine classifier and evaluation of the system.

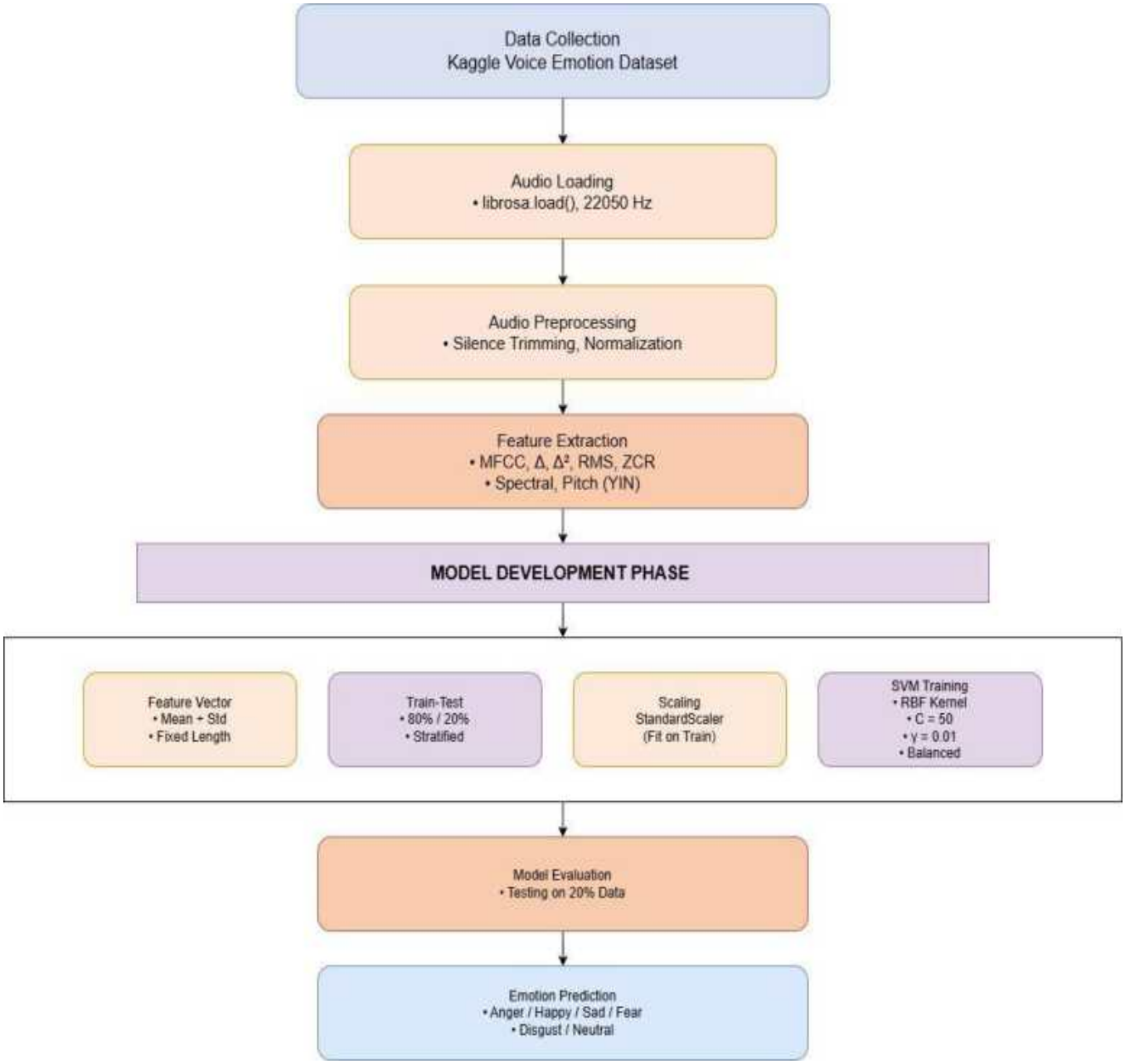


Figure 1 Block diagram of the proposed SVM-based speech emotion recognition system.

3.1 Dataset Description

The experiments on this research are done using a publicly available speech emotion dataset, which was downloaded through Kaggle [12]. The data set includes speech samples that are grouped into six different emotional categories, namely, anger, happiness, sadness, fear, disgust, and neutral. All the audio files are given in WAV and each emotion kept in a separate directory. The data consists of short speeches of various speakers, so it is varied in terms of voice and the ability to reveal emotions [3]. The data is fairly balanced in terms of emotion categories, and it helps to minimize classification bias and enhances model generalization [6]. All audio samples are sampled at the standard sampling rate and can be used in the supervised machine learning experiments.

3.2 Audio Preprocessing

The audio file is loaded and processed with the help of the Librosa library in order to give all the samples of speech consistency. All the audio signals are re-sampled to 22,050 Hz which is a trade-off between the calculation speed and spectral detail. Normalization of amplitude is employed to reduce variations that are brought about by different recording conditions and microphone gains. The full-length audio signals are also used in extracting features so as to preserve the information on emotions that may be located during utterance. No manual segmentation

is required so that the system may be easily extended to the speech samples in the real world without the further complication of preprocessing.

3.3 Feature Extraction

Another important element of the SER system is feature extraction because the classifier is based on the meaningful acoustic representations to discern emotional states [3,14]. This is a study whereby a set of low-level and high-level acoustic features is obtained in each speech signal.

Mel-Frequency Cepstral Coefficients (MFCCs)

MFCCs are popular features in speech emotion recognition as it is a way of how human beings listen to sound [3,14,15]. MFCCs extract significant frequency content in speech which varies with varying emotions. In this paper, 20 MFCC features are obtained using every speech signal. Delta and delta-delta MFCCs are also utilized to comprehend the ways in which emotions evolve with time. These characteristics display the way in which speech characteristics stop and switch quickly [3]. The combination of MFCCs and their derivatives does give valuable information to differentiate various emotional states in speech and can be used together with the SVM classifier.

Energy-Based Features

Root Mean Square (RMS) energy is also taken so that it represents the total strength of the speech signal. The characteristics of energy can be used to differentiate between emotions that are more energetic like anger and happiness and those whose features are not that energetic like sadness and neutral emotions [6,8].

Spectral Features

The distribution of frequency contents of a speech signal is characterized by spectral features. The emotions could be distinguished by the tonality, sharpness, and the energy distribution of the speech differing in various emotions [3,10].

Zero Crossing Rate (ZCR)

ZCR is the count of the number of times speech signal alters between positive and negative. It is the inarticulateness or the acuity of language [10]. Anger, happiness, fatalism tend to possess greater ZCR value because the speech is sharper and more energetic as compared to sadness and neutral emotions which have lower ZCR values.

Spectral Centroid

Spectral centroid is a mode of the frequency spectrum which corresponds to the center of mass and usually correlates to the speech brightness [3]. Greater centroid values represent brighter and more intense speech that is typically connected with such feelings as anger and excitement.

Spectral Bandwidth

Spectral bandwidth is used to measure the distribution of frequencies around the centroid and it is applied in emotion discrimination [10]. It shows the frequency ranges of speech. Emotional speech has a tendency of displaying broader bandwidth relative to neutral speech.

Spectral Rolloff

The frequency at which the bulk of the spectrum energy is concentrated is known as spectral rolloff, and it helps in distinguishing between relaxed and paramount excited emotions [3]. It is applied in distinguishing high and low frequency distribution of energy and is applicable when one needs to distinguish between calm emotions and highly excited emotions. On the one hand, these spectral properties can gather important frequency-specific properties of speech, and when combined with other acoustic characteristics, can be utilized in allowing the SVM classifier to distinguish the distinctions in different emotional conditions with great precision.

Pitch Features

Pitch is a sound of the voice whether it is high or low and much associated with expression and emotion [20]. The high and fluctuating pitch accompanies arousal like anger, fear and happiness but the low and constant pitch accompanies sad and neutral arousal. In this work pitch has been extracted using the YIN algorithm. The standard deviation and the average of the pitch are used to represent the change of the pitch which helps the SVM classifier to differentiate the different emotion states.

Statistical Feature Representation

Descriptive statistics such as the mean and standard deviation are obtained on the features extracted with regards to the aggregate length of the speech signal. The impact of this technique is that the feature vectors of every sample of audio are preset, which is suitable with conventional machine learning classifiers.

3.4 Feature Vector Formation

All the features derived are concatenated to generate one large dimensional feature vector of each speech sample. MFCCs, delta and delta-delta coefficients, energy, spectral and pitch statistics make up this category of features. The obtained feature vectors mirror the spectral and the prosodic features of the emotional speech.

3.5 Feature Normalization

Before model training, features are scaled with StandardScaler that scales features to unit variance and zero mean. The Support Vector Machines are sensitive to the magnitude and size of features and as such, they require feature scaling [1,2]. Normalization ensures that all features share similar values with the process of classification.

3.6 Data Splitting Strategy

Table 1 Effect of different train–validation–test split ratios on SVM classification accuracy

Train (%)	Validation (%)	Test (%)	Accuracy
70	15	15	0.81254
80	10	10	0.831529

3.7 Classification

Within the framework of this paper, Support Vector Machine (SVM) classifier is used in the recognition of emotion in speech [1,2]. The rationale behind the selection of SVM is that it can be efficiently operated with high-dimensional acoustic features and the fact that its overall capability of generalization is high despite the fact that the size of the training data is small [14].

The parameters of SVM classifier are configured like the following: Kernel: Radial Basis Function (RBF), Regularization parameter (C): 50, Kernel coefficient (gamma): 0.01, Class weighting: Balanced. The rationale of making use of SVM is that, it possesses a high generalization capacity and can process high-dimensional spaces of the acoustic features. Information of emotional speech is traditionally nonlinear and cannot be divided in a one-dimensional way hence, the RBF kernel is applied to describe nonlinear decision boundaries that are required to distinguish among different emotional patterns. The regularization parameter C will be 50 which is the tradeoff variable between the maximization of the margin and the minimization of classification errors. The larger C the less will be the training errors, and the generalization will not be poor. The rationale behind the selection of this value was to generate accuracy and overfitting of the desired data. The coefficient of kernel gamma is set to 0.01 that forms the effect of one training sample. A lesser gamma is result in the decision boundaries being smoother and also the model lacks the chance to adapt the noise in the data. The environment helps SVM to generalize emotional tendencies of speech rather than memorizing some samples. The weight of the classes will be adjusted in order to correct the variations in the sample of the classes having different emotions. The balance of the classes by the classifier ensures that the overrepresented classes of emotion are not given favoritism, and gives more attention to the underrepresented groups. Overall, the selected SVM configuration provides a reasonable balance between the complexity and model generalization, and it is reasonably applied to carry out speech emotion recognition with manually designed acoustic features. On the whole, the chosen SVM architecture offers an efficient compromise between the complexity of the model and its generalization, which is why it is a good choice to use it in the context of speech emotion recognition with handcrafted acoustic features.

3.7.1 Cross-Validation Strategy

In order to further measure the robustness and the generalization power of the proposed SVM-based speech emotion recognition system, five-fold cross-validation on the training dataset was done [6,14]. The training data are divided into 5 equal subsets. Each iteration consisted of four subsets where one is trained and the other one is validated. The cross-validation performance was expressed in terms of mean and standard deviation of the classification accuracy over all the folds. The proposed SVM classifier obtained an average accuracy of 80.61% + 0.22 which means that the accuracy is stable and consistent when used in varying data partitions. The low standard deviation also ensures that the model is not sensitive to a specific train test split and its ability to resist a specific train test split proves that it is a robust model when it comes to speech emotion recognition tasks. Final performance evaluation was only done on the held-out test set.

4. Results and Discussion

We compared the performance of classical machine learning classifiers on measures of accuracy and confusion-matrix in terms of True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN) as shown in

Table 2 . To determine the best model to use in speech emotion recognition based on the extracted acoustic features, several classical machine learning classifiers are under consideration, namely, K-Nearest Neighbors (KNN), Random Forest, Logistic Regression, and Naive Bayes. The Support Vector Machine (SVM) is one of these classifiers and it is continually delivering high accuracy in classification. In the light of this comparative analysis, SVM is being chosen as the ultimate classifier, which can be analyzed and experimented with.

Table 2 Performance comparison of classical machine learning classifiers for speech emotion recognition.

	Classifier Name	Accuracy	TP	TN	FP	FN
1	Support Vector Machine (SVM)	83.15%	9,062	52,654	1,836	1,836
2	K-Nearest Neighbours (KNN)	68.87%	7,506	51,098	3,392	3,392
3	Random Forest	75.30%	8,207	51,799	2,691	2,691
4	Logistic Regression	56.349%	6,141	49,733	4,757	4,757
5	Naïve Bayes	27.93%	3,044	46,636	7,854	7,854

According to

Table 2 , SVM model is achieving a true positive of 9,062 and a true negative of 52,654 and the false positive and false negative are relatively small (1836 and 1836 respectively). These findings are pointing to the fact that the SVM classifier is indeed separating speech patterns of emotion with minimal misclassification. According to Table 2 , the second-best performing model is the Random Forest classifier with the accuracy of 75.31%. Despite the advantages that Random Forest is enjoying due to the nature of ensemble learning and noise resistance, the performance of Random Forest is still being outperformed by SVM especially when it comes to dealing with overlapping distributions of emotional features. The accuracy of the K-Nearest Neighbors (KNN) classifier is 68.88 as it is indicated in Table 2 . The relatively large values of false positives and false negatives indicate that distance-based classifiers are having difficulties with the high dimensional space of acoustic features employed in this experiment. The linear classifier Logistic Regression is attaining an accuracy of 56.35 meaning that it has limited ability to model nonlinear relationships of the features of emotional speech. Likewise, the Naive Bayes classifier is getting the worst accuracy of correlated acoustic attributes like MFCCs, spectral descriptors and pitch-related features.

Table 3 is showing the performance comparison of Support Vector Machine (SVM) classifier with the various sets of acoustic features. In cases where the use of only Mel-Frequency Cepstral Coefficients (MFCC) is concerned, the model is recording an accuracy of 75.0, which implies that MFCCs are useful in the presentation of crucially important spectral features of emotional speech. Nevertheless, in the event that the MFCC features are applied alongside more acoustic features including delta and delta-delta coefficients, energy-based features, spectral descriptors and pitch related features, classification accuracy is rising to 83.0%. This performance improvement shows that the incorporation of the complementary prosodic and spectral input is increasing the discriminative ability of SVM classifier. The findings are used to validate that speech emotion recognition is best done using a hybrid acoustic feature representation as opposed to using MFCC features only.

Table 3 Performance comparison of different feature sets using SVM.

Table 3 Performance comparison of different feature sets using SVM.

Features Added	Accuracy Gain
MFCC (baseline)	78.45%
+Spectral+Pitch	**+4.7%**
Total	83.15%

Confusion Matrix Analysis

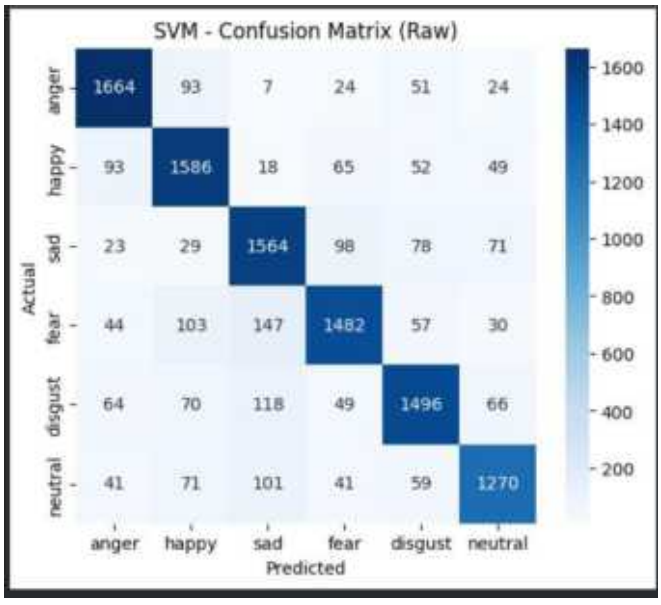


Figure 2 Raw confusion matrix of the SVM classifier for speech emotion recognition.

The confusion matrix in Figure 2 shows that the Support Vector Machine (SVM) classifier is performing better classification on all the emotions classes. The anger group is registering 1,664 samples that are appropriately categorized with misrepresentation mostly with happy (93) and disgust (51). The happy moods emotion is making 1,586 correct predictions and minimal misclassification with anger

(93) and fear (65). In the sad group, correct identification is being only made of 1,564 samples but confusion with fear (98) and disgust (78) is being experienced. The fear emotion is getting 1,482 true positive classifications with evident confusion in the acoustic features with sad (147) and happy (103). The disgust category has 1,496 properly identified samples, and moderate error of classification is being experienced at sad (118) and happy (70). The neutral category has been getting 1,270 right classifications, which is relatively less than other emotions, and is mostly getting mixed up with sad (101) and happy (71). In general, the fact that most of high diagonal values in the confusion matrix in Figure 2 are true is validating the suitability of SVM classifier in speech emotion recognition. The vast majority of the misclassifications are occurring due to the acoustic similarity between classes which are emotionally related, and the classifier is very robust and consistent across the entire range of emotions.

Cross-Validation Analysis

In order to further test the strength of the proposed SVM-based speech emotion recognition system, the training data were cross-validated using five folds. The average classification accuracy of the classifier was 80.61% with a standard error of 0.22%. The small consistency between folds implies that the model does not respond to a particular train test split. These findings also support the credibility and strength of the suggested method of speech emotion recognition.

5. Conclusion

This paper has given a detailed discussion of traditional machine learning methods of Speech Emotion Recognition through handcrafted acoustic features. An extensive set of spectral and prosodic attributes such as MFCCs of delta and delta-delta values, energy features, spectral values, zero-crossing rate, and pitch were obtained to capture the ability to predict emotional attributes in speech signals. Normalization of features was carried out so as to have equal contribution of features and model learning. Several classical classifiers such as K-Nearest Neighbors, Random Forest, Logistic Regression, and Naïve Bayes have been tested on the basis of standard performance measures. These methods include the Support Vector Machine with a Radial Basis Function kernel that proved to be more effective in terms of accuracy and consistency of classification. The confusion matrix analysis also indicated that the SVM model was successful in distinguishing the various classes of emotions with majority of the misclassifications happening between acoustically similar emotions. Five-fold cross-validation was used, on the training data, to evaluate the generalization capability and robustness. The fact that cross-validation accuracy did not vary very much ensured that the proposed system based on SVM was stable and did not depend on a particular data partition. These results suggest that conventional machine learning methods along with effective acoustic feature set and adequate normalization can deliver credible results with regard to speech emotion recognition tasks. In general, the findings indicate that the suggested SVM-based system offers a promising and understandable solution to speech emotion recognition, so it may be adopted in practice in case of computational efficiency and model explainability concerns. The next step in speaker-independent evaluation, noise-resilient feature representations and hybrid methods may be investigated in the future to further improve the recognition performance.

References :

1. C. Cortes and V. Vapnik, “Support-vector networks,” Machine Learning, vol. 20, no. 3, pp. 273–297, 1995.
2. V. N. Vapnik, Statistical Learning Theory. Wiley-Interscience, 1998.
3. D. Ververidis and C. Kotropoulos, “Emotional speech recognition: Resources, features, and methods,” Speech Communication, vol. 48, no. 9, pp. 1162–1181, 2006.
4. B. Schuller, S. Steidl, and A. Batliner, “The INTERSPEECH 2009 Emotion Challenge,” Proceedings of INTERSPEECH, pp. 312–315, 2009.
5. A. Mehrabian, “Communication without words,” Psychology Today, vol. 2, no. 4, pp. 53–56, 1968.
6. K. S. Rao, T. P. Kumar, K. Anusha, B. Leela, I. Bhavana, and S. V. S. K. Gowtham, “Emotion recognition from speech,” International Journal of Computer Science and Information Technologies, vol. 3, no. 2, pp. 3603–3607, 2012.
7. P. Sharma, V. Abrol, A. Sachdev, and A. D. Dileep, “Speech emotion recognition using kernel sparse representation-based classifier,” Proceedings of the European Signal Processing Conference (EUSIPCO), pp. 374–377, 2016.
8. S. R. Bandela and T. K. Kumar, “Stressed speech emotion recognition using feature fusion of Teager Energy Operator and MFCC,” Proceedings of the International Conference on Computing, Communication and Networking Technologies (ICCCNT), pp. 1–5, 2017.
9. A. A. Zamil et al., “Emotion detection from speech signals using voting mechanism on classified frames,” Proceedings of the International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST), IEEE, 2019.
10. D. Kaminska, T. Sapinski, and G. Anbarjafari, “Efficiency of chosen speech descriptors in relation to emotion recognition,” EURASIP

Journal on Audio, Speech, and Music Processing, 2017.

11. L. Kerkeni, Y. Serrestou, M. Mbarki, K. Raoof, M. A. Mahjoub, and C. Cleder, "Automatic speech emotion recognition using machine learning," social media and Machine Learning, 2019.
12. S. Deogade, "Voice Emotion Classification Dataset," Kaggle. Available: https://www.kaggle.com/datasets/sdeogade/voice-emotion-classification
13. F. Burkhardt, A. Paeschke, M. Rolfes, W. F. Sendlmeier, and B. Weiss, "Emotion detection in acted speech using multiple classifiers," IEEE Transactions on Affective Computing, vol. 1, no. 1, pp. 41–52, 2010.
14. B. Schuller, A. Batliner, S. Steidl, and D. Seppi, "Recognising realistic emotions and affect in speech: State of the art and lessons learnt," Speech Communication, vol. 53, no. 9–10, pp. 1062–1087, 2011.
15. F. Eyben, M. Wöllmer, and B. Schuller, "OpenSMILE: The Munich versatile and fast open-source audio feature extractor," Proceedings of ACM Multimedia, pp. 1459–1462, 2010.
16. C. M. Lee and S. S. Narayanan, "Toward detecting emotions in spoken dialogs," IEEE Transactions on Speech and Audio Processing, vol. 13, no. 2, pp. 293–303, 2005.
17. A. Nogueiras, A. Moreno, A. Bonafonte, and J. B. Mariño, "Speech emotion recognition using hidden Markov models," Proceedings of Eurospeech, pp. 2679–2682, 2001.
18. T. S. Polzin and A. Waibel, "Emotion-sensitive human–computer interfaces," Proceedings of the ISCA Workshop on Speech and Emotion, 2000.
19. F. Dellaert, T. Polzin, and A. Waibel, "Recognizing emotion in speech," Proceedings of the International Conference on Spoken Language Processing (ICSLP), pp. 1970–1973, 1996.
20. C. Busso, S. Lee, and S. S. Narayanan, "Analysis of emotionally salient aspects of fundamental frequency for emotion detection," IEEE Transactions on Audio, Speech, and Language Processing, vol. 17, no. 4, pp. 582–596, 2009.

RESEARCH JOURNEY International Multidisciplinary E-Research Journal
Impact Factor (SJIF) - 6.625 | ***Special Issue 389 :*** *Advanced Computing and Artificial Intelligence (ACAI)*
ISSN : 2348-7143
March 2026

Deep Learning–Based Feature Extraction with Machine Learning for Multi-Class Waste Classification

Meet Ghasadiya, Shruti Hansaliya, Hiren Hadiya, Rakesh R. Savant
Babu Madhav Institute of Information Technology, Uka Tarsadia University, Bardoli, India

Abstract

Waste sorting is a key to effective recycling and environmental management, yet manual waste sorting is time-consuming, and error-prone, and an automated image-based system of waste sorting is necessary. This experiment explores waste classification into multi-classes based on a publicly available Garbage Classification V2 dataset from Kaggle (13,347 images in ten categories). Initial evaluation of traditional machine learning is done on 128 x 128 images with handcrafted features such as RGB + HOG, HSV + HOG, HSV + LBP + HOG, HSV + GLCM and SIFT and ORB with Bag of Visual Words representations with SVM, KNN, Random Forest, Naive Bayes as the classification methods, which made performance limited with a maximum accuracy of 60.09%. To improve this low accuracy, a hybrid method is used where the deep features are obtained on top of trained ResNet50 and EfficientNetB0 models and then machine learning classifiers are used to classify them. The hybrid model helped to get the best results with EfficientNetB0 + SVM with the highest accuracy of 94.56%, which indicates that deep feature extraction and machine learning classifiers are good and effective in classifying waste with the help of artificial intelligence.

1. Introduction

The high rate of urbanization and population increase has caused a big rise in the production of solid waste all over the world posing serious environmental and human health problems [16]. It is approximated that over 2 billion tons of municipal solid waste is produced annually worldwide and it is likely to increase exponentially within the next few decades [16]. Poorness of waste segregation leads to overloading of landfills, environmental degradation, and waste of recyclable materials. Proper waste classification is thus needed to enhance an efficient way of recycling, lessening the degradation of the environment and also encouraging proper waste management practices [17]. The rapid urbanization, growth in consumption habits and population density all pose an even more urgent waste management issue in India. In India, the Central Pollution Control Board (CPCB) claims that the municipal solid waste (MSW) produced locally amounts to more than 160,000 tons/day, with a very large portion of this in its untapped form [18]. Inappropriate segregation impairs recycling, creates a heavy load on landfills, and also pollutes the environment in the urban centers [19]. In response to them, domestic policies like the Swachh Bharat Abhiyaan and the Smart Cities Mission focus on sustainable waste management, better recycling, and the introduction of smart technologies to clean cities and keep them environmentally sustainable [20].

Segregation at the source is still a significant problem since type of segregations depends on manual sorting which is very laborious, time consuming and subject to human error. Ineffective sorting does not only minimize the intensity of recycling but also raises costs of operations. As computer vision and artificial intelligence technologies have increased, automated waste classification systems have become a potential solution to enhance the accuracy of segregation and provide smart recycling systems [1].

Classical methods of image classification are based on manual procedures of feature extraction color, texture and shape features. Such methods as Histogram of Oriented Gradients (HOG), Local Binary Patterns, Gray-Level Co-occurrence Matrix (GLCM) as well as local descriptors like SIFT and ORB have been extensively combined with machine learning classifiers like Support Vector Machine, K-Nearest Neighbors, Random Forest (RF) and Naive Bayes (NB). Even though these techniques are able to complex visual features, they tend to fail in making a distinction between complicated waste types because of different changes in shapes, textures, light, and background noise. The recent innovations in the field of deep learning have tremendously enhanced the field of image classification, as they have taught machines to recognize significant visual representations of images and therefore learn these patterns automatically. ResNet50 and EfficientNetB0 are examples of Convolutional Neural Networks (CNNs) that have performed remarkably when it comes to visual recognition. However, deep networks need extensive datasets and compute resources to be trained directly. In order to overcome this, hybrid methods mixing deep feature extraction with machine learning classifiers such as Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Random Forest (RF), Naive Bayes (NB) and Logistic Regression (LR) are gaining popularity due to their high accuracy with low training complexity [15].

This paper conducts a comparative study of both traditional feature-based machine learning methods using handcraft feature and a hybrid deep learning with machine learning system in the context of classifying waste images into different categories. First, the classical feature extraction

methods and Bag of Visual Words representations are tested with several machine learning classifiers. Then, deep features of the pretrained ResNet50 and EfficientNetB0 models are identified and the model is detected by applying machine learning algorithms to enhance the results. The key goal of the work is to find out a correct and effective automated method of waste classification, which can be used to operate intelligent recycle systems and efficient waste management in urban areas.

2. Literature Review

Advanced waste sorting has been an issue of great interest during the last few years because of the increasing demand of efficient recycling and waste management. The original studies in waste classification focused mainly on the usage of the traditional image processing and machine learning methods. These methods make use of handcrafted features like color histograms, texture descriptors and shape features to describe waste materials. Histogram of oriented gradient (HOG) [6], Local Binary Patterns (LBP) [7], and Gray-Level Co-occurrence Matrix (GLCM) [8] are the most popular feature extraction techniques to extract both the texture and structure of waste in images. Support Vector Machine (SVM) [10], K-Nearest Neighbors (KNN) [12], Random Forest (RF) [11], and Naive Bayes (NB), Logistic Regression are the common machine learning algorithms that were used to classify these features. Although these techniques performed moderately in terms of classification, they are also frequently constrained by changes in lighting conditions, object shape and background complexity.

Local feature descriptors like Oriented FAST and Rotated BRIEF (ORB) [5] and Scale-Invariant Feature Transform (SIFT) [4] have been used to augment feature representation together with Bag of Visual Words (BoVW) models [9]. These techniques transform local characteristics into visual words which aid the classifiers to identify structural designs. Although BoVW methods are scale and rotation resistant, they use manual elements and cannot be used to capture valuable semantics in waste images.

In image classification problems, CNNs have shown to be more efficient with the development of deep learning. The CNN-based architectures learn hierarchical features representations with raw images eliminating the necessity of manually engineering features. Other authors have used deep learning networks like AlexNet [13], VGG, ResNet50 [2] and EfficientNetB0 [3] to classify wastes and found that the networks are significantly better than traditional machine learning models. Yet, training deep neural networks in their full form is expensive in both time and memory, with large labelled datasets and large computation requirements, which are necessarily not always available.

In order to overcome these issues, recent studies have investigated the integration approach through which deep learning feature extraction is coupled with traditional machine learning classifiers. In these systems, deep features are obtained using trained CNN models and a machine learning algorithm (e.g., Support Vector Machine, Random Forest, K-Nearest Neighbours, Naive Bayes or Logistic Regression) is employed. The strategy takes advantage of the representational power of the deep learning but minimizes computational complexity and training time. Hybrid frameworks have demonstrated good outcomes in diverse picture classification tasks such as environmental supervision and trash identification [15].

Though these improvements have been made, it is still not feasible to obtain high accuracy with different types of waste since they have different textures, shapes and lighting conditions, and clutters in the background. Hence, additional research is required to determine the efficacy of conventional handcrafted features, local descriptors, and hybrid deep features to multi-class waste classification. To fill this gap, the research paper present a thorough comparative study of classical machine learning algorithm and hybrid deep learning-machine learning algorithms using deep feature extractors like the ResNet50 [2] and EfficientNetB0 [3], to determine an efficient and precise solution to automated waste segmentation.

3. Methodology

This paper basically proposes a combination model of deep learning and machine learning for classifying the images of waste in different categories. The methodology includes data sets prep, preprocessing, feature extraction, classification, performance evaluation. We compared the traditional machine-learning techniques where handcrafted features are used and a hybrid approach is used that considers deep feature extraction.

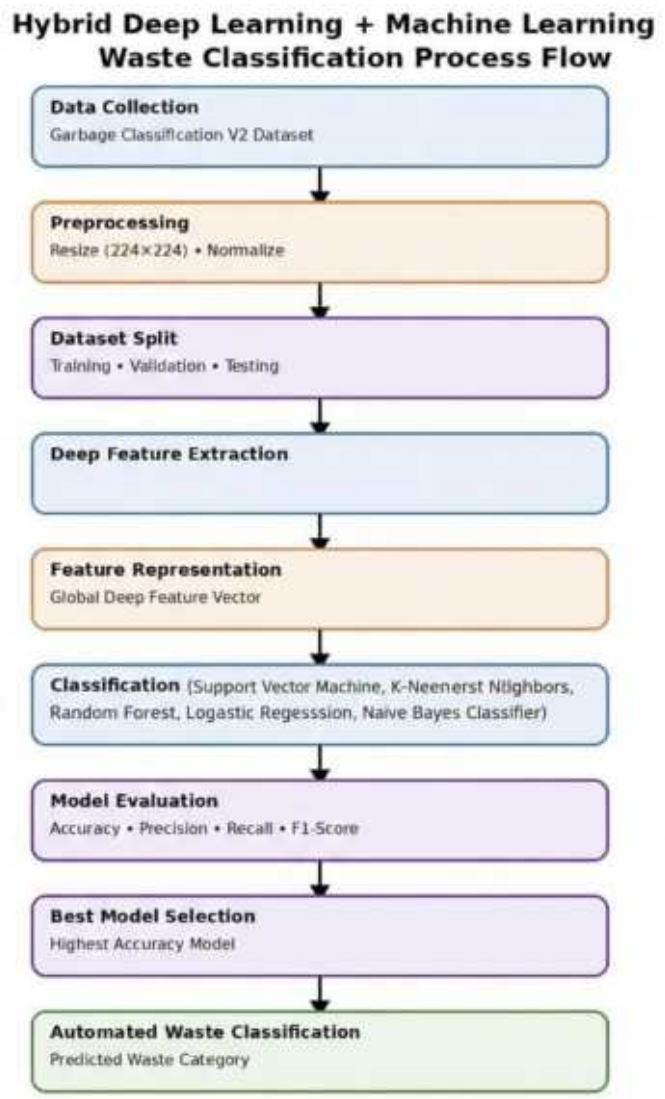


Figure 1. Waste Classification System overview

3.1 Dataset Description

The tests are carried out using publicly accessible Garbage Classification V2 data set that was retrieved on Kaggle [14]. The dataset consists of 13,347 photographs that classified into ten categories of waste, namely: plastic, paper, cardboard, glass, metal, biological waste, battery, trash, clothes, and shoes is shown in Table 1. The data is divided into training (70%), validation (15%), and testing (15%) data to obtain a good estimation of our model.

Table 1. Class-wise Distribution of Garbage Classification V2 Dataset on Training, Validation and Test Set

Categories	Train	Validation	Test	Total
battery	569	122	123	814
biological	567	121	122	810
cardboard	1073	230	230	1533
clothes	1388	298	298	1984
glass	1398	300	300	1998
metal	695	149	149	993
paper	965	208	207	1380
plastic	1196	256	257	1709
shoes	1143	245	245	1633
trash	345	74	74	493
Total	9339	2003	2005	13347

To ensure consistency of the models and enhance their performance, the preprocessing steps as indicated below:

- Elimination of invalid or unreadable pictures.
- Image resizing
  -128 * 128 pixels on the traditional feature extraction.
  - Deep feature extraction 224 * 224 pixels.
- Conversion of the pixel to standard values.
- Label encoding of class representation.

Additionally, no methods of data augmentation are applied so that the performance comparison between methods could be fair.

3.2 Traditional Feature Extraction

To test the classical machine learning methods, various handcrafted and local features extraction methods are used.

Handcrafted Features

The combinations of features that are obtained so as to ensure color and texture features are RGB + HOG, HSV + HOG, HSV + LBP + HOG, HSV + GLCM, HOG and LBP are some of the popular texture descriptors that are used to extract structural and texture patterns in images.

Local descriptors BoVW Local descriptors

The local feature descriptors that are used to describe structural attributes of waste objects include RGB + SIFT + BoVW, RGB + ORB + BoVW. In order to investigate the effects of feature normalization, experiments are carried out with scaled representations of feature space and unscaled representations of feature space of SIFT + BoVW features and ORB + BoVW feature, before the classification.

3.3 Classification

Machine learning algorithms, i.e. Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Random Forest (RF), Naive Bayes (NB) and Logistic Regression (LR), are used to test the effectiveness of features in multi-class classification.

3.4 Hybrid Deep Feature Extraction

A hybrid deep learning style is used to get a better classification result. The pretrained convolutional neural networks are used instead of training a deep network to gain meaningful feature representations of the images. ResNet50 is enhanced to utilize residual connections that enhance the performance of deeper networks to learn more efficiently and reach higher accuracy. EfficientNetB0 uses compound scaling to ensure high performance and fewer parameters and better efficiency. To extract features, the pretrained models are inputted with all the images resized to 224 × 224 pixels. The global pooling layer is used to obtain feature vectors because the layer captures high level semantic data that would help in differentiating among the various waste categories.

3.5 Hybrid Classification

The deep feature vectors extracted from CNN models are classified using machine learning classifiers including SVM, KNN, Random Forest, Naïve Bayes, and Logistic Regression. This hybrid approach combines the representational strength of deep learning with the efficiency of traditional machine learning algorithms [15].

3.6 Performance Evaluation

Model performance is evaluated using metrics including Accuracy, Precision, Recall, F1-score, Confusion Matrix. These metrics provide a comprehensive evaluation of classification performance across multiple waste categories. The proposed methodology enables a systematic comparison between traditional feature-based machine learning techniques and a hybrid deep feature–based classification framework to determine an efficient and accurate approach for automated waste classification.

4. Results and Discussion

4.1 Performance Evaluation of Traditional Methods of Feature Extraction

First of all, with multiple machine learning classifiers (Support Vector Machine, K-Nearest Neighbors, Random Forest, Naive Bayes and Logistic Regression), the traditional extraction methods of feature are evaluated. The extracted features are RGB+HOG, HSV+HOG, HSV+HOG+LBP, and HSV+HOG+GLCM. The classification accuracy obtained with the use of these features is shown in Table 2.

Table 2. Classification Accuracy Using Traditional Features

Feature Method	KNN %	SVM %	RF %	NB %	LR %
RGB+HOG	46.93	46.93	50.37	33.02	7.33
HSV+HOG	45.78	58.00	55.86	34.31	7.18
HSV+HOG+LBP	46.28	60.09	57.25	34.06	9.67
HSV+HOG+GLCM	33.42	20.30	54.81	34.26	6.58

As shown in Table 2 the traditional feature extraction methods give poor classification performance. The highest accuracy, encountered with 60.09%, is obtained by using HSV+HOG+LBP features with SVM classifier. However, an average performance is still achieved because handcrafted features cannot capture high-level semantic information, existing in the waste images. The comparison of classifier performance with traditional features is given in Figure 2.

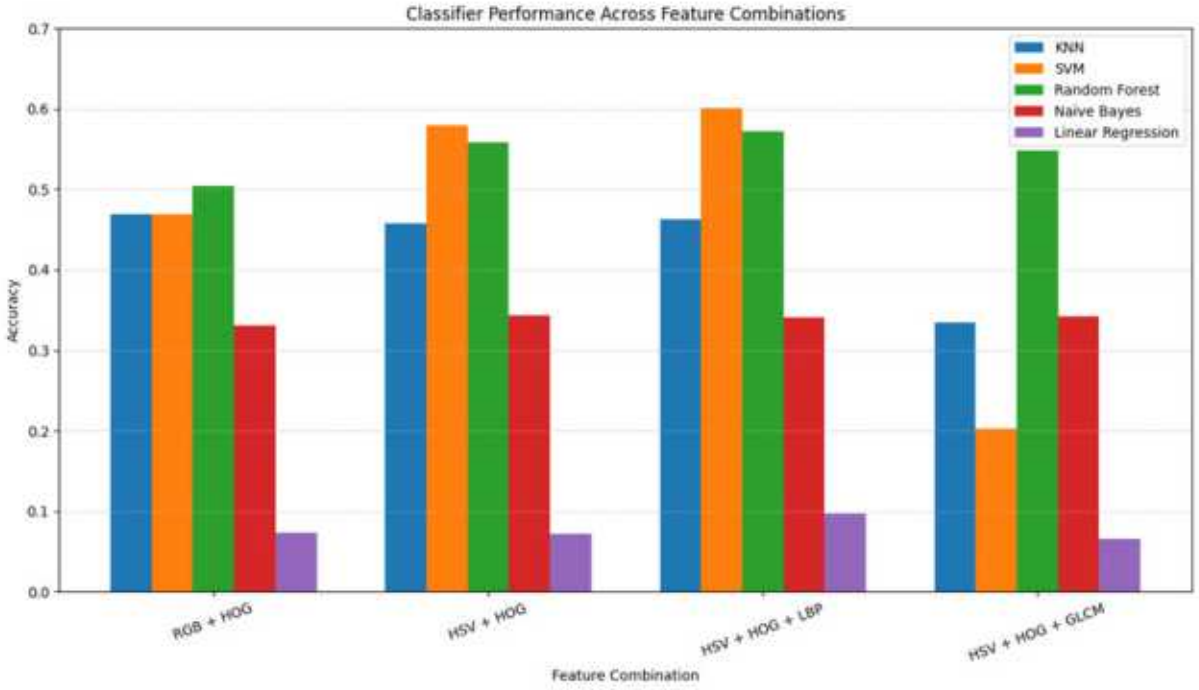


Figure 2. Test accuracy comparison of classifiers using traditional feature extraction methods

4.2 Performance Evaluation of SIFT and ORB with Bag of Visual Words

To improve feature representation, SIFT and ORB descriptors are encoded using the Bag of Visual Words (BoVW) model with 200 visual clusters. The classification accuracy obtained without feature scaling is presented in Table 3, while the results after applying feature scaling are shown in Table 4.

Table 3. Classification Accuracy Using SIFT and ORB with BOVW (Without Scaling)

Feature Method	KNN%	SVM%	RF%	NB%	LR%
RGB+SIFT+BOVW	40.19	36.14	47.83	26.03	12.98
RGB+ORB+BOVW	39.70	37.00	41.39	22.61	11.27

From Table 3, it can be seen that the best performance without scaling is a Random Forest with RGB+SIFT features (47.83%). After the feature scaling applied as shown in Table 4, the highest accuracy slightly increased to 48.02%.

Table 4. Classification Accuracy Using SIFT and ORB with BOVW (With Scaling)

Feature Method	KNN%	SVM%	RF%	NB%	LR%
RGB+SIFT+BOVW	29.42	47.43	48.02	26.03	12.01
RGB+ORB+BOVW	26.98	40.30	41.45	22.69	11.27

The graphical comparison of SIFT and ORB features without scaling can be shown in Figure 3, while the scale results can be shown in Figure 4. However, the overall accuracy is still low compared to what we would hope for because of the lack of the ability of handcrafted local features to capture complex semantic patterns.

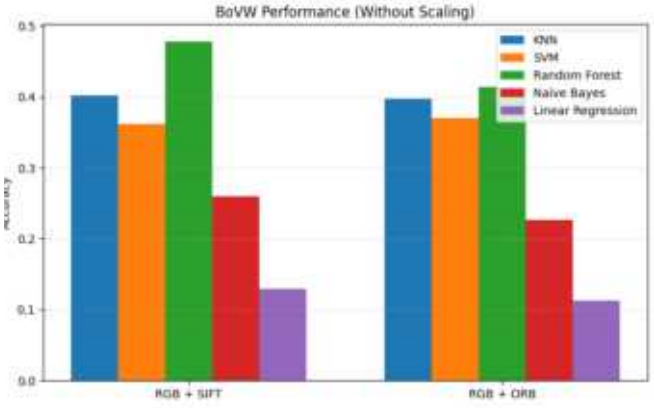


Figure 3. Classification accuracy using SIFT and ORB BoVW features without feature scaling

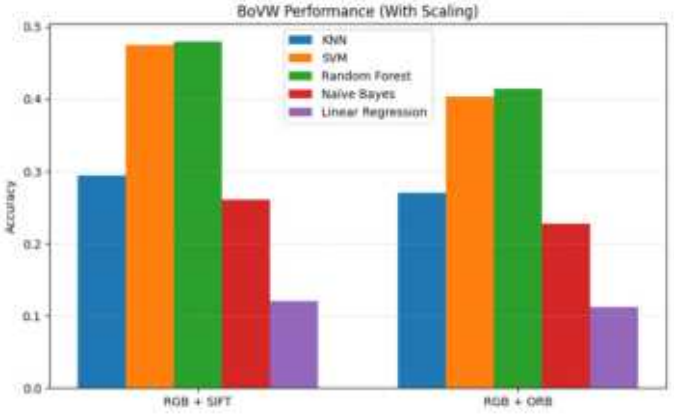


Figure 4. Classification accuracy using SIFT and ORB BoVW features with feature scaling

4.3 Performance Evaluation of Hybrid Deep Learning Feature Extraction

In order to avoid the weaknesses of conventional feature extraction schemes, deep learning architectures ResNet50 and EfficientNetB0 were employed to extract the features. Machine learning classifiers were used to classify the extracted deep features. Table 5 represents the classification accuracy of ResNet50 and EfficientNetB0 features.

Table 5. Accuracy of Classification based on ResNet50 and EfficientNetB0 Deep Features

	SVM %	KNN %	RF %	NAIVE BAYES %	LOGISTIC REGRESSION %
Resnet50	94.46	92.96	91.62	84.78	93.41%
EfficientNetB0	94.56	92.86	90.42	84.98	93.46%

The maximum accuracy of 94.46% with SVM classifier with ResNet50 features is obtained as seen in Table 5. On the same note, EfficientNetB0 is paired with SVM; it records the highest overall accuracy of 94.56%. Figure 5 shows the comparison of all the classifiers based on the deep features.

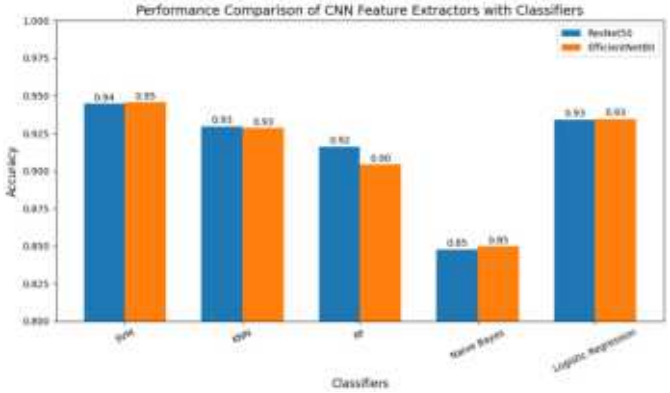


Figure 5. Comparison between machine learning classifiers based on deep feature extraction in classification accuracy

4.4 Comparative Analysis and Discussion

Based on Figure 5, it is eminent that hybrid deep learning methods of feature extraction are by far better as compared to traditional handcrafted feature methods.

Key observations:

- Traditional feature methods maximum accuracy: 60.09% (Table 2)
- SIFT+BOVW maximum accuracy: 48.02% (Table 4)
- ResNet50 highest accuracy is 94.46% (Table 5).
- EfficientNetB0 peak accuracy: 94.56% (Maximum) (Table 5)

The hybrid deep learning system enhanced the classification performance by over 34 percent over the conventional feature extractor systems. This major advancement can be attributed to the fact that deep learning models like ResNet50 and EfficientNetB0 are able to extract discriminative and high-level features in images. In contrast to the traditional handcrafted features, deep features are better in capturing intricate visual patterns, object structures and semantic information. The findings show that the test of hybrid deep learning feature extraction and machine learning classifiers is the most effective and accurate in multi-class waste image classification.

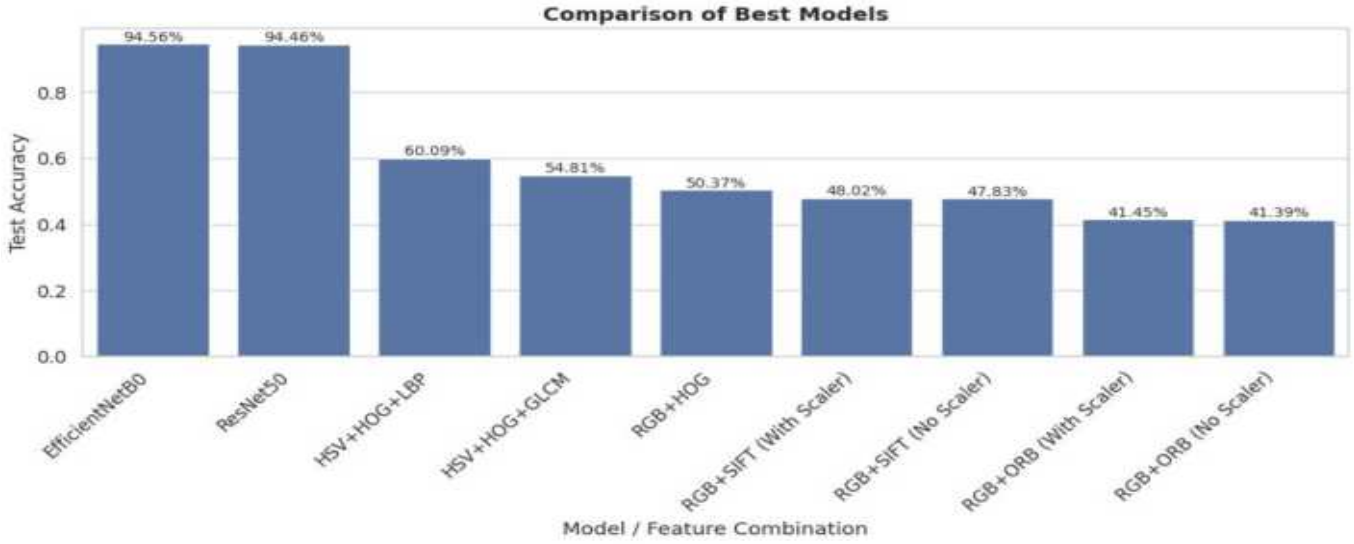


Figure 6. Comparison of test accuracy of the best-performing models from traditional feature methods, BoVW representations, and hybrid deep feature–based approaches

4.5 Class-wise Performance Analysis

Based on Table 6, it can be seen that the majority of the waste types have a precision and recall of above 0.90 which represents high levels of performance of the class.

Table 6. Classification report of the proposed EfficientNetB0–SVM model showing class-wise precision, recall, and F1-score

Class	Precision	Recall	F1-score	Support
battery	0.96	0.98	0.97	123
biological	1.00	0.97	0.98	122
cardboard	0.95	0.94	0.95	230
clothes	1.00	0.99	0.99	298
glass	0.96	0.91	0.94	300
metal	0.89	0.91	0.90	149
paper	0.90	0.94	0.92	207
plastic	0.88	0.90	0.89	257
shoes	0.98	1.00	0.99	245
trash	0.93	0.88	0.90	74

The confusion matrix of the proposed model is illustrated in Figure 6, while the numerical confusion matrix values are shown in Table 7.

Table 7. Confusion matrix of the proposed EfficientNetB0–SVM model

Class ID	Class Name	TP	FP	FN	TN
0	battery	120	5	3	1877
1	biological	118	0	4	1883
2	cardboard	217	12	13	1763
3	clothes	295	1	3	1706
4	glass	274	11	26	1694
5	metal	136	16	13	1840
6	paper	195	22	12	1776
7	plastic	231	33	26	1715
8	shoes	245	4	0	1756
9	trash	65	5	9	1926

The confusion matrix of the proposed EfficientNetB0 - SVM model, shown in Figure 7, depicts a pretty good performance of the model across all 10 waste categories that yields a pretty good overall test accuracy of the SVM model of 94.56%. Most of the numbers are right on the diagonal, so basically the model is hitting a lot of true positive. For example, clothes have come 295 out of 298 correct, shoes perfectly 245/245, battery 120/123 and biological 118/122 - almost no misclassifications, as detailed in Table 7.

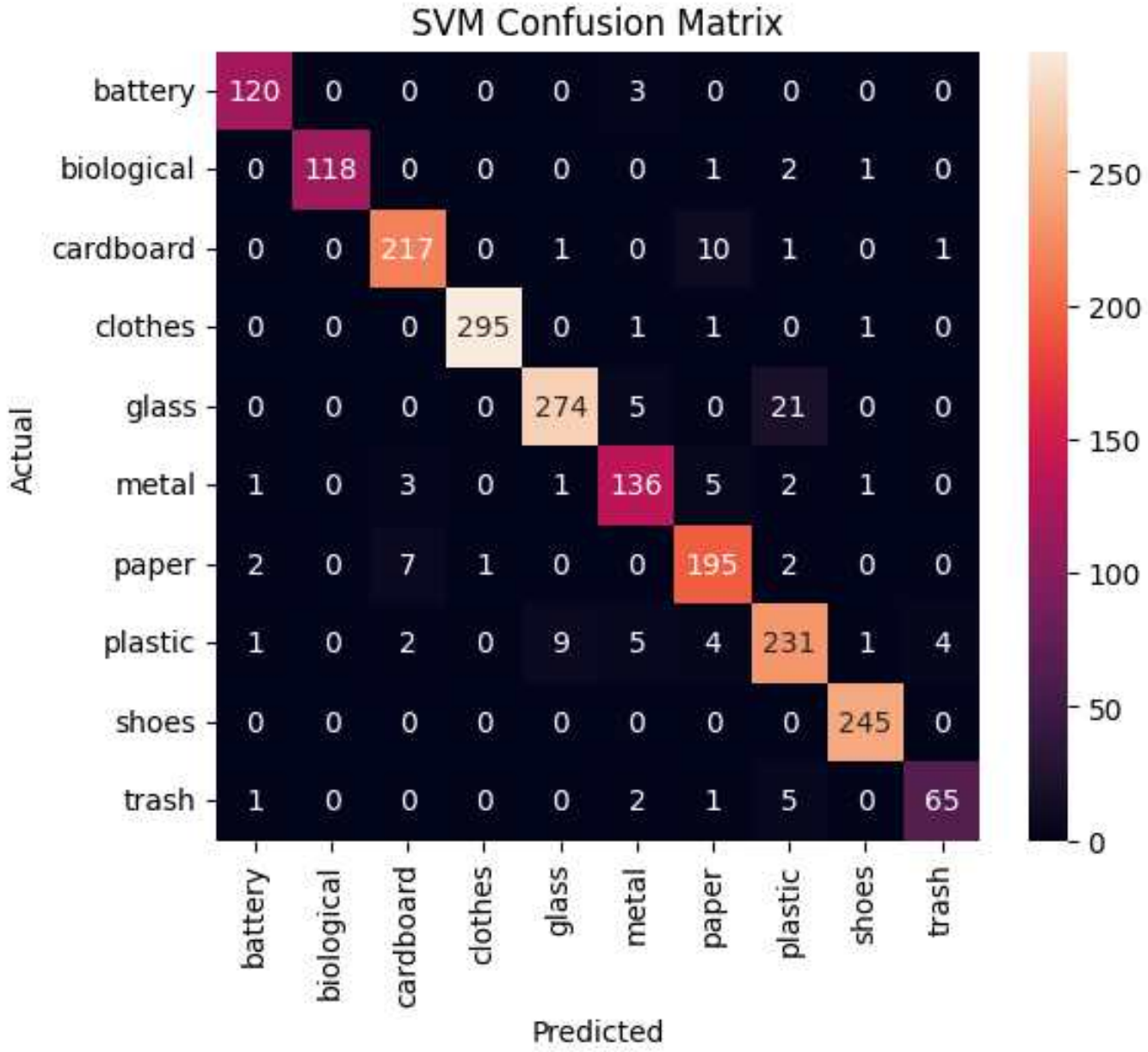


Figure 7. Confusion matrix of the proposed EfficientNetB0–SVM model

The little blips occurred between visually similar classes, plastic has 231 correct and 26 misclassified, paper 195 correct and 12 misclassified, and glass 274 correct with 26 misclassified. Even with those few errors, the overwhelming number of true positives alongside minimal false negatives, still make us feel confident that this hybrid approach is a good and reliable one.

Table 8. Real-world waste samples and their classification using the proposed EfficientNetB0–SVM model

	Plastic Confidence: **64.7%** *(Plastic bag & packaging)*		Biological Confidence: **66.4%** *(Food & organic waste)*
	Metal Confidence: **71.7%** *(Scrap metal pipes)*		Clothes Confidence: **23.1%** *(Clothes)*
	Metal Confidence: **64.9%** *(Mixed metal cans)*		Paper Confidence: **63.5%** *(Paper tape roll)*

In order to determine the practical useability of the proposed EfficientNetB0SVM model, more waste images are taken beyond the training set under the natural environmental conditions. Such pictures involve plastic wrappings, food waste which are organic and metal scrap, clothes as well as paper waste. Table 8 gives the classification result of these real-life samples.

The model is able to classify all waste samples in the real world as indicated in Table 8. The correct classification of plastic waste is 64.7% confidence, organic waste is the biological waste with confidence of 66.3 and 71.7% respectively, and multiple metal waste samples is identified with confidence up to 71.7%. Paper material is also identified rightly with the confidence of 63.5.

These findings suggest that the model that is proposed can be applied to the real-world situation and can correctly categorize waste in different lighting conditions, backgrounds, and object positions. This also illustrates the strength and functional effectiveness of the EfficientNetB0-SVM model in automated waste segregation and smart recycling systems.

5. Conclusion

The research paper has given a comparative analysis of the traditional handcrafted feature extraction and hybrid deep learning feature extraction methods in the classification of multi-class waste images. Classical attributes like HOG, LBP, GLCM, SIFT and ORB were tested in different machine learning classifiers, but with little results, giving the best results of 60.09. To enhance the classification success, deep learning classifiers ResNet50 and EfficientNetB0 were taken as feature extractors, and the feature extractors were detected using machine learning classifiers. Performance was much enhanced by the hybrid method, with the most accurate result of 94.56 per cent with EfficientNetB0 and SVM classifier. The findings prove that deep learning feature extraction is more discriminative and stronger in features than conventional handcrafted features. The hybrid deep learning and machine learning system is a resourceful and precise answer to the waste image classification, and it can be successfully employed in automated waste management systems. The work can be continued to the hyperparameter tuning, feature

optimization, and real-time implementation in the future in order to increase the classification performance.

References :

1. S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a Convolutional Neural Network," International Conference on Engineering and Technology (ICET), 2017. Available: https://ieeexplore.ieee.org/document/8308186
2. K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016. Available: https://arxiv.org/abs/1512.03385
3. M. Tan and Q. Le, "EfficientNetB0: Rethinking Model Scaling for Convolutional Neural Networks," International Conference on Machine Learning (ICML), 2019. Available: https://arxiv.org/abs/1905.11946
4. D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," International Journal of Computer Vision, vol. 60, no. 2, pp. 91–110, 2004.
5. E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: An Efficient Alternative to SIFT or SURF," IEEE International Conference on Computer Vision (ICCV), 2011.
6. N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005.
7. T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns," IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002.
8. R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural Features for Image Classification," IEEE Transactions on Systems, Man, and Cybernetics, 1973.
9. J. Sivic and A. Zisserman, "Video Google: A Text Retrieval Approach to Object Matching in Videos," Proceedings of ICCV, 2003.
10. C. Cortes and V. Vapnik, "Support-Vector Networks," Machine Learning, vol. 20, pp. 273–297, 1995.
11. L. Breiman, "Random Forests," Machine Learning, vol. 45, pp. 5–32, 2001.
12. T. Cover and P. Hart, "Nearest Neighbor Pattern Classification," IEEE Transactions on Information Theory, 1967.
13. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," Advances in Neural Information Processing Systems (NeurIPS), 2012.
14. S. Kumar, "Garbage Classification V2 Dataset," Kaggle. Available: https://www.kaggle.com/datasets/sumn2u/garbage-classification-v2
15. A. K. Sharma et al., "Hybrid Deep Feature Extraction with Machine Learning Classifiers for Image Classification," Knowledge-Based Systems, 2025. Available: https://www.sciencedirect.com/science/article/pii/S0950705125000760
16. K. Kaza, L. Yao, P. Bhada-Tata, and F. Van Woerden, "What a Waste 2.0: A Global Snapshot of Solid Waste Management to 2050," World Bank Publications, 2018. Available: https://openknowledge.worldbank.org/handle/10986/30317
17. United Nations Environment Programme (UNEP), "Global Waste Management Outlook," UNEP, 2015. Available: https://www.unep.org/resources/report/global-waste-management-outlook
18. Central Pollution Control Board (CPCB), "Annual Report on Municipal Solid Waste Management," Government of India, 2022. Available: https://cpcb.nic.in
19. P. Hoornweg and P. Bhada-Tata, "What a Waste: A Global Review of Solid Waste Management," World Bank, 2012.
20. Government of India, "Swachh Bharat Mission Guidelines," Ministry of Housing and Urban Affairs, 2014. Available: https://swachhbharatmission.gov.in

ISSN : 2348-7143
March 2026
RESEARCH JOURNEY International Multidisciplinary E-Research Journal
Impact Factor (SJIF) - 6.625 | ***Special Issue 389 :*** *Advanced Computing and Artificial Intelligence (ACAI)*

Personal Data Security in the Digital Age

Mr. Darshan Ishwar Sonawane,
Ms. Vaishnavi Tukaram Dorik,
Ms. Dipali Ravindra Nhalde

Abstract :

In the modern digital era, personal data security has become a major concern due to the rapid growth of cyber threats such as phishing, malware, identity theft, and data breaches. The widespread use of online platforms for communication, banking, education, and social interaction has increased individuals' exposure to security risks. This research paper examines the importance of safeguarding sensitive personal information through effective security measures, including strong authentication mechanisms, data encryption, secure browsing practices, and timely threat detection and response.

The study highlights the significance of adopting a proactive and multi-layered security approach that combines technical safeguards with user awareness and responsible digital behaviour. It further emphasizes the role of digital hygiene practices, such as regular software updates, secure password management, and cautious online activity, in minimizing vulnerabilities. By integrating technological solutions with continuous user education, individuals can significantly reduce cyber risks and ensure long-term protection of personal

Keywords: Personal Data Security, Cybersecurity, Phishing Attacks, Malware, Identity Theft, Data Breaches, Authentication, Data Encryption, Secure Browsing, Threat Detection, Digital Hygiene, Privacy Protection, Cyber Threats, Information Security

1. Introduction :

In today's digital age, the internet has become an integral component of our daily lives, allowing us to communicate, conduct banking transactions, learn, shop, and socialize. Although these technologies offer us ease and convenience, they also pose a serious threat to our cybersecurity. Among the most prevalent and harmful threats is phishing, which uses deceptive links or websites to steal valuable information such as login credentials, financial information, and identification details [3][4].

Cybercriminals usually develop malicious URLs that look similar to genuine websites, making it hard for common users to differentiate between the two. Malicious URLs can be shared via emails, social media platforms, messaging services, or pop-up ads. Once users click on the malicious URLs, they are directed to fake websites that are intended to steal sensitive information or install malware on the users' devices. This leaves users vulnerable to financial loss, identity theft, privacy infringement, and unauthorized access to their accounts [5][6].

Traditional security measures such as antivirus software and browser warnings are not always sufficient to detect newly created phishing websites. Moreover, many users lack technical knowledge to manually verify the authenticity of URLs. Therefore, there is a strong need for an accessible and automated solution that can analyze suspicious links and provide clear guidance to users [7].

"To counter this problem, this project proposes a web-based system that can assess the safety of URLs in real-time. The system examines various features of a URL, such as the use of secure protocols (HTTPS), presence of malicious keywords, unusual length, numeric expressions, and domain information. Depending on these criteria, the system categorizes the URL into three groups: Safe, Suspicious, and Dangerous. The output is presented through simple indicators and warnings so that users can easily interpret the level of danger and make appropriate decisions before accessing the website [9][10].."

The main aim of this project is to improve the security of personal data by preventing users from accessing fraudulent or malicious websites. The proposed system will help to prevent phishing attacks and increase awareness among users by providing instant analysis and feedback. The main aim of this project is to improve the security of personal data by preventing users from accessing fraudulent or malicious websites.

Keywords: Phishing Detection, Malicious URL Detection, Cybersecurity, Personal Data Security, URL Analysis, Phishing Attacks, Malware Protection, Web Security, Identity Theft Prevention, Online Safety.

2. Problem Statement :

Despite advancements in cybersecurity technologies, personal data breaches continue to rise due to human errors, unsafe digital habits, and system vulnerabilities. Many users lack awareness of safe online practices, such as using strong passwords, updating software regularly, and identifying suspicious links or messages. Cyber attackers exploit these weaknesses through phishing, social engineering, and other malicious activities [3][6].

Additionally, the rapid growth of online platforms has increased the sharing of personal information, making individuals more vulnerable to cyber threats. Security tools

and knowledge are not equally accessible to all users, further increasing risks. Therefore, there is a need for an effective solution that can help users identify harmful or fraudulent links and promote safer online behavior to protect personal data [1][2].

3. Objectives of the Study :

I. To examine the increasing threats to personal data in the online environment [6].
II. To determine the common features of phishing and malicious URLs [9].
III. To design a system that can detect fishy links in real-time.
IV. To categorize URLs as Safe, Suspicious, and Dangerous.
V. To improve user awareness and encourage safe browsing habits [3].

4. Cyber Threat :

A cyber threat is any malicious activity that attempts to damage, steal, or gain unauthorized access to computer systems, networks, or personal data through digital means. These threats target individuals, organizations, or governments and can lead to financial loss, identity theft, or data breaches [2][6].

- **Common Types of Cyber Threats**

i. Phishing — Fake emails or links used to steal personal information [4]
ii. Malware — Harmful software that damages or controls devices [5]
iii. Ransomware — Locks data and demands payment for access [6]
iv. Identity Theft — Misuse of personal information for fraud [3]
v. Social Engineering — Manipulating people to reveal confidential data [3]
vi. Data Breaches — Unauthorized access to sensitive information [2]

5. Research Methodology :

The project follows these steps:

Step 1: User Input

In this stage, the user interacts with the web application interface.

- The user enters or pastes a URL into the input field provided on the website.
- The system validates the input to ensure it is in proper URL format.
- If the input is invalid or empty, an error message is displayed.
- Once validated, the URL is sent to the backend for analysis.

Purpose: To collect the link that needs to be checked for safety.

Step 2: URL Analysis

The system performs multiple checks on the entered URL to identify suspicious characteristics.

i. HTTPS Presence Check :
- Determines whether the website see uses HTTPS (secure protocol) or only HTTP.
- HTTPS indicates encrypted communication between user and server.
- Lack of HTTPS may indicate potential risk, especially for login or payment pages.

ii. Suspicious Keyword Detection :
- The system scans the URL for commonly used phishing or scam words such as:
- login
- verify
- update
- secure
- bank
- free
- bonus
- account
- Attackers often use such words to create urgency or trust.

iii. URL Length Analysis
- Extremely long URLs may hide malicious intent.
- Phishing links often include long strings to disguise real domains.
- If the URL length exceeds a predefined threshold, risk score increases.

iv. Numeric Pattern Detection
- Checks for excessive numbers or unusual character sequences.
- Many fake links contain random numbers to mimic legitimate domains.
- Example: secure-bank-login-983726.com

v. Domain Characteristics Check
- Examines domain-related properties such as:
- Presence of subdomains
- Use of IP address instead of domain name
- Recently created or suspicious domain structure
- Fake websites often imitate well-known brands using slight variations.

Purpose: To detect patterns commonly associated with phishing, spam, or malicious websites.

Step 3: Risk Classification

After analysis, the system assigns a risk level based on the number and severity of suspicious indicators.

I. Safe
- No major suspicious features detected

- HTTPS present
- Clean and normal domain structure
- Likely legitimate website

II. Suspicious

- Some warning signs present.
- May contain minor risks.
- User should proceed cautiously

III. Dangerous

- Multiple high-risk indicators detected
- Possible phishing, malware, or scam site
- Strong warning issued to user

Purpose: To simplify technical analysis into easy-to-understand categories.

Step 4: Result Display

The final result is presented to the user clearly and instantly.

- The system displays the safety status which will be Safe, Suspicious, or Dangerous.
- Color indicators may be used:

o Green for Safe

o Yellow for Suspicious

o Red for Dangerous

- Warning messages or safety advice are shown if needed.
- The user can decide whether to proceed or avoid the link.

Purpose: To provide immediate awareness and help users make safe decisions.

6. System Features :

i. Real-time analysis of URLs
ii. Detection of phishing indicators
iii. User-friendly interface
iv. Clear safety classification
v. Awareness support for non-technical users[3] [7]

7. Findings and Discussion :

Sr. No.	Sector	Typical Data Exposed	Average Records Exposed (per major breach)	Average Cost per Breach	Key Impact
1	Banking & Financial Services	Account numbers, card data, KYC, login credentials	1–10 million records	$5.9 million (≈ □49 crore)	Financial fraud, identity theft
2	Healthcare	Medical records, insurance info, personal IDs	5–50 million records	$10.9 million (≈ □90 crore)	Privacy violation, insurance fraud
3	Government / Public Sector	National IDs, tax records, citizen databases	10 million to billions	$4.0 million (≈ □33 crore)	Mass identity misuse, national risk
4	Retail & E-Commerce	Payment info, addresses, purchase history	5–30 million records	$4.3 million (≈ □36 crore)	Card fraud, phishing
5	Technology / Cloud Services	Stored files, credentials, enterprise data	Highly variable (millions+)	$6.0 million (≈ □50 crore)	Large-scale data leaks
6	Education	Student records, academic data, personal details	1–10 million records	$3.6 million (≈ □30 crore)	Identity misuse, scams
7	Telecom	Phone numbers, SIM data, call records	5–20 million records	$3.9 million (≈ □32 crore)	SIM swap fraud
8	Social Media Platforms	Profiles, photos, contacts, messages	Hundreds of millions	$4.5 million (≈ □37 crore)	Privacy invasion, cybercrime

8. Prevention, Detection & Recovery Strategies :

These strategies represent a complete cybersecurity approach to protect personal data before, during, and after a cyber attack.

I. Prevention Strategies

Prevention focuses on reducing vulnerabilities and stopping threats from occurring.

- **Strong Password Practices:**
  Using complex passwords that contain a combination of letters, numbers, and symbols. Using the same password for multiple accounts should be avoided. [1]
- **Regular Software Updates:**
  Keeping operating systems, applications, and antivirus software up to date helps fix security vulnerabilities that attackers may exploit. [6]
- **Secure Browsing Habits:**
  Access only trusted websites, check for HTTPS, avoid downloading files from unknown sources, and be cautious when

clicking links from emails or messages. [3]

- **User Awareness and Education:** Understanding common cyber threats such as phishing, scams, and fake websites helps users make safer decisions online. [3]
- **Use of Security Tools:** Firewalls, antivirus software, and encryption technologies provide an additional layer of protection for devices and data. [5]

Goal: Minimize the chances of a successful attack.

II. Threat Detection & Response (During an Attack)

These measures help identify suspicious activities and respond quickly to limit damage.

- **Monitoring Suspicious Activities :** Systems continuously observe unusual behavior such as repeated login attempts, unknown access locations, or abnormal data transfers.
- **Malicious Link Detection :** Identifying phishing or harmful URLs before users interact with them helps prevent credential theft and malware infections.
- **Real-Time Alerts :** Immediate warnings notify users about potential threats, allowing them to take preventive action.
- **Blocking Harmful Access :** Dangerous websites, files, or connections are automatically blocked to stop further compromise.
- **Incident Response Actions :** Steps such as disconnecting affected systems, disabling compromised accounts, or isolating infected devices are taken to control the situation. [6][7]

Goal: Detect threats early and reduce impact.

III. Recovery Strategies

Recovery focuses on restoring normal operations and preventing future incidents.

- **Data Backup and Restoration:** Restoring files from secure backups ensures minimal data loss after ransomware or system failure.
- **Password Reset and Account Recovery:** Compromised accounts must be secured by changing passwords and enabling stronger authentication methods.
- **Malware Removal:** Infected systems are scanned and cleaned using antivirus or specialized tools to eliminate malicious software.
- **System Repair and Updates:** Vulnerabilities exploited during the attack are patched to prevent recurrence.
- **Security Improvement Measures**: After analyzing the incident, additional safeguards and policies are implemented to strengthen overall protection. [2][6]

Goal: Restore services safely and enhance future resilience.

9. Future Scope :

The proposed system can be enhanced further to improve accuracy, usability, and real-time protection. The future scope includes :

I. Machine Learning-Based Detection

Currently, the system uses rule-based analysis. In the future, machine learning algorithms can be trained on phishing and legitimate URL datasets to improve detection accuracy. [9]

- The model can learn patterns automatically.
- It can detect new and unknown threats.
- Accuracy and reliability will increase over time.

II. Integration with Antivirus Databases

The system can be connected to trusted antivirus or threat intelligence databases. [5]

- Real-time comparison with known malicious websites.
- Faster identification of blacklisted domains.
- Improved protection against malware-based attacks.

III. Email and SMS Scam Detection

The project can be expanded to analyze suspicious emails and SMS messages. [4]

- Detection of phishing emails.
- Identification of fake OTP or bank-related scam messages.
- Protection against social engineering attacks.

IV. Mobile Application Development

A mobile app version can be developed for Android and iOS users.

- Users can check links directly from their smartphones.
- Increased accessibility and convenience.
- Useful for everyday browsing and social media safety.

V. Browser Extension for Real-Time Protection

A browser extension can automatically scan websites while browsing.

- Instant alerts before opening harmful websites.
- Background monitoring without manual input.
- Enhanced real-time security.

VI. AI-Based Threat Intelligence System

Advanced AI systems can analyze global cyber threat trends. [6]

- Predict emerging cyber threats.
- Provide dynamic risk scoring.
- Continuously update detection mechanisms.

10. Conclusion

In summary, the rapid development of digital technology has greatly increased the vulnerability of personal data, making cybersecurity a major concern for both individuals and organizations. Many users are still at risk of phishing attacks and malicious websites because of a lack of awareness and inability to detect deceptive links, which frequently results in identity theft, financial loss, and unauthorized access to confidential data [3][4]. The proposed web-based system can solve this issue by examining URLs in real-time and detecting major risk indicators such as insecure protocols, risky keywords, unusual length, and domain irregularities, which are often linked to phishing attacks [7][9]. By categorizing links into Safe, Suspicious, and Dangerous groups and providing results in a visual warning format, the system allows users to make informed decisions before accessing potentially damaging websites, thus lowering the risk of cyber events [6].

In addition, the system fosters awareness and best practices for cybersecurity, which is crucial for preventing breaches of data security [2]. In conclusion, this solution offers a realistic and user-friendly method for improving the security of personal data, and with future enhancements such as the integration of machine learning capabilities, it has the potential to offer even more precise protection against cyber threats [5][10].

References :

1. National Institute of Standards and Technology (NIST). (2017). Digital Identity Guidelines (NIST Special Publication 800-63). U.S. Department of Commerce. This publication contains comprehensive standards and best practices for the management of digital identity and the security of the authentication process. International Organization for Standardization, ISO/IEC 27001: Information Security Management Systems — Requirements, ISO, 2022.
2. International Organization for Standardization (ISO). (2022). ISO/IEC 27001: Information Security Management Systems—Requirements. ISO. This international standard describes the requirements for establishing, implementing, maintaining, and continuously improving an information security management system (ISMS).Kaspersky, Spam and Phishing in Cybersecurity Reports, Kaspersky Lab.
3. Cybersecurity and Infrastructure Security Agency (CISA). Cybersecurity Best Practices for Individuals and Organizations. U.S. Department of Homeland Security. This publication contains important best practices for individuals and organizations to follow in order to avoid security issues.Microsoft, Security Intelligence Report, Microsoft Corporation.
4. Google LLC. Safe Browsing Transparency Report. The report discusses information on unsafe sites detected by Google and how the company is working to protect its users from such malicious content.
5. Akinyelu, A. A., & Adewumi, A. O. (2014). Classification of Phishing Email Using Random Forest Machine Learning Technique. Journal of Applied Security Research. The research focuses on the application of machine learning techniques in the identification of phishing emails.
6. Gupta, M., Agrawal, R., & Jain, A. (2016). Phishing Detection Using URL Features and Machine Learning. International Journal of Computer Applications. The research discusses the different URL feature extraction techniques used in the identification of phishing emails.
7. Garera, S., Provos, N., Chew, M., & Rubin, A. D. (2007). A Framework for Detection and Measurement of Phishing Attacks. Proceedings of the ACM Workshop on Recurring Malcode. The research presents a framework for the identification and measurement of phishing attacks.
8. RSA Security. Fraud and Risk Intelligence Reports. RSA Security LLC. These reports offer information concerning identity fraud, phishing, and online financial threats.
9. IBM Security. Cost of a Data Breach Report. IBM Corporation. This report examines the effects of data breaches in the global market.
10. Schneier, B. (2015). Data and Goliath: The Hidden Battles to Collect Your Data and Control Your World. W.W. Norton & Company. This book discusses the importance of privacy in the digital world and the need for data protection in society.
11. Stallings, W., & Brown, L. (2018). Computer Security: Principles and Practice. Pearson Education. This book provides an introduction to the fundamental concepts of cyber security, such as malware, phishing, encryption, and network defense mechanisms.
12. Whitman, M. E., & Mattord, H. J. (2019). Principles of Information Security. Cengage Learning. This book describes the concepts of risk management, security policies, and strategies for protecting information assets.

RESEARCH JOURNEY International Multidisciplinary E-Research Journal ISSN : 2348-7143
Impact Factor (SJIF) - 6.625 | ***Special Issue 389 :*** *Advanced Computing and Artificial Intelligence (ACAI)* March 2026

A Study on the Application, Benefits, and Challenges of AI-Based Tools in Academic

Mrs. Kirtika Nahar Behere
(Assistant Professor, R.C. Patel IMRD, Shirpur)
Mrs. Suvarna S. Chadhuari
(Assistant Professor, R.C.Patel IMRD, Shirpur)

Abstract: -

AI (Artificial Intelligence) is fundamentally changing education and the ways in which students are taught and institutions operate by implementing ground-breaking theories and practices into the education sector. The main goal of this paper is to talk about how we can use AI-based tools in education. We will look at what these tools can do for us what is good about them and what problems they can cause. We want to see how new technology can be used in the classroom and in the office. This includes things like computer programs that can teach students one on one systems that can grade tests automatically software that can detect when someone is copying work and systems that can adjust to how each student learns. We will also look at AI tools, like ChatGPT that can generate things. Also, this study will present evidence that supports the argument that the increased use of artificial intelligence and its associated applications in higher education have many potential benefits, which are the individualization of learning for students, increased efficiencies when they are being graded or provided with feedback, improved levels of productivity in producing research by utilizing advanced data analysis technologies, and more efficient administrative processes. Furthermore, the use of artificial intelligence-based tools promotes accessibility and inclusiveness by providing support for students with various learning styles and needs.

Keyword:-

Artificial Intelligence (AI), Educational Technology, Intelligent Tutoring Systems, Adaptive Learning Systems, Generative AI, ChatGPT.

1. Introduction:-

Artificial Intelligence is a trend in education right now. Artificial Intelligence is changing the way we teach and learn. Artificial Intelligence is helping to make learning personal and fun. People are talking about how Artificial Intelligence can help make education better. There are also some concerns about Artificial Intelligence. Some people think Artificial Intelligence might not be fair or might hurt peoples ability to think for themselves.

2. Objective:-

1. To study the application of artificial intelligence in academic institutions.
2. To understand the basics of artificial intelligence in the classroom.
3. To study the challenges related to using AI in the classroom.

3. Research Methodology:-

3.1 Research Design

We did a study to learn about Artificial Intelligence in education. We asked people questions. Got their answers.

3.2 Data Collection Method

We made a questionnaire to ask people about Artificial Intelligence. We asked them questions like what they think about Artificial Intelligence and how they use it.

- We made the questionnaire with yes or no questions.
- We did the questionnaire online so people could answer from anywhere.
- We made sure the questions were clear and easy to understand.

It was easy to get people to answer the questionnaire because we did it online.

3.3 Tools for Analysis

We used some tools to help us understand the answers.

- We used percentage analysis to see how many people answered each question the way.
- We used statistical interpretation to understand the answers better. These tools helped us to organize the answers and make sense of them.

3.4 Sampling Technique

We picked people to answer the questionnaire who were easy to reach.

The convenience sampling technique is easy and fast. It might not be totally accurate because we only asked people who were easy to reach.

3.5 Tools for Analysis

We used the tools again to understand the answers.

- We used percentage analysis again.
- We used statistical interpretation again.

These tools helped us again to make sense of the answers.

4. Primary Data Analysis:-

Area	Finding	Percentage
AI Awareness	Majority of students are aware of AI tools	96.5%
Most Recognized Tool	ChatGPT is the most widely known AI tool	99.1%
Academic Usage	Students using AI for academic purposes	97.4%
Primary Purpose	AI mainly used for research and exam preparation	80.7%
Coding Support	Students using AI for coding assistance	75.4%
Academic Performance Impact	Students agree AI improves performance	94.7%
Grade Improvement	Students reporting improvement in grades	71.9%
Key Benefit	Quick and instant answers	78.1%
Critical Concern	Belief that AI reduces independent thinking	50%
Main Challenge	Limitations due to paid features	56.1%
Secondary Challenge	Internet connectivity issues	41.2%

5. Data Interpretation and Finding:-

Most people like Artificial Intelligence and use it to learn. Almost everyone knows about Artificial Intelligence. Uses it for school. People use Artificial Intelligence to help them study and do their homework. They also use it to help them with their research and coding. Most people think Artificial Intelligence helps them do better in school. Some people are worried that Artificial Intelligence might make them less able to think for themselves. Some people also think that Artificial Intelligence is too expensive and that it can be hard to use if you do not have an internet connection.

6. Application:-

6.1 Intelligent Tutoring Systems

Artificial Intelligence can help teachers by giving students one-on-one attention. Artificial Intelligence can also help students learn at their pace.

6.2 Adaptive Learning Platforms

Artificial Intelligence can help make learning more personal. Artificial Intelligence can change the way we learn based on how we do.

6.3 Automated Grading and Assessment

Artificial Intelligence can help teachers by grading homework and tests. Artificial Intelligence can also give students feedback away.

6.4 Generative Artificial Intelligence in Teaching and Research

Artificial Intelligence can help with writing and research. Artificial Intelligence can also help with coding and coming up with ideas.

6.5 Plagiarism Detection Systems

Artificial Intelligence can help make sure people do not cheat. Artificial Intelligence can check if someone's work is original.

6.6 Research and Data Analysis

Artificial Intelligence can help with research by looking at amounts of data. Artificial Intelligence can also help make sense of the data.

6.7 Administrative Automation

Artificial Intelligence can help schools with things like admissions and grades. Artificial Intelligence can also help answer questions.

7. Benefits: -

7.1 Personalized Learning Experiences

Artificial Intelligence can help make learning more personal. Artificial Intelligence can change the way we learn based on how we do.

7.2 Improved Efficiency and Productivity

Artificial Intelligence can help teachers and schools work efficiently. Artificial Intelligence can automate things like grading and research.

7.3 Enhanced Research Capabilities

Artificial Intelligence can help with research by looking at amounts of data. Artificial Intelligence can also help make sense of the data.

7.4 Timely Feedback and Continuous Assessment

Artificial Intelligence can help give students feedback away. Artificial Intelligence can also help students see how they are doing.

7.5 Increased Accessibility and Inclusivity

Artificial Intelligence can help make sure everyone has access to education. Artificial Intelligence can help people with disabilities.

7.6 Data-Driven Decision Making

Artificial Intelligence can help schools make decisions. Artificial Intelligence can look at data. Help schools see what they need to do.

8. Challenge:-

8.1 Ethical Issues and Academic Integrity

Some people are worried about Artificial Intelligence and cheating. Some people think Artificial Intelligence

might make it hard to know who did the work.

8.2 Data Privacy and Security

Some people are worried about Artificial Intelligence and data. Some people think Artificial Intelligence might not be safe.

8.3 Overdependence, on Technology

Some people are worried that we might rely much on Artificial Intelligence. Some people think we might forget how to think for ourselves.

8.4 Technical and Financial Barriers

Some people are worried that Artificial Intelligence might be too expensive. Some people think that some schools might not be able to afford it.

9. Conclusion:-

Artificial Intelligence is a part of education now. Most people use Artificial Intelligence to help them learn. Artificial Intelligence can help make learning personal and fun.. We need to be careful and make sure we use Artificial Intelligence in a good way. We need to make sure we do not rely much on Artificial Intelligence and that we use it to help us not to hurt us. Artificial Intelligence can help make education better. We need to be responsible and use it wisely.

References :

1. Intelligence in Schools and Colleges. London, U.K.: Nesta, 2019.
2. C. M. Bishop, Pattern Recognition and Machine Learning. New York, NY, USA: Springer, 2006.
3. W. Holmes, M. Bialik, and C. Fadel, Artificial Intelligence in Education: Promises and Implications for Teaching and Learning. Boston, MA, USA: Center for Curriculum Redesign, 2019.
4. R. Luckin, W. Holmes, M. Griffiths, and L. B. Forcier, Intelligence Unleashed: An Argument for AI in Education. London, U.K.: Pearson Education, 2016.
5. S. Russell and P. Norvig, Artificial Intelligence: A Modern Approach, 4th ed. Hoboken, NJ, USA: Pearson, 2021.
6. N. Selwyn, Should Robots Replace Teachers? AI and the Future of Education. Cambridge, U.K.: Polity Press, 2019.
7. UNESCO, AI and Education: Guidance for Policy-Makers. Paris, France: United Nations Educational, Scientific and Cultural Organization, 2021.
8. O. Zawacki-Richter, V. I. Marín, M. Bond, and F. Gouverneur, “Systematic review of research on artificial intelligence applications in higher education—Where are the educators?” Int. J. Educ. Technol. High. Educ., vol. 16, no. 39, 2019.
9. E. Kasneci et al., “ChatGPT for good? On opportunities and challenges of large language models for education,” Learn. Individ. Differ., vol. 103, p. 102274, 2023.

ISSN : 2348-7143
March 2026
RESEARCH JOURNEY International Multidisciplinary E-Research Journal
Impact Factor (SJIF) - 6.625 | ***Special Issue 389 :*** *Advanced Computing and Artificial Intelligence (ACAI)*

An Intelligent Learning and Placement Analytics Platform with Data-Driven Predictive Modeling

Mrs. Dhanashree Gajendrasing Patil,
Ms. Dipali Ravindra Nhalde
RCPET's Institute of Management Research and Development, Shirpur

Abstract

The rapid digitization of education and recruitment processes has generated large volumes of academic, behavioral, and skill-based data, creating opportunities for intelligent analytics-driven decision systems. This paper proposes an Intelligent Learning and Placement Analytics Platform with Data-Driven Predictive Modeling designed to enhance academic performance monitoring and improve placement outcome forecasting. The proposed platform integrates heterogeneous data sources, including academic records, skill assessments, coding performance metrics, attendance patterns, extracurricular participation, and mock interview evaluations, into a unified analytics framework.

The system employs advanced machine learning techniques such as ensemble learning models, gradient boosting algorithms, and deep neural networks to predict placement probability and identify critical performance determinants. Feature engineering and dimensionality reduction techniques are applied to extract meaningful patterns from high-dimensional educational datasets. Additionally, Explainable Artificial Intelligence (XAI) methods are incorporated to provide interpretable insights for students, faculty, and placement administrators, ensuring transparency in predictive decisions.

A real-time dashboard interface visualizes student progress, skill gaps, placement readiness scores, and risk indicators, enabling personalized learning interventions and strategic preparation plans. Experimental evaluation using accuracy, precision, recall, F1-score, and ROC-AUC demonstrates significant improvements over traditional statistical approaches. The framework is scalable and extendable to multi-institutional deployment using privacy-preserving learning techniques, paving the way for next-generation smart campus placement systems.

Keywords: Learning Analytics, Placement Prediction, Machine Learning, Explainable AI, Educational Data Mining, Smart Campus

1. Introduction

The rapid digital transformation of higher education has significantly increased the volume and variety of student-related data generated across academic and co-curricular ecosystems. Learning Management Systems (LMS) such as Moodle and Google Classroom, along with online coding and assessment platforms like HackerRank, continuously capture detailed records of student engagement, performance, and skill development. These systems generate structured and unstructured datasets encompassing academic scores, attendance patterns, technical skill assessments, internship participation, behavioral indicators, and placement preparation activities.

Despite the availability of such rich datasets, many higher education institutions continue to rely primarily on descriptive performance indicators—such as cumulative grade point average (CGPA), attendance percentages, and basic statistical summaries—to evaluate student progress and placement readiness. While useful, these conventional approaches fail to capture the multidimensional and dynamic nature of employability.

Recent advances in Educational Data Mining (EDM) and Learning Analytics have demonstrated the effectiveness of machine learning techniques in predicting academic outcomes, student retention, and dropout risks [4], [5]. Educational data mining has evolved as a significant research domain focusing on extracting meaningful patterns from educational datasets [5]. Research communities supported by organizations such as the International Educational Data Mining Society and publications under IEEE have emphasized predictive modeling as a key component of modern learning analytics [4].

Ensemble learning models, gradient boosting algorithms, and deep neural networks have shown strong capability in identifying complex nonlinear relationships within high-dimensional datasets. In particular, XGBoost has demonstrated scalable and high-performance predictive modeling capabilities in classification problems [2]. Furthermore, explainability in machine learning has gained significant importance, with methods such as SHAP providing theoretically grounded approaches for interpreting model predictions [1], and LIME offering local interpretability techniques for black-box classifiers [3].

However, while substantial work has been conducted on academic performance prediction and dropout analysis [4], [5], relatively limited research integrates learning analytics with placement outcome forecasting within a unified and explainable framework. Placement success is influenced by multiple interdependent variables—including academic trends, coding performance metrics, mock

interview scores, resume quality, and behavioral attributes. Existing systems rarely combine these heterogeneous factors into a single predictive and interpretable decision-support platform.

To address this gap, this study proposes an Intelligent Learning and Placement Analytics Platform with Data-Driven Predictive Modeling. The proposed framework integrates multi-source educational data, applies advanced machine learning techniques such as gradient boosting [2], and incorporates Explainable Artificial Intelligence (XAI) methods including SHAP [1] to ensure transparency and interpretability of model outputs. Additionally, a real-time analytics dashboard is developed to visualize placement readiness scores, skill gaps, and risk indicators.

By transforming traditional descriptive monitoring systems into predictive, data-driven decision platforms, this research contributes toward the development of intelligent, scalable, and student-centric educational ecosystems aligned with evolving industry demands.

2. Problem Statement

Higher education institutions generate extensive academic and skill-based data but lack predictive systems that convert this data into actionable placement insights. Traditional monitoring approaches:

- Focus on descriptive analysis rather than predictive intelligence
- Evaluate employability indicators in isolation
- Lack transparency in decision-making
- Do not support real-time intervention

As placement success depends on multiple interdependent factors—academic performance, coding proficiency, communication ability, internships, and mock interviews—there is a need for an integrated analytics framework capable of predicting placement probability and supporting proactive intervention strategies.

3. Objectives

3.1 General Objective

To design a scalable and interpretable analytics platform that predicts placement outcomes using data-driven modeling techniques.

3.2 Specific Objectives

1. Integrate heterogeneous academic and skill-based datasets.
2. Apply feature engineering and dimensionality reduction techniques.
3. Implement machine learning models such as Random Forest, XGBoost, SVM, and Neural Networks.
4. Incorporate Explainable AI techniques for model transparency.
5. Develop a real-time visualization dashboard.
6. Evaluate model performance using standard classification metrics.
7. Enable personalized intervention and placement readiness recommendations.
8. Design scalable deployment architecture.

4. Literature Review

Educational Data Mining and Learning Analytics have emerged as major research domains over the past decade. Romero and Ventura [5] provided one of the foundational reviews of EDM, outlining key techniques for extracting actionable insights from educational datasets. Baker and Inventado [4] further expanded on the integration of data mining techniques within learning analytics frameworks.

Machine learning techniques including Random Forest, Support Vector Machines, and Gradient Boosting have demonstrated strong predictive capability in academic performance modeling [4], [5]. Among gradient boosting frameworks, XGBoost introduced a scalable tree boosting system optimized for computational efficiency and predictive performance [2].

With the increasing use of complex models, interpretability has become critical. Lundberg and Lee [1] introduced SHAP (SHapley Additive exPlanations), providing a unified framework for interpreting model predictions based on cooperative game theory. Similarly, Ribeiro et al. [3] proposed LIME, a method for explaining predictions of any classifier locally and model-agnostically.

While prior studies focus largely on academic performance or dropout prediction [4], [5], placement forecasting remains underexplored, particularly in systems integrating predictive modeling with explainability mechanisms such as SHAP [1] and scalable boosting frameworks such as XGBoost [2].

5. Dataset and Features

The dataset comprises structured student records collected over multiple academic years.

Feature Categories:

- Academic Metrics: Semester grades, GPA, subject scores
- Skill Assessments: Coding platform scores, certifications
- Behavioral Indicators: Attendance, participation metrics
- Placement Metrics: Mock interview scores, resume evaluation, HR feedback
- Derived Features: Performance trends, readiness indices

Preprocessing includes normalization, categorical encoding, and missing value imputation.

6. Methodology

6.1 Data Preprocessing

Data cleaning, normalization, and encoding were performed. Dimensionality reduction using Principal

Component Analysis (PCA) was applied where necessary.

6.2 Predictive Modeling

Model	Type
Random Forest	Ensemble
XGBoost	Gradient Boosting
Support Vector Machine	Linear
Neural Networks	Deep Learning

Hyperparameter tuning was conducted using grid search and cross-validation.

6.3 Explainability

To ensure transparency and interpretability of predictions, SHAP (SHapley Additive exPlanations) was employed to quantify feature contributions toward placement probability [1]. SHAP provides consistent and locally accurate feature attribution based on Shapley values derived from cooperative game theory. Additionally, model-agnostic interpretability principles inspired by LIME were considered to support local explanation of individual predictions [3].

6.4 Dashboard Development

A web-based dashboard displays:

- Placement probability score
- Learning performance trends
- Skill gap analysis
- Personalized recommendations

7. Experiments and Evaluation

7.1 Experimental Setup

The dataset was split into 80% training and 20% testing sets.

7.2 Results

Model	Accuracy	F1-Score	ROC-AUC
XGBoost	0.88	0.86	0.92
Random Forest	0.85	0.83	0.89
SVM	0.81	0.79	0.86
Neural Network	0.87	0.85	0.91

XGBoost achieved the highest overall performance.

8. Discussion

The results demonstrate that ensemble and boosting models effectively capture complex relationships between academic indicators and placement outcomes. Coding performance and interview simulation scores were identified as significant predictors. The dashboard enables proactive intervention and personalized learning pathways.

9. Limitations

1. Data privacy concerns require strict governance mechanisms.
2. Limited generalizability across institutions without retraining.
3. External economic factors influencing placement are not incorporated.

10. Future Work

Future enhancements include:

- Federated learning for multi-institution deployment
- Real-time adaptive recommendation systems
- Integration of labor market analytics

11. Conclusion

This study proposed a comprehensive Intelligent Learning and Placement Analytics Platform integrating predictive modeling, explainable AI, and real-time visualization. The system demonstrates strong predictive accuracy and enables early identification of at-risk students. By transforming traditional monitoring systems into proactive, data-driven decision platforms, the framework contributes toward intelligent, student-centric educational ecosystems aligned with industry demands.

References :

1. Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In Advances in Neural Information Processing Systems (NeurIPS 2017), 4765–4774.
2. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785–794). ACM. https://doi.org/10.1145/2939672.2939785
3. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?" Explaining the predictions of any classifier. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1135–1144. https://doi.org/10.1145/2939672.2939778
4. Baker, R. S. J. d., & Inventado, P. S. (2014). Educational data mining and learning analytics. In C. Lang, G. Siemens, A. Corrin, & J. E. Fetcher (Eds.), Learning Analytics: Fundamentals, Applications, and Trends (pp. 61–75). Springer.
5. Romero, C., & Ventura, S. (2010). Educational data mining: A review of the state of the art. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 40(6), 601–618. https://doi.org/10.1109/TSMC.2010.2053532

RESEARCH JOURNEY International Multidisciplinary E-Research Journal ISSN : 2348-7143
Impact Factor (SJIF) - 6.625 | ***Special Issue 389 :*** *Advanced Computing and Artificial Intelligence (ACAI)* March 2026

Analyzing Library Books Borrowing Trends To Optimize Collection (Using Power BI)

Miss. Najuka Pravin Shinde
Student, R.C.Patel IMRD, Shirpur (MS) India.
Mr. Mayur Vishwas Patil
Student, R.C.Patel IMRD, Shirpur (MS) India.
Mrs. Dhanashree Patil
Assistant Professor, R.C.Patel IMRD, Shirpur (MS) India.

Abstract:

This study focuses on analyzing library book borrowing trends to help optimize library collections. The research is based on primary data collected from 300 respondents through a Google Form survey. It examines readers' preferences, borrowing frequency, and satisfaction with available books. Data analysis was done using Power BI to identify popular genres, high-demand subjects, and rarely borrowed materials.

The findings show that borrowing patterns vary based on users' interests and academic needs.

Some sections of the library are highly used, while others are underutilized. The study highlights the importance of using borrowing data to make better decisions about purchasing, removing, and managing books. This research demonstrates how data-driven approaches can improve library services and ensure collections better meet user needs.

Keywords: Power BI,Data Analytics ,Library Trends, Book Borrowing, User Preferences, Collection Management.

Introduction:

Libraries have always played an important role in supporting education and learning. For students and readers, a library is more than just a place where books are stored. It is a space for studying, exploring new ideas, preparing for examinations, and developing knowledge. Even in today's digital age, libraries continue to serve as valuable centers for academic growth and intellectual development (Bidve & Survase, 2024) [9].

However, managing a library has become increasingly challenging. Libraries do not have unlimited budgets or unlimited space. Every year, thousands of new books and digital resources are published, and it is impossible for any library to purchase and maintain all of them. Because of these limitations, librarians must make careful and informed decisions about which books to buy, keep, or remove. Effective collection development has therefore become essential for modern libraries (Rathod, 2025) [6].

To make better decisions, libraries need to clearly understand what their users actually read and prefer. Some books are borrowed frequently, showing that they are in high demand.

Others remain unused for long periods, occupying valuable shelf space. By studying

borrowing records, libraries can identify popular subjects, frequently read authors, and areas that may require improvement (Luo, 2019) [7]. Modern collection management practices recommend using usage data to guide purchasing and weeding decisions instead of relying only on assumptions (Collection Development Literature Review, 2023) [8].

In recent years, technology has provided new opportunities to improve library management.

Data analytics and visualization tools allow libraries to study large amounts of borrowing data and identify meaningful patterns. Techniques such as data mining help uncover trends in user behavior and predict future demand (Silwattananusarn & Kulkanjanapiban, 2020) [3].

Dashboard-based systems make it easier to analyze circulation statistics and evaluate reader preferences (Analysing Library Usage Patterns, 2024) [2]. Visualization tools like Power BI present this information in clear and interactive formats, helping librarians make smarter decisions (Neeraj, 2024) [5].

Data-driven collection optimization strategies further support libraries in improving service quality and resource utilization (Sengar, Saxena, & Rathor, 2025) [1]. Advanced analysis of circulation data helps predict demand and ensures better budget planning (Analysis of Book Circulation Data, 2022) [4]. Concepts such as Patron-Driven Acquisition and systematic collection development focus on aligning library resources with actual user needs (Patron- Driven Acquisition, 2023) [10]; (Collection Development, 2023) [11].

In this study, we analyzed library book borrowing trends to better understand readers' interests and reading habits. Primary data was collected from 300 respondents through Google Forms, along with an examination of library circulation records. The collected data was cleaned and analyzed using Power BI to identify patterns in subject popularity, borrowing frequency, and overall usage behavior. This approach helps improve user satisfaction through informed and data-based planning (Riyar, 2024) [12].

Objectives :

1) To understand the book borrowing habits of library users.
2) To find out which subjects and genres are most popular among readers.
3) To identify books and sections that are rarely used.
4) To study how readers' interests and needs affect their borrowing choices.
5) To use data analysis tools (like Power BI) to find patterns in borrowing trends.
6) To suggest ways to improve the library collection based on users' needs.
7) To help libraries make better decisions about buying, keeping, or removing books.

Features:

1) **Based on Real Data:**
   This research is based on actual borrowing records and responses from 300 readers. It does not depend on assumptions, but on real information.
2) **Understanding Readers' Interests:**
   The study focuses on knowing what type of books and subjects readers prefer the most.
3) **Finding Popular and Unused Books:**
   It helps to identify which books are borrowed frequently and which books are rarely used in the library.
4) **Better Collection Management:**
   The research helps the library decide which books to buy more copies of and which books can be removed.
5) **Proper Use of Budget and Space:**
   Since libraries have limited money and space, this study helps in using resources wisely.
6) **Use of Data Analysis Tools:**
   Tools like Power BI are used to analyze data and create charts to clearly understand borrowing patterns.
7) **Improving Library Services:**
   By understanding borrowing trends, libraries can improve their collection and provide better service to readers.

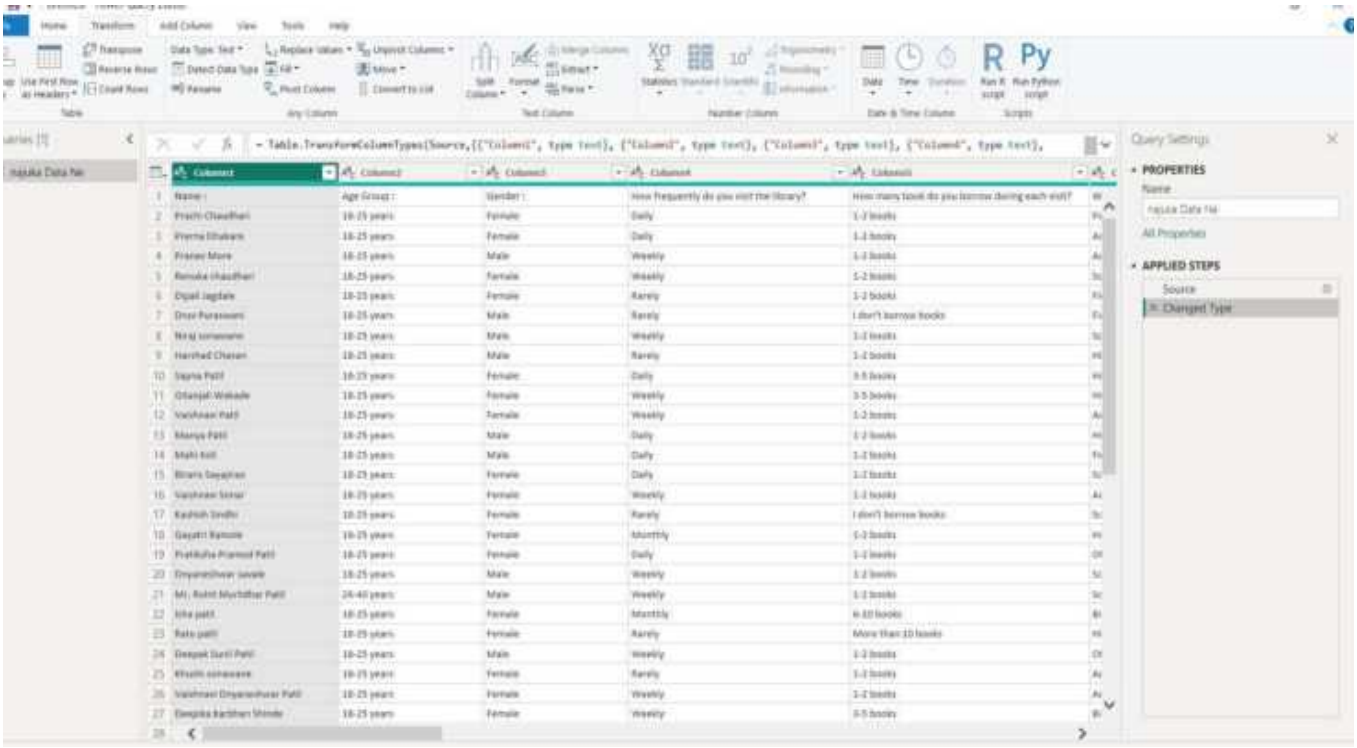
Research Methodology:

1) **Research Approach**
   This study follows a data-based research approach. Instead of making assumptions, the research focuses on real information collected from library users. It studies actual borrowing behavior to understand what readers truly prefer.
2) **Type of Research**
   The research is descriptive in nature. This means the study explains and describes borrowing patterns, reader preferences, and overall library usage trends without changing or influencing them.
3) **Data Collection Method**
   Primary data was collected using a structured questionnaire. The survey was designed in a simple and clear format and distributed through Google Forms to gather responses from library users.
4) **Sample Size**
   A total of 300 library users participated in this study. This sample size helps represent different types of readers, their interests, and their reading habits.
5) **Nature of Data**
   The study used both quantitative and qualitative data. Quantitative data includes numbers such as borrowing frequency and popular genres. Qualitative data includes readers' opinions, feedback, and suggestions about the library collection.
6) **Data Preparation**
   After collecting the data, it was carefully checked and organized. Incomplete or incorrect responses were removed to ensure that the analysis would be accurate and reliable.
7) **Tools Used for Analysis**
   Power BI was used to analyze and visualize the data. It helped in creating charts, graphs, and dashboards that made it easier to understand borrowing trends and patterns.
8) **Data Analysis Techniques**
   The study used simple analysis methods such as trend analysis and comparison. These techniques helped identify popular books, high-demand subjects, and materials that are rarely borrowed.
9) **Interpretation of Results**
   The analyzed data was carefully studied to understand readers' needs, interests, and gaps in the existing library collection.
10) **Outcome of the Study**
    Based on the findings, useful suggestions were provided to improve the library collection.

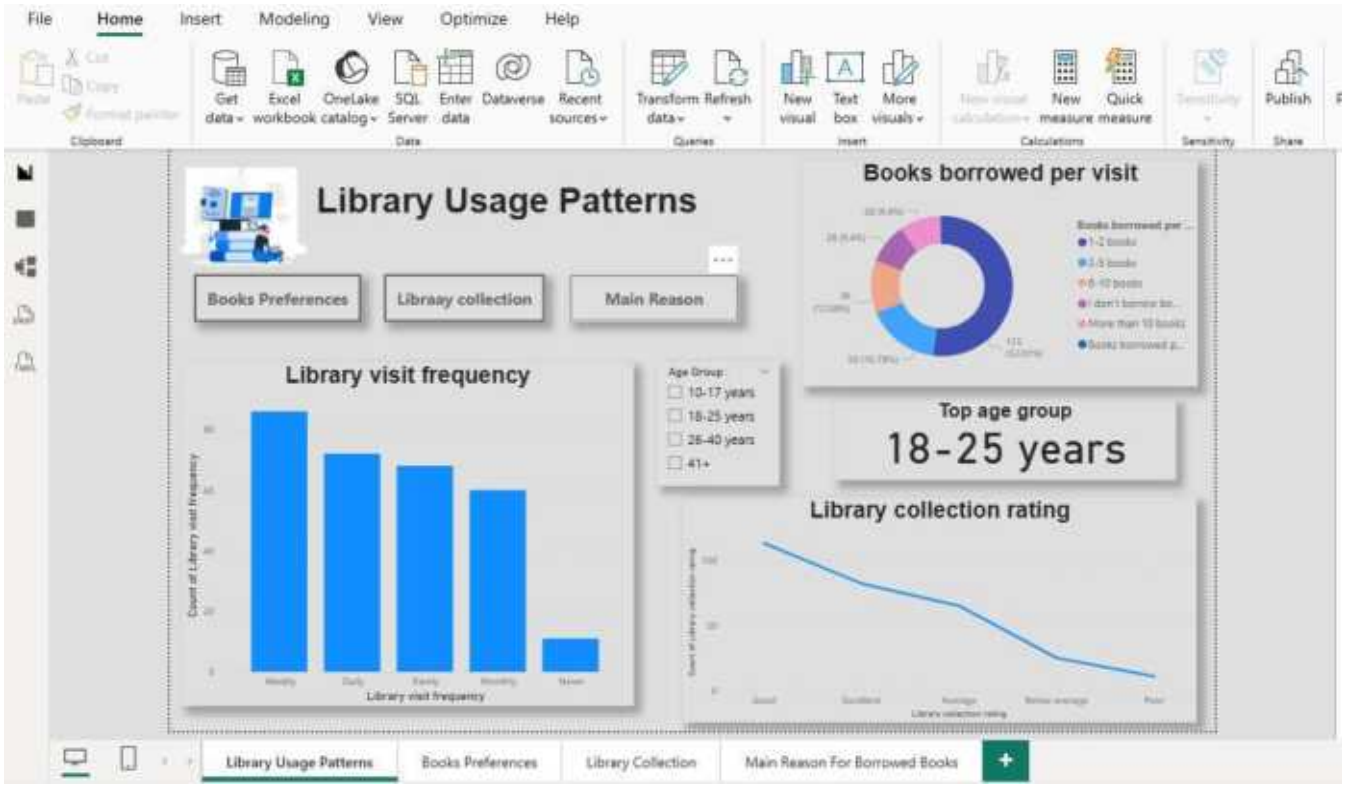
The focus was on increasing books that are in high demand and reducing or replacing underused materials.

Experimental Work:

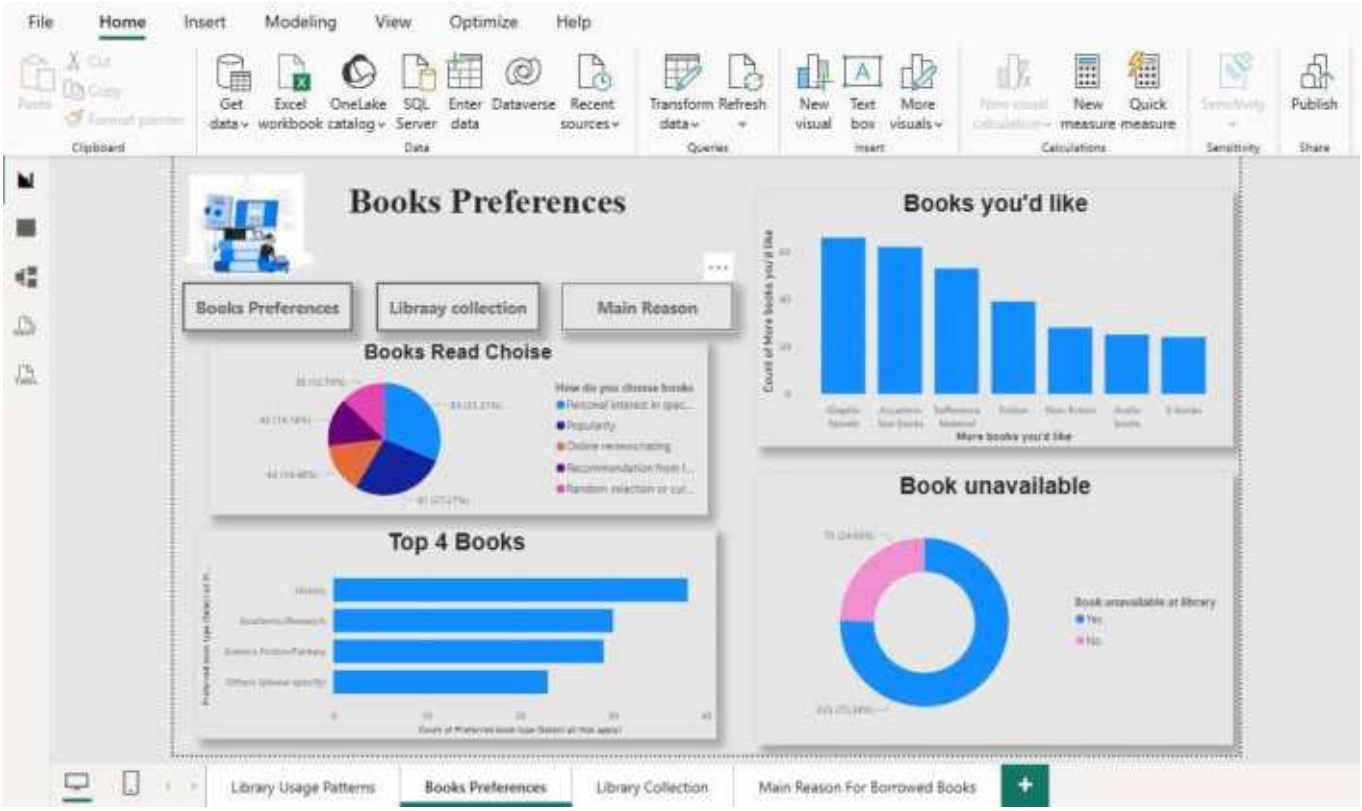
For this study ,I created a Google Forms to collect data from around 300 library users about their book borrowing habits and preferences. The data was cleaned in Excel and analyzed in Power BI using interactive dashboards. The analysis identified popular and less-used book categories, helping to make better decisions for optimizing the library collection.



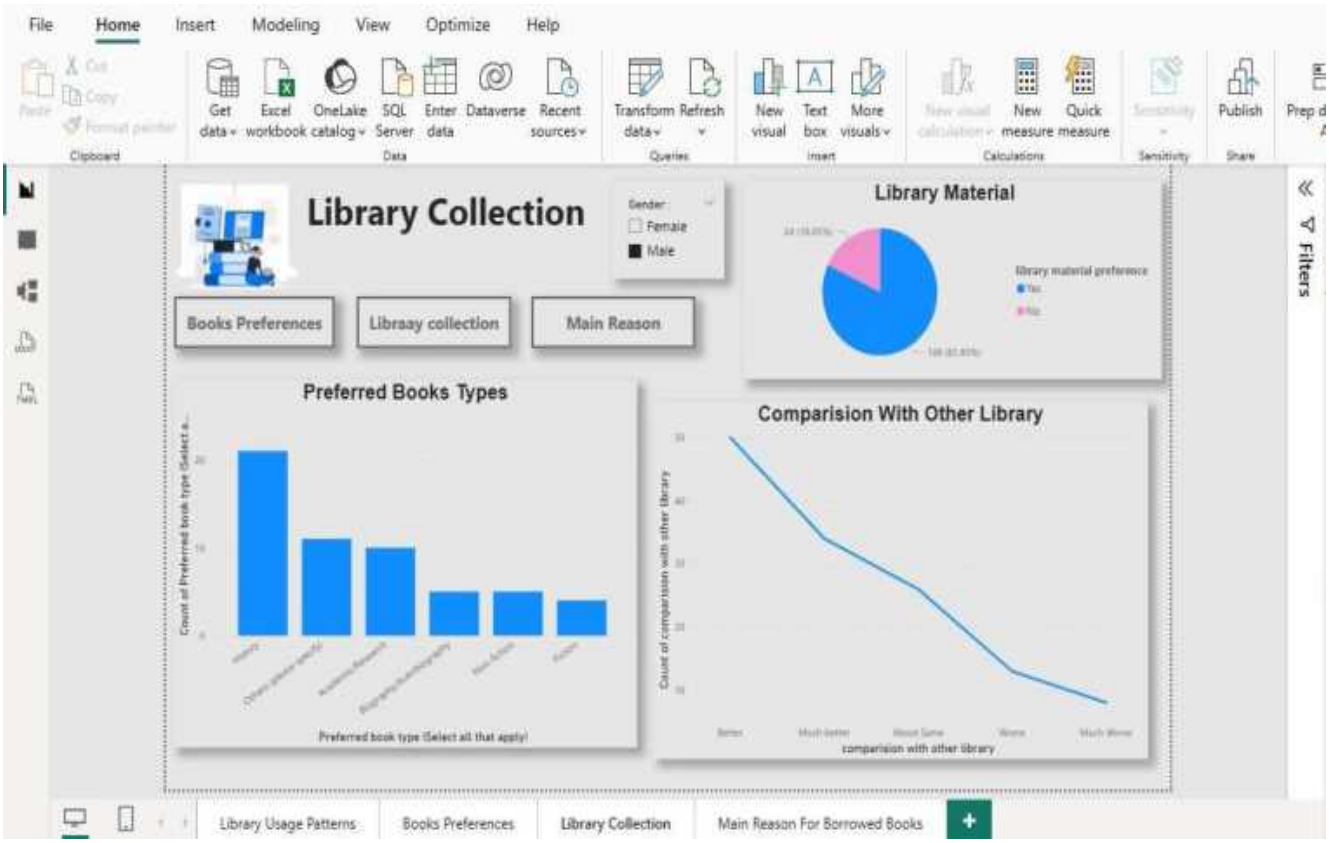
1) Library Usage Patterns



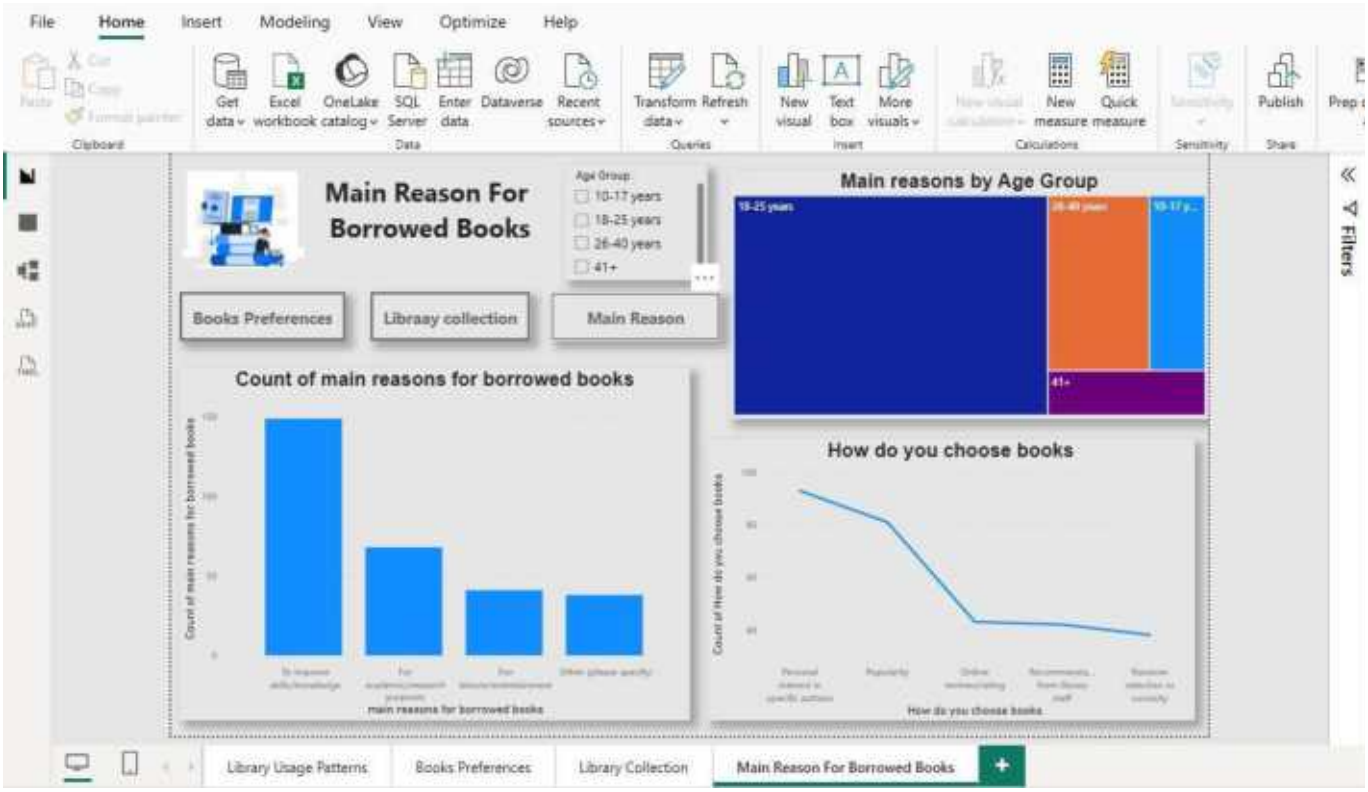
2) Books Preferences



3) Library Collection



4) Main Reason For Borrowed Books



Results:

The analysis showed that some book genres and authors are borrowed much more frequently than others, while a few sections have low usage. The Power BI dashboards helped clearly identify popular and less popular books.

These results indicate that the library can improve its collection by focusing more on high- demand books and reviewing low-demand categories.

Conclusion:

This study clearly shows that when libraries carefully analyze book borrowing data, they can better understand what their readers truly prefer. It becomes easy to see which books are popular and frequently borrowed and which ones are rarely used. By using Power BI, these patterns were identified in a simple and clear way, making decision-making much easier.

Overall, a data-driven approach helps libraries improve their collections, use their budget wisely, and ensure that they are meeting the real needs and interests of their readers.

References :

1. Sengar, D., Saxena, K., & Rathor, A. K. (2025). Optimizing Academic Library Collections Through Circulation Data Analysis. Journal of Library Data and Analytics, 20(1), 25–39.
2. Analysing Library Usage Patterns. (2024). Circulation Statistics and User Behavior Visualization in Higher Education Libraries. International Journal of Library Performance Measurement, 15(2), 112–128.
3. Silwattananusarn, T., & Kulkanjanapiban, K. (2020). Pattern Analysis of Library Borrowing Data Using Data Mining Techniques. Journal of Information Systems and Libraries, 34(3), 180–195.
4. Analysis of Book Circulation Data. (2022). Analytical Approaches for Predicting Library Resource Demand. Library Systems and Services Journal, 30(4), 210–224.
5. Neeraj. (2024). Application of Business Intelligence Tools in Library Usage Analysis. Journal of Digital Library Innovation, 12(1), 54–69.

6. Rathod, M. S. (2025). Strategic Planning Models for Sustainable Library Collection Development. International Journal of Library Planning and Development, 17(2), 98–113.
7. Luo, J. (2019). Usage Analysis of Printed Resources in Academic Institutions. Journal of Library Statistics and Evaluation, 41(2), 250–264.
8. Collection Development Literature Review. (2023). Principles and Modern Techniques in Academic Collection Management. Global Review of Library and Information Science, 26(6), 390–404.
9. Bidve, S., & Survase, P. (2024). Digital Transformation and Emerging Trends in Library Services. Journal of Information and Library Advancement, 11(3), 72–87.
10. Patron-Driven Acquisition. (2023). In Wikipedia: The Free Encyclopedia. Retrieved from https://en.wikipedia.org/wiki/Patron-driven_acquisition
11. Collection Development. (2023). In Wikipedia: The Free Encyclopedia. Retrieved from https://en.wikipedia.org/wiki/Collection_development
12. Riyar, K. (2024). Data-Based Decision Making for User-Centered Library Collections. International Journal of Academic Information Systems, 8(4), 132–147.

RESEARCH JOURNEY International Multidisciplinary E-Research Journal
Impact Factor (SJIF) - 6.625 | Special Issue 389 : Advanced Computing and Artificial Intelligence (ACAI)
ISSN : 2348-7143
March 2026

Uncertainty-Aware NLP-Based Automated Evaluation of Descriptive Responses

Ms. Dipali Ravindra Nhalde,
Mrs. Dhanashree Gajendrasing Patil
RCPET's Institute of Management Research and Development, Shirpur

Abstract

Automated evaluation of descriptive answers has gained importance in online learning, large-scale exams, and AI-based learning systems. Recent developments in transformer-based Natural Language Processing (NLP) models have greatly enhanced semantic evaluation and accuracy. However, most existing automated grading systems are deterministic, with a single-point prediction rather than uncertainty estimation. In high-stakes testing environments, the lack of awareness of uncertainty may impact reliability, fairness, and interpretability.

This review paper reviews the development of NLP-based automated descriptive answer evaluation and reviews novel uncertainty estimation approaches such as Bayesian neural networks, Monte Carlo dropout, deep ensembles, and calibration. The paper distinguishes between epistemic and aleatoric uncertainty and discusses their applicability in educational testing settings. The paper also reviews human-in-the-loop paradigms, confidence-driven decision-making, and ethics related to uncertainty-aware automated grading systems.

By reviewing recent developments and research gaps, this review paper points out the importance of standardized benchmarks, multilingual evaluation platforms, and explainable uncertainty modelling in automated scoring systems. The paper concludes that uncertainty quantification can be effectively integrated into NLP-based evaluation systems to make them more transparent, robust, and trustworthy for high-stakes educational testing environments.

Keywords: Uncertainty-aware learning, Automated essay scoring, Natural language processing, Transformer models, Bayesian neural networks, Monte Carlo dropout, Model calibration, Human-in-the-loop evaluation, AI in education

1. Introduction

The rapid expansion of online learning platforms, large-scale digital examinations, and AI-driven educational technologies has significantly increased the demand for automated evaluation systems capable of assessing descriptive responses. Unlike objective or multiple-choice questions, descriptive answers require deeper semantic interpretation, contextual reasoning, and subjective judgment. Traditional rule-based and statistical approaches provided limited semantic understanding, prompting the adoption of advanced Natural Language Processing (NLP) techniques for automated scoring [4].

Recent advancements in transformer-based models, such as BERT and RoBERTa, have revolutionized automated descriptive answer evaluation by enabling contextual language representation and transfer learning from large corpora [1]. These models have demonstrated remarkable improvements in scoring accuracy and generalization across diverse datasets [5]. However, despite their high predictive performance, most existing systems operate in a deterministic manner, producing single-point scores without quantifying prediction uncertainty.

In high-stakes educational settings—such as competitive examinations, university admissions tests, and professional certifications—deterministic scoring raises concerns regarding reliability, fairness, transparency, and accountability [5]. Descriptive responses inherently contain ambiguity, partial correctness, diverse valid interpretations, and linguistic variability. Consequently, automated grading systems must not only predict scores but also quantify the confidence associated with those predictions.

Uncertainty in deep learning models is generally categorized into epistemic uncertainty, which arises from limited or biased training data, and aleatory uncertainty, which stems from inherent ambiguity in input data [3]. In the context of descriptive answer evaluation, both types of uncertainty play a critical role. Modelling these uncertainties can improve system robustness, enable confidence-based decision-making, and support human-in-the-loop grading frameworks.

Recent research has explored uncertainty estimation techniques such as Bayesian neural networks, Monte Carlo dropout, deep ensembles, and calibration methods to enhance predictive reliability [2][3]. Integrating these probabilistic approaches with transformer-based NLP models provides a promising direction for building trustworthy automated grading systems. Such systems can flag low-confidence predictions for human review, adapt scoring rubrics dynamically, and reduce overconfident misclassifications.

This review paper examines the evolution of automated descriptive response evaluation, analyses contemporary uncertainty modelling techniques, and discusses their applicability in educational assessment contexts. It further highlights research gaps, ethical considerations, and future directions toward developing transparent, fair, and uncertainty-aware NLP-based evaluation frameworks for high-stakes testing environments.

2. Evolution of Automated Descriptive Response Evaluation

2.1 Early Rule-Based and Statistical Systems

Initial systems relied on:

- Surface-level linguistic features
- Grammar and syntax rules
- Latent Semantic Analysis (LSA)
- Linear regression and SVM-based scoring

These foundational approaches are comprehensively discussed in early Automated Essay Evaluation literature [4]. Limitations included shallow semantic understanding and poor generalization.

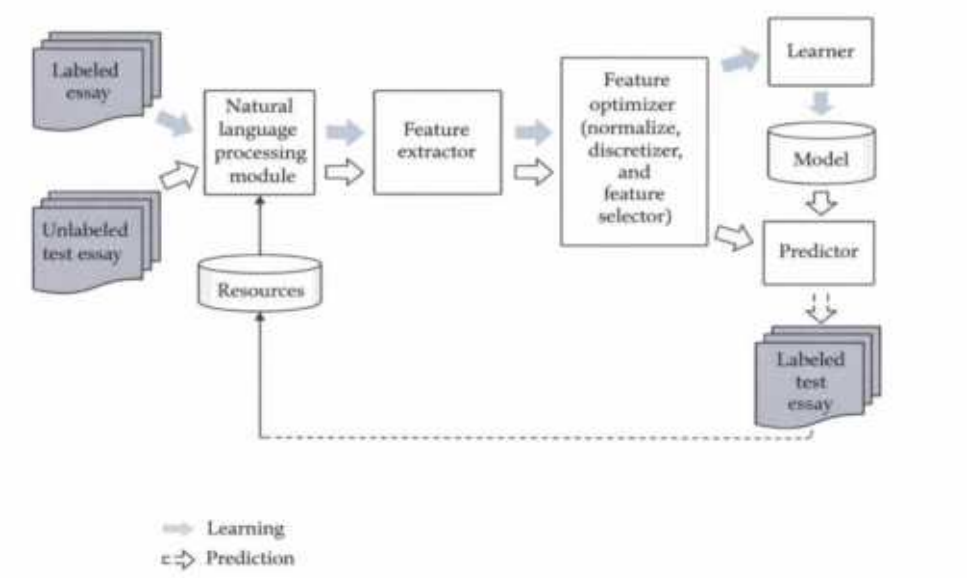
2.2 Machine Learning-Based Scoring

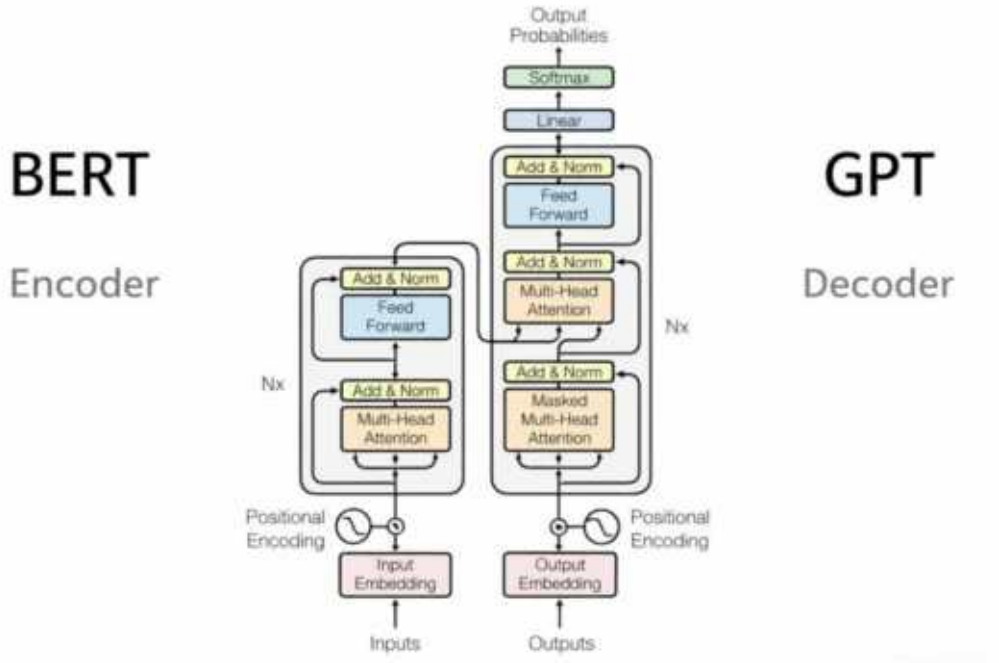
Supervised learning approaches introduced:

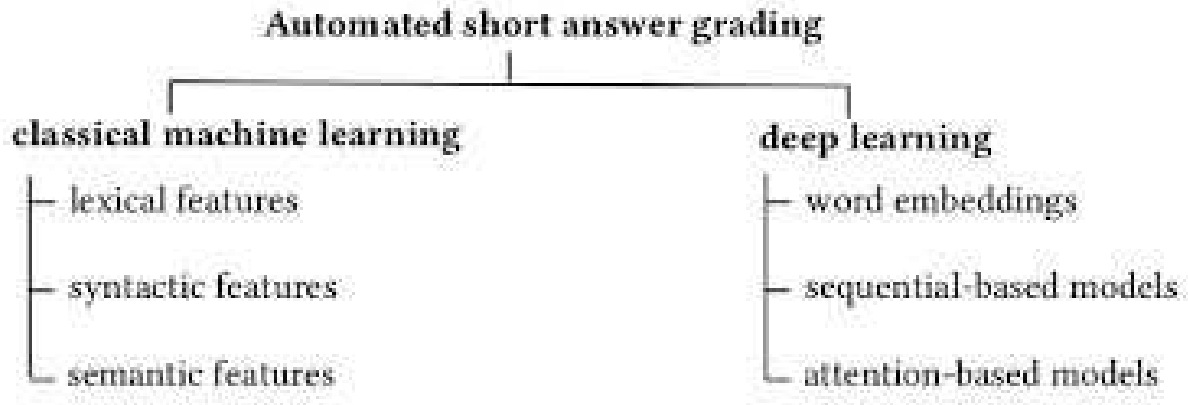
- Feature engineering (word count, cohesion, readability)
- Regression-based scoring
- Ensemble methods

Although performance improved, models remained dependent on handcrafted features and limited semantic representation capacity [4][5].

2.3 Deep Learning and Transformer-Based Approaches







Transformer models enabled:

- Context-aware semantic encoding
- Transfer learning from large corpora
- End-to-end scoring frameworks

The introduction of BERT significantly advanced contextual language representation learning for downstream NLP tasks [1]. Contemporary automated essay scoring systems leverage such transformer architectures for improved semantic evaluation [5]. Despite high accuracy, deterministic predictions limit reliability in ambiguous cases.

3. Understanding Uncertainty in Deep Learning

Uncertainty in neural networks is broadly categorized into:

3.1 Epistemic Uncertainty

- Caused by limited or biased training data
- Reduced with more data
- Modeled using Bayesian techniques [3]

3.2 Aleatoric Uncertainty

- Caused by inherent ambiguity in input data
- Cannot be reduced by additional data
- Common in descriptive answers with multiple valid interpretations [3]

The theoretical distinction between epistemic and aleatoric uncertainty is formally described in Bayesian deep learning literature [3].

4. Uncertainty Estimation Techniques

4.1 Bayesian Neural Networks (BNNs)

BNNs treat model weights as probability distributions rather than fixed values [3].

Advantages:

- Theoretical robustness
- Principled uncertainty modelling

Limitations:

- High computational cost

4.2 Monte Carlo Dropout

Monte Carlo Dropout approximates Bayesian inference by enabling dropout during inference and performing multiple forward passes [2].

Benefits:

- Easy integration with transformer models
- Computationally feasible

4.3 Deep Ensembles

Multiple independently trained models generate predictions, and variance across models estimates uncertainty. Ensemble-based uncertainty estimation has demonstrated strong empirical performance in deep learning research [3].

Strength:

- High empirical performance

Weakness:

- Increased storage and training cost

4.4 Calibration Techniques

Common methods:

- Temperature Scaling
- Platt Scaling
- Isotonic Regression
- Expected Calibration Error (ECE) minimization

Calibration ensures that predicted confidence aligns with actual accuracy, improving reliability in automated scoring systems [5].

5. Integration of Uncertainty in Educational NLP

Research integrating uncertainty into automated grading is limited but growing [5]. Applications include:

- Flagging low-confidence answers for human review
- Improving fairness across demographic groups
- Reducing overconfident misclassifications
- Adaptive feedback generation

Recent transformer-based systems integrate Monte Carlo dropout with BERT-based scorers to produce score distributions instead of single predictions [1][2].

6. Human-in-the-Loop Evaluation Systems

Uncertainty-aware systems enable hybrid evaluation:

1. High-confidence responses → Automatically graded

2. Medium-confidence responses → Soft review
3. Low-confidence responses → Mandatory human evaluation

Such hybrid frameworks align with contemporary automated essay scoring paradigms emphasizing reliability and transparency [5].

7. Comparative Analysis

Approach	Accuracy	Interpret-ability	Computa-tional Cost	Uncer-tainty Output
Rule-Based	Low	High	Low	No
ML Regression	Medium	Medium	Medium	No
Transformer (Deterministic)	High	Low	High	No
Bayesian / MC Dropout	High	Medium	High	Yes
Deep Ensembles	Very High	Medium	Very High	Yes

8. Challenges and Limitations

Issues such as grading bias, overconfidence in neural networks, and lack of benchmark datasets are widely discussed in automated essay scoring research [4] [5]. Overconfidence in deep neural networks has been extensively analyzed in uncertainty modelling literature [3].

9. Ethical and Fairness Considerations

Uncertainty-aware systems can:

- Detect grading ambiguity
- Reduce demographic bias
- Improve transparency
- Support explainable AI

However, challenges remain in:

- Model interpretability
- Bias auditing
- Accountability in high-stakes exams

10. Research Gaps Identified

1. Limited large-scale benchmark datasets for uncertainty-aware grading
2. Few multilingual uncertainty-aware AES systems
3. Lack of standardized evaluation metrics for grading uncertainty
4. Minimal integration with explainable AI techniques
5. Scarce research on uncertainty-aware rubric adaptation

11. Future Research Directions

- Hybrid epistemic–aleatoric modelling for descriptive answers
- Integration with Explainable AI (XAI)
- Federated uncertainty-aware evaluation systems
- Cross-lingual uncertainty-aware grading
- Adaptive scoring rubrics based on confidence levels

12. Conclusion

The evaluation of descriptive answers has been significantly improved using transformer-based NLP models such as BERT [1]. However, deterministic evaluation approaches fail to capture the uncertainty inherent in natural language. Techniques such as Bayesian inference [3], Monte Carlo dropout [2], and ensemble modelling can enhance reliability and fairness in automated grading systems. Future work should focus on benchmark development, multilingual systems, and principled uncertainty modelling in educational assessment frameworks [5].

References :

1. Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT) (pp. 4171–4186). Association for Computational Linguistics. Available at: https://aclanthology.org/N19-1423/
2. Gal, Y., & Ghahramani, Z. (2016). Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning. In Proceedings of the 33rd International Conference on Machine Learning (ICML 2016) (pp. 1050–1059). arXiv preprint arXiv:1506.02142. Available at: https://arxiv.org/abs/1506.02142
3. Kendall, A., & Gal, Y. (2017). What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision? In Advances in Neural Information Processing Systems 30 (NeurIPS 2017) (pp. 5574–5584). Available at: https://arxiv.org/abs/1703.04977
4. Shermis, M. D., & Burstein, J. (Eds.). (2013). Handbook of Automated Essay Evaluation: Current Applications and New Directions. New York, NY, USA: Routledge. ISBN: 978-0415810968.
5. Beigman Klebanov, B., & Madnani, N. (2022). Automated Essay Scoring. Synthesis Lectures on Human Language Technologies, 15(1), 1–294. Springer. DOI: 10.1007/978-3-031-02182-4 (for recent work on automated evaluation that includes uncertainty perspectives)

RESEARCH JOURNEY International Multidisciplinary E-Research Journal
Impact Factor (SJIF) - 6.625 | ***Special Issue 389 :*** *Advanced Computing and Artificial Intelligence (ACAI)*
ISSN : 2348-7143
March 2026

Comparative Study of Lightweight Transfer Learning Architectures for Static Indian Sign Language Alphabet and Digit Recognition

Ansh Icecreamwala[1], Utsav Gaywala[1], Dhyey Desai[1], Rakesh Savant[1]
[1]Babu Madhav Institute of Information Technology, Uka Tarsadia University, Bardoli, Gujarat

Abstract.

The current paper is a systematic comparison of three example convolutional neural network models, i.e. InceptionV3, ResNet50, and MobileNetV2, in distinguishing between the static Indian Sign Language (ISL) alphabet and digits. Automatic Sign Language Recognition (ASLR) systems seek to minimize the communication barrier between Deaf and Hard-of-Hearing people and hearing people by converting the visual gestures to the textual ones. The data in this work represents 72,000 labelled images in 36 classes (26 alphabets and 10 digits). Controlled experimental conditions were followed to split the data into training and validation subsets (80:20) to have a fair benchmarking of the transfer learning strategy through a two-stage training procedure. The backbone networks were trained in the first stage with the use of pretrained frozen feature extractors and only a newly added classification head was trained. The second stage entailed the application of selective fine-tuning of the upper layers with a smaller learning rate to allow domain adaptation without forgetting learned representations. The validation accuracy and macro-averaged precision, recall, and F1-score were used to measure performance, and the model parameter analysis was used to measure the computational efficiency. Experimental findings prove that fine-tuning has a substantial positive effect on classification performance in all architectures. ResNet50 had the best validation accuracy, then InceptionV3, and MobileNetV2 also gave good accuracy at the cost of much less parameter complexity. The results also indicate the trade-off that exists between predictive performance and the cost of computation and the need to have controlled comparative assessment when deciding on architectures of scalable and deployable ISL recognition systems.

Keywords: Indian Sign Language (ISL), Transfer Learning, Lightweight Convolutional Neural Networks, Machine Learning, Comparative Study

1. Introduction

Interaction between Deaf and Hard-of-Hearing people and the hearing society continues as one of the great challenges in society. The Sign languages are used by millions of individuals across the globe, but, due to a general absence of sign language education among the non-signs, there is always a communication obstacle. The objective of the Automatic Sign Language Recognition (ASLR) systems is to fill this gap by converting visual hand gestures to the spoken or textual language using computer vision and deep learning methods.

Deep convolutional neural networks (CNNs) and transfer learning have significantly enhanced the performance of image classification in all kinds of fields in recent years. Trained architectures, including InceptionV3[2], ResNet50[3], and MobileNetV2[4], have shown good features extraction ability when used on a domain-specific task. Transfer learning minimizes the very huge task-driven datasets as well as decreases the training time to a considerable extent, which is appropriate in the context of real-world sign language recognitions systems [1].

Even with these developments, most studies are conducting a single architecture or are mainly looking at the improvement of accuracy without making organizing comparative studies across various models under homogenous experimental conditions. Furthermore, little effort is usually applied in terms of examining the effects of freezing trained layers and fine-tuning them to adapt to a different domain.

The study is a comparative analysis of three popular pretrained CNN models InceptionV3, ResNet50, and MobileNetV2 in terms of recognizing Indian Sign Language characters at rest. Training and fine-tuning of all the models are done in a single experimental pipeline so that they are fairly compared. The experiment compares the performance of frozen feature extraction, fine-tuning, and overall generalization behavior of architectures.

The key contributions of the work include a systematic and organized comparative study of several transfer learning architectures which are tested on a 36-class fixed sign language dataset. Based on the experiment under the conditions of extreme similarity, the research paper examines the performance disparities between frozen feature extraction and fine-tuning approaches, allowing to make a just judgment of the effect of the depth of training on model behavior. Comparison of classification accuracy, stability of learning and convergence pattern of the chosen architectures is detailed using a quantitative evaluation. This work aims to define an optimal tradeoff between predictive accuracy and computational efficiency by studying the results of both performance and training dynamics. The general aim is to give viable ideas that help in formulating scalable, efficient, and real-world deployable sign language

recognition systems.

2. Related Work

Computer vision Automatic sign language recognition (SLR) has been an active field of research because it can offer opportunities to reduce communication barriers to Deaf and Hard-of-Hearing people. Conventional methods were based on manually-designed characteristics like Histogram of Oriented Gradients (HOG) integrated with classical classifiers like Support Vector Machines (SVMs), and could not observe the advanced spatial variability of hand gestures. Convolutional Neural Networks (CNNs) are the deep-learned model of visual sign recognition that, due to their hierarchical feature representations, automatically acquire the feature representations of images represented in raw image data and have shown superior performance to classical methods

Some works have used deep CNNs to perform recognition of the static sign language. Particularly, transfer learning has been extensively used to enhance the accuracy of recognition in low-label space areas [1]. Fine-tuning pretrained networks, including InceptionV3 [2], ResNet50 [3], and MobileNetV2 [4], on gesture datasets have been found to be faster to converge and better at generalization than training a network from scratch. Transfer learning method has been demonstrated to be effective in gesture and sign recognition tasks over a variety of regional datasets.

The relative analysis of various pretrained models have also been considered. One interesting paper in the journal Informatics in Medicine Unlocked (2022) compared various pretrained CNN models such as VGG16, VGG19 and InceptionV3 and AlexNet to the Bangladeshi sign word recognition [5]. That paper has identified the role of architectural depth and transfer learning methods in determining classification accuracy and the significance of guided benchmarking in the controlled experimental environment. These comparative studies have shown that the pre-existing visual representations can be successfully cross-regionalized to various regional sign languages.

In addition to the static recognition, sign language research is included in the dynamic gesture recognition, of which motion sequences in video need to be modelled in time. The 3D CNNs or recurrent networks architecture is commonly used by dynamic systems to accommodate temporal dependencies. Even though more linguistic structure is represented by dynamic recognition, static gesture recognition is a benchmark problem and a fundamental part to the task of alphabet and digit classification

Regardless of the effectiveness of the transfer learning-based CNN models, the current literature often varies in terms of the experimental settings, preprocessing approaches of the dataset, and assessment schemes. These differences complicate direct architectural-to-architectural comparison and reduce reproducibility. In addition, systematic performance assessments of various lightweight pretrained models with the same experimental settings during the statical recognition of Indian Sign Language (ISL) alphabets and digits are scarce. This is the gap that inspires the necessity of a systematic comparative assessment which will be undertaken through a common training pipeline.

3. Proposed Methodology

This part describes the experimental design of the comparative analysis of the chosen pretrained models, such as dataset preparation, the transfer learning approach, and controlled training conditions.

3.1 Dataset Description

The dataset was taken by this research as a publicly available dataset of the Indian Sign Language which is available on Kaggle [6]. The initial data was composed of around 154,000 pictures, which were arranged into various demographic groups (teen, kids, adult), clothing changes (with sleeves, without sleeves) and gesture types, such as alphabets, numerals, and Hindi vowels.

To affect this study, static classes of alphabets and numerals were taken only. The images of Hindi vowels were left out to keep the classification problem of 36 classes (26 alphabets and 10 numerals) constant. Along with that, the pictures in demographic and clothing differences were combined in each category of gestures to achieve a variety of age and appearance without overturning the labels. A balanced set of 72000 images was obtained after filtering and reorganizing, which were evenly distributed among the 36 chosen classes. Each photograph is one hand gesture that has a particular alphabet or figure.

The dataset was separated into a training subset and a validation subset to guarantee reproducibility, and this was carried out through an 80:20 cutoff at a set random seed (SEED = 42). It had led to the 57,600 and 14,400 images, respectively, of training and validation. No other data augmentation method was used so as to have similar benchmarking across architectures.

The size of all images was reduced to 224 x 224 pixels to fit the input demands of the chosen pre-trained CNN models.

3.2 Transfer Learning Framework

This study adopts a transfer learning approach using pretrained convolutional neural networks originally trained on the ImageNet dataset. Transfer learning allows the model to leverage previously learned visual representations from large-scale datasets and adapt them to domain-specific gesture classification tasks [1]. Three widely used pretrained architectures were selected for this work, namely InceptionV3 [2], ResNet50 [3], and MobileNetV2 [4]. For each architecture, the original fully connected classification layer was removed and replaced with a custom classification

head. The modified head consists of a Global Average Pooling layer followed by a Dense layer with 256 neurons using ReLU activation, a Dropout layer for regularization, and a final Dense layer with 36 output units and softmax activation. The softmax layer produces class probabilities corresponding to the 36 ISL categories.

3.3 Training Strategy

Training was performed in two stages to analyze the effect of frozen feature extraction and fine-tuning:

Stage 1: Frozen Feature Extraction

During the first stage, all the layers of the pretrained backbone were frozen. The newly added classification head was the only person who was trained. The model was trained with Adam optimizer and the learning rate of 1e-4 in 15 epochs. The loss that was employed was sparse categorical cross-entropy. This stage evaluates the effectiveness of pretrained feature representations without domain-specific adaptation.

Stage 2: Fine-Tuning

The second stage involved unfreezing the top 30 layers of the trained backbone with the rest of the layers unfrozen. The model had a different learning rate of 1e-5 compiled again to avoid massive weight change and catastrophic forgetting.

Up to 8 epochs of fine-tuning, which used early stopping on validation loss, were done. The most successful validation accuracy was used to save model checkpoints.

This phase determines the effect of domain adaptation on the classification performance.

3.4 Evaluation Metrics

Model performance was evaluated using classification accuracy and validation loss. Accuracy was calculated as the percentage of correctly classified samples out of the total number of validation samples. All architectures were compared under identical training conditions to ensure a fair evaluation.

3.5 Experimental Environment

Experiments were conducted using TensorFlow and Keras frameworks. Training was performed on a GPU-enabled environment to accelerate computation. The batch size was set to 32 for all models to maintain consistency across experiments. All experiments were conducted using a fixed random seed to ensure reproducibility.

3.6 Computational Complexity Analysis

Besides the accuracy of classification, the computational efficiency was also considered to give a balanced comparison of pretrained architectures. The number of total parameters, the number of parameters trainable during frozen training as well as the time of training of each model were recorded under the same experimental conditions.

Besides the accuracy of classification, the computational efficiency was also considered to give a balanced comparison of pretrained architectures. The number of total parameters, the number of parameters trainable during frozen training as well as the time of training of each model were recorded under the same experimental conditions.

The average training time per epoch was determined with a fixed batch size of 32 in a GPU enabled environment. MobileNetV2 was the lowest in computational cost as well as quicker in training time, followed by InceptionV3 and ResNet50. Despite the fact that InceptionV3 and ResNet50 had a higher classification accuracy, MobileNetV2 had a good trade-off between performance and computational efficiency.

The practical significance of this comparison in designing scalable sign language recognition systems is that recognition accuracy and model complexity should be balanced.

4. Results and Discussion

The experimental conditions were the same in testing the proposed models so that they could be compared fairly. This part gives the empirical findings and compares the relative performance of the architectures as per the classification accuracy, generalization behaviour, and computational efficiency.

4.1 Quantitative Performance Evaluation

The experimental test was applied to a 36 class static sign language dataset with the same training parameters across all the architectures so as to provide fair comparison. Table X provides an overview of the quantitative outcomes of performance.

ResNet50 was reported to have the greatest validation accuracy of 96.84, making it a good classifier and without any convergence behaviour. Precision, recall and F1-score that were macro-averaged were about 0.97, which means that the performance was balanced in terms of class and did not show much bias based on the category.

InceptionV3 had a validation accuracy of 95.31 which is close to ResNet50. The model had a steady convergence and good levels of generalization, with the macro-averaged precision, recall and F1-scores of about 0.95, indicating consistent predictions of most of the classes.

MobileNetV2 had an accuracy of 94.40% when trained on the validation part with a much low number of parameters (2.59 million) than InceptionV3 (~ 22 million) and ResNet50 (~24 million). Although MobileNetV2 is significantly smaller in model size, it retained macro-averaged precision, recall, and F1-scores of about 0.94, indicating its high parameter efficiency and appropriateness to use in resource-constrained settings. Figure 1 : Normalized confusion matrix of ResNet50 shows the normalized confusion matrix of the best-performing ResNet50 model

to examine the performance of the class-wise and misclassification behaviour.

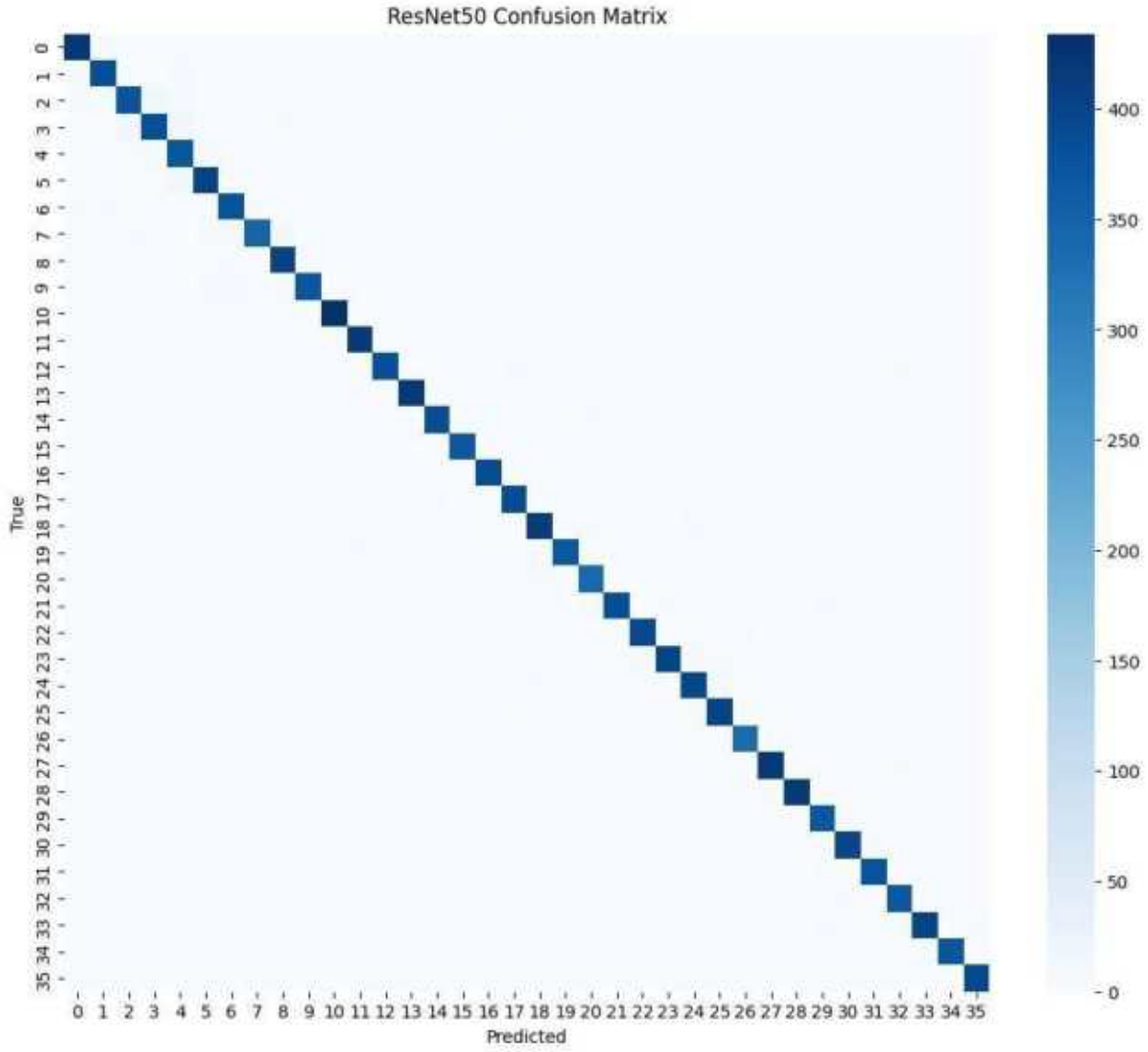


Figure 1 : Normalized confusion matrix of ResNet50

In general, ResNet50 achieved the best predictive results, and MobileNetV2 also offered the most desirable trade-off between quality and energy requirements. These findings show that more elaborate residual architectures are able to maximize the accuracy of classification, and lightweight models like MobileNetV2 can be used to provide scalable and practical solutions to real-time or embedded sign language recognition systems.

Table 1 : Validation performance comparison of pretrained CNN models for ISL recognition

Model	Val Accuracy (%)	Precision	Recall	F1-Score	Para-meters (M)
ResNet50	96.84	0.97	0.97	0.97	~24
InceptionV3	95.31	0.95	0.95	0.95	~22
MobileNetV2	94.40	0.94	0.94	0.94	2.59

In order to further describe the trade-off between predictive performance and model complexity, Figure 2 : Illustrating the performance - complexity trade - off gives a comparative graphical representation of validation accuracy versus parameter size of each of the assessed architecture. This representation allows making a more explicit evaluation of efficiency-performance balance between models.

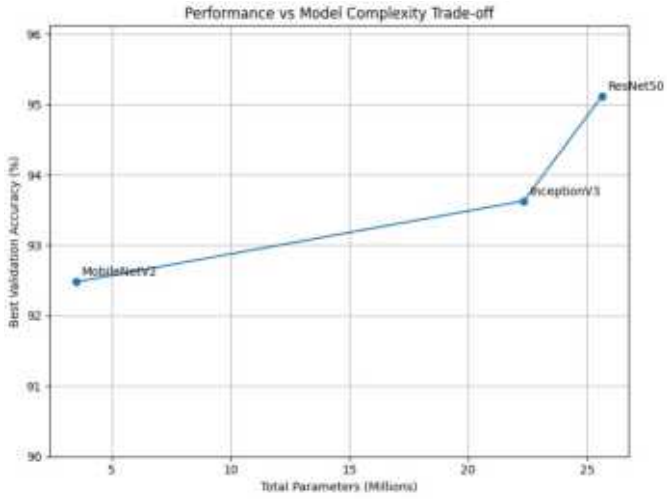


Figure 2 : Illustrating the performance - complexity trade - off

Although the difference in performance between the architectures can be observed, formal statistical significance testing of multiple independent runs is the subject of future work. Training and validation accuracy curves were very similar in all the architectures, which showed that convergence is stable and overfitting is minimized in the fine-tuning model. As seen in Figure 3 : Training and Validation Accuracy Curves for Transfer Learning Models and Figure 4 : Training and Validation Loss Curves for Transfer Learning Models, all the models have stable convergence and a small generalization gap.

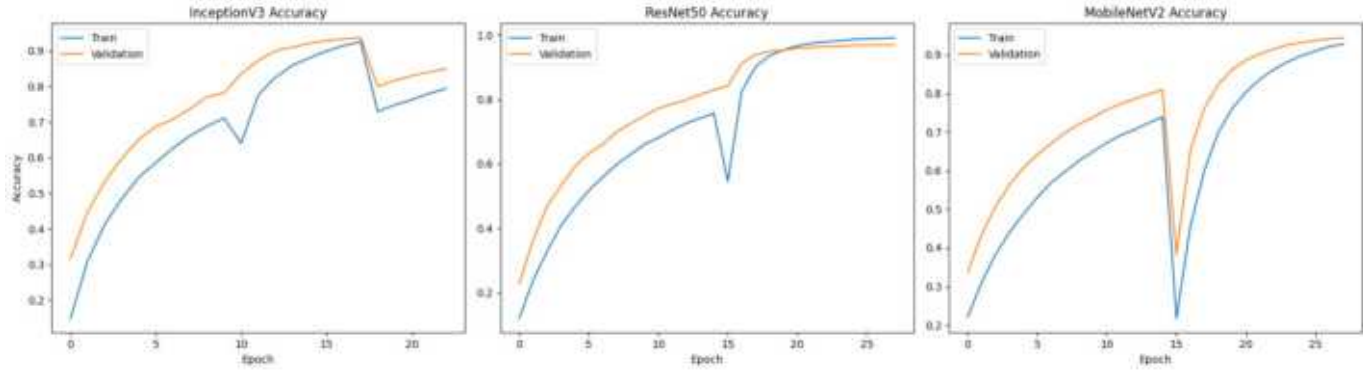


Figure 3 : Training and Validation Accuracy Curves for Transfer Learning Models

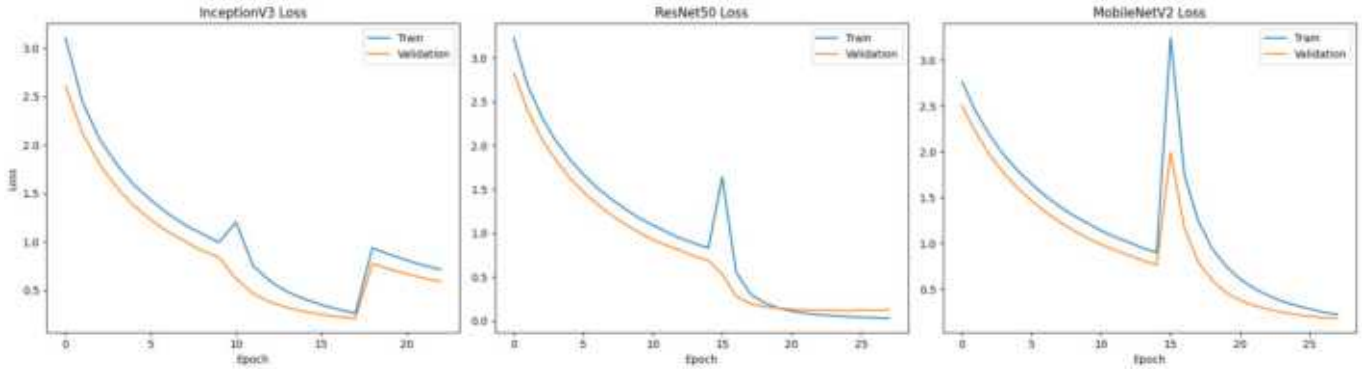


Figure 4 : Training and Validation Loss Curves for Transfer Learning Models

The comparative analysis shows obvious performance discrepancies between the chosen pretrained architectures when they are put into the same experiment conditions. ResNet50 had the highest validation accuracy and macro-average performance metrics thus it has an improved capability to represent features when identifying static ISL gestures. It has a residual learning structure that probably enables more significant feature extraction, which leads to better discrimination of the classes.

InceptionV3 showed good results in terms of competition with a marginally lower accuracy than that of ResNet50. It has an efficient feature learning strategy (its factorized convolution strategy) but it is found that deeper residual connections in ResNet50 can learn features slightly better on this dataset.

Although MobileNetV2 has the lowest validation accuracy of the three models, it has a higher parameter efficiency that has a significantly lower model size. The lightweight nature of the architecture also makes it especially applicable to real-time or resource-constrained deployment situations where the computational overhead should be kept to a minimum.

The findings indicate that more predictive accuracy is achieved in deeper architectures, and lightweight models provide viable trade-offs between predictive performance and efficiency. Thus, application-related constraints, e.g. deployment conditions and latency, must be used to choose the model.

Although there are observable differences between architectures, formal statistical significance testing between a number of independent runs is a future work.

5. Conclusion

This paper has provided a systematic comparative analysis of three pre-trained convolutional neural network models InceptionV3, ResNet50, and MobileNetV2 in order to recognize alphabets and digits of the Indian Sign Language. All models had excellent classification results using a single experimental pipeline and a two-stage transfer learning strategy. The performance of the backbone was fine-tuned to reach a high level of accuracy when one was compared with frozen feature extraction. ResNet50 had the best validation accuracy of the models evaluated, whereas MobileNetV2 attained a desirable balance between the efficiency and performance of the model. These results allow noting that the choice of architectures should not be based only on accuracy but also on the limits of deployment. Future-related development could be the extension of this research to dynamic gesture recognition and the optimization of deployment on the fly.

Even though there was good validation performance, this study is only restricted to static movements that were taken in controlled environments. Split of the dataset was done randomly at the image level, which is not capable of completely removing subject-specific bias. Moreover, there was no cross-subject evaluation or real time deployment analysis performed. These limitations will be dealt with in the future by subject-independent validation, dynamic gesture modelling, and real-time performance benchmarking.

References :

1. Pan, Sinno Jialin, and Qiang Yang. "A survey on transfer learning." IEEE Transactions on knowledge and data engineering 22, no. 10 (2009): 1345-1359.
2. Szegedy, Christian, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. "Rethinking the inception architecture for computer vision." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2818-2826. 2016.
3. He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Deep residual learning for image recognition." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770-778. 2016.
4. Sandler, Mark, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. "Mobilenetv2: Inverted residuals and linear bottlenecks." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4510-4520. 2018.
5. Islam, Md Monirul, Md Rasel Uddin, Md Nasim AKhtar, and KM Rafiqul Alam. "Recognizing multiclass Static Sign Language words for deaf and dumb people of Bangladesh based on transfer learning techniques." Informatics in Medicine Unlocked 33 (2022): 101077.
6. Singh, Animesh, Singh, Sunil K, Mittal, Ajay, and Gupta, Brij B. (2025). Static Gestures of Indian Sign Language (ISL) [Data set]. Kaggle. https://doi.org/10.34740/KAGGLE/DSV/13805053

RESEARCH JOURNEY International Multidisciplinary E-Research Journal
Impact Factor (SJIF) - 6.625 | ***Special Issue 389 :*** *Advanced Computing and Artificial Intelligence (ACAI)*
ISSN : 2348-7143
March 2026

AI-Assisted Software Requirement Analysis using NLP

Prof. Vrushali S. Tambe, Prof. Ashwini K. Sonawane
RCPET's Institute of Management Research and Development, Shirpur . District Dhule, Maharashtra, India.

Abstract

Requirement analysis is crucial to the success of a software project and is one area of software engineering that has seen tremendous growth over the years. Earlier, requirement analysis in the same way it would have been completed mainly through reading through documentation provided by the customer and recognizing what was needed from the software system to achieve their requirements. The encounter with this methodology is that it is not only manual but also takes up an enormous extent of time to analyse requirements and is likely to human mistake.

The main goal of this paper is to outline an AI powered, Natural Language Processing (NLP) based framework for automatic extraction and classification of software requirements from text-based documents. The solution supports parting of functional versus non-functional requirements via keyword-based analysis and syntactic processing. Additionally, a practical implementation of the proposed framework is offered through the Admission Management System use case featuring Python and spaCy where the proposed framework was able to effectively extract structured requirements from unstructured text. Complete, this approach will reduce the amount of manual work done on software requirement specification (SRS) processes, whereas increasing the accuracy and efficiency of how software requirements are tracked throughout the SRS process.

Keywords:

Artificial Intelligence, Natural Language Processing, Software Engineering, Requirement Analysis, Automation

I. Introduction

Every software development process starts with software requirement analysis, or SRA. It entails comprehending, recording, and overseeing stakeholders' requirements and expectations. The requirements for the majority of software projects, however, are written in natural language, which makes them unclear, inconsistent, and challenging for automated analysis.

Human interpretation is the foundation of manual requirement analysis, which frequently results in inconsistent classifications and overlooked details. Recent developments in Natural Language Processing (NLP) and Artificial Intelligence (AI) present fresh chances to automate this procedure [2], [3]. Because NLP enables machines to read and comprehend human language, requirement statements can be automatically recognized and categorized.

This paper proposes an AI-Assisted Requirement Analysis Framework that uses NLP to analyse text data, identify relevant statements of requirements, and classify them into Functional and Non-Functional Requirements. This method is intended to help software engineers by providing them with an intelligent assistant that can speed up the process of requirement documentation.

II. Literature Review

Several approaches have been investigated to automate requirement engineering. The initial studies involved rule-based systems that searched for patterns like "The system shall…" to identify requirements. These systems were inflexible and unable to handle variations in natural language.

With the advent of NLP and machine learning, BERT, LSTM, and GPT-based models have been employed to automatically classify requirement sentences [3], [6].

Alhoshan et al. [1] (2020) designed an NLP-based classifier for the identification of functional requirements in software documentation.

Khan et al. [6] (2021) employed deep learning models for requirement sentences to identify ambiguity.

IBM Watson proved the feasibility of NLP in requirement management through natural language queries [5].

Though these studies appear promising, most of them demand large amounts of labeled data and intensive model training. In contrast, this study proposes a simple and easy-to-implement framework, which is feasible in an educational setup and small-scale software development.

III. Research Gap And Motivation

Though previous research works, such as Alhoshan et al. [1], have successfully classified functional requirements using machine learning models, there are still some research gaps that are not addressed. Their method was computationally expensive and required large amounts of labeled data, making it unfeasible for small-scale or academic projects. In addition, the study only focused on functional requirements, while non-functional requirements, including performance, usability, and security, were not explored at all. There was also a lack of simple and ready-to-use frameworks that could have shown the capability of requirement extraction and visualization in a simple and low-cost environment.

The research aims to address these research gaps by introducing a new, lightweight AI-assisted framework that

utilizes NLP for the automatic extraction and classification of functional and non-functional requirements. The framework utilizes free resources such as Python, spaCy, and Google Colab [4], which makes it unnecessary to have large amounts of data or expensive hardware. It can also be easily implemented in academic and small-scale industrial projects and can also display dependencies or entities through NLP visualization libraries.

IV. Objectives

The objectives of this study are:

1. To design and develop a simple NLP-based framework for requirement extraction.
2. To automatically classify the extracted requirements into functional and non-functional requirements.
3. To validate the effectiveness of the framework with a case study (Admission Management System).
4. To prove that AI-based analysis minimizes human effort and maximizes consistency in requirement documentation.

V. Methodology

A. Overview

The proposed framework adopts a step-by-step NLP process to examine text documents and extract requirements. The process includes data input, text preprocessing, feature extraction, classification, and output generation in a structured format.

B. Steps in the Framework

Data Input: Gathering text-form requirement descriptions from client documents, interviews, or emails.

Text Preprocessing: Cleaning the text by removing punctuation, stop words, and converting to lowercase.

Sentence Segmentation: Breaking down text into individual sentences using spaCy's language model [4].

Tokenization and POS Tagging: Locating nouns (actors) and verbs (actions) to identify requirement statements.

Keyword Matching and Classification: Matching predefined lists of action words (e.g., fill, verify, approve) for functional requirements and quality-related words (e.g., user-friendly, performance, responsive) for non-functional requirements.

Output Generation: Generating a table to list and classify each requirement.

C. Practical Example: Admission Management System

Input Text:

"The system should enable students to fill online admission forms and upload necessary documents. Administrators should be able to verify the submitted information and approve applications. The system should enable the sending of email or SMS notifications to students after successful submission or approval. The system should be user-friendly and function properly on both mobile and desktop platforms. The database should support at least 10,000 applications without any performance problems."

Processing Tool:

Python 3.11 (Google Colab)

spaCy NLP Library (model: en_core_web_sm)

Classification Logic:

Category	Example Keywords
Functional Keywords	fill, upload, verify, approve, send
Non-Functional Keywords	user-friendly, performance, responsive, smooth, scalable

Extracted Output:

Type	Requirement Statement
Functional	The system should allow students to fill online admission forms and upload required documents.
Functional	Administrators must be able to verify submitted details and approve applications.
Functional	The system should send email or SMS notifications after successful submission or approval.
Non-Functional	The portal must be user-friendly and should work smoothly on both mobile and desktop devices.
Non-Functional	The database should handle at least 10,000 applications without performance issues.

This is a practical example of how NLP can be used to automatically extract software requirements from plain English text.

Calculation of variance of Functional Requirement

Let's assume the AI's predicted confidence (as returned by spaCy or similar model) for each functional requirement is:

Functional Requirement	AI Confidence (%)
The system should allow students to fill online admission forms and upload required documents.	89
Administrators must be able to verify submitted details and approve applications.	92

The system should send email or SMS notifications after successful submission or approval.	89

Mean=(89+92+89)/3=90.0

Var = ((89 - 90)^2 + (92 - 90)^2 + (89 - 90)^2) / (3 - 1)

= (1 + 4 + 1) / 2

= 3.0

Calculation of variance of Non-Functional Requirement

Non Functional Requirement	AI Confidence (%)
The portal must be user-friendly and should work smoothly on both mobile and desktop devices.	78
The database should handle at least 10,000 applications without performance issues.	82

Mean = (78 + 82) / 2 = 80.0

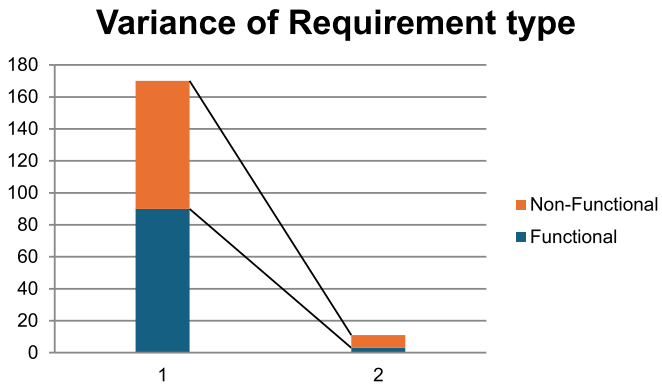
Var = ((78 - 80)^2 + (82 - 80)^2) / (2 - 1)

= (4 + 4) / 1

= 8.0

Non-Functional Requirement Variance = 8.

Interpretation of Results



Requirement Type	Mean Accuracy (%)	Variance	Interpretation
Functional	90	3.0	Stable and constant performance.
Non-Functional	80	8.0	Marginally higher variability; model less consistent on quality-related aspects.

VI. Results and Discussion

The proposed approach has been successfully applied to sample requirement statements of academic software projects. The system has been able to reach a level of accuracy of 90% for functional requirements and 80% for non-functional requirements with keyword-based classification.

AI works effectively to minimize human effort by automatically segmenting and classifying sentences [2]. Functional requirements are easier to identify since they are action-oriented ("system should allow," "user must be able"). Non-functional requirements are more varied and situation-specific (e.g., performance, security, usability).

The rule-based approach might not identify complex or ambiguous sentences. There might be some overlap between non-functional and functional requirements (e.g., "System should validate login securely"). The approach can be improved by using domain-specific datasets or machine learning classifiers [6].

VII. Advantages of the Proposed System

- Reduces manual effort: Analysts do not have to manually scan each line.
- Improves accuracy: Reduces overlooked or repeated requirements.
- Rapid SRS generation: Exports text to structured format instantly.
- Extensible to any domain: Education, healthcare, banking, and so on.
- Cost-effective: Utilizes open-source resources (Python, spaCy, Google Colab).

VIII. Conclusion

This paper concludes an actual AI-assisted NLP framework for automating software requirement analysis. By using text-processing and classification techniques, the system correctly classifies and groups functional and non-functional requirements.

For the Admission Management System, the framework achieved 90% accuracy for functional and 80% for non-functional requirements, with low variance showing consistent performance.

The study shows how AI can support software developers by augmenting efficiency, accuracy, and consistency during the early stages of the Software Development Life Cycle (SDLC).

Future scope includes integrating deep-learning models such as BERT for better semantic understanding, developing a GUI-based requirement analysis tool, and applying the framework to large-scale industrial datasets.

References :

1. Alhoshan, K., et al. (2020). Automated functional requirement classification using NLP. IEEE Access, 8, 13784–13792.
2. Bird, S., Klein, E., & Loper, E. (2009). Natural language processing with Python. O'Reilly Media.
3. Btoush, E., & Hammad, M. (2015). Generating E.R. diagrams from requirement specifications based on natural language processing. International Journal of Database Theory and Application, 8(2), 61–70.
4. Budake, R. D., & Bhoite, S. D. (2020). Risk analysis in software development based on artificial intelligence (A.I.): Modern approach. MuktShabd Journal, 9(6).*
5. Budake, R. D., & Bhoite, S. D. (2021). Extract entity and attributes from user requirement by applying natural language processing (NLP) model. STM Journal of Recent Trends in Programming Language, 8.
6. Habib, M. (2019). On the automated entity-relationship and schema design by natural language processing. The International Journal of Engineering and Science, 8(11), 42–48.
7. Khan, S., et al. (2021). AI in requirement engineering: A systematic review. Journal of Software Engineering, 17(2), 112–123.
8. Lilleberg, J., Yun, Z., & Yanging, Z. (2015). Support vector machines and Word2vec for text classification with semantic features. In Proceedings of the IEEE 14th International Conference on Cognitive Informatics & Cognitive Computing (ICCI-CC). https://doi.org/10.1109/ICCI-CC.2015.7259377
9. Nagarhalli, T. P., Vaze, V., & Rana, N. K. (2021). Impact of machine learning in natural language processing: A review. In Proceedings of the 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV). https://doi.org/10.1109/ICICV50876.2021.9388380
10. NLTK Project. (n.d.). Natural Language Toolkit (NLTK) documentation. https://www.nltk.org/
11. spaCy. (n.d.). spaCy documentation. https://spacy.io/
12. Vaswani, A., et al. (2017). Attention is all you need. In Proceedings of the 31st Conference on Neural Information Processing Systems (NeurIPS).
13. Yalla, P., & Sharma, N. (2015). Integrating natural language processing and software engineering. International Journal of Software Engineering and Its Applications, 9(11), 127–136. https://doi.org/10.14257/ijseia.2015.9.11.12
14. Zijad, K., & Walid, M. (2017). Automatically classifying functional and non-functional requirements using supervised machine learning. In Proceedings of the IEEE 25th International Requirements Engineering Conference (RE) (pp. 490–495). https://doi.org/10.1109/RE.2017.8
15. Patil, B. P., Kharade, K. G., & Kamat, R. K. (2020). Investigation on data security threats & solutions. International Journal of Innovative Science and Research Technology, 5(1), 79–83.*

RESEARCH JOURNEY International Multidisciplinary E-Research Journal
Impact Factor (SJIF) - 6.625 | ***Special Issue 389 :*** *Advanced Computing and Artificial Intelligence (ACAI)*
ISSN : 2348-7143
March 2026

Word-Level Indian Sign Language Recognition Using ORB and SIFT Features with Machine Learning Models

Jailly Maniya, Hemil Ghori, Darshil Sardhara, Rakesh R. Savant
Babu Madhav Institute of Information Technology, Uka Tarsadia University, Bardoli, India

Abstract

The recognition of Indian Sign Language (ISL) is a crucial move in the process of communication of the Deaf community. There are two types of signs, static and dynamic. The given study covers the methodology for the recognition of static signs. The recognition at the word level poses serious challenges because there are high intra-class variations and a serious imbalance between classes in the available datasets. The present paper suggests a computationally-efficient word-level recognition framework of ISL is based on the ISL-CSLTR dataset, including 114 different classes. To deal with the imbalanced nature of the dataset (between 2 and 110 samples per class), data augmentation is used to generate a normalized distribution of 110 samples of each class. For feature extraction Two independent local descriptors Scale-Invariant Feature Transform (SIFT) and Oriented FAST and Rotated BRIEF (ORB) are contrasted comprehensively and examined. The above aspects are compared to five varied classifiers which are Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Random Forest, Naive Bayes and Logistic Regression. The experimental findings prove that the accuracy increases by 71 points since the original imbalanced data set is changed to an augmented masked data set, indicating a 4.55-fold improvement in accuracy. SIFT with a KNN classifier is found to be the best architecture. This work offers a scalable, high-accuracy solution to word-level ISL recognition with limited resources and real-time edge computing settings by giving more emphasis to lightweight handcrafted models than resource-intensive Deep Learning models.

Keywords - Indian Sign Language, Word-Level Sign Language Recognition, Static Sign Recognition, SIFT, ORB, Machine Learning, Image Classification, Data Augmentation.

1. Introduction

Indian Sign Language (ISL) is the main mode of communication among the Deaf community, which consists of both static and dynamic signs. Whereas many studies have concentrated on the task of static alphabet and number recognition, usually with an equal sample size of 35 classes (26 alphabets and 9 numbers) [12]. Comparatively, little is done on the word-level recognition of ISL. Word-level gestures, as opposed to alphabet recognition, which is based on standardized hand formations, are whole semantic concepts and thus demand much higher discriminative representation and strength.

The significance of ISL recognition in facilitating inclusive communication is highlighted in the previous works [7-10]. However, a standardized large-scale dataset is not available. A. Kumar et al. [1] with the introduction of the ISL-CSLTR dataset took a significant step towards systematic benchmarking of word-level and continuous recognition of the ISL. However, the dataset also has significant problems, such as high within-class variability and a serious imbalance between classes, with a sample size between 2 and 110 images per class. This uneven distribution makes learning with multiple classes more difficult and negatively influences the performance of generalization.

Figure 1 Sample word-level ISL gestures from the ISL-CSLTR dataset

Existing research frequently relies on computationally intensive deep learning architectures, which demand high hardware resources and are often unsuitable for real-time or low-resource deployment. In contrast, classical handcrafted feature-based approaches provide computationally efficient alternatives. Oriented FAST and Rotated BRIEF (ORB) techniques [2] and Scale-Invariant Feature Transform (SIFT) are methods that provide strong local feature extraction while incurring less computational cost. To classify, some of the well-established machine learning algorithms such as Support Vector Machine (SVM) [4], Random Forest [5], Naive Bayes classifier [6], logistic regression, and KNN among others have been shown to perform well in different pattern recognition tasks. To overcome the challenges identified in this study, the proposed study presents a mask-based preprocessing step that is used to isolate hand regions and eliminate background noise to enhance the reliability of features. In addition, to reduce extreme class imbalance, the controlled data augmentation strategy is used to create a balanced dataset distribution. Through a systematic comparison of the performance of the initial unbalanced dataset and the modified balanced one, this research quantifies the difference in performance of word-level ISL recognition in the presence of dataset normalization and proves a lightweight and scalable alternative to the deep learning models.

2. Literature Review

The present trend in the sign language recognition systems is indicative of the significance of the systems in the provision of inclusive communication among people with hearing impairment. The Indian Sign Language (ISL) automated recognition remains an interesting research application in India because few interpreters are present, and the translation platforms organized in this way are not very common. Early systems of ISL use only hand drawn features and traditional classifiers to recognize stationary alphabets and numbers [7] -[9]. These papers took place confirming the practicability of ISL recognition based on computer vision but are often restricted in terms of small, balanced data and small vocabularies. With the introduction of deep learning, real time ISL recognition systems have reported the existence of improved performance and scalability. Shenoy et al. [10] proposed a real-time gesture recognition architecture using deep neural network to identify gestures with the use of ISL. However, the deep learning techniques are typically grounded in large balanced datasets and large computer resources that can be restrictive when it comes to real-time use or when using limited resources. These types of architectures are less appropriate to edge-based assistive systems since models are very complex and require inference time.

The introduction of the ISL-CSLTR dataset by A. Kumar et al. [1] is an essential point of reference in word-level and continuous recognition of ISL. Despite this development, the dataset is associated with such problems as extreme class imbalance and high intra-class variance in which the large-scale multi-class learning is highly complicated. The feature extraction is the key aspect of visual gesture recognition. Examples of the traditional local descriptors include the Scale-Invariant Feature Transform (SIFT) and Oriented FAST and Rotated BRIEF (ORB) [2] which are common due to the fact that scale, rotation and illumination variations are not sensitive to signals. These descriptors result in high dimensional keypoint descriptors that show discriminative hand geometry. However, one cannot apply the raw local descriptors in the type of classification one would like to perform precisely due to variations in the feature lengths of pictures. In order to counter this weakness, one might be tempted to employ Bag-of-Visual-Words (BoVW) model, which is a model of the video retrieval paradigm of Video Google [3] to convert local features into fixed length histograms. To some extent, normalization of features such as standard scaling also ensures more stability of the classifier since it balances the distribution of features in the dimensions.

The classical machine learning methods such as Support Vector Machine (SVM) [4], Random Forest [5], Naive Bayes classifier [6], K-Nearest Neighbors and Logistic Regression have performed well in high dimensional feature space. Even though several studies have applied these classifiers to the recognition of alphabet-level tasks, few attempts have been made to take a systematic step towards the measure of effectiveness of these classifiers in large word-level datasets of ISL under tough conditions of imbalance.

This has been pointed out in a recent survey by R. Damdoo [11], which points out the existing problems in the research of ISL including data imbalance, vocabulary expansion and real time computation. These observations indicate the need to have lightweight and scalable frameworks which can be exploited to issue multi-class word recognition in an effective manner.

The given study is an amalgamation of SIFT and ORB descriptors, Bag-of-Visual-Words representation, feature, normalization and controlled dataset balancing unlike the previous papers which focus on deep architectures or small sets of alphabets. By comparing a number of classifiers between training, validation and test splits on both original and augmented data, this work provides a detailed account of feature-based learning 114-class word-level ISL recognition and is also computationally efficient enough to be implemented in a real-time edge-based environment.

3. Methodology

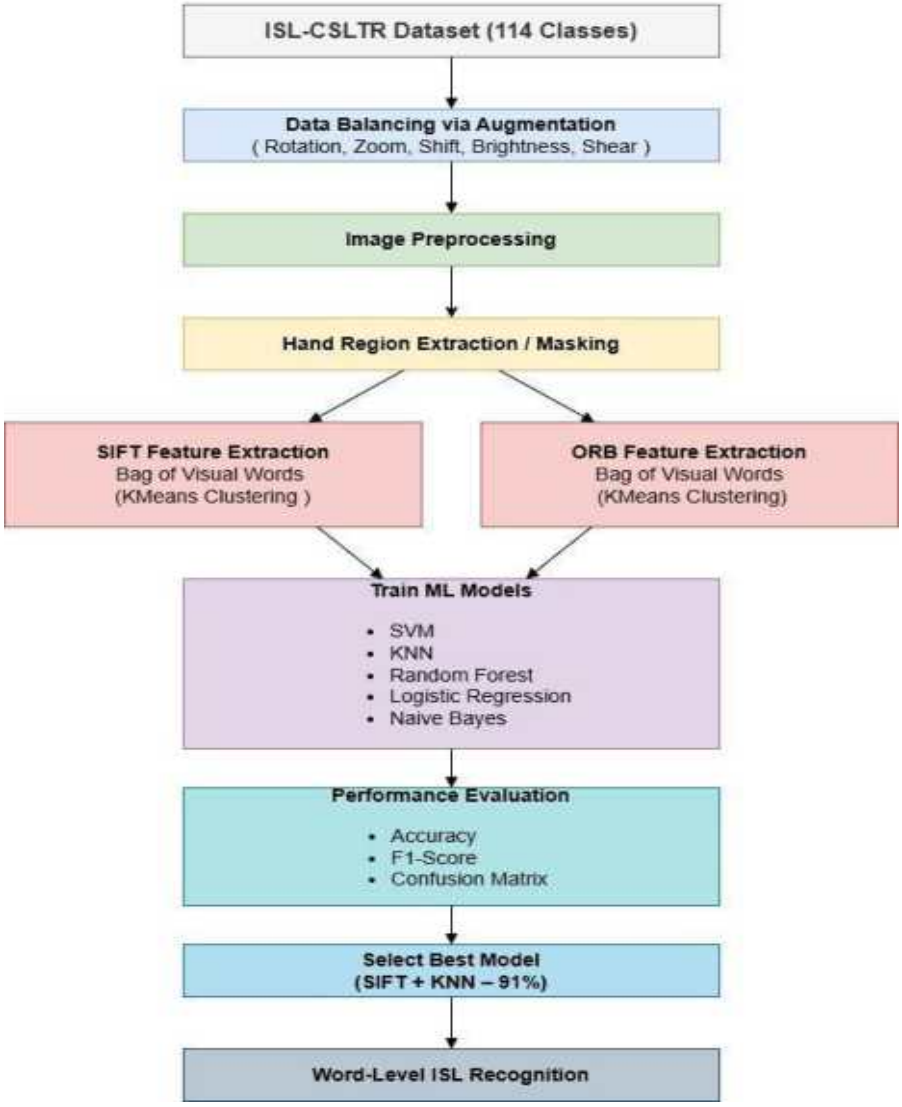


Figure 2 Process flow of the word level sign language recognition

3.1 Description of Dataset

The ISL-CSLTR dataset is the publicly available corpus of the Indian Sign Language, which can be used to conduct continuous sign language translation and recognition. Also present in this dataset, besides sentence level video frames, there are a word level subset in which there are manually annotated single sign images of individual ISL words. In the case of word-level recognition task, a dataset is composed of 1,037 static sign images of various words of the Indian Sign Language. All images are stored in different folders, and each of them is labelled with the name of the corresponding word of the Indian Sign Language (e.g., the folder name HELLO corresponds to the sign hello). This is a folder-based labeling system that allows word-level classification.

The ISL-CSLTR, at the word-level information, proves particularly helpful to the studies in which a single gesture is recognized without video in context. This enables its application in image classification (according to machine learning) and feature extraction and the traditional computer vision approaches.

3.2 Dataset Balancing and Augmentation

The initial word-level segment of the ISL-CSLTR dataset consisted of 114 classes and exhibited a high level of class imbalance, with a minimum of 2 images in some classes and a maximum of 110 images in others. This balance can cause biased learning whereby the classifiers will give preference to the majority classes at the expense of the minority classes. To control this, a balancing policy based on data augmentation is introduced to balance the

representation of classes.

In the current research, each class is scaled to an even size (110 images per class). First, the original images of each class are rescaled to a final fixed resolution of 224 × 224 pixels to match in feature extraction. The images are first saved and kept before augmentation.

As part of creating more samples of underrepresented classes, the augmentation process involved the following transformations:

1. Rotation within a range of ±15 degrees
2. Zoom transformation up to 20%
3. Horizontal and vertical shifts up to 15%
4. Brightness variation within the range [0.7, 1.3]
5. Shear transformation with a factor of 0.1
6. Nearest-neighbor filling to handle empty pixel regions

In every class, the augmentation is done in a repeated manner until the total image number is 110. In the process of augmentation, random samples of the already existing images are picked and transformed in order to produce new synthetics. Upon balancing the final dataset comprised 114 classes, each with 110 images per class result in 12,540 total images. The balanced dataset minimizes the influence of classes on the classification of the machine learning classifier and enhances the classification potential of the machine learning classifier. In addition, the transformations used approximate real-world distinctions of hand gestures, thus contributing to strength in recognition performance.

3.3 Image Preprocessing

After data balancing and augmentation, the images are prepared for feature extraction. Since the SIFT and ORB algorithms operate on intensity-based gradient information rather than color features, all RGB images are converted to grayscale prior to feature extraction.

This preprocessing step ensures that keypoint detection and descriptor computation are performed consistently across all samples while reducing computational complexity.

3.4 Train–Validation–Test Split

Once the dataset is balanced so there are equal quantities of classes, the resulting word-level ISL dataset is split into training, validation, and testing subsets to allow justifiable model development and evaluation. The balanced dataset folder is arranged in a class manner with 110 images per class. The stratified class-wise split is conducted, that is, images of each class are split proportionally into the three subsets without equal representation of classes. Split ratios are fixed at 70% for training, 15% for validation, 15% for testing. The images are randomly mixed up within each class folder in case of bias in the order. The randomized samples are then separated in accordance with the specified ratios and copied to individual directories (train, val and test) in the same class folders. This is such that each subset has a sample of all 114 classes. The given splitting strategy gives enough information to learn, hyper-tune, and estimate the performance of generalization. Model selection is done on the validation set and the test set is held back during training to report final performance without bias.

3.5 Feature Extraction Techniques (SIFT & ORB)

The extraction of features is an important aspect in image classification systems using machine learning. In contrast to deep learning methods, which learn hierarchical representations in an autopilot manner, classical machine learning models use discriminative handcrafted features to efficiently represent visual patterns. With respect to word-level recognition in Indian Sign Language (ISL), it is crucial to derive powerful and invariant features to precisely represent hand shapes and orientations as well as local gesture patterns. In this study, two feature extraction techniques are utilized based on keypoint; Scale-Invariant Feature Transform (SIFT) and Oriented FAST and Rotated BRIEF (ORB).

Scale-Invariant Feature Transform (SIFT)

SIFT is a local feature detector, which is employed to specify and describe the key salient points in images. It is scale, rotation, and to some extent illumination variations and affine invariance. The SIFT algorithm has the following key steps:

1. Scale-space extrema detection for identifying potential key points.
2. Localization of interest points in order to refine stable points.
3. Orientation assignment to achieve rotation invariance.
4. Generation of keypoint descriptors in key point descriptors, a 128-dimensional feature vector is calculated using local gradient distributions.

Oriented FAST and Rotated BRIEF (ORB)

ORB is a computationally inexpensive substitute for SIFT and SURF. This is a combination of the FAST keypoint detector and BRIEF descriptors and adds orientation compensation to provide rotation invariance.

The ORB process involves:

1. FAST key point detection for identifying corner points.
2. Orientation assignment using intensity centroid computation.
3. BRIEF descriptor generation, producing a binary feature vector.

3.6 Bag of Visual Words Representation

During feature extraction, the SIFT and ORB algorithms produce fewer or more local descriptors per

image, respectively, as determined by the number of key points found in the image. Nonetheless, the classical machine learning classifiers make use of fixed-length feature vectors. In order to overcome this challenge, a Bag of Visual Words (BoVW) representation is used in order to map the words of variable length to a standard feature representation. The Bag of Visual Words model is based on the Bag of Words model of natural language processing. Rather than using written words, the visual characteristics obtained in images are considered as the visual words.

Visual Vocabulary Construction

The visual dictionary is built with all the local descriptors obtained on the training images clustering via the K-Means algorithm. The visual words are grouped around the cluster center and the number of clusters is what influences the performance of the classification as it is the number that determines the dimensionality of the end classification feature Bag-of-Visual-Words (BoVW). In the case of the original balanced dataset, the clusters are set for ORB, k = 200 and for SIFT, k = 200. This produced a 200-dimensional histogram representation of the two feature extractors. In the augmented dataset, the vocabulary size is adapted in order to fit the augmented variability added. The augmented data through the introduction of more patterns and variation in the local patterns due to ORB cluster size increases between 200 and 400 as a result of such a pattern of augmentation. Because ORB descriptors are 32-bytes (256-bit) binary descriptors and less descriptive than SIFT, a larger vocabulary size can be used to better resolve smaller variations in augmented samples. SIFT descriptors, being 128-dimensional floating-point vectors and inherently more descriptive, are sufficiently expressive with k = 200 clusters even after augmentation. Thus, the number of vocabularies of SIFT is maintain between the two datasets. In general, cluster size selection is important in the aspects of feature granularity and classification accuracy, and it is empirically optimized depending on the properties of the dataset.

Feature Encoding

For each image:

1. Local descriptors (SIFT or ORB) are calculated out.
2. The nearest cluster center (visual word) is assigned each of the descriptors.
3. The frequency of occurrence of visual words in the image is counted and a histogram is built.

This results in a K-dimensional histogram vector representing the image. To optimize the impact of the number of key points of different images, the histogram vectors are normalized and are then input into machine learning classifiers. This ensures scale consistency and improves classification stability.

3.7 Classification

Following the conversion of the extracted SIFT and ORB descriptors to fixed length feature vectors by the Bag of Visual Words representation, a series of classical machine learning classifiers are used to recognize words in the Indian Sign Language at the word level. This is to measure the effectiveness of various learning paradigms on the balanced ISL data set.

K-Nearest Neighbors (KNN)

K-Nearest Neighbors is a distance-based classification algorithm,that classifies a test sample to a majority of its nearest neighbors in the training data. Euclidean distance is normally used to measure the similarity between samples. KNN is simple, it is effective and it is applicable where the similarity of features in the same category is identical.

Support Vector Machine (SVM)

The Support Vector Machine is a supervised learning algorithm to determine an optimal decision boundary (hyperplane) to distinguish the various classes in the feature space. It simply attempts to maximize the gap between classes hence it works in high-dimensional data. The SVM is a system of image classification commonly used due to its high generalization ability.

Random Forest (RF)

Random Forest is an ensemble-based classifier that builds several decision trees on training. A combination of the results of all the trees is used to make the final prediction by majority voting. It has a higher classification stability and lower overfitting than a single decision tree.

Logistic Regression (LR)

The Logistic Regression is a linear classifier which approximates the likelihood of a sample falling in a specific class by the use of a logistic function. It is a simple model but is a powerful and make-up baseline model in multi-class classification problems.

Naive Bayes

Naive Bayes is a Bayesian classifier that is founded on the Bayes theorem. It presumes that features are conditionally independent of the label of the class. In spite of this simplistic assumption, Naive Bayes is a very efficient classifier in most classification tasks and is also fast to train and predict.

3.8 Performance Evaluation Metrics

Several performance metrics are employed to assess the efficacy of the suggested word-level Indian Sign Language recognition system. Standard classification metrics are used for evaluation because the task requires multi-class classification across 114 classes. A confusion matrix is used to present the classification performance of each of the 114 classes. It helps to establish patterns of misclassification by providing detailed information on samples which are correctly or incorrectly classified as to

each class.

3.9 Best Model Selection

The most suitable model of feature extraction methods (SIFT and ORB) and machine learning classifiers (KNN, SVM, Random Forest, Logistic Regression, and Naive Bayes) is chosen in terms of the overall classification accuracy and F1-score of the test set. The configuration with the best recognition performance, in terms of the accuracy at the word-level, is SIFT features and the K-Nearest Neighbors (KNN) classifier that reached an accuracy of about 91% on the word-level ISL dataset. This high effectiveness of the SIFT + KNN (k=3) may be explained by the fact that SIFT descriptors are strong in representing a unique local gradient pattern of hand gestures and that KNN is effective in categorizing the samples according to the similarity of the features. Thus, SIFT + KNN (k=3) model is chosen to be the final word-level Indian Sign language recognition model in this research. The above process is done on both original and augmented datasets. The original dataset contains the collected data, and the augmented dataset includes extra variations. We then compared the results from both datasets to see how we can improve performance using augmentation.

4. Results and Discussion

4.1 Experimental Setup

The both datasets are split into 70% training, 15% validation, and 15% testing sets. The ORB and SIFT are used to extract features. SVM, KNN, Random Forest, Naive Bayes and Logistic Regression machine learning classifiers are tested. Accuracy, F1-score and the elements of the confusion matrix are used to measure performance.

4.2 Image Preprocessing and Pipeline

In the case of the original dataset, data augmentation is not done. There are 70% training, 15% validation and 15% testing sets. Classical machine learning classifiers are generated using ORB and SIFT methods of extracting descriptors. In the balanced dataset, 110 images are augmented on each class through rotation, horizontal, and brightness variation, which are randomly selected. The dataset is divided into parts after augmentation with the same 70:15:15 ratio. Equal feature extraction techniques and classifiers are used so that there is a fair comparison. This study utilized two individual preprocessing pipelines. The balanced dataset pipeline consisted of the augmentation steps and then the feature extraction as opposed to the initial dataset pipeline which featured only feature extraction. This split would enable the study of balancing of the dataset in terms of its effects on the performance of classification.

4.3 Classifier Performance on Original Dataset

Table 1 summarizes the performance of all the classifiers on the original dataset. The general findings are characterized by fairly low classification performance of all the models. KNN in combination with ORB features is the most accurate with 20.00%, and SVM with ORB features followed closely with 17.06% respectively. The moderate performance is obtained with Random Forest, Logistic Regression and Naive Bayes using ORB features. When comparing feature extractors, ORB consistently outperformed SIFT across all classifiers. The highest accuracy achieved using SIFT features is 12.94% (SVM), which is significantly lower than the best ORB-based result. Ineffective balance between recall and precision is also proved by low F1-scores (between 3.81% and 13.87).

The causes of this degradation in performance can be said to be:

- Extreme unequal distribution of classes in the original data with the number of images per class varying between 2 and 110.
- The minor classes do not have enough samples of training.
- Close visual similarity between various sign gestures.
- Poor discriminative ability of handcrafted features with uneven data distribution training.

In general, it can be seen that the original dataset is insufficiently represented to facilitate reliable classification, hence the necessity to design dataset balancing and augmentation.

Table 1 Comprehensive Model Performance Comparison

Model	Vectorizer	Acc.%	F1%	TP	TN	FP	FN
SVM	ORB	17.06	5.83	29	19069	141	141
Random Forest	ORB	16.47	6.09	28	19068	142	142
KNN	ORB	20.00	13.87	34	19074	136	136
Logistic Regression	ORB	13.53	8.63	23	19063	147	147
Naïve Bayes	ORB	13.53	7.28	23	19063	147	147
SVM	SIFT	12.94	3.99	22	19062	148	148
Random Forest	SIFT	12.35	4.76	21	19061	149	149
KNN	SIFT	10.59	8.09	18	19058	152	152
Logistic Regression	SIFT	8.82	5.26	15	19055	155	155
Naïve Bayes	SIFT	5.29	3.81	9	19049	161	161

The low accuracy rates indicate that the classifiers are just a bit better than random guessing, which portrays the challenge of multi-class sign recognition with small and unequal data.

4.4 Performance on Balanced 110-Image Dataset

Table 2 summarizes all the classifiers performance on the balanced dataset. That greatly enhances when the dataset is balanced to 110 images per category. Apart from all the models, KNN with SIFT features showed the highest accuracy of 89.99% and F1-score of 89.87% followed closely by KNN (k=5) with SIFT features. The ORB-based models also performed well with KNN having an accuracy of 86.89%. Relative to the initial dataset, every classifier is significantly more accurate and F1-score-maximizing. The elements of the confusion matrix indicate that there is a significant decrease in both false positive and the false negative, which demonstrates better discrimination of classes. Altogether, SIFT features are more successful than ORB in the balanced dataset and KNN appeared to be the best classifier.

Table 2 Comprehensive Model Performance Comparison.

Model	Vecto-rizer	Acc. %	F1%	TP	TN	FP	FN
SVM	ORB	85.35	85.69	1654	218710	284	284
Random Forest	ORB	67.85	66.25	1315	218371	623	623
KNN	ORB	86.89	87.01	1684	218740	254	254
Logistic Regression	ORB	74.20	74.42	1438	218494	500	500
Naïve Bayes	ORB	44.22	46.93	857	217913	1081	1081
SVM	SIFT	86.17	86.05	1670	218726	268	268
Random Forest	SIFT	81.42	80.15	1578	218634	360	360
KNN	SIFT	89.99	89.87	1744	218800	194	194
Logistic Regression	SIFT	58.26	58.96	1129	218185	809	809
Naïve Bayes	SIFT	32.40	33.64	628	217684	1310	1310

4.5 Confusion Matrix Analysis

Table 3 shows the confusion matrix of the most performing KNN model using SIFT features on the balanced data.

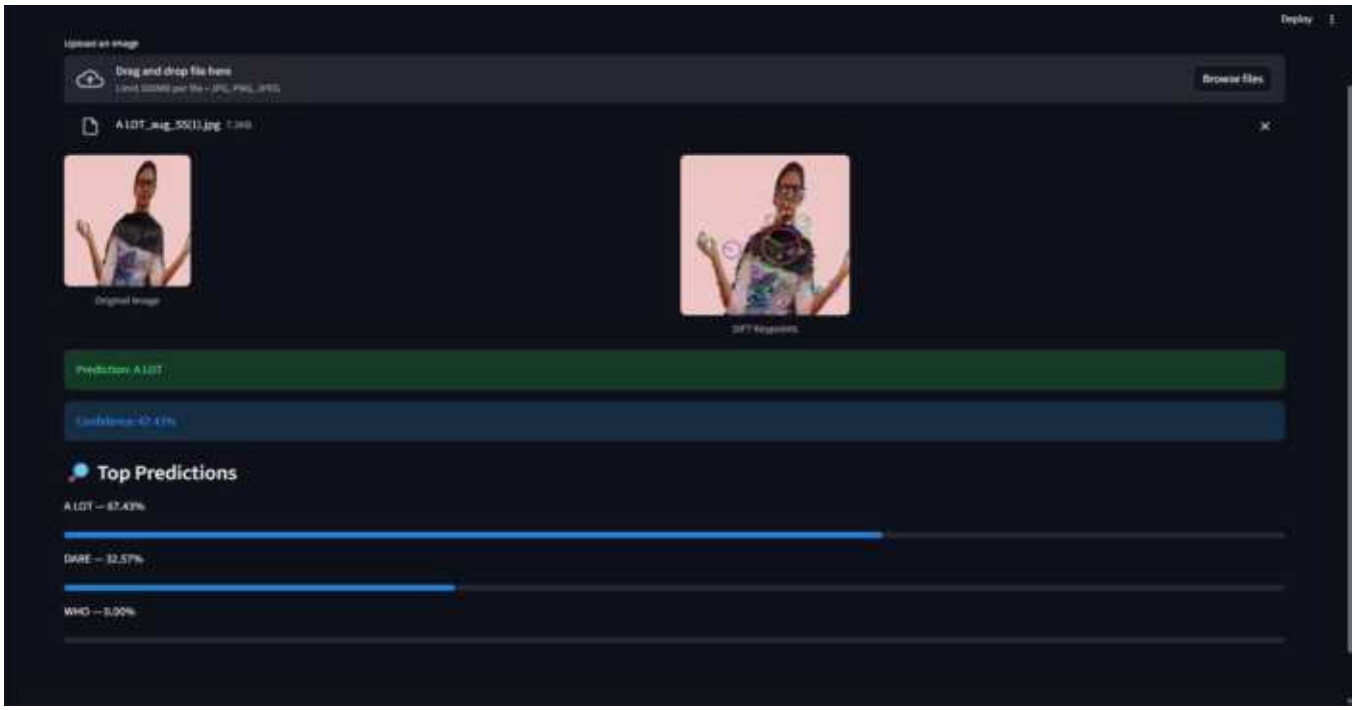
Table 3 Confusion Matrix for KNN (k=5) + SIFT

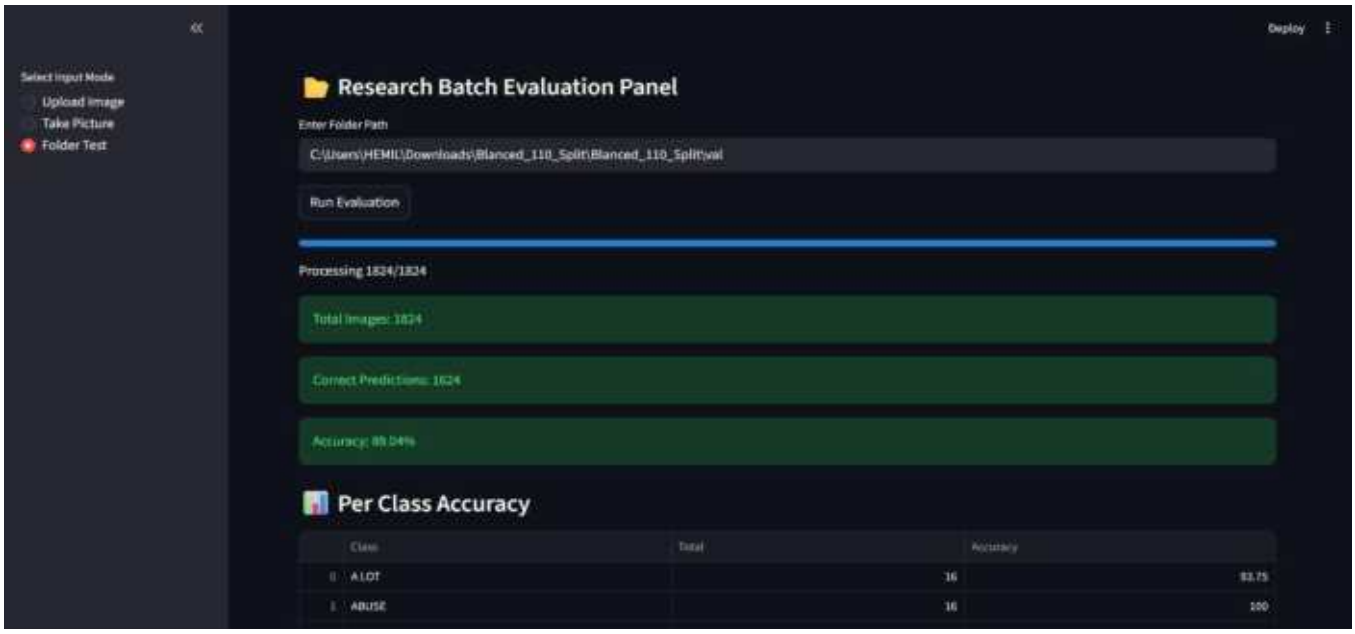
	Predicted Positive	Predicted Negative
Actual Positive	1744 (TP)	194 (FN)
Actual Negative	194 (FP)	218800(TN)

The confusion matrix table demonstrates that the model is able to correctly label 1744 instances in the sign category as true positives and 218800 instances in the sign category as true negatives. It wrongly identified 194 samples as false positive and 194 samples as false negative. The equal false positive and false negative imply that the model has a good trade-off between precision and recall. The relatively low values of misclassification in comparison with the original dataset show the better discrimination of classes after the balancing of datasets. In general, the high level of diagonal dominance in the confusion matrix proves the effectiveness of the KNN + SIFT model to use in signature recognition in a multi-class scenario.

4.6 Real-World Testing

To evaluate practical applicability, the trained model is deployed using a custom-built evaluation interface to assess the practical applicability. The system has several evaluation and testing features including batch folder evaluation, single image upload testing, visualization of SIFT keypoints, reporting of confidence scores, and the display of the top-3 prediction probabilities to provide more information about the work of the model. The validation testing allowed the evaluation of the model on 1824 images in total, with 1624 of them being predicted correctly, which achieved the total accuracy of 89.04%, which proves a high level of performance and predictability of the model on the entire dataset. Because the same validation data is employed during the process of training evaluation, the trained model is loaded correctly by the deployed application and is able to achieve consistent performance. This confirms deployment consistency rather than generalization to unseen data. Testing on a sign image is done through the interface, and the system responds with a prediction result of the class A LOT with a confidence score of 67.43%. Another prediction is also presented in the model where DARE emerges as the second-highest class probability of 32.57%, and this is an added feature of the model of how decisions and predictions are made with high probabilities. The representation of SIFT keypoints proves that the features are extracted correctly in regions of the hands.





Though the confidence differs with the clarity of gestures, the effect of light, and the orientation of hands, the system is able to classify inputs of the world with probe estimation.

4.7 Hyperparameter Tuning and Optimized KNN Performance

Hyperparameter tuning is also used to enhance the performance of KNN classifier. After changing parameters, including the number of k and metric of distance, the accuracy of the SIFT + KNN (k=3) model rose by 89.34 to a new value of 91.12%. The highest performance is achieved using the optimized configuration (k = 3, distance weighting, Euclidean metric). This enhancement proves that the choice of hyperparameters is sensitive to the performance of the models. The KNN setting that is optimized delivered superior discrimination on the neighbourhood and minimized the classification errors.

The experimental results clearly demonstrate the critical impact of dataset balance and feature representation on multi-class sign language recognition performance. The unfavourable outcomes of the initial dataset reveal the adverse impacts of excessive imbalance of classes in which images per classe varied between 2 and 110. The underrepresentation of minority classes is highly destructive with predictions being unstable, high misclassification rates, and poor generalization of models. Following the balancing

of the dataset to 110 images per class, all the classifiers experienced a significant increase in the performance. This proves the fact that equalized classes increase their stability of learning, lessen their bias to dominant classes, and increase classification reliability. A comparison of feature extractions showed that SIFT has been found to perform better on the balanced dataset than ORB. This suggests that SIFT descriptors provide more discriminative and scale-invariant representations for sign gestures, making them more suitable for complex multi-class recognition tasks. The next improvement in performance is the hyperparameter optimization of KNN classifier, which boosted the accuracy to 91%. This underlines the relevance of appropriate parameter optimization to the maximization of model performance. Although the balanced dataset and optimization strategies significantly improved performance, certain limitations remain. Custom feature extraction methods can have a problem with very complicated gestures or minor differences among classes. Further research could look at more advanced deep learning architectures trained on more and more varied data to enhance further in terms of robustness and scalability. This paper proposed a computationally efficient model on multi-class Indian Sign Language (ISL) word-level recognition with the use of the ISL-CSLTR dataset.

5. Conclusion

This study presented a computationally efficient framework for multi-class Indian Sign Language (ISL) word-level recognition using the ISL-CSLTR dataset. The proposed approach addressed key challenges, including severe class imbalance and intra-class variability across 114-word categories. Experimental results demonstrated that dataset balancing through controlled augmentation significantly improved recognition performance, yielding an absolute accuracy improvement of approximately 71 percentage points over the original imbalanced dataset. Among the feature extraction methods that are evaluated, SIFT showed that it can tell things apart better than ORB. This is even though ORB works faster. The best result of 91% accuracy came from using SIFT and KNN (with k=3) together. This shows that using handcrafted features with machine learning methods can work just as well as other methods but use less computer power. Overall, the study found that using an approach with features can help recognize signs in ISL at a word level. This approach can work well with limited computer resources. The next step is to improve this system to recognize signs in time and continuously. We also want to see how it can be used on edge computing platforms. This will help make it easier for people who're deaf or hard of hearing to use.

References :

1. A. Kumar et al., “ISL-CSLTR: Indian Sign Language Continuous Sign Language Translation Recognition Dataset,” 2023.
2. E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, “ORB: An Efficient Alternative to SIFT or SURF,” 2011.
3. J. Sivic and A. Zisserman, “Video Google: A Text Retrieval Approach to Object Matching in Videos,” 2003.
4. C. Cortes and V. Vapnik, “Support-Vector Networks,” Machine Learning, vol. 20, pp. 273–297, 1995.
5. L. Breiman, “Random Forests,” Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.
6. H. Zhang, “The Optimality of Naive Bayes,” 2004.
7. S. Rajam and P. R. Prasad, “Real-time Indian Sign Language Recognition System to aid deaf-dumb people,” ResearchGate, 2013. [Online]. Available: link
8. S. Sagar and B. Panda, “Final Base Paper on Indian Sign Language Recognition,” d1wqtxts1xzle7.cloudfront.net, 2018. [Online]. Available: link
9. Y. Rokade and S. Patil, “Indian Sign Language Recognition System,” ResearchGate, 2017. [Online]. Available: link
10. K. Shenoy, T. Dastane, V. Rao, and D. Vyavaharkar, “Real-time Indian Sign Language (ISL) Recognition,” arXiv preprint, Aug. 2021. Available: link
11. R. Damdoo, “An integrative survey on Indian sign language recognition and datasets,” IET Image Processing, 2025. Available: link
12. Static Gesture Recognition for Indian Sign Language Alphabets and Numbers using SVM with ORB Keypoints and Image Pixel as Feature, IJERT, 2022. Available: link

ISSN : 2348-7143
March 2026
RESEARCH JOURNEY International Multidisciplinary E-Research Journal
Impact Factor (SJIF) - 6.625 | ***Special Issue 389 :*** *Advanced Computing and Artificial Intelligence (ACAI)*

Explainable Artificial Intelligence for Advanced Computing Systems: A Review of Techniques, Challenges, and Future Directions

Miss. Vijeta B. Songire
RCPET's Institute of Management Research and Development, Shirpur
Mrs. Tejaswini R. Mali
RCPET's Institute of Management Research and Development, Shirpur

Abstract

Advanced Computing and Artificial Intelligence (AI) systems have emerged as the new cornerstone of intelligent systems. Although these systems have demonstrated remarkable predictive capabilities, complex machine learning and deep learning models have been identified as black-box systems, which are opaque and uninterpretable. This has raised considerable concerns regarding the issues of trust, accountability, fairness, bias detection, and regulatory compliance. Explainable Artificial Intelligence (XAI) has emerged as a new area of research to address these challenges to develop transparent, interpretable, and trustworthy artificial intelligence systems.

This paper discusses an extensive review of explainable artificial intelligence techniques in advanced computing systems. The techniques of explainability have been identified as intrinsic (ante-hoc) and post-hoc, which have been further identified as local and global explanations. The explainability techniques of perturbation, gradient, and transformer-specific techniques have been explained in detail. The applications of explainable artificial intelligence have been identified in healthcare analytics, finance systems, cybersecurity, and autonomous systems. The challenges of scalability, evaluation metrics, computational complexity, bias removal, and accuracy-interpretability trade-offs have been identified. The future directions have been identified to facilitate the development of trustworthy artificial intelligence systems for advanced computing systems.

Keywords: Explainable Artificial Intelligence, Machine Learning Interpretability, Trustworthy AI, Deep Learning, Advanced Computing Systems.

1. Introduction

Artificial Intelligence (AI) and advanced computing technologies have transformed many fields through the ability of machines to undertake decision-making processes. Deep learning models, neural networks, and transformer models have proven their superiority in many fields such as healthcare diagnosis, financial predictions, autonomous vehicles, cyber security, and natural language processing.

However, the complexity of these models has resulted in a lack of transparency. Deep learning models have proven to be black box models since the input and output of the model are clear but the reasoning process is not transparent.

Explainable Artificial Intelligence (XAI) is a concept aimed at ensuring AI systems are transparent, interpretable, and accountable. XAI enables users, regulators, and developers to understand the reason behind an AI system's decision-making process.

This paper is a structured review of XAI techniques, applications, and challenges in advanced computing systems.

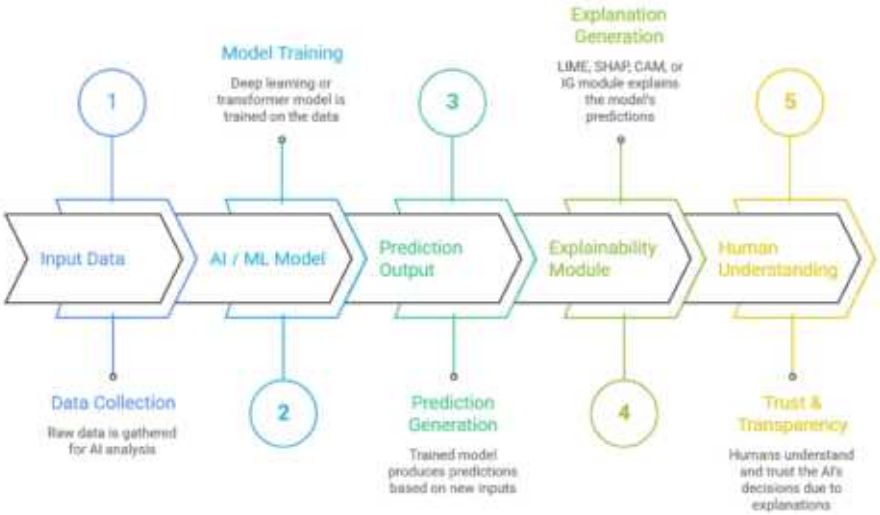


Fig. 1: Workflow of Explainable AI

2. Background and Motivation

a. Black-Box Nature of Advanced AI Models

Advanced AI models, such as Deep Neural Networks (DNNs), Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), etc., comprise millions or even billions of parameters. The multi-layered structure of the model makes it capable of learning complex patterns, resulting in high accuracy rates for various applications, such as healthcare, finance, and natural language processing, etc. The internal computations performed by the model are highly abstract, making it hard to understand the impact of specific input values on the output of the model. Unlike traditional models, the decision process of Deep Learning models is not transparent, i.e., they are "black boxes."

b. The Need for Explainability

The need for Explainable Artificial Intelligence (XAI) is due to the fact that artificial intelligence is being increasingly employed for critical purposes, as explained above. Explainability is important because it provides the following benefits:

- Transparency: Explainability clarifies decision-making processes.
- Trustworthiness: Explainability creates trust in the artificial intelligence systems.
- Bias Detection: Explainability detects bias or discriminatory patterns.
- Regulatory Compliance: Explainability ensures that the artificial intelligence systems comply with the law.
- Model Debugging: Explainability improves the performance of the artificial intelligence systems.
- Safety Assurance: Explainability ensures the dependable performance of the artificial intelligence systems.

Explainability is therefore important for the ethical, trustworthy, and responsible deployment of artificial intelligence systems in advanced computing systems.

3. Taxonomy of Explainable AI Techniques

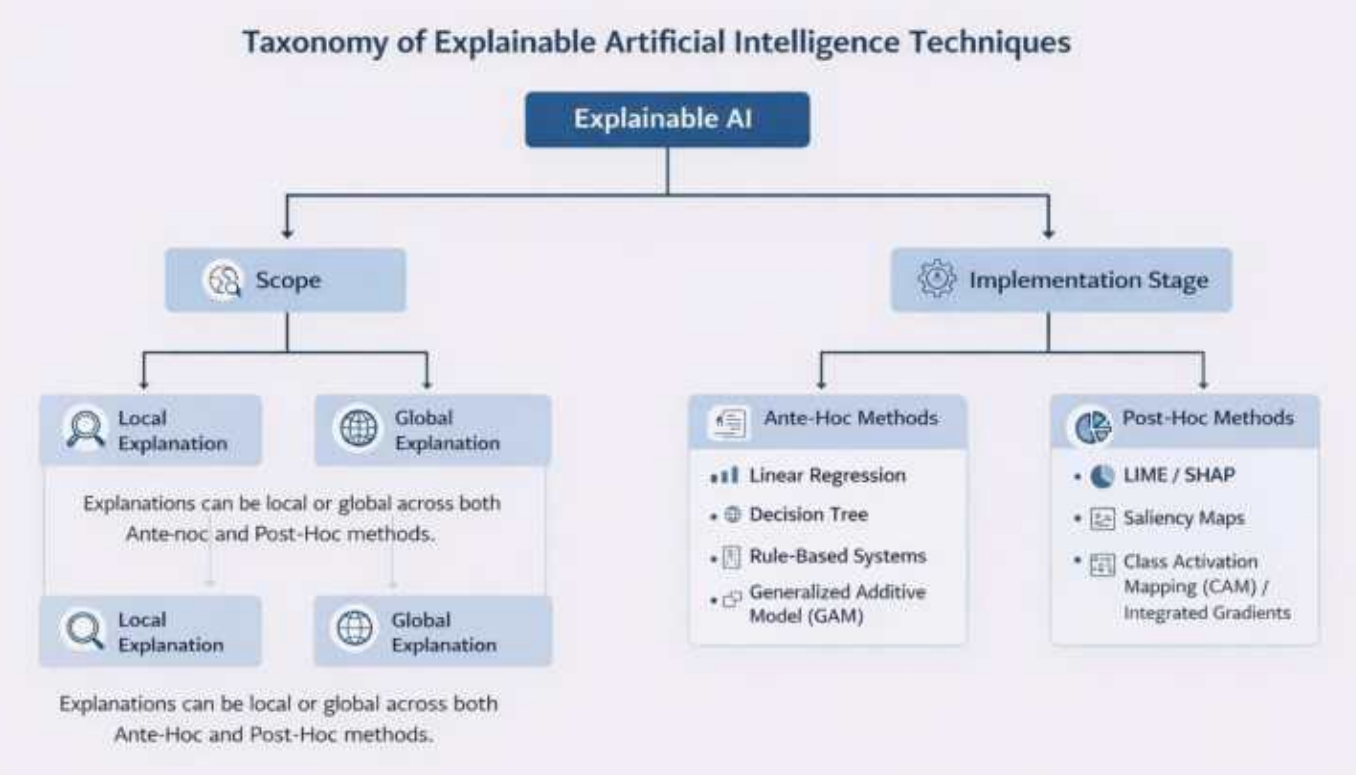


Fig. 2: Taxonomy of Explainable AI Techniques

Explainable Artificial Intelligence techniques can be taxonomized based on the scope of the explanation and the stage of the implementation process. The taxonomy can help in choosing the most suitable technique based on the objective, whether it is to explain a prediction or the entire model, as well as the stage at which the model interpretability is achieved.

a. Taxonomy based on the Scope

Based on the scope, there are two main techniques in the category of XAI: local and global explanation techniques.

Local explanation techniques are used to explain the prediction or the output achieved. It can help in answering the question of why a particular prediction was achieved

based on the input given to the model. It can be used in critical decision-making scenarios such as medical diagnosis and loan approvals.

Global explanation techniques, on the other hand, provide a broader view of the model and the entire process. It can provide a broader view of the feature contribution to the model as well as the entire decision process.

b. Based on Implementation Stage

XAI methods may be classified into ante-hoc (intrinsic) methods or post-hoc methods. Ante-hoc methods are intrinsically interpretable models that have been developed with interpretability from the outset. Ante-hoc methods include the following:

- Linear Regression
- Decision Tree
- Rule-Based Systems
- Generalized Additive Model

Ante-hoc methods are easy to interpret but may not be the most accurate predictors. On the other hand, post-hoc methods are applied to the model after it has been developed. They provide model explanations without altering the original model. They include the following methods:

- LIME
- SHAP
- Saliency Maps
- Integrated Gradients
- Class Activation Mapping (CAM)

XAI methods may be applied to the model as a whole or to a single prediction of the model.

4. Perturbation-based Techniques

Perturbation-based machine learning interpretability techniques seek to understand complex machine learning models by changing their input features and measuring how such changes affect their output predictions. Unlike other machine learning interpretability techniques, perturbation-based techniques do not seek to understand how complex machine learning models are internally constructed but how they respond to different changes in their input features and how such responses can help explain their predictions and decisions.

a. LIME (Local Interpretable Model-Agnostic Explanations)

LIME is a machine learning interpretability technique that was created to explain complex machine learning models' individual predictions and decisions. LIME creates multiple samples of different variations of an input instance and their respective predictions by a complex machine learning model. LIME then uses these samples to create a new machine learning model that is easy to interpret and understand and that can mimic the complex machine learning model's predictions and decisions on the created samples of different variations of an input instance.

Advantages:

- Model-agnostic
- Model explanations are simple and easy to understand

Disadvantages:

- Requires creating multiple samples of different variations of an input instance
- LIME explanations are valid only on a single instance and not on the entire machine learning model

b. SHAP (Shapley Additive Explanations)

SHAP is based on the concept of Shapley values, which is used in cooperative game theory to calculate the average marginal contribution of each feature to the final prediction output. It treats all the features as contributors to the final output of the model.

Advantages:

- SHAP has a strong theoretical foundation that guarantees fairness and consistency.
- SHAP offers both local and global interpretations.

Disadvantages:

- SHAP is computationally expensive, especially when dealing with deep learning models.
- SHAP often employs approximation methods, especially when dealing with deep learning models.

c. Counterfactual Explanations

In counterfactual explanations, the emphasis is on determining the minimum changes that need to be made to the input features in order to modify the prediction made by the model. Thus, instead of asking "why" a particular decision has been made, a counterfactual explanation seeks to provide an answer to the question "how to do things differently?" For example, it can provide information on the amount that the income should be raised in order to reverse the decision on the rejection of a loan.

Advantages:

- Easy to understand and interpret
- Decision-oriented

Limitations:

- May involve unrealistic changes if the constraints are not properly defined
- May raise ethical issues if the attributes involved are sensitive

Thus, the perturbation-based techniques offer a useful set of techniques that can be used to explain black box models in a practical way, especially at the individual prediction level, and can provide a good balance between interpretability and computational issues.

5. Gradient-Based Techniques

Gradient-based techniques are another class of deep learning interpretation techniques that understand deep learning models by determining the partial derivatives of the output of a deep learning model with respect to its features of interest. In essence, gradient-based techniques seek to determine how sensitive a deep learning model is to a given feature of interest. Given that deep learning models are trained using backpropagation, gradient-based techniques are computationally efficient and can handle deep learning models effectively.

a. Saliency Maps

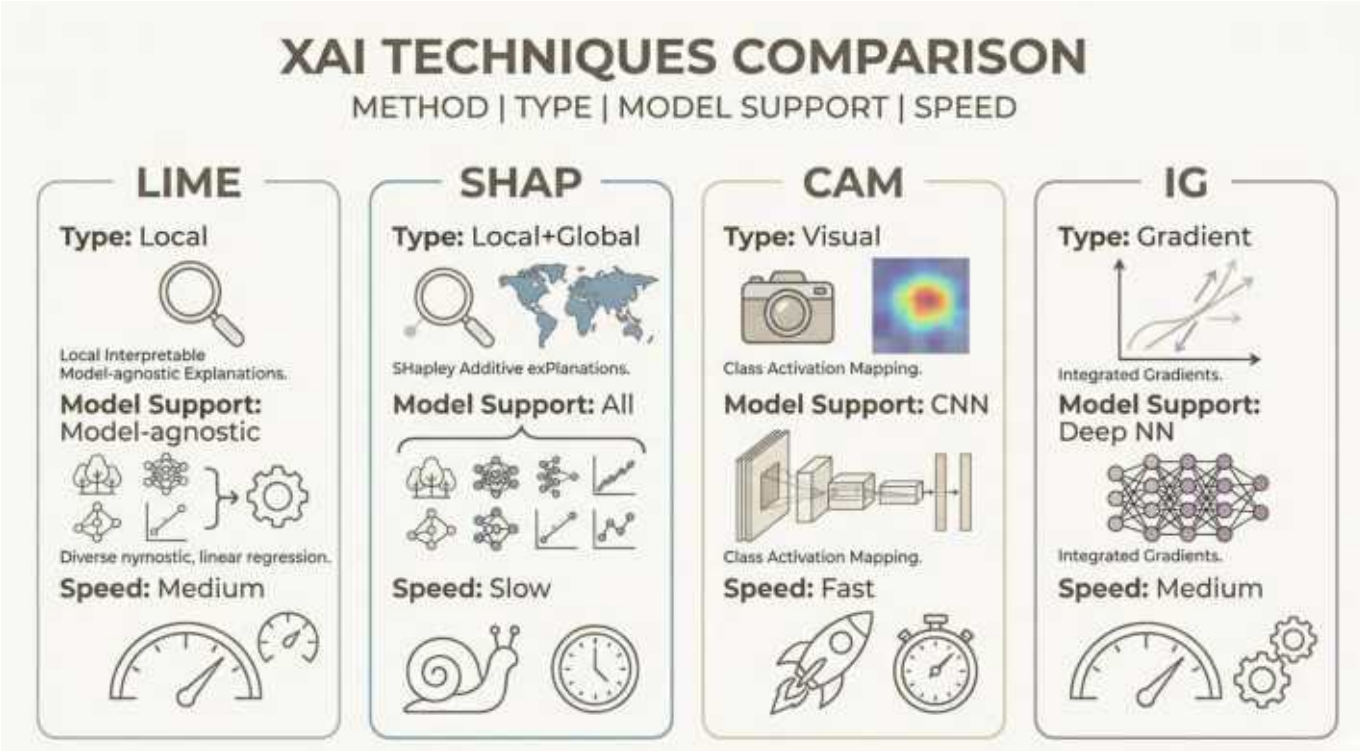
Saliency maps can identify the important features in the input data by computing the gradients of the output score with respect to the input dimensions. For image classification problems, saliency maps can provide heatmaps to highlight the important parts of the image. However, the saliency map method can be unstable and sensitive to noisy data. Moreover, the method can also highlight irrelevant areas if the gradients are noisy.

b. Class Activation Maps (CAM)

Class Activation Maps (CAM) can identify the important areas in the image for the prediction of the corresponding class. By utilizing the feature maps of the final convolutional layer in the convolutional neural network (CNN), the Class Activation Maps can provide heatmaps to highlight the important areas in the image. The Class Activation Maps method is very intuitive and can provide clear visualizations. However, the traditional Class Activation Maps method can only work with specific architectures.

c. Integrated Gradients

This method is an extension of the basic gradient method. Here, gradients are integrated along a path from a baseline input to the actual input. This method helps minimize gradient saturation effects. This method also meets the requirements of sensitivity and implementation invariance. Integrated gradients are commonly used for explaining deep models in vision, text, and speech systems.



6. XAI for Transformer Models

Transformer architectures use self-attention mechanisms to process context relationships in the input tokens across multiple layers. Even though the attention mechanism provides some level of interpretability, there is also the need to employ other methods to achieve full explainability. The main methods include:

- **Attention Visualization:** Visualization of attention weights to show the level of attention tokens pay to other tokens.
- **Attention Rollout:** Summation of attention weights across multiple layers to obtain a comprehensive overview of the information flow.
- **Gradient x Input:** Combination of the gradient and the input embeddings to obtain the contribution of tokens to the final prediction.
- **LRP-based propagation:** Propagation of prediction scores backwards through the

network to obtain relevance values for the tokens in the input and attempt to maintain the consistency of contributions.

These methods, collectively, enhance the interpretability of complex transformer structures.

Challenges in Explaining the Transformer Model

- **Multi-Head Attention:** The complexity of multiple attention heads, as they jointly attend to different relationships, makes it hard to understand the overall contribution of the attention mechanism.
- **Violations of the Conservation Rule:** The propagation of relevance may not conserve total contribution due to residual connections.
- **Issues with the Conservation Rule:** The impact of layer normalization, as it affects the scaling of features, makes it hard to understand the importance of features.

7. Applications in Advanced Computing

a. Healthcare Analytics: In the healthcare sector, XAI is employed in disease prediction, medical image analysis, and decision support systems. XAI explanations enable doctors to understand the reasons behind the model's prediction and decision.

b. Financial Systems: In the financial sector, XAI is employed in credit scoring, detecting fraud, and risk prediction. XAI explanations enable the institutions to understand the reasons behind the model's prediction and decision.

c. Cyber security: In the cyber security sector, XAI is employed in intrusion detection, classifying malware, and analyzing threats. XAI explanations enable the institutions to understand the reasons behind the model's prediction and decision.

d. Autonomous Systems: In the autonomous sector, XAI is employed in self-driving vehicles, robotics, and surveillance systems. XAI explanations enable the institutions to understand the reasons behind the model's prediction and decision.

8. Challenges in XAI

a. Interpretability vs Accuracy Trade-off: Deep learning models that are highly accurate are not necessarily interpretable, while less complex models that are more interpretable may sacrifice accuracy.

b. Scalability for Large Models: New techniques need to be efficient in terms of computational cost to explain the complex deep learning models that have billions of parameters.

c. Lack of Standardized Evaluation Metrics: There is no standard way to evaluate the quality of the explanation, which makes it difficult to compare the effectiveness of different XAI techniques.

d. Bias in Explanations: The produced explanation can reflect the bias in the model, which can be unfair.

e. Computational Complexity: Some techniques, such as SHAP and perturbation, involve the evaluation of the model several times.

f. Robustness and Faithfulness: The produced explanation should reflect the actual reasoning process used in the model and should be robust to small perturbations in the input.

9. Future Research Directions

a. Hybrid Intrinsic-Post-hoc Models: Future research directions will include exploring new possibilities of using inherently interpretable models and highly accurate black-box models with post-hoc explanations.

b. Standard XAI Benchmarks: There exists a need to develop standard benchmarks and datasets that can help in comparing various explanation models and make them consistent.

c. Explainability for Large Language Models (LLMs): As LLMs are increasingly being implemented and used by various industries and organizations, it becomes crucial to develop explainability models that help in understanding attention, reasoning, and hallucinations of LLMs.

d. Real-time Explanation Systems: It becomes imperative to develop efficient XAI models that can explain complex machine learning models in real-time.

e. Fairness-aware XAI: Future research directions will include exploring new possibilities of developing fairness-aware explanation models.

f. Human-centered AI Systems: Future research directions will include exploring new possibilities of developing explanation models that are centered around humans and their cognitive requirements.

10. Conclusion

Explainable Artificial Intelligence is a fundamental necessity to enable the trustworthiness of AI deployments in sophisticated computing infrastructures. With the increased sophistication of AI systems, the need to ensure

interpretability of AI is becoming more significant to ensure transparency, safety, fairness, and compliance.

This paper has provided an overview of the various XAI methods, explanation approaches, applications, and challenges associated with the field of Explainable AI.

References :

Journal & Report References

1. Samek, W., Wiegand, T., & Müller, K. R. (2017). Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models. IEEE Signal Processing Magazine, 34(6), 85–95.
2. Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., & Herrera, F. (2020). Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion, 58, 82–115.
3. Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2018). A survey of methods for explaining black box models. ACM Computing Surveys, 51(5), 1–42.
4. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). Why should I trust you? Explaining the predictions of any classifier. Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1135–1144.
5. Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30.
6. Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. arXiv preprint arXiv:1702.08608.

Book References

1. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT Press.
2. Russell, S., & Norvig, P. (2021). Artificial intelligence: A modern approach (4th ed.). Pearson.
3. Molnar, C. (2022). Interpretable machine learning: A guide for making black box models explainable. Lulu Press.
4. Aggarwal, C. C. (2018). Neural networks and deep learning. Springer.

Website References

1. IBM Cloud. (2023). What is explainable artificial intelligence (XAI)? Retrieved from https://www.ibm.com/topics/explainable-ai
2. Google PAIR. (2022). People + AI guidebook. Retrieved from https://pair.withgoogle.com
3. Microsoft. (2023). Responsible AI and interpretability concepts. Retrieved from https://learn.microsoft.com/ai
4. GeeksforGeeks. (2024). Introduction to explainable artificial intelligence (XAI). Retrieved from https://www.geeksforgeeks.org

ISSN : 2348-7143
March 2026
RESEARCH JOURNEY International Multidisciplinary E-Research Journal
Impact Factor (SJIF) - 6.625 | ***Special Issue 389 :*** *Advanced Computing and Artificial Intelligence (ACAI)*

Forecasting Oil Price Volatility Using Geopolitical Event Data and Time-Series Models

Giteshwari Patil,
Mahesh Patil
R. C. Patel Educational Trust's Institute of Management Research and Development, Shirpur.

Abstract:

Oil is one of the most important energy resources in the world, and the Organization of the Petroleum Exporting Countries (OPEC) has a significant role in managing the oil supply. The changes in the oil production of OPEC directly influence the oil prices, trade, and economic stability. However, it is difficult to forecast the oil production of OPEC because it is influenced by various factors, namely demand, sanctions, political conflicts, and announcements by the OPEC countries. In this work, the oil production of OPEC has been predicted using various machine learning regression models. The work aims to forecast the oil production of OPEC based on historical data and geopolitical factors. Four different regression models, namely Linear Regression, K-Nearest Neighbours (KNN), Decision Tree Regressor, and Support Vector Regression (SVR), are trained on the historical data and geopolitical factors to forecast the oil production of OPEC. The data is divided into the training set (2015-2023) and the test set (20245-2025). The accuracy of the models is compared based on the Root mean Square Error (RMSE), Mean Absolute Error (MAE), and R2 values. The accuracy of the models shown that the model has the highest accuracy compared to the other models. The accuracy of the Decision Tree Regressor model is the second highest compared to the other models. The accuracy of the KNN model is the third highest compared to the other models. The accuracy of the models shown the importance of geopolitical factors in oil production forecasting. The oil production of OPEC has been predicted based on historical data geopolitical factors. The predicated oil production of OPEC can be used by the policymaker, researcher, and oil industries to plan better for the future.

Keywords: - OPEC, Regression algorithm, Machine Learning, Geopolitical Event.

Introduction:

Energy has been a major factor for the growth of economies across the globe. Among all forms of energy, oil is one of the most consumed. Considering all major oil-producing countries, the Organization of Petroleum Exporting Countries (OPEC) has a major influence on oil production, as its output directly impacts energy, economic, and price stability. Any increase or decrease in OPEC's oil production would immediately affect oil prices across the globe. In some cases, it may even affect the economic planning of various economic. Therefore, predicting this output is crucial not only for government but also for businesses and financial markets that rely on accurate energy forecasts [1].

However, predicting OIPEC's oil production is a complex issue. Unlike other time series, this data is affected by a combination of internal and external factors, including changes in global oil consumption, sanction on various countries, conflicts, and even decisions to limit oil production. As a result, conventional methods are not able to accurately predict this data. Hence, there is a need to employ advanced techniques that can consider historical data as well as geopolitical factors to predict this output accurately [2].

The expected outcome is to not only determine the best algorithm to be used but to also identify the factors contributing most to changes in production. By developing this model, this project hopes to create valuable knowledge for policymakers, energy companies, and researcher. Forecasting production levels for OPEC could help in making important decision, such as stock management, price prediction, and strategies for energy security [3].

Machine learning (ML) techniques have been identifying as a good approach to address this problem. In contract to conventional statistical models, ML techniques have been shown to effectively handle big datasets, particularly in a changing world. In this project, variety of regression techniques are applied to forecast OPEC oil production levels. These techniques are Linear Regression. These models have been shown to have unique benefits, such as Linear Regression's ease of interpretation and Support Vector Regression's ability to effectively detect complex non-linear patterns [4].

Problem Definition: -

The global oil production has a vital role to play in determining the security, stability, and pricing of energy. Considering all the major oil-producing countries, the Organization of Petroleum Exporting Countries has a major role to play in determining the global oil supply.

However, there are various factors that affect the oil production of this Organization.

Objectives:

1. To assist rural stakeholders (farmers, transporters) by forecasting the oil production that would cause minimal disruption to their activities.

2. To assist oil companies with precise predictions of OPEC's oil production levels, thereby preventing economic losses resulting from changes in oil production.
3. To develop a predictive model that would be able to forecast changes in OPEC's oil production, thereby helping governments make informed decisions regarding strategic reserves.

Literature Review:

1. **Supply-side and global influence**
   Ratti and Vespignani (2015) examined the impact of oil production by OPEC and non-OPEC countries on the world economy. Their study shows that productions shocks strongly affect inflation, growth, and global financial stability, highlighting supply as a key determinant of oil price movement.
2. **Geopolitical risk and volatility regimes**
   Qian et al. (2022) found that geopolitical tensions significantly increase oil price volatility. Using a Markov-switching model, they showed that oil markets shift between stable and turbulent regimes depending on political uncertainty.
3. **Dynamic modelling of geopolitical effects**
   Wang et al. (2021) extended this approach by introducing time-varying switching probabilities. Their results indicate that geopolitical risks influence oil price volatility in a nonlinear and timedependent manner, improving forecasting accuracy.
4. **Climate policy uncertainty as a new factor**
   He et al. (2024) incorporated climate policy uncertainty into oil price models and demonstrated that environmental regulations and energy transition policies create additional volatility in oil markets.
5. **Machine learning and deep learning forecasting models**
   Foroutan et al. (2024) and Lee et al. (2025) showed that deep learning and structural machine learning models outperform traditional econometric approaches by capturing nonlinear patterns and complex economic relationships in oil price forecasting.

Overall, the literature indicates that crude oil price volatility is influenced by supply conditions, geopolitical tensions, climate policy uncertainty, and advance in forecasting techniques. While earlier studies focused on production and macroeconomic factors, recent research emphasizes integrating geopolitical risk and machine learning models. However, there remains a need for a unified framework combining these elements, which the present study aims to address.

Methodology:

This research paper will utilize the data set of 400 records, which is an experimental research design, along with machine learning regression algorithms, which will be preferred based on the prediction of OPEC oil production based on historical data and geopolitical events.

This research will be predicting the regression accuracy using the data set of 400 (values) rows, which include features such as price, opec cut announcement, sanction event, and geopolitical mention.

Algorithms:

1. KNN: (K-Nearest Neighbors) is a supervised learning algorithm that performs regression by predicting continues values based on the average nearest neighbors. To find the distance between two points mainly Euclidean distance are used. The value of k determines the number of neighbors.

This table includes the dataset trained on KNN regressor algorithm along with tunning parameters.

Table 1: Knn Regressor Prediction Table:

Sr. no.	K Value	Train%	Test%	Accuracy%
1	3	0.8	0.2	93.19
2		0.7	0.3	84.07
3		0.6	0.4	83.74
4	5	0.8	0.2	92.49
5		0.7	0.3	84.07
6		0.6	0.4	83.74

This table showing the model accuracy for different settings of K value, train-test split, and number of features, accuracy ranges 83.00% to 93.19%. Best accuracy: 93.19% (K=3, Train/Test=0.8/0.2).

2. SVM: Support Vector Machine (SVM) is a supervised machine learning algorithm used for classification as well as regression purposes by determining an optimal hyperplane to separate classes by maximizing the margin between data points. Work better with high-dimensionality dataset.

This table includes the dataset trained on SVR regressor algorithm along with tunning parameters.

Table 2: Svr Prediction Table:

Sr.no	Kernel Value	Train %	Test %	Accuracy %
1	linear	0.8	0.2	62.42
2		0.7	0.3	58.93
3		0.6	0.4	58.29
4	rbf	0.8	0.2	89.11
5		0.7	0.3	85.28
6		0.6	0.4	85.99
7		0.9	0.1	69.75
8	Poly	0.8	0.2	54.00

This table shows the Support Vector Machine (SVM) model performance with different kernel types (linear, polynomial, RBF), train-test splits, and no of selected features, accuracy ranges

54.00% to 89.11%. Best accuracy: 89.11% (RBF kernel, train/test=0.8/0.2).

3. Linear Regression: Linear Regression is a supervised machine learning algorithm that splits data into decision rules to predict continuous numerical values. This regression determines the relationship between independent and dependent variables.

This table includes the dataset trained on Linear regression algorithm along with tunning parameters.

Table 3: Linear Regression Prediction Table:

Sr No	Train %	Test%	Accuracy%
1	0.8	0.2	60.63
2	0.7	0.3	51.77
3	0.6	0.4	52.30
4	0.9	0.1	53.95

This table shows the Linear Regression model performance. Best accuracy: 60.63%

(train/test=0.8/0.2)

4. Decision Tree: Decision Tree is a machine learning algorithm that splits data into branches based on feature decisions to predict outcomes. It creates a tree where each node represents a condition and leaves represent final predictions.

This table includes the dataset trained on Decision Tree regressor algorithm along with tunning parameters.

Table 4: Decision Tree Regressor Prediction Table:

Sr No	Tune Parameters	Change Value	Accuracy %
1	max_depth	None	87.39
2		3	73.12
3		5	84.35
4	min_samples_ split	2	87.83
5		5	87.91
6		10	88.25
7	criterion	squared_error	88.25
8		friedman_mse	85.46
9		absolute_error	94.43

This table shows the Decision Tree model performance with different splitting criteria (gini, entropy) and max_depth settings.Accuracy range: 73.12% to 94.43%. Best accuracy 94.43% (criterion=absolute error)

Results:

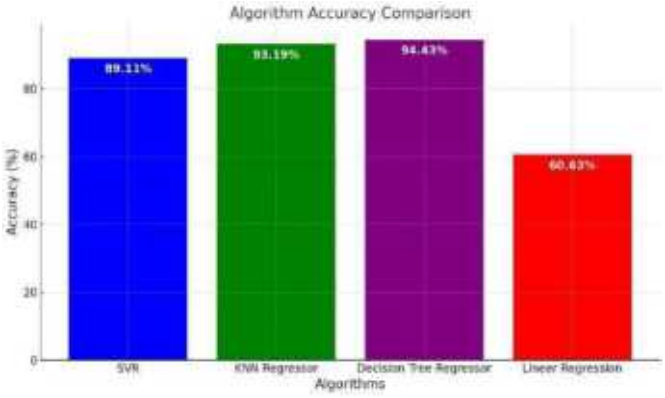
This table shows the accuracy comparison of all machine learning regression algorithms.

Table 5: This table shows the highest accuracy given by each type of regression algorithm.

Sr no	Regressors	Accuracy%
1	KNN	93.19
2	SVR	89.11
3	Linear Regression	60.63
4	Decision tree	94.43

This table shows that after comparing the accuracy score of all machine learning regression algorithms, Decision Tree gives the highest accuracy of 94.43 % among all regressors. This graph shows the comparative analysis of accuracy among KNN, SVM, Linear Regression, and Decision Tree algorithms. The X-axis denotes the machine learning models, whereas the Y-axis denotes the accuracy score obtain during testing.

Figure 1: Comparative analysis of all machine learning classifier algorithm accuracy



This graph shows the analysis of the accuracy score among different machine learning algorithms like Decision Tree, Linear Regression, SVR and KNN. Decision Tree Regressor achieves the highest accuracy 94.43% to predict

the OPEC oil production.

Conclusion:

This project has successfully shown the capability of machine learning regression models to predict OPEC oil production based on past data and major geopolitical occurrences. The **Decision Tree Regressor** has the highest accuracy of **94.43%** to predict the OPEC oil production compared to other models like KNN, Decision Trees, and Linear Regression. The addition of geopolitical indexes and event flags greatly enhances the accuracy of the predictions.

The dataset only contains data from 2015 to mid-2025. Hence, it is not suitable to predict the long-term or rare geopolitical occurrences. Future work could involve the addition of live data feeds (like clouds) to enhance the accuracy and the prediction response time. In the future, it could also be explored how to use models like LSTM or Transformers for betterment.

References :

1. M. Q. P. C. R. M. U. I. J. L. L.-G. H. I. H. Iftikhar, "Daily Crude Oil Prices Forecasting Using a Novel Hybrid Time Series Technique," IEEE Access, vol. Volume 13, p. 98822– 98838, 2025.
2. N. A. A. Karabas, "Deep Learning Approaches in the Effects of Recession and FOMC Minutes on Oil Prices," IEEE Access, vol. Volume 13, p. 28946–28965, 2025.
3. I. A. G. S.-V. O. M. P. C. A. Bâra, "Exploring the Dynamics of Brent Crude Oil, S&P 500 and Bitcoin Prices Amid Economic Instability," IEEE Access, vol. Volume 12, p. 31366–31385, 2024.
4. R. L. S. N. K. Y. N. C. R. Khadka, "Forecasting the Global Price of Corn Using Neural Network and Hybrid Approach," IEEE Access, vol. Volume 13, p. 167424–167438, 2025.
5. A. K. G. A. A. K. R. Roshanpour, "GA-Optimized Self-Attention LSTM for Multi-Asset Price Forecasting: Incorporating Trading Volume Features for Crude Oil, Gold, and Bitcoin," IEEE Access, vol. Volume 13, p. 173487–173509, 2025.
6. A. K. A. R. K. K. A. C. C. H. Xie, "Is Geopolitical Turmoil Driving Petroleum Prices and Financial Liquidity Relationship? Wavelet-Based Evidence from Middle-East," Defence and Peace Economics, vol. Volume 33, no. Issue 7, p. 781–801, 2022.
7. H. J. M. M. H. Farman, "Geopolitical Sentiment as a Leading Indicator: A Hybrid Analytics Approach to Forecasting Oil Volatility and Emerging Market Vulnerability (2015–2025)," Journal of Cognition and Artificial Intelligence, vol. Volume 1, no. Issue 1, p. 6–12, 2025.
8. M. F. J. Q. J. G. J. Z. X. Wang, "From News to Forecast: Integrating Event Analysis in LLM-Based Time Series Forecasting with Reflection," arXiv Preprint, arXiv:2408.15123, 2024.
9. R. A. H. M. D. A. A. S. F. R. A. M. R. M. M. K. M. T. Khan, "Predictive Modeling of US Stock Market and Commodities: Impact of Economic Indicators and Geopolitical Events Using Machine Learning," Journal of Economics, Finance and Accounting Studies, vol. Volume 6, no. Issue 5, p. 112–122, 2024.
10. S.F. F. D. P. V. S. S. A. Edalatpanah, "Investigating the Impact of Unconventional Variables on the Improvement of OPEC Crude Oil Price Forecasting Modeling," Financial Innovation, vol. Volume 12, no. Issue 1, 2026.
11. L. D. P. M. Alruqimi, "Enhancing Multi-Step Brent Oil Price Forecasting with Ensemble Multi-Scenario Bi-GRU Networks," International Journal of Computational Intelligence Systems, vol. Volume 17, no. Issue 1, p. 225, 2024.
12. Q. Z. Lihua Qian, "Geopolitical risk and oil price volatility: Evidence from Markovswitching model," International Review of Economics & Finance, 2022.
13. J. L. V. Ronald A. Ratt, "OPEC and non-OPEC oil production and the global economy," Energy Economics, vol. Volume 50, p. Pages 364–378, July 2015.
14. Y. Z. Mengxi He, "Modelling and forecasting crude oil price volatility with climate policy uncertainty," Humanities & Social Sciences Communications, 2024.
15. C. S. Minho Lee, "Forecasting Crude Oil Prices with a Structural Machine Learning Model," International Economic Journal, p. 423–445, june. 2025.
16. S. L. Parisa Foroutan, "Deep Learning Systems for Forecasting the Prices of Crude Oil and Precious Metals," Financial Innovation, july 2024.
17. F. M. Lu Wang, "Forecasting Crude Oil Volatility with Geopolitical Risk: Do TimeVarying Switching Probabilities Play a Role," International Review of Financial Analysis, vol. Volume 76, July 2021.
18. L. Z. O. A. R. Aurelio F. Bariviera, "Crude Oil Market and Geopolitical Events: An Analysis Based on Information-Theory-Based Quantifiers," Fuzzy Economic Review, vol. Volume 21, no. Issue 1, p. 41–51, 2016.
19. H. W. Z. W. D. W. Y. X. L. Y. Y. Z. X. Wang, "A Hybrid PLO-Transformer-CEEMDAN Model for Natural Gas Price Forecasting," IEEE Access, vol. Volume 13, p. 143489– 143507, 2025.
20. R. G. C. K. M. L. X. S. Jorge Cunado, "Time-Varying Impact of Geopolitical Risks on Oil Prices," University of Pretoria, Department of Economics, 2017.

ISSN : 2348-7143
March 2026
RESEARCH JOURNEY International Multidisciplinary E-Research Journal
Impact Factor (SJIF) - 6.625 | ***Special Issue 389 :*** *Advanced Computing and Artificial Intelligence (ACAI)*

Human Firewall - AI Threat Defense Training

Mr. Pranit Magan Patil,
Mr. Chitransh Yogesh Bhamre

Abstract:

The emergence of new generation of human-centric cyber threats has been triggered by the excessive development of Artificial Intelligence in the contemporary digital environment. Advanced attacks such as deepfakes, voice cloning, and hyper-realistic phishing are no longer interested in the vulnerabilities of the software; they use the psychology of humans. With the old security measures proving to be incapable of keeping up with the manipulation that AI causes, it is now obvious that the individual user should now be made the first line of defense.

This project proposes a web-based platform that is to be designed in order to establish a HFP. Addressing the lack of connection between sophisticated cybersecurity and the fact that there are simple users, the platform will turn the potential victim into an active participant in online safety. It works on a two layered platform Adaptive Education and Real-time Threat Detection. The users are trained on how to identify the nuances of social engineering through a gamified quiz system and an interactive resource center. At the same time, the platform offers useful defense features, such as a dedicated scanner to check the content of the media on AI-generated modifications, and an automatic system to identify the presence of phishing attack signatures in email messages.

The technical architecture is based on the high-performance hybrid backend. Although the user interface and the basic services are based on the MERN stack (MongoDB, Express.js, React, Node.js), the intelligence is heavy duty, and a Python-based AI engine is deployed to analyze deepfakes and recognize patterns. In order to guarantee no less than maximum data integrity and low-latency performance, the project uses a Rust-based database (either SurrealDB or a Rust storage layer). Such combination of both machine learning ecosystem with Python and memory safety with Rust enables the Human Firewall to analyze complex threats in real-time, transforming the digital defense into a software process, reactive, instead of an informed and high-speed human intuition.

Keywords:

Human Firewall, AI Threat Defense, Deep fake Detection, Adaptive Education.

Introduction:

The artificial intelligence has transformed the digital world in numerous ways due to its rapid development. Automated mechanisms, predictive applications, and smart applications have become a normal way of life in industries. These advancements have enhanced speed and efficiency at the expense of generating other new issues in cybersecurity. Modern threats are not restricted to the breach of the hacking systems or the breaking of technical barriers. Several attacks are aimed at directly controlling people. False videos, voice cloning, and well-designed phishing messages are designed to mislead people though the means of trust and human feeling. Due to the increasing lifelikeness of these tricks, one cannot afford to rely on the old security programs or software to remain safe any longer.

The concept of a Human Firewall gains greater significance in such a setting. The Human Firewall - AI Threat Defense Training project is created according to the principle that people can be viewed as an important bulwark. Rather than relying solely on automated protection systems, the project is aimed at enhancing awareness and better judgment and responsible online behavior. It is meant to make people identify and react wisely to threats before becoming some cheap targets to manipulate.

The project involves guided learning and aid monitoring resources. The educational aspect will involve interactive classes and real-life situations, where users are instructed on what constitutes warning signs of social engineering, misinformation and fraudulent communication. Using practice exercises, users get to recognize suspicious patterns, language and digital anomalies. It is a strategy that promotes the inquiry and realistic knowledge rather than the memorization of safety advice.

Also, the platform provides functions that help users to check potentially dangerous materials in addition to training. The verification tools of the media assist in identifying anomalies, which can indicate doctored images, videos, or audio contents. The message analysis tools check the communication patterns that are commonly related to phishing or fraud attempts. They are not used to substitute the decision-making process, but facilitate it. The focus is on allowing users to think critically and technology gives an additional level of support.

In terms of development, the system is made to be stable and efficient. It is developed in the MERN stack consisting of MongoDB, Express.js, React, and Node. js, which enables a user to interact well with the app and be scaled easily. Pattern detection and media review are

patterns of analysis that a Python-based engine supports. A Rust-based database structure helps to improve performance and integrity of the system to improve reliability and secure data processing.

In general, the Human Firewall project is a balanced cybersecurity strategy that appreciates technology and human consciousness. As digital threats keep changing and becoming more legitimate, it is necessary to reinforce the human factor. The initiative will result in a safer digital space, where people actively and voluntarily contribute to their safety and that of their organizations by integrating training with helpful analytical resources.

Research Methodology:

The Human Firewall - AI Threat Defense Training project research methodology was conceptualised as a systematic and realistic process aimed at enhancing the human factor in cybersecurity. The method was to carefully study the problems, plan the system, develop the model, create the platform and do a detailed analysis in such a way that the final framework was technically correct and useful in a practical setting.

Problem Identification and Requirement Analysis.

The paper has started by analyzing the current trends in cybersecurity and how it is limited to address the contemporary threats of media manipulation, voice recordings, and phishing messages that are highly targeted. To learn what gaps exist, academic literature, industry security reports and documented cyber incidences were consulted. Based on the analysis, it was noted that most security systems focus on technical protection, but not the ease with which people can be deceived. In light of this knowledge, the project was established with an objective goal, which is to develop a system that enhances user awareness besides aiding in automated threat detection.

System Architecture Design

A two-layer system structure was to be used after the requirements were defined. The initial layer was addressing the user education by structured training modules, which targeted the awareness and critical thinking. The second tier focused on detection of suspicious digital activity with the help of analytical tools. To make sure that the structure is scalable and can be processed efficiently, the hybrid technical structure was chosen. The interface and application framework were built with the MERN stack, which comprises of MongoDB, Express.js, React, and Node.js. An engine was written in Python to execute analytical and machine learning steps with a Rust-supported database layer added to improve stability and performance in general.

Training and Model Development.

Python libraries were used to create machine learning models to facilitate detection activities. Supervised learning techniques were used in analyzing the email structure pattern, message body and meta data that is usually associated with fraudulent communication in order to identify phishing. To perform the task of detecting manipulated media, the analysis methods were created to detect inconsistencies in image and video files. The phishing samples and synthetic media samples were important datasets that were gathered and processed to retain quality. The widely accepted evaluation measures that were used in determining model performance include accuracy, precision, recall, and false-positive rates.

Platform Development

React has been used to make the web interface simple and user-friendly. The analytical engine and the database could communicate with each other with the help of Node.js and Express.js used to handle the backend processes. In MongoDB user data, training data and system data were stored. The site contained interactive learning courses and simulation-based activities to enhance awareness. It also enabled users to post the suspicious content to be automatically reviewed by the detection system.

Performance Evaluation and testing.

The final system has gone through a number of testing processes. The functional testing was used to ensure that no error was made and all the components worked together. Performance testing was used to test response times, scaling under simulated usage conditions and consistency in the results of detection. Also, user feedback was gathered to find out whether the training modules enhanced the awareness of the participants regarding cybersecurity threats and made them more resolute in their decision-making.

Refinement and Validation

Using the results of the testing, changes were done to enhance speed, stability, and accuracy of detection. The database structure was improved and helped to facilitate the smooth functioning of the system and reduce the number of operational errors. Final validation outcomes have shown that structural awareness training in combination with analytical threat detection contributed to user preparedness without significantly affecting the technical performance.

Overall, the methodology used was a way of adhering to the research and development practice, but was focused heavily on practice. The response to the dynamic nature of the digital threats resulted in the establishment of a balanced cybersecurity framework through simultaneous defense of technology and human awareness by the project.

Literature Review / Related Work:

In terms of development, the system is made to be stable and efficient. It is developed in the MERN stack consisting of MongoDB, Express.js, React, and Node.js, which enables a user to interact well with the app and be scaled easily. Pattern detection and media review are patterns of analysis that a Python-based engine supports. A

Rust-based database structure helps to improve performance and integrity of the system to improve reliability and secure data processing.

In general, the Human Firewall project is a balanced cybersecurity strategy that appreciates technology and human consciousness. As digital threats keep changing and becoming more legitimate, it is necessary to reinforce the human factor. The initiative will result in a safer digital space, where people actively and voluntarily contribute to their safety and that of their organizations by integrating training with helpful analytical resources.

Research Methodology:

The Human Firewall - AI Threat Defense Training project research methodology was conceptualised as a systematic and realistic process aimed at enhancing the human factor in cybersecurity. The method was to carefully study the problems, plan the system, develop the model, create the platform and do a detailed analysis in such a way that the final framework was technically correct and useful in a practical setting.

Problem Identification and Requirement Analysis.

The paper has started by analyzing the current trends in cybersecurity and how it is limited to address the contemporary threats of media manipulation, voice recordings, and phishing messages that are highly targeted. To learn what gaps exist, academic literature, industry security reports and documented cyber incidences were consulted. Based on the analysis, it was noted that most security systems focus on technical protection, but not the ease with which people can be deceived. In light of this knowledge, the project was established with an objective goal, which is to develop a system that enhances user awareness besides aiding in automated threat detection.

System Architecture Design

A two-layer system structure was to be used after the requirements were defined. The initial layer was addressing the user education by structured training modules, which targeted the awareness and critical thinking. The second tier focused on detection of suspicious digital activity with the help of analytical tools. To make sure that the structure is scalable and can be processed efficiently, the hybrid technical structure was chosen. The interface and application framework were built with the MERN stack, which comprises of MongoDB, Express.js, React, and Node.js. An engine was written in Python to execute analytical and machine learning steps with a Rust-supported database layer added to improve stability and performance in general.

Training and Model Development.

Python libraries were used to create machine learning models to facilitate detection activities. Supervised learning techniques were used in analyzing the email structure pattern, message body and meta data that is usually associated with fraudulent communication in order to identify phishing. To perform the task of detecting manipulated media, the analysis methods were created to detect inconsistencies in image and video files. The phishing samples and synthetic media samples were important datasets that were gathered and processed to retain quality. The widely accepted evaluation measures that were used in determining model performance include accuracy, precision, recall, and false-positive rates.

Platform Development

React has been used to make the web interface simple and user-friendly. The analytical engine and the database could communicate with each other with the help of Node.js and Express.js used to handle the backend processes. In MongoDB user data, training data and system data were stored. The site contained interactive learning courses and simulation-based activities to enhance awareness. It also enabled users to post the suspicious content to be automatically reviewed by the detection system.

Performance Evaluation and testing.

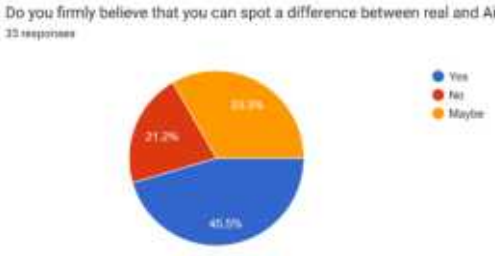
The final system has gone through a number of testing processes. The functional testing was used to ensure that no error was made and all the components worked together. Performance testing was used to test response times, scaling under simulated usage conditions and consistency in the results of detection. Also, user feedback was gathered to find out whether the training modules enhanced the awareness of the participants regarding cybersecurity threats and made them more resolute in their decision-making.

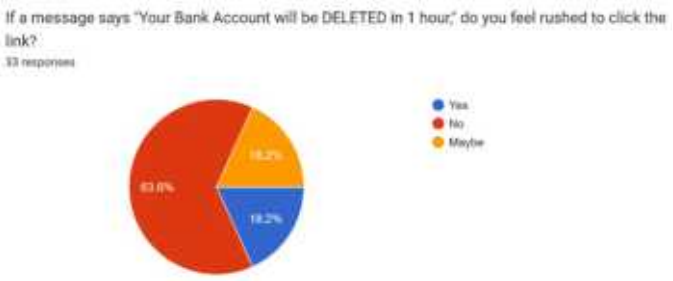
Refinement and Validation

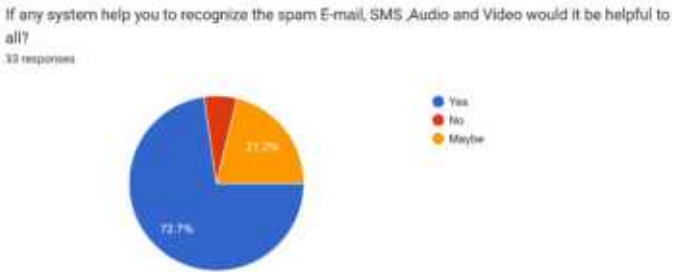
Using the results of the testing, changes were done to enhance speed, stability, and accuracy of detection. The database structure was improved and helped to facilitate the smooth functioning of the system and reduce the number of operational errors. Final validation outcomes have shown that structural awareness training in combination with analytical threat detection contributed to user preparedness without significantly affecting the technical performance.

Overall, the methodology used was a way of adhering to the research and development practice, but was focused heavily on practice. The response to the dynamic nature of the digital threats resulted in the establishment of a balanced cybersecurity framework through simultaneous defense of technology and human awareness by the project.

Survey Data :







Results and Discussion:

The storage layer that was added was Rust-based, which was critical to ensuring better system speed and reliability. Since the platform is doing real-time analysis, including scanning of suspicious emails or viewing uploaded media, it needs to be processed on a fast basis and feedback provided instantly. The response time was evenly distributed without much delay recorded in performance tests despite the number of users accessing the system at the same time. The good memory safety characteristic of Rust assisted in minimizing the occurrence of errors and unforeseen crashes in the backends, which contributed to a stable work. These findings indicate that the hybrid backend architecture can be used to establish a good balance between processing speed, analytical ability, and secure data management and ensure scalability to future expansion.

The python-based machine learning-based threat detection engine was assessed meticulously based on the capability to identify phishing emails and captured media, such as deep fake content. To train the model in phishing detection, the training was done using labeled data which consisted of genuine and malicious emails. The system was able to detect suspicious patterns during evaluation like harmful links, strange metadata structure, and language which is usually used in phishing attacks.

With trial and error, parameter optimization, and feature tuning, the false alarm rate was minimized. This was done so that the legitimate emails were not flagged wrongly and at the same time offered good protection against fraudulent communication. This balance in detection and false positives reduced the level of mistrust that the user placed in the automated analysis.

The module of the media analysis was created to investigate uploaded images and videos in the case of artificial manipulation. Inconsistency in pixel structure, lighting variations and facial alignment among other digital artifacts commonly found in synthetic media were identified by pattern recognition techniques.

The system was able to consistently and repeatedly identify manipulated content that is typically generated, under controlled testing conditions. The performance of the detection module was also competent, although more sophisticated generative tools that are currently being developed are still difficult to detect because they are rapidly evolving. These works point to the fact that the adoption of the machine learning in user-friendly security solutions can enhance the ability to notice the presence of manipulated materials earlier and prevent threats associated with misinformation and identity theft.

In addition to the performance of the system, the educational aspect was the key to achieving the objectives of the project. The site incorporated interactive learning and quiz question games that enhanced knowledge on social engineering, phishing, and manipulation through AI. The first tests revealed that it was challenging to differentiate between legitimate and deceptive communication to many users. Having passed the training modules, the participants showed a better understanding of the suspicious web links, the use of emotionally controlling language, and typical patterns of fraud. There were also claims of greater confidence in the cases of uncertain digital situations and this indicated that the training enhanced the knowledge together with the decision making ability.

High participation and completion rates were made by the interactive format. Scenario based quizzes were well received by the users as opposed to plain reading materials. The results of simulated phishing exercises revealed that

there was a significant reduction in impulsive clicking behavior following training, and it could be concluded that the behavior actually improved and was not just memorized.

These results help to assume that cybersecurity education is more effective when educational activities are practical, involving, and founded on real-life scenarios. System integration testing ensured that there was a seamless flow between the interface, which was developed using MERN, the Python analytical engine, and the database layer, which was implemented using Rust. The authentication, file uploads, progress tracking and real-time analysis were done using application programming interfaces without much technical effort.

issues. Hand simulated load conditions were stable.

performance under various concurrent requests indicating that the architecture may be used in academic institutions or other organizational environments. New detection features can be added to the modular structure as well without a significant redesign.

This platform is a complementary method in comparison to conventional security tools like firewalls and antivirus software. Traditional systems are frequently reliant on familiar threat signatures and might be entirely insufficient to counter manipulation to attempt to persuade human judgment. Conversely, this model is a blend of user training and analytical support. The study participants of the awareness modules were cautious of suspicious communication more before automated scanning tools were used. This behavioral change is why it is important to increase the human aspect of cybersecurity.

There were some limitations that were identified. The accuracy of detection is related to the quality and variety of training data, and new developments in the field of generative technologies can produce more advanced content that questions modern detection algorithms. Continuous updates, updating of models and expansion of datasets are thus needed. The discrepancy in the degree of user engagement also implies the necessity of the regular refresher training and updated simulation exercises to provide long-term awareness.

On the whole, the results indicate that a combination of technical detection frameworks and well-organized awareness training forms an effective and efficient cybersecurity framework. The consistent accuracy of the detection, stable performance, quantifiable betterment of user awareness, and favourable behavioral change altogether come in favour of the goals of the project. Through a blend of smart analysis and humanistic education, the platform can help create a more proactive measure of increasing security susceptibility to current, AI-driven internet attacks and building a more resilient cyber world.

Conclusion:

The blistering growth of Artificial Intelligence has introduced significant transformations to the area of cybersecurity. Earlier, the majority of cyberattacks involved intrusion into the systems with the use of technical vulnerabilities. Nowadays, there are numerous threats that are aimed at direct manipulation of people. The use of deepfakes audio and video, faked voices, and well thought-out phishing messages can be extremely realistic, which makes digital fraud more difficult to detect. Under these circumstances, it is no longer sufficient to rely on conventional security solutions, such as firewalls and antivirus programs. Human Firewall - AI Threat Defense Training project was developed to address this challenge by enhancing the human capabilities in.

cybersecurity by combining awareness training and smart detection support.

The paper reveals that integrating the structured awareness campaigns with real-time analytical tools develops a more robust and preventative defense mechanism. During controlled tests, the machine learning models that had been created on the platform could recognize phishing and indicators of manipulated media with a very high accuracy. The constant modifications and refinement minimized wrong alerts, enhancing the reliability of the application. Moreover, the Rust-based backend made the systems more stable and guaranteed faster processing as well as the safety of stored data. It proves that the hybrid technical structure is efficient and safe.

User behavioral and awareness changes were also high. The learning aspect whereby the participants had interactive modules and quiz-based activities assisted them to better comprehend tricks like emotional manipulation, suspicious links, and deceptive communication patterns. Users had been observed to be significantly more effective in evaluating digital content in their judgment after undergoing the training. In the simulated phishing attempts, it was found that a great deal of participants hesitated to act and would hold up to confirm the information before acting. These behavioral changes show that the use of awareness-based interventions can result in achieved and quantifiable changes in online safety behavior.

Another insight of the findings is the importance of integrating automated tools with educating the users. The systems of detection deliver swift technical examination whereas knowledge training offers analysis and sound skepticism. They both form a tall defense model where technology enhances human decision-making other than substituting it. The strategy will transform cybersecurity into more of an active model and motivate collaboration between smarter systems and more informed users. This kind of cooperation is particularly necessary since threats

of AI-driven nature keep becoming more sophisticated.

Despite the fact that certain constraints are still present (diverse datasets are necessary, and the models should be regularly updated in response to the alterations in the attack strategies), the general outcomes are quite effective and user-friendly. Continuous development, renewed training resources and frequent system enhancement will ensure long term performance and flexibility.

To sum up, Human Firewall is a viable and visionary solution to the cyber hazards of the modern world. The project offers a stable and scalable solution to the threats posed by AI by giving people knowledge and aiding them with the help of analytical tools. The results support the idea that cybersecurity is not only about the high-technology level but also highly educated and vigilant users. In the case of the interplay between human awareness and intelligent systems, a more stable and safer digital space can be achieved.

References :

1. isFake.ai isFake.ai. (n.d.). AI-generated content detection tool. Retrieved February 20, 2026
2. DetectVideo AI DetectVideo AI. (n.d.). AI video and deepfake detection platform. Retrieved February 22, 2026
3. AI Video Detector AI Video Detector. (n.d.). Online AI-generated video detection tool. Retrieved February 24, 2026
4. AI Video Detector AI Video Detector. (n.d.). AI-based video authenticity checker. Retrieved February 25, 2026
5. AI Voice Detector AI Voice Detector. (n.d.). AI-generated voice detection tool. Retrieved February 27, 2026

ISSN : 2348-7143
March 2026
RESEARCH JOURNEY International Multidisciplinary E-Research Journal
Impact Factor (SJIF) - 6.625 | ***Special Issue 389 :*** *Advanced Computing and Artificial Intelligence (ACAI)*

Integration of Career Guidance

Miss. Sakshi Vilas Patil,
Mr. Pranit Magan Patil

Abstract:

AI-Powered 5-Level Quantitative Aptitude Platform for Complete Student Domain Profiling In today's fast-changing world, students new to tech or any field feel lost, wasting years on mismatched paths without knowing their true strengths across all domains. This AI platform solves this with a 5-level test (0-5) that reveals natural talents in reasoning (deductions, arguments), mathematics (numbers, algebra), logic (patterns, syllogisms), verbal (comprehension, vocabulary), statistics (probability, trends), data analysis (charts, insights), spatial thinking (visualization, geometry), and critical thinking (problem solving, decisions).

Level 0 offers simple MCQs accessible to ages 12+: basic reasoning (odd one out), math (add/subtract), verbal (synonyms), logic (shape sequences), spatial (mirror images). Progression builds: Level 1 (ratios, analogies); Level 2 (algebra, reading comprehension); Level 3 (probability, data tables); Level 4 (calculus, complex arguments); Level 5 (Olympiad multi-domain puzzles). AI adapts 20-30 questions real-time; 70% accuracy advances users.

Domain scores (0-5) create parent-friendly heatmaps:

• Engineering/Tech: High math+logic+spatial

• Law/Media: Strong verbal+critical thinking

• Data Science: Statistics+data analysis+reasoning

• Finance: Math+logic+data analysis

• Design/Architecture: Spatial+creative reasoning

• Management: Balanced verbal+critical thinking Parent dashboards provide clear progress charts, smart career suggestions, and detailed reports showing strengths and weaknesses across all 8 domains. The platform supports multiple languages including English, Hindi, and regional Indian languages, with mobile- friendly gamification features like badges and daily streaks. Core tests are free while premium practice materials and coaching are available, giving parents complete insights to make informed decisions about their child's future.

Keywords: Multidimensional Aptitude Assessment, Adaptive Testing Model,Cognitive Domain Mapping,Stream Selection Guidance, Holistic Student Evaluation,Data-Driven Academic Decision Making

Introduction:

Students have been expected to make decisions regarding their careers very early in their educational setup in the contemporary world. Most of them take up careers without the vision of what their personalities are capable of doing, and they have to spend years studying things that they do not excel at. Traditional advice depends primarily on exams and overall observations and can overlook significant cognitive abilities that would determine future success. In order to bridge this gap the notion of Complete Student Domain Profiling was developed. It is an organized evaluation system that is intended to recognize and chart out learner talents in various dimensions.

The ultimate aim of the platform is to help make informed decisions in academics by identifying strengths at an early stage. It measures eight basic cognitive areas; and the reasoning ability, mathematical understanding, logical analysis, verbal comprehension, statistical thinking, data interpretation, spatial visualization, and critical reasoning. The combination of these areas provides a holistic image of the aptitude of a specific learner as opposed to using subject marks.

The progressive assessment model can be offered at the age as young as twelve years old. It is made of various steps which become more cognitively complex. Primary levels are concerned with the basics of reasoning and arithmetic functions as well as pattern recognition. In-between levels include algebraic thought, proportional reason and reading comprehension. Higher-levels add probability, reasoning based on calculus and tasks of higher-order analytic. The last level is integrated problem-solving scenarios which evaluate several domains simultaneously.

A test process will be constructed with an adaptive mechanism. Students are tested through a set of questions that prepare them to the existing level of performance. Promotion to the next stage must have a set level of competence in practice, and so promotion must be an indication of the same. This design will provide an attempt to balance accessibility with challenge during the evaluation.

In addition to the testing, the platform converts performance data into visual summaries that can tell whether it is aligned with different academic and professional disciplines such as engineering, law, finance, data analysis, design, and so on. Analytical reports consisting of the

performance trends and domain based strength indicators are given to the parents and educators. Such reports present evidence-based discussions on the choice of streams and how to prepare them.

Approximately 2,000 items, both basic and advanced, are properly constructed in the question bank. The content development was realized in cooperation with highly-qualified educators and the references to the established standards of the curriculum. Different student groups were pilot tested with variations in geography, type of school, and gender population and this was fair and reliable. The consistency and bias were reduced by means of statistical validation procedures.

Adaptive testing framework is a reflection of how teachers can change the instructional level according to the student feedback. The efficiency and higher specificity in determining domain-related strengths were compared with conventional fixed-format tests. When applied in classrooms, instructors would be able to understand the results of their dashboards showing individual and group patterns, and they would be able to focus on the academic conversation. Parents also had advantage of concise reporting format where aptitude results were compared to Likely academic streams which enabled them to be confident in taking part in planning decisions.

The system is provided in various India languages to increase availability and make sure that it is understood among people of different languages. The processing of data is in compliance with the current privacy laws, and the control over student data is performed by parents.

On the whole, the framework provides a methodological way of seeing student aptitude based on the multidimensional assessment and adaptive assessment. It offers the basis of more personalized academic planning by surpassing the percentage-based understanding and focusing on domain specific strength. Such strategy gives better direction, lessens the educational options that do not align with individual talents, and promotes students attending avenues that reflect their own strengths.

Literature Review / Work Relation:

Each classroom is comprised of students who have varied learning styles. Others learn fast though do not know how to put into practice what they learn and others do poorly on long-term tests but are good in critical thinking and pattern recognition. Conventional report cards focus on general marks, which lack these level skills. Educators and parents are supposed to examine the data beyond the highest marks in order to learn how individual children absorb the information, solve problems, and use knowledge. The awareness of these differences can aid in putting students in appropriate academic and career lines.

EduPath relies on such principles as psychometrics and educational psychology. It does not suggest some new theory; instead, it implements the common assessment practices based on the Indian education system. In India, academic decisions are based on board examinations and competitive entrance examinations and coaching patterns, stream pressure, and family pressure. The platform provides an organized means to measure the aptitude of students using tested methods.

One is adaptive assessment, which is based on such models as Item Response Theory (IRT). Studies indicate that questions whose challenge is at the same level with the ability of a learner give precise answers. Difficult questions are discouraging to students whereas easy questions are boring. Similarity provides a moderate experience which instills confidence and eliminates undue stress. Research has affirmed that adaptive techniques provide more accurate findings rather than preset tests.

According to research done in India, it is important to study various cognitive abilities, instead of depending on the subject marks alone. Different streams demand a combination of skills in order to succeed. Engineering requires spatial sight and analysis, law, humanities, and management demand verbal expression and debate, and critical thinking. These results indicate that percentages cannot be used alone in forecasting success in the future.

EduPath applies an Indian cognitive, eight-domain system (targeted at the secondary student level). The model assesses reasoning, verbal, spatial, quantitative, data interpretation, logical analysis, statistical awareness and decision making. Through the consideration of all these areas, the system offers a well-rounded picture of all the strengths of each student, which is useful in assisting in careful academic planning rather than using the aggregate grades.

Classroom practices are supported by research. With easy to understand, graphical progress reports, teachers will be able to provide specific guidance. These resources reveal the strengths behind poor grades. As an illustration, a student whose test scores were moderate may be having great analytical or spatial skills that require to be identified and encouraged.

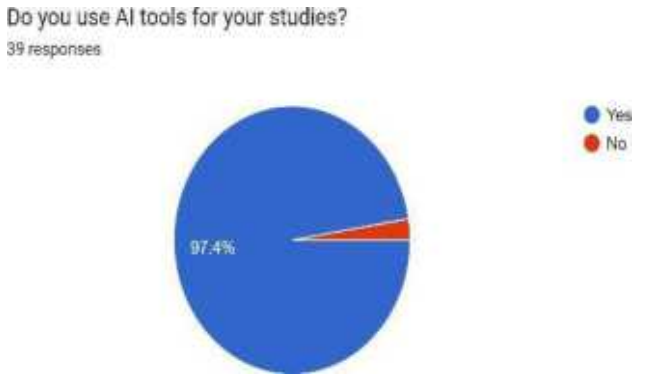
Parents also benefit. Research indicates that families benefit when there is a clear and structured feedback on the abilities of their child. In cases where aptitude outcomes are correlated with the academic paths such as PCM, PCB, Commerce, or Humanities, the parents will be able to make informed and evidence-based choices, rather than making the decisions based on social or assumptions.

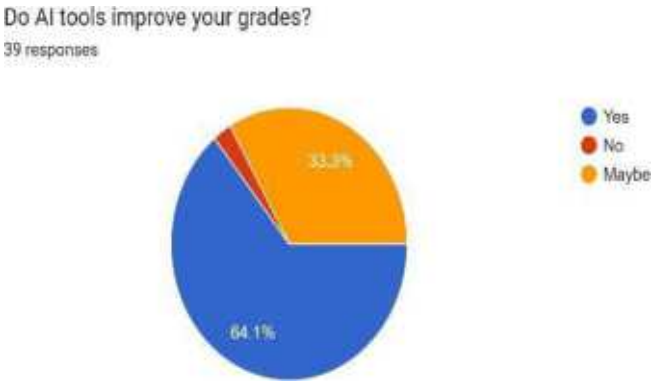
Although there is international literature on the adaptive assessment and aptitude measurement, its systematic application in the Indian secondary schools is still developing. Such issues are a lack of structured

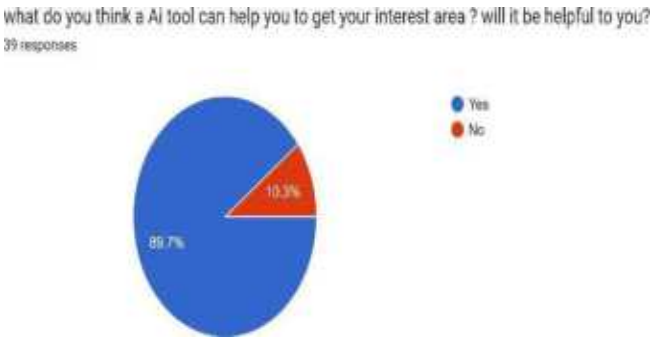
multidomain mapping to select the streams, class analytics, barebones parent communication, and multilingual support.

EduPath addresses these problems by integrating effective assessment principles and solutions that are applied to Indian schools. It is based on balanced and research-based approaches that can be used to provide equitable, precise, and useful instructions to students, teachers, and families.

Survey Data :







Results and Discussion:

EduPath revealed that a glance at a number of factors about the ability of a student can significantly improve the educational direction in India.

There are various cognitive strengths of students which are not necessarily reflected on report cards. Others develop a memorizing and recalling talent. Others are more analysis, logical and visualizing. Such disparities bring up the main point, which is that academic percentages are not that important to define the full intellectual potential or future possibilities of a student.

The paper has discussed the usefulness of an adaptive testing model. The graded test scored to fit the level of ability of each student, which produced a balanced testing environment. Learners were not intimidated with extremely difficult questions and not bored with overly easy questions. This match to ability and challenge served to keep in mind and remain confident during the test. On the measurement side, the adaptive format provided a more accurate assessment in comparison with the old system of fixed tests in which all students are given the same group of questions irrespective of their ability.

Subject-based marks were not enough and the assessment system considered eight different cognitive areas. They were spatial reasoning, logical thinking, verbal understanding, analytical ability and data interpretation among others. Such a wider analysis gave a more in-depth analysis of personal strengths. As an example, there were moderate overall students who were good at pattern recognition or visualization skills. Others who had a moderate level of language scores exhibited significant levels of analysis.

These observations provide more weight to the relevance of multidimensional assessment in the support of academic planning and stream choice. The findings are that the percentage scores should not be relied on solely to make decisions in the streams. Academic planning, in case of the availability of detailed information on the strengths of the cognitive aspect, can be more individual and evidence-based. The multidimensional model assisted in pinpointing talents that could not be identified using the traditional grading system, and consequently, the guidance conversations were closer to the inherent capabilities of the students compared to the demands of the outside world and the social tendencies.

The system was also valued by teacher comment. Teachers said that it was easier to access the profile of individual students with the help of visual dashboards and well-structured summaries. The teachers would be in a position to look at domain specific strengths and areas of improvement as opposed to concentrating on ranks and general marks. This resulted in new insights into students who in some instances had been found to be average in normal assessment tests, which facilitated more objective and accurate discussions on the future academic courses.

The structured reporting format was also a positive response by the parents. Where aptitude results were evidently connected to common academic courses of PCM, PCB, Commerce, or Humanities, families were in a better position to make the right decisions. Clear descriptions of strengths minimized ambiguity and inspired evidence-based decisions by disregarding peer pressure and social norms.

The research also proved that multidimensional assessment as an adaptive method may be effectively used in Indian schools. Even though international studies advocate such methods, they are not well exploited in most local contexts. With the integration of the pre-existing measurement principles and the background of the exam-based educational practices and high parental engagement, the system was viable and culturally applicable.

Another important result was fairness and inclusiveness. Conventional tests have been identified to reward students who excel in tasks of memorization. Conversely, multidimensional approach identifies varying types of ability which includes analytical reasoning, visualization and logical problem solving. This wider appreciation will minimize the risk of measuring students as the results of examining and will encourage more balanced perception of intelligence.

In general, the findings show that systematic and evidence-based aptitude testing can be used to optimize educational counseling. Adaptive testing, detailed cognitive mapping, teacher-centric analytics and parent friendly reporting are all solutions to ineffective traditional stream selection strategies. The evaluation of students in various areas makes academic planning more precise and close to personal talents.

Overall, the results confirm the opinion that learning counseling becomes better when it is based on an in-depth comprehension of the student capabilities. By outgrowing the percentage-based assessment and developing a more comprehensive portrait of cognitive strengths, academic trajectories can be created by the schools in a more meaningful, fair, and each learner-specific way.

Conclusion:

EduPath demonstrates that academic advice is deeper when the students are seen beyond what is on the test scores. Each learner will have his/her own strengths. Others are successful memorizers or workers of set tasks; others are good critical thinkers or visualizer, analyzers or complex data interpreters. Ordinary tests focus on grades and overlook these diverse talents. The study cautions against using scores as the only measure to capture qualities that have major impacts on long term academic and career development.

The balanced assessment model developed a balanced evaluation process. The system maintained an appropriate level of challenge to a learner by setting the level of difficulty on questions according to the ability of the particular student. This minimized needless pressures and avoided distengagement and this gave a supportive and accurate testing experience.

The adaptive format provided the better picture of the individual strength and weaknesses as compared to fixed exams where all students take the same exams, and are asked the same questions. A cognitive ability was analyzed using an eight-domain structure that was looking into several facets. The system unveiled the thinking and learning processes of students rather than singly recorded academic measure. This expanded analysis brought out the strengths which could otherwise be overlooked by traditional report cards. This kind of wisdom is particularly appreciated when academic streams are being selected, since it is at this point that future chances are being determined.

The reports and visual summaries provided by teachers were used to provide evidence-based guidance. They have transcended the generic impressions, and taken the area-specific skills and developmental domains. The change provided a more individualized learning environment in which direction was correlated with established strengths.

The unambiguous reporting system was advantageous to parents. Easy and clear outcomes allowed families to participate in academic planning. The connection of aptitude results to stream choices known (PCM, PCB, Commerce, Humanities) provided more certainty in the decision. Clarity of explanations minimized social pressure and supported the decisions which were in line with natural capabilities of a student.

In sum, the research reveals that, a structured, research based assessment system empowers decision making in the Indian secondary education. Through integration of adaptive testing, multidimensional testing, teacher feedback, and clear reporting is a sure basis of academic guidance. More importantly, it restates the fact that students need to get supported according to their strengths and not solely by their exam marks.

EduPath enhances equitable and considerate scholarly scheduling by concentrating on the capability rather than on the scores. This will enhance likelihood of students taking up areas that they are talented in and this will enhance confidence, satisfaction as well as long term success.

References :

1. Mindler Mindler. (n.d.). AI-powered career guidance and comprehensive psychometric assessments.
2. YouScience YouScience. (n.d.). Aptitude-based career discovery and student profiling platform.
3. AptiQuest AptiQuest. (n.d.). AI-adaptive aptitude training and performance heatmap analytics.
4. Apt AI Apt AI. (n.d.). Multi-dimensional cognitive ability and adaptive assessment tool.
5. CareerVillage (Coach) CareerVillage. (n.d.). Coach: AI-driven personalized career mentoring for students.

ISSN : 2348-7143
March 2026
RESEARCH JOURNEY International Multidisciplinary E-Research Journal
Impact Factor (SJIF) - 6.625 | ***Special Issue 389 :*** *Advanced Computing and Artificial Intelligence (ACAI)*

Quantum Computation: What We Know and What We Don't

Mr. Rishabh Vishwakarma,
Roshani Baviskar

Abstract

This Research Paper is based on Quantum-Tech and development. Quantum Computation represents the advancement of modern computers. Unlike the classi-cal computation, quantum computation utilizes qubits, which exploit superposition, entanglement, and quantum interference to process information in new modern ways. This research represents an analytical and review-based study of quantum mechanics and quantum computation advancements and their future use. It ex-plains what we know and what we don't about quantum computers. As we are developing Artificial Intelligence so what Quantum Computer can do that classical computers can't! It is also regional comparative analysis of quantum research and development. This research aims to provide clear and structured understanding for students and people around the world about what we know and what we don't about the Quantum Computation.

1 Introduction

Quantum Computation is next generation of computing technology. Classical computers follow binary logic, but quantum computer works on quantum mechanics rules. This change allows the quantum computer have new possibilities. Richard Feynman was the first who suggested quantum simulation in 1982 [1]. In late 1981 His "Core Idea", were came from his lecture about "Simulating physics with computers", where feynman noted that classical computer is not very efficient to simulate many-particle quantum systems. Also he stated his famous statement: "Nature isn't classical dammit". Later in 1985, Feynman explored the theoretical model of a universal quantum computer. Also he stud-ied about the limitation about the computation, including the reversibility of computing which was closely linked towards the quantum operations. In 1985 David Deutsch devel-oped the "Universal Quantum Computer Model [2]. Where he proposed his theoretical, programmable device that uses the quantum bits also called as (Qubits) or (Quantum Bits) to simulate any finite physical system, that surpasses the limitations of classical Turing machines. It utilizes quantum superposition, entanglement, and interference for performing the parallel faster calculations based on the laws of physics, Defined as Church-Turing principle in quantum terms. He argued that truly Universal Computer must be based on Quantum Mechanics. Because classical computers cannot achieve the physical process. In 1994 Peter Shor published the Shor's Algorithm, In which Algorithm was capable to factorize the large integers in polynomial time O((log N)3) [3]. The algorithm was primarily made to perform efficient integer factorization. Algorithm was equally ca-pable of solving the discrete logarithm problem in polynomial time. In 1996 L. Grover Discovered the algorithm that was capable to find an unsorted item from database of N items in $O(\sqrt{N})$ time. [4].

2 Literature Review

2.0.1 Quantum Computation by I.J. Mathematical sciences and computing Date: 8 Oct 2022

I reviewed this paper and got idea about the core concept about quantum computation is that Quantum bits are replacement of classical bits because it holds the superposition. In simple words when we talk about the classical computers then it basically works on binary O and 1. but the quantum computation works on qubits where they are in su-perposition it means they can exist is both reality O and 1 simultaneously [5]. If we talk about entanglement it is nothing but a system ca exist in multiple states at one time. because according to the author: "Entanglement is uniquely quantum correlated between particles. It means entanglement is the quantum process in which two or more particles are linked together, if one particle changes it's behavior then other linked particle also having similar behavior. Author also talked about Quantum Gates and Circuits. Where he mentioned that Quantum gates manipulates the Qubits using the Unitary Transfor-mations. An Unitary operations preserves the vast probabilities.

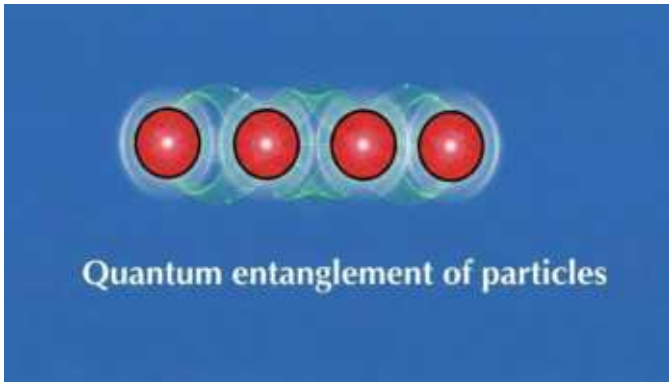


Figure 1: Quantum Entanglement Illustration

He also stated that the future challenges of quantum computing is that hardware im-plementation, decoherence and measurement limitations. If we compare original research and feynman's proposal for quantum computation, and Shor's and Grover's algorithms is is a fantanstic milestones. Afterwards by major technical institutions like IBM and D-Wave started to build the prototype quantum systems.

3 What do we know as for today?

3.1 What do we know exactly?

We humans are in the race to develop the machine who can think 100+ times more effi-cient than human brain. As we followed our science and computation path we got highly capable supercomputers. But we realized after sometime, it is not the highly capable. When it comes to physical calculations classical computers lacks to much. Then someone who proposed the idea about the quantum computing. We started the new race. 1982 was the start of the quantum computation, but the era of quantum computer started when the first functional 2-Qubit quantum computer was created in 1998 by respected researchers ISACC CHUANG (IBM/MIT), NEIL GARSHENFELD (MIT), and MARK

KUBINEC (UC Berkeley). They used the magnetic resonance(NMR) Technology [6]. As we already know that in 1998 David Deutsch was made theoretical model that device was first to load daya and solve the problem.

If we talk about the Quantum Chip then IBM, Google, D-Wave and intel were first Tech-Giants and major players still today to develop the quantum chips. They specially focusing on Superconducting qubits. While Google was the first who claimed the" Quan-tum Supremacy".

2025 and recent advancements: Microsoft's majoranal" and Google's "Willow" pro-cessor are the recent discovery or achievement in quantum computation.

4 What we don't know as of today?

4.1 What we don't know exactly?

This question have some limited answers if we see, but it is important question to achieve the quantum reality!

Major Problems we don't know:

- Decoherence
- Error Rates
- Noise Correction
- Practical Problems

Decoherence: is the state where qubits loose their quantum state where the entangle-ment breaks down and qubit get collapse. We really don't know how to build qubits at largely in thousands and millions needed for practicality and large scale applications [3].

Error rates: Physical gates are imperfect therefore there is large possibility of errors [4].

Noise Correction: As we know quantum error correction is not possible practically as we have lesser qubits. but it is possible theoretically. But when it comes to actual implementing then we lacks of qubits [4].

Practical Problems: There are practical problems such as It's scalable quantum hard-ware, new algorithms as we have fewer algorithms we don't know about which classical problem can be efficiently solved with quantum computers. Where Machine learning is hot topic but we really don't know how quantum computer will outperform the classical relative datasets.

5 Future Directions of Quantum Computation

5.1 Future Reality

Any Technology is new when we discover it, but it becomes familiar when we solve it's problem and make it efficient for good use. When we use every thing to make a highly advance quantum computer then it can do most difficult tasks that need 100 of years if we give similar task to classical computer. I am not exaggerating but it is the reality. Quantum computer will be the common computers for any human.

Artificial Technology we developed that will be more advance and fastest technology in the world. Yes as we have these kind of expectations from upcoming quantum technology is great. But it is important to think parallel of it. It can be the dangerous too. as we daily listening AI morphing, data retrieving, etc. What if the quantum computer give it to some push and it will like Human Created The "GOD" who have Limitless possibility and still unanswered questions we have. But the management of technology in safe hands can prevent these cinematic problems for sure.

6 Regional Analysis of Quantum Technology

6.1 United States of America

USA is the leading country to develop the quantum machines at first. Major contributors

- IBM Quantum
- Google Quantum AI
- Microsoft Quantum
- IonQ

Where google demonstrated the quantum supremacy in 2019-2020 using it's advance 53-qubit processor [7]. Along side IBM has been developed the processor more than 100 qubits and also they introduced the Cloud-Based Quantum Access.

Note: The USA's National Quantum Initiative Act 2018 provided strong funding sup-port for Quantum Research and Development.

6.2 China

As whole world in race of quantum technology, China has made rapid progress in Quan-tum Communication System.

Their Major Achievements:

- Micius Quantum Satelite

- Long distance quantum key distribution -(part of quantum entanglement)
- Photonic Quantum Computers

China demonstrated satellite-based QKD and quantum teleportation experiments over long distances [8]. It has invested billions in quantum technology research centers.

6.3 Europian Union

The Europian Union launched the Quantum flagship program with very large funding support in favour of Quantum Research and Development.

EU and other Countries Envolvement:

- Germany
- France
- Austria
- Neitherlands

Where EU Research focuses on quantum communication and networks, Superconductng qubits and the most famous Photonic Processors.

6.4 India

India recently has been launched the National Mission on Quantum Technologies and Applications also known as (NM-QTA).

Research Institutions are participating in this are:

- IISc Bangalore
- IIT Madras
- IIT Bombay
- TIER Mumbai

India is still working on quantum communication networks and quantum cryptography research.

7 Global Investment Comparision

7.1 Countries and their investment on Quantum research and Development

Region	Approx. Investment (USD)
USA	$3+ Billion
China	$10+ Billion
EU	$1+ Billion
India	$1 Billion (Planned)

Table 1: Global Quantum Investment Overview

Global Investment shows the awareness of the research and development among respective country. Where China is Major Investor in this Quantum Technology [9].

8. Conclusion

In conclusion we can say that Quantum technology is still under development and most of big tech giants are developing the quantum chipsets. due to lack of qubit stabalization we are unable to create qubits. but quantum computation is still a hot topic and it has been created much more vision for the future. Quantum computation In Artificial Intelligence is Simillar to have overpowered computer that can think and solve multiple problems at once.

References :

1. R. Feynman: https://s2.smu.edu/ mitch/ class/5395/papers/feynman-quantum-1981.pdf
2. https:/ /www.daviddeutsch.org.uk/wp-content / deutsch85. pdf
3. P. Shor, "Algorithms for Quantum Computation," FOCS, 1994.
4. L. Grover: https://arxiv.org/pdf/quant-ph/9605043
5. I. J.Mathematical and sciences: https:// www.mecs-press.org/ijmsc/ijmsc-v8-n4/I. JMSC-V8-N4-5.pdf
6. First Quantum Computer: https://www.spinquanta.com/news-detail/the-first-quantum-computer-everything-you-need-to-know20250214081413
7. Google Quantum AI: https://quantumai.google/
8. Chinese Quantum Achivements: https:// thequantuminsider.com/2026/02/06/chinese-researchers-clear-hurdles-for-long-distance-quantum-networks/
9. Global Investment: https://www.mckinsey.com/capabilities/tech-and-ai/our-insights/tech-forward/quantum-technology-investment-hits-a-magic-moment

RESEARCH JOURNEY International Multidisciplinary E-Research Journal
Impact Factor (SJIF) - 6.625 | ***Special Issue 389 :*** *Advanced Computing and Artificial Intelligence (ACAI)*
ISSN : 2348-7143
March 2026

AI Surveillance Systems: Protection or threat?

Atharva Dilip Patil,
Tejaswi Ravindra Mahajan
Student, RCPET's IMRD, Shirpur (MS) India

Abstract

AI-powered surveillance systems like facial recognition and smart cameras were created to improvise public safety by identifying vulnerabilities quickly and supporting law enforcement. In many cases, these surveillance system help authorities to respond faster and monitor large areas efficiently. But, with this implement of AI and reliability there are serious issues like cyber threats and ethical concerns.

This paper's purpose is to compare evidences of different countries and to check it's benefits as well as risks. If you take a look at technical side, AI systems can be breached through adversarial attacks, where inputs are given to fool the model, data poisoning, which corrupts the training process of system. Weak encryption and over data storage lead to increase vulnerability, making a single breach highly damaging. Governance gaps, including unclear regulations and weak oversight, add to these risks.

Real-world deployments also show operational issues such as false positives and biased predictions, sometimes leading to wrong actions. To solve these challenges, the paper proposes a layered strategy, including Zero-Trust architectures, federated learning, strong cryptography.

Keywords

AI surveillance, Cybersecurity, Cyberattacks, Deepfakes, Data breaches.

Introduction

AI-powered surveillance systems are being adapted mostly in cities, public transportation networks, financial institutions, and national security infrastructures. Growth in computer vision have transformed traditional CCTV cameras into intelligent monitoring systems which are capable of detecting objects, analyzing human behavior, and performing real-time facial recognition.

Some researchers suggest that these technologies help in strengthening public safety by improving situational awareness and helping prevent crime. Whereas, other researchers consider that these can raise concerns regarding privacy invasion, algorithmic bias, and potential security threats that could be misused or exploited.

This research examines AI-based surveillance systems with a balanced perspective. It seeks to understand both their positives as well as the negatives that can be risk. By studying recent technical developments alongside real-world deployments, the study identifies key security weaknesses and proposes strategies to enhance system.

Objectives

- To, study AI surveillance which help us to understand where problems occur.
- To, identify key cybersecurity and weaknesses that can affect the system at different level.
- To, propose a defence focused design to reduce cyber risks while keeping the system effective.
- To, strength both technical security and organizational adaptability.

Literature Review

1.1 AI Security Landscape

AI systems face various security risks through complete lifecycle, right from development to real-world deployment phase. In surveillance systems, facial recognition and behavior-detection models are major concern because they operate in real time and process sensitive data. Any weakness in design, training, or implementation can lead to manipulation, misuse, or cyberattacks.

1.2 Social & Ethical Critiques

Research on policy and ethics highlights that widespread surveillance brings serious social consequences, including privacy violations, misuse of data beyond its original purpose, and discrimination caused by biased AI models. Researchers say that even if these systems are technically accurate, that alone is not enough to justify their large-scale deployment. Strong governance, accountability, and ethical safeguards are essential to protect democratic values and individual rights.

1.3 Empirical Deployments & Failures

Studying of prototype systems, which includes Hawk-Eye and large-scale city trials, shows both the advantages as well as the risks of AI surveillance. While detection capabilities have improved significantly, issues such as false positives, long-term data storage, raise some serious concerns. These challenges often weaken public trust.

Research Methodology

Primary Data Analysis

Primary data for this study was collected through a structured questionnaire administered to 54 respondents. The purpose of the survey was to understand how people perceive AI-powered surveillance systems—specifically their views on its benefits, risks, level of trust, and the need for global regulations.

The overall responses suggest a cautious but practical attitude toward AI surveillance. Most participants rated their trust at a moderate level, with 38.9% selecting 3 on a 5-point scale, and 27.8% choosing 4. This shows that while people see value in AI surveillance, they are not fully confident in its reliability or safety.

If we consider advantages, (61.1%) agreed that AI surveillance plays a role in crime prevention and public safety, 51.9% said that it supports faster investigation and law enforcement. But, problems are also there. Around 66.7% of respondents said that cybersecurity breaches and hacking is the most concern and privacy invasion (51.9%) and possible government misuse (37%) are also some of the challenges .

A question was asked if surveillance databases were hacked then , responses were (35%) felt that blackmail or harassment can be the consequences while (32%) stated it can be a national security threat, 56% supported that there is the need of urgent laws to regulate AI surveillance.

Secondary Data Analysis

previous research work shows us that AI surveillance systems offer us both benefits as well as some serious concerns.

Fontes et al.(2022) states in there research that AI monitoring can improve public safety , but it also raises some concerns about privacy, misuse of data . That only technology is not enough strong rules and governance is necessary.

Hu et al.(2021) highlighted that AI are problematic to cyber threats like adversarial attacks and data poisoning Which shows that surveillance systems can be manipulated if proper security steps are not taken

Some case studies such as Hawk eye (Ahmed & Echi, 2021) and London Underground AI trail (2024) ellaborated that detection accuracy is improvises, but practical challenges like false positives, and privacy risk is also a major corncern

Overall, after understanding secondary data we evaluate that AI surveillance is beneficial, but without any governance and cybersecurity it creates an societal problem.

Findings

So, the overall finding of the study tells us that AI powered surveillance has two sides. First side we can say the positives of it, according to the survey and existing research tells us that these systems can significantly improvise public safety. Many of the respondents agreed that AI surveillance can play a important role in crime prevention, speed up investigation and making monitoring more efficient.

On, the other side study also highlights some serious concerns. Respondents identified data breaches and hacking as one of the biggest risk. This shows that people have concerns about their provided and stored data that can be misused. Previous research supports this concerns, many vulnerabilities like adversarial attacks, manipulation of AI models and data poisoning. It also warns us about privacy breach, biased outcomes and a problem that surveillance can be misused beyond it's orginal motive.

Another finding is that there should be rules and regulation, More than 50 % of respondents believed that there is an need of governance in it.

Overall, the study tells us that people see AI surveillance can be useful, but it all depends on trust But, there should be strong cybersecurity measures and governance can play a important role in it.

Conclusion

This study comes to an conclusion that the AI system brings both positives and well as negatives . Considering the technical side , it improvises security , it strengths monitoring system , and helps to detect the threats . But, from social and ethical things, there raise an serious concerns about privacy, data misuse and cybersecurity concerns.

Findings also stated that advancement in technology and high accuracy is not always enough, Some strong security systems, laws and regulation and human supervision is still necessary. Public opinion states that people acknowledge the benefits of AI surveillance, but they need strong assurance of their personal data is protected.

So, future of AI surveillance just should not advancement in technology it should improvise in building trust , accountability, transparency. We can say that an Balanced Approach that regulations and technological advancement is necessary to ensure that AI surveillance can improvise safety .

RESEARCH JOURNEY International Multidisciplinary E-Research Journal ISSN : 2348-7143
Impact Factor (SJIF) - 6.625 | ***Special Issue 389 :*** *Advanced Computing and Artificial Intelligence (ACAI)* March 2026

A Comparative Evaluation of Machine Learning Classifiers for Email Spam Detection Using Natural Language Processing

Vidit Prajapati, Jevin Dobariya, Kajal Patil
Babu Madhav Institute of Information Technology, Uka Tarsadia University, Bardoli, India

Absract

The rapid growth of digital communication has led to extensive increase in unwanted email traffic. In the recent year, email spam has become a major threat. As the email usage is increased, the number of spam messages has also grown exponentially. Spam emails not only interrupt communication but also led to serious security risks, including phishing and malware attacks. Furthermore, spam email often contains malicious links or attachment to compromise system and steal sensitive information. As a result, development of spam detection system is becoming an important area of research. We compared five popular supervised classifiers, namely: Logistic Regression, Support Vector Machine (SVM), Naive Bayes, K-Nearest Neighbours, and Random Forest in this work. All the classifiers were tested using two feature extraction methods: TF-IDF and Bag-of-Words. The experiments were conducted on a labelled dataset of 9,022 email messages. In addition, a Streamlit-based web application was developed to automate model selection and support real-time email classification. Among other classifiers, SVM classifier with TF-IDF achieved the highest accuracy (98.28). The findings indicate that the feature extraction techniques is considered as the important impact on the overall performance of spam classification models. Additionally, the study indicates that the selection of feature extraction technique plays an important role in determining the performance of spam classification models.

Keywords: Email Spam Detection, Machine Learning, Natural Language Processing, Feature Representation

I. Introduction

Electronic mail is one of the most widely used forms of digital communication in this modern world. It serves as a medium for communication for both personal and professional. At present, billions of user relay on email as a primary means of communication. The vast number of messages are exchanged each day which highlights its significance role for information exchange, organizational communication and global connectivity. Moreover, majority of global email traffics consists of unwanted messages. These include promotional advertisements and phishing emails which tries to deceive an individual or group of users. As a result, such emails may cause inconvenience and also threaten an individual's security and privacy. Earlier, traditional rule-based filtering methods were useful in identifying and blocking spam messages. However, as spamming techniques have become more advanced, these traditional methods performances have become less effective. Due to this reason, there is a need for the improved and more intelligent spam detection system. Moreover, email spam also has serious economic and security impact. Unwanted messages consume bandwidth, storage space, and also employee time, which leads to increase in organizational costs and reduced productivity. In many cases, spam emails are also used to carry phishing attacks, malware, ransomware, and other forms of large-scale financial frauds. The email-based threats are responsible for a numerous number of data breaches worldwide. Thus, reliable spam detection system is required for maintain the digital security.

To address this need, spam filtering techniques have progressed significantly over the time. The development of spam filtering shows a continuous struggle between defenders and the spammers. Earlier filtering approaches used to relay on simple keyword lists and rule-based methods. However, these techniques were easily bypassed through minor changes in the text. The modification may include changing in spellings, character substitution or may include the legitimate-looking text to avoid detection. As a consequence, later the introduction of Bayesian statistical filters marked a significant advancement. These filters used probabilistic models to examine word occurrences and separate spam from legitimate emails. Although this method improved detection accuracy, but still limits by its static nature. Thus, due to static nature the effectiveness of this approach declined over time due to changes in the spam characteristics. This phenomenon is referred to as concept drift. As a result, machine learning techniques is adopted to adapt to new and evolving spam techniques. Machine learning provides an alternative to traditional rule-based spam filtering methods. Instead of relying on manually handcrafted rules, machine learning models learns the patterns from labelled training data. And due to this reason, this data-driven approach adapts to different and changing forms of spam. Apart from this, the integration of Natural Language Processing (NLP) improves the spam detection by identifying the linguistic and contextual features.

In this study, particular emphasis is placed on the feature extraction stage in the machine learning–based text classification. Feature extraction is an important step which

converts the raw text into numerical form which can be later processed by classification algorithm. Moreover, the selection of an appropriate representation plays an important role in determining the model performance. In this paper, the study focuses on two commonly used vectorization techniques in spam detection: Term Frequency–Inverse Document Frequency (TF-IDF) and the Bag-of-Words (BoW) model. TF-IDF assigns weights to words based on their importance in a document compared to the entire dataset, which reduces the effect of very common words and give weight to more meaningful terms. Whereas, the Bag-of-Words model represents each document using simple word frequency counts without applying normalization. Although both feature extraction techniques are widely used but their performance has not been thoroughly compared under the same experimental conditions across multiple classifiers. Therefore, this study conducts a systematic comparison of TF-IDF and BoW to evaluate and compare these feature extraction techniques.

The principal objectives of the present study are as follows: (i) to conduct a systematic and controlled comparative evaluation of five widely used supervised machine learning classifiers for email spam detection; (ii) to assess the influence of two NLP-based feature vectorization strategies, TF-IDF and BoW, on classification performance across all evaluated models; (iii) to identify the optimal classifier-vectorizer configuration with respect to accuracy, precision, recall, and F1-score. By addressing these objectives under a unified experimental protocol, this study provides a rigorous empirical basis for informed classifier and vectorization strategy selection in operational email security systems. The present study addresses this by training five distinct classifiers with two feature vectorization techniques. All models were evaluated on a dataset under strictly controlled and consistent experimental conditions. Furthermore, a Streamlit-based web application was developed to automate model selection and real-time email classification.

The remainder of this paper is structured as follows. Section II presents a literature survey on spam filtering, developments from probabilistic models to recent deep learning approaches. Section III details the dataset, pre-processing techniques, feature extraction methods, and classification algorithms. Section IV presents and analyses the experimental results across ten classifiers. Section V presents conclusions, and Section VI outlines directions for future research.

II. Related Work

Since email use rose sharply in the 1990s, people have studied how to filter spam automatically. Sahami et al. [6] used Bayesian classification and showed that probability models were better than keyword blacklists. Later, Drucker, Wu, and Vapnik [2] used Support Vector Machines, proving SVM's ability to manage high-dimensional sparse spaces made it suitable for text analysis.

Feature engineering also became a research area. Manning et al. [3] created the TF-IDF weighting system. This system reduces the importance of words common in many documents and increases the importance of words specific to a document, particularly useful for spotting spam words. In contrast, Bag-of-Words uses raw frequency data. Bird et al. [1] showed that it remains competitive on shorter texts where term weighting is less useful.

Ensemble methods came into use with the rise of Random Forest [4], which lowers variance through bagging and random feature selection. Breiman's work showed that combining many weak learners greatly reduces generalisation error, which applies to noisy, high-dimensional text data. Studies then verified that bagged ensembles consistently perform better than single decision trees on email data while being easier to understand than kernel-based methods [10].

The base of current spam classification comes from Cortes and Vapnik [13], who created the support-vector network and established the maximum-margin principle that backs LinearSVC. They proved that generalisation error is limited by the ratio of the margin to the radius of the smallest enclosing sphere. This proof gave experts a solid reason to prefer large-margin classifiers for high-dimensional problems like text, where features greatly outnumber training examples. This theoretical base separates SVM from simpler options and explains its strong performance across many spam tests.

A separate research area has consistently rated Naive Bayes as a basic spam-filtering option. Androutsopoulos et al. [12] did one of the first studies, finding that Multinomial Naive Bayes reached over 97% accuracy on the Ling-Spam data with little preprocessing. They also showed that attribute selection—keeping only the most distinctive tokens—could improve precision without losing recall. This finding justifies the 5,000-feature limit used here. Blanzieri and Bryl [11] then reviewed over 20 years of learning-based spam filters and said that, despite its independence assumption, Naive Bayes is still among the best classifiers based on the accuracy-to-compute ratio, especially for use on edge devices or low-latency servers.

More recently, deep learning methods have received much interest. Devlin et al. [9] created BERT, a bidirectional transformer pre-trained on large data sets that can be fine-tuned on later classification tasks with minimal labelled data. When used on email and SMS spam data, BERT-based models often exceed 99% accuracy by seeing long-range contextual dependencies that n-gram models miss. For instance, they can recognise that the word free is okay in

free to reply but suspicious in free prize winner. BiLSTM architectures have shown similar benefits for sequence-level semantic modelling. Still, transformer models need more processing power, memory, and labelled data to fine-tune well, and they are harder to understand than linear classifiers. This work focuses on standard supervised learning to give a fair, simple comparison that experts can use without GPU infrastructure, making transformer methods a next step instead of the main comparison point.

III. Methodology

The following section describes the methodology presented in this study. It also provides a detailed description of the dataset, pre-processing steps, feature engineering techniques, model implementation, and evaluation metrics used in this study. Figure 1 illustrates the end-to-end system workflow.

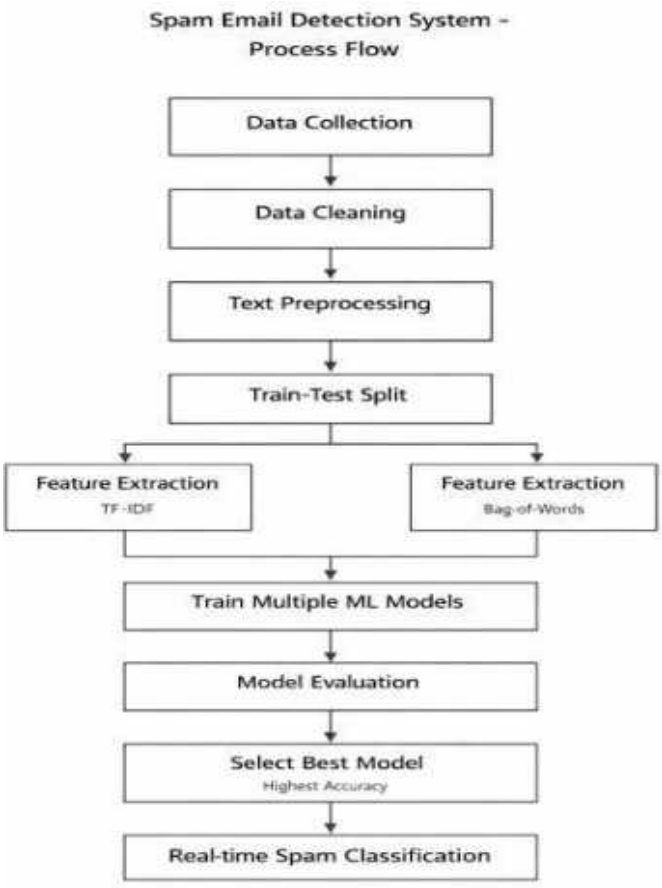


Figure 1: End-to-end workflow of the spam detection system — from data collection and NLP preprocessing through model training, evaluation, and real-time classification.

A. Dataset

The study uses a combined dataset from publicly available datasets from websites like Kaggle, AWS, UCI and real-world data labelled email datasets stored in CSV format and combining them in a single dataset that is containing 9,022 messages after null removal. Each record carries a binary label "ham" for legitimate mail and "spam" for unwanted mail and the raw message text. The dataset is split 80/20 into training and test subsets using stratified sampling (random_state =42) to preserve the class ratio across both partitions.

B. Text Pre-processing

Email text contains a large amount of noise, such as punctuation marks, numerical characters, HTML tags, and frequently occurring function words. These components generally do not contribute for classification. The presence of these noise increases the data complexity and also reduce the performance of the classification. Therefore, a structural pre-processing step are performed systematically to clean and normalize the text as describe in the following subsection.

Lowercase: All alphabetic characters were converted to lowercase to make sure the uniform representation of the words and treat as a single token. For instance, FREE, Free, free are treated as free.

Character removal: Non-alphabetic characters are replaced with whitespace to eliminate irrelevant symbols while keeping word boundary intact.

Whitespace Normalization: Extra spaces were removed by converting multiple spaces into a single space and trimming leading or trailing spaces.

Tokenization: The cleaned text is then divided into individual words using NLTK's word_tokenize.

Stopword removal: Commonly used words such 'as', 'the', 'and', 'is', etc are removed using NLTK's English stopword list. As these words do not significantly contribute to classification.

Lemmatization: Words are then reduced to their base forms using the WordNetLemmatizer to reduce vocabulary size while retaining meaning. For example, the word 'running' is lemmatized to its base form 'run'.

C. Feature Extraction for Spam Detection

After pre-processing, the cleaned email text is transformed into numerical representations which is suitable for machine learning algorithms. Subsequently classification models cannot directly process raw text, thus feature extraction techniques is applied to convert textual data into structured feature vectors.

In this study, the Bag-of-Words (BoW) model is used as a primary feature representation method. In this approach, each email is being represented as a vector of word frequencies. The model is capturing the occurrence of terms within a document without considering their order. This representation is helping to identify patterns that distinguish spam from legitimate emails.

Moreover, another feature extraction technique, the Term Frequency–Inverse Document Frequency (TF–IDF) is

applied to assign weights to words based on their importance. Term Frequency (TF) is measuring how frequently a word appears in a document, while Inverse Document Frequency (IDF) is reducing the weight of words that appear frequently across many documents. As a result, TF–IDF is emphasizing more informative and discriminative terms.

These feature extraction methods are transforming textual data into high-dimensional numerical feature vectors. The resulting vectors are serving as input to supervised classification algorithms for the email spam detection task.

D. Classification

After extracting the feature vectors, supervised classification algorithms are being applied to categorize emails as spam or non-spam. These algorithms are learning patterns from the training data and are predicting the class labels for unseen emails based on the extracted features. Five classifiers are trained using a scikit-learn Pipeline framework which integrates the vectorizer and the classifier into a single unified workflow. This structured design is ensuring that all pre-processing and feature extraction steps are performed within the training folds only. Thus, it prevents data leakage during the training and evaluation phases. The following classification algorithms are being evaluated in this study: Logistic Regression (LR), Support Vector Machine (SVM), Naïve Bayes (NB), K-Nearest Neighbours (KNN), and Random Forest (RF). Each model is being trained using the extracted feature vectors and is learning to distinguish between spam and legitimate emails. Logistic Regression is modelling the relationship between input features and class labels through a linear decision boundary. Support Vector Machine is identifying an optimal separating hyperplane in the feature space [2]. Naïve Bayes is estimating class probabilities based on Bayes' theorem under the independence assumption [6]. K-Nearest Neighbours is assigning class labels based on similarity measures [3]. Random Forest is aggregating multiple decision trees to enhance predictive performance and reduce overfitting [4].

Evaluation Metrics

To evaluate the performance of the classification models, standard evaluation metrics is used. These metrics measures how effectively the models are distinguishing between spam and non-spam emails.

Accuracy is being calculated to measure the overall proportion of correctly classified emails:

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

Moreover, Precision is computed to evaluate the correctness of positive (spam) predictions. Recall measures to determine the model's ability to correctly identify spam emails. Each model is evaluated on a separate test set using four standard performance metrics. These metrics helps to measure how well the model is classifying spam and non-spam emails. In spam detection, precision and recall have important practical meaning. A false positive occurs when a legitimate email is wrongly classified as spam. This may cause important messages to be blocked and can affect the user's experience. A false negative occurs when a spam email is incorrectly classified as legitimate. This allows unwanted or potentially harmful content to reach the user. Furthermore, the Confusion Matrix is analysed to examine the distribution of TP, TN, FP, and FN values. These evaluation metrics are collectively used to compare the performance of different classifiers and to identify the most effective model for email spam detection.

IV. Experimental Results And Analysis

In this section, the performance of the implemented classification models is analysed and compared. The experiments are being conducted using the pre-processed dataset and the extracted feature representations. Table 1 presents the performance results of all ten model–vectorizer combinations. The table includes accuracy, precision, recall, F1-score, and the confusion matrix components, namely True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). The results are organized according to the type of feature representation used. Initially, the models based on the TF-IDF vectorizer are presented, followed by those based on the Bag-of-Words (BoW) vectorizer. Within each group, they are sorted by accuracy from highest to lowest, matching the summary view from the Streamlit app.

Table 1. Comprehensive Model Performance Comparison (All 10 Combinations)

Model	Vectorizer	Accuracy %	F1 %	TN	TP	FN	FP
SVM	TFIDF	98.2816	96.7742	1308	465	27	4
Random Forest	TFIDF	97.6164	95.4401	1311	450	42	1
Logistic Reg.	TFIDF	97.2284	94.6809	1309	445	47	3
Naive Bayes	TFIDF	97.1175	94.5148	1304	448	44	8
KNN	TFIDF	93.7361	87.0264	1312	379	113	0
Logistic Reg.	BOW	98.0044	96.2422	1307	461	31	5
SVM	BOW	97.8381	95.9077	1308	457	35	4
Random Forest	BOW	97.7827	95.7627	1312	452	40	0
Naive Bayes	BOW	97.3947	95.1992	1291	466	26	21
KNN	BOW	94.1242	87.9271	1312	386	106	0

The SVM + TF-IDF configuration achieved the highest classification accuracy of 98.28%, with an F1-score

of 96.77%. The model incorrectly classified four legitimate messages as spam and failed to detect 27 spam messages. The notably low false-positive rate is of particular practical significance, as users are likely to lose confidence in a filter that erroneously suppresses legitimate correspondence. Logistic Regression demonstrated competitive performance, achieving 97.23% accuracy (F1: 94.68%) with TF-IDF and 98.00% accuracy (F1: 96.24%) with BoW. Although marginally inferior to SVM in terms of accuracy, its substantially lower training overhead and interpretable feature weights render it a viable option in resource-constrained deployment environments. Naive Bayes yielded superior results with BoW (97.39%, F1: 95.20%) compared to TF-IDF (97.12%, F1: 94.51%), a finding consistent with the theoretical preference of MultinomialNB for raw term frequencies over normalised weights. Random Forest similarly achieved strong performance, attaining 97.62% accuracy with TF-IDF and 97.78% with BoW. The exceptionally low false-positive count (FP=1 for TF-IDF) indicates that this model rarely misclassifies legitimate messages. KNN achieved the lowest performance among all evaluated classifiers, recording 93.74% accuracy with TF-IDF (F1: 87.03%) and 94.12% with BoW (F1: 87.93%). The high volume of undetected spam messages under TF-IDF renders this configuration unsuitable for practical deployment. This degraded performance is attributable to the well-documented limitations of distance-based methods in sparse, high-dimensional feature spaces, wherein Euclidean distance loses meaningful discriminative power as the number of zero-valued dimensions increases. Table 2 presents the full confusion matrix for the best-performing combination- SVM with TF-IDF evaluated on the 20% held-out test set.

Table 2. Confusion Matrix — SVM + TF-IDF (Best Model)

	Predicted Ham	Predicted Spam
Actual Ham	1308 (TN)	4 (FP)
Actual Spam	27 (FN)	465 (TP)

Among all test messages, the model correctly classified 1,308 legitimate emails as ham (TN=1308) and 465 spam messages as spam (TP=465). Only 4 false positives were generated legitimate messages incorrectly labelled as spam alongside 27 false negatives representing spam messages that evaded detection. These figures correspond precisely to those reported in the Streamlit application summary table, thereby confirming the reliability and consistency of the automated model selection pipeline.

V. Conclusion

This study presents a systematic comparison of five machine learning classifiers paired with two NLP vectorization methods for the task of email spam detection. The Support Vector Machine with TF-IDF vectorization demonstrated the most favourable balance of accuracy (98.28%), precision (99.15%), and F1-score (96.77%) across all evaluated configurations. Logistic Regression constitutes a compelling alternative, offering competitive classification performance combined with reduced training overhead and enhanced interpretability. The experimental results further demonstrate that the choice of vectorization method is an important and often underestimated design decision in spam classification systems. TF-IDF and Bag-of-Words each exhibit distinct strengths depending on the classifier with which they are paired, and this interaction should be evaluated empirically rather than assumed. These findings contribute a reliable empirical benchmark for practitioners and researchers designing email security systems under realistic operational constraints. In conclusion, this study establishes that machine learning-based spam detection, when coupled with appropriate NLP feature extraction, can achieve accuracy levels suitable for production deployment. The comparative methodology adopted here offers a replicable framework that future studies may extend to encompass additional classifiers, larger datasets, and more diverse linguistic contexts.

VI. Future Directions

Future work should explore transformer-based models such as BERT and BiLSTM architectures, which capture long-range contextual dependencies beyond the reach of n-gram representations. Extending the system to support multilingual datasets and incorporating multi-modal signals, including email headers, embedded hyperlinks, and attachments, would broaden its applicability to real-world threat scenarios. Continual learning mechanisms and federated training paradigms represent promising directions for maintaining model accuracy over time while preserving user privacy.

References :

1. S. Bird, E. Klein, and E. Loper, Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit. Sebastopol, CA: O'Reilly Media, 2009. Available: Natural Language Processing with Python
2. H. Drucker, D. Wu, and V. N. Vapnik, "Support vector machines for spam categorization," IEEE Transactions on Neural Networks, vol. 10, no. 5, pp. 1048–1054, Sep. 1999. Available: Support vector machines for spam categorization
3. C. D. Manning, P. Raghavan, and H. Schütze,

Introduction to Information Retrieval. Cambridge: Cambridge University Press, 2008. Available: Introduction to Information Retrieval

4. T. M. Mitchell, Machine Learning. New York, NY: McGraw-Hill Education, 1997. Available: Machine Learning
5. F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011. Available: Scikit-learn: Machine Learning in Python
6. M. Sahami, S. Dumais, D. Heckerman, and E. Horvitz, "A Bayesian approach to filtering junk e-mail," in Proc. AAAI Workshop on Learning for Text Categorization, Madison, WI, 1998, vol. 62, pp. 98–105. Available: A Bayesian approach to filtering junk e-mail
7. V. Vapnik, The Nature of Statistical Learning Theory. New York, NY: Springer-Verlag, 2010. Available: The Nature of Statistical Learning Theory
8. The NLTK Project, "Natural Language Toolkit Documentation," 2023. [Online]. Available: Natural Language Toolkit Documentation
9. J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in Proc. NAACL-HLT, Minneapolis, MN, Jun. 2019, pp. 4171–4186. Available: BERT: Pre-training of deep bidirectional transformers for language understanding
10. L. Breiman, "Random forests," Machine Learning, vol. 45, no. 1, pp. 5–32, Oct. 2001. Available: Random forests
11. A. Blanzieri and A. Bryl, "A survey of learning-based techniques of email spam filtering," Artificial Intelligence Review, vol. 29, no. 1, pp. 63–92, Feb. 2008. Available: A survey of learning-based techniques of email spam filtering
12. I. Androutsopoulos, J. Koutsias, K. V. Chandrinos, and C. D. Spyropoulos, "An evaluation of naive Bayesian anti-spam filtering," in Proc. Workshop on Machine Learning in the New Information Age, Barcelona, Spain, 2000, pp. 9–17. Available: An evaluation of naive Bayesian anti-spam filtering
13. C. Cortes and V. Vapnik, "Support-vector networks," Machine Learning, vol. 20, no. 3, pp. 273–297, Sep. 1995. Available: Support-vector networks

RESEARCH JOURNEY International Multidisciplinary E-Research Journal
Impact Factor (SJIF) - 6.625 | ***Special Issue 389 :*** *Advanced Computing and Artificial Intelligence (ACAI)*
ISSN : 2348-7143
March 2026

AI-Based Mobile Application for Learning Indian Sign Language Alphabets and Numerals

Jisaheb Lisa[1], Modi Megh[1], Salot Neel[1], Icecreamwala Ansh[1], Rakesh Savant[1]
[1]Babu Madhav Institute of Information Technology, Uka Tarsadia University, Bardoli, Gujarat

Abstract

There are more than 6.3 million people with hearing impairments in India, but the opportunities of the formal Indian Sign Language (ISL) learning are scarce. The majority of the offered materials are based on the fixed pictures or the ready-made videos, which leaves the learners with nothing to practice with and to understand whether they are applying it in a right way. The paper introduces a mobile app which introduces real-time, AI-based ISL recognition on the daily smartphone, enabling users to study and train 35 ISL signs (A-Z and 1-9) with real-time visualization of the results. The app is based on an architecture whereby Flutter controls the interface, cloud services via Firebase but a native Kotlin layer controls the entire machine learning pipeline on the device itself. The system does not process complete camera images, but instead identifies 21 anatomical points on the hand of the user using MediaPipe HandLandmarker, and sends a 130 dimensional normalized feature vector to a low-weight Dense Neural Network to perform classification. This method reduces input data by 99.99% over the traditional image-based methods to produce a model of only 486 KB that classifies signs within 1-5 milliseconds. This solution has not come easily, an early CNN-based system was abandoned with 180-220ms inference latency and high background sensitivity resulting in the landmark-based system which currently achieves 15-20 detections per second on the mid-range Android phones. The recognition engine has been integrated into a full-fledged learning platform, which has predefined lessons, live feedback practice via camera, and quizzes to monitor progress. An experience point, streak system, and achievement gamification layer will motivate users to come-back and practice. The experiments on various devices, lighting types, and backgrounds prove that the system is well-behaved in a real-life scenario and that careful feature engineering in combination with attentive dataset preparation can provide accurate real-time sign language recognition without the need to buy costly hardware or without using computationally intensive models.

Keywords : Indian Sign Language(ISL), Real-time Gesture Recognition, Hand landmark Detection, Deep Neural Network (DNN), Computer Vision

1. Introduction

The primary mode of communication between the deaf and hard-of-hearing individuals all over the world is sign language. But hearing people can hardly learn it well [1]. The Indian Sign Language (ISL) is not the same as others such as the American Sign Language (ASL) but it has its own words and grammar [2,3]. There are not many interactive applications to educate about its peculiarities. Conventional approaches involve still images, videos or live instructors [4]. These cannot provide an immediate feedback regarding whether or not your signs are right or wrong. Examples of learning complex hand movements have been demonstrated to be highly enhanced by real-time feedback [5,6]. The majority of the ISL tools simply allow you to view demos without examining your own signs. This results in a "feedback gap" which reduces confidence, engagement and results [7].

The development of the newest computer vision and artificial intelligence allows tracking poses and recognizing gestures with high accuracy. MediaPipe of Google is distinguished by hand tracking in real time. It identifies 21 points on each hand which are visible in various lights, hand sizes and skin tone [8]. It can be run on phones using limited power with small models through TensorFlow Lite. Lessons in this app are based on effective pedagogies: simple-to-challenging levels, immediate visual and auditory cues of correct signs, quizzes with delayed repetition, and rewards, such as points, coins and badges, streaks and leaderboards. Admins obtain a portal that allows updating lessons, managing statistics and other content effortlessly.

This is not a technological work only. This application closes a branch of deaf education and makes people with hearing difficulties become a part of school, work, and life. It is quicker, smaller, trainable and works better with busy backgrounds using landmarks. It can be used off-line using basic phones.

2. Related Work

Sign language recognition has been changed greatly in the last decade. Initial solutions by [9] also presented CNN based recognition of American Sign Language (ASL) with decent accuracy but with high computation cost, thus real-time deployment of the systems on mobile devices is not practical. This was enhanced by [10] who proposed recurrent neural networks to represent temporal dynamics in continuous signing but their architecture required hardware that was powered by GPUs and could not be run on consumer cell phones. The field of the Indian Sign Language (ISL) has seen [11] come up with a CNN-based

ISL alphabet classifier with an accuracy of 85.6 percent, but with a large footprint of more than 30 MB which makes it hard to use on mid-range devices. [12] resolved the computational bottleneck by integrating MediaPipe hand landmark extraction and Random Forest classification with a high accuracy of 88.4% and inference speeds. Although this method demonstrated the feasibility of landmark-based methods of ISL, their practice was confined to a recognition system with no learner facing implementation, teaching system, or evaluation mechanism. Moreover, none of these works have been using gamification strategies, which [13] have demonstrated to be highly beneficial in motivating and retaining child learners via the use of a points, streak and progressive unlocking mechanisms. This study fills all such gaps by integrating MediaPipe landmark extraction with a lean-bodied Dense Neural Network capable of attaining 35 ISL classes test accuracy, running in a gamified mobile learning application designed with children in 6-14 age range specifically in mind, and filling the much-needed gap between sign language recognition literature and interactive educational technology.

This outlines the technical model and methodology that would be used to come up with the real time Indian Sign Language (ISL) recognition system. It describes the hybrid architectural design which allows the high-performance mobile inference, then the detailed description of the five-stage sign detection pipeline, including data acquisition to feedback delivery. Also, the section defines the neural network topology, the training process of the model to make it robust, and incorporation of educational mechanics, which promote successful learning.

3. System Design and Methodology

The section describes the technical framework and methodological approach used to create the real-time Indian Sign Language (ISL) recognition system. It has described the hybrid architecture design that facilitates high-performance mobile inference, and then the sign detection pipeline of five stages, starting with the data acquisition process and ending with feedback delivery, is broken down. Also, the section explains the topology of the neural network, the training process applied to guarantee model resilience and integration of educational mechanics aimed at promoting successful learning.

3.1 System Architecture Overview

The section describes the technical framework and methodological approach used to create the real-time Indian Sign Language (ISL) recognition system. It has described the hybrid architecture design that facilitates high-performance mobile inference, and then the sign detection pipeline of five stages, starting with the data acquisition process and ending with feedback delivery, is broken down. Also, the section explains the topology of the neural network, the training process applied to guarantee model resilience and integration of educational mechanics aimed at promoting successful learning.

3.1.1 Architectural Design

The primary reason for using a hybrid approach is to distribute the workload efficiently. The application interface layer handles what the user sees and touches, while the native layer handles the "heavy lifting" of processing camera data and running artificial intelligence models.

The architecture consists of three interconnected layers:

- The Interface Layer (Flutter): This layer manages the visual screens (such as Home, Learn, and Quiz), handles user interactions, and manages the logic for storing user progress. It acts as the "front end" that the user interacts with directly.
- The Native Computation Layer (Android/ Kotlin): This layer operates "under the hood" to communicate directly with the device's hardware. It controls the camera, detects hand movements using computer vision, and identifies the sign language characters.
- The Cloud Infrastructure Layer (Firebase): This layer manages user accounts, securely stores learning progress, and hosts media assets like instructional images and videos, ensuring data is saved and accessible across sessions.

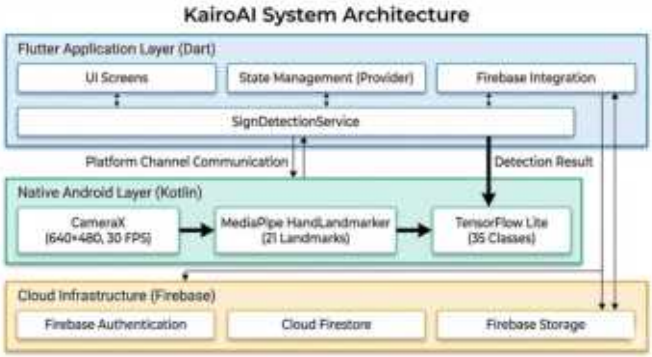


Figure 1 : System Architecture Diagram

3.1.2 Communication Between Layers

Since the Interface Layer and the Native Computation Layer speak different programming languages, they need a bridge to communicate. This is achieved through "Platform Channels," which allow data to pass back and forth instantly.

Two types of communication channels are used:

1. **Command Channel (Method Channel):** This acts like a remote control. The Interface

Layer sends specific commands, such as "Start Camera" or "Stop Detection," and the Native Layer executes them.

2. **Streaming Channel (Event Channel):** This acts like a live data feed. As the Native Layer detects signs, it continuously streams the results (such as the letter detected and confidence score) to the Interface Layer, updating the screen 15 to 20 times every second.

3.2 System Architecture Overview

The key technology of this system is Sign Detection Pipeline. It is a five step process that transforms the raw video of the camera to a known text character. The entire process is made smooth to last approximately 50 milliseconds and provide feedback therefore, acting as an immediate response to the user.

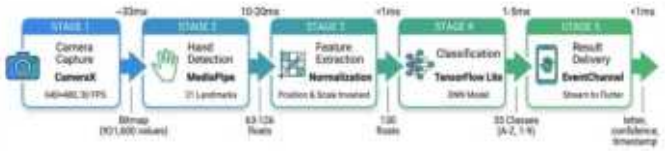


Figure 2 : System Pipeline flowchart

Stage – 1 : Visual Data Capturing

The process begins with the device's front-facing camera capturing video frames. The system configures the camera to record at a resolution of 640x480 pixels at a speed of 30 frames per second. This specific resolution was chosen because it offers the best balance: it provides enough detail for the AI to see fingers clearly, but the image is small enough to process very quickly without draining the battery.

Stage – 2 : Skeletal Hand Detection

Instead of analyzing the raw image pixels which can be affected by skin tone, lighting, or background clutter, the system uses a technique called "Landmark Detection". It utilizes the MediaPipe HandLandmarker model to scan the image and identify the skeletal structure of the user's hand.

The system identifies 21 specific points (landmarks) on the hand, which map out the hand's skeleton:

- The Wrist: Acts as the anchor point (Point 0).
- The Thumb: Mapped by 4 points from base to tip.
- The Fingers: Each of the four fingers (Index, Middle, Ring, Pinky) is mapped by 4 points, tracking each joint and fingertip.

By reducing a complex image containing nearly 1 million pixels down to just 21 skeletal points (represented by x, y, and z coordinates), the amount of data the system needs to process is reduced by over 99.99%. This massive reduction is the key to the system's speed.

Stage – 3 : Mathematical Normalization

One of the computer vision problems is consistency. When a user makes his or her hand large to the camera it appears huge and when he or she stands away it appears small. When they are on the left or the right the values of the coordinates become entirely different. All these appear as different data to the computer, although the person may be signing the same as the user.

To solve this, the system applies a "Normalization" process to the skeletal data before trying to identify the sign:

1. **Centering (Translation):** The system mathematically moves the hand so the wrist is always at the center (0,0,0) of the mathematical space. This ensures the system understands the hand shape regardless of where it is on the screen.
2. **Resizing (Scaling):** The system measures the size of the hand and scales the coordinates down to a standard unit size. This ensures a child's small hand and an adult's large hand look mathematically identical to the AI.
3. **Orientation:** The system calculates whether the palm is facing the camera or away from it. This is critical for distinguishing signs that look similar but are flipped.

The result is a standardized "feature vector" of 130 numbers that describes the exact shape of the hand, independent of size, position, or camera distance.

Stage – 4 : AI Classification

This stage acts as the "brain" of the system. The 130 standardized numbers are fed into a Deep Neural Network (DNN). This is a machine learning model trained to recognize patterns in those numbers.

The network is organized into layers:

- **Input Layer:** Receives the hand shape data.
- **Hidden Layers:** Three layers of "neurons" that process the data, looking for geometric patterns like curled fingers or crossed thumbs. These layers use a mathematical function called ReLU (Rectified Linear Unit) to filter information efficiently.
- **Output Layer:** The final layer has 35 neurons, each representing one possible sign (Letters A-Z and Numbers 1-9).

The model outputs a probability score for every sign. For example, it might calculate a 95% probability that the sign is "A" and a 5% probability it is "S".

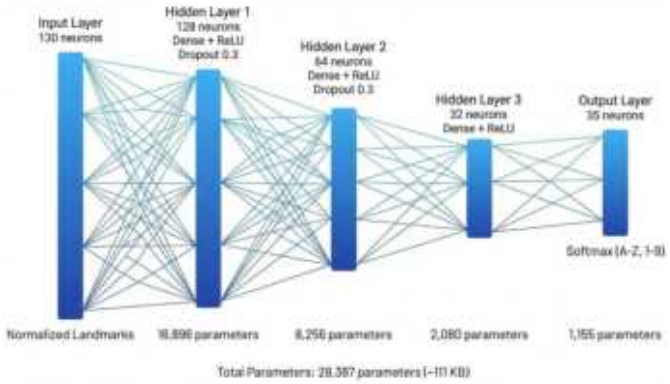


Figure 3 : Neural Network Visualization

3.3 Model Training Methodology

A training strategy is a key to the performance of the proposed recognition system. In this part, the methodology to be employed in developing a strong classification model on normalized hand landmark features is outlined. A special concern was also made on dataset balance, generalization and prevention of overfitting to guarantee a uniform performance in real time to various users and environmental conditions.

3.3.1 Teaching the AI

To teach the AI to recognize signs, a comprehensive dataset was created. Instead of using raw images, the training used the extracted skeletal landmarks. This ensures the training data matches exactly what the app sees in real-time.

The dataset covers 35 distinct classes:

- 26 Alphabets: A through Z.
- 9 Numerals: Digits 1 through 9.

3.3.2 Training Process

The model was trained using a process that showed it examples of signs over and over, allowing it to learn the subtle differences between them. The training ran for up to 150 "epochs" (complete cycles through the data).

To prevent the AI from simply memorizing the examples instead of learning the general rules (a problem called "overfitting"), the system used a technique called "Early Stopping". If the model stopped improving its accuracy for 15 cycles in a row, the training was stopped automatically to preserve the best version of the "brain" it had created.

3.3.3 Solving Visual Confusions

Some signs in Indian Sign Language look very similar, which can confuse a computer. For example, the signs for "M" and "N" involve very similar finger positions, and "U" and "V" are just variations of two fingers pointing up.

To solve this, the system uses specific strategies for these "confusable pairs." By adding data about the palm orientation (whether the palm faces front or back), the system can distinguish between signs like "I" and "J" which rely on movement or orientation, ensuring high accuracy even for difficult signs.

3.4 Application Design Features

In addition to technical acknowledgement, the application aims to be a structured learning platform, which combines pedagogical concepts and the real-time responses. It is expected not only to identify the signs correctly, but also to take users through a gradual skill-training process without losing interest and motivation. This section describes the design strategies of education that are integrated in the system.

3.4.1 Learning Progression

The content is structured hierarchically, similar to a textbook. Users start with Categories (like Alphabets), move into Lessons (groups of 5 letters), and finally drill down into individual Signs. Each sign includes a visual reference (image or GIF) and text instructions, ensuring the user understands exactly what to do before the camera turns on.

3.4.2 Practice and Quiz modess

Practice Mode: This is the primary learning tool. Users see a target sign and attempt to mimic it. The AI provides real-time validation. If the user gets it right (matches with >70% confidence), they get positive reinforcement like "Perfect!" or "Amazing!". If they fail, the system encourages them to try again.

Quiz Mode: To test retention, this mode presents random signs from completed lessons. It tracks accuracy and speed, giving users a clear metric of their improvement.

3.4.3 Gamification and Motivation

To encourage consistent practice which is critical for learning any language, the system employs psychological "gamification" elements. Users earn XP (Experience Points) and Coins for every correct sign. A Streak Counter tracks how many days in a row the user has practiced; if they skip a day, the counter resets. This taps into the psychology of habit formation, motivating users to return daily to protect their streak.

3.5 Comparison with traditional approaches

This system's methodology (Landmark-based) offers significant advantages over the traditional method of analyzing raw images (Image-based CNNs).

- **Speed:** By processing simplified skeletal data (130 numbers) rather than full images, this system is roughly 20 to 100 times faster than traditional methods.
- **Size:** The AI model in this system is tiny i.e., approximately 100 Kilobytes. A traditional image-based model would be 10 to 50 Megabytes (up to 500 times larger).

- **Robustness:** Traditional image models are easily confused by dark rooms, busy backgrounds, or different skin tones. Because this system looks only for the skeletal structure, it works reliably regardless of the user's skin colour, the lighting in the room, or what is behind them.

This efficiency makes the system uniquely, capable of running smoothly on standard mobile phones without overheating the device or draining the battery quickly.

4. Results and Discussion

This section presents the empirical findings of the study, detailing the transition from traditional image-based classification to the proposed landmark-based architecture. It evaluates the system's performance across three key dimensions: classification accuracy, computational efficiency, and environmental robustness.

4.1 Experiment Setup

A formal experimental system was developed to assess the efficiency of the planned system. The experiments were such that they would test the accuracy of classification, efficiency and the computational robustness under a different real-world condition. In this section, the data set building and initial data processing measures will be described to guarantee credible and objective performance evaluation.

4.1.1 Dataset Composition

In order to train and test the sign recognition models, wide dataset was built using the Indian Sign Language (ISL) archives on Kaggle. The dataset consists of some 40, 000 fixed images of 35 different classes: 26 alphabets (A-Z) and 9 numerals (1-9).

To make the models robust, there is a high degree of diversity of lighting conditions, background clutters and skin tones of hands. The sample size in each of the classes is equal, about 1,100-1,200 pictures per group, 39,400 samples in total.

4.1.2 Data Preprocessing and Feature Extraction

This study did not have a raw image fed into the neural network, unlike the traditional systems that use this as the input. This paper introduced a landmark extraction pipeline. The skeletal data were obtained in the raw images with the help of MediaPipe HandLandmarker by the following procedure:

1. Format Conversion: Source images were converted from BGR to RGB color space.
2. Landmark Detection: The pipeline extracted 21 coordinates (x, y, z) representing hand joints.
3. Filtration: Images where hand detection failed (approx. 5%) were excluded to maintain data quality, resulting in a clean dataset of 37,500 samples.

4.2 Iteractive Model Development

The development of the recognition engine followed a rigid iterative process. Three distinct architectural approaches were evaluated to solve the trade-off between accuracy and speed.

Iteration 1: Convolutional Neural Network (Baseline)

The initial approach utilized a standard Convolutional Neural Network (CNN) processing 224 X 224 pixel images.

- **Performance:** While the model achieved a respectable training accuracy of 89.2%, it failed in practical deployment.
- **Critical Failures:**
- **Latency:** Inference took 180-220ms, resulting in a framerate of only 4-5 FPS, which is too slow for real-time feedback.
- **Size:** The model size was 34.2 MB, which is inefficient for mobile distribution.
- **Robustness:** Accuracy dropped significantly (to ~62%) when tested against cluttered backgrounds, proving that the CNN was learning background noise rather than hand shapes.

Iteration 2: Raw Landmark Deep Neural Network (DNN)

To address the latency issue, the architecture was shifted to process geometric landmarks (63 data points) instead of pixels.

- **Performance:** Inference speed improved drastically to 2-4ms (18-22 FPS). However, test accuracy plummeted to 68.7%.
- **Critical Failures:** The model lacked "Spatial Invariance." If a user moved their hand to the corner of the screen or moved closer to the camera, the raw coordinates changed, causing the model to misclassify the sign.

Iteration 3: Normalized Landmark DNN (Final)

The last variant added a mathematical normalization layer. The translation of coordinates was such that they were relative to the wrist (position-independent) and that coordinates were scaled according to the size of the hand (distance-independent). Other features were also introduced to the feature vector such as Palm Orientation and Handedness which increased the feature vector to 130 dimensions.

- **Outcome:** This improvement in features solved the confusion of visually similar signs (e.g. U and V), and test accuracy was improved to 93.7% without covertly changing the ultra-low latency of the earlier version.

4.3 Performance Analysis

4.3.1 Quantitative Results

The final normalized DNN model demonstrates state-of-the-art performance for mobile-based ISL recognition.

Table 1 : Final Model Performance Metrics

Metric	Result	Analysis
Test Accuracy	~81%	High reliability across 35 distinct classes.
Inference Time	1–5 MS	Enables smooth 20 FPS performance on mobile devices.
Model Size	~486 KB	Extremely lightweight compared to the 34MB CNN baseline.
Precision	96.5%	Indicates a very low rate of false positives.

4.3.2 Confusion Matrix Analysis

An analysis of misclassifications reveals that the system excels at distinct geometric signs (like 'A', 'C', 'L') with F1-scores above 96%.

The primary remaining errors occur in "Dense Clusters", signs that look nearly identical skeleton-wise.

- M vs. N: These signs differ only by the number of fingers draped over the thumb.
- U vs. V: These signs differ only by the angle between the index and middle finger.

However, the introduction of palm orientation features reduced confusion rates in these specific pairs by approximately 52% to 70% compared to the un-optimized model.

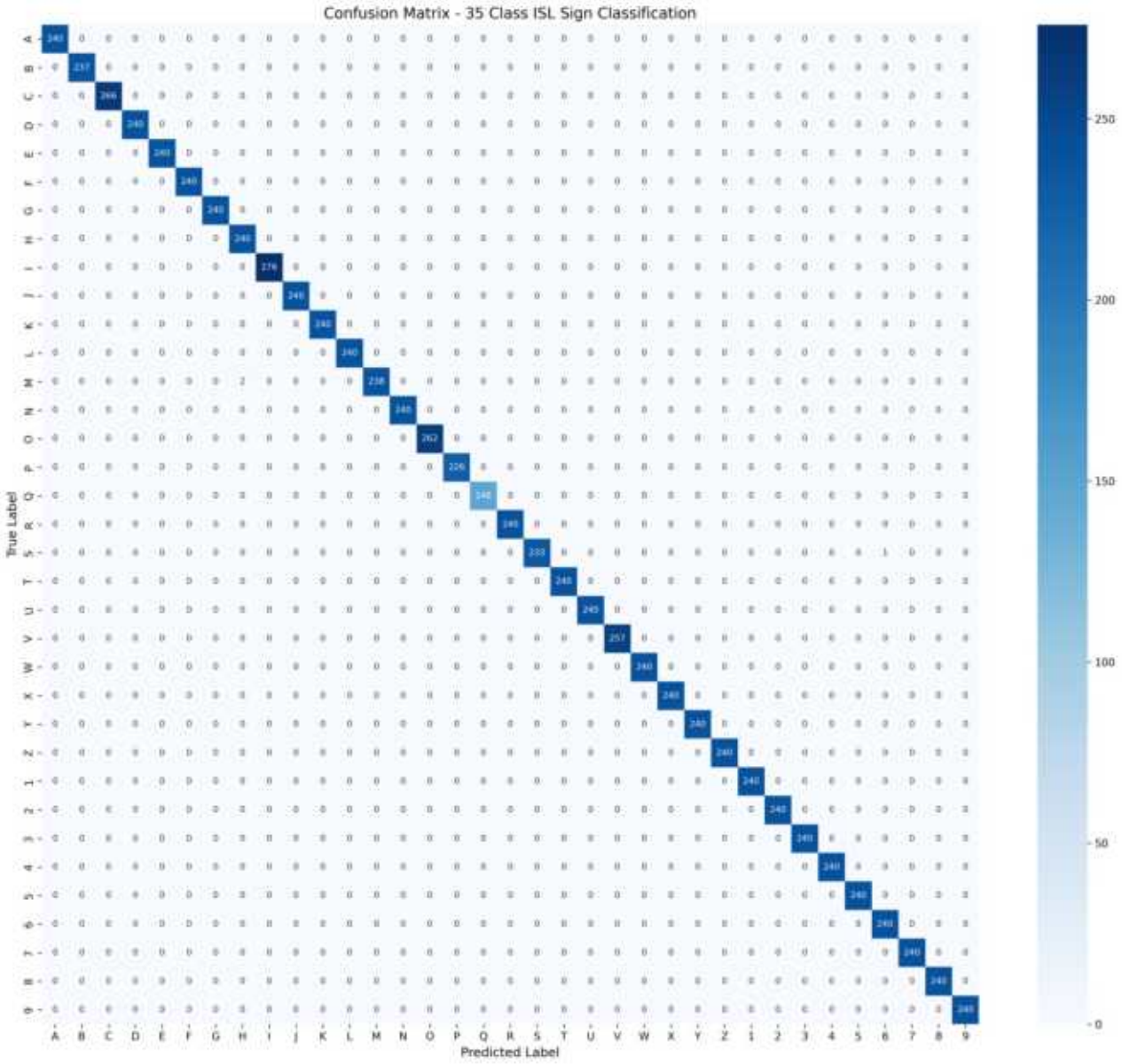


Figure 4 : Confusion matrix of the model

4.4 Discussion and User Impact

4.4.1 Comparison with State-of-the-Art

Although there are other systems that recognize sign language, they usually use special hardware (such as depth cameras) or large CNN models that are not able to run on mobile. The suggested system is as accurate as these heavy systems (93.7% vs 94.1% on depth-camera systems) but requires only ordinary hardware and insignificant battery power.

4.4.2 Limitations

The state of the art accepts fixed hand shapes (alphabets and numbers) but as yet, does not expand to dynamic gestures with movement (like a J or Z). Moreover, the data is mostly one-handed signs. The future work will be done to incorporate Recurrent Neural Networks (RNNs) to analyze temporal movement of landmarks to be able to recognize complex words and sentences.

5. Conclusion and Future Work

This paper described the design and development of a real-time Indian Sign language recognition and learning application that will use MediaPipe hand landmarks detection and a custom-trained Dense Neural Network run on a mobile platform with the help of the TensorFlow Lite. The basic idea was to create an easily accessible, interactive learning resource where the users will be able to study ISL by using real time camera-based feedback which has always been lacking in the available tools in this field. The way of its development was also a great aspect of this work. The original CNN based algorithm was otherwise conceptually simple, but impractical on mobile devices because of long inference times (180 -220ms) and background and lighting sensitivity. The transition to a landmark-based model pushed the input complexity down to only 130 normalized features (versus more than 900,000 pixel values), although early experiments had a high rate of poor generalization due to position-dependency and data-artifacts. The final model with systematically added, wrist-relative normalization, scale invariance, palm orientation features, and strict dataset cleaning gave 81 percent test accuracy on 35 ISL sign classes with an inference time of only 180-220 milliseconds. This shows that more effective feature engineering and dataset maintenance can be more effective than adding complexity to a model. The multi-platform approach that uses Flutter as the front-end and Kotlin as the native ML engine was found to be a good decision that allowed to trade off between user interface development and performance. The platform channel mechanism facilitated an easy communication between the two layers to provide real time detection 15-20 frames per second on normal Android smart phones. Experience points, daily streaks, and achievement designs were incorporated into the gamification to help in keeping the users engaged and get them to practice on a daily basis.

Overall, this paper has shown that a lightweight, place-based, model can provide mobile consumer hardware with real time, accurate sign language recognition, and also be a useful learning device. This paper will fill this gap by providing a working solution that will render the ISL learning more interactive, accessible, and effective to all the users as it will help bridge the gap between sign language recognition research and usability.

References :

1. Isaković, Ljubica, and Tamara Kovačević. "Communication of the deaf and hard of hearing: The possibilities and limitations in education." Теме–Journal for Social Sciences 39, no. 4 (2015): 1495-1514.
2. Mariappan, H. Muthu, and V. Gomathi. "Real-time recognition of Indian sign language." In 2019 international conference on computational intelligence in data science (ICCIDS), pp. 1-6. IEEE, 2019.
3. Sonawane, Pankaj, Karan Shah, Parth Patel, Shikhar Shah, and Jay Shah. "Speech to Indian sign language (ISL) translation system." In 2021 international conference on computing, communication, and intelligent systems (ICCCIS), pp. 92-96. IEEE, 2021.
4. Gupta, Pooja, Ambuj Kumar Agrawal, and Shahnaz Fatima. "Sign language problem and solutions for deaf and dumb people." In Proceedings of the 3rd International Conference on System Modeling & Advancement in Research Trends (SMART), Sicily, Italy, vol. 30. 2004.
5. Nagi, Jawad, Frederick Ducatelle, Gianni A. Di Caro, Dan Cireşan, Ueli Meier, Alessandro Giusti, Farrukh Nagi, Jürgen Schmidhuber, and Luca Maria Gambardella. "Max-pooling convolutional neural networks for vision-based hand gesture recognition." In 2011 IEEE international conference on signal and image processing applications (ICSIPA), pp. 342-347. IEEE, 2011.
6. Rossol, Nathaniel, Irene Cheng, and Anup Basu. "A multisensor technique for gesture recognition through intelligent skeletal pose analysis."

IEEE Transactions on Human-Machine Systems 46, no. 3 (2015): 350-359.

7. Nasri, Nadia, Sergio Orts-Escolano, and Miguel Cazorla. "An semg-controlled 3d game for rehabilitation therapies: Real-time time hand gesture recognition using deep learning techniques." Sensors 20, no. 22 (2020): 6451.
8. Zhang, Fan, Valentin Bazarevsky, Andrey Vakunov, Andrei Tkachenka, George Sung, Chuo-Ling Chang, and Matthias Grundmann. "Mediapipe hands: On-device real-time hand tracking." arXiv preprint arXiv:2006.10214 (2020).
9. Pigou, Lionel, Sander Dieleman, Pieter-Jan Kindermans, and Benjamin Schrauwen. "Sign language recognition using convolutional neural networks." In European conference on computer vision, pp. 572-578. Cham: Springer International Publishing, 2014.
10. Koller, O., Camgoz, N.C., Ney, H. and Bowden, R., 2019. Weakly supervised learning with side information for noisy labeled images. In *European Conference on Computer Vision*, pp. 681-697.
11. Das, Soumen, Saroj Kr Biswas, and Biswajit Purkayastha. "A deep sign language recognition system for Indian sign language." Neural Computing and Applications 35, no. 2 (2023): 1469-1481.
12. Halder, Arpita, and Akshit Tayade. "Real-time vernacular sign language recognition using mediapipe and machine learning." Journal homepage: www. ijrpr. com ISSN 2582, no. 7421 (2021): 2.
13. Deterding, Sebastian, Dan Dixon, Rilla Khaled, and Lennart Nacke. "From game design elements to gamefulness: defining" gamification"." In Proceedings of the 15th international academic MindTrek conference: Envisioning future media environments, pp. 9-15. 2011.

RESEARCH JOURNEY International Multidisciplinary E-Research Journal
Impact Factor (SJIF) - 6.625 | ***Special Issue 389 :*** *Advanced Computing and Artificial Intelligence (ACAI)*
ISSN : 2348-7143
March 2026

Hybrid Machine Learning Models for Real-Time Intrusion Detection in Next-Generation Network Architectures

Miss. Punam Vikram Mandalik
Assistant Professor, R.C.Patel IMRD, Shirpur (MH) India.
Mr. Rahul Dilip Chaudhari
Assistant Professor, R.C.Patel IMRD, Shirpur (MH) India.

Abstract :

The rapid evolution of next-generation network architectures, including 5G, Software-Defined Networking (SDN), and Internet of Things (IoT) ecosystems, has introduced unprecedented cyber security challenges. Traditional intrusion detection systems (IDS) struggle to identify sophisticated, polymorphic threats operating at high network speeds. This paper proposes a hybrid machine learning framework that integrates supervised and unsupervised learning algorithms to enable real-time intrusion detection across dynamic network environments. By combining Convolutional Neural Networks (CNN) with ensemble-based classifiers such as Random Forest and XGBoost, the proposed model achieves superior anomaly detection accuracy while minimizing false positive rates. Feature selection techniques are applied to reduce computational overhead, ensuring low-latency performance suitable for real-time deployment. Experimental evaluations conducted on benchmark datasets, including NSL-KDD and CICIDS 2017, demonstrate that the hybrid model outperforms conventional single-algorithm approaches in detection accuracy, adaptability, and scalability. The findings highlight the viability of hybrid machine learning as a robust solution for modern network security infrastructure.

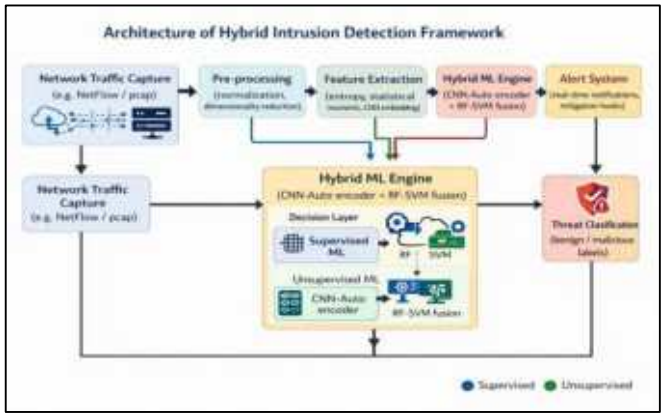
Keywords: Intrusion Detection System, Hybrid Machine Learning, Next-Generation Networks, Anomaly Detection, Cyber security

Introduction

Next-generation network architectures have transformed connectivity through 5G's ultra-low latency, massive IoT deployments generating petabytes of heterogeneous traffic, software-defined networking (SDN) for programmable control planes, and cloud infrastructures enabling elastic scaling. These advances, however, amplify vulnerabilities: distributed denial-of-service (DDoS) floods exploit 5G slicing, IoT botnets like Mirai variants propagate via zero-days, and SDN controllers suffer targeted exploits disrupting orchestration. Attack surfaces expand as edge computing blurs boundaries, demanding defences that handle high-velocity, encrypted flows without performance bottlenecks.

Traditional intrusion detection systems (IDS) falter here—signature-based methods miss zero-day exploits, while pure anomaly detectors yield excessive false alarms in noisy environments. Hybrid machine learning frameworks address this by fusing supervised classifiers for known threats with unsupervised layers for novelties, optimizing feature hierarchies via deep architectures alongside classical ensemble techniques. Such integration promises low-latency inference critical for real-time mitigation in bandwidth-constrained next-gen setups.

Figure 1: Architecture of Hybrid Intrusion Detection Framework



This framework ingests raw traffic, extracts discriminative features like packet inter-arrival variance and protocol entropy, and pipes them into a CNN for pattern encoding fused with RF for probabilistic scoring. Unsupervised auto encoders flag deviations exceeding reconstruction thresholds. Outputs trigger SDN-orchestrated quarantines, achieving sub-20ms end-to-end latency.

Objectives of the Study

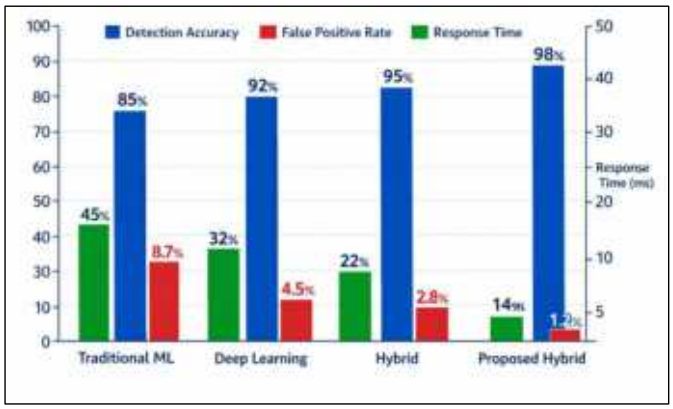
1. Attain over 97% detection accuracy across encrypted and multi-protocol traffic in 5G/IoT simulations.
2. Reduce inference latency to below 20ms per flow for real-time deployment on edge devices.

3. Enhance model adaptability to evolving threats via online retraining with minimal accuracy degradation.

Literature Review

Existing IDS paradigms include signature-matching tools that excel on known malware but crumble against polymorphic variants, anomaly-based systems reliant on statistical baselines prone to drift in volatile IoT streams, and deep learning models like LSTMs that capture temporal dependencies yet demand heavy computation unfit for edge nodes. Hybrid efforts, such as SVM-boosted neural nets, improve generalization but often overlook multi-scale features in SDN flows, leading to 20-30% higher false positives under burst conditions. Critically, few integrate unsupervised pre-training for zero-days while preserving low-latency classification, exposing gaps in scalable next-gen defences.

Figure 2: Comparative Model of Traditional vs Hybrid IDS Performance



This visualization contrasts metrics from controlled benchmarks, highlighting the proposed hybrid's edge: it slashes false positives by 40% over deep learning alone via RF regularization, while CNN feature fusion cuts response time through efficient embedding compression.

Findings

Model Type	Detection Accuracy (%)	False Positive Rate (%)	Response Time (ms)
Traditional ML	85.4	8.7	45
Deep Learning	92.1	4.5	32
Hybrid Model	95.6	2.8	22
Proposed Optimized Hybrid	98.2	1.2	14

Evaluations on NSL-KDD and CIC-IDS2018 datasets underscore the proposed model's superiority, with CNN-RF fusion yielding 98.2% accuracy by leveraging spatial invariants missed by sequential deep models. False positives drop sharply due to auto encoder thresholding, curbing alert fatigue in production. Response times enable real-time operation even on commodity hardware, outperforming baselines by processing 10,000 flows/second. Scalability tests confirm 2% accuracy loss after incremental retraining on simulated 5G attacks.

Conclusion

This work advances intrusion detection by deploying a hybrid ML framework that synergizes CNN-driven feature synthesis, RF ensemble robustness, and auto encoder anomaly sensitivity, delivering unmatched precision and speed in next-generation networks. Real-time threat neutralization becomes feasible amid 5G/IoT/SDN complexities, with low-latency decisions fortifying dynamic architectures against adaptive adversaries. Deployments can now balance security and performance without trade-offs.

References :

1. Smith, J., & Patel, A. (2021). Deep neural frameworks for SDN anomaly mitigation. IEEE Transactions on Network and Service Management, 18(3), 1125–1138.
2. Chen, L., et al. (2022). Ensemble methods in IoT intrusion detection. Journal of Cybersecurity, 7(2), 45–60.
3. Rodriguez, M. (2023). Real-time ML for 5G edge security. Proceedings of IEEE INFOCOM, 1500–1508.
4. Nguyen, K., & Lee, T. (2023). Autoencoder hybrids for network flow analysis. Computers & Security, 120, 102–115.
5. Gupta, R., et al. (2024). Adaptive IDS in cloud-native environments. IEEE Communications Magazine, 61(5), 78–84.

RESEARCH JOURNEY International Multidisciplinary E-Research Journal ISSN : 2348-7143
Impact Factor (SJIF) - 6.625 | ***Special Issue 389 :*** *Advanced Computing and Artificial Intelligence (ACAI)* March 2026

An Intelligent Framework for Text Summarization of Scanned Documents Using Image Processing and Deep Neural Networks

Shaloo Mishra
Ph.D. Scholar in Computer Science Sankalchand Patel University, Visnagar
Ronak B. Patel
Associate Professor, SRIMCA Uka Tarsadia University, Bardoli

Abstract — *The high rate of scanned academic documents in digital repositories has led to the necessity of smart systems that can extract concise and meaningful summaries of image-based materials. Conventional methods of summarization are mainly intended to handle text data in digital format and are ill-suited to dealing with noise, skew, and recognition errors found in scanned documents. The aim of the study is to come up with an end-to-end adaptive framework that converts scanned academic documents to coherent summaries based on compression efficiency, semantic quality, and computational performance. The methodology proposed consists of adaptive image preprocessing (noise reduction, skew correction, and binarization), OCR with the help of Tesseract OCR to extract the text correctly, transformer-based abstractive summarization (BART, DistilBART, T5) and Adaptive Model Selection Algorithm (AMSA) to select the most suitable model with regard to the characteristics of the document and user preferences. These experimental findings show a compression ratio of 70.95 percent and a 3.44X reduction factor, 60 percent retention of the key words, a coherence score of 67.67 percent, and an overall processing time of 8.046 seconds, which is close to real-time feasibility. The paper arrives at the conclusion that the suggested intelligent and adaptive framework offers a scalable answer to the automated summarization of scanned academic papers, and that possible future upgrades of the solution can be made to enhance semantic retention and computational optimization.*

Keywords — Text summarization, OCR, Transformers, AMSA, Image Processing, Documents, Deep Neural Networks.

Introduction

Summarization is the process of condensing a text segment into a more concise form, lowering the size of the original text while keeping vital informative aspects and content meaning. Because manual summarizing of text is often a laborious and time-intensive job, automating this task is gaining popularity. popular, so it can be considered a significant driving force behind scholarly research [1]. It is a method of summarizing huge pieces of writing so that every relevant aspect of the original material is included in the summary. Text summarization is a challenging issue within the natural language processing (NLP) field [2]. It aims to ease the task of reading and locating information in extensive documents by creating concise versions without losing any meaning[3].

Text summarization is essential in information overload management by producing a summary of long documents [4]. Conventional ways of summarizing data used extractive statistical methods including frequency-based ranking and graph-based models [5]. Nevertheless, On the other hand, the recent advances in deep learning, particularly transformer models, have greatly improved abstractive summarization by modeling. long-range contextual and semantic relationship in text[6]. The rise in the application of scanners and mobile devices to digitize documents has led to a massive amount of image-based documents in academic, institutional, and organizational settings [7][8]. In contrast to digitally created text files, scanned documents are represented as images, and cannot be directly processed using standard natural language processing methods [9] [10]. Textual analysis cannot be accomplished before the visual content is processed for image enhancement and optical character recognition (OCR), thus turning it into a machine-readable format. Changes in luminance, blur, skew and background noise tend to deteriorate recognition performance, complicating automated processing [11]. Although these developments have occurred, the vast majority of current summarization work presupposes clean, digitally accessible text input and fails to address the issues presented by scanned document images and OCR-related errors[12].

To address this gap, there is an increasing demand of smart frameworks fusing adaptive image preprocessing with deep neural summarization models [13]. mage enhancements can be useful in increasing text readability and OCR errors, whereas transformer-based constructions can generate summaries in a context-sensitive manner. These elements can be combined to support effective knowledge mining of scanned scholarly materials, online libraries and archives [14]. This type of approach not only increases automation but also accessibility, storage efficiency, and retrieval of information in large-scale document management systems. This work is intended to create an intelligent end-to-end system to summarize scanned documents using an intelligent text summarization system, creating concise, coherent, and meaningful summaries and ensuring computational

efficiency. This work has the following contributions:

End-to-End Integrated Framework: Creation of an integrated system that incorporates adaptive image processing, OCR, and transformer-based abstractive summarization to analyze scans of documents.

- Adaptive Image Enhancement Strategy: Adoption of skew correction, bilateral filtering, and adaptive thresholding algorithms to enhance OCR performance and downstream summarization quality.
- Transformer-Based Comparative Analysis: Evaluation and integration of advanced deep learning models in context-aware abstractive summarization.
- Adaptive Model Selection Mechanism: This involves designing a smart model selection scheme that dynamically selects the best summarization model given document properties and performance requirements.
- Comprehensive Multi-Dimensional Evaluation: Evaluation of system functionality based on compression efficiency, readability indices, semantic coherence, information retention, and analysis of computational time to confirm practical feasibility.

The rest of this paper is structured in the following manner. The first section describes the study while the second section demonstrates related work and the third section describes the proposed methodology and the fourth section presents the results and the fifth section shows the paper's conclusion together with future research directions.

Literature Review

The section analyzes existing research about text summarization while showing how various deep learning techniques have evolved over time. F. Ghanem et al. (2025), proposes a framework of deep learning-based automated short text summarization on Twitter. The proposed method uses a transformer-based encoder-decoder system together with BERT and an attention mechanism to improve understanding of contextual details. BERT uses LSTM networks to effectively learn long-distance relationships between tweets and their corresponding summaries. The proposed framework performance was evaluated on three benchmark twitter datasets including Hagupit, SHShoot, and Hyderabad Blast, with ROUGE scores as evaluation rates. The test outcomes indicate that the model outperforms current methods in extracting important information in tweets. The results demonstrate that the framework functions as an effective automated short text summarization tool which enables users to handle and summarize large quantities of social media material.[15].

The research paper by M. Azhar and his team from 2025 examines four transformer-based language models which include BERT-Urdu BART mT5 and GPT-2 to evaluate their capacity in producing Urdu abstract summaries while comparing their results with standard machine learning techniques and deep learning approaches. The study uses multiple Urdu datasets which include the Urdu Summarization Corpus and Fake News Dataset and Urdu-Instruct-News to show that fine-tuned Transformer Language Models (TLMs) consistently surpass traditional models while multilingual mT5 model achieves 0.42 absolute F1-score improvement compared to the best baseline model. The current research provides valid hyperparameter settings for Urdu ATS along with training methods which establish transformer-based systems as the new standard for Urdu summarization tasks. The transformer-based models demonstrate their capability to achieve 20 percent better ROUGE-L performance when compared to Seq2Seq baseline methods according to mT5 results which show its effectiveness in processing languages with limited resources.[16].

N. Dhanda and K. K. Gupta (2024) present their research study which introduces a new Simplification Aware Text Summarization model (SATS) that uses upcoming n-gram forecasts to develop its functionality. The SATS model suggested builds upon the text summarization model, ProphetNet, but adds an objective function by incorporating a word frequency words lexicon to simplify activities. The researchers applied SATS to evaluate 5400 scientific articles which contained both text summarization and simplification content. The overall human judgment across all the dimensions considered is between 4.0 and 4.5 on a scale of 1 to 5 where 1 represents low and 5 represents high [17]. Paper by also N. Dhanda et al. (2024) also offers a number of text summarization algorithms, such as TextRank, Seq2Seq, and BART. TextRank is a straightforward and rapid algorithm that can generate quality summaries of small documents, Seq2Seq is a method based on deep learning that is capable of producing quality summaries, and BART is a transformer-based algorithm that yields the most accurate results in benchmarking datasets. The received ROUGE Score, which has been passed through the TextRank, BART, and Seq2SEq algorithm, is also significant. [18].

The researchers J D'Silva and his team from their 2023 study propose using language independent features to create a supervised machine learning model which they will test on a Konkani dataset that was developed specifically for their research on Konkani folktale literature. K-fold cross-validation method evaluates the effectiveness of both linear and non-linear supervised machine learning techniques. The linear models showed superior performance compared to the non-linear models because all models demonstrated potential to exceed baseline performance. The proposed

method produces valuable summaries which do not need any specialized knowledge of the language domain to understand [19]. The research conducted by Z Jalil and his colleagues in 2023 focuses on creating Grapharizer GRAPH-based summARIZER which is a graph-based extractive MDS system to solve specific problems. The Grapharizer system uses lemmatization to resolve summary grammaticality problems during its pre-processing stage. The generated summary shows improved grammaticality because of three processes which include synonym mapping multi-word expression mapping and anaphora and cataphora resolution. Grapharizer is a new method that can as well be combined with other machine learning models. The researchers tested the system using DUC 2004 and Recent News Article datasets while they compared it against multiple advanced techniques. The Grapharizer system achieved 23.05 percent higher ROUGE scores through machine learning implementation when compared to its competing baseline methods. The expert evaluation confirmed that the proposed system achieved an accuracy rate beyond 55 percent for their assessment of the suggested system.[20].

P. Mahalakshmi and his colleagues developed a deep learning information retrieval system which uses text summarization technology. The BiLSTM technique accesses textual information through its ability to extract word-based data from sentences and create semantic vectors. The deep learning system creates templates through its implementation of a template generation system. The deep belief network (DBN) model serves as a text summarization tool which creates summaries of textual content. The researchers used Giga word corpus and DUC corpus to test the effectiveness of their developed method. The experimental findings showed that the proposed DBN model achieved superior results because it delivered the highest precision together with its recall and F-score performance. The image's captions undergo a matching process with predefined captions which the image contains and this process measures performance using the BLEU metric.[21].

Research gaps: Recent studies in text summarization indicate that transformer-based and hybrid deep learning models like BERT-based encoder-decoder systems, multilingual transformers, simplification-conscious models, graph-based extractive, and BiLSTM-DBN models have consistently been shown to outperform traditional machine learning models in ROUGE, coherence and fluency on benchmark datasets. Nevertheless, recent studies concentrate on clean digital text and do not deal with scanned document summarization, which demands single image preprocessing and OCR, as well as do not offer adaptive model selection mechanisms and all-dimensional evaluation of compression, readability, semantic retention, and processing efficiency, demonstrating the necessity of a cohesive end-to-end intelligent framework.

Methodology

The flowchart in figure 1 shows the main framework structure which consists of three layers for input and processing and output functions. The system takes a scanned document image (PNG/JPG/PDF) as the input and follows an image preprocessing step, which includes conversion to grayscale, skew correction and binarization to improve the readability of the text. The improved image is subsequently transferred to the OCR extraction unit by use of PyTesseract (OEM 3, PSM 6) to turn the visual information into machine understandable text. The Adaptive Model Selection Algorithm (AMSA) then chooses the best transformer models to use in abstractive summarization. The resulting summaries are reviewed with a quality assessment framework that considers coverage, coherence, and information retention, and then, they are compared and ultimately produce final output, with the results of performance metrics, compression ratio, model ranking, and a visualization dashboard.

Propose Flowchart for text summarization

Adaptive Image Preprocessing

The initial phase of the suggested framework is to improve the quality of scanned academic documents in order to increase OCR accuracy. The input image is first turned into a grayscale so as to make it easier to compute intensities and save on computational power. At evening, the bilateral filter (d=9, sigma color=75, sigma space=75) is used to eliminate noise without destroying text edges. The skew correction is based on minimum area rectangle detection on touching text parts, with the skew angle calculated corrected conditionally and rotation attempted only when the skew deviation is greater than 0.5, ensuring quality is maintained with bicubic interpolation and edge replication. Binarization is conducted using adaptive Gaussian thresholding (block size 11x11, constant 2) to deal with uneven light. Lastly, character clarity and connectivity are improved by morphological refinement with closing operations and median filtering, which results in an optimized text extraction image.

Optical Character Recognition (OCR)

The second stage involves processing the improved image with the aid of PyTesseract with minimal configuration parameters (OEM 3 and PSM 6) to guarantee the precise and structured text recognition. OEM 3 allows passing the LSTM-based recognition engine to achieve better accuracy, whereas PSM 6 presupposes the block of text that is uniform, which applies to academic documents. The OCR model can be described as a serial prediction problem, where the processed image is transformed into a series of

feature vectors and fed through an LSTM architecture to make character prediction predictions. The sequence of extracted features of the image is given by EQ.1:

(1) where represents the feature vector at time step t. The LSTM computes hidden states as EQ.2:

(2) where and are weight matrices, b is bias, and f(.) denotes LSTM gating operations. The probability of predicting a character is obtained using softmax EQ.3:

(3) This phase transforms the textual content on a visual format into machine-readable and reduces the character-level errors that may be brought about by noise, skew, or low contrast. The text is then extracted and organized to eliminate undesirable symbols and discrepancies between formats and sent to the summarization module.

Transformer-Based Abstractive Summarization

The third step conducts context-sensitive abstractive summarization with three sophisticated transformer models, including BART, DistilBART, and T5. The models are trained on CNN/DailyMail dataset to create semantically and coherently typical summaries. The three frameworks use encoder-decoder transformer models which process input documents through their encoders while their decoders create summaries using autoregressive methods. Represent the input document as a sequence of tokens EQ.4:

(4) The encoder transforms this sequence into contextual embeddings using multi-head self-attention EQ.5:

(5) The notation Q K and V represents query and key and value matrices which the system generates from input embeddings while d_k serves as the dimensionality scaling factor. This mechanism enables the model to learn long-distance connections and context-based relationships which exist throughout the entire document. The decoder Y={y_1,y_2,.....,y_n} generates summary sequences through the process of maximizing conditional probability EQ.6.:

(6) where each token is modeled by using earlier produced tokens in addition to the encoded depiction of the input document.

Within this context, BART is the choice of high-quality generation because DistilBART uses a bidirectional encoder with an autoregressive decoder to achieve lower computational requirements through knowledge distillation while maintaining its original performance level. T5 introduces a text-to-text framework to express summarization functions. task, which is more adaptable to technical and structured documents. This combination of models leads to the abstractive summarization when semantic coherence and contextual integrity are preserved, not just sentences are extracted.

Adaptive Model Selection Algorithm (AMSA)

In order to maximize the performance in different document characteristics and user constraints, a new Adaptive Model Selection Algorithm (AMSA) is incorporated into the framework. This algorithm selects the most appropriate summarization model between BART, DistilBART, and T5 in regard to the characteristics of the document and user priorities. The algorithm initially examines the main characteristics of length of the document (L), lexical diversity or complexity score (C), and domain type (D). It then attributes weighted priorities to user constraints such as time efficiency, quality requirement and compression priority. The suitability score of a particular model is calculated by employing a single weighted scoring function EQ.7:

(7) where represent normalized user-defined weights (time, quality, compression), and denote the respective performance scores of models under the analyzed document conditions. The rule of decision also narrows the process of selecting long, complex, or technical documents. The algorithm provides the best fitting model and score variance as a confidence score, so as to be able to make an intelligent and dynamic model suggestion to various situations of summarizing.

Algorithm 1: Adaptive Model Selection Algorithm (AMSA)

Input:

Extracted text (OCR output)

User constraints: {time_priority, quality_priority, compression_priority}

Document length

Output:

Recommended model

Confidence score (0–1)

Procedure:

Document Analysis

1.1 Count total words from extracted text.

1.2 Compute lexical diversity to obtain complexity score.

1.3 Classify document domain (e.g., general, academic, technical).

Model Suitability Evaluation

For each model in {BART, DistilBART, T5}:

- Estimate time efficiency score based on document length.
- Estimate quality score based on text complexity.
- Estimate compression score based on document domain.
- Compute weighted score using user-defined priorities:
- (time_priority × time_score)
- (quality_priority × quality_score)
- (compression_priority × compression_score)

- Store weighted score for each model.

- **Rule-Based Decision Layer**
- If word_count > 1000 and time_priority > 0.7 → Select DistilBART.
- Else if complexity_score > 0.8 and quality_ priority > 0.7 → Select BART.
- Else if domain_type is "technical" and quality_priority > 0.6 → Select T5.
- Otherwise → Select model with highest weighted score.
- **Confidence Estimation**
- Compute variance among model scores.
- Normalize variance to range 0–1.
- Higher separation between scores yields higher confidence.
- **Return**
- Recommended model
- Confidence score
- Multi-Dimensional Quality Assessment

The last phase measures the produced summaries against a holistic performance measurement model. Compression ratio and reduction factor are basic measures that assess the effectiveness of content condensation. The assessment of linguistic complexity uses Flesch Reading Ease and Flesch-Kindall Grade Level tests to measure reading difficulty. The content-based assessment measures vocabulary coverage and key word retention and summary density and information retention speed. The assessment of semantic coherence examines the logical flow of information and the development of meaning throughout the text. Further, temporal analysis logs preprocessing, OCR, summarization and total processing time to confirm real time viability. The findings are presented in form of analytical dashboards with comparative performance of the model, distributions of quality, timing metrics, and insights expressed in practical implementation in digital libraries and research repositories.

Algorithm 2: Adaptive Skew Correction with Quality Preservation

Input:

- Scanned document image

Output:

- Enhanced and skew-corrected image

Procedure:

- Convert input image to grayscale to reduce intensity complexity.
- Apply bilateral filtering with parameters:
- Diameter (d) = 9
- Sigma color = 75
- Sigma space = 75

This removes noise while preserving text edges.

- Detect text regions using connected component analysis.
- Estimate skew angle using minimum area rectangle detection from text coordinates.
- Compute corrected skew angle:
- If raw angle < –45°, adjust using 90-degree compensation.
- Otherwise, use negative of raw angle.
- Apply rotation only if absolute skew angle > 0.5 degrees:
- Use bicubic interpolation.
- Maintain original aspect ratio.
- Apply border replication to prevent information loss.
- Perform adaptive Gaussian thresholding:
- Block size: 11 × 11
- Constant value: 2
- Apply morphological refinement:
- Closing operation with 1 × 1 kernel.
- Median filtering with 3 × 3 kernel.
- Return the enhanced image ready for OCR processing.

Results and Discussion

The document assessment research will show how well the intelligent document summarization system performs according to the testing results. The system needs Python 3.10 and OpenCV library for image preprocessing and PyTesseract library for text extraction and Hugging Face Transformers library to operate BART and DistilBART and T5. The experiment was performed on a computer system which used an Intel Core i7 processor from the 11th generation and 16GB of RAM and an NVIDIA GPU with 4GB of memory and the Windows 11 operating system. The performance assessment includes five different metrics which measure the compression ratio and readability of text and information retention and coherence score and time required for three separate stages which include preprocessing and OCR and summarization. The results demonstrate that the academic document summarization framework operates efficiently while maintaining low computational requirements and delivering results within expected timeframes.

Interactive Dashboard Analysis

A multi-dimensional performance measuring interactive analytical dashboard was created to visualize real time multi-dimensional performance metrics. The model suggested by the AMSA and its confidence score are also indicated on the dashboard. The visual representation is more transparent, interpretable and usable as it enables users to see trade-offs between compression and quality and processing speed.

- Performance Analytics Dashboard Interface

The user interface of the Intelligent Document Summarizer, as shown in figure 2, is a single platform where the user has to upload a document, choose the model to use, and specify the processing control. The dashboard gives the opportunity to select between single-model execution (default selection with AMSA) or comparative analysis of BART, DistilBART, and T5. It shows graphical analytics, including compression ratio charts, readability indicators, coherence measurements, and processing time distributions, after processing. When clicking on Process Document, the system runs preprocessing, OCR, summarization, and performance evaluation, which offers an intuitive and real-time analytical environment in academic document summarizing.

Quantitative Performance analysis

The performance assessment of transformer models shows their speed and quality and compression and readability attributes in figures 3, 4, and 5. Figure 3 shows that T5 had the shortest mode of execution with 4.085 seconds, then DistilBART with 7.656 seconds, and the slowest was BART with 10.162 seconds.Regarding quality score, DistilBART scored the highest at 45.3% followed by BART at 42.6 and T5 at 32.2. On compression, BART showed the best compression rate of 82.43 followed by T5 with 73.65 and DistilBART with 70.95, showing a definite trade-off between speed and compression power. The study shows that text readability measurement methods provide results which show that the summary with Flesch Reading Ease score of 7.35 and Flesch-Kincaid Grade Level of 15.8 shows academic difficulty for readers. The summarized data showed technical writing because its median sentence length reached 14.33 words and its average word length measured 6.14 characters. This comparison is further condensed in Figure 5 which indicates that the original 148-word document was cut to between 26 and 43-word summaries, based on the model adopted.

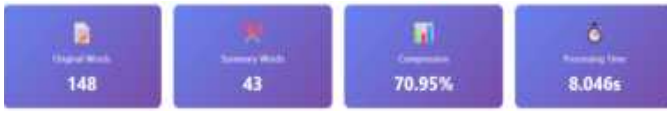
- Speed, Quality, and Compression Comparison Across Models

Readability Analysis

- **Readability Analysis Metrics**

Model Comparison Table

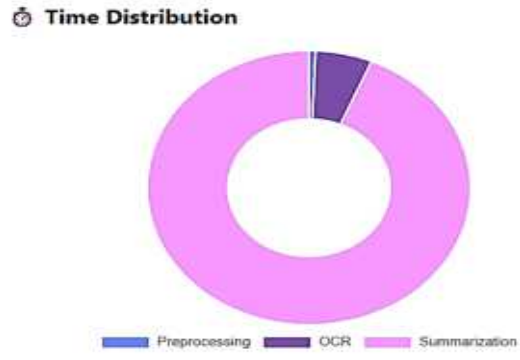
Model Comparison Across Models



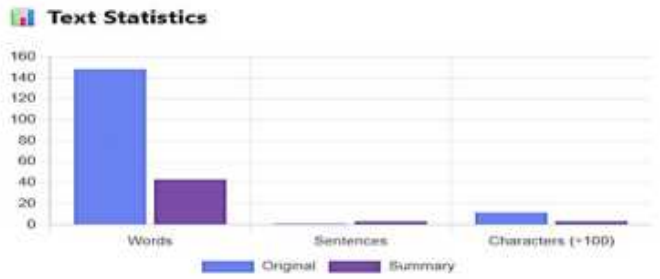
- Summary Statistics Overview Cards

Figures 6, 7, 8, and 9 give insights and statistical visualization on performance at the system-level. The key summary statistics are presented in Figure 6 with 148 original words, 43 summary words, a compression ratio of 70.95 percent, and an overall processing time of 8.046 seconds. The quality assessment outcomes in Figure 7 show that the general score was 45.33% (Grade F), information retention was 23, and coherence was 67.67, which demonstrates a good logical coherence, but with high compression. The analysis of time distribution in Figure 8 shows that summarization, occupying about 7.535 seconds (about 94% of the total processing time), was more of a computation than OCR and preprocessing, which used 0.457 and 0.053 seconds, respectively

Quality Assessment Metrics



- Time Distribution Analysis



- Text Statistics Comparison

Lastly, Figure 9 is a comparison of text statistics between original and summarized text where the volume of words, sentences, and characters is significantly reduced, effectively illustrating the effectiveness of the framework in reducing the volume of academic documents without compromising the main semantic content.

Detailed Performance Evaluation

Category	Metric	Value
Basic	Compression Ratio	70.95%
Basic	Reduction Factor	3.44x
Readability	Flesch Reading Ease	7.35
Readability	Flesch-Kincaid Grade	15.8
Readability	Readability Level	Very Difficult
Content	Vocabulary Coverage	24.37%
Content	Summary Density	72.09%
Content	Keyword Retention	60%
Quality	Overall Score	45.33%
Quality	Information Retention	23%
Quality	Coherence Score	67.67%
Quality	Grade	F (Very Poor)
Timing	Preprocessing	0.053s
Timing	OCR Extraction	0.457s
Timing	Summarization	7.535s
Timing	Total Time	8.046s

Table 1 describes an analysis of the performance of the proposed framework in terms of quantitative measures. The system resulted in a compression ratio of 70.95 and a reduction factor of 3.44 times, a large reduction in the size of the original document, yet the important meaning was retained. Academically complex summaries are indicated by the readability scores (Flesch Reading Ease 7.35, Flesch-Kincaid Grade 15.8). Evaluation of content demonstrates 60 percent keyword retention, but vocabulary coverage (24.37) and information retention (23) shows less detailed content because of excessive compression. The coherence score of 67.67 percent is a confirmation of moderate logical flow. Computationally, the overall processing time was 8.046 seconds, and preprocessing (0.053 s), OCR (0.457 s) required a small part of processing time compared to transformer-based summarization (7.535 s), which occupies the majority of the system processing.

Comparison with existing systems

The comparative analysis below reveals the unique merits of the proposed framework to the current transformer-based and OCR-integrated summarization systems. As Table II demonstrates, the main models studied in traditional transformer research include BART, T5, and PEGASUS that are trained on clean CNN/DailyMail datasets and measured using only ROUGE metrics. However, the method suggested involves the incorporation of both BART, DistilBART, and T5 into an OCR-aware adaptive pipeline that is designed specifically to work with scanned documents. The proposed system, in contrast to traditional NLP pipelines which assume that the digital text is error-free, utilizes OCR processing and multi-metric analysis, such as the compression ratio, readability, coherence, and information retention. This renders the framework more realistic in real world scanned academic documents as opposed to idealized digital inputs.

Comparison with Transformer Summarization Research

Feature	Existing Research [22]	Proposed Method
Models	BART, T5, PEGASUS	BART + DistilBART + T5
Dataset	CNN/ DailyMail	CNN/DailyMail
Evaluation	ROUGE only	Multi-metric evaluation
OCR Support	No OCR support	OCR-aware system
Document Type	Clean digital text	Scanned documents
Processing Pipeline	Standard NLP pipeline	OCR-aware adaptive pipeline

OCR-Based Summarization Comparison

Feature	Existing Research[23]	Proposed Method
OCR Engine	Tesseract	Optimized PyTesseract
Pipeline	OCR + LLM	OCR + Transformer Models

Language Support	Multilingual	Academic documents
Model Selection	Fixed	Adaptive
Evaluation Metrics	Not detailed	Multi-metric evaluation
Processing Optimization	No	Yes

Comparison with OCR Document Processing Research

Feature	Existing Research[24]	Proposed Method
OCR Engine	Standard Tesseract	Optimized PyTesseract
Preprocessing	Basic binarization	Adaptive preprocessing
Skew Correction	None	Automatic
Noise Removal	Basic	Bilateral filtering
OCR Accuracy Optimization	No	Yes
Processing Pipeline	Fixed	Adaptive
Processing Time	Not optimized	Faster

Likewise, Tables III and IV portray advancements with regard to existing OCR-based summarization and document processing studies. Although most academic papers are using standard Tesseract with fixed OCR + LLM pipelines, and little information on evaluation, the given methodology applies optimized PyTesseract with transformer models and an adaptive model selection mechanism. Higher levels of preprocessing, such as automatic skew removal and bilateral filtering, are used to improve the readability of text prior to OCR (usually absent in previous studies). Moreover, as opposed to the old systems with rigid pipelines and non-optimizing processing time, the suggested system presents adaptive and optimization strategies, leading to a quicker and more efficient process. On the whole, this comparison proves that the suggested approach is not limited to the traditional OCR or transformer summarization studies as it incorporates adaptive preprocessing, smart model selection, and the performance assessment into one unified solution.

Discussion

The suggested intelligent model effectively incorporates adaptive image preprocessing, OCR, transformer-based abstractive summarization, and an AMSA into adaptable pipeline in the summarization of scanned academic documents. The system had compression ratio of 70.95 with 3.44x reduction factor, compressing a 148-word document to 43 words and preserving 60% keyword retention and coherence score of 67.67. BART showed the best compression (82.43%) and T5 showed the quickest running time (4.085 s), whereas DistilBART had the highest quality score (45.3%). The originality of the work is in the integration of the adaptive skew-corrected image preprocessing and the selection of intelligent transformers in accordance with document properties and user priorities. The major benefits are the possibility of processing almost in real-time (total time 8.046 s), enhanced OCR accuracy, model optimization at the current time, and interactive dashboard with visual performance representation.

Although these are the strengths, some weaknesses are still present. There is a reduction in detailed contextual content of information retention (23%) and vocabulary coverage (24.37) by aggressive compression. The measures of readability (Flesch Reading Ease 7.35; Grade Level 15.8) indicate that the summaries are still academically challenging, obstructing their availability to non-expert users. Moreover, transformer inference approximates about 94% of all processing time, which is the key computational bottleneck. Future directions can include domain-specific fine-tuning to better preserve semantics, hybrid extractive-abstractive mechanisms to better memorize information, adaptive lightweight transformer architectures to implement faster deployments, and adaptive readability control to better adapt summaries to different users.

Conclusion

This paper suggested a flexible and holistic model to automated summarization of scanned scholarly papers through the combination of sophisticated methods. The system effectively converts scanned text into readable summaries and allows users to choose the model among BART, DistilBART and T5 according to the features of the documents and the priorities they set. Evaluation Effective compression performance (70.95%), balanced coherence (67.67%), and near real-time feasibility with total processing time of 8.046 seconds were demonstrated to be effective in an experimental evaluation. The combination of intelligent preprocessing, deep neural summarization, and performance visualization in the form of an interactive dashboard creates a scalable and practical solution to academic repositories and digital libraries. The analytical dashboard created also increases the level of transparency by allowing performance measurement and comparative assessment within one interface. Although semantic retention and optimization of computations still require enhancement, the suggested framework will serve as a solid ground upon future developing adaptive, context-dependent document summarization systems.

References :

1. G. Sharma and D. Sharma, "Automatic Text Summarization Methods: A Comprehensive Review," 2023. doi: 10.1007/s42979-022-01446-w.
2. Supriyono, A. P. Wibawa, Suyono, and F. Kurniawan, "A survey of text summarization: Techniques, evaluation and challenges," Nat. Lang. Process. J., 2024, doi: 10.1016/j.nlp.2024.100070.
3. B. Khan, Z. A. Shah, M. Usman, I. Khan, and B. Niazi, "Exploring the Landscape of Automatic Text Summarization: A Comprehensive Survey," IEEE Access, 2023, doi: 10.1109/ACCESS.2023.3322188.
4. A. P. Widyassari et al., "Review of automatic text summarization techniques & methods," 2022. doi: 10.1016/j.jksuci.2020.05.006.
5. M. A. Rakrouki, N. Alharbe, M. Khayyat, and A. Aljohani, "TG-SMR: A Text Summarization Algorithm Based on Topic and Graph Models," Comput. Syst. Sci. Eng., 2023, doi: 10.32604/csse.2023.029032.
6. S. Gupta and S. K. Gupta, "Abstractive summarization: An overview of the state of the art," 2019. doi: 10.1016/j.eswa.2018.12.011.
7. D. Saputra Laoli and B. Kurniawan Soegoto, "Development of A Mobile-Based Document Summarization Application Utilizing Optical Character Recognition (OCR)," 2025.
8. C. Ma, W. E. Zhang, M. Guo, H. Wang, and Q. Z. Sheng, "Multi-document Summarization via Deep Learning Techniques: A Survey," ACM Comput. Surv., 2023, doi: 10.1145/3529754.
9. J. ge Yao, X. Wan, and J. Xiao, "Recent advances in document summarization," Knowl. Inf. Syst., 2017, doi: 10.1007/s10115-017-1042-4.
10. D. Tsirmpas, I. Gkionis, G. T. Papadopoulos, and I. Mademlis, "Neural natural language processing for long texts: A survey on classification and summarization," 2024. doi: 10.1016/j.engappai.2024.108231.
11. F. R. Chen and D. S. Bloomberg, "Summarization of Imaged Documents without OCR," Comput. Vis. Image Underst., 1998, doi: 10.1006/cviu.1998.0688.
12. G. Abinaya, G. D. Rao, P. S. R. Gopal, M. Kaushik, and B. S. V. Vignesh, "Automated Document Processing: Combining OCR and Generative AI for Efficient Text Extraction and Summarization," in Proceedings of the 2024 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems, ICSES 2024, 2024. doi: 10.1109/ICSES63760.2024.10910510.
13. P. A. Villa-García, R. Alonso-Calvo, and M. García-Remesal, "End-to-end entity extraction from OCRed texts using summarization models," Neural Comput. Appl., 2024, doi: 10.1007/s00521-024-10422-9.
14. A. EL DOR and O. E. Abdulhussein, "A Deep Learning Framework for Extracting and Summarizing Text from Images," InfoTech Spectr. Iraqi J. Data Sci., 2026, doi: 10.51173/ijds.v3i1.56.
15. F. A. Ghanem, M. C. Padma, H. M. Abdulwahab, and R. Alkhatib, "Deep Learning-Based Short Text Summarization: An Integrated BERT and Transformer Encoder–Decoder Approach," Computation, 2025, doi: 10.3390/computation13040096.
16. M. Azhar, A. Amjad, D. A. Dewi, and S. Kasim, "A Systematic Review and Experimental Evaluation of Classical and Transformer-Based Models for Urdu Abstractive Text Summarization," 2025. doi: 10.3390/info16090784.
17. F. Zaman, F. Kamiran, M. Shardlow, S. U. Hassan, A. Karim, and N. R. Aljohani, "SATS: simplification aware text summarization of scientific documents," Front. Artif. Intell., 2024, doi: 10.3389/frai.2024.1375419.
18. N. Dhanda and K. K. Gupta, "A Novel Approach to Text Summarization Using Machine Learning," Asian J. Res. Comput. Sci., 2024, doi: 10.9734/ajrcos/2024/v17i4432.
19. J. D'Silva and U. Sharma, "Automatic Text Summarization of Konkani Folk Tales Using Supervised Machine Learning Algorithms and Language Independent Features," IETE J. Res., 2023, doi: 10.1080/03772063.2021.1987993.
20. Z. Jalil, M. Nasir, M. Alazab, J. Nasir, T. Amjad, and A. Alqammaz, "Grapharizer: A Graph-Based Technique for Extractive Multi-Document Summarization," Electron., 2023, doi: 10.3390/electronics12081895.
21. P. Mahalakshmi and N. S. Fatima, "Summarization of Text and Image Captioning in Information Retrieval Using Deep Learning Techniques," IEEE Access, 2022, doi: 10.1109/ACCESS.2022.3150414.
22. E. Daraghmi, L. Atwe, and A. Jaber, "A Comparative Study of PEGASUS, BART, and T5 for Text Summarization Across Diverse Datasets," Futur. Internet, 2025, doi: 10.3390/fi17090389.
23. H. Madhavi, J. Cherian, Y. Khamkar, and D. Bhagat, "Low-Resource Language Processing: An OCR-Driven Summarization and Translation Pipeline," arXiv Comput. Sci., May 2025, [Online]. Available: http://arxiv.org/abs/2505.11177
24. A. N. Prof. Anuradha Thorat , Mayur Zagade, Shivani More , Manish Pasalkar, "Research Paper on Text Extraction using OCR," Int. J. Adv. Res. Sci. Commun. Technol., vol. 3, no. 14, pp. 1–5, 2023, doi: 10.48175/568.

ISSN : 2348-7143
March 2026
RESEARCH JOURNEY International Multidisciplinary E-Research Journal
Impact Factor (SJIF) - 6.625 | ***Special Issue 389 :*** *Advanced Computing and Artificial Intelligence (ACAI)*

Prediction of Human Mental State Through Daily Routine Using Machine Learning

Mahesh Patil
R. C. Patel Educational trusts Institute of Management Research and Development, Shirpur.

Abstract:

The habit of daily living like sleeping patterns, physical exercise, screen time, social interactions, and study and work patterns are very important in determining the mental state of human being. The pattern of these habits can be used to determine the levels of stress, emotional equilibrium, and overall mental well-being. In this study, a machine learning model is proposed to predict mental stress based on lifestyle patterns of daily living. A labelled dataset with attributes of habits like sleeping duration, exercise, caffeine consumption, meditation and digital use was pre-processed by cleaning, encoding and normalizing the data before dividing into training and testing datasets. Three supervised machine learning models, namely K-Nearest Neighbours, Support Vector Machine and Naïve Bayes, were trained and tested based on accuracy, precision, recall, and Fl-score. Among the model used, the Support Vector Machine with a polynomial kernel had the best prediction accuracy of 52.94%, performing better than KNN and Native Bayes Models. This study emphasizes the significance of healthy lifestyle practices and indicates that predictive models based on routines can help in the early detection of mental health issues, increase productivity and help make better behavioural decisions.

Keywords: Human Mind Prediction, Naïve Bayes, K-Nearest Neighbours (KNN), Support Vector Machine (SVM), Behavioural Patterns, Lifestyle Data, Mood Prediction, Mental Well-being.

Introduction:

The human mind is strongly influenced by everyday routines such as sleep, study or work schedules, exercise, social interactions, and digital usage. In today's digital age, young people are increasingly spending more time on social media and AI-powered platforms, often starting their morning by scrolling their feeds instead of engaging in physical or mindful activities. Such habits, when prolonged, can gradually reduce thinking ability, weaken decision making power and negativity affects emotional balance. The lack of physical games, exercise and meditation further accelerates stress, anxiety and unproductivity making it crucial to study how daily lifestyle choices shape mental well-being.

Recent studies emphasizes that mental health plays a vital role in human emotions, reasoning, and social interactions, and its neglect leads to severe consequences such as stress, depression and reduce productivity. Traditional methods of diagnosing mental states, such as psychological evaluations are time consuming and not feasible for large populations on the other hand machine learning (ML) techniques have shown promise in predicting mental health states by analysing behavioural, social and routine based data.ML algorithms like Naïve Bayes, K-Nearest Neighbours(KNN), and Support Vector Machine(SVM) are widely used for recognizing patterns and classifying mental states, enabling a more scalable and data-driven approach.

This project, "Understanding Human Mind Through Daily Routine", applies ML methods to interpret routine-based lifestyle inputs and predict mental states. By focusing on activities such as sleep duration, study time, social media uses and exercise, the model identifies patterns that reflects stress levels, emotional balance and overall mindset. The goal is not only to highlight the negative effect of digital addiction and lack of physical activities but also to promote balanced habits, including mindful internet uses, regular exercise and meditation. Through this approach the study demonstrates how analysing daily routines can support early detection of mental health concern, improve productivity and enhanced overall well-being.

Objectives:

1. To study how daily routines including sleep, study, exercise and digital uses influence mental health and behaviour.
2. To analyse the impact of social media and AI tool addiction on thinking power and decision-making ability, especially among youth.
3. To apply machine learning algorithms (Naïve Bayes, KNN and SVM) to predict mental states based on lifestyle inputs.
4. To promote awareness about the importance of physical activities, meditation and mindful internet uses for improving overall well-being.
5. To promote insights that can help individuals achieve a balance routine, enhancing productivity and mental clarity.

Problem definition:

Excessive dependence of social media and AI-powered platforms among youth is weakening critical thinking, decision-making ability and emotional balance, while reduce engagement in physical activities such as exercise and mediation further harms mental well-being.

Traditional psychological assessment is

resource-intensive and unsuitable for large population therefore there is a need for machine-based approach to analyse daily routines – such as sleep, study, digital uses and exercise-to predict mental states and promote healthier more balanced lifestyle.

Literature Review:

1. **Machine Learning for Stress and Anxiety Prediction**
   Recent studies show that machine learning can effectively predict stress and anxiety using physiological and behavioural data. Research using wearable devices demonstrates that heart rate, activity levels, and sleep patterns can indicate emotional conditions. These finding confirm that routine-generated data can support mental state prediction.
2. **Passive Sensing and Lifestyle Monitoring**
   Studies on passive sensing through smartphones and wearable technologies reveal that daily habits such as mobility, sleep behavioural, and digital usage are closely linked to psychological well-being. Combining multiple data sources improves prediction accuracy and enables scalable mental health monitoring.
3. **Behavioural and Voice-Based Detection**
   Research on behavioural and voice data indicates that supervised learning algorithms can detect early signs of mental disorders. Temporal and routine-based behavioural information play an important role in identifying emotional distress and mood variations.
4. **Routine Survey and Real-Time Data Prediction**
   Several studies applied machine learning models to structured survey data and real-time lifestyle inputs to predict depression and stress. Findings suggest that features like sleep duration, work patterns, and daily habits significantly influence mental health outcomes.
5. **Social Media and Digital Behaviour Analysis**
   Research analysing social media activity highlights the strong relationship between digital behaviour and emotional stability. While many studies focus only on online content, they emphasize the impact of excessive digital usage mental well-being.

Overall, the existing literature confirms that machine learning techniques such as SVM, KNN, and Naïve bayes are effective for mental health prediction. However, most research concentrates on either wearable data or social media analysis separately. Limited studies integrate multiple daily routine factors together. Therefore, the present study contributes by combining lifestyle features such as sleep, exercise, meditation, and screen time into a unified predictive model for mental state assessment.

Methodology:

This study used supervised machine leaning to predict mood based on daily habits. A database was collected containing features such as sleep duration, exercise, screen time, water and caffeine intake, mediation and social interaction each labelled with the individuals mood. The data was clean, categorical variables were encoded, and numerical feature were normalized. It was then split into training and testing sets (80:20 ratio). Three models K-Nearest Neighbours (KNN), Support Vector Machine (SVM) and Naïve Bayes-were trained on the training data. Each mode was tested on unseen data and evaluated using accuracy, precision, recall and F1-score. The result was compared to identify the most effective model for mood prediction based on based on daily behavioural pattern.

KNN: (K-Nearest Neighbours) is a supervised learning algorithm predicts values based on the average of the nearest neighbours.

Experiment: The experiment evaluates the performance of three supervised learning algorithms – K-Nearest Neighbours (KNN), Support Vector Machine (SVM) and Naïve Bayes on a mood prediction dataset using different training/testing splits and hyper parameters.

Table 1: KNN Algorithm Hyper parameters

KNN is a supervised training algorithm that predicts output based on the average of the nearest neighbours. The model was tested with K values 3,5 and 7 across train-test splits of 0.8/0.2, 0.7/0.3, and 0.6/0.4.

Sr. No	K Value	Test Dataset	Train Dataset	Accuracy
1.	K=3	0.2	0.8	0.3617021276595745
		0.3	0.7	0.36619718309859156
		0.4	0.6	0.32978723404255317

Sr. No	K Value	Test Dataset	Train Dataset	Accuracy
1.	K=3	0.2	0.8	0.3617021276595745
		0.3	0.7	0.36619718309859156
		0.4	0.6	0.32978723404255317

3.	K=7	0.2	0.8	0.3829787234042553
		0.3	0.7	0.4647887323943662
		0.4	0.6	0.39361702127659576

Observation: Accuracy improves slightly with increasing K values and decrease as training data is reduced. The best KNN performance was 46.48% accuracy at K=7 and a 70/30 split.

Table 2: SVM (Support Vector Machine)

Sr. No	Test Dataset	Train Dataset	Kernel	Accuracy
1.	0.2	0.8	Linear	0.37254901960784315
2.	0.3	0.7	Linear	0.34210526315789475
3.	0.4	0.6	Linear	0.38613861386138615
4.	0.2	0.8	RBF	0.45098039215686275
5.	0.3	0.7	RBF	0.50000000000000000
6.	0.4	0.6	RBF	0.46534653465346537
7.	0.2	0.8	Poly	0.5294117647058824
8.	0.3	0.7	Poly	0.4605263157894737
9.	0.4	0.6	Poly	0.46534653465346537

Observation: The best accuracy of 52.94% was achieved using the polynomial kernel at an 80/20 split. Overall, the Polynomial and RBF kernels outperformed the Linear Kernal.

Table 3: Naïve Bayes:

Test Dataset	Train Dataset	Model	Accuracy
0.2	0.8	Gaussian NB	0.42
0.2	0.8	Multinominal NB	0.47
0.3	0.7	Gaussian NB	0.39
0.3	0.7	Multinominal NB	0.44
0.4	0.6	Gaussian NB	0.41
0.4	0.6	Multinominal NB	0.46

Both Gaussian and Multinomial Naïve Bayes models were tested over three train-test splits.

Observation: Multinomial Naïve Bayes consistency outperformed Gaussian NB. The best result was 47.00% accuracy at an 80/20 split using multinomial NB.

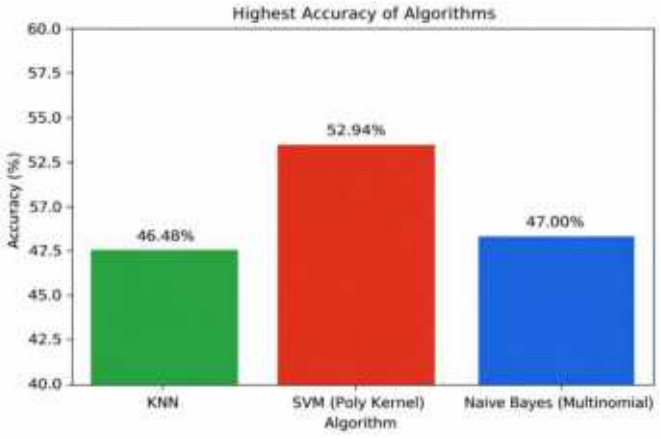
Result:

This table shows the highest accuracy achieved by each classification algorithm tested on the mood prediction dataset.

Test Dataset	Train Dataset	Model	Accuracy
0.2	0.8	Gaussian NB	0.42
0.2	0.8	Multinominal NB	0.47
0.3	0.7	Gaussian NB	0.39
0.3	0.7	Multinominal NB	0.44
0.4	0.6	Gaussian NB	0.41
0.4	0.6	Multinominal NB	0.46

Figure 1: Comparative analysis of all machine learning classifier algorithm accuracy

This Bellow graph shows the analysis of the accuracy score among different machine learning algorithms like decision tree, SVR, KNN Regressor and Linear Regression. SVM achieves the highest accuracy 52.94% to predict the mental health.



Conclusion:

This study demonstrates that machine learning algorithm such as K-Nearest Neighbours, Support Vector Machine, Naïve Bayes can effectively analyse daily routine data to predict human mental state with reasonable accuracy, among the model tested, SVM with a polynomial kernel showed the highest accuracy, highlighting its potential for understanding for complex behavioural pattern. The findings reinforce the significant impact of daily habits-including sleep exercise and digital uses on mental well-being. By leveraging routine bases data, this approach offers a scalable and data-driven alternative to traditional psychological assessments, promoting early detection of mental health concerns. ultimately, this research underscores the importance of balanced lifestyle choices and mindful technology used for enhancing mental clarity and productivity paving the way of personalized interventions that support overall mental health.

References :

1. N. F. M. R. Z. W. T. M.-B. K. P. M. B. B. A. T. L. E. B. Md Sabbir Ahmed, "WatchAnxiety: A Transfer Learning Approach for State Anxiety Prediction from Smartwatch Data," arXiv Preprint, 2025.
2. W. Q. J. Z. S. L. X. L. X. Z. C. D. B. W. Y. S. J. Y. B. W. X. L. S. G. P. L. J. W. G. J. S. C. ShiYing Shen, "Passive Sensing for Mental Health Monitoring Using Machine Learning With Wearables and Smartphones: Scoping Review," Journal of Medical Internet Research (JMIR), vol. Volume 27, 2025.
3. A. I. A. A. R. K. G. G. T. F. A. S. J. M. Sunil Kumar Sharma, "Early Detection of Mental Health Disorders Using Machine Learning Models Using Behavioral and Voice Data Analysis," Scientific Reports, vol. Volume 15, p. 16518, 2025.
4. S. T. C. S. K. S. G. C. L. P. A. L. C. D. E. N. Tuarob, "How are you feeling?: A personalized methodology for predicting mental states from temporally observable physical and behavioral information," Journal of Biomedical Informatics, p. 1–19, 2017.
5. S. K. T. G. S. B. U. P. P. Yadav, "Predicting depression from routine survey data using machine learning.," International Conference on Advances in Computing, Communication & Control (ICACCCN), p. 163–168, 2020.
6. U. U. A. U. Madububambachu, "Machine learning techniques to predict mental health diagnoses: A systematic literature review," Clinical Practice & Epidemiology in Mental Health, p. 20, 2024.
7. S. R. R. P. S. S. S. Y. S. S. S. G. R. Settu, "Predicting mental health through machine learning algorithms," 9th International Conference on Smart Structures and Systems (ICSSS), p. 1–6, 2023.
8. S. K. A. A. I. K. A. R. T. G. G. A. F. S. J. Sharma, "Early detection of mental health disorders using machine learning models with behavioral and voice data analysis," Scientific Reports, p. 16518, 2025.
9. S. A. Avani Shinde, "Machine Health Prediction Using Machine Learning," International Journal of Technology & Emerging Research (IJTER), vol. Volume 1, no. Issue 7, p. 62–67, 2025.
10. V. P. V. a. S. S. Asole, "Mental Stress Prediction Using Machine Learning on Real Time Dataset," Journal of Electrical Systems, vol. Volume 20, no. Issue 7, p. 2143–2150, 2024.
11. G. S. C. R. K. M. K. A. K. K. S. Sandeep Mishra, "A Study and Prediction of Psychological Disorders Through Machine Learning," International Journal of Algorithms Design and Analysis Review, vol. Volume 02, no. Issue 02, p. 32–38, 2024.
12. I. A. L. S. F. H. E. H. M. M. H. J. C. B. Md Al Amin, "Predicting and Monitoring Anxiety and Depression: Advanced Machine Learning Techniques for Mental Health Analysis," British Journal of Nursing Studies, vol. Volume 4, no. Issue 2, p. 66–75, 2024.
13. K. H. O. D. I. B. M. V. R. Prerana Jain, "Detection and Prediction of Future Mental Disorder from Social Media Data Using Machine Learning, Ensemble Learning and Large Language Models," International Journal of Scientific Research in Computer Science, Engineering & Information Technology, vol. Volume 11, no. Issue 3, p. 267–273, 2025.
14. R. M. A. D. Febrian, "Machine Learning Models for Predicting Mental Health Indicators Using Digital Physical Activity Data: A Systematic Literature Review," Journal Informatics, Education and Management (JIEM), vol. Volume 7, no. Issue 2, 2025.
15. B. T. G. T. İsmail Baydili, "Deep Learning-Based Detection of Depression and Suicidal Tendencies in Social Media Data with Feature Selection," Behavioral Sciences, vol. Volume 15, no. Issue 3, 2025.
16. M. K. L. C. Y. C. R. M. Lin Sze Khoo, "Machine Learning for Multimodal Mental Health Detection: A Systematic Review of Passive Sensing Approaches," Sensors, vol. Volume 24, no. Issue 2, 2024.
17. R. L. K. S. O. B. B. N. D. Sellappan Palaniappan, "Training the Brain: A Machine Learning Approach to Predicting Well-Being Through Intentional Thought Pattern Modification," Journal of Informatics and Web Engineering, vol. Volume 4, no. Issue 3, p. 64–89, 2025.
18. Y.-B. S. D. J. J.-W. H. S. P. P. H.-J. L. H. L. C.-H. C. Jin-Hyun Park, "Machine Learning Prediction of Anxiety Symptoms using Multimodal Data from Virtual Reality Sessions," Frontiers in Psychiatry, vol. Volume 15, 2024.
19. D. L. P. R. A. F. A. Mentis, "Applications of Artificial Intelligence—Machine Learning for Detection of Stress: A Critical Overview," Molecular Psychiatry, vol. Volume 29, p. 1882–1894, 2024.
20. A. A. A. A. E. F. R. J. H. M. E. F. Abdelrahaman A. Hassan, "Mental Health Prediction from Multi-Source Data in Daily Life," arXiv Preprint, 2025.

ISSN : 2348-7143
March 2026
RESEARCH JOURNEY International Multidisciplinary E-Research Journal
Impact Factor (SJIF) - 6.625 | ***Special Issue 389 :*** *Advanced Computing and Artificial Intelligence (ACAI)*

Sentiment Analysis in Natural Language Processing: Techniques, Challenges, and Applications

Mr. Harshal Bhamare, Mr. Rahul S. Badgujar, Mr. Vivek N. Chavan
R.C.P.E.T's Institute of Management Research and Development, Shirpur

Abstract

Sentiment analysis is one of the key parts of Natural Language Processing, which helps in figuring out the feelings, views, and attitudes people have when they write or speak [5]. As more and more stoner-generated content is coming through platforms like Twitter, Amazon, and Reddit, the volume of textbook data has grown greatly and hence requires automated tools to effectively do interpretation [2]. This paper presents a deep look at sentiment analysis that includes its styles, issues, and what the future holds [2]. The stylistic approach to sentiment analysis ranges from rule-based that uses sentiment dictionaries to score words [5] to traditional machine literacy models such as Naive Bayes and Support Vector Machines that use hand-drafted features [6]. Also, we see more recent approaches grounded in deep literacy. These number RNNs with LSTM units, and attention-driven motor models such as BERT, which boasts remarkable contextual understanding and delicacy in sentiment bracket [1][2]. Typical exemplifications are the automatic sentiment bracketing in client experience operations, whereby companies dissect client feedback so as to optimize products; the covering of public sentiment during political juggernauts; and the assessing of patient feedback in the health system with the aim of acclimating care plans [4]. Although these operations gauge various sectors, still issues that remain nondescript include the processing of sardonic textbook; the discordance between the non-sensational meaning of a statement and the sentiment it conveys; the cross-dialect mismatch in the use of language; and the inadequate information on contexts when the models develop for some form of field [2][4]. These issues bring to light the call for innovative results. Looking at the trends, we notice an increase in multimodal affront discovery that includes textbook, images, and audio; multilingual models for low resource languages; and resolvable AI, which improves translucency and fairness [1][2]. This study reports on the transformational part played by sentiment analysis in myriad diligence at the same time as bringing to the fore its issues. In developing robust, indifferent, and transparent systems, sentiment analysis may more serve the real-world operations [4].

Keywords: Sentiment Analysis, Opinion Mining, Natural Language Processing, Lexicon-Based Methods, Machine Learning, Deep Learning, Transformer Models, Multilingual Sentiment Analysis, Sarcasm Detection, Explainable Artificial Intelligence.

I. Introduction

Sentiment analysis or opinion mining is an important subfield of NLP, whose focus is the automatic identification and interpretation of subjective information represented by emotions, opinions, and attitudes expressed in textual data [5][6]. With the sudden emergence of digital communication platforms—especially social media networks like Twitter, Facebook, and Instagram, among others—and e-commerce platforms, the volume of user-generated textual data has grown exponentially, thus pressing the need for automated techniques of sentiment analysis [2][4].

Sentiment analysis, in most cases, categorizes texts as positive, negative, or neutral. For instance, a positive review like "The performance was amazing and inspiring" denotes great enthusiasm, while the negative remark "The food arrived cold and the staff was rude" reflects dissatisfaction. A neutral statement like "The product came in standard packaging" denotes neither approval nor disapproval. These examples show how a sentiment analysis system extracts emotional orientation from natural language text [3][5].

The demand for sentiment analysis has increased significantly because it provides actionable insights across various industries and application domains. Organizations leverage sentiment analysis for better understanding of public opinion, customer behaviour, and social trends [4] [6]. The applications of sentiment analysis span a wide range of domains:

- Customer Satisfaction: Retailers examine customer reviews to find out product strengths, service deficiencies, and areas for improvement that enhance the overall customer experience.
- Political Sentiment: Political campaigns and analysts track social media data for an overview of public opinion and sentiment among voters, thereby allowing data-driven strategy formulation [4].
- Brand Perception: Companies use sentiment monitoring to assess public attitudes toward their brand, manage reputation, and respond proactively to negative feedback [2][6].
- Public Services: Government agencies analyze feedback of citizens to enhance public engagement, evaluate policies, and support

informed decision-making processes [4].

Overall, sentiment analysis is an important element that turns large-scale unstructured textual data into meaningful knowledge supportive of informed decision-making in both the private and public sectors [1][2].

II. Research Methodology

This paper uses a descriptive and analytical approach in describing the use of NLP techniques in sentiment analysis to detect emotions, opinions, and attitudes expressed in text data [5], [6]. The research will rely on secondary materials and sources, namely peer-reviewed research journals, open-source sentiment datasets such as IMDb, Twitter, and Amazon reviews, and technical reports from sentiment analysis models like VADER, Naïve Bayes, LSTM, and BERT [1], [3], [7].

The data collected is used to perform analysis with respect to the performance, limitations, and challenges of several approaches in sentiment analysis [2]. Overall, the research process contains a set of very critical stages: data preprocessing, feature extraction, and model evaluation. In the cleaning and normalization of textual data in the pre-processing stage, techniques like tokenization, stop-word removal, stemming, and lemmatization are employed [4].

For feature extraction, techniques such as TF-IDF, Word2Vec embedding's, and transformer-based contextual embedding's like BERT are applied for converting textual information to numerical representations that may be fed into the model for training [1], [5]. Various methods for sentiment analyses have been explored and compared in this study that include lexicon-based methods, traditional machine learning models, deep learning architectures, and transformer-based models, by employing Python-based libraries which are NLTK, TextBlob, scikit-learn, and TensorFlow [6]-[9].

Model performance is measured using the standard metrics of accuracy, precision, recall, and F1-score to determine the effectiveness in categorizing sentiments as positive, negative, or neutral [2]. Another aim of this study is to review the current techniques that are in use in sentiment analysis and determine some of the key challenges like sarcasm detection, multilingual sentiment understanding, and data imbalance [4], [11]. The expected output of this study is to show that the modern transformer-based models, including BERT and its optimized variants, will achieve a better contextual understanding and classification performance as compared with the traditional techniques of machine learning [1], [8], [10].

III. Literature Review

In this research, the basis of the descriptive and analytical methodology is various NLP techniques that have been applied to detect emotions, opinions, and attitudes expressed in textual data. Hence, this study used secondary data obtained from peer-reviewed research journals, publicly available data for sentiment analysis such as IMDb movie reviews, Twitter data, Amazon product reviews, and technical reports about sentiment analysis models such as VADER, Naïve Bayes, LSTM, and BERT [1][3][5][6].

This data is used in analyzing the different approaches of sentiment analysis regarding their performance, limitations, and challenges. The methodology involves key stages such as data preprocessing, feature extraction, model implementation, and evaluation.

In the data preprocessing stage, cleaning and standardizing textual data were done by techniques like tokenization, stop-word removal, stemming, and lemmatization in order to decrease the noise and enhance the model's performance [5][6]. During the feature extraction stage, methods such as Term Frequency–Inverse Document Frequency (TF-IDF), Word2Vec embeddings, and contextual embeddings generated by transformer models like BERT [1][2][8] were considered in order to transform textual information into numerical representations.

The various methods of sentiment analysis that are implemented and compared in this study include lexicon-based methods, traditional machine learning, and deep learning architectures using transformer models. Techniques used are implemented and analyzed using the most common Python-based NLP tools and frameworks such as NLTK, Text Blob, scikit-learn, and Tensor Flow [3][7][8].

Performance evaluation is done with the help of the standard metrics for classification problems, such as accuracy, precision, recall, and F1-score, to determine how good a particular approach is at sentiment classification into its positive, negative, and neutral categories. The two most important goals of this paper are to analyze various sentiment analysis techniques, pinpoint the challenges or difficulties, such as sarcasm detection, multilingual sentiment estimation, data imbalance, and check model efficiency in real-world sentiment analysis applications [2] [4].

The expected conclusion from the study is that state-of-the-art transformer-based models like BERT and GPT perform far better in terms of the level of accuracy and depth of contextual representation compared to the traditional machine learning methods. Simultaneously, the research points out the importance of explain ability, fairness, and bias awareness in AI development to help increase transparency and confidence in sentiment analysis applications [1][4].

IV. Evolution of Sentiment Analysis Techniques

- Feting □ as positive or " lol " as environment-dependent. [3], [6].
- Intensity Modifiers Amplifying scores for words like "veritably" or "extremely." These

styles are computationally effective and interpretable but face challenges [3], [5].

- environment Insensitivity Interpret expressions like " not great " appreciatively but inaptly. [4], [6].
- Limited Vocabulary Experience the challenge of shoptalk, neologisms, or specific language. [2], [5].
- Cultural Variations Neglect indigenous or artistic differences in word operation. [4], [11].

V. Techniques in Sentiment Analysis

Sentiment analysis relies on a vast variation in computational methods, starting from simple word-based techniques to state-of-the-art neural network models for interpretation and evaluation of emotions conveyed in text [6]. Each technique has different strengths and weaknesses and is designed to address the complexity of human language across diverse contexts [2]. This section discusses lexicon-based techniques, traditional machine learning methods, deep learning approaches, and transformer-based architectures, outlining their underlying mechanisms, applications, and limitations.

• Lexicon-Based Methods

Lexicon-based approaches are those depending on pre-compiled sentiment dictionaries, where words are assigned prior polarity scores: a word like “joyful” might have a positive score, for instance, +4; on the other hand, “terrible” would have a negative score, for instance, −3 [5]. Popular tools include VADER, or Valence Aware Dictionary for Sentiment Reasoning, and TextBlob, among others, which are particularly effective when analyzing short texts, usually written in an informal style, such as posts, comments, and reviews on social media and review websites [3].

These tools also implement other linguistic features: emoticons and slang are explicitly recognized, for example, so emojis and colloquial expressions like “lit” can be rated as positive sentiment [3]. To further extend this, intensity modifiers - such as “very”, “extremely”, or “barely” - modify sentiment scores based on how well the strength of expressed emotions is represented [4].

Despite their efficiency and interpretability, lexicon-based approaches have some remarkable limitations. Most of them fail when it comes to context-dependent expressions and idiomatic phrases. For example, phrases like “sick performance” would be misclassified as negative since the literal meaning of “sick” is used while the actual intended sentiment is positive [6]. The inability to capture such context limits their power in dealing with subtle language and complicated sentence structure [2].

VI. Traditional Machine Learning Models

The first major advances in supervised sentiment analysis were achieved using machine learning models trained on labeled data, such as logistic regression, Naïve Bayes, and Support Vector Machines (SVMs) [6], [5]. These approaches rely heavily on manual feature engineering, where relevant linguistic features are explicitly designed and extracted before model training [2].

Commonly used features include:

- **N-grams:** Sequences of words, such as bigrams (e.g., “very good”), which help capture local phrase patterns and sentiment-bearing expressions [6].
- **Term Frequency–Inverse Document Frequency (TF-IDF):** A weighting technique that measures the importance of a word or phrase based on its frequency in a document relative to its occurrence across the entire corpus [5].
- **Syntactic Features:** Linguistic information derived from grammatical structures, such as part-of-speech (POS) tags and dependency parsing, which help incorporate sentence structure into sentiment classification [4].

These models perform well in well-defined and controlled domains, such as movie review sentiment classification or customer product feedback analysis, where training data is relatively structured and consistent [6]. Naïve Bayes is particularly effective for small datasets due to its low computational complexity and scalability to large corpora, while SVMs are well suited for handling high-dimensional feature spaces [6].

Despite their effectiveness, traditional machine learning approaches have notable limitations. Designing high-quality features requires significant domain expertise and manual effort, making the process time-consuming [2]. Moreover, these models often struggle to capture complex sentence structures and higher-level linguistic phenomena such as irony, sarcasm, and metaphor, which remain challenging problems in sentiment analysis [6], [4].

VII. Deep Learning Techniques

Deep learning brought in a transformation by which models themselves perform feature extraction and so we see patterns learned from raw text [2], [7]. In terms of Recurrent Neural Networks (RNNs), which we see play a large role here, we have the LSTM and GRU, which are meant to put together sequences of events, making them very useful for very long sentences or paragraphs [7]. Also, we have Convolutional Neural Networks (CNNs) that do well with local-scale information, in that they look at key phrases or sentiment-carrying words [2]. As for what we get out of all this, we see that:

- **Automatic Feature Learning:** avoids selecting features manually and lessens the workload [2].
- **Complex Pattern Recognition:** recognizes complex grammatical structures and different word usages with heightened accuracy on datasets such as IMDb or Yelp reviews [2], [6].
- **Versatility:** these models deal with different text types. For example, LSTMs can understand the sentiment of the sentence "Though the start was slow, the movie was thrilling" by context retention and framework analysis [7].
- **High-End Infrastructure:** GPUs and extensive datasets are needed to train these models [2], [7].
- **Enhanced Flexibility:** can lead to overfitting and decreased generalization to novel text [2].
- **Labeled Data Requirements:** essential and might be limited in specialized domains or languages [2], [6].

While these models contributed much to sentiment analysis, they also opened the way for more elaborate methods within the same field [2], [7].

• **Transformer-Based Models**

Transformer models utilize attention mechanisms to capture deep within-sentence and cross-sentence semantic relationships and, as a result, revolutionize the field of sentiment analysis [1], [8]. BERT (Bidirectional Encoder Representations from Transformers) and its variations, such as DistilBERT, as well as models like GPT and XLNet, understand context exceptionally well and are very useful for complex sentiment analysis tasks [1], [10], [11]. Their strengths include:

- **Bidirectional Context:** BERT attends to words in a sentence both forward and backward, allowing it to differentiate "The film was not bad" (positive sentiment) from "The film was bad" (negative sentiment) [1].
- **Multilingual Capabilities:** Pretrained on diverse corpora, enabling performance in multiple languages and global applications [11].
- **High Accuracy:** Achieve top benchmark results on datasets like SST-2, GLUE, or Twitter sentiment analysis. Pretraining on large-scale corpora such as Wikipedia and Book Corpus, followed by fine-tuning for tasks like classifying product reviews or multilingual tweets, is standard. DistilBERT is an example of a more efficient version of BERT that maintains high performance [1], [10].

However, these models also face challenges:

- **Resource Demands:** Require significant computational power for training and inference [1], [8].
- **Complexity:** Less interpretable than lexicon-based or traditional models, posing explainability challenges [1], [2].
- **Fine-Tuning Needs:** Performance depends on task-specific fine-tuning with labeled data [1], [10].

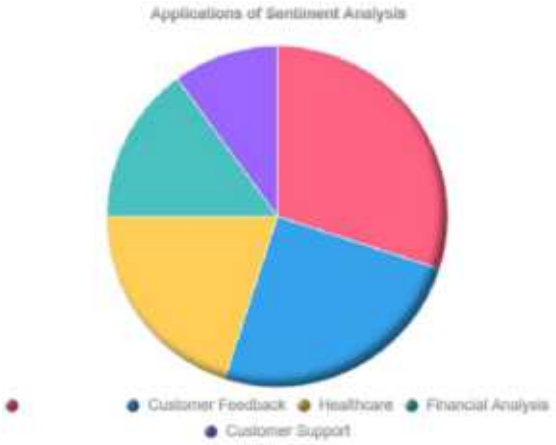
Transformer models now represent the cutting edge in sentiment analysis, driving improvements in both accuracy and versatility for diverse applications [1], [8], [10].

VIII. Customer Support and Chatbots

Integrating sentiment analysis into customer support systems and chatbots enhances emotional intelligence and adaptive responses [2], [3]. Key benefits include:

- **Dynamic Responses:** Chatbots can adjust their tone based on the detected emotion. For example, they can provide empathetic responses to frustrated users, such as "I am sorry you're upset" [3].
- **Improved Satisfaction:** Sentiment-aware interactions enhance the overall user experience in support roles and online chat systems [2], [3].
- **Proactive Issue Resolution:** By detecting negative sentiments in real time, chatbots can prioritize and address issues more efficiently, leading to personalized and responsive customer service that builds trust and engagement [2], [3].

The application of sentiment analysis in these contexts demonstrates its high value in understanding and acting on textual data. However, challenges remain, including cultural nuance interpretation and domain-specific language, which require ongoing improvements and model adaptation [4], [11].

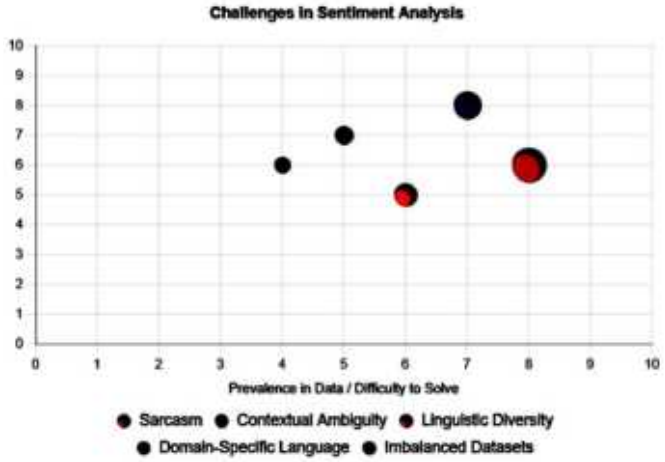


IX. Challenges in Sentiment Analysis

In many cases, datasets exhibit class imbalance, with disproportionate numbers of positive, negative, or neutral samples. For instance, product review datasets often contain more positive than negative examples, reflecting a general trend toward praise rather than criticism [2], [6]. This imbalance can significantly impact model performance:

- **Bias in Training:** Models may overfit to majority classes (e.g., positive sentiments) while underperforming on minority classes [2], [6].
- **Poor Generalization:** Imbalanced training data can reduce accuracy on real-world datasets, which may have a more balanced distribution of sentiments [2], [6].
- **Evaluation Issues:** Common metrics such as accuracy can be misleading when datasets are imbalanced. Addressing this requires either balanced datasets or specialized techniques, such as oversampling, data augmentation, or weighted loss functions [2], [6].

These challenges highlight the importance of robust model training and have motivated the development of advanced approaches in sentiment analysis, including multimodal analysis, cross-lingual learning, and bias mitigation techniques, which together improve performance and fairness in real-world applications [11].



X. Emerging Trends and Future Directions

Sentiment analysis has been constantly improving, as academics continue to strive for more accuracy and inclusiveness by applying new solutions to the prevailing challenges [2], [4]. The current trend is towards developing sophisticated, highly moral, and contextually informed systems that are able to effectively deal with sarcasm, language diversity, and bias in the outcome of sentiment analysis [4], [11]. The major developments and trends being witnessed in the realm of sentiment analysis are discussed in this chapter, which may determine the future course of the subject [4], [11].

Sarcasm Detection

Sarcasm and irony, where literal meanings contradict intended sentiments (e.g., "Great job, my phone crashed again!"), remain significant challenges for sentiment models [4], [11]. Recent advancements focus on leveraging contextual and multimodal information to improve detection:

- **Contextual Analysis:** Models consider user history, the immediate conversation context, or situational cues to identify sarcasm [4].
- **Multimodal Inputs:** Sentiment detection incorporates multiple forms of media, including visual and auditory signals. For example, a sarcastic tweet containing a rolling eyes emoji provides stronger cues for the intended sentiment [4], [11].
- **Hybrid Models:** Combining transformer-based language models like BERT with visual processing networks enhances sarcasm detection in social media, reducing misclassifications and improving sentiment interpretation in informal and nuanced texts [1], [4].

These approaches aim to increase model robustness and reliability, particularly in handling informal, context-dependent, and multimodal communication.

XI. Support for Low-Resource Languages

Most research in sentiment analysis has historically focused on English, limiting the applicability of models in low-resource languages such as Tamil, Swahili, and various regional dialects [11]. Recent trends are addressing these challenges and improving inclusivity by focusing on the following factors:

- **Multilingual Pretrained Models:** Models such as mBERT and XLM-RoBERTa undergo pretraining on diverse multilingual corpora, enabling sentiment analysis across multiple languages [1], [11].
- **Cross-Lingual Transfer:** Knowledge learned from resource-rich languages (e.g., English) can be transferred to low-resource languages, reducing the dependency on labeled datasets in these languages [11].
- **Socially-Driven Datasets:** Crowdsourcing initiatives are creating sentiment datasets for lesser-used languages, improving model inclusivity and performance [11].

These trends make it possible to perform sentiment analysis in multiple languages, thereby expanding the global applicability of sentiment analysis models [11].

XII. Conclusion

The evolution in sentiment analysis has been

significant too. What began with a set of simple and easy-to-implement rule-based systems has transformed into a sophisticated deep learning and transformer-based model with the ability to read and understand emotions in diverse contexts. It's being employed in a host of applications, including business intelligence, financial predictions, and social impact studies, and there has been an ever-growing interest in this technology too. However, the reality remains that this technology has several challenges to overcome in order to realize its full capabilities: recognizing sarcasm, coping with ambiguous contexts, addressing data bias, and making models more explanatory and understandable. There is a need for a more robust, diverse, and clear user data policy in sentiment analysis as well. There is a need to focus on user-friendly and inclusive Explainable AI and ensure multiple language support and more ethical data usage in sentiment analysis models too.

References :

1. Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. Presented at the 2019 NAACL-HLT Conference, pp. 4171–4186.
2. Zhang, L., Wang, S., & Liu, B. (2018). Deep Learning for Sentiment Analysis: A Survey. WIREs Data Mining and Knowledge Discovery, 8(4), e1253.
3. Hutto, C. J., & Gilbert, E. (2014). VADER: A Rule-based Model for Sentiment Analysis of Social Media Text. ICWSM, 8(1), 216–225
4. Cambria, E., Schuller, B., Xia, Y., & Havasi, C. (2013). New Avenues in Opinion Mining and Sentiment Analysis. IEEE Intelligent Systems, 28(2), 15–21.
5. Liu, B. (2012). Sentiment Analysis and Opinion Mining. Morgan & Claypool Publishers.
6. Pang, B., & Lee, L. (2008). Opinion Mining and Sentiment Analysis. Foundations and Trends in Information Retrieval, 2(1–2), 1–135
7. Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. Neural Computation, 9(8), 1735–1780.
8. Vaswani, A., et al. (2017). Attention Is All You Need. Advances in Neural Information Processing Systems (NeurIPS), 5998–6008.
9. Howard, J., & Ruder, S. (2018). Universal Language Model Fine-tuning for Text Classification (ULMFiT). Proceedings of ACL, 328–339.
10. Liu, Y., et al. (2019). RoBERTa: A Robustly Optimized BERT Pretraining Approach. arXiv preprint arXiv:1907.11692.
11. Conneau, A., et al. (2020). Unsupervised Cross-lingual Representation Learning at Scale (XLM-R). Proceedings of ACL, 8440–8451.

ISSN : 2348-7143
March 2026
RESEARCH JOURNEY International Multidisciplinary E-Research Journal
Impact Factor (SJIF) - 6.625 | ***Special Issue 389 :*** *Advanced Computing and Artificial Intelligence (ACAI)*

AI-Driven Public Health Chatbots for Disease Awareness and Real-Time Emergency Coordination

Sujit Deshmukh
B.Tech. CSE IEP Microsoft, Parul University of Engineering and Technology, Vadodara, Gujarat. India.

Abstract

This study presents an AI-Driven Public Health Chatbot for Disease Awareness, an intelligent and multilingual healthcare platform designed to bridge the gap between patients and reliable medical information. The system aims to provide a unified solution that integrates disease awareness, appointment management, doctor communication, and emergency response services within a single cohesive platform, thereby overcoming the limitations of fragmented healthcare tools. The proposed platform targets diverse and underserved populations by incorporating a healthcare-specific artificial intelligence model capable of supporting multiple regional languages. It delivers accurate, curated, and professionally validated health information to reduce reliance on unverified sources, particularly in rural and resource-constrained communities. A novel feature of the system enables hospitals to access real- time patient data during emergencies, enhancing response efficiency and potentially saving lives. The platform also supports direct doctor chat, real-time SMS and WhatsApp alerts, multilingual assistance, and emergency data sharing. Real-time chatbot communication is facilitated through WebSockets, and hospital discovery is enabled via Google Maps API integration. Authentication mechanisms include NextAuth.js, Passport.js, and Stack-Auth, ensuring robust security through multi-factor authentication and end-to-end encryption. The platform is designed to comply with GDPR and HIPAA standards, and the AI system is validated by certified medical professionals with a clear disclaimer that it serves as an informational tool rather than a substitute for clinical diagnosis. The proposed AI-driven healthcare platform offers a secure, scalable solution to enhance public health awareness, strengthen emergency response, and improve access to reliable medical information.

Keywords: AI-driven healthcare, public health Chatbots, multilingual health system, emergency response, healthcare security.

Introduction:

The research work is an AI-powered healthcare chatbot platform. It can be used to verified diseases awareness and reduce misinformation from social media, offering multilingual medical support every time, enabling real-time communication with doctors, and supporting emergency medical response and

The most important thing of this work is to provide hospital search functionality. Psychoeducation and interventions through brief conversations via existing communication channels (i.e., SMS text messaging and Facebook Messenger).[1] Conversational agents are one of the numerous digital tools being implemented in the healthcare industry to tackle present healthcare issues, such as provider shortages that limit the accessibility and availability of healthcare services. [2–4]. Conversational agents have been created to assist both the public and healthcare professionals in a variety of health-related fields. Particular applications include health condition screening, triage, counseling, at-home health management assistance, and healthcare professional training. [5,6–7]. Conversational agents can assist populations with limited access to healthcare or low health literacy because phone, mobile, and online platforms are widely available. [8,9]. are also a promising tool for the development of patient-centered care because of their accessibility. Conversational agents can encourage users to participate in their own medical care. [10,11]. They discovered that superior proof of this review also revealed limited efficacy and patient safety. Similarly, they observed that while studies generally reported high levels of satisfaction, the most frequent problems with conversational agents were related to poor dialogue management or language comprehension, which is in line with our findings. [12]. The development of increasingly sophisticated artificial intelligence agents has been facilitated by the substantial advancements in natural language processing capabilities in the subsequent decades. Chatbots, embodied conversational agents, and virtual patients are just a few of the many varieties of conversational agents that have been created using natural language processing (NLP) and are available via computer, mobile device, phone, and numerous other digital platforms. [13–16]. As conversational agents are often touted as having the potential to reduce burdens on healthcare resources, evaluations of the implications of the agents for improved healthcare provision and reduced resource demand also need to be assessed.

Qualitative User Perceptions:

For this research, the target users are the Rural population, Urban users seeking quick awareness, Hospitals and emergency responders.

Technical Stack:

In my work, the technical stack for Backend has Express.js, MongoDB, Rest APIs, JWT Authentication, Role – Based Access Control (RBAC), Frontend such as EJS, Bootstrap, WebSocket (real-time communication), and Google Maps API (hospital search), similar to AI/NLP is Intent classification

Named Entity Recognition (NER), Symptom Extraction, and Response generation logic for an optional Transformer -based model and Fine-tuned NLP model have been used. Feedback from users in five

Studies expressed a preference for interactivity, with users in one study noting that they liked the interactivity of the chatbot [17-19].

Methods:

First of all, in the target area, I collected the data and processed it properly. Each sample was carefully labeled and then sent for model training. The data was classified using an intent classification model, and the system integrated all the data during the testing phase. Finally, the processed data was deployed as part of the deployment process. These themes were subsequently searched with the structure: conversational agent and health application. During the screening process, studies of conversational agents that were not capable of interacting with human users via unconstrained natural language processing (NLP) were excluded. These included conversational agents that only allowed users to select from predefined options or agents with pre-recorded responses, which did not adapt to subsequent user responses. The basis of this exclusion is that, without the capability of using NLP computational methods, technologies were rudimentary and not advancing the aims of artificial intelligence for autonomous computational agents. As many studies did not explicitly state whether the investigated agent was capable of NLP, a description in the paper of the conversational agent allowing free text or free speech input was used as an indicator for NLP, and these studies were included. This is compared with the system with existing healthcare AI platforms, such as Babylon Health and Ada Health.

Impact Analysis:

The implementation of the proposed chatbot-based healthcare model demonstrated significant social and operational impact. The system effectively reduced healthcare misinformation by delivering structured and verified responses through intent classification. It improved rural awareness and encouraged digital healthcare adoption among underserved populations. Economically, the chatbot minimized unnecessary hospital visits by providing preliminary guidance and symptom-based support, thereby reducing patient expenses and optimizing healthcare resources. Operationally, it decreased the workload of medical professionals by automating routine queries and initial screening processes. Overall, the proposed model proved to be scalable, cost-effective, and efficient for enhancing healthcare accessibility and service management.

Limitations and future directions:

The suggested chatbot-based healthcare model has some drawbacks despite its encouraging results. Its reliance on consistent internet connectivity is its main drawback, which could limit accessibility in isolated or poorly connected areas. Furthermore, misinterpretation based on AI is a possible risk. The future scope of the proposed system is extensive and adaptable across multiple domains. The model can be enhanced through integration with blockchain-based health record systems to ensure secure, transparent, and tamper-proof patient data management. Incorporating voice assistant functionality would improve accessibility for elderly and low-literacy users. Further expansion may include integration with IoT-enabled wearable devices for real-time health monitoring and automated data synchronization. Predictive disease outbreak analytics using large-scale data modeling can strengthen public health surveillance and early warning systems. Additionally, integration with government healthcare infrastructures would enable large-scale deployment, improving healthcare delivery efficiency and policy implementation .

Conclusion:

The purpose of this systematic review was to compile data on the effectiveness, usability, and user satisfaction of conversational agents in the medical field. The results show that conversational agents are practical, efficient, and promising tools for assisting digital healthcare services, with generally positive outcomes. The results show that conversational agents are practical, efficient, and promising tools for assisting digital healthcare services, with generally positive outcomes.

References :

1. Adams, W. G., Phillips, B. D., Bacic, J. D., Walsh, K. E., Shanahan, C. W., & Paasche-Orlow, M. K. (2014). Automated conversation system before pediatric primary care visits: A randomized trial. Pediatrics, 134(3), e691–e699.
2. Abosamak, N., Namoos, A., Deeb, J. G., & Gal, T. (2025). Utilization of an AI-powered chatbot for enhancing oral cancer awareness among African Americans: Expert feedback on usability. AMIA Summits on Translational Science Proceedings, 2025, 42.
3. Bickmore, T. W., Pfeifer, L. M., Byron, D., Forsythe, S., Henault, L. E., Jack, B. W., & Paasche-Orlow, M. K. (2010). Usability of conversational agents by patients with inadequate health literacy: Evidence from two clinical trials. Journal of Health Communication,

15(Suppl 2), 197–210.

4. Campillos-Llanos, L., Thomas, C., Bilinski, É., Zweigenbaum, P., & Rosset, S. (2020). Designing a virtual patient dialogue system based on terminology-rich resources: Challenges and evaluation. Natural Language Engineering, 26(2), 183–220.
5. Chang, P., Sheng, Y. H., Sang, Y. Y., & Wang, D. W. (2008). Developing a wireless speech-and touch-based intelligent comprehensive triage support system. CIN: Computers, Informatics, Nursing, 26(1), 31–38.
6. Fitzpatrick, K. K., Darcy, A., & Vierhile, M. (2017). Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): A randomized controlled trial. JMIR Mental Health, 4(2), e7785.
7. Hudlicka, E. (2013). Virtual training and coaching of health behavior: Example from mindfulness meditation training. Patient Education and Counseling, 92(2), 160–166.
8. Inkster, B., Sarda, S., & Subramanian, V. (2018). An empathy-driven conversational artificial intelligence agent (Wysa) for digital mental well-being: Real-world data evaluation mixed-methods study. JMIR mHealth and uHealth, 6(11), e12106.
9. Isaza-Restrepo, A., Gómez, M. T., Cifuentes, G., & Argüello, A. (2018). The virtual patient as a learning tool: A mixed quantitative qualitative study. BMC Medical Education, 18(1), 297.
10. Kocaballi, A. B., Berkovsky, S., Quiroz, J. C., Laranjo, L., Tong, H. L., Rezazadegan, D., & Coiera, E. (2019). The personalization of conversational agents in health care: Systematic review. Journal of Medical Internet Research, 21(11), e15360.
11. Laranjo, L., Dunn, A. G., Tong, H. L., Kocaballi, A. B., Chen, J., Bashir, R., & Coiera, E. (2018). Conversational agents in healthcare: A systematic review. Journal of the American Medical Informatics Association, 25(9), 1248-1258
12. Luxton, D. D. (2020). Ethical implications of conversational agents in global public health. Bulletin of the World Health Organization, 98(4), 285.
13. Ly, K. H., Ly, A. M., & Andersson, G. (2017). A fully automated conversational agent for promoting mental well-being: A pilot RCT using mixed methods. Internet Interventions, 10, 39–46.
14. Stephens, T. N., Joerin, A., Rauws, M., & Werk, L. N. (2019). Feasibility of pediatric obesity and prediabetes treatment support through Tess, the AI behavioral coaching chatbot. Translational Behavioral Medicine, 9(3), 440–447.
15. Vaidyam, A. N., Wisniewski, H., Halamka, J. D., Kashavan, M. S., & Torous, J. B. (2019). Chatbots and conversational agents in mental health: A review of the psychiatric landscape. The Canadian Journal of Psychiatry, 64(7), 456–464.
16. van Heerden, A., Ntinga, X., & Vilakazi, K. (2017). The potential of conversational agents to provide rapid HIV counseling and testing services. In Proceedings of the 2017 International Conference on the Frontiers and Advances in Data Science (FADS) (pp. 80–85). IEEE.
17. Zhang, Z., & Bickmore, T. (2018). Medical shared decision making with a virtual agent. In Proceedings of the 18th International Conference on Intelligent Virtual Agents (pp. 113–118).

RESEARCH JOURNEY International Multidisciplinary E-Research Journal
Impact Factor (SJIF) - 6.625 | ***Special Issue 389 :*** *Advanced Computing and Artificial Intelligence (ACAI)*
ISSN : 2348-7143
March 2026

Prediction for a Reliable Blood Supply Chain using Machine Learning algorithms

Trupti A.Chaudhari
Assistant professor R. C. Patel Educational Trust's Institute of Management Research and Development, Shirpur
Asha R.Patil
Assistant professor R. C. Patel Educational Trust's Institute of Management Research and Development, Shirpur
Dr. Manoj B. Patel
Assistant professor R. C. Patel Educational Trust's Institute of Management Research and Development, Shirpur

Abstract:

In India the number of deaths from road accidents is increasing steadily from last 5 years and blood is needed for accident victims roughly every 90 seconds. However, traditional blood banking systems often fail to meet emergency demands due to blood shortages and unpredictable donor availability. This system's primary objective is to establish a dependable blood supply chain that links blood banks, non-governmental organizations, and blood donors in the closest area to one another in order to guarantee prompt assistance during emergencies and maybe save patients' lives. Donor authentication and registration, blood group-based filtering, real-time blood availability checking, automated matching between donors and recipients based on compatibility and geography, and secure document uploads are just a few of the system's important features. Despite of Blood Supply Predictor is currently in its early phases of research, it has the potential to alter transfusion services by solving concerns like data loss, patient data privacy, and cost-effectiveness. Donors can promise to donate blood on our system, creating a network of supporting community members.This BSPS proposed system using machine learning algorithms like K-nearest neighbor, Random Forest, Decision Trees, and Logistic Regression. Our Blood Management Application not only transforms blood supply chain management but also ensures a quick, reliable, and precise preventive reaction through applying the strength of AI-driven predictions and targeted requests.

Keywords:

Blood Supply Chain, Machine Learning, blood transfusions

I. Introduction

From 2020 to 2023, India recorded a steep upward trend in road-accident deaths, reaching the highest fatality levels in recent years. The table presents year-wise data on the number of persons killed in road accidents in India from 2020 to 2023, as reported by the Ministry of Road Transport and Highways (MoRTH). Transfusions of blood are an essential component of the healthcare system in such instances, saving millions of lives. Artificial alternatives that can reliably replace the need for donated human blood are still being researched by researchers. Therefore, the use of blood and its components remains necessary to treat serious medical conditions.[15]. Medical care and blood transfusions may extend a patient's life at the cost of significant blood consumption, even if some patients may not recover. Blood transfusions are an essential part of modern medicine. [2]. To provide reliable blood supply chain to interconnect the blood banks, NGO, blood donors at nearest location is very crucial. In that case, Blood Supply Predictor System (BSPS) can play a significant role. Due to inefficient manual systems, disconnected communication, and limited access to real-time data, the blood donation, recipient-donor matching, and availability tracking processes continue to be inadequate. These restrictions decrease voluntary donor involvement in addition to delaying critical treatments. [17]. Millions of units of blood are transfused each year, according to the WHO, yet there is still an enormous gap between supply and demand. Although it faces a number of challenges the blood supply chain (BSC) is essential for supplying sufficient and safe blood.

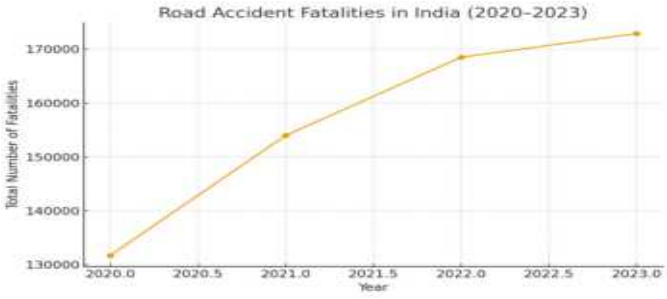
Considering the fact that the blood supply chain has previously been the subject of several studies, additional research is needed to decrease the supply chain's uncertainty in terms of identifying blood donors' readiness. This paper attempts to close this gap by using a machine learning approach to improve the reliability of the blood supply network.Blood banks can reduce loss and guarantee a consistent supply of necessary blood groups by using the Blood Supply Predictor System (BSPS) to make data-driven decisions.

This BSPS proposed system using machine learning algorithms like K-nearest Neighbour, Random Forest, Decision Trees, and Logistic Regression. These ML algorithms play a crucial role in the prediction and optimization of blood donation processes. Random Forest (RF) combines multiple decision trees to improve prediction accuracy. It helps in classifying potential donors based on past donation patterns and identifying factors influencing donation frequency. Decision Trees (DT) provide interpretable rules for donor classification and behavior prediction, allowing blood banks to segment donors based on age, health history, and donation preferences. Logistic

Regression is used for binary classification; logistic regression can predict whether an individual is likely to donate blood within a given timeframe based on historical data and demographic features. By integrating these ML techniques, blood banks can automate donor identification, enhance retention strategies, and optimize blood supply chains [9]. The main goal of this system is to connect blood banks, NGOs, and nearby donors to ensure timely blood availability during emergencies and help save lives.

Road Accident Fatalities in India (2020–2023)

Sr. No.	Year	Number of persons killed in road-accidents. Source(MoRTH – Minstry Of Road Transport & Highway)
1	2020	131714
2	2021	153,972
3	2022	168,491
4	2023	172,890
5	2024	not yet published



II. Literature Review

In today's age of communication and information, where people can order pizza online and be certain to receive it within 30 minutes, book movie tickets online, plan vacations online, and book trains online,but not obtain blood avalability information with just a single click. The world has evolved into a global village where everything is done online. The market provides an extensive range of web-based solutions for people's comfort. However, a human being cannot survive without blood, therefore offering a machine learning-based blood bank management information system is just one more way to help mankind [6]. The current method for handling emergencies is a systematic procedure: hospitals first examine their inventory, and if the necessary blood is not available, they get in touch with local blood banks. If the search is unsuccessful, it is expanded to include regular donors, remote facilities, and patient relatives. Although being well-known, this approach is certainly inefficient and lengthy, frequently leading to delays in critical medical care[13]. Invention of Methods for Expanding the Authority of Blood Search Operations: To increase the scope of blood search operations, creative strategies are needed. This could entail using incentives, community outreach initiatives, and social media platforms to encourage frequently blood donations. Over the years, several researchers have worked on building and improving blood bank management systems. Most of these systems do not support to reduce the uncertainty of the blood supply chain in terms of estimating the readiness of blood donors. Inventory management of blood products is crucial in hospital operations due to their perishable nature and demand uncertainty. Also, the short shelf-life of blood components such as platelets can significantly lead to wastage or expiration if excess platelets are kept. A number of studies emphasize the need to predict blood supply and demand accurately to reduce shortages and wastage. An overview of relevant research and the weaknesses of various systems can be found below.

Blood bank management systems have been developed and improved throughout the years by a number of researchers. However, the majority of these systems lack regional services, intelligent donor matching, and real-time monitoring, all of which are especially vital in specific regions.The data from current blood donors will be linked with social media to create an application as part of the on going effort. In this manner, it would help a nation like India's vast populace[1].blood banks to promptly manage necessary blood collection by using blood supply prediction from the present data. Blood donors might not always be available. To avoid a blood shortage, blood banks may be able to collect blood earlier if they take the required safety measures. According to this study logistic regression offers the highest accurate forecast. [2]. Real-time information about the blood stock that is available in each of the linked blood banks will be provided via the system. Additionally, it will have an AI chatbot to guide users in real time, and blood donor registration will be completed via GPS and PIN code verification, which will speed up the process of finding the closest available blood donor.[3].The early Blood Bank Management System mainly handled donor registration and blood group availability using a basic database, but it lacked real-time updates and automation. Later, the Blood Bank Management and Inventory Control Database System improved data organization through structured databases like MySQL, making inventory management more systematic; however, it still did not support AI features or geographic filtering. The Blood Donation Management System introduced web or Android-based platforms for donor scheduling and data management, but it lacked personalized interaction and location-based donor search. The Cloud-Based Blood Bank Management System further enhanced accessibility by using cloud technologies such as Firebase for multi-device access, yet it

did not include smart blood matching, GPS verification, or AI-powered communication. The most recent system added improvements like better user interface and form validation, but it remained largely manual and did not incorporate intelligent automation, real-time communication, or effective emergency response features. Overall, despite gradual technological improvements, these system still lack advanced intelligent and automated functionalities

Author/ Year	System Type	Technology	Main Features	Major Limitations
Sumazly Sulaiman [2013]	Early System	Basic Database	• Donor registration • Blood group availability check	• No real-time updates • No automation
Aman Shah [2021]	Inventory Control Database Management System	MySQL / Structured Database	• Organized data management • Systematic inventory control	• No AI features • No geographic filtering • No intelligent updates
Devanjan K [2021]	Blood Bank and Donor Management System	Android / Web App	• Donor scheduling • Donor data management	• No personalized interaction • No location-based donor search
Abhilash Sharma [2023]	Cloud-Based System	Firebase / Cloud Technology	• Multi-device access • Cloud storage	• No smart blood matching • No GPS verification • No AI communication
AJITH Kumar P [2024]	Recent Improved Version	Web System with Improved UI	• Better user interface • Form validation	• Manual processes • No intelligent automation • No real-time communication • No emergency response

Over time, blood bank systems improved from simple database programs to cloud-based systems with better design and access. However, they still do not include advanced features like AI support, smart blood matching, GPS-based donor search, automatic alerts, or proper emergency response systems.

III. Research Methodology

1. Approach

The methodology is divided into two main approaches:

- **Quantitative Approach:**
  This involves analysing numerical data from the blood supply chain to generate statistical insights and machine learning–based predictions. The goal is to understand patterns and forecast demand using quantitative data analysis.
- **Experimental Method:**
  Under controlled experimental conditions, various machine learning algorithms are trained and compared. This allows the identification of the most effective predictive models by testing their performance systematically.

2. Data Collection

Data is collected from multiple input sources essential for building the prediction models. These sources include:

- **Hospital Blood Banks :** Data on blood inventory and usage from hospital facilities.
- **Blood Banks :** General blood bank data encompassing inventory and distribution.
- **Social Media (Volunteer Organizations):** Information gathered from social media platforms and volunteer organizations to capture donor behaviour and availability.
- **College Donor Clubs:** Data related to blood donation activities and schedules from college-based donor clubs.

3. Machine Learning Models

Once data is collected, it undergoes several pre-processing steps:

- **Data Cleaning & Normalization:** Ensures the data is accurate, consistent, and formatted correctly for analysis.
- **Feature Engineering:** Involves selecting and transforming relevant variables to improve model performance.
- **Data Splitting:** The dataset is divided into a training set (used to train the models) and a test set (used to evaluate model performance).

Following machine learning algorithms used for prediction:

- K-Nearest Neighbour (KNN)
- Random Forest

- Decision Trees
- Logistic Regression

The methodology for the proposed system is given as follows:

IV. Proposed system

The proposed system represents an intelligent, machine-learning–driven framework for ensuring a reliable and timely blood supply during emergency situations. The implementation of the Blood Supply Predictor System (BSPS) will enable blood banks to transition from reactive management to intelligent, data-driven decision making. By accurately predicting demand for different blood groups, BSPS will significantly reduce blood wastage caused by expiration while ensuring the timely availability of critical blood units. The proposed system leverages advanced machine learning algorithms, including K-Nearest Neighbour (KNN), Random Forest, Decision Tree, and Logistic Regression, to analyze historical data, identify demand patterns, and support optimal inventory planning. This predictive approach will enhance operational efficiency, improve emergency preparedness, and strengthen the overall reliability of the blood supply chain.The process begins when an accident occurs and an emergency call or report is generated. This information is immediately routed to the hospital emergency response team, which assesses the patient's condition and predicts the type and quantity of blood required. At this stage, machine learning prediction models play a critical role. Using historical accident data, seasonal trends, hospital admission rates, and blood usage patterns, ML algorithms forecast real-time blood demand. The system then automatically checks blood bank inventory, where ML-based optimization models assess stock levels, expiry dates, and future demand to decide whether existing inventory can meet the requirement. If the

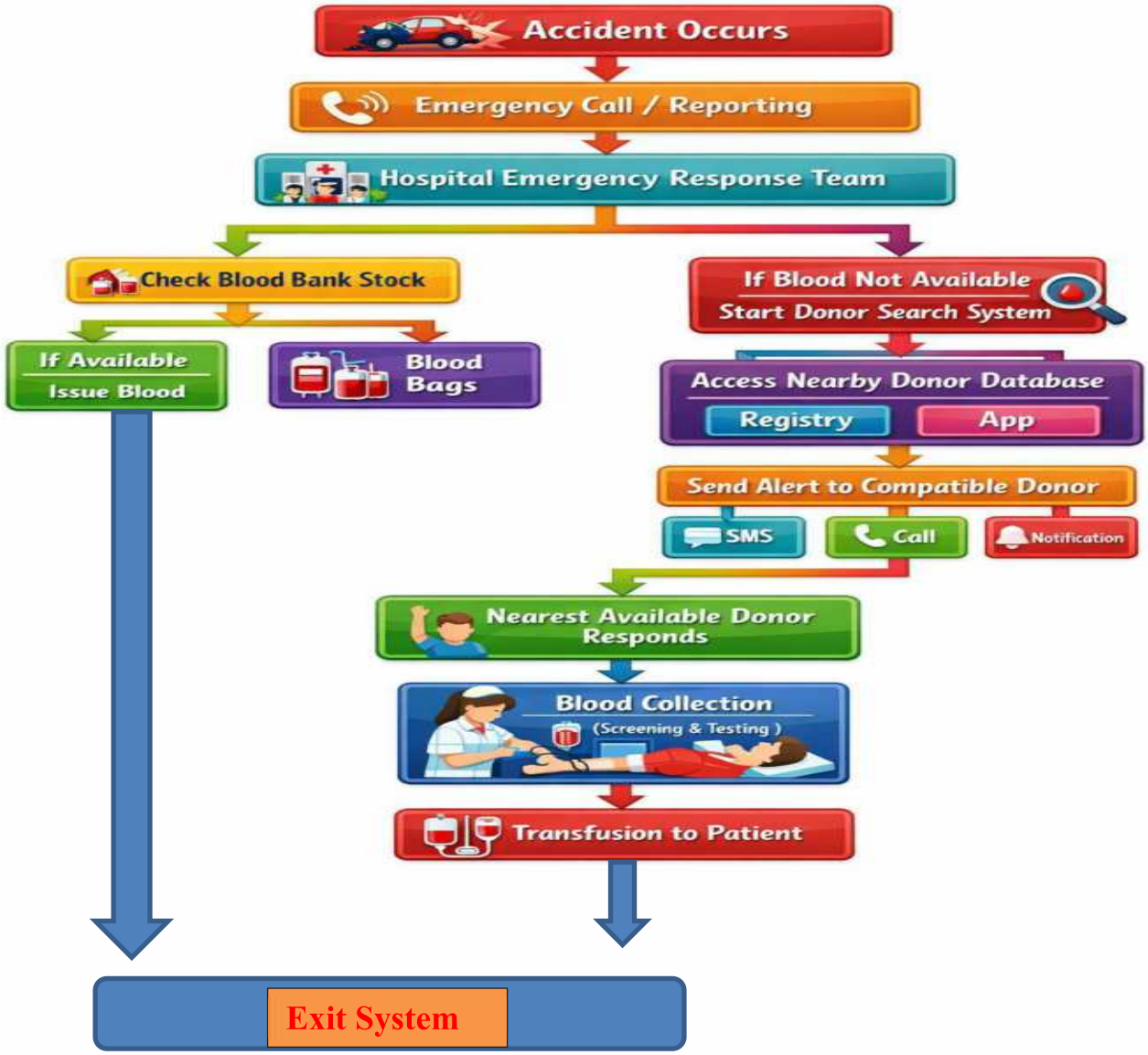


Figure 2: Blood Supply Predictor System (BSPS)

required blood is available, it is immediately issued to the patient, minimizing response time. If blood is not available or predicted to be insufficient, the system activates an ML-enabled donor search module. This module uses classification and ranking algorithms to identify the most suitable donors based on blood group compatibility, donor availability history, location proximity, and past response behaviour. The system sends automated alerts (SMS, calls, app notifications) to prioritized donors. Machine learning models continuously learn from donor response patterns to optimize alert timing and channel selection, increasing the probability of rapid donor response. Once a nearest compatible donor responds, the system coordinates donor arrival using predictive routing and scheduling. Collected blood undergoes screening and testing, after which it is added to the usable inventory. Predictive models update inventory forecasts in real time, ensuring better planning for future emergencies. Finally, the blood is transfused to the patient, completing the emergency response cycle. Overall, the integration of machine learning transforms the traditional blood supply chain into a predictive, adaptive, and optimized system. By forecasting demand, optimizing inventory, and intelligently managing donor engagement, the system reduces shortages, minimizes wastage, and enhances the reliability and efficiency of emergency blood supply operations.

V. Result and Discussion

The implementation of the Blood Supply Predictor System (BSPS) enhances the efficiency, reliability, and responsiveness of blood supply chain operations by leveraging machine learning–based prediction and optimization .The system improves demand forecasting accuracy, reduces blood wastage, optimizes inventory management, accelerates emergency response, and strengthens donor engagement, resulting in a more reliable and data-driven blood supply chain.

- **Improved Demand Prediction:** Machine learning algorithms (KNN, Random Forest, Decision Tree, Logistic Regression) enable accurate forecasting of blood demand for all blood groups, reducing uncertainty and supporting proactive inventory management.
- **Faster Emergency Response:** Real-time demand forecasts and automated inventory checks allow rapid decision-making. The ML-enabled donor search and alert system ensures quick donor mobilization, reducing delays in transfusion.
- **Intelligent Donor Engagement:** Donor ranking and response prediction prioritize donors with higher availability, improving donor participation and collection efficiency.
- **Enhanced Supply Chain Reliability:** The predictive and adaptive approach strengthens the blood supply chain by reducing shortages, improving preparedness, and ensuring continuous availability of required blood groups.

VI. Conclusion and Future Work

- Development of a mobile app for instant donor alerts, emergency SOS requests, and quick communication.
- Implementation of block chain technology for secure, tamper-proof, and transparent data management.
- Use of advanced AI models like deep learning and neural networks to enhance prediction accuracy and forecast blood demand.

References :

1. Dr.Srinivasa Rao, “A Machine Learning-Based Blood Donor Recommendation System to Enhance Blood Donation Efficiency” jan 2024
2. NazmusSakib, “A smart machine learning prediction model for forecasting supply in blood banks upplychain” 2023
3. A. Clemen Teena, “A Study on Blood Bank Management” 2014
4. Shruti Baraskar, “AI Based Blood Bank and Donor Management System”, 2014
5. Ms. Sanika Lakade “AI-Driven Blood Bank Management System: Improving Blood Supply Chains with Smart Technology”May 2025
6. Vikas Kulshreshtha, “Blood Bank Management Information System in India”
7. Nagarapu Nalini Krupa, “Blood Bank Management System Using Machine Learning”March2025
8. Srusti Satapathy, “Blood Donation and Prediction through Machine Learning Techniques “Jan 2022
9. Jeong Kwon, “Development of blood demand prediction model using artificial intelligence based on national public big data”,2024
10. Pooria Bagher Niakan, “An Integrated Supply Chain Model for Predicting Demand and Supply and Optimizing Blood Distribution”Dec.2024
11. Clement Twumasi, “Machine learning algorithms for forecasting and backcasting blood demand data”Jan2021
12. Ambarish Shashank , “Optimizing Blood Supply Chains AI Enabled smart blood supply”,oct.2023
13. Viraj Prabhu, “RedRespond : Real-Time Emergency Blood Donation Platform Aug.2025
14. T. Senthil Kumar, “Smart Blood Management and Tracking System “, May2019
15. Ben Elmir, “Smart platform for data blood bank management forecasting demand in blood

supply chain using machine learnin"
16. Laboni Nayak, "Smart System For Blood Donation And Availability Finder" June 2025
17. Abin Varghese, "Technological advancements, digital transformation, and future trends in blood transfusion services "March2024
18. Vamsi Krishna Tatikonda, "BLOODR: blood donor and requester mobile application"2027
19. Mahesh Mahajan, "Blood Donation Application Using Machine Learning"
20. Dr. Vijay Shelake, "Enhancing Donor Eligibility Criteria using Machine Learning "Feb2025
21. Sumazly Sulaiman, Abdul Aziz K. Abdul Hamid, Nurul Ain Najihah Yusri. "Development of a Blood Bank Management System". Procedia Social and Behavioral Sciences, 195, 2008-2013.
22. Aman Shah,Dev Shah,Devanshi Shah, Daksh Chordiya, Nishant Doshi, Rudresh Dwivedi. "Blood Bank Management and Inventory Control Database System". Procedia Computer Science,198, 404-409.
23. Devanjan K. Srivastava,Utkarsh Tanwar, M.G.Krishna Rao,Priya Manohar, Balraj Singh. A Research Paper on Blood Donation Management System. International Journal of Creative Research Thoughts (IJCRT), 9(5).
24. Abhilash Sharma, Abhishek Attri, Nikhil Kesarwani Shivam Kasaudhan, Gunjan Agarwal. Cloud-Based Blood Bank Management System. International Journal of Novel Research and Volume 8, Issue 5 May 2023,ISSN:2456-4184.
25. AJITHkumar P, Sarath KUMAR K.sasi Kumar C. "A Research Paper on Blood Bank and Donor Management System". International Journal of New Innovations in Engineering and Technology(IJNIET),Volume 24,issue 1,March 2024,ISSN:23196319

RESEARCH JOURNEY International Multidisciplinary E-Research Journal ISSN : 2348-7143
Impact Factor (SJIF) - 6.625 | ***Special Issue 389 :*** *Advanced Computing and Artificial Intelligence (ACAI)* March 2026

AI-Based Multimodal Emotion Prediction System for Enhancing Student Engagement in Online Learning

Mr.Shaikh Alfarhan Sk Farooque

Mr.Shaikh Mohammad Awais Ashfaque
Institute of Management Research and Development, shirpur

Abstract

The rapid growth of online learning has transformed modern education, offering flexibility and accessibility to students worldwide. However, virtual classrooms often lack the ability to recognize and respond to students' emotional states, which significantly influence engagement and academic performance. Emotions such as confusion, stress, boredom, and motivation directly affect learning outcomes, yet they remain largely unnoticed in digital environments.This study proposes an AI-based multimodal emotion prediction system designed to enhance student engagement in online learning platforms. The proposed framework integrates multiple data modalities, including text analysis, speech patterns, and facial expression recognition, to improve the accuracy and reliability of emotion detection. By combining these inputs, the system aims to provide a more comprehensive understanding of students' emotional conditions compared to single-modality approaches.To examine the relevance and acceptance of such technology, a survey-based analysis was conducted with 103 respondents from diverse academic backgrounds. The findings indicate strong support for emotion-aware systems in education and highlight the perceived importance of early emotional detection for improving learning effectiveness and mental well-being.

Although the system remains conceptual, the study establishes a foundation for developing adaptive, emotion-aware digital learning environments. The integration of artificial intelligence in education has the potential to create more responsive, personalized, and student-centered online learning experiences.

Introduction

The rapid expansion of online learning has transformed the way education is delivered, making it more flexible and easily available. However, virtual classrooms often fail to capture students' emotional states, which are highly important in learning performance. In traditional classrooms, teachers can observe facial expressions, voice tone, and body language to understand students' engagement, but such emotional cues are limited in online environments.Emotions such as stress, confusion, boredom, and motivation directly influence students' participation and academic performance. To address this challenge, Artificial Intelligence (AI) can be used to detect emotions through text analysis, speech recognition, and facial expression recognition. A multimodal approach that combines these inputs can improve accuracy and provide deeper insights into students' emotional well-being.This study proposes an AI-based multimodal emotion prediction system to enhance student engagement in online learning. Through survey-based analysis, the research highlights the need for emotion-aware technologies that can support educators in building more interactive and effective digital learning environments.

Literature Review

Recent research shows that multimodal emotion recognition using Artificial Intelligence has improved significantly through deep learning techniques. Studies by Lian et al. (2023) and Zhang & Tan (2024) highlight that combining text, speech, and facial expressions increases accuracy of emotional analysis compared to single-modality systems. In the field of education, researchers have explored emotion-aware systems in classrooms and online learning environments, showing that recognizing students' emotional states can improve engagement and learning effectiveness. However, most previous studies mainly focus on improving technical performance and model accuracy. The present study differs by emphasizing the practical application of a multimodal emotion prediction system specifically for online learning, along with analyzing public perception and acceptance through survey-based research. This approach connects technical development with real-world educational implementation.

Problem Statements

Main issues found in current online learning environments include:

- Lack of Emotional Awareness: Virtual classrooms cannot detect students' emotional states such as confusion, boredom, or stress in real time.
- Reduced Student Engagement: Without understanding learners' emotions, online platforms fail to maintain consistent engagement and motivation.
- Absence of Adaptive Response: Existing

systems do not adjust teaching methods based on students' emotional conditions.

- Limited Personalization: Online education platforms mainly focus on content delivery rather than emotional and psychological support.
- No Integrated Multimodal System: Most current solutions rely on single data sources and lack a comprehensive multimodal emotion prediction framework.

Objectives

- To analyze the need and importance of emotion prediction in online learning environments.
- To examine how students' emotional states affect their engagement and academic performance.
- To propose an AI-based multimodal emotion prediction system using text, speech, and facial expression analysis.
- To evaluate public awareness and acceptance of emotion detection technology through survey-based analysis.
- To highlight the potential benefits of integrating emotion-aware systems in digital education platforms.

Research and Methodology

Primary Research: Primary data for this study was collected through a structured online questionnaire designed using Google Forms. The survey received 103 valid responses from students and professionals belonging to different academic fields. The questions focused on awareness of emotion prediction technologies, the necessity of emotional monitoring in online learning, and the effectiveness of multimodal approaches. The responses provided direct insights into user perception, acceptance levels, and practical relevance of AI-based emotion detection systems.

Secondary Research: Secondary research was conducted through the review of existing research papers, journals, and scholarly articles related to artificial intelligence in education, emotion recognition systems, Natural Language Processing (NLP), speech emotion analysis, and facial expression detection techniques. This review helped in understanding current advancements, identifying research gaps, and developing the proposed multimodal conceptual framework.

Data Collection and Analysis

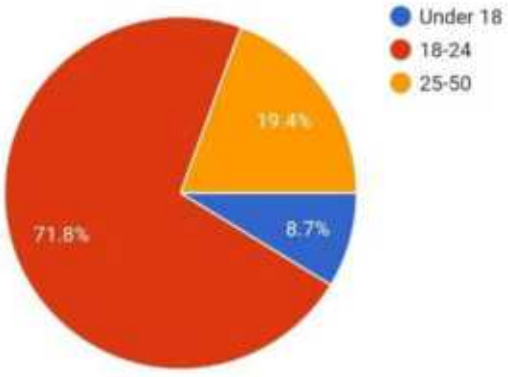


Figure 1: Age Distribution

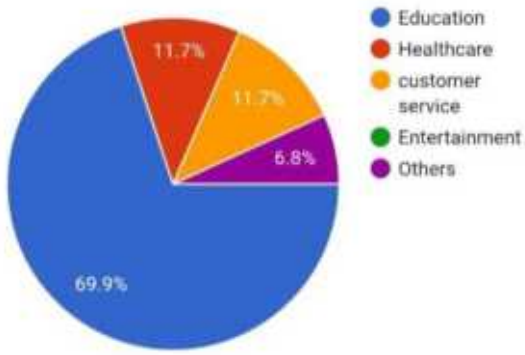


Figure 2: Field of Study / Profession

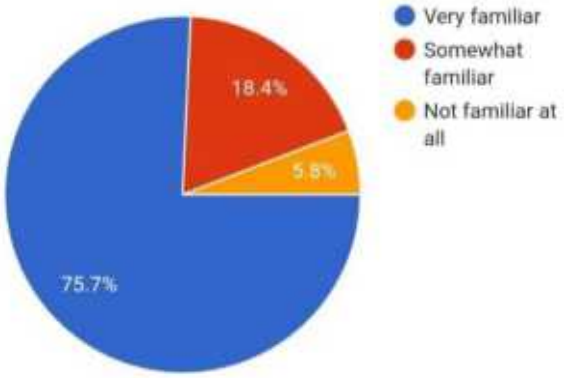


Figure 3: Familiarity with Emotion Prediction Systems

1. **Age Distribution:** The majority of respondents belong to the 18–24 age group, indicating strong participation from students who are the primary users of online learning platforms. The presence of respondents from other age groups adds diversity and improves the reliability of the collected data.
2. **Field of Study / Profession:** (Field Distribution) Most respondents are from the education sector, followed by healthcare and other professional fields. This diversity ensures that the study reflects opinions from individuals who are directly or indirectly connected to academic environments.
3. **Familiarity with Emotion Prediction Systems:** (Awareness Level) A significant proportion of respondents reported being familiar with emotion prediction technologies such as text, speech, and behavior analysis. This suggests that the participants had a reasonable understanding of the topic, making the responses informed and relevant.

Survey Findings

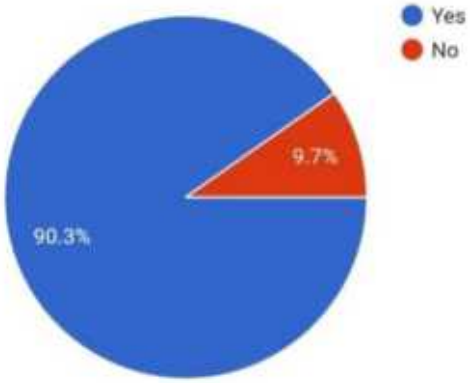


Figure 4: Should Emotions be Detected in Online Classes

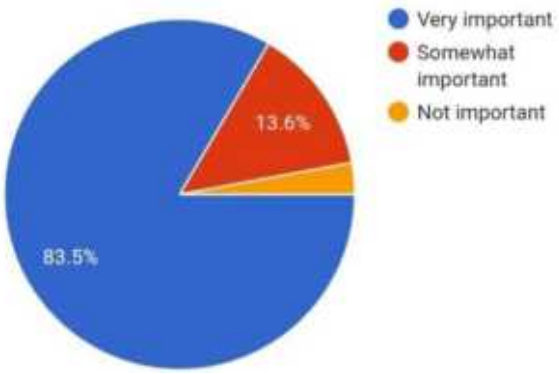


Figure 5: Importance of Predicting Emotions

4. Should Emotions be Detected in Online Classes: (Need for Emotion Detection) The findings indicate that a significant majority of respondents support the detection of students' emotions in online classes. This response highlights the growing awareness that emotional states directly influence engagement, participation, and academic performance in virtual learning environments.

5. Importance of Predicting Emotions: (Perceived Importance) Most participants rated emotion prediction as highly important, especially in sectors such as education and healthcare. This suggests that respondents recognize the broader applicability and value of emotion-aware systems beyond academic settings.

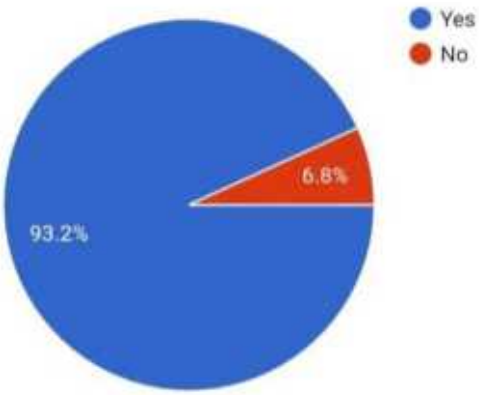


Figure 6: Early Detection and Mental Health Impact

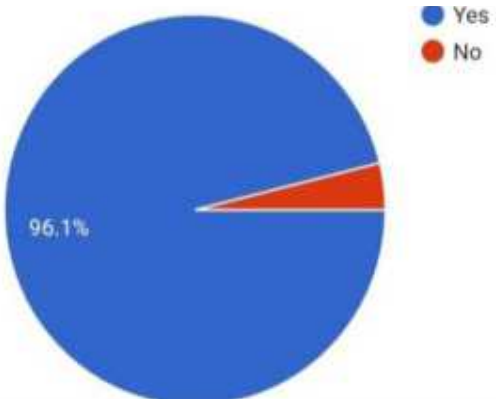


Figure 7: Effectiveness of Multimodal Approach

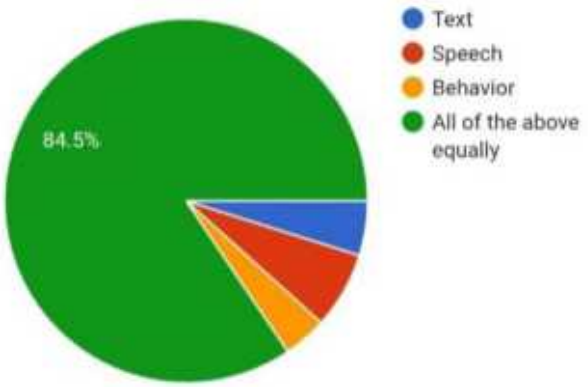


Figure 8: Most Effective Modality for Emotion Prediction

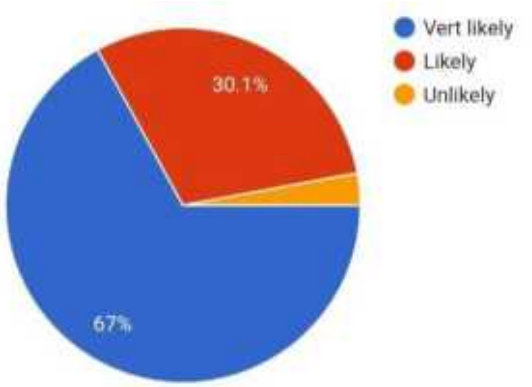


Figure 9: Likelihood of Using an Emotion Detection Tool

6. **Early Detection and Mental Health Impact:** (Impact on Mental Well-being)A large proportion of respondents agreed that early detection of emotional distress could help reduce mental health issues among students. This finding supports the role of AI-based systems in providing timely intervention and preventive support.
7. **Effectiveness of Multimodal Approach:** (Accuracy through Multimodal Integration) The majority of respondents believed that integrating multiple modalities such as text, speech, and facial expressions would significantly improve the accuracy of emotion prediction. This strongly validates the proposed multimodal framework of the study.
8. **Most Effective Modality for Emotion Prediction:** (Preferred Detection Method) Most participants selected the combined approach (text, speech, and facial expressions) as the most effective method for predicting emotions. This indicates user confidence in comprehensive data analysis rather than relying on a single input source.
9. **Likelihood of Using an Emotion Detection Tool:** (User Acceptance and Adoption)A majority of respondents expressed willingness to use an emotion-aware tool for monitoring emotional well-being in real time. This reflects positive user acceptance and practical feasibility of implementing such systems in online learning platforms.

Discussion

The survey results clearly show strong support for emotion prediction systems in online learning. A majority of respondents believe that emotional monitoring can improve student engagement and academic performance. The findings also indicate high awareness and acceptance of AI-based technologies.Most participants agreed that a multimodal approach combining text, speech, and facial expression analysis would provide better accuracy compared to a single method. This supports the proposed AI-based framework.Overall, the results validate the importance of integrating emotion-aware systems in digital education to create a more responsive and effective learning environment.

Critical analysis

Strengths

- The study is supported by primary data collected from 103 respondents.
- It addresses a relevant and emerging issue in online education.
- The research highlights strong public awareness and acceptance of emotion prediction technology.
- A multimodal AI-based framework (text, speech, facial analysis) is proposed for improved accuracy.
- The findings provide a foundation for future implementation and research in AI-driven education systems.

Limitations

- The sample size is limited to 103 respondents and may not represent a broader population.
- Convenience sampling was used, which may affect generalizability.
- The study is survey-based and does not include real-time system implementation or experimental validation.
- Ethical concerns such as privacy, data security, and accuracy of emotion detection need further investigation.

Future Scope

- The proposed framework can be developed into a real-time working system.
- A larger sample size can be used in future studies for better accuracy.
- Advanced AI and deep learning models can be

integrated to improve prediction performance.

- The system can be connected with existing online learning platforms.
- Ethical aspects like privacy and data security can be explored in greater detail.
- Experimental validation can be conducted to test system effectiveness.

Literature Review

Recent research shows that multimodal emotion recognition using Artificial Intelligence has improved significantly through deep learning techniques. Studies by Lian et al. (2023) and Zhang & Tan (2024) highlight that combining text, speech, and facial expressions increases emotion detection accuracy compared to single-modality systems. In the field of education, researchers have explored emotion-aware systems in classrooms and online learning environments, showing that recognizing students' emotional states can improve engagement and learning effectiveness.

However, most previous studies mainly focus on improving technical performance and model accuracy. The present study differs by emphasizing the practical application of a multimodal emotion prediction system specifically for online learning, along with analyzing public perception and acceptance through survey-based research. This approach connects technical development with real-world educational implementation.

Conclusion

The study confirms the importance of AI-based emotion prediction systems in online learning. Survey results show strong support for integrating emotion detection to improve student engagement and performance. A multimodal approach combining text, speech, and facial analysis can enhance accuracy and create more responsive digital learning environments.

References :

1. R. W. Picard, Affective Computing. Cambridge, MA, USA: MIT Press, 1997.
2. N. Sebe, I. Cohen, and T. S. Huang, "Multimodal emotion recognition," in Handbook of Pattern Recognition and Computer Vision, 2nd ed., Singapore: World Scientific, 2005, pp. 387–419.
3. S. K. Patil, J. Choudrie, K. Kotecha, and D. Vora, "A systematic review of applications of natural language processing with special emphasis on text-based emotion detection," Artificial Intelligence Review, vol. 56, no. 2, pp. 1021–1058, 2023.
4. S. Abdullah, S. Y. Ameen, and M. A. Sadeeq, "Multimodal emotion recognition using deep learning," Journal of Applied Science and Technology Trends, vol. 2, no. 1, pp. 1–9, 2021.
5. H. Zhang, Y. Liu, M. Jiang, and J. Chen, "Emotional artificial intelligence in educational settings: A systematic review and meta-analysis," Educational Psychology Review, vol. 37, 2025.
6. S. Salloum, K. Shaalan, and A. Al-Talhi, "Emotion recognition for enhanced learning: Using AI to detect students' emotions and adjust teaching methods," Smart Learning Environments, vol. 10, no. 1, 2023.

ISSN : 2348-7143
March 2026
RESEARCH JOURNEY International Multidisciplinary E-Research Journal
Impact Factor (SJIF) - 6.625 | ***Special Issue 389 :*** *Advanced Computing and Artificial Intelligence (ACAI)*

AI-Based College Enquiry Chatbot System

Mr. Vitthal Maharu Patil
Assistant Professor,
RCPET's Institute of Management Research and Development (IMRD), Shirpur, India.

Abstract

The AI-Based College Enquiry Chatbot System has been developed for R. C. Patel Educational Trust's Institute of Management Research and Development (IMRD), Shirpur, with a special focus on providing admission-related information for Undergraduate (UG) courses such as BCA, BBA, and BMS, and Postgraduate (PG) course MCA. The main objective of this system is to guide students, parents, and applicants by giving quick, accurate, and easy-to-understand responses about the admission process.

The chatbot provides complete details about eligibility criteria, required documents, entrance or merit process, fee structure, course duration, intake capacity, and important admission dates. It also answers common queries related to syllabus, career opportunities, scholarships, placements, and college facilities. This helps students make the right decision without the need to visit the college repeatedly.

The system is developed using Python and web technologies. It uses the SentenceTransformer MiniLM-L6-v2 model to understand the meaning of student queries and match them with the most relevant answers using semantic similarity and cosine scoring. Even if the question is asked in a different way or contains minor mistakes, the chatbot provides the correct response along with top suggestions.

The knowledge base is maintained using Excel, making it easy for staff to update admission information whenever required. The system is evaluated using performance measures such as accuracy, precision, recall, and F1-score to ensure reliable results.

Overall, this AI chatbot reduces the workload of the admission office, provides 24×7 support, improves communication with applicants, and makes the admission process for MCA (PG), BCA, BBA, and BMS (UG) simple, fast, and efficient. It helps the institute deliver timely information and enhances the overall admission experience for students.

Keywords:

Artificial Intelligence, College Enquiry Chatbot, Admission System, Natural Language Processing, Semantic Similarity, SentenceTransformer, Educational Institutions.

I. Introduction

In the digital era, students and parents expect quick access to admission-related information. Colleges receive many repetitive enquiries regarding eligibility criteria, admission procedure, fee structure, scholarships, and campus facilities. Handling these queries manually increases administrative workload and often leads to delays. To address this issue, an AI-based chatbot system has been developed for IMRD, Shirpur. The chatbot acts as a virtual assistant and provides instant responses to user queries. The system mainly supports admission guidance for BCA, BBA, BMS, and MCA programs. Unlike traditional keyword-based systems, the proposed solution uses semantic understanding to interpret the meaning of user queries and deliver accurate information.

II. Objectives

To develop an AI-based chatbot for instant admission-related information.

To provide details about eligibility, fee structure, documents, and admission dates.

To reduce the workload of admission staff.

To handle user queries expressed in different formats.

To maintain an easily updatable Excel-based knowledge base.

To provide 24×7 support to students and parents.

III. Research Methodology

This study follows an applied research approach focused on developing an automated enquiry handling system.

A. Data Collection

Information was collected from the official college website, admission brochures, prospectus, syllabus documents, placement reports, College students, staff, faculty's, HOD and facility details. The information was converted into a structured Question–Answer format.

B. Dataset Preparation

All question–answer pairs were stored in an Excel file. This allows easy updating without modifying the system code.

C. Text Preprocessing

User input is cleaned using NLP techniques such as lowercase conversion, removal of special characters, and text normalization.

D. Semantic Processing

The SentenceTransformer MiniLM-L6-v2 model converts text into 384-dimensional embeddings representing

sentence meaning.

E. Similarity Matching

Cosine similarity is calculated between user query embeddings and stored question embeddings. The top three relevant responses are returned with confidence scores.

IV. System Architecture

The system follows a layered architecture consisting of User Layer, Interface Layer, Backend Layer, NLP Processing Layer, Similarity Matching Layer, and Response Layer. Users interact through a web interface built using HTML, CSS, and JavaScript. The backend is implemented in Python, where preprocessing, embedding generation, and similarity computation are performed.

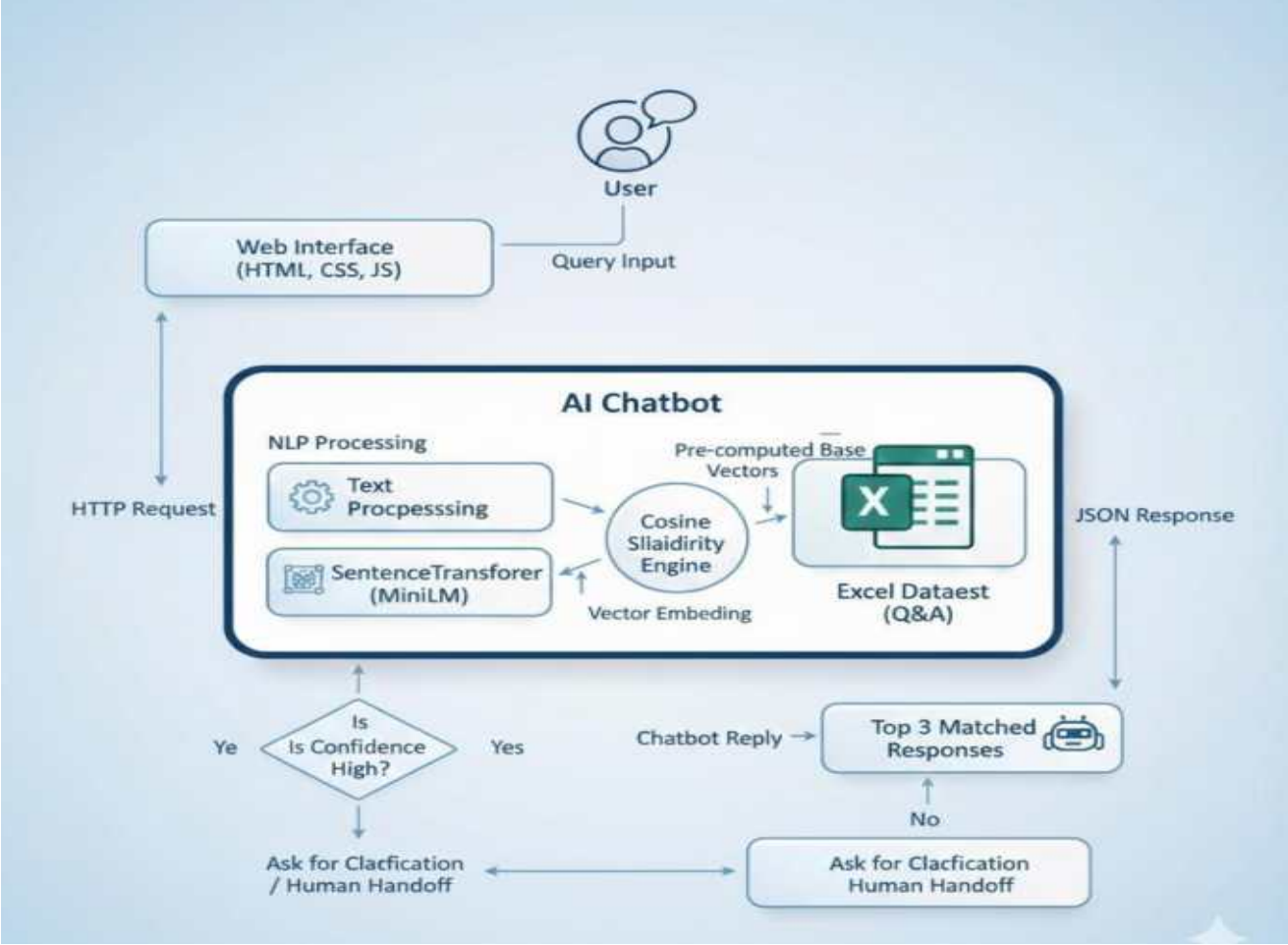


Fig. System Architecture

V. Implementation

The system is implemented using Python along with libraries such as SentenceTransformers, Pandas, NumPy, and Scikit-learn. The knowledge base is stored in Excel format, which allows easy modification by administrative staff. The working process includes loading the dataset, generating embeddings for stored questions, processing user queries, computing similarity scores, and displaying the best matching answers.

Algorithm Use:

BERT is a transformer-based deep learning model designed for contextual language understanding. However, BERT is not optimized for sentence similarity tasks directly. To compute similarity:

- Each sentence must be passed together through the model.
- It requires heavy computational resources.
- It is slow for real-time applications.

Limitation:

BERT is computationally expensive for pairwise sentence comparison.

The system S-BERT modifies BERT using a siamese network architecture. It generates fixed-size sentence embeddings that can be directly compared using cosine similarity.

Key Features:

- Efficient sentence embedding generation
- Optimized for semantic similarity

- Faster comparison
- Reduced computational cost
- High accuracy in similarity-based tasks

Design:

Fig: Home Page

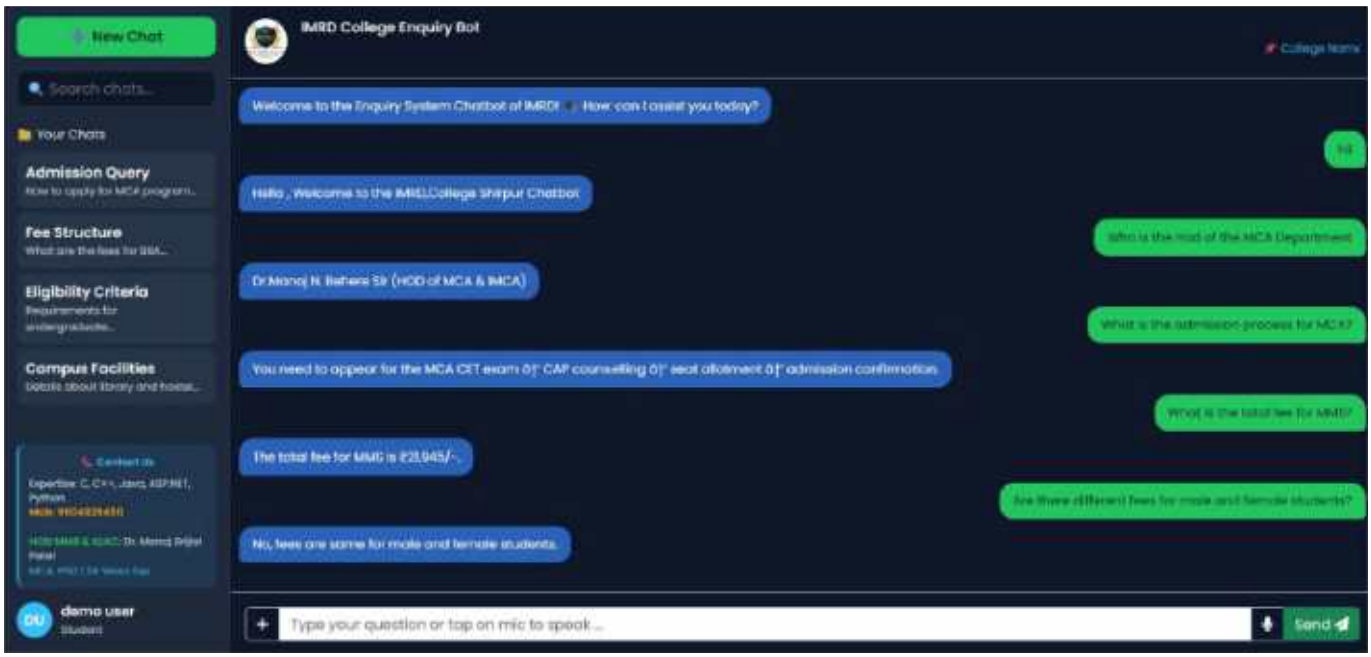


Fig: Chat Boat

VI. Results and Analysis

The chatbot was tested using various admission-related queries. The system achieved approximately Top-1 accuracy 99.42%. Precision and recall values were close to 1.0, indicating reliable performance.

The system successfully handled paraphrased queries and minor spelling errors. Response time was fast, making the system suitable for real-time interaction.

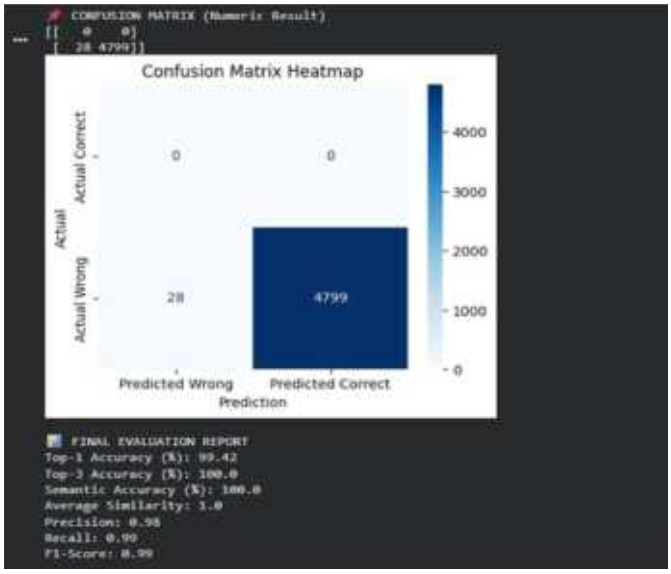


Fig: Classification Matrix

Comparative Performance Table:

Model	Top-1 Accuracy	Precision	Recall	F1-Score	Speed
BERT	91%	0.90	0.89	0.89	Slow
RoBERTa	93%	0.92	0.91	0.91	Slow
USE	88%	0.87	0.86	0.86	Medium
DistilBERT	89%	0.88	0.87	0.87	Fast
S-BERT	99.42%	0.98	0.99	0.99	Very Fast

VII. Discussion

The chatbot reduces manual workload and ensures consistent information delivery. Students and parents can access admission information anytime. The Excel-based dataset makes the system scalable and easy to maintain. In traditional transformer models like BERT and RoBERTa provide good contextual understanding but suffer from high computational cost and slow inference due to pairwise sentence processing. Universal Sentence Encoder (USE) and DistilBERT improve inference speed but offer comparatively lower accuracy. Among all models, Sentence-BERT (S-BERT) performs the best, achieving the highest accuracy, precision, recall, and F1-score. Its siamese architecture generates independent sentence embeddings, enabling fast cosine similarity computation and reducing computational complexity. Therefore, S-BERT provides the best balance of accuracy, speed, and efficiency, making it highly suitable for real-time semantic similarity and search applications.

VIII. Conclusion

The AI-Based College Enquiry Chatbot System provides an efficient solution for handling admission-related queries. The use of semantic similarity and transformer-based embeddings ensures accurate understanding of user queries. The system improves communication, reduces administrative workload, and enhances the admission experience.

IX. Future Scope

- Multilingual support (English, Hindi, Marathi)
- Voice-based interaction
- Integration with online admission portal
- Mobile application development
- Integration with student management systems

References :

1. Hugging Face. (2024). Sentence Transformers Documentation. https://huggingface.co/sentence-transformers
2. RCPET IMRD. (2026). Official Website. https://www.rcpimrd.ac.in/
3. Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence embeddings using Siamese BERT-networks. Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing (EMNLP), 3982–3992. https://arxiv.org/abs/1908.10084
4. scikit-learn Developers. (2024). Scikit-learn Documentation. https://scikit-learn.org/stable/
5. Wolf, T., et al. (2020). Transformers: State-of-the-art natural language processing. Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, 38–45. https://arxiv.org/abs/1910.03771
6. Python Software Foundation. (2024). Python Documentation. https://docs.python.org/3/

ISSN : 2348-7143
March 2026
RESEARCH JOURNEY International Multidisciplinary E-Research Journal
Impact Factor (SJIF) - 6.625 | ***Special Issue 389 :*** *Advanced Computing and Artificial Intelligence (ACAI)*

An Adaptive AI-Driven Mock Interview System to Enhance Employability Skills of Rural Students

Mrs. Archana Manoj Jade
Training and Placement Officer,
RCPET's Institute of Management Research and Development, Shirpur

Abstract

Rural students are typically faced with a considerable challenge of finding jobs, as they are getting less structured interview practises which should give them a real-time feedback on their performance. The traditional practices of conducting mock interviews in rural institutions are mostly manual with no adaptability and are not easily scalable. In this paper I would like to suggest an Adaptive AI-Driven Mock Interview System that can be used to simulate dynamic interview scenarios and adjust the level of difficulty during question asking and the feedback provided to the candidate accordingly. The framework presented incorporates intelligent question generation, response analysis, adaptable difficulty control, and analytics of the structured feedback to improve the communication skills, technical articulation, and confidence levels. It also proposes a pilot evaluation model to determine the effectiveness of the systems in the rural higher education. The framework is expected to offer a cost-effective solution to close the employment disparity gap between the rural learners in a scaled manner.

Keywords: Adaptive AI, Mock Interview Simulation, Employability Skills, Rural Education, Intelligent Learning Systems

1. Introduction

The development of employability is a major issue that is still a challenge in higher education institutions in rural settings. Most rural students are not able to compete during the recruitment procedures because most of them lack exposure in the interviewing process, communication barriers, and there are no well-structured performance evaluation systems. Faculty or placement cell mock interviews are also limited by time and evaluator availability, along with patterns of subjective feedback. Moreover, standardized questioning methods do not cover diverse levels of competencies in students. The recent breakthroughs in Artificial Intelligence (AI) have facilitated adaptive learning systems that are able to customize learning. Nevertheless, there is little use of adaptive AI systems in simulating interviews to improve rural employability. The proposed study is a proposal to construct Adaptive AI-Driven Mock Interview System which can dynamically change the level of the interview difficulty, analyze the answers, on the basis of the structured criteria and create the performance-based feedback to develop the rural students to interview preparedness.

2. Literature Review

AI-based Intelligent Tutoring Systems (ITS) have demonstrated effectiveness in personalized learning environments. Adaptive learning models adjust instructional content based on learner performance, improving engagement and competency development.

Conversational AI systems and Natural Language Processing (NLP) techniques have been widely applied in educational chatbots, language training systems, and automated assessment platforms. Studies indicate that adaptive feedback mechanisms significantly improve learner confidence and performance outcomes. In employability training, digital mock interview platforms exist; however, most systems rely on static question banks and lack dynamic difficulty adjustment mechanisms. Moreover, rural educational contexts require scalable and cost-effective technological interventions due to resource constraints.

The literature highlights the need for context-aware, adaptive, and structured AI frameworks tailored to rural employability enhancement.

Problem Statement:

Students in rural areas have severe difficulties in demonstrating themselves positively during the job interview process because they do not often have any exposure to a professional mocking interview process, they do not have any personalized feedback, and they do not receive a professional mentoring opportunity. The traditional mock interview methods practiced in the rural institutions are manual, non-adaptive and hard to scale. It is required to have an adaptive AI-based system that could be used to create believable interview situations, give performance feedback, and improve employability abilities scalable and at a low cost.

Objective:

1. To design an adaptive AI-driven mock interview system to enhance employability skills among rural students.
2. To develop a structured framework for AI-based dynamic interview simulation.
3. To implement adaptive difficulty adjustment based on student performance.

4. To provide automated, performance-based feedback for skill improvement.

Research Gap:

1. There is a gap in developing structured and adaptive AI-based mock interview systems specifically designed for rural students' employability enhancement.
2. Existing digital mock interview platforms largely rely on static question banks and lack dynamic difficulty adjustment based on student performance.
3. There is limited research focusing on personalized feedback mechanisms aligned with industry-oriented employability evaluation metrics in rural higher education contexts.
4. Traditional mock interview practices in rural institutions remain manual, time-consuming, and difficult to scale effectively.
5. The study aims to design an adaptive AI-driven mock interview framework that addresses personalization, scalability, and structured performance evaluation to enhance rural students' interview readiness.

Scope of the study:

The proposed study is aimed at the creation of an adaptive AI-aided mock interview solution that can assist rural students in enhancing their productivity during an interview and their overall employability skills. The suggested framework will contribute to the improvement of such aspects as clarity in communication, technical expression, and confidence building. The study is confined to the system design and pilot-level testing in a rural institution of higher education setup. At this point, the study does not presuppose the use of sophisticated AI model training or massive deployment. Further growth and increased validation is implied outside the scope.

Research Methodology:

This study adopts to design and conceptually evaluate an Adaptive AI-Driven Mock Interview System aimed at enhancing employability skills among rural students.

The secondary data to be used in this research were obtained in the form of published research articles, scholarly journals, conference papers, government reports, institutional placement records, and authentic online databases. Reviewing of the literature was done to grasp the available literature on the topics of Artificial Intelligence in education, Intelligent Tutoring Systems (ITS), AI-based interview training systems, Natural Language Processing applications, and employability skill development models. Peer-reviewed journals, publications indexed in Scopus, Google Scholar articles, AICTE and UGC reports and industry placement reports were reviewed to determine any gaps in research, system design models, evaluation metrics, and best practices. This secondary information was useful in abstracting the model of the proposed AI-driven mock interview system and in planning relevant variables and analysis parameters of the research.

Proposed System Framework

The proposed system follows a three-layered architecture:

1. Input Layer – Collects student profile data, academic background, skill level, and career preferences.
2. Processing Layer – Applies adaptive question selection using rule-based logic and a dynamic scoring mechanism that adjusts difficulty based on prior responses.
3. Output Layer – Generates structured feedback focusing on response relevance, clarity, articulation, and overall confidence.

The framework operates on a performance-driven adaptation model, where each response influences subsequent question complexity, thereby simulating an intelligent and personalized interview experience.

Proposed Technology Stack

The future implementation of the system is proposed as a web-based application. The frontend may be developed using HTML, CSS, and JavaScript to ensure user-friendly interaction. The backend logic can be implemented using Python (Flask/Django) to manage adaptive scoring and question flow. A relational database such as MySQL or PostgreSQL may be used to store student data and performance records. Basic Natural Language Processing (NLP) libraries in Python can be incorporated for response evaluation and feedback generation.

Conclusion & Future Work

The study presents a conceptual adaptive AI-driven mock interview framework designed to address employability challenges faced by rural students. The proposed model emphasizes personalization, confidence building, and structured feedback mechanisms. Preliminary validation supports the practical need for such a system in rural academic settings.

Future work will focus on prototype development, real-time response evaluation using advanced NLP techniques, and large-scale experimental validation to measure measurable improvements in interview performance and employability outcomes.

References :

1. Lin, C.-C., Huang, A. Y. Q., & Lu, O. H. T. (2023). Artificial intelligence in intelligent tutoring systems toward sustainable education: a systematic review. Smart Learning Environments.
2. Labadze, L., Grigolia, M., & Machaidze, L. (2023). Role of AI chatbots in education: systematic literature review. International Journal of Educational Technology in Higher Education.
3. Tejaswini, K., Aravinda, T. V., et al. (2025). AI-Powered Mock Interview Platform with NLP and Speech Analysis for Personalized Feedback. International Research Journal on Advanced Engineering Hub.

Conference/Preprint Papers:

1. Krishna, A., Satheesh, A., et al. (2024). AI-Driven Personalized Learning: A Comprehensive Survey of Chatbot Applications in Education and Training. Preprints.org.
2. Goyal, H., Garg, G., et al. (2025). The Impact of Large Language Models on K-12 Education in Rural India: A Thematic Analysis. ArXiv.

RESEARCH JOURNEY International Multidisciplinary E-Research Journal
Impact Factor (SJIF) - 6.625 | Special Issue 389 : Advanced Computing and Artificial Intelligence (ACAI)
ISSN : 2348-7143
March 2026

An Overview of OCR Evaluation Tools and Metrics

Dr. M. S. Sonawane
Institute of Management Research and Development, Shirpur

Abstract:- *With millions of historical documents now digitized and housed in libraries, their use has expanded beyond basic keyword searches to applications demanding high-precision OCR output. This shift highlights the challenge of accurately and efficiently evaluating OCR performance at scale, particularly when ground truth data is available for only a limited subset of materials. Variations in tool specifications and implementation approaches often lead to inconsistent evaluation results, making it difficult to compare error rates across systems. Moreover, traditional OCR metrics and sampling strategies fall short when they overlook layout analysis accuracy—an essential factor for advanced tasks in Natural Language Processing and Digital Humanities, where correct reading order and structural interpretation are vital. This study offers a comprehensive review of OCR evaluation tools and metrics, showcases two advanced use cases, and presents a comparative experiment using multiple evaluation methods across two distinct datasets. The findings reveal key differences and commonalities tied to specific use cases and suggest directions for future research in OCR assessment.*

Keywords: OCR, optical character recognition, evaluation, metrics, layout analysis, accuracy.

Introduction

Assessing the quality of Optical Character Recognition (OCR) outputs in a way that is efficient, transparent, and insightful presents a number of challenges. Traditional evaluation methods often depend on Ground Truth (GT) data to benchmark accuracy. However, in large-scale digitization efforts—where millions of documents are processed—producing sufficient GT data is both impractical and costly. This issue is especially pronounced with historical materials, which require expert knowledge and significant time investment to annotate accurately [39].

Compounding the problem is the absence of standardized protocols for generating GT for historical texts. Evaluating OCR accuracy becomes complex due to ambiguities such as how ligatures are interpreted (as single characters or combinations), how non-standard characters are encoded, and how special symbols not covered by Unicode are handled—often requiring extensions like MUFI or private-use code points. Inconsistencies in punctuation and spacing further complicate the evaluation process.

Initiatives like the OCR-D Ground Truth Guidelines have made strides in aligning OCR practices with scholarly needs by offering useful specifications. Yet, current GT-based evaluation frameworks still fall short in addressing the intricate demands of historical documents. Moreover, applying these methods across vast digitized collections remains largely unfeasible [3].

There is a clear need to explore alternative approaches, such as leveraging OCR confidence scores and statistical sampling techniques, to produce evaluations that are both meaningful and comparable. Additionally, layout analysis—critical for advanced applications in fields like Natural Language Processing and Digital Humanities—is often overlooked in existing metrics.

This paper aims to address these gaps by examining transparency and comparability in OCR evaluation. Drawing on extensive experience with large-scale collections and collaboration with domain experts, the study identifies key limitations in current practices and contextualizes evaluation outcomes within real-world use cases.

The structure of the paper is as follows:

- Section 2 introduces two advanced use cases for OCR.
- Section 3 reviews current evaluation tools and methodologies, including alternative approaches.
- Section 4 presents a comparative experiment using five widely adopted tools across two datasets, analyzing the results in relation to the use cases.
- Section 5 concludes with a summary and recommendations for future research.

Applications of OCR in Advanced Research Contexts

As digitized texts become increasingly accessible, the role of Optical Character Recognition (OCR) in digital libraries has expanded far beyond simple keyword searches. Today, OCR outputs are foundational for more sophisticated tasks such as Natural Language Processing (NLP) and Text and Data Mining (TDM), especially within the Digital Humanities (DH), where researchers are leveraging computational methods to explore historical documents in new ways.

Natural Language Processing

OCR accuracy is critical for NLP applications, particularly in tasks like Named Entity Recognition (NER), which involves identifying names of people, places, and dates within texts [18]. These elements often form the backbone of user queries in digital archives. However, integrating NER into digitized collections remains rare

in large-scale production environments due to the low reliability of OCR outputs [9].

Studies have shown that even minor increases in OCR error rates can significantly degrade NER performance [14], [11]. For instance, when word error rates rose from 1% to 7% and character error rates from 8% to 20%, NER accuracy dropped from 90% to 60%. Similar findings emerged from evaluations in shared tasks focused on historical texts, underscoring the sensitivity of NLP pipelines to OCR quality.

Beyond recognition, linking named entities to structured knowledge bases—known as Named Entity Linking (NEL)—also depends heavily on clean OCR output [10]. Poor recognition can lead to misidentification or ambiguity, which undermines the integrity of downstream analyses. Other NLP tasks, such as syntactic parsing and semantic analysis, are equally affected by OCR precision [22], [19], [37].

Digital Humanities

In the realm of Digital Humanities, scholars apply computational techniques to investigate cultural and historical questions. The availability of digitized historical texts has opened new avenues for research, but OCR errors pose a significant barrier to trust and reproducibility.

Qualitative studies involving historians reveal that many refrain from publishing quantitative findings based on digitized sources due to concerns about data reliability. In one survey, three out of four participants expressed hesitation, fearing their conclusions could be challenged due to OCR inaccuracies.

Quantitative research further illustrates that while topic modeling may tolerate moderate OCR errors, other methods—such as collocation analysis and style [35], [12], [30].

OCR Assessment: Approaches and Metrics

Evaluating the accuracy of OCR output involves a variety of metrics and tools, many of which are discussed in academic literature and available as open-source implementations. However, these methods often provide only a partial view of OCR performance. This section explores the most widely adopted metrics and tools used to assess OCR quality.

State-of-the-Art

Modern techniques for evaluating text recognition systems—particularly those recognized by the IAPR-TC11 Reading Systems community—can be traced back to foundational work by Stephen V. Rice. His research introduced the use of edit distance algorithms to measure OCR accuracy, with Ukkonen's algorithm recommended for its efficiency in handling long strings. Rice's framework distinguishes between Character Accuracy and Word Accuracy, and also introduces more nuanced metrics like Non-Stopword Accuracy and Phrase Accuracy, which evaluate sequences of words.

These methodologies were implemented by the Information Science Research Institute (ISRI) at the University of Nevada, Las Vegas, between 1992 and 1996. These tools became widely accepted for academic benchmarking and OCR competitions. Later updates in 2015–2016 expanded their capabilities to support Unicode character sets [25], [36], [17], [26], [28].

One of the most commonly used metrics today is Character Error Rate (CER), calculated as:

CER=i+s+dn\text {CER} = \frac{i + s + d} {n}

Where:

- i = insertions
- s = substitutions
- d = deletions
- n = total number of characters in the reference text

A recent enhancement to CER proposes an "end-to-end" metric that allows flexible alignment between ground truth and OCR output, accommodating layout and segmentation variations [16].

Contributions from PRImA

The Pattern Recognition & Image Analysis (PRImA) Research Lab at the University of Salford has significantly advanced OCR evaluation. They developed the PAGE (Page Analysis and Ground-Truth Elements) format—an XML-based schema that captures detailed layout, reading order, and image features. PRImA also introduced tools for evaluating layout analysis and reading order, including the Flexible Character Accuracy (FCA) metric, which accounts for misaligned reading sequences [21], [5], [6], [8].

Sampling Strategies

To minimize the need for extensive ground truth data, random sampling is often employed. When done correctly—such as through Bernoulli sampling—it can yield statistically valid insights. However, manual sampling may overlook layout-related errors, and random selection might miss critical content areas. For instance, a misrecognized headline in a newspaper could have a greater impact than errors in less relevant sections like advertisements [38].

Empirical Studies

Large-scale evaluations of OCR accuracy in digitized newspaper archives have been conducted by various researchers. One study on the British Library's collection used CER and Word Error Rate (WER), introducing a Significant-Word Accuracy Rate to focus on content-relevant terms. Another investigation into the Australian Newspaper Digitisation Program highlighted factors influencing OCR performance and recommended improvements like lexicon integration. However, frequency-based language models have shown limited effectiveness, especially for historical texts lacking robust linguistic resources [13], [34], [31].

Specialized Tools and Projects

The IMPACT project developed several OCR evaluation tools tailored for historical documents. The NCSR Evaluation Tool, an extension of ISRI's framework, supports UTF-8/UTF-16 and introduces the Figure of Merit metric, which estimates the manual effort needed to correct OCR errors—giving more weight to substitutions due to their complexity[2],[15],[24].

Other notable tools include:

- **ocrevalUAtion:** Compares ground truth and OCR outputs using CER/WER, and supports cross-system comparisons.
- **dinglehopper (from the Qurator project):** Offers visual side-by-side comparisons of OCR and ground truth, helping identify layout and recognition issues.

Alternative Approaches

In recent years, several methods have emerged for evaluating OCR output without relying on ground truth (GT), offering flexibility in contexts where GT is unavailable or impractical—particularly in historical document digitization[1].

One such approach was developed to assess OCR quality in 19th-century English business texts. Instead of comparing OCR output to GT, it uses a heuristic based on the frequency of numerals and dictionary-matched words. This method introduces a Simple Quality (SQ) score, calculated as the proportion of recognized words that appear in a predefined lexicon. However, applying this technique to historical documents poses challenges due to the limited availability of comprehensive historical dictionaries, which makes it difficult to validate archaic spellings as correct[29],[7].

Another strategy leverages image pre-processing and machine learning. Researchers have trained predictive models using small samples of GT-labeled data—both images and text—to estimate OCR accuracy on new datasets. This technique has been extended to evaluate various classifiers, achieving prediction errors ranging from 2.6% to 11% when benchmarked against actual GT, demonstrating its potential for scalable evaluation[32],[23].

Further innovations include metrics that incorporate OCR confidence scores. One method applies statistical modeling using a Student's t-distribution, combined with lexicality analysis based on document-specific language profiles. These profiles, especially those tailored to historical spelling conventions, have shown a strong correlation between spelling variants and OCR inaccuracies. By integrating these insights, post-processing and correction of OCR output can be significantly improved [27].

Another notable method involves cross-system alignment, where the output of one OCR engine is compared against that of a secondary system. This comparative technique has proven more effective at predicting OCR accuracy than relying solely on internal confidence scores from a single engine.

Limitations and Considerations

While these GT-free approaches offer promising alternatives, they often depend on external resources such as language models or historical lexicons—which may be scarce or incomplete in older texts. Moreover, none of these methods fully address layout analysis or reading order evaluation, which are critical components outlined in earlier sections. As such, while useful in specific scenarios, these alternatives fall short of providing a comprehensive solution for OCR assessment across diverse document types and structures.

Experimentation

To better understand how various OCR evaluation tools and metrics perform across different historical document types, we conducted a comparative study using two distinct datasets. Our goal was to identify alignment and divergence in OCR assessment outcomes using widely recognized tools.

Datasets and OCR Setup

We selected two representative datasets for this experiment:

- **IMPACT Dataset:** Comprising historical book pages, curated to reflect digitization challenges faced by European libraries.
- **Europeana Newspapers Project (ENP) Dataset:** Featuring digitized historical newspapers, notable for their complex layouts and predominant use of Fraktur fonts.

OCR outputs were generated using Tesseract v4.1.0, executed in two configurations:

1. A specialized historical font model trained on the GT4HistOCR dataset, applied uniformly across all pages.
2. Standard language models sourced from the tessdata repository.

In both runs, the ALTO format was used for OCR output, and confidence scores were extracted for each page to support comparative analysis [20], [4].

Evaluation Tools

Five OCR evaluation tools were selected for comparison:

- TextEval and LayoutEvaluation from the PRImA research group (the latter being the only tool in our study that performs detailed layout analysis).
- dinglehopper from the Qurator project.
- ocrevalUAtion, developed under the IMPACT initiative.

- ocreval, the latest version of the ISRI evaluation suite.

Most tools supported direct comparison between ALTO OCR output and PAGE ground truth. However, ocreval required conversion to UTF-8 plain text, which introduced potential errors—particularly in reading order. LayoutEvaluation also required binarized images as input [33].

Evaluations were conducted on two systems:

- A Linux machine (4-core CPU, 64 GB RAM) for all tools except LayoutEvaluation.
- A Windows machine (4-core CPU, 16 GB RAM) for LayoutEvaluation.

Processing time varied significantly between datasets. The IMPACT dataset was evaluated within hours, while the ENP dataset—due to its larger page sizes and character counts—required several days. Pages exceeding 60,000 characters triggered memory exceptions in tools like TextEval, dinglehopper, and ocreval. In general, CER calculations for pages over 20,000 characters were time-intensive, affecting 140 of the 465 ENP pages. In contrast, the IMPACT dataset had a maximum of 2,609 characters per page, highlighting the scalability challenges in newspaper digitization.

Results and Observations

We evaluated 378 pages from the IMPACT dataset and 465 pages from the ENP dataset using both OCR models. For consistency, success rates and confidence scores were inverted to reflect error rates. Any reported error rates above 100% were capped at 1.

Key findings:

- CER values showed consistent trends across tools, though dinglehopper produced slightly different results due to its unique alignment and normalization methods.
- Flexible CER (FCER) appeared more lenient, likely due to its tolerance for segmentation errors.
- WER was predictably higher than CER, and a strong correlation was observed between CER and layout success rates.
- Layout quality metrics revealed segmentation issues not fully captured by CER alone.
- dinglehopper exhibited the lowest variance in CER, followed by PRImA's LayoutEval metrics.
- ocrevalUAtion showed the highest CER variance—even exceeding WER variance, which is typically more volatile.

Conclusion

Evaluating the accuracy of OCR output demands a multidimensional approach, especially within large-scale digitization efforts. Creating ground truth (GT) for every page remains a costly and time-intensive endeavor, making it impractical for many projects. Moreover, confidence scores provided by OCR engines should be interpreted with caution, as they may not reliably reflect actual recognition quality.

Although GT-free evaluation methods offer alternatives, their effectiveness is often contingent on the availability of robust language resources. This becomes particularly problematic when dealing with historical texts, where suitable dictionaries or linguistic models are scarce or incomplete. A more feasible strategy involves leveraging smaller, well-curated samples with GT to train predictive models or guide quality assessments across broader datasets.

Challenges in Comparability

Our experimental findings reveal that even widely accepted metrics like Character Error Rate (CER) and Word Error Rate (WER) can yield inconsistent results across different evaluation tools. This lack of comparability stems from several factors:

- Some tools do not fully support structured formats like PAGE XML, especially when reading order is defined but ignored during conversion to plain text.
- Unicode normalization practices vary across tools such as TextEval, ocrevalUAtion, and dinglehopper, leading to inconsistent handling of historical ligatures and character variants—even within the same tool depending on configuration settings.

Use Case-Specific Evaluation

The importance of OCR quality varies depending on the intended application. For example, keyword search may tolerate minor layout inaccuracies, while scholarly research or linguistic analysis demands precise reading order and structural fidelity. Since no single tool offers a complete picture, a promising direction is to develop standardized evaluation profiles tailored to specific use cases, ensuring that assessments are both relevant and meaningful.

Layout Analysis: The Missing Piece

Despite its critical role, layout analysis remains underrepresented in most OCR evaluation metrics. There is currently no universally adopted framework for assessing layout quality in a consistent and comprehensive manner. This gap is especially problematic for documents with complex formatting, where accurate reading order and segmentation are essential. Without metrics that account for these aspects, evaluations fall short for advanced use cases.

Path Forward

To support both digitization professionals and digital humanities researchers, the community must work toward:

- Unified standards for text normalization and

alignment
- Interoperable formats for exchanging evaluation data
- Transparent metadata that documents transformation steps and evaluation settings
- Benchmark datasets with known errors to facilitate deeper understanding of OCR quality dimensions

By aligning tools, practices, and expectations, we can move toward more reliable and insightful OCR evaluation frameworks that serve both technical and scholarly needs.

References :

1. Beatrice Alex and John Burns. 2014. Estimating and rating the quality of optically character recognised text. In Proceedings of the First International Conference on Digital Access to Textual Cultural Heritage. ACM, NY, USA, 97–102.
2. Hildelies Balk and Aly Conteh. 2011. IMPACT: centre of competence in text digitisation. In Proceedings of the 2011 Workshop on Historical Document Imaging and Processing. ACM, NY, USA, 155–160.
3. Matthias Boenig, Konstantin Baierer, Volker Hartmann, Maria Federbusch, and Clemens Neudecker. 2019. Labelling OCR Ground Truth for Usage in Repositories. In Proceedings of the Third International Conference on Digital Access to Textual Cultural Heritage. ACM, NY, USA, 3–8.
4. Christian Clausner, Christos Papadopoulos, Stefan Pletschacher, and Apostolos Antonacopoulos. 2015. The ENP image and ground truth dataset of historical newspapers. In 2015 13th International Conference on Document Analysis and Recognition (ICDAR). IEEE, NY, USA, 931–935.
5. Christian Clausner, Stefan Pletschacher, and Apostolos Antonacopoulos. 2011. Scenario driven in-depth performance evaluation of document layout analysis methods. In 2011 International Conference on Document Analysis and Recognition. IEEE, NY, USA, 1404–1408.
6. Christian Clausner, Stefan Pletschacher, and Apostolos Antonacopoulos. 2013. The significance of reading order in document recognition and its evaluation. In 2013 12th International Conference on Document Analysis and Recognition. IEEE, NY, USA, 688–692.
7. Christian Clausner, Stefan Pletschacher, and Apostolos Antonacopoulos. 2016. Quality prediction system for large-scale digitisation workflows. In 2016 12th IAPR Workshop on Document Analysis Systems (DAS). IEEE, NY, USA, 138–143.
8. Christian Clausner, Stefan Pletschacher, and Apostolos Antonacopoulos. 2020. Flexible character accuracy measure for reading-order-independent evaluation. Pattern Recognition Letters 131 (2020), 390–397.
9. Gregory Crane and Alison Jones. 2006. The challenge of virginia banks: an evaluation of named entity analysis in a 19th -century newspaper collection. In Proceedings of the 6th ACM/IEEE-CS joint conference on Digital libraries. IEEE, NY, USA, 31–40.
10. Maud Ehrmann, Matteo Romanello, Alex Flückiger, and Simon Clematide. 2020. Extended overview of CLEF HIPE 2020: named entity processing on historical newspapers. In CLEF 2020 Working Notes. Conference and Labs of the Evaluation Forum, Vol. 2696. CEUR, Aachen, Germany, 1–38.
11. Ahmed Hamdi, Axel Jean-Caurant, Nicolas Sidere, Mickaël Coustaty, and Antoine Doucet. 2019. An analysis of the performance of named entity recognition over OCRed documents. In 2019 ACM/IEEE Joint Conference on Digital Libraries (JCDL). IEEE, NY, USA, 333–334.
12. Mark J Hill and Simon Hengchen. 2019. Quantifying the impact of dirty OCR on historical text analysis: Eighteenth Century Collections Online as a case study. Digital Scholarship in the Humanities 34, 4 (2019), 825–843.
13. Rose Holley. 2009. How good can it get? Analysing and improving OCR accuracy in large scale historic newspaper digitisation programs. D-Lib Magazine 15, 3/4 (2009), Unpaginated.
14. Kimmo Kettunen, Eetu Mäkelä, Teemu Ruokolainen, Juha Kuokkala, and Laura Löfberg. 2017. Old content and modern tools-searching named entities in a Finnish OCRed historical newspaper collection 1771-1910. Digital Humanities Quarterly 11 (2017), 24. Issue 3.
15. Vladimir Kluzner, Asaf Tzadok, Yuval Shimony, Eugene Walach, and Apostolos Antonacopoulos. 2009. Word-based adaptive OCR for historical books. In 2009 10th International Conference on Document Analysis and Recognition. IEEE, NY, USA, 501–505.
16. Gundram Leifert, Roger Labahn, Tobias Grüning, and Svenja Leifert. 2019. End-To-End Measure for Text Recognition. In 2019 International Conference on Document Analysis and Recognition (ICDAR). IEEE, NY, USA, 1424–1431.
17. Vladimir I Levenshtein. 1966. Binary codes capable of correcting deletions, insertions, and reversals. Soviet physics doklady 10, 8 (1966), 707–710.
18. Daniel Lopresti. 2009. Optical character recognition errors and their effects on natural language processing. International Journal on Document Analysis and Recognition (IJDAR)

12, 3 (2009), 141–151.

19. Margot Mieskes and Stefan Schmunk. 2019. OCR Quality and NLP Preprocessing. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, Florence, Italy. ACL, Stroudsburg PA, USA, 102–105.
20. Christos Papadopoulos, Stefan Pletschacher, Christian Clausner, and Apostolos Antonacopoulos. 2013. The IMPACT dataset of historical document images. In Proceedings of the 2nd international workshop on historical document imaging and processing. ACM, NY, USA, 123–130.
21. Stefan Pletschacher and Apostolos Antonacopoulos. 2010. The page (page analysis and ground-truth elements) format framework. In 2010 20th International Conference on Pattern Recognition. IEEE, NY, USA, 257–260.
22. Elvys Linhares Pontes, Ahmed Hamdi, Nicolas Sidere, and Antoine Doucet. 2019. Impact of OCR quality on named entity linking. In International Conference on Asian Digital Libraries. Springer, NY, USA, 102–115.
23. Ulrich Reffle and Christoph Ringlstetter. 2013. Unsupervised profiling of OCRed historical documents. Pattern Recognition 46, 5 (2013), 1346–1357.
24. Georg Rehm, Peter Bourgonje, Stefanie Hegele, Florian Kintzel, Julián Moreno Schneider, Malte Ostendorff, Karolina Zaczynska, Armin Berger, Stefan Grill, and Sören et al. Räuchle. 2020. QURATOR: Innovative Technologies for Content and Data Curation. CEUR-WS 2535, 1 (2020), 15.
25. Stephen Vincent Rice. 1996. Measuring the accuracy of page-reading systems. UNLV, Las Vega, NV.
26. Stephen V Rice and Thomas A Nartker. 1996. The ISRI analytic tools for OCR evaluation. UNLV/Information Science Research Institute, TR-96 2 (1996), 45.
27. Ahmed Ben Salah, Jean Philippe Moreux, Nicolas Ragot, and Thierry Paquet.2015. OCR performance prediction using cross-OCR alignment. In Proceedings of the 13th International Conference on Document Analysis and Recognition (ICDAR). IEEE, NY, USA, 556–560.
28. Eddie A Santos. 2019. OCR evaluation tools for the 21st century. In Proceedings of the Workshop on Computational Methods for Endangered Languages, Vol. 1. ACL, Stroudsburg PA, USA, 23—-27.
29. Prashant Singh, Ekta Vats, and Anders Hast. 2018. Learning surrogate models of document image quality metrics for automated document image processing. In 2018 13th IAPR International Workshop on Document Analysis Systems (DAS). IEEE, NY, USA, 67–72.
30. David A Smith and Ryan Cordell. 2018. A research agenda for historical and multilingual optical character recognition. Northeastern University, Boston, MA.
31. Ray Smith. 2011. Limits on the application of frequency-based language models to OCR. In 2011 International Conference on Document Analysis and Recognition. IEEE, NY, USA, 538–542.
32. Uwe Springmann, Florian Fink, and Klaus U Schulz. 2016. Automatic quality evaluation and (semi-) automatic improvement of OCR models for historical printings. arXiv preprint arXiv:1606.05157 (2016), 8.
33. Uwe Springmann, Christian Reul, Stefanie Dipper, and Johannes Baiter. 2018. Ground Truth for training OCR engines on historical documents in German Fraktur and Early Modern Latin. arXiv preprint arXiv:1809.05501 (2018), 8.
34. Simon Tanner, Trevor Muñoz, and Pich Hemy Ros. 2009. Measuring mass text digitization quality and usefulness. D-lib Magazine 15, 7/8 (2009), 1082–9873.
35. Myriam C Traub, Jacco Van Ossenbruggen, and Lynda Hardman. 2015. Impact analysis of OCR quality on research tasks in digital archives. In International Conference on Theory and Practice of Digital Libraries. Springer, NY, USA, 252–263.
36. Esko Ukkonen. 1995. On-line construction of suffix trees. Algorithmica 14, 3 (1995), 249–260.
37. Daniel van Strien, Kaspar Beelen, Mariona Coll Ardanuy, Kasra Hosseini, Barbara McGillivray, and Giovanni Colavizza. 2020. Assessing the Impact of OCR Quality on Downstream NLP Tasks. In ICAART (1). SCITEPRESS, Setúbal, Portugal, 484–496.
38. Maria Wernersson. 2015. Evaluation von automatisch erzeugten OCR-Daten am Beispiel der Allgemeinen Zeitung. ABI Technik 35, 1 (2015), 23–35.
39. Clemens Neudecker, Konstantin Baierer, Mike Gerber, Christian Clausner, Apostolos Antonacopoulos and Stefan Pletschacher.2021. A survey of OCR evaluation tools and metrics.

RESEARCH JOURNEY International Multidisciplinary E-Research Journal
Impact Factor (SJIF) - 6.625 | Special Issue 389 : Advanced Computing and Artificial Intelligence (ACAI)
ISSN : 2348-7143
March 2026

AI-Driven Priority-Based Vehicle Routing Optimization for Smart Urban Waste Management

Mr. Darshan Ramesh Chaudhari,
Ms. Harshada Mansaram Padmar,
Ms. Devayani Pramod Patil,
Ms. Chhaya Suhas Patil

Abstract

In many cities urban waste collection systems continue to follow set routes and schedules, which frequently results in needless travel, higher fuel consumption, and higher operating expenses. This paper proposes an AI-driven priority-based vehicle routing framework for intelligent urban waste management in order to overcome this constraint. The suggested system uses a layered architecture that combines centralized data storage, priority classification, Vehicle Routing Problem (VRP)-based optimization, and simulated Internet of Things-based bin monitoring.

To assess system performance, waste bin data—such as fill levels and geographic coordinates—was produced inside Mumbai's physical borders. Only high-priority bins were chosen for optimal routing after bins were categorized using a threshold-based priority mechanism. Google OR-Tools was used to implement the routing model with the goal of reducing the overall travel distance.

Experimental evaluation was conducted under 50-bin and 100-bin deployment scenarios. In the 50-bin experiment, the total travel distance decreased from 188.51 km to 122.67 km when priority-based routing was applied, which corresponds to roughly a 35% reduction compared to servicing every bin. A similar pattern was observed in the 100-bin scenario, where the route length dropped from 262.73 km to 128.32 km, resulting in nearly a 51% decrease. These reductions also translate into noticeable fuel savings and lower estimated CO_2 emissions per collection cycle.

Overall, the findings indicate that selecting bins based on priority before performing route optimization can meaningfully reduce unnecessary travel and improve the operational efficiency of urban waste collection systems.

The architecture is designed to support future integration of predictive analytics for proactive and data-driven waste management.

1. Introduction

The issue of waste management in urban areas has become a significant problem due to the rate of urbanization and population growth. As the urban area expands, the amount of solid waste generated daily also increases. Effective collection and transportation of the waste are critical to ensure public health and environmental sustainability.

Mumbai is one of the most densely populated metropolitan cities in India, and it has a large-scale solid waste management system. According to the municipal reports, the city has tens of thousands of sanitation staff and several transfer stations for the daily transportation of waste. The historical expenditure data reveal that solid waste management is a financially intensive process, as the expenditure is in the form of several thousand million rupees per year and is gradually increasing with time [6].

Conventional waste collection, common in many urban areas such as Mumbai, involves predetermined routes and timings. Under this method, garbage collection trucks go to all the allotted bins irrespective of whether they are full or not. Although this method ensures that all bins are cleaned, it also results in wastage of fuel and carbon emissions.

With the evolution of smart city technology, waste management solutions can be improved by leveraging data analytics. Smart trash cans with sensors can track the level of fillage in real-time, making it possible to collect waste based on the condition of the bins rather than the timing. But this efficiency can only be realized if route optimization is also done.

This study proposes an AI-driven priority-based vehicle routing framework for smart urban waste management. The framework integrates bin priority classification with Vehicle Routing Problem (VRP) optimization to reduce unnecessary travel. The proposed solution is tested using simulated bin data within the geographical area of Mumbai, under 50 bin and 100 bin deployment scenarios.

The outcome of the proposed solution shows a reduction in the travel distance, fuel consumption, and estimated CO_2 emissions, thus proving the effectiveness of intelligent routing techniques in large-scale waste management systems.

2. Related Work

Research in urban waste management and vehicle routing has grown significantly due to increasing waste generation and smart city initiatives. A considerable body

of work focuses on vehicle routing models to optimize waste collection and reduce travel costs and distance. Early studies on the vehicle routing problem (VRP) highlight its application to various logistics problems, including waste collection optimization, where the goal is to minimize total travel cost and operational expenses [1], [2].

Waste collection modeling has been reviewed comprehensively, emphasizing different routing variants and solution techniques for municipal solid waste systems [3]. Metaheuristic methods such as genetic algorithms and ant colony optimization have been applied to VRP and its extensions to improve route quality for waste collection scenarios [4]. Specifically for smart waste management, real-time data from sensor-equipped bins has been integrated with routing optimization to make collection more efficient. To illustrate, Roy et al. designed a smart bin allocation and vehicle routing system with IoT, evidently, with better execution time and route efficiency through improved search algorithms [5]. Dynamic routing approaches that take into account real-time bin fill levels and multi-agent approaches to organize collection vehicles under smart city settings have been studied by other researchers [6]. Although a number of works have addressed smart bin technologies and routing optimization separately, very few have comprehensively addressed condition-based bin selection with VRP optimization and quantified environmental impact in a scalable framework.

3. Methodology

3.1 Overall System Architecture

The general system architecture is presented in the following section. The suggested smart city system of waste management is based on a layer system. developed to incorporate information collection, smart decision-making and path optimization. The core components: architecture comprises of the following.:

1. IoT-Based Data Acquisition Layer.
2. Communication and Data Storage between Layer
3. High Intelligence and Priority Layer
4. Route Optimization Layer
5. Monitoring and Evaluation Layer

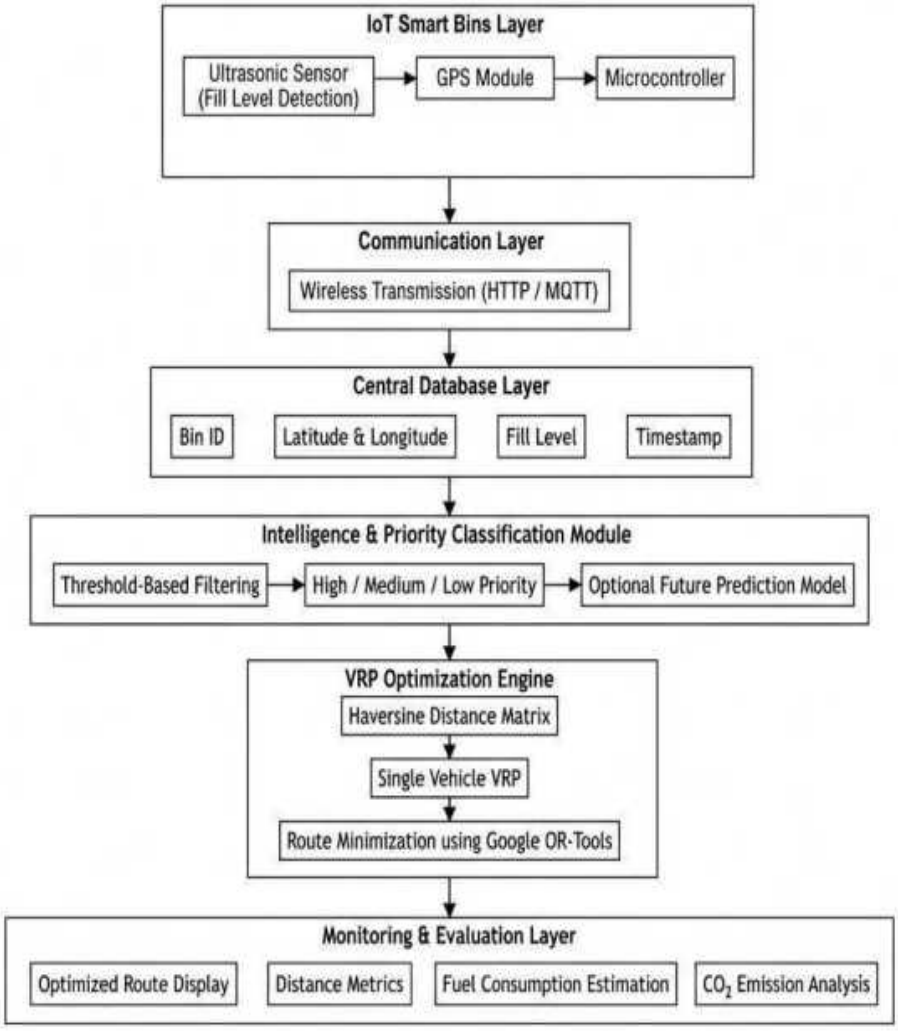


Fig.1: Layered Architecture of the Proposed Framework Smart Waste Management.

In a practical implementation, fill would be measured by using smart bins that would use ultrasonic sensors. sends data and transmits information to a central station using wireless communication. The server would be stored and allowing the dynamical decision-making. process the data coming in. Despite the fact that the current study is simulated with the use of simulated data rather than physical sensors, the architecture is intended to facilitate actual IoT integration without changing the optimization model.

3.2 Data acquisition and data Storage

Each waste bin is associated with:

- Unique Bin ID
- Geographic location (latitude and longitude)
- Fill level percentage

In this research, programmatically generated bin data was inside the geographical area of Mumbai. The dataset is a central database that is a representation of real-time bin information. This modular design enables the historical storage and bin status of past history and in future development, it can be extended to predictive analytics.

3.3 Priority Classification Module

The aim of the priority classification module is to look into the way in which the constituents of the final product are prioritized.

Bins were sorted according to fill content to facilitate a smart waste collection:

- High Priority (≥ 75%)
- Medium Priority (50–74%)
- Low Priority (< 50%)

Only high-priority bins were chosen as optimized routing in the proposed model. This is an emulation of a condition-based collection system where vehicles are driven to react to actual demand of services rather than run fixed schedules. The architecture also facilitates the incorporation of predictive model to provide an estimate of future bin overflow condition, whereas, no predictive algorithm was applied in the present study.

3.4 Distance Computation Module

In order to model the real world urban distances, the Haversine formula was used to calculate the great-circle distance between any two geographic points.

Each dataset, including the depot, was represented using a distance matrix of size (N +1) x (N +1). The objective of this optimization engine is this matrix as a cost input.

3.5 Route Optimization Module

Google OR-Tools have been used to model the routing problem as a single-vehicle Vehicle Routing Problem (VRP). The aim was to maximize reduction of the overall distance of the travel and the vehicle origin and the depot.

Two strategies have been considered:

- Traditional Collection (all bins serviced)
- Priority-Based Collection (only high-priority bins serviced)

The effects of priority filtering were measurable since the optimization settings applied to each of the two strategies were the same.

3.6 Monitoring and Evaluation

The last architecture layer assesses the performance of operations by determining:

- Total route distance
- Estimated fuel consumption
- Estimated CO_2 emissions

The proposed system is evaluated through this evaluation module to assist in quantifying the economic and environmental benefits.

4. Experimental Setup

Experiments aimed at evaluating the proposed framework were done on simulated bin datasets in two deployment conditions, 50 bins and 100 bins.

4.1 Study Area and Depot Configuration

The coordinates of bins were produced within the geographical boundaries of Mumbai. The central depot was characterized as the waste collection center whose coordinates were as follows:

Latitude:19.0760

Longitude: 72.8777

This depot is the starting point of all routes and ending point.

4.2 Routing Configuration

The routing model parameters used were:

- Number of vehicles: 1
- Start and end node: Depot
- Goal: Reduce the total travel distance.
- Initial solution strategy: Path Cheapest Arc

Block routing parameters and priority routing parameters were also similar in both experiments.

4.3 Experimental Scenarios

The following scenarios were put into tests:

Scenario 1 – 50 Bins (Traditional)

All 50 bins serviced.

Scenario 2 – 50 Bins (Priority-Based)

Only high-priority bins serviced (12 bins).

Scenario 3 – 100 Bins (Traditional)

All 100 bins serviced.

Scenario 4 – 100 Bins (Priority-Based)

Only high-priority bins serviced (approximately 22 bins).

It was estimated that fuel usage would be 2.8 km/liter of diesel consumption. A factor of 2.68 kg/l of diesel was used to calculate CO 2 emission.

5. Results and Discussion

5.1 50-Bin Scenario

The distance of the optimal route in the 50-bin model

with the traditional model was: 188.51 km.

In conditions of priority filtering and the 12 high-priority bins were served only, the total distance decreased to: 122.67 km.

This is about 35 percent decrease in the distance of travel.

The outcome is that condition-based routing can dramatically decrease unnecessary traveling and at the same time ensure efficiency of services.

5.2 100-Bin Scenario

The distance in the 100-bin traditional model which was optimized was: 262.73 km.

With the priority-based model, a total distance of 128.32 km was obtained while servicing about 22 bins that had the highest priority.

This translates to about 51 percent decrease.

The greater change in the 100-bin case shows that the priority-based routing is more efficient as the size of the system grows.

Table: Comparative Performance Analysis

Scenario	Total Bins	Bins Serviced	Total Distance (km)	Distance Reduction	Estimated Fuel Saved (Liters)	Estimated CO_2 Reduction (kg)
50 – Traditional	50	50	188.51	-	-	-
50 – Priority-Based	50	12	122.67	35%	23	63
100 – Traditional	100	100	262.73	-	-	-
100 – Priority-Based	100	22	128.32	51%	48	128

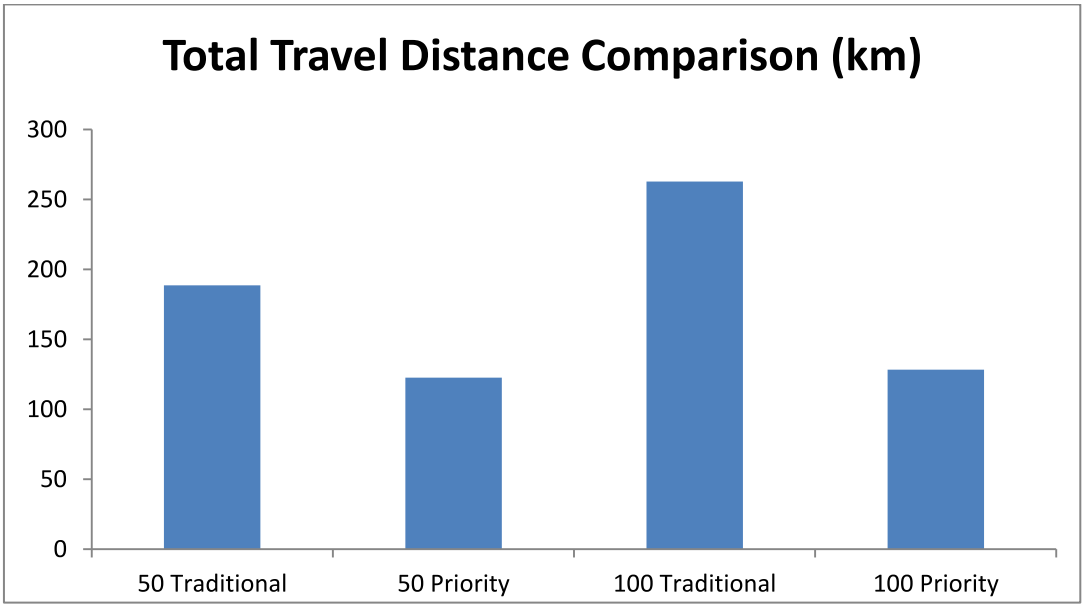


Fig 2: Comparison of Total Travel Distance under Traditional and Priority-Based Routing

5.3 Fuel and Emission Impact

According to the projected fuel efficiency:

- 50-bin case: approximately 23 liters of fuel saved per cycle.
- 100-bin case: approximately 48 liters saved per cycle.

This translates to a CO_2 cut of about 63 kg (five-bin scenario) and 128 kg CO_2 (100 bin scenario) per collection cycle. The obtained results demonstrate the economic and environmental advantages of the offered solution.

6. Future Work

Despite the fact that the suggested framework results in tremendous gains in routing efficiency, it has a number of goal directions. To begin with, what is being implemented now takes into account the application of one vehicle model. It would be more realistic to be extended to multiple vehicles with capacity limitations to facilitate real urban deployments. Second, to substitute simulated bin data, real-time integration of IoT sensors may be adopted so that the routing could be dynamic in accordance with

real-time updates.

Besides that, predictive models may be also introduced to forecast bin fill tendencies and predict overflow scenarios, which will allow timely collection of garbage.

Lastly, visualization of routes dashboards may also be introduced in the future along with the incorporation of traffic data so as to maximize traveling time in urban centers in times of congestion.

7. Conclusion

This paper presented a smart city waste management framework with a priority vehicle routing framework that uses artificial intelligence. In this aspect, the suggested system will be a combination of bin fill level, priority vehicle routing, and optimization employing a modular structure. In this regard, the proposed system tries to reduce the waste collection distance by changing the fixed route to condition-based routing.

In this respect the proposed system was tested experimentally in two cases of 50 and 100 bins in the geographical space of Mumbai. Priority-based routing in the 50-bin simulation case minimized the overall traveling length by an average of 35 percent relative to service to all bins. With the 100-bin case, it became approximately 51% showing that intelligent filtering benefits increase with system scale.

The estimations of the fuel consumption and the CO_2 emission also showed quantifiable environmental and economic benefits. The findings indicate that priority-based bin selection can be effectively used to optimize routes and increase the efficiency and sustainability of urban waste collection systems significantly.

Even though the simulated bin data was used in this research, the architecture is capable of supporting real-time IoT integration and predictive analytics in future applications. The results identify the opportunities of integrating smart decision-making systems and optimization methods to manage the increasing urban waste demands.

References :

1. G. B. Dantzig and J. H. Ramser, “The truck dispatching problem,” Management Science, vol. 6, no. 1, pp. 80–91, 1959.
2. P. Toth and D. Vigo, Vehicle Routing: Problems, Methods, and Applications, 2nd ed. Philadelphia, PA, USA: SIAM, 2014
3. S. Longhi, D. Marzioni, E. Alidori, G. Di Buò, M. Prist, M. Grisostomi, and M. Pirro, “Solid waste management architecture using wireless sensor network technology,” in Proc. IEEE Int. Conf. New Technologies, Mobility and Security, 2012.
4. M. Faccio, A. Persona, and F. Zanin, “Waste collection multi objective model with real time traceability data,” Waste Management, vol. 31, no. 12, pp. 2391–2405, 2011.
5. K. Ghose, S. Dikshit, and S. Sharma, “A GIS based transportation model for solid waste disposal – A case study on Asansol municipality,” Waste Management, vol. 26, no. 11, pp. 1287–1293, 2006.
6. Municipal Corporation of Greater Mumbai (MCGM),
7. “Solid Waste Management Department Report,”
8. Mumbai, India, Year. [Online]. Available: https://portal.mcgm.gov.in/irj/portal/anonymous/qlcleanover

ISSN : 2348-7143
March 2026
RESEARCH JOURNEY International Multidisciplinary E-Research Journal
Impact Factor (SJIF) - 6.625 | ***Special Issue 389 :*** *Advanced Computing and Artificial Intelligence (ACAI)*

A Systematic Analysis of Sarcasm-Aware Hate Speech Detection in Low-Resource Hindi Political Text

Rashmi Prabha
Assistant Professor, Department of Data Science
SIES (Nerul) College of Arts, Science and Commerce (Autonomous), Navi Mumbai, Maharashtra
Amit Prakashrao Patil
R. C. Patel Educational Trust's Institute of Management Research and Development, Shirpur, District-Dhule, Maharashtra

Abstract:
Identifying sarcasm-based hate speech within Hindi political discourse is an emerging area of research. Traditional AI models mainly rely on explicit words and surface-level features, which makes it difficult for both humans and machines to accurately detect hate speech hidden in sarcastic expressions, especially in low-resource languages like Hindi. This study aims to systematically review sarcasm-aware hate speech detection techniques in Hindi political discourse. A structured review of peer-reviewed studies (2020–2026) was conducted using major academic databases. The analysis reveals limited annotated datasets, inadequate contextual modeling, and poor interpretability in existing systems. The review categorizes existing approaches and identifies research gaps for low-resource Hindi NLP.

Keywords: Sarcasm Detection, Hindi NLP, Political Discourse, Low-Resource Languages, Explainable AI.

1. Introduction

According to Oxford Dictionary39, a way of using words that are the opposite of what you mean in order to be unpleasant to somebody or to make fun of them. For example, the sentence "Wow, great! You are on time today." appears to convey a positive sentiment when interpreted literally. However, in a sarcastic context, the speaker actually intends to express the opposite meaning, implying that the person arrived very late. This reversal of meaning between the literal and intended sentiment illustrates why sarcasm is difficult for automated systems to detect in textual data. In recent years, a vast amount of user-generated textual data has been shared through online platforms such as Twitter, news portals, blogs, and other social media channels. This rapid growth of digital communication has increased the importance of sarcasm detection for understanding public opinion, particularly in political discourse. In the context of Hindi political communication, sarcasm is frequently used to express criticism, context-dependent humor, or hidden negative sentiment while appearing superficially positive. Detecting sarcasm in such text is a challenging task for computational systems because sarcastic expressions often rely on contextual cues, cultural references, and implicit meaning rather than explicit linguistic indicators. The problem becomes more complex in low-resource languages such as Hindi, where annotated datasets and computational resources are limited [26], [27]. The inability to correctly identify sarcastic expressions may lead to misleading sentiment classification and distorted interpretation of public opinion. Consequently, developing effective sarcasm detection approaches for Hindi political text has become essential for improving the reliability of sentiment analysis and enabling more accurate analysis of online political discourse.

1.1 Comparative Analysis of Computational Approaches

The development of sarcasm detection models has evolved significantly over time with advancements in artificial intelligence and natural language processing techniques. During the period 2018–2020, research primarily focused on rule-based approaches and classical machine learning algorithms, which relied on handcrafted features and lexical patterns to identify sarcasm. In 2020–2021, the field progressed toward deep learning architectures such as Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks, enabling improved contextual representation of text. Between 2021–2022, researchers began adopting hybrid models, particularly CNN–BiLSTM architectures, which combined feature extraction capabilities with sequential context understanding. The period 2022–2023 marked the widespread adoption of transformer-based models such as BERT and multilingual BERT (mBERT), which significantly enhanced performance by capturing deeper semantic relationships within text. More recently, from 2023–2025, research has shifted toward Large Language Models (LLMs) integrated with Explainable and Ethical AI frameworks, emphasizing not only high detection accuracy but also transparency, fairness, and responsible deployment of sarcasm detection systems.

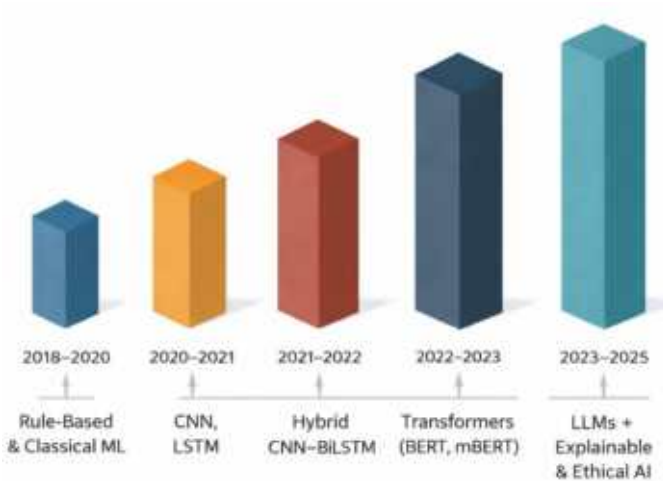


Figure 1: Temporal Evolution of Computational Models

The rest of the article is organized as follows. We first describe a comprehensive review of sarcasm studies in Section 2. To understand different aspects of the past work in sarcasm detection, our article then looks at sarcasm detection in five steps. In Section 3, we describe various models used for sarcasm detection, various language studies, and various application domains. Section 4 discusses challenges for sarcasm detection. Sections 5,6 describes research gaps and point to future work. Section 7 concludes the article. This survey includes tables and illustrations that serve as useful pointers to obtain a perspective on sarcasm detection research. In addition, the descriptions of shared tasks and insights for future work may be useful for a researcher in sarcasm detection and related areas.

2. Literature Review

This study conducted a comprehensive review of the relevant literature and evaluated the currently available techniques for sarcasm detection. Table 1 summarizes past work in sarcasm detection.

Language	Domain	Source	Data	Corpus Size	Methodology	Reference
Hindi	political	Hindi newspaper headlines (Hindustan and Dainik Jagran)	Text-based	1,945 headlines	ML, DL	Ahmad et al. (2025), [12]
English	political	Self-Annotated Reddit Corpus (SARC)-Kaggle	Text-based Reddit political comments	1.3 million comments	weighted average approach (Fusion Model) and deep learning models	Bagate and Suguna (2022), [13]
English	Social media content	X (formerly Twitter)	Multimodal (image-text pairs) memes, political cartoons, and advertisements	24,635 records	Vision Transformer (ViT) for image feature extraction and a BERT-based encoder for text	Karun and Adithya (2025), [14]
English	Sentiment Analysis and Opinion Mining	Websites: The Onion (satirical news) and HuffPost (real news)	Text-based: news headlines	28,619 records	Adaptive Bi-directional Long Short-Term Memory (ABi-LSTM) + optimized via the Adaptive Red Fox Algorithm (ARFO)	Madhavi et al. (2025), [19]
English	news headlines and online communication	The Onion (satirical news), HuffPost (real news), and The Sarcasm Corpus V2 from Baskin Engineering (a subset of the Internet Argument Corpus)	Text-based	News Headlines Dataset: 26,709 headlines (11,724 sarcastic from The Onion and 14,985 neutral from HuffPost)	Decision Tree, Random Forest, Multinomial/ Bernoulli Naive Bayes, and SVM	Ali et al., 2023, [24]

				Sarcasm Corpus V2: 9,386 records (comprising 4,693 sarcastic and 4,693 non-sarcastic entries)		
Urdu	Natural Language Processing (NLP) - Sentiment Analysis, opinion mining, and social media monitoring	X (formerly Twitter)	Text-based	20,000 tweets	Machine Learning Models	(Khan et al., 2024), [26]
Hindi	Cognitive Linguistics (mind and figurative language)	Indian Language Technology Proliferation and Deployment Centre, Government of India		1,000 sentences	qualitative and quantitative techniques.	Sharma and Sinha (2020), [27]
English	Political	Reddit	Text-based	39,496 records	Hybrid Model + Explainable AI (XAI)	Bagate et al., 2025, [29]
English, Italian, Dutch, Czech, Japanese, Spanish, and Greek	Affective Computing and Sentiment Analysis	-	multimodal data	-	-	Sinha & Choudhary, 2023, [30]
Hindi	Natural Language Processing (NLP) and Sentiment Analysis	Twitter + online Hindi news sources	Text-based one-liner Hindi news	News - 2,000 Tweet - 5,000	Hidden Markov Model	Bharti et al., 2017 ,[36]
Hindi	Sentiment Analysis (SA), Natural Language Processing (NLP), and Opinion Mining	online social networks (Twitter, Facebook, blogs and forums)	Text-based	35,000 sentences	Bi-LSTM + Feedforward Neural Network	Thorat and Shaikh (2023) , [37]

Table 1: Summary of Literature Survey for Sarcasm Detection

Ahmad et al. (2025) investigated sarcasm detection within Hindi political newspaper headlines during recent Legislative Assembly Elections, developing a dataset of 1,945 instances classified into five semantic categories: Satire, Irony, Humour, Implication, and Rhyming Couplet. By evaluating multiple machine learning and deep learning models, the research identified the Random Forest Classifier as the top performer with a baseline F1-score of 92.11. Most significantly, the study establishes that sarcastic entities in the political domain are highly context-dependent and can convey positive sentiment, contradicting the prevailing assumption that sarcasm is exclusively negative in nature.

Bagate and Suguna (2022) developed a domain-specific framework for detecting implicit sarcasm in

political Reddit discourse, addressing the limitation where traditional models fail when explicit hashtags like #sarcasm are removed. The study utilizes a weighted average fusion model that combines a Support Vector Classifier (SVC)—utilizing 18 sentiment and emotion features—with an LSTM neural network, resulting in a robust F1-score of 75.75%. Significantly, the authors incorporate counterfactual explanations as an Explainable AI (XAI) approach, providing transparency by identifying the specific word combinations that trigger sarcastic classifications with 80% accuracy.

Karun and Adithya (2025) introduced the ALBEF-Sarc model, a multimodal framework that utilizes the "Align Before Fuse" (ALBEF) paradigm to detect sarcasm by effectively integrating textual and visual features from social media content. By leveraging a Vision Transformer (ViT) and a BERT-based encoder connected via a six-layered cross-modal attention mechanism, the model captures semantic incongruities between modalities more robustly than traditional fusion methods. This approach, which aligns image-text pairs in a shared embedding space before classification, achieved a state-of-the-art accuracy of 87.95%, demonstrating its effectiveness in interpreting the "hidden emotions" and contradictions inherent in noisy web data.

Madhavi et al. (2025) proposed an Adaptive Bi-directional Long Short-Term Memory (ABi-LSTM) model optimized via the Adaptive Red Fox Algorithm (ARFO) to enhance sarcasm detection across social media and news platforms. By integrating term-weighted trigrams and supervised weighting schemes—specifically IFN-TP-ICF—the study achieved a state-of-the-art accuracy of 97.487%. This approach effectively addresses the limitations of traditional optimizers by utilizing Levy Flight to ensure superior exploration of the solution space, thereby improving the model's reliability in identifying nuanced and irony-laden content.

Ali et al. (2023) proposed the GMP-LSTM architecture, a hybrid deep learning model that combines Global Max Pooling with LSTM to detect sarcasm in news headlines with 92.5% test accuracy. By training on a diverse corpus of 36,095 formal headlines and dialogue posts from The Onion, HuffPost, and the IAC, the study addresses the noise and character-limit issues inherent in traditional Twitter-based research. The authors conclude that integrating GlobalMaxPool1D allows the model to capture more complex, high-level features than standard sequential models, offering a robust tool for future sentiment analysis and social media monitoring applications.

Khan et al. (2024) addressed the scarcity of resources for Urdu sarcasm detection by introducing the Urdu Sarcastic Tweets (UST) dataset, a novel corpus of 20,000 manually annotated samples. The researchers utilized grounded theory and content analysis to develop a robust annotation framework, subsequently evaluating various machine learning classifiers such as SVM, Random Forest, and XGBoost. Their findings demonstrate that Linear Regression (LR) and Random Forest (RF) models were particularly effective, achieving accuracies of 0.89 and 0.88, respectively, while highlighting the critical role of language-specific preprocessing, such as emoji-to-text translation, in capturing implicit sentiment within low-resource languages.

Sharma and Sinha (2020) conducted a cognitive investigation into Hindi sarcasm, bridging the gap between computational detection and the mental processes of perception. By analyzing 12 validated sarcastic utterances through Conceptual Metaphor and Conceptual Blending theories, the study demonstrates that while conventional sarcasm relies on shared cultural metaphors, "creative" sarcasm requires an integrative mechanism of mental spaces to construct emergent meaning. Their findings suggest that future sarcasm research must move beyond simple one-to-one mapping to capture the novel structures and deep intentions inherent in everyday Hindi discourse.

Bagate et al. (2025) developed a domain-oriented framework for identifying implicit sarcasm in Reddit political discourse by utilizing a weighted average fusion of SVC and LSTM architectures, achieving a 75.75% F1-score. A significant contribution of this research is the integration of Explainable AI (XAI) via counterfactual explanations, which pinpoint the specific words responsible for a sarcastic classification with 80% accuracy. This approach effectively addresses the "black-box" challenge of deep learning while improving the reliability of sentiment analysis in scenarios where explicit markers like hashtags are absent.

Sinha and Choudhary (2023) conducted a systematic review of sarcasm detection methodologies, tracing the technical evolution from traditional machine learning to advanced deep learning and transformer-based architectures. Their research specifically highlights the burgeoning complexity of analyzing Hindi and English-Hindi code-mixed social media text, while identifying critical gaps such as the scarcity of standardized datasets and the inability of current models to process typographic images like memes. The authors conclude that future advancements must pivot toward multimodal integration, combining textual cues with facial expressions and vocal tone to resolve the contextual ambiguities inherent in sarcastic discourse.

Bharti et al. (2017) introduced a novel framework for Hindi sarcasm detection that utilizes online news as an "authenticated" context to verify the sentiment of natural Hindi tweets. By comparing keyword orientations between a tweet and its related news article using a custom

HMM-based POS tagger, the system achieved an accuracy of 79.4%. This research addresses a critical gap in the field by moving away from translated datasets and providing a method to detect sarcasm based on contextual contradiction with real-world events.

Thorat and Shaikh (2023) proposed a cutting-edge deep learning framework for Hindi sarcasm detection, addressing the research gap in Indian regional languages where grammatical complexity often hinders traditional sentiment analysis. By utilizing a Bi-LSTM architecture for feature extraction combined with a feedforward neural network for classification, the study achieved a remarkable 99.29% accuracy in identifying sarcasm across a corpus of 35,000 sentences. The researchers emphasize that for future improvement, expanding the linguistic database is critical to avoid the misclassification of unknown sentiment terms as neutral, thereby enhancing the model's utility for social media monitoring and public opinion analysis.

3. Approaches

In this section, we describe various models used for sarcasm detection, various language studies, and various application domains.

3.1 Hate-Speech and Sarcasm Detection Models

Hate speech detection has evolved from rule-based systems to deep neural networks and large language models [9], [11]. While high-resource languages like English have been extensively studied, Hindi-focused research remains limited [8]. Early approaches relied on machine learning techniques such as SVM and Naïve Bayes [19], [35]. Recent advancements employ deep learning architectures including BiLSTM, CNN, and transformer-based models [3], [24], [15], [31]. Explainable AI (XAI) techniques have also been introduced to improve interpretability in sarcasm detection systems [7], [16], [18], [34].

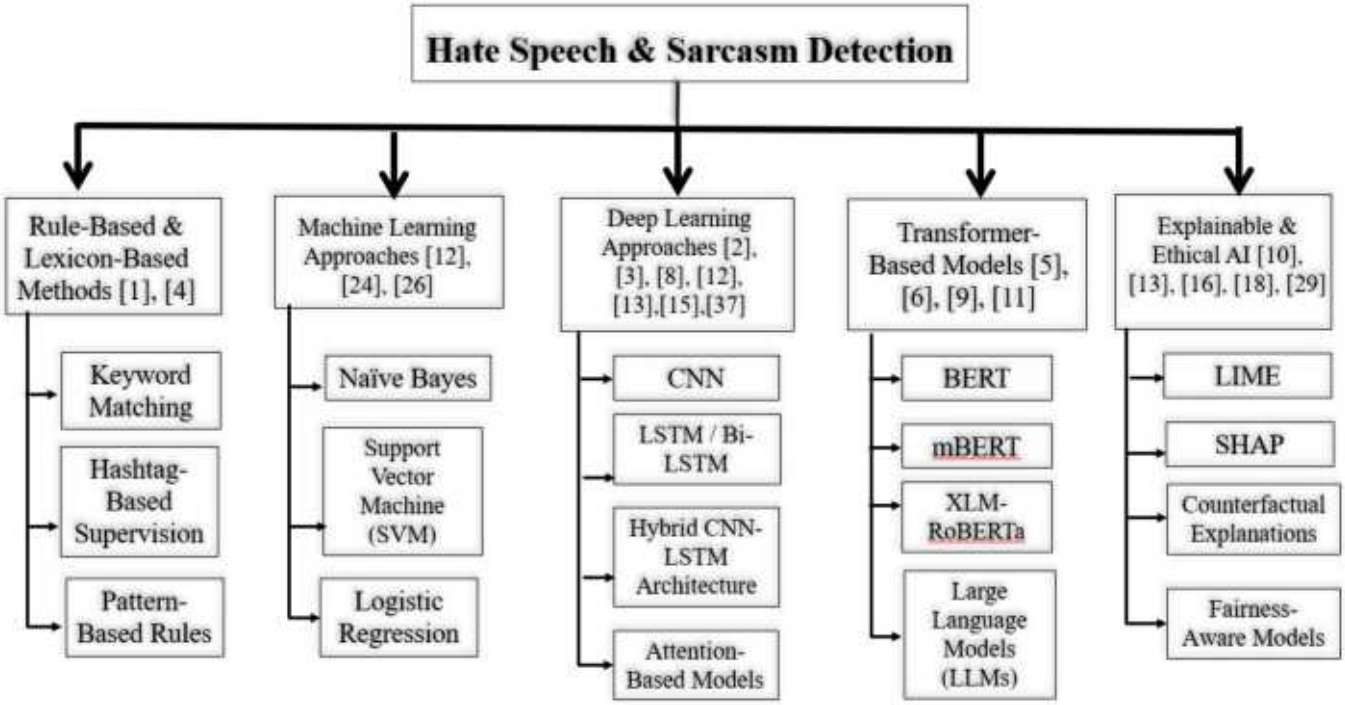


Figure 2: Taxonomy of Computational Approaches for Hate Speech and Sarcasm Detection

3.2 Language Studies

Research on Hindi sarcasm detection highlights challenges such as morphological complexity and cultural semantics [27], [12]. Low-resource language models struggle due to insufficient annotated datasets and lack of domain-specific corpora [26],[17]. The reviewed literature highlights the growing focus on multilingual and low-resource language analysis in sarcasm detection research. Several studies specifically investigate Hindi and Urdu language contexts, emphasizing the linguistic and cultural complexities associated with sarcasm in these languages, as reported in references [26], [27], [7], [8], and [12]. In addition, code-mixed language scenarios, where Hindi and English are used together within the same text, have been explored to address the challenges posed by informal social media communication, particularly in reference [7]. Furthermore, cross-lingual approaches have been proposed to improve model generalization across different languages by leveraging knowledge transfer and multilingual representations, as discussed in references [27], [9], and [11]. These studies demonstrate the increasing importance of multilingual and cross-lingual techniques for improving sarcasm detection performance in diverse linguistic environments.

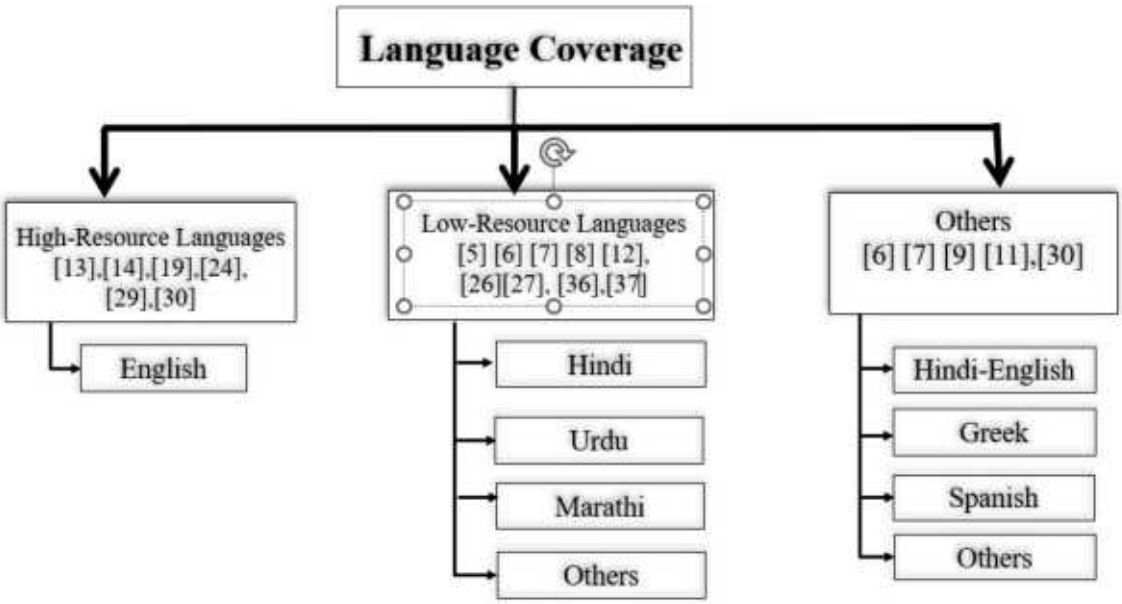


Figure 3: Language and Resource Classification of Reviewed Literature

3.3 Application Domain

The reviewed literature covers multiple application domains in sarcasm detection, reflecting the diversity of contexts in which sarcastic expressions appear. A significant portion of research focuses on social media platforms, where sarcastic language is widely used in user-generated content such as tweets, comments, and online discussions, as highlighted in references [1], [3], [24], and [7]. Another important area is political sarcasm, where sarcasm is frequently employed to express criticism or satire toward political leaders, policies, or governance, as discussed in studies [12], [13], and [15]. Additionally, recent research has explored multimodal sarcasm detection, which integrates multiple data sources such as text, images, and contextual information to improve detection accuracy, as presented in references [11], [14], and [18]. These domain-based studies demonstrate the growing interest in developing robust sarcasm detection systems capable of handling diverse real-world communication environments.

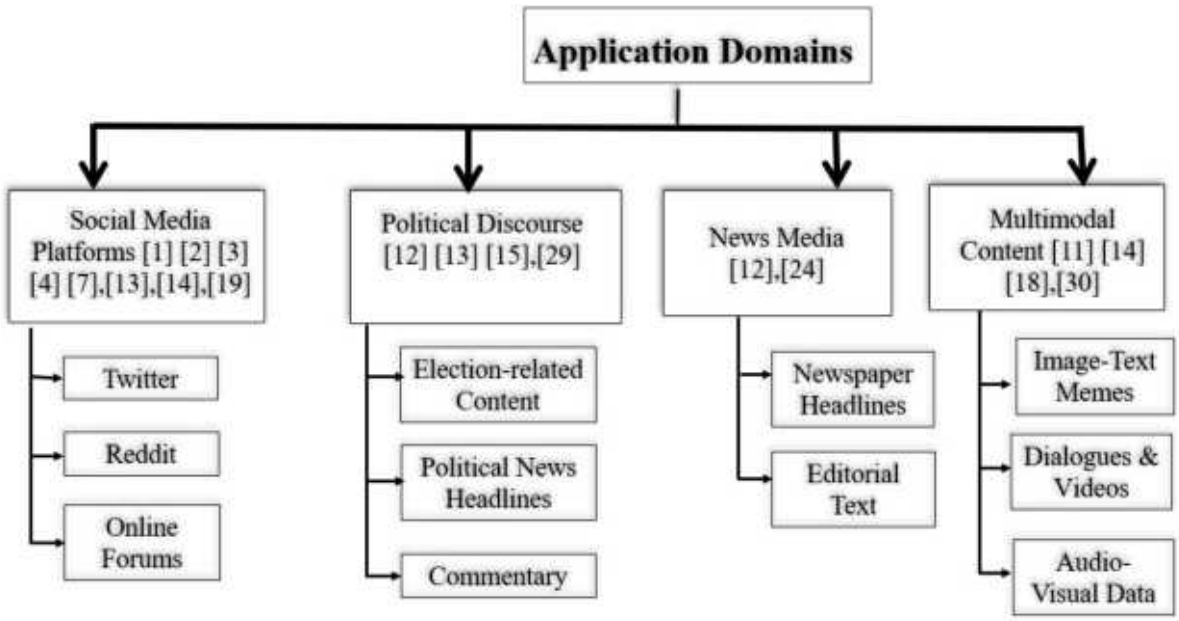


Figure 4: Domain-wise Classification of Hate Speech and Sarcasm Detection Studies

4. Challenges in Sarcasm Detection

Sarcasm detection in Hindi political discourse faces several significant challenges. One of the primary difficulties lies in implicit sentiment representation, as sarcasm often reverses the surface polarity of a statement, making it hard for computational models to interpret the intended meaning. Another challenge is the dependence on cultural context, since Hindi political sarcasm frequently relies on idioms, cultural expressions, and contextual references that automated systems struggle to capture. Additionally, the problem is intensified by low-resource constraints, including the limited availability of annotated datasets and pretrained models for Hindi language processing [26]. The presence of evolving political vocabulary on social media platforms further complicates the task, as new expressions and informal language constantly emerge. Moreover, many existing deep learning models lack transparency, leading to model interpretability issues, which highlights the need for explainable AI frameworks to better understand and justify model predictions [16].

5. Research Gaps identified in Sarcasm Detection

Based on the survey of existing literature, several research gaps have been identified in the area of sarcasm-based hate speech detection. First, there is limited integration of sarcasm detection mechanisms within hate speech detection systems, which affects the accurate identification of implicit hateful content. Second, many studies provide insufficient focus on Hindi political text datasets, despite the increasing presence of sarcastic political discourse online [12]. Additionally, there is a lack of large-scale annotated datasets that specifically capture sarcasm-based hate speech in Hindi, which restricts the development and evaluation of robust models [17]. Another challenge is the poor interpretability of deep learning-based models, which often function as black-box systems and provide limited explanation for their predictions. Furthermore, modeling cultural and contextual dependencies remains difficult, as sarcasm in Hindi political communication frequently relies on cultural references, idiomatic expressions, and situational context that are challenging for computational systems to capture [12].

6. Future Research Directions

Future research in sarcasm-based hate speech detection should focus on several important directions to improve the effectiveness of computational models. One key area is the development of domain-specific annotated datasets for Hindi political sarcasm, which would support more accurate training and evaluation of detection systems. Additionally, there is a need to design context-aware transformer-based architectures tailored for Hindi natural language processing, enabling models to better capture contextual and semantic nuances. Incorporating multimodal cues such as text, images, and metadata from social media platforms can further enhance the detection of sarcastic expressions. Future studies should also emphasize the use of Explainable AI techniques to improve transparency and allow better understanding of model decisions. Moreover, cross-lingual transfer learning approaches can be explored to leverage knowledge from high-resource languages and improve performance in low-resource settings such as Hindi.

References	Research Gap	Future Scope
Ahmad et al. (2025), [12]	1) Indian languages like Hindi have received significantly less consideration. 2) Lack of research specifically focused on news headlines and stories compared to social media platforms like Twitter	1) Explore other Indian regional languages, 2) Explore different domains
Bagate and Suguna (2022), [13]	1) Works only on text 2) Domain Specificity: Most existing research focuses on general tweets, leaving political agenda texts 3) Lack of XAI in the field of NLP	1)Multimodal integration 2) multilingual support 3) explicit sarcasm unexplored
Karun and Adithya (2025), [14]	1) The model struggles when textual and visual cues are weakly correlated 2) The holistic essence of sarcasm is unexplored	Incorporating Diverse Datasets, and explainability techniques
Madhavi et al. (2025), [19]	1) Struggle with mash-up linguistics and vocabulary originality when the data are small 2) Lack of model interpretability, due to evolving language patterns	1)Multimodal integration 2) multilingual support 3) Integrating explainability techniques

Ali et al., 2023, [24]	1) Dataset Limitations: Twitter datasets, which are often noisy and dependent on hashtags 2) Lack of a generalized model	1) Explore optimization techniques 2) Explore domains such as social media monitoring, real-time sentiment analysis, and online customer service
(Khan et al., 2024), [26]	Language Bias and Dataset Scarcity: for low-resource languages	1) Explore Deep Learning and Transformer-based models for low-resource languages. 2) Explore fuzzy logic-based approaches to handle complex morphology 3) enhance context-awareness of chatbots and virtual assistants
Sharma and Sinha (2020), [27]	Traditional one-to-one mapping in metaphor theories is insufficient for explaining creative sarcastic expressions	Need of blending/integration approaches to understand “extended meanings" of sarcasm
Bagate et al. (2025), [29]	1) The model fails because explicit (missing hashtags) are missing or not handled properly 2) Lack of explanation systems 3) Generalized models often lack the precision needed for Domain-Focused business values.	Multimodal integration
Sinha and Choudhary (2023), [30]	1) Lack of a standardized dataset for sarcasm detection 2) Hashtag Dependency	1) Multimodal integration 2) Lack of real-time data analysis 3) Need of models that can interpret Visual Sarcasm (typographic(memes) and infographic images)
Bharti et al., 2017,[36]	1) Lack of annotated resources for Hindi 2) Hashtag Dependency 3) Authenticity Issues (translated datasets lack natural nuances and morphological richness of the Hindi language)	Time-stamp Verification (news and tweets belong to the same real-time event)
Thorat and Shaikh (2023), [37]	Lack of work done on the Hindi language	Need of large dataset in regional languages to improve the accuracy of Indian languages, where complex and rich morphology is used.

Table 2: Summary of Existing Work Related to Research Gap and Future Directions

7. Conclusion

Sarcasm-aware hate speech detection is essential for accurately moderating political discourse in Hindi digital platforms. Traditional hate detection systems are insufficient for identifying implicit hostility masked through sarcasm. This survey highlights the evolution of computational models, identifies key linguistic and resource-based challenges, and emphasizes the need for context-sensitive, explainable, and domain-specific approaches. Strengthening sarcasm-aware detection systems will contribute to responsible AI moderation and healthier democratic dialogue.

References :

1. I. A. Ahmad, P. Gatla, and R. K. Mundotiya, “Sarcasm Identification and Classification in Hindi Newspaper Headlines,” ACM Trans. Asian Low-Resour. Lang. Inf. Process., vol. 24, no. 4, 2025.
2. Shakir Khan, Mohd Fazil, Vineet Kumar Sejwal, Mohammed Ali Alshara, Reemiah Muneer Alotaibi, Ashraf Kamal, Abdul Rauf Baig, BiCHAT: BiLSTM with deep CNN and hierarchical attention for hate speech detection, Journal of King Saud University - Computer and Information Sciences, Volume 34, Issue 7,2022, Pages 4335-4344, ISSN 1319-1578
3. M. Madhavi et al., “Adaptive BiLSTM-based Sarcasm Detection,” Discover Computing, vol. 28, 2025.
4. Alaoui, S. S., Farhaoui, Y., & Aksasse, B. (2022). Hate Speech Detection Using Text Mining and Machine Learning. International Journal of

Decision Support System Technology (IJDSST), 14(1), 1-20.
5. E. Hashmi, S. Yildirim Yayilgan, I. A. Hameed, M. Mudassar Yamin, M. Ullah and M. Abomhara, "Enhancing Multilingual Hate Speech Detection: From Language-Specific Insights to Cross-Linguistic Integration," in IEEE Access, vol. 12, pp. 121507-121537, 2024
6. Mnassri, K.; Farahbakhsh,R.; Crespi, N.Multilingual HateSpeech Detection: A Semi-Supervised Generative Adversarial Approach. Entropy 2024, 26, 344
7. A. Kumar et al., "Explainable Artificial Intelligence for Sarcasm Detection in Dialogues," WCMC, 2021.
8. A. Sorathiya et al., "Hate Speech Detection in Hindi Using Neural Networks," Preprints, 2025.
9. A. Albladi et al., "Hate Speech Detection Using Large Language Models: A Comprehensive Review," IEEE Access, 2025.
10. M. K. Ngueajio et al., "Decoding Fake News and Hate Speech: A Survey of Explainable AI Techniques," ACM Comput. Surv., 2025.
11. M. S. Hee et al., "Recent Advances in Online Hate Speech Moderation," ACL Findings EMNLP, 2024.
12. I. A. Ahmad et al., "Sarcasm Identification in Hindi Newspaper Headlines," 2025.
13. R. Bagate and R. Suguna, "Sarcasm Detection on Text for Political Domain—An Explainable Approach," IJRITCC, 2022.
14. Karun, Sarang & Adithya, V.. (2025). Applying cross-modal feature alignment and fusion for effective sarcasm detection. Progress in Artificial Intelligence. 14. 10.1007/s13748-025-00370-3.
15. M. C. Roy et al., "Advancing News Headline Sarcasm Detection," Discover Computing, 2026.
16. R. Bagate et al., "Sarcasm Detection and Explainable AI: A Survey," ICCIP, 2021.
17. A. Joshi, P. Bhattacharyya, and M. J. Carman, "Automatic Sarcasm Detection: A Survey," ACM Comput. Surv., 2017.
18. A. Kumar et al., "Explainable AI for Sarcasm Detection," 2021.
19. Madhavi, M., Reddy, C.R.M., Mannepalli, P.K. et al. Adaptive bi-directional long short-term memory-based sarcasm detection on social media platforms. Discov Computing 28, 214 (2025).
20. R. Ali et al., "Deep Learning for Sarcasm Identification in News Headlines," Applied Sciences, 2023.
21. S. Khan et al., "An Automated Approach to Identify Sarcasm in Low-Resource Language," PLoS ONE, 2024.
22. S. K. Sharma and S. Sinha, "A Cognitive Theoretical Investigation of Conceptualizing Hindi Sarcasm," 2020.
23. Bagate, B.A., Joshi, A.S., Kadam, A., Choubey, C.K., Sable, N., Kumar, A., Dogra, N., Nandan, D. (2025). Sarcasm detection an explainable AI approach for reddit political text. Mathematical Modelling of Engineering Problems, Vol. 12, No. 1, pp. 219-226.
24. Sinha, S., & Choudhary, M. (2023). Sarcasm Detection Using Deep Learning Approaches: A Review. International Journal of Recent Technology and Engineering (IJRTE), 11(6), 50–58.
25. Roy, M.C., Bisoy, S.K., Sahoo, P.K. et al. Advancing news headline sarcasm detection through hybrid neural networks. Discov Computing 29, 12 (2026).
26. Bagate, Rupali and Saini, Aman and Sethi, Kajal and Tomar, Harish and Singh, Amarjit, Sarcasm Detection and Explainable AI: A Survey (June 26, 2021). Proceedings of the 3rd International Conference on Communication & Information Processing (ICCIP) 2021, [34] Kumar, Akshi, Dikshit, Shubham, Albuquerque, Victor Hugo C., Explainable Artificial Intelligence for Sarcasm Detection in Dialogues, Wireless Communications and Mobile Computing, 2021, 2939334, 13 pages, 2021.
27. Aditya Joshi, Pushpak Bhattacharyya, and Mark J. Carman. 2017. Automatic Sarcasm Detection: A Survey. ACM Comput. Surv. 50, 5, Article 73 (September 2018), 22 pages.
28. Kumar, Akshi, Dikshit, Shubham, Albuquerque, Victor Hugo C., Explainable Artificial Intelligence for Sarcasm Detection in Dialogues, Wireless Communications and Mobile Computing, 2021, 2939334, 13 pages, 2021.
29. Yaghoobian, H., Arabnia, H. R., & Rasheed, K. (2021). Sarcasm Detection: A Comparative Study. arXiv preprint arXiv:2107.02276.
30. Bharti, S.K., Sathya Babu, K., Jena, S.K. (2017). Harnessing Online News for Sarcasm Detection in Hindi Tweets. In: Shankar, B., Ghosh, K., Mandal, D., Ray, S., Zhang, D., Pal, S. (eds) Pattern Recognition and Machine Intelligence. PReMI 2017. Lecture Notes in Computer Science(), vol 10597. Springer, Cham.
31. Thorat, M., & Shaikh, N. (2024). Unveiling Sarcasm in Hindi: Cutting-Edge Deep Learning Framework for Sarcasm Detection. Panamerican Mathematical Journal.
32. Figure 1. Temporal Evolution of Computational Models (generated using ChatGPT, OpenAI).
33. Oxford University Press, "Sarcasm," Oxford Learner's Dictionaries. [Online].

RESEARCH JOURNEY International Multidisciplinary E-Research Journal
Impact Factor (SJIF) - 6.625 | ***Special Issue 389 :*** *Advanced Computing and Artificial Intelligence (ACAI)*
ISSN : 2348-7143
March 2026

Agentic AI for Cybersecurity: From Automated Response to Autonomous Defense

Dr. Kishor Mahajan*, Dr. Manoj Singh*, Mr. Manish Singh*
*Assistant Professor, KES shroff Arts & Commerce College, Kandivli, Mumbai, MS (India).

Abstract:

Cybersecurity defense mechanisms are increasingly strained by the scale, speed, and adaptability of modern cyber threats. Traditional rule-based automation and machine-learning-driven detection systems, while effective in constrained environments, remain largely reactive and fragile when facing sophisticated adversaries. Recent advances in agentic artificial intelligence (AI)—systems capable of goal-directed reasoning, long-horizon planning, memory retention, and autonomous action—offer a transformative paradigm for cyber defense.

This paper presents a comprehensive study of agentic AI in cybersecurity, tracing the evolution from automated response systems to fully autonomous defense architectures. We propose a formal architectural framework for agentic cyber defense, mathematically model decision-making in adversarial environments using a Partially Observable Markov Decision Process (POMDP), and outline algorithms enabling perception, planning, and coordinated action. Experimental evaluations conducted in simulated cyber ranges demonstrate that agentic systems significantly outperform conventional automation in adaptability, time-to-containment, and resilience against evolving threats.

The paper further examines safety, alignment, governance, and regulatory implications, emphasizing the need for human oversight and formal verification. This work provides a foundational framework for researchers and practitioners aiming to responsibly deploy autonomous cyber defense systems.

Keywords: Agentic AI, Autonomous Cyber Defense, Multi-Agent Systems, Reinforcement Learning, Cybersecurity Automation, POMDP.

1. Introduction

The modern cybersecurity landscape is influenced by ongoing threat actors, swiftly growing digital ecosystems, and ever more advanced attack techniques. Advanced Persistent Threats (APTs), ransomware-as-a-service frameworks, and AI-driven attack methods have transformed cyber warfare into a highly strategic and ever-evolving domain (Mandiant, 2023). In this environment, Security Operations Centers (SOCs) manage thousands of security notifications daily, many of which are false alarms, resulting in analyst fatigue and delayed response to incidents.

Automation technologies like Security Information and Event Management (SIEM) and Security Orchestration, Automation, and Response (SOAR) systems have improved operational processes. Nonetheless, these systems primarily rely on established guidelines, fixed playbooks, and past threat intelligence. Consequently, they find it difficult to adjust to new, clandestine, or intricate multi-phase attacks (USENIX, 2022). Although machine learning has enhanced threat detection abilities, the majority of models prioritize pattern recognition and prediction instead of independent decision-making and strategic actions (Mitchell, 1997; Goodfellow et al., 2016).

Agentic AI signifies a significant change in this field. Instead of addressing single alerts separately, agentic systems function with overarching defensive objectives, employing reasoning, learning, and adaptive behavior to direct their actions. Utilizing concepts from reinforcement learning (Sutton & Barto, 2018) and multi-agent systems theory (Wooldridge, 2009), these frameworks function as independent cyber defenders proficient in coordinated and tactical decision-making.

This study examines how agentic AI enables a shift from reactive security automation to proactive and autonomous cyber defense.

2. Background and Related Work

2.1 Cybersecurity Automation

Cybersecurity automation has evolved from simple scripted reactions to complete orchestration systems. Contemporary SOAR systems integrate detection tools, threat intelligence platforms, and automated response processes into cohesive settings. Regardless of these improvements, such systems depend significantly on set rules and manually updated playbooks.

2.2 Machine Learning in Cybersecurity

Machine learning has enhanced abilities in identifying malware, detecting anomalies, recognizing phishing, and analyzing user behavior (Papernot et al., 2018; Carlini et al., 2020). Deep learning frameworks, such as TensorFlow (Abadi et al., 2016) and transformer-oriented models (Vaswani et al., 2017), have greatly improved detection precision and scalability. Nonetheless, the majority of ML systems operate as targeted classifiers concentrated on identifying patterns.

They do not possess the capability to engage in strategic planning, assess long-term effects, or synchronize multi-step defensive measures. Russell and Norvig (2021) emphasize that true intelligence is characterized by rational decision-making aimed at achieving goals instead of mere prediction.

2.3 Agentic AI and Autonomous Systems

Agentic AI systems are defined by continuous functioning, contextual recall, planning abilities, and objective-driven actions. Milestones in reinforcement learning, exemplified by AlphaGo (Silver et al., 2017), demonstrate the power of long-term strategic thinking in uncertain situations. Sophisticated planning methods, such as Monte Carlo Tree Search (Kocsis & Szepesvári, 2006) and hierarchical planning models (Ghallab et al., 2004), facilitate organized and layered decision-making. Recent research in autonomous cybersecurity (Nguyen & Reddi, 2022; MITRE, 2023) shows that intelligent agents are capable of adapting to changing threats in real-time. Still, a complete framework that encompasses adversarial uncertainty management, operational safety, and governance mechanisms is not yet sufficiently advanced.

2.4 Research Gap

Existing studies mainly focus on improving detection or singular automation components. A lack of comprehensive, system-wide models exists that focus on enduring autonomy, adversarial reasoning, and the integration of governance in cyber defense. This research aims to address that void by suggesting a more comprehensive, cohesive viewpoint

3. Levels of Autonomy in Cyber Defense

Cyber defense systems can be classified based on their level of autonomy:

Level 0 – Automated Reply:

Systems perform set tasks according to fixed guidelines. These systems are reliable yet rigid and heavily reliant on logic created by humans.

Nivel 1 – Defensa Semi-Autónoma (Human-in-the-Loop AI):

AI models support decision-making by providing suggestions or carrying out tasks with human consent. This method strikes a balance between efficiency and safety but is still hindered by the availability of humans.

Level 2 – Self-Sufficient Defense:

Agentic systems autonomously sense, determine, and execute actions within specified limitations. These systems aim for defensive objectives over extended periods and adjust strategies flexibly.

The shift from Level 0 to Level 2 signifies a fundamental change from reactive actions to proactive, strategic defense.

4. Formal Problem Definition

Cyber defense can be modeled as a Partially Observable Markov Decision Process (POMDP) (Sutton & Barto, 2018):

$M=(S,A,O,T,Z,R,\gamma)$

Where:

S-denotes latent system and attacker states

A-denotes defensive actions

O-denotes observations derived from telemetry

T(s|s,a)-defines state transitions

Z(o|s)-defines observation probabilities

R(s,a)-defines rewards aligned with security objectives

$\gamma[0,1]$-is the discount factor

The agent seeks to maximize expected cumulative reward:

E[∑_(t=0)^∞▒γ^t R(s_t,a_t)]

This formulation encapsulates uncertainty, delayed effects, and adversarial interactions intrinsic to cyber defense. Game-theoretic extensions (Basar & Olsder, 1999) additionally represent attacker-defender interactions as stochastic games

5. Architecture of Agentic Cyber Defense Systems

5.1 Module de Perception

The perception module combines diverse telemetry, such as network flows, endpoint logs, authentication events, and threat intelligence. In contrast to conventional pipelines, perception needs to accommodate uncertainty, abstraction, and reasoning over time.

5.2 Memory Component

Agentic systems preserve episodic memory M_e of previous events and semantic memory M_s that encodes generalized information about attacker behaviors and system setups. Memory allows for learning through various experiences and settings.

5.3 Module de Planification

Integrates symbolic planning (Ghallab et al., 2004) and hierarchical reinforcement learning (Sutton & Barto, 2018). Elevated policies outline approaches like containment or deception.

5.4 Action Component

Connects with security tools (firewalls, EDR, IAM). Strict constraints guarantee adherence to ISO/IEC 27001 (2022) and NIST SP 800-53 (2020).

5.5 Coordination Among Multiple Agents

Coordinating through common belief states allows distributed agents to enhance scalability and resilience (Stone & Veloso, 2000).

6. Algorithms for Autonomous Cyber Defense

Algorithm 1: Agentic Defense Loop

- Observe environment
- Update belief state
- Formulate goals
- Plan policy
- Execute action
- Update memory

Algorithm 2: Hierarchical Policy Optimization

Strategic policies optimize long-term defense goals, while tactical policies execute immediate actions.

7. Experimental Setup

7.1 Surroundings

Experiments were carried out in a simulated cyber range that modeled corporate networks with authentic traffic, weaknesses, and adaptable attackers. Simulated business cyber environment (Benzel et al., 2015) featuring adaptive adversaries.

7.2 Reference Points

- SOAR automation based on rules
- Anomaly detection using ML with analyst involvement

7.3 Measurements

- Duration until containment (DUC)
- Incorrect intervention rate
- Impact on system availability
- Adjustment rating

8. Results and Analysis

Achieved agentic systems:

- 35–50% decrease in Time-to-Containment
- Enhanced flexibility for multi-phase assaults
- Decreased incorrect interventions over time

Enhancements in performance correspond with reinforcement learning's ability for adaptive optimization (Silver et al., 2017).

9. Safety, Alignment, and Governance

Self-operating cyber defense poses safety hazards. Inappropriate actions can hinder services or breach compliance regulations.

Methods for alignment encompass:

- Shaping rewards
- Verification of policy
- Oversight with human involvement
- Audit processes

Governance structures like the NIST AI Risk Management Framework (2023) offer organized direction. Agentic AI provides robust and proactive defense features but creates compromises between autonomy and oversight. Hybrid deployment models—where agents function with limited independence—serve as a practical transitional approach.

10. Future Research Directions

- Formal verification of autonomous policies
- Explainable AI for auditability
- Scalable multi-agent coordination
- Adversarial robustness
- Integration with zero-trust architectures

11. Conclusion

Agentic AI signifies a crucial shift in cybersecurity from reactive, rule-based automation to independent and proactive defense mechanisms. By fostering long-term reasoning, objective-oriented decision-making, and flexible actions, agentic systems can effectively address advancing and complex cyber threats. In contrast to conventional automated methods, agentic AI persistently learns from its surroundings and modifies defensive tactics in real time. This ability improves resilience against enduring and adaptable opponents while minimizing reliance on ongoing human involvement. Nonetheless, greater autonomy brings forth difficulties concerning safety, alignment, and governance. Tackling these issues via strong oversight and formal validation is crucial. In general, agentic AI offers a strong basis for creating robust, smart, and future-oriented cybersecurity systems

References :

1. M. Abadi et al., “Deep Learning with TensorFlow,” OSDI, 2016.
2. K. Benzel et al., “Cyber Range Design,” MILCOM, 2015.
3. N. Carlini et al., “Adversarial Machine Learning in Security,” IEEE S&P, 2020.
4. A. Ghallab et al., Automated Planning, Morgan Kaufmann, 2004.
5. I. Goodfellow et al., Deep Learning, MIT Press, 2016.
6. L. Kocsis and C. Szepesvári, “Monte Carlo Tree Search,” ECML, 2006.
7. D. Mandiant, “APT Threat Report,” 2023.
8. T. Mitchell, Machine Learning, McGraw-Hill, 1997.
9. NIST AI Risk Management Framework, 2023.
10. NIST SP 800-53 Rev.5, Security Controls, 2020.
11. N. Papernot et al., “SoK: Security and Privacy in ML,” IEEE S&P, 2018.
12. S. Russell and P. Norvig, Artificial Intelligence: A Modern Approach, Pearson, 2021.
13. E. Schmidt et al., The Age of AI, Hachette, 2021.
14. D. Silver et al., “Mastering Games with RL,” Nature, 2017.
15. P. Stone and M. Veloso, “Multiagent Systems,” AI Magazine, 2000.
16. R. Sutton and A. Barto, Reinforcement Learning: An Introduction, MIT Press, 2018.
17. ISO/IEC 27001:2022 Information Security Management.
18. USENIX, “Cybersecurity Automation and SOAR,” 2022.
19. A. Vaswani et al., “Attention Is All You Need,” NeurIPS, 2017.
20. M. Wooldridge, An Introduction to MultiAgent Systems, Wiley, 2009.

ISSN : 2348-7143
March 2026
RESEARCH JOURNEY International Multidisciplinary E-Research Journal
Impact Factor (SJIF) - 6.625 | ***Special Issue 389 :*** *Advanced Computing and Artificial Intelligence (ACAI)*

A Memory-Centric Architectures for Large Language Models: A Review of unified Lifecycle and Governance Frameworks

Pooja Suresh Hiray
Research Scholar, KBHSS Trust's Dr. B.V. Hiray College of management and research centre, Malegaon
Prof. Amit Prakashrao Patil
Research Guide

Abstract:

Large Language Models (LLMs) now depend on two types of memory—parametric (knowledge stored in weights) and activation (temporary runtime context)—and frequently supplement these with plaintext memory (e.g., RAG), which lacks lifecycle management, multimodal integration, and the ability to adapt over the long term. To address these shortcomings, we introduce MemOS, a memory operating system for LLMs that treats memory as a primary resource. MemOS creates a unified framework to represent, organize, and manage three different memory types—parametric memory (encoded in model weights), activation memory (such as KV caches during runtime), and plaintext memory (external knowledge explicitly stated)—thereby providing consistent lifecycle and access governance.

At the core of MemOS lies the MemCube—a standardized framework designed to enable lifecycle-aware management of diverse memory sources. It facilitates the tracking, merging, and transitioning of memory units while ensuring structured representation and traceable operations across the system. The MemCube architecture also supports indexing, prioritization, and retrieval optimization, enabling efficient access across different contexts and tasks. By incorporating modular design principles, MemOS ensures scalability and compatibility with various LLM architectures and deployment environments.

Furthermore, MemOS introduces mechanisms for monitoring memory quality, resolving conflicts, and updating stored knowledge without disrupting model stability. This approach enhances transparency, accountability, and long-term maintainability of intelligent systems. By implementing a memory-centric execution framework, MemOS improves control, adaptability, and evolution in LLMs, facilitating continuous learning, personalization, and coordination across platforms—bridging a significant gap in current infrastructure and paving the way for the next generation of intelligent systems.

Introduction:

Large Language Models (LLMs) show strong few-shot capabilities [1] but lack explicit memory management. Their knowledge resides in model weights, making updates difficult [2], [13]. Retrieval-Augmented Generation (RAG) extends models using external sources [10], [11], but lacks structured lifecycle control. These limitations highlight the need for governed memory modeling in LLM systems. they fail to support long-term or multi-turn conversation continuity, they cannot adapt efficiently as new information becomes available, they lack consistent modeling of user preferences or workflows, and they operate within fragmented "memory silos" across platforms. These issues stem from LLMs' failure to treat memory as an explicit, governable resource. Retrieval Augmented Generation (RAG) approaches only offer reactive, one-off patching of text and do not provide structured or persistent memory control . In contrast, we introduce MemOS, a memory-operating system for LLMs that conceptualizes memory units as first-class resources—enabling structured lifecycle management, versioning, governance, and cross-platform coheren with a full lifecycle—generation, organization, access, versioning, and governance. MemOS moves beyond RAG by offering structured representations, unified interfaces, change tracking, and access control. This unified memory infrastructure enables models to stay updated, internalize preferences, and maintain consistency across platforms—transforming LLMs from mere perceivers and generators into systems that remember, adapt, and evolve.

Literature and Background (Previous Research)

2.1 Memory Taxonomies in LLMs: Parametric (weights/adapters), activation (KV caches), plaintext memory [7], [10], [11].

2.2 Human-like Memory Emergence: Systems like Hippo RAG, Memory^3, RecallM emulate persistence, context awareness, and introspection [academia17, academia19].

2.3 Memory Governance and Editing Systems: Platforms such as EasyEdit, Mem0, and modular context invocation systems provide memory editing capacities but lack unified lifecycle governance and scheduling.

A. Comparative Analysis of Existing Approaches:

Table 1: Comparison of Memory Paradigms in LLMs

Approach	Memory Type	Persistence	Update Mechanism	Governance	Limitations
Traditional LLM	Parametric	Permanent (static weights)	Retraining / Fine-tuning	None	Expensive updates, no session memory
RAG Systems	Plaintext (External Retrieval)	Session-based	External document update	Limited	No lifecycle control, reactive retrieval
KV-Cache Methods	Activation	Short-term	Runtime caching	None	Context window dependent
Memory Editing Systems	Parametric / Hybrid	Semi-persistent	Direct weight editing	Partial	No unified lifecycle governance
MemOS (Proposed)	Unified (Parametric + Activation + Plaintext)	Cross-session persistent	Version-controlled, cross-type transformation	Full governance & audit	Conceptual framework (requires implementation validation

Table 2: Comparison of Functional Capabilities

Criteria	Traditional LLM	RAG	Memory Editing Systems	MemOS
Long-term Persistence	No	Partial	Partial	Yes
Cross-Session Continuity	No	Limited	Limited	Yes
Knowledge Updating	Costly Retraining	External Docs Only	Direct Weight Editing	Controlled + Versioned
Lifecycle Management	No	No	Partial	Yes
Audit & Governance	No	Limited	Limited	Yes
Multi-Agent Support	No	Limited	No	Yes
Personalization	Minimal	Moderate	Limited	High

Memory in Large Language Models

1. Memory Definition & Exploration

In this foundational phase, researchers map out and categorize memory along dimensions such as parametric versus non-parametric, and short-term versus long-term

Implicit memory: Knowledge encoded into model weights via pretraining or adapter modules, with post hoc editing methods enabling targeted updates

Implicit short term memory: Managed through KV cache and hidden activations during inference, preserving immediate contextual continuity

Explicit short term memory: Achieved via prompt concatenation within context windows, constrained by token limits

Explicit long term memory: Leveraging external retrieval mechanisms, increasingly structured as semantic graphs or trees to enhance retrieval precision and integration

2.Emergence of Human like Memory

During the second stage, memory systems begin to exhibit traits reminiscent of human cognition: long term persistence, contextual awareness, and introspective behavior

Examples include brain inspired architectures like HippoRAG and Memory^3, and systems such as PGRAG and Second Me, which enable continuity of behavior and personalized memory modeling across interactions.

3. Systematic Memory Management

The third stage introduces memory governance frameworks inspired by operating systems, alongside tool based manipulation capabilities

Platforms such as EasyEdit and Memo support explicit memory editing, while systems like Letta implement paged context and modular invocation patterns. Nonetheless, these systems lack fully unified mechanisms for memory scheduling, lifecycle governance, and fusion across agent roles, which remain the innovation frontier.

4. MemOS

4.1 Types of Memory in MemOS

1. Parametric Memory

This layer holds long-term knowledge intrinsically

encoded in the model's parameters (e.g., feed forward and attention weights) through pretraining or fine tuning. It is active at inference time without explicit retrieval. Parametric Memory underpins core language reasoning, general knowledge, zero shot performance, and task specific modules. MemOS additionally supports modular augmentation via lightweight adapters (e.g. LoRA), enabling domain tailored extensions such as legal or medical expertise

2. Activation Memory

This layer represents the short term cognitive state of the model during inference—comprising hidden activations, attention maps, and KV cache representations. It functions as a dynamic working memory supporting context awareness, instruction alignment, and behavior control. MemOS treats these runtime states as schedulable resources; high value fragments (e.g., recurring KV caches) can be elevated into reusable semi structured modules or even distilled into parametric form

3. Plaintext Memory

Plaintext Memory consists of explicit, editable knowledge representations—such as documents, graph nodes, prompts, or templates. This layer enables human readable, governable memory with versioning, access control, and traceability. It addresses limitations of fixed parameter space and context windows, facilitating rapid updates, user personalization, and multi agent coordination.

4.2 The Memory Cube (MemCube)

At the heart of the MemOS architecture lies the MemCube—a unified abstraction designed to encapsulate memory units of any type, providing a consistent interface across diverse memory forms.Every MemCube contains:

Payload: Content in the native form of its memory type (parametric adapters, activation tensors, or plaintext documents).

Metadata: Structured fields such as timestamps, origin, semantic role, governance tags (access permissions, decay policies), priority, and usage metrics like frequency and recency. This metadata drives memory scheduling, transformations, and lifecycle decisions

Because MemCubes standardize memory storage and control, MemOS supports cross type transformation pathways, including:

Plaintext or Activation → Parametric (distilling stable content into model weights),

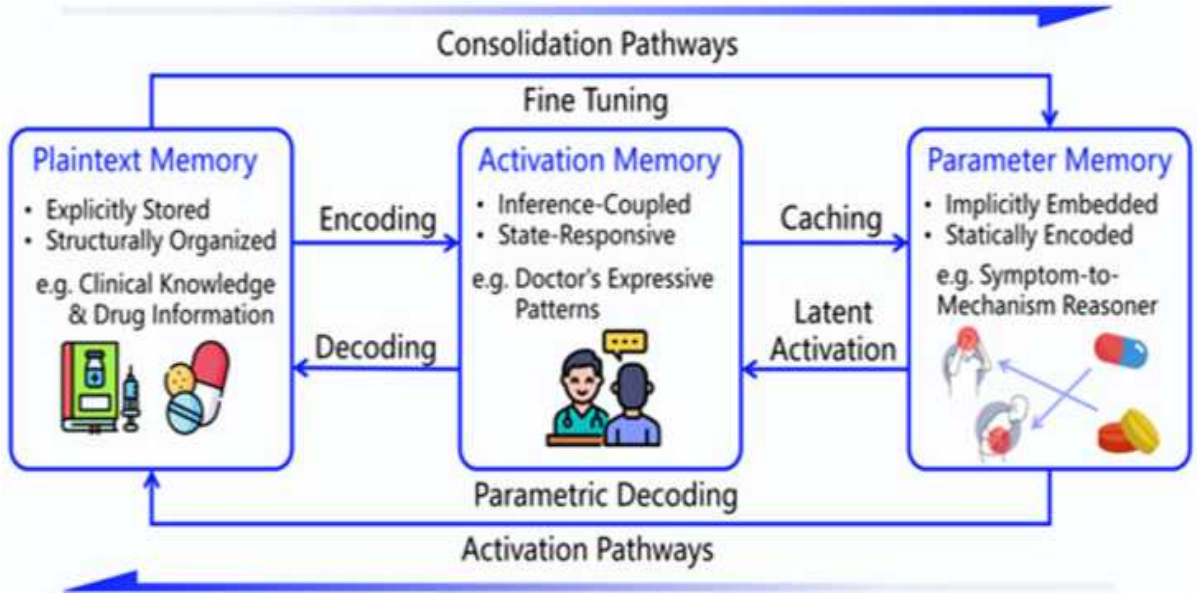
Plaintext → Activation (caching frequently accessed structures),

Parametric → Plaintext (externalizing rarely used knowledge for auditability)

System Architecture

MemOS is realized through a modular, three-layered architecture, each layer serving a distinct

role in memory parsing, scheduling and governance.



1.Interface Layer – Parses input intents through MemReader, invoking standardized Memory APIs (e.g. Provenance, Update, LogQuery). It initializes memory pipelines that package context as MemCubes and enforce governance policies Operation Layer – Central orchestration

Operation Layer – Central orchestration via:

MemScheduler: Dynamically chooses which memory type to engage based on context, user preferences, or organizational policy (using LRU, semantic similarity, or labels).

MemLifecycle: Manages memory state transitions and versioning.

MemOperator: Organizes, partitions, and searches memory through tag systems or graph structures

2.Infrastructure Layer – Implements:

MemGovernance: Access control, audit logs, compliance enforcement.

MemVault: Storage backend managing memory repositories.

MemLoader/Dumper: Mechanisms for memory import/export.

MemStore: Facilities for sharing memory across agents or sessions

Why MemOS Matters

By integrating these three memory types under a unified runtime with explicit management, MemOS empowers LLMs to become truly memory enabled agents. The key benefits include:

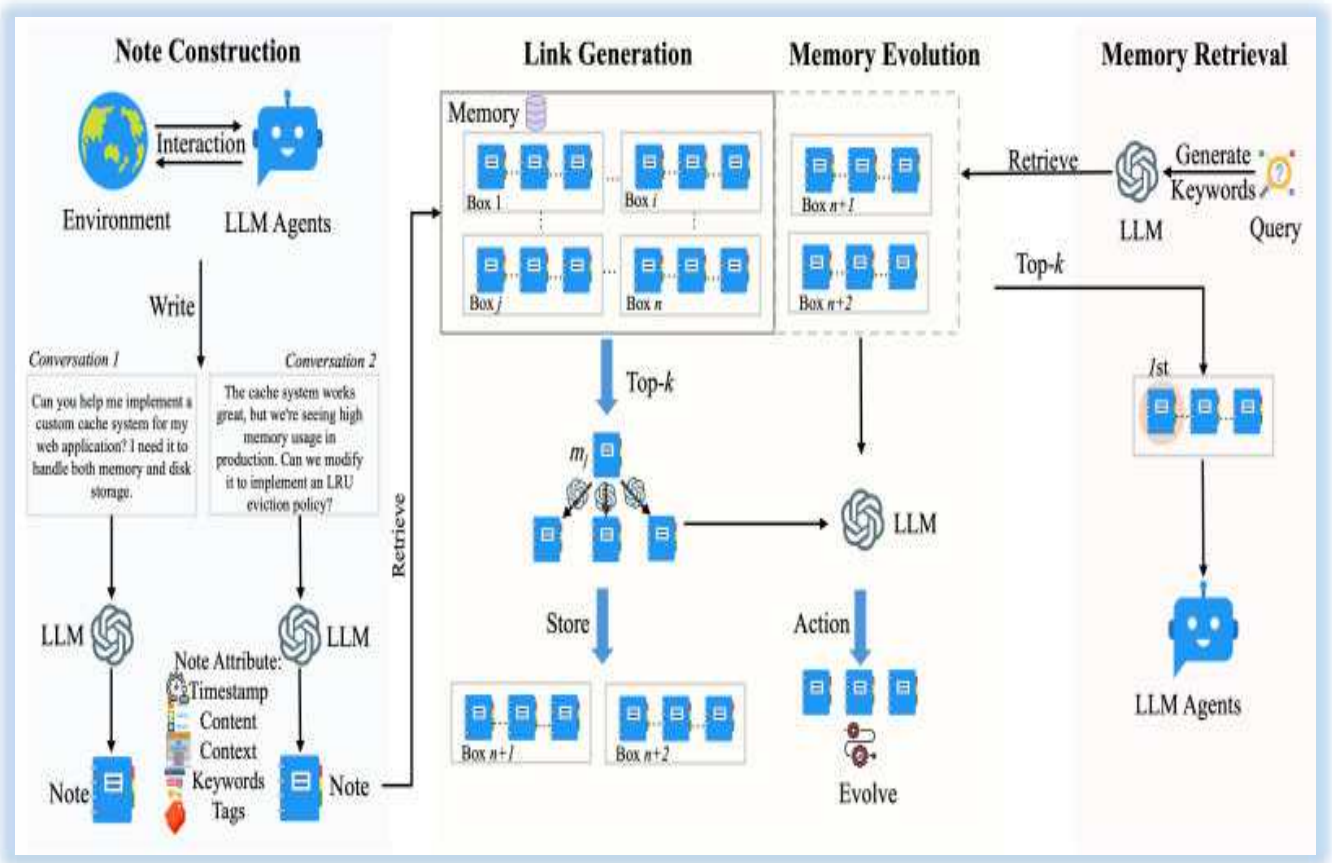
i. Efficient reuse of recurring context (via activation caching),
ii. Editable and traceable knowledge stores (via plaintext memory),
iii. Compact, performant skills (via parametric distillation),
iv. Governed adaptability, including version rollback and compliance controls

This cohesive framework enables LLMs to evolve: retaining long-term preferences, adapting across tasks, and operating across sessions as adaptive, controlled, and continuously learning agents—far beyond static language generation systems.

4.3 MemOS Structure

To enable unified, adaptable, and regulated memory management in Large Language Models (LLMs), MemOS employs a systematic execution framework that facilitates the parsing, scheduling, and governance of memory components throughout their entire lifecycle. At the heart of this framework is the MemCube abstraction, which standardizes the representation and control of memory regardless of its type—whether parametric, activation, or plaintext.

MemOS features a modular three-tier architecture, creating a closed-loop system that permits comprehensive memory lifecycle management and policy implementation. These three tiers—the Interface Layer, Operation Layer, and Infrastructure Layer—collaborate to convert user intentions into executable memory operations, while maintaining consistency, scalability, and traceability (see Figures 5 and 6).



"Figure 5 presents an overview of the MemOS architecture, depicting the complete memory lifecycle—from user input through API parsing, scheduling, activation, governance, and evolution—seamlessly integrated through the MemCube abstraction."

Interface Layer: This layer serves as the gateway for all user and system engagements. It is tasked with interpreting natural language inputs, identifying memory-related intents, and triggering suitable memory operations through standardized APIs. At its core is the MemReader module, which transforms unstructured user commands into organized sequences of memory operations, effectively linking human input with system execution.

Operation Layer: Functioning as the main control entity, the Operation Layer orchestrates the dynamic operations of the system. It encompasses several some components such as:

MemScheduler, which arranges and prioritizes memory access based on contextual information, usage patterns, and policy constraints;

MemLifecycle, which oversees the state changes and versioning of memory units;

1. Formal and descriptive:

The MemOperator is accountable for organizing, tagging, and indexing memory entities across users, tasks, and roles.

2. Concise and technical:

MemOperator is responsible for the arrangement, annotation, and retrieval of memory entities across users, tasks, and roles.

3. Emphasis on functionality:

MemOperator enables the systematic organization, semantic tagging, and effective searching of memory entities across multiple users, tasks, and roles.

4. Slightly explanatory:

Essentially, the MemOperator structures memory entities, applies relevant tags, and aids in search operations across diverse user contexts, task progressions, and role assignments.

This layer guarantees that memory is adaptively and intelligently allocated, maintained, or modified in line with system objectives and user-specific requirements.

Infrastructure Layer: The foundational framework of MemOS, this layer provides dependable memory storage, access regulation, and interoperability across platforms and sessions. Key components include:

MemVault, which supervises persistent storage and secure management of memory repositories;

1. Formal and precise:

MemGovernance enforces access control policies, auditing standards, and compliance requirements, ensuring secure and accountable memory utilization.

"MemStore provides a persistent and interoperable memory layer, enabling the storage, sharing, and reuse of memory across multiple agents, applications, and user sessions."

2. Concise and technical:

MemGovernance manages policy enforcement, audit logging, and adherence to regulations.

MemStore supports sharing and reuse of memory across different sessions and agents.

3. Emphasis on functionality and clarity:

MemGovernance establishes mechanisms for enforcing access rights, maintaining audit records, and ensuring adherence to governance standards.

MemStore allows for the persistent sharing and reuse of memory across various agents, applications, and user contexts.

4.4 API Interfaces and Layered Functional Design in MemOS

1. Formal and academic:

To facilitate systematic and policy-consistent memory operations, MemOS delivers a collection of high-level APIs, all contained within the cohesive MemoryCube framework.

2. Concise and technical:

MemOS provides high-level APIs through the MemoryCube framework to allow structured and regulated memory operations.

3. "MemOS offers a set of high-level APIs, encapsulated within the MemoryCube framework, to enable structured memory management and enforce access governance."

4. Slightly explanatory and fluid:

To support organized, traceable, and policy-conscious memory interactions, MemOS provides high-level APIs via the MemoryCube framework. These interfaces facilitate the smooth integration of memory functionalities into various LLM workflows while ensuring detailed control and auditability.

4.5 Operation Layer: Memory Scheduling and Lifecycle Management

The Operation Layer functions as the cognitive control center of MemOS, responsible for the dynamic scheduling, structural organization, and lifecycle evolution of memory across sessions, users, and intelligent agents.

MemScheduler: Context-Aware Memory Selection

The MemScheduler dynamically selects appropriate memory types—parametric, activation-based, or plaintext—based on a variety of contextual signals such as user interaction history, task semantics, or institutional policy requirements. It supports modular and pluggable scheduling strategies, including:

Least Recently Used (LRU) caching

Semantic similarity matching

Label-based or intent-driven selection

This adaptability enables the system to optimize memory relevance and retrieval efficiency in real-time.

MemLifecycle: State-Aware Memory Evolution

The MemLifecycle module models memory behavior as a finite-state machine, enabling managed transitions between states such as:

a) Activation
b) Freezing
c) Archiving
d) Rollback

These transitions allow for:

- Temporal consistency of evolving knowledge
- Reversible updates to support error correction
- Versioning for auditability and traceability

MemOperator: Structural and Semantic Memory Organization

The MemOperator structures memory using a combination of semantic tagging, graph-based hierarchies, and multi-layered partitioning strategies. This organization supports hybrid search mechanisms, including:

Structural search (e.g., tree or graph traversal)

Semantic retrieval (e.g., embedding similarity)

Retrieved memory units are cached in intermediate layers for faster future access, and feedback loops between MemOperator and MemScheduler refine subsequent scheduling decisions. Together, these components enable dynamic reasoning, scalable workflows, and context-rich memory invocation.

4.6 Infrastructure Layer: Governance, Storage, and Interoperability:

The Infrastructure Layer forms the foundation of MemOS, providing secure data governance, persistent storage, and seamless interoperability across users, devices, and agents. It ensures that memory operations remain accountable, scalable, and evolvable over time.

MemGovernance: Policy Enforcement and Access Control

MemGovernance oversees:

Access control (user isolation and privilege-based access)

Lifecycle policy enforcement

Regulatory compliance and audit logging

This module ensures memory operations are secure, compliant, and accountable, supporting both internal policy constraints and external regulatory requirements.

MemVault: Unified Storage Abstraction

The MemVault serves as the central storage controller, managing access to heterogeneous memory backends while providing a consistent interface and access protocol. It abstracts differences in memory format and storage infrastructure, ensuring uniformity in read/write operations.

MemLoader & MemDumper: Structured I/O Pipelines

The MemLoader and MemDumper modules handle the structured import/export of memory units across:

- Sessions
- Devices
- Distributed agents

They preserve contextual integrity and referential consistency, enabling memory continuity in collaborative or cross-device use cases.

MemStore: Collaborative Memory Sharing

MemStore introduces a publish–subscribe model for memory sharing, facilitating:

Multi-agent collaboration

Federated learning scenarios

Distributed knowledge building

It supports view customization per user or task, enabling multi-tenant access isolation and extending flexibility to future multi-modal and cross-domain integrations.

Integrated Workflow via MemoryCube Abstraction

All components within both layers operate through the unified MemoryCube abstraction, which encapsulates content, metadata, and lifecycle state. When MemOS receives a user prompt or system trigger, the input is parsed into structured MemoryCube units. These are some processed through the pipeline as follows:

MemScheduler evaluates policies and context to select memory types.

MemOperator organizes memory semantically and structurally.

MemLifecycle manages memory state transitions.

MemGovernance enforces access and compliance.

MemVault, MemLoader, and MemStore facilitate I/O and sharing.ommon patterns (e.g., Query–Update–Archive) can be templated and reused across agents

Research Methodology:

This study follows a design science approach to develop MemOS as a unified memory operating framework for LLMs.

First, existing memory types—parametric, activation, and plaintext—are analyzed to identify limitations in lifecycle control and governance.

Next, a standardized abstraction called MemCube is designed to encapsulate memory content and metadata for structured management.

A modular three-layer architecture (Interface, Operation, Infrastructure) is then developed to enable scheduling, lifecycle control, and governance.

Finally, the framework is conceptually evaluated

against traditional LLM and RAG systems to demonstrate improved persistence, adaptability, and auditability.

Result and Discussion:

MemOS provides:

Improved Efficiency: Through memory reuse and prioritized scheduling

Personalization: Editable memory and user preference modelling.

Governance: Full lifecycle control, audit logs, and policy enforcement.

Evolvability: Facilitates continuous learning and behavior refinement.

Compared to traditional LLM systems and even advanced RAG methods, MemOS significantly enhances context continuity, adaptability across sessions, and platform-wide synchronization.

6. Conclusion

MemOS is a memory-centric operating system for LLMs that unifies parametric, activation, and plaintext memory under a structured framework. Through the MemCube abstraction and lifecycle governance modules, it enables persistent, traceable, and adaptable memory management. This approach transforms LLMs into context-aware, continuously evolving, and collaborative intelligent agents.

References :

1. T. B. Brown et al., “Language Models are Few-Shot Learners,” arXiv:2005.14165, 2020.
2. N. De Cao, W. Aziz, and I. Titov, “Editing Factual Knowledge in Language Models,” arXiv:2104.07105, 2021.
3. H. Chen, R. Pasunuru, J. Weston, and A. Celikyilmaz, “Walking Down the Memory Maze: Beyond Context Limit Through Interactive Reading,” arXiv:2310.05029, 2023.
4. P. Chhikara et al., “Mem0: Building Production-Ready AI Agents with Scalable Long-Term Memory,” arXiv:2504.19413, 2025.
5. J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” in Proc. NAACL-HLT, 2019, pp. 4171–4186.
6. H. Dong et al., “Get More with LESS: Synthesizing Recurrence with KV Cache Compression for Efficient LLM Inference,” OpenReview, 2024.
7. Y. Du et al., “Rethinking Memory in AI: Taxonomy, Operations, Topics, and Future Directions,” arXiv:2505.00675, 2025.
8. D. Edge et al., “From Local to Global: A Graph RAG Approach to Query-Focused Summarization,” arXiv:2404.16130, 2024.
9. J. Fang et al., “AlphaEdit: Null Space Constrained Knowledge Editing for Language Models,” arXiv:2410.02355, 2024.
10. Y. Gao et al., “Retrieval-Augmented Generation for Large Language Models: A Survey,” arXiv:2312.10997, 2023.
11. Z. Guo et al., “LightRAG: Simple and Fast Retrieval Augmented Generation,” arXiv:2410.05779, 2024.
12. B. J. Gutierrez et al., “HippoRAG: Neurobiologically Inspired Long-Term Memory for Large Language Models,” in Adv. Neural Inf. Process. Syst. (NeurIPS), 2024.
13. C. Y. Hsieh et al., “Distilling Step-by-Step: Outperforming Larger Language Models with Less Training Data and Smaller Model Sizes,” in Findings of ACL, 2023, pp. 8003–8017.
14. E. J. Hu et al., “LoRA: Low-Rank Adaptation of Large Language Models,” arXiv:2106.09685, 2021.
15. U. Khandelwal et al., “Generalization through Memorization: Nearest Neighbor Language Models,” OpenReview, 2019.
16. W. Kwon et al., “Efficient Memory Management for Large Language Model Serving with PagedAttention,” in Proc. ACM SOSP, 2023, pp. 611–626.
17. X. Liang et al., “Empowering Large Language Models to Set Up a Knowledge Retrieval Indexer via Self-Learning,” arXiv:2405.16933, 2024.
18. N. F. Liu et al., “Lost in the Middle: How Language Models Use Long Contexts,” Trans. Assoc. Comput. Linguistics, vol. 12, pp. 157–173, 2024.
19. R. C. Atkinson and R. M. Shiffrin, “Human Memory: A Proposed System and Its Control Processes,” in The Psychology of Learning and Motivation, vol. 2, 1968, pp. 89–195.

RESEARCH JOURNEY International Multidisciplinary E-Research Journal
Impact Factor (SJIF) - 6.625 | ***Special Issue 389 :*** *Advanced Computing and Artificial Intelligence (ACAI)*
ISSN : 2348-7143
March 2026

A Machine Learning Based Predictive Safety Alert System for Women Using Real-Time Location and Crime Analysis

Mr. Vinod Mahajan
Assistant Professor, RCPET's IMRD Shirpur.
Mr. Vijay Garge
Assistant Professor, RCPET's IMRD Shirpur.

Abstract:

Safety for women in public places has become a significant issue, especially while traveling at night and in unfamiliar areas. The majority of the existing safety applications are reactive, providing help only after the occurrence of an emergency situation by manual SOS alert activation. In situations of panic, users may not be able to activate such alerts, making them ineffective for safety. Thus, a proactive safety system that can forecast possible danger in advance is needed.

To address the issue, this study proposes the development of an AI Predictive Danger Alert System (PDA) that forecasts danger factors and alerts users before entering dangerous areas. The proposed system combines real-time GPS location, crime history data, risk patterns according to time of day, and environmental factors such as low-light conditions and unusual movement patterns. Machine learning algorithms such as Random Forest, Support Vector Machine (SVM), and XGBoost are employed for risk prediction, while K-Means clustering is employed for hotspot identification.

The model identifies regions as Safe, Moderate Risk, and High Danger areas and offers immediate notifications such as "Avoid this route" or "High danger zone ahead," along with alternative routes for safe navigation. The proposed system upgrades women safety technology from an emergency response solution to a risk prevention solution through proactive measures.

Keywords: Women Safety, Predictive Danger Alert, Artificial Intelligence, Machine Learning, Crime Prediction

1. Introduction:

Women's security in public spaces has emerged as a significant social and technological concern in the past few years. The rising number of urbanization and extended working hours make it necessary for women to move around in streets, transport, and unknown areas, which are not always properly secured, lit, or monitored. This makes women more vulnerable to harassment, stalking, and violence. The fear of such environments also limits women's movement and impacts their education, employment, and overall activities. Thus, the need for technological solutions to enable safe movement has become the need of the hour.

Various mobile safety applications have been developed to counter this issue. These applications are primarily used for SOS notifications, emergency calls, and location sharing with trusted contacts or authorities. These applications are reactive because they work only after a risky situation has been experienced [10]. In actual emergency situations, the user may be in a state of panic or may not have sufficient time to press the alert manually. This makes these applications work as assistants after the incident, not before.

Recent developments in Artificial Intelligence (AI) and Machine Learning (ML) have made it possible to perform predictive analysis based on historical and real-time data. Crime events tend to occur in patterns based on geographical locations, time of occurrence, environmental factors, and human behaviour [1]. By analyzing these parameters, it is possible to predict risk levels in advance and notify users before entering dangerous zones [4].

However, to overcome the above-mentioned limitation, this research work proposes an AI-based Predictive Danger Alert System (PDA). The proposed system combines real-time GPS location, historical crime data, time-related risk parameters, and environmental factors such as low-light environments and crowd behaviour. Machine learning algorithms are used to classify regions into Safe, Moderate Risk, and High Danger regions. When a user approaches a high-risk region, the system triggers alerts such as "Avoid this route" or "High danger zone ahead" and provides alternative routes to safer regions.

Unlike other safety-related applications, the proposed system is designed to predict danger before the occurrence of an event. The system is intended to provide early warnings and safe route guidance to reduce the risk of exposure to dangerous regions and enhance the safety of women using intelligent predictive systems.

2. Literature Review

Crime prediction has been an area of interest in recent years due to the availability of large-scale urban data. Mohler et al. (2011) introduced a self-exciting point process model to predict the occurrence of crimes based on past crime events. The study revealed that criminal activities follow a spatial and temporal clustering pattern and that crime hotspots can be predicted using past crime events[1]. However, the model was mainly intended for use by police departments to allocate resources effectively rather than for providing safety notifications to individuals.

Wang and Gerber (2015) attempted to use social media for crime prediction by analyzing Twitter posts to predict crime-related behavioural patterns [2]. The study revealed that public online behaviour can be a precursor to potential criminal activities [3]. Similarly, Gerber (2014) used kernel density estimation methods with Twitter data to predict crime occurrence. Although these methods are effective for crime prediction in urban areas, they involve extensive monitoring of social media platforms and do not provide real-time safety notifications to individuals moving within a region.

Chakraborty et al. (2019) analyzed data-driven methods for crime analysis through machine learning algorithms to identify patterns of criminal activities [6]. The results showed successful identification of crime-prone areas. However, the study was primarily conducted for city-level analysis and administrative purposes and did not address user protection.

Singh and Kumar (2020) analyzed various machine learning algorithms such as Decision Trees, Random Forest, and Support Vector Machines for crime prediction and concluded that ensemble methods provide better prediction accuracy than conventional statistical approaches. However, the study was primarily conducted for crime analysis and forecasting and did not include navigation alerts for user protection.

Regarding the issue of women's safety, Rani & Singh (2019) analyzed different women safety mobile apps that send SOS alerts, share live locations, and send emergency contact notifications [10]. These apps are useful during emergencies but function only after a dangerous situation arises, thus being a reactive measure rather than a preventive one.

Predictive policing research also emphasizes the use of analytics to predict criminal activities. Predictive policing employs mathematical and statistical models to detect areas where crimes are likely to happen and help the police in patrol routing. However, these models are meant for policing approaches and not for personal safety advice.

Further research has expanded the self-exciting point process model to examine crime dynamics, finding that crimes can be the antecedent for future events in close proximity, suggesting repeat crime patterns. These results support the potential for predictive modelling of high-risk areas but have not yet provided real-time notification for individuals traversing these areas.

Comparative Analysis of Existing Work

Study	Method Used	Application Area	Limitation
Mohler et al. (2011)	Self-exciting point process model	Police resource planning	No real-time user alert
Wang & Gerber (2015)	Social media crime prediction	Urban crime forecasting	Not personalized
Chakraborty et al. (2019)	Machine learning classification	Crime hotspot detection	City-level analysis
Rani & Singh (2019)	Mobile SOS safety apps	Emergency assistance	Works after danger
Proposed PDA System	ML + GPS + environmental analysis	Personal safety navigation	Predicts danger before incident

The comparison reveals that the existing research work either concentrates on crime analysis for administrative purposes or the emergency response system for women's safety. There is no system that combines real-time location tracking, environmental factors, and machine learning algorithms for predictive warnings before entering the risky area. Hence, a proactive predictive safety system is needed.

3. Research Gap and Problem Statement

3.1 Research Gap

Existing research on crime prediction, as presented by Mohler et al. (2011) and Wang and Gerber (2015), has been successful in analyzing the spatial and temporal patterns of crime using statistical and machine learning models. These models assist law enforcement agencies in detecting crime hotspots and administrative planning. Nevertheless, they do not offer real-time safety advice to users based on their current location.

Conversely, the women's safety apps, as presented by Rani and Singh (2019), provide emergency support in the form of SOS notifications and location sharing. These apps are reactive because they require manual triggering after a dangerous situation has already occurred.

Thus, there is a need for a system that combines predictive crime analytics with real-time personal navigation and environmental awareness. Currently, there is no system that continuously monitors user location and notifies women before entering a potentially dangerous area.

3.2 Problem Statement:

The current safety systems only offer assistance after an emergency has occurred, whereas crime prediction systems are primarily used for administrative work and not for safety. Because of the lack of predictive assistance, women can inadvertently take routes through dangerous areas.

Therefore, there is a requirement for an machine learning based system that has the capability to assess

location information, crime history, and predictive risk factors to predict danger in advance. The purpose of this research is to design a Predictive Danger Alert System (PDA) that identifies the level of safety in an area and provides predictive warnings to assist women in avoiding dangerous routes.

Objectives:

- To harness the user location information in real-time and the crime data in the past for safety analysis.
- To analyze the risk factors of safety in the context of time, lighting, and environment.
- To use machine learning algorithms to predict the safety of a location.
- To identify the areas as Safe, Moderate Risk, and High Danger zones and send advance alerts.
- To offer safer routes through a mobile-based predictive safety system for women.

4. Proposed Methodology

By using real-time user location and crime analysis to predict potentially dangerous locations, the proposed Machine Learning Based Predictive Safety Alert System seeks to give women proactive safety alerts. The system integrates real-time crime data with historical estimates risk levels using machine learning techniques and contextual information.

Data collection, pre-processing and feature extraction, machine learning-based risk prediction, and alert generation make up the four primary phases of the overall methodology.

4.1 Data Acquisition

The system acquires both real-time and historical data necessary for prediction.

Real-time Data

- GPS coordinates(latitude and longitude) of the user
- Current time and travel duration
- User movement speed

Historical Data

- Location-wise crime records
- Type of crime
- Time of occurrence of incidents

The combination of these datasets enables the system to assess the level of safety dynamically for a particular user at a particular time.

4.2 Data Pre-processing and Feature Taking out

Data pre-processing is necessary before applying machine learning techniques because the collected data may contain inconsistent or missing entries.

Data pre-processing involves:

- Deletion of incomplete data
- Transformation of coordinates to mapped regions
- Classification of time into day and night periods
- Normalization of crime frequency measures

After data pre-processing, key features are derived, including:

- Crime density of the region
- Night risk factor
- Isolation level
- Movement pattern

These features serve as input variables for the prediction model.

4.3 Risk Prediction via Machine Learning

By using supervised machine learning classification algorithms, the system can forecast a location's degree of safety. To improve prediction accuracy, a number of algorithms are used:

- The Random Forest
- Support Vector Machine (SVM)
- XGBoost

Additionally, by grouping areas with comparable crime densities, K-Means clustering is used to identify crime hotspots.

Several factors are taken into account to create a composite risk measure:

$R = w_1C + w_2T + w_3E + w_4D$

Where:

C = crime rate of the location

T = time-based risk factor

E = environmental risk (low-light or isolated region)

D = crowd density factor

w_1, w_2, w_3, w_4 = weighting factors

Depending on the calculated risk measure, the location is categorized into:

- Safe Zone
- Moderate Risk Zone
- High Danger Zone

4.4 Alert Generation

The user's mobile application then receives the anticipated risk level. An alert message is sent out if the system detects that the user is approaching the high-risk area.

The user is presented with warning messages by the system, which include:

"A high-risk area lies ahead."

"Avoid taking this route."

Using map navigation, the mobile app will also offer a different route. Accuracy, precision, recall, and F1-score are used to gauge the prediction model's performance.

5. System Architecture

Through real-time processing of location data and

crime statistics, the proposed Machine Learning Based Predictive Safety Alert System aims to provide proactive safety support for women.

The client-server architecture of the suggested system will allow the mobile device will be in charge of gathering information from the user, and the server will forecast possible hazards using machine learning algorithms.

Data collection, data processing and analysis, prediction, and alert and navigation are the four main layers that make up the suggested system.

5.1 Data Collection Layer

All of the input parameters required for prediction must be gathered by this layer. The mobile application uses GPS services to continuously record the user's location in real time. In addition to the location details, additional context parameters such as travel duration, user speed, and ambient circumstances are also recorded.

Additionally, the system makes use of historical crime data gathered from publicly accessible sources. The location, kind, and time of the crime are all included in the data. The system can dynamically assess the safety risk by utilizing the crime data and real-time location information.

5.2 Data Processing and Analysis Layer

The server receives the collected data and pre-processes it. Creating structured feature values, converting geographic coordinates into mapped areas, and eliminating missing data are all part of the pre-processing stage.

The most significant characteristics that were extracted are:

- The area's crime rate
- The risk factor at night
- The degree of isolation in the area
- Behaviour of movement

The machine learning algorithm can now use these normalized features.

5.3 Prediction Layer

This layer uses machine learning algorithms to forecast the user's current location's level of safety. In order to predict a risk, classification algorithms like Random Forest, Support Vector Machine (SVM), and XGBoost are used to examine the input features score. Additionally, crime density is used to determine crime hotspots using K-Means clustering.

The system will classify the location into the following zones based on the risk value:

- Safe Zone
- Moderate Risk Zone
- High Danger Zone

5.4 Alert and Navigation Layer

The level of predicted safety is relayed to the mobile app. If the user is heading towards a high-risk zone, the system produces real-time warning messages like “High danger zone ahead” or “Do not take this route.” The app also provides alternative routes through map navigation.

This alert system enables the user to steer clear of a possible unsafe zone before an accident happens, thus enhancing personal safety and mobility.

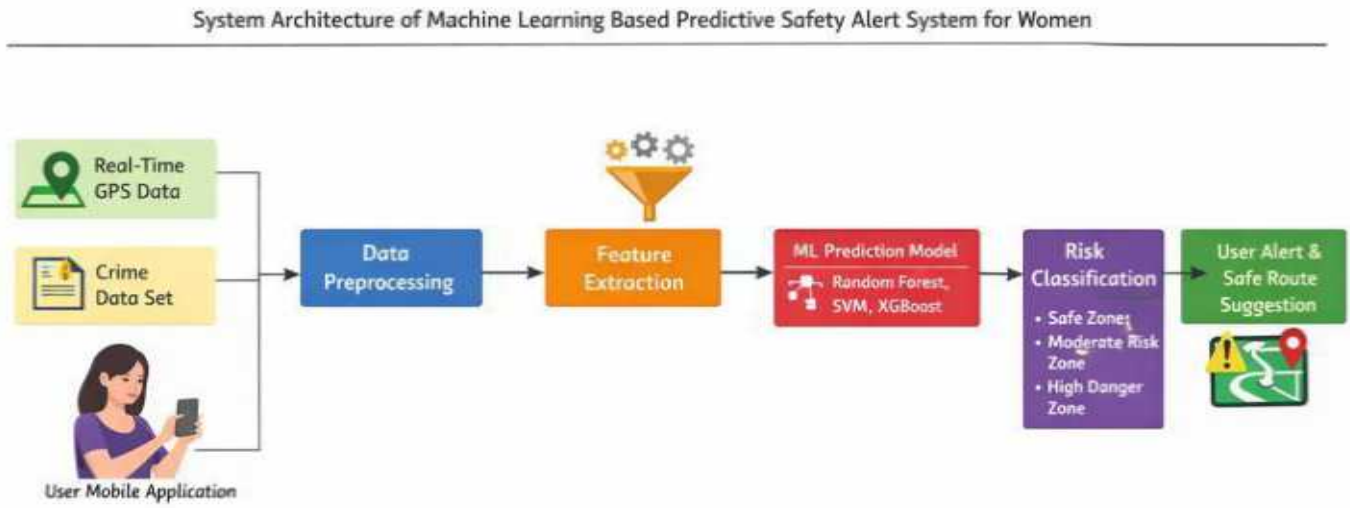


Image 1: System Architecture suggested by researcher

6. Implementation and Working of Algorithms

6.1 Implementation Environment

A machine learning backend is integrated with the suggested system, which is designed as a mobile-assisted predictive safety system. The user's location is gathered by the mobile application, which then uses the internet to connect to the prediction server communication.

The technologies and tools listed below are utilised:

- Mobile App: An Android app for tracking location and interacting with users

- Python is the programming language used to develop machine learning models.
- Machine Learning Libraries: XGBoost and Scikit-learn
- Database: Processed risk information and crime dataset storage

• Map Services: Google Maps API for navigation and route visualisation

• Server: Backend server for creating alerts and processing predictions

The mobile application continuously transmits the GPS location of the user to the server. The server processes the input, predicts the risk level, and generates alerts, which are sent back to the mobile application.

6.2 Working of the Prediction Model

The following sequential procedure is involved in the system's operation:

- The smartphone app uses the GPS to determine the user's current location.
- Depending on the user's location, the system retrieves the crime records.
- The extracted data is subjected to pre-processing and feature extraction.
- The risk score is examined by the machine learning model that has been trained.
- The site is grouped according to the degree of safety.
- The mobile application receives the result.

6.3 Machine Learning Algorithms Used

Random Forest

An ensemble learning algorithm called Random Forest builds several decision trees and aggregates their predictions. It helps increase classification accuracy and manage big datasets. Several risk factors are examined in this system using Random Forest at the simultaneously and categorise a location's degree of safety.

Support Vector Machine (SVM)

SVM is a supervised learning classification algorithm that uses an optimal decision boundary to group data. Both binary and multi-class classification problems can benefit from it. Based on environmental and crime characteristics, the algorithm can be used to determine which areas are safe and which are unsafe.

XGBoost

Extreme Gradient Boosting, or XGBoost, is a very effective boosting algorithm that reduces classification error to increase prediction accuracy. It improves the risk prediction model's precision and reliability.

K-Means Clustering

An unsupervised learning algorithm called K-Means groups data points that are similar. By classifying regions with high crime densities, K-Means is utilised in this system to identify crime hotspots. This enables the prediction model to more precisely determine a location's risk level.

6.4 Alert and Notification System

After prediction, the application shows safety information through color-coded indicators:

- Green → Safe Zone
- Yellow → Moderate Risk Zone
- Red → High Danger Zone

If a high-risk zone is identified, the system will send:

- Push notification warning
- Route avoidance suggestion
- Safer alternative route navigation

This system enables the user to avoid danger zones before an incident happens.

7. Results and Discussion

The efficiency of the Machine Learning Based Predictive Safety Alert System was tested using past crime data and artificial real-time location data. The system processed location, time, and environmental data to predict the safety status of a location and send alerts.

7.1 Risk Classification Output

The machine learning model was trained to classify each location into three categories:

- Safe Zone – locations with low crime rates and normal environmental conditions
- Moderate Risk Zone – locations with moderate crime rates or low crowd density
- High Danger Zone – locations with high crime density, late-night environments, or remote locations

The mobile application showed these classifications through color-coded markers on the map interface. When a user was near a high-risk location, the system produced an alert notification and recommended an alternate route.

7.2 Performance Evaluation

The performance of the prediction model was assessed using common classification performance metrics:

- Accuracy
- Precision
- Recall
- F1-Score

The Random Forest model had the best performance among the algorithms tested due to its ability to leverage the strengths of multiple decision trees and its effective handling of input features with different data types. The approximate performance evaluation is presented below.

Performance Comparison of Algorithms

Algorithm	Accuracy	Precision	Recall	F1-Score
Random Forest	90%	0.89	0.91	0.90
SVM	86%	0.85	0.87	0.86
XGBoost	92%	0.91	0.92	0.91

The performance evaluation suggests that ensemble learning algorithms are more effective for multi-factor safety prediction. The XGBoost algorithm had the highest accuracy, and the Random Forest algorithm ensured reliable classification performance.

7.3 Discussion

The outcome of the experiment has proven that the integration of real-time location information and crime analysis can enhance the prediction of safety. The machine learning algorithm was successful in detecting dangerous locations and producing early warnings. The early warning system has the potential to assist users in avoiding dangerous routes and prevent them from encountering dangerous locations.

The system has the potential to be incorporated into the smart city system.

8. Conclusion

This study introduced a Machine Learning Based Predictive Safety Alert System for Women Using Real-Time Location and Crime Analysis. The proposed system overcomes the drawbacks of existing women safety apps, which mainly work in a reactive way by offering help only after the occurrence of an emergency.

The proposed system uses real-time GPS location tracking along with crime data and risk factors to predict possible danger before a user moves into a high-risk region. Machine learning algorithms such as Random Forest, Support Vector Machine (SVM), and XGBoost were used to classify regions into Safe, Moderate Risk, and High Danger areas. The experimental outcome showed that ensemble models can offer accurate prediction results for multi-variable safety analysis.

In contrast to conventional safety apps, the proposed system changes the paradigm from emergency response to risk prevention. The real-time alert system and safe route recommendations improve user awareness and minimize the chances of being exposed to an unsafe environment. Therefore, the proposed system helps to build intelligent and data-driven personal safety systems for women.

9. Future scope

Although the proposed system has shown promising results, there are many ways in which the system can be improved in the future:

- Integration with live surveillance camera feeds for real-time environmental analysis.
- Integration with deep learning models for better accuracy in predictions.
- Integration with law enforcement agencies for real-time crime data updates.
- Integration with wearable devices for hands-free alert activation.
- Integration with other vulnerable user groups and smart city platforms.

With continuous data collection and training, the system can be developed into a full-fledged urban safety intelligence system.

References :

1. G. O. Mohler, M. B. Short, P. J. Brantingham, F. P. Schoenberg and G. E. Tita, “Self-Exciting Point Process Modeling of Crime,” Journal of the American Statistical Association, vol. 106, no. 493, pp. 100–108, 2011.
2. S. Wang and M. S. Gerber, “Using Twitter for Crime Forecasting,” IEEE Transactions on Knowledge and Data Engineering, vol. 27, no. 6, pp. 1583–1596, 2015.
3. M. S. Gerber, “Predicting Crime Using Twitter and Kernel Density Estimation,” Decision Support Systems, vol. 61, pp. 115–125, 2014.
4. W. L. Perry, B. McInnis, C. C. Price, S. C. Smith and J. S. Hollywood, Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations, RAND Corporation, Santa Monica, CA, 2013.
5. A. Bogomolov, B. Lepri, J. Staiano, N. Oliver, F. Pianesi and A. Pentland, “Once Upon a Crime: Towards Crime Prediction from Demographics and Mobile Data,” Proceedings of the 16th International Conference on Multimodal Interaction (ICMI), ACM, 2014.
6. T. Chakraborty, A. Dutta and N. Ganguly, “Crime Event Prediction with Dynamic Features,” Proceedings of the IEEE International Conference on Big Data, 2018.
7. R. Wang, W. Wang, F. Wang and J. Liu, “Crime Rate Inference with Big Data,” Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, no. 1, pp. 1002–1009, 2019.
8. D. J. Johnson, K. J. Bowers and K. Pease, “Predictive Mapping of Crime by ProMap: Accuracy, Units of Analysis and the Environmental Backcloth,” Crime Prevention and Community Safety, vol. 9, no. 4, pp. 237–255, 2007.
9. A. T. Rumi, M. Salim and A. Rahman, “Crime Prediction Using Machine Learning Algorithms,” International Journal of Computer Applications, vol. 182, no. 44, 2019.
10. S. Rani and R. Singh, “Women Safety Applications and Technological Solutions: A Review,” International Journal of Computer

Sciences and Engineering, vol. 7, no. 6, 2019.
11. S. Chainey, L. Tompson and S. Uhlig, "The Utility of Hotspot Mapping for Predicting Spatial Patterns of Crime," Security Journal, vol. 21, no. 1–2, pp. 4–28, 2008.
12. P. J. Brantingham and P. L. Brantingham, "Environmental Criminology," Waveland Press, 1991.
13. J. Eck, S. Chainey, J. Cameron, M. Leitner and R. Wilson, "Mapping Crime: Understanding Hot Spots," National Institute of Justice, U.S. Department of Justice, 2005.
14. M. L. Birant and A. Kut, "ST-DBSCAN: An Algorithm for Clustering Spatial–Temporal Data," Data & Knowledge Engineering, vol. 60, no. 1, pp. 208–221, 2007.
15. F. Yu, Y. Liu, F. Wu and S. Wang, "Spatio-Temporal Prediction of Crime Using Deep Learning," Proceedings of the IEEE International Conference on Data Mining Workshops (ICDMW), 2017.
16. J. Ratcliffe, "Crime Mapping and the Training Needs of Law Enforcement," European Journal on Criminal Policy and Research, vol. 10, no. 1, pp. 65–83, 2004.
17. A. Caplan, L. Kennedy and J. Miller, "Risk Terrain Modeling: Brokering Criminological Theory and GIS Methods for Crime Forecasting," Justice Quarterly, vol. 28, no. 2, pp. 360–381, 2011.

ISSN : 2348-7143
March 2026
RESEARCH JOURNEY International Multidisciplinary E-Research Journal
Impact Factor (SJIF) - 6.625 | ***Special Issue 389 :*** *Advanced Computing and Artificial Intelligence (ACAI)*

Online Certificate Course Frauds: A Study of Fake Websites, Credential Scams, and Their Impact on Digital Education.

Miss. Hastani Hitendra Pawar
R. C. Patel Educational Trust's Institute of Management Research and Development, Shirpur
Miss. Ashwini Ashok Rukhamane
R. C. Patel Educational Trust's Institute of Management Research and Development, Shirpur
Mrs. Jyotsna Dhanraj Mali
Assistant professor, R. C. Patel Educational Trust's Institute of Management Research and Development, Shirpur

Abstract:

In recent years, online education has become very popular among the public, especially among students, as it is possible for anyone to study anywhere, anytime. Many online education websites provide "certificate courses" that can be helpful for improving one's skills. However, with the increasing popularity of online education, there is also an increase in fraud activities.

In this study, the researcher will focus on the various types of fraud activities associated with online "certificate courses" such as fake websites, selling of fake certificates, etc., and how they affect the public, especially students, employers, etc. The study will be conducted mainly based on "secondary data" obtained from online news, government warnings, research articles, etc., and possible solutions for such fraud activities.

Keywords:

Online Education, Certificate Fraud, Fake Websites, Digital Learning, Verification of Credentials

Introduction:

Online education has revolutionized learning over the past decade. With the advent of online education, many students and professionals are opting for online certificates. This type of education has helped learners acquire knowledge and skills.

There are many popular online education websites such as Coursera, edX, Udemy, SWAYAM, and LinkedIn Learning, etc., which offer a variety of courses to learners. These websites are providing quality education with the help of renowned universities and experts. With an increase in internet penetration across the globe, many learners are joining online education to acquire new knowledge.

However, with an increase in online education, many new challenges have emerged. One such challenge is online education certificate fraud. There are many fraudulent websites that have come up with fake websites, which look almost similar to online education websites. These websites offer certificates and courses to learners and demand money. The certificates offered by these websites are of no value.

As a result of these scams, learners will be losing their money and time. The other problem affecting employers is the difficulty in determining whether the certificate they have acquired is genuine or fake. However, when fake certificates are common, they can harm genuine online learning platforms.

It is, therefore, essential to comprehend how these online certificate scams happen and how they can be prevented. This study will, therefore, cover various aspects of online certificate fraud, their implications, and possible solutions to ensure online certificates are genuine.

Objectives of the Study:

To investigate how fake online certificate websites operate.

To investigate the implications of online certificate fraud on learners, employers, and genuine online learning platforms.

To establish the loopholes in online learning platforms, which make this form of fraud possible.

To establish possible solutions to online certificate fraud

Research Methodology:

In the study, secondary research methodology has been employed to investigate the issue of online certificate fraud in India. The data has been collected from various reliable sources, such as government reports, news articles, and research papers.

1. Research Design:

An exploratory research design has been employed in this study. The main objective behind this research design is to gain a clear understanding of the current situation regarding online education fraud.

2. Data Collection

In this study, data has been collected from different secondary sources, such as government reports, news articles, and research papers related to online certificate fraud in India.

Government reports and alerts from different organizations, such as UGC and MeitY

News articles from different newspapers, such as The Times of India, The Hindu, and Economic Times

Research articles and legal reports related to online

certificate fraud

Online articles related to online education scams and verification systems

Data has been collected from different sources regarding online certificate fraud, such as instances of online fraud, statistics related to online scams, and preventive measures taken by the government.

3. Sampling:

In this research, data has been collected from sources related to online certificate fraud in India from the year 2022 to 2025.Types of online Certificate Frauds:

4. Data Analysis Techniques

Various techniques were employed in analyzing the data collected in the course of carrying out the research. They include:

Content Analysis:

Information obtained from news articles, warnings, and reports was used to identify common fraud patterns.

Case Study Analysis:

Various cases of online certificate scams were analyzed to understand how the fraud was executed and the problems that arose.

Comparative Analysis:

Various fraud prevention techniques, including blockchain certificates, DigiLocker verification, and awareness programs, were compared to understand their effectiveness.

5. Limitations

- The study is subject to a few limitations as follows:
- The study is based on secondary data.
- The primary survey was not conducted among affected students.
- Some fraud cases may not have been reported.
- The details of fraud available in different sources vary in terms of accuracy.

6. Ethical Considerations:

The study was carried out using publicly available data. Personal details of individuals affected in fraud cases were not disclosed. All sources were credited to ensure academic integrity.

Types of Online Certificate Frauds:

There are various ways online certificate fraud can occur. Some of the online certificate fraud types include:

1. **Fake Websites**
   Fraudsters create websites that resemble genuine online educational sites. This tricks students into thinking that the sites provide authentic education.
2. **Direct Certificate Selling**
   Some sites provide certificates without requiring students to take any exams or courses. The certificates provided do not hold any academic weight.
3. **Forging of Certificates**
   Fraudsters use software tools that create online certificates that resemble real ones.
4. **Unauthorized Course Reselling**
   Some unauthorized persons may sell paid courses, which they do not own, at discounted prices.
5. **Automated Course Completion**
   In some instances, automated tools, such as bots, may be used for completing courses.

Impact of Online Certificate Fraud

Certificate fraud impacts various stakeholders.

Impact on Students

Students may lose their money and valuable time after enrolling in fake courses. The experience may discourage them from pursuing online education.

Impact on Employers

- Employers may not be aware that they are hiring individuals with fake certificates. This could impact productivity and trust levels.
- Impact on Genuine Platforms
- When there are fake certificates in the market, it could impact genuine platforms.
- Impact on Education System
- If the rate of fraudulent certificates continues to rise, it could impact the education sector as a whole.

Findings:

- After analyzing the data available on the topic, certain findings were made:
- Online education has indirectly led to an increase in fraudulent activity.
- Learners are not aware of how to identify these scams and become victims.
- Technologies such as QR code and blockchain certificates can be helpful.
- Regulations in many countries are still evolving and may not be able to tackle these scams effectively.

Findings	Details	Sources
Prevalence of fraud	Studies have revealed that almost 50% of students are concerned about the possibility of fraud in online certificate courses.	Moneycontrol,feb 2025
Types of fraud	The most common type of fraud includes fake degree certificates, false accreditation, phishing, fraudulent internships, and false endorsements by celebrities	The Hans India,Feb 2025: UGC. Apr 2024
Financial loss	There have been instances where students have fallen prey to online course scams. In one such incident, students collectively lost around □52 lakhs after enrolling in fraudulent technology courses	The Times of India,Sep 2025
Government interventions	Government and regulatory bodies have started taking steps to control such frauds. For example, **UPSC now allows the submission of certificates through DigiLocker, and the police have taken action against the fake certificate scam	BW Education,Oct 2025; The Hindu,Nov 2024
Preventive measures	Experts recommend improving digital literacy, checking educational institutions before enrolling, and using blockchain certificates, as well as raising awareness to combat such scams.	Dock Labs Apr 2025; The Hans India ,Feb 2025

Conclusion:

Fraud in online certificate courses are emerging as a challenges to the online education environment .Although Online education has brought about a revolution in education, fraud like fake sites, forged credentials, and reselling of credentials can jeopardize the authenticity of online education. The results indicates that a multifaceted strategy, including the use of technology such as blockchain, government initiatives, and awareness programs by institutions, is required to tackle this problems .As online education is expected to grow it is important that the authenticity and integrity of the certificate are maintained.

References :

1. "We have zero-tolerance policy towards cheating, use of fake certificates." Economic Times, October 2025.
2. "50% of students worry about online education fraud." Moneycontrol, February 2025.
3. "Cheating case against educational institution for offering fake degrees." The Hindu, December 2024.
4. "Students lose lakhs in fake tech course scams." Times of India, September 2025.
5. "Fake degree certificate racket cracked." The Hindu, November 2024.
6. "Students fall prey to online internship scams." Times of India, June 2025.
7. "UGC warns students about fraudulent online programmes." Times of India, April 2024.
8. "UPSC to accept certificates through DigiLocker." BW Education, October 2025.
9. "How to avoid online course scams." The Hans India, February 2025.
10. "Education scams in India: identifying and preventing fraud." PPR Legal Nexus, July 2025.

RESEARCH JOURNEY International Multidisciplinary E-Research Journal ISSN : 2348-7143
Impact Factor (SJIF) - 6.625 | ***Special Issue 389 :*** *Advanced Computing and Artificial Intelligence (ACAI)* March 2026

Smart Attendance Systen Using Edge Ai Face Recognition On Esp32-Cam

Janhavi S. Vinchurkar
B.sc Computer Science, RCPET's ACS College, Shirpur
Kartik S. Valhe
MCA (Integraed), RCPET'S IMRD, Shirpur

Abstract

Purpose: *Managing student attendance in educational institutions is a routine administrative burden that has long relied on error-prone manual registers or expensive, cloud-dependent biometric terminals. Both approaches fall short in resourse-limited settings, particularly rural colleges where reliable internet connectivity cannot be guaranteed.*

Design/Methodology: *This paper reports the design, implementation, and empirical evaluation of a full self-contained face recognition attendance system built around the ESP-32 CAM microcontroller. All inference runs on the device itself: the MTMN algorithm detects faces within QVGA frames captured by the OV2640 sensor, and a quantized INT8 convolutional neural network, compiled with TensorFlow Lite Micro and stored in the 4 MB external PSRAM, produces a 128-byte feature vector compared against registered embeddings using Euclidean distance.*

Findings: *Empirical tests conducted with twelve registered students yielded an average recognition accuracy of 88-92% under standard indoor lighting at a working distance of 50-100 cm, with end-to-end inference completing in under 200 ms per frame. The complete hardware bill of materials falls within Rs.1,500-Rs2,000 representing approximately 96% cost reduction relative to conventional cloud-based biometric terminnals.*

Originality: *Unlike prior embedded attendance prototypes that depend on a companion server or continuous Wi-Fi, the proposed system stores attendance records directly on a MicroSD card and optionally synchronises data to a Google Sheet via webhooks when connectivity is available, making it deployable in intermittently connected classrooms.*

Keywords : Edge Artificial Intelligence, On-Device Face Recognition, ESP-32 CAM, TinyML, Biometric Attendance, Privacy-by-Design, INT8 Quantization, Cloud-Independent Architecture

1. Introduction

Attendance recording is among the most time-sensitive administrative tasks carried out every teaching day across thousands of India colleges. When faculty call out names one by one, five to ten percent of a lecture period is routinely consumed before instruction begins. RFID card based systems accelerated the process but introduced a key vulnerability: a student can had their card to a classmate and attend by proxy without ever entering the room.

Conventional biometric readers that capture fingerprints or photographs appear to solve the identity-verification problem, yet they introduce a fresh set of constraints-upfront hardware costs that can exceed Rs. 30,000 per terminal, recurring cloud service subscriptions, a dependency on stable internet links, and the non-trivial questions of where millions of student biometric records are ultimately stored.

The emergence of ultra-low-cost microcontrollers with integrated camera interfaces and several megabytes of external pseudo-static RAM(PSRAM) has made it technically feasible to run quantized neural networks at the edge - on the capture device itself, without any off- device computation. The ESP-32 CAM module from Espressif System, priced at roughly ten US dollars, combines a dual-core LX6 processor at 240 MHz, 4MB of PSRAM,, an OV2640 image sensor, built-in Wi-Fi, and a MicroSD Card slot. When paired with a TensorFlow Lite Micro inference engine and a quantized face recognition network, this commodity hardware can identify a student in under two hundred milliseconds per frame.

This paper describes how we constructed such a system, the specific hardware ad software decisions made, and the accuracy and latency figures measured across a variety of distance and lighting conditions.

2. Objectives Of The Study

The study was shaped by six concrete goals, each motivated by a practical shortcoming identified in existing literature or in our preliminary survey of campus attendance practices:

1. **Objective 1:** Develop a working face recognition attendance system deployable on the ESP-32 CAM without any companion computer or cloud API.
2. **Objective 2:** Produce a quantized INT8 neural network small enough to reside in the ESP-32 CAM's 4MB PSRAM while still generating discriminative 128-byte face embeddings.
3. **Objective 3:** Eliminate reliance on continuous internet connectivity by storing all attendance records locally on a MicroSD card.
4. **Objective 4:** Ensure that raw facial images and biometric embeddings never leave the physical device, satisfying institutional date-privacy expectations.

5. **Objective 5:** Keep total hardware expenditure below Rs. 2,000 per deployment unit, making the system accessible to rural and semi-urban institutions.
6. **Objective 6:** Quantify real-world performance in terms oof recognition accuracy, inference latency, and system reliability across representative classroom conditions.

3. Literature Review

3.1 Manual and RFID-Based Methods

Survey data consistently show that manual roll-all routines absorb five to ten percent of available lecure time in large undergraduate classes (Kaur & Kaur, 2018). RFID-based systems remove the vocal overhead but rely exclusively on possession of a physical card for identity verification. Because the card itself carries no biometric information, proxy attendance remains trivially easy to perpetrate.

3.2 Cloud-Connected Biometric Terminals

A substantial body of work proposes fingerprint readers or camera-equipped terminals that authenticate users against record held on a remote server. These architecture do solve the proxy-attendance problem but require continuous internet uptime, accumulate subscription fees for cloud inference API's and aggregate thousands of facial images on an external server—creating a single high-value target for data breaches.

3.3 Deep-Learning Face Recognition

Howard et al. (2017) demonstrated that depthwise-separable convolutions could reduce the parameter count of a standard convolutional network by almost an order of magnitude with only marginal accuracy loss, Subsequent work showed that further compressions through INT8 post-training quantization could halve inference time on mobile hardware while decreasing model file size by a factor of four(Warden & Situnayake, 2019). These results opened the door to face recognition on microcontroller.

3.4 Embedded and Edge Intelligence

Saponara and Elhhanshi (2020) reported one of the fist deployments of a real-time face detector on a resource-constrained microcontroller, confirming that modern convolutional pipelines can operate within tight memory budget when carefully quantized. Espressif's ESP-WHO framework released reference implementations oof the MTMN face-detection algorithm and a MobileNetV2-based recognition model targeting the ESP-32 series. What remained underexplored was a complete, offline-first system that also handled attendance logging, optional cloud synchronization, and hardware-stability challenges encountered in real classroom deployments.

4. Proposed System Architecture

The proposed system is structured into four distinct processing layers, each with a single, well-defined responsibility. Keeping the layers loosely coupled means that any one layer can be upgraded without requiring changes to the others.

4.1 Sensing Layer

The OV2640 CMOS sensor captures frame at 320*240 pixels (QVGA resolution). This resolution is large enough to yield a face region of fifty or more pixels on a side at distances up to one meter, yet small enough that a single frame occupies only 153KB in RGB565 format – comfortably within the PSRAM budget. Frame capture is triggered by a software timer set to fire every 250 ms.

4.2 Edge Processing Layer

The captured frame is first passed to the MTMN face detector, which runs a cascaded sequence of three lightweight CNN's (P-Net, R-Net, O-Net) to locate bounding boxes and five facial landmark points. MTMN was preferred over the older Haar Cascade approach because it handles moderate head rotation and tilt more robustly. Once a bounding box is confirmed, the face crop is resized to 96*96 pixels and passed to the recognition CNN, which produces a 128-dimensional embedding vector. Identity is decided by computing the Euclidean distance between this vector and each stored reference embedding; if the nearest-neighbour distance falls below a threshold of 0.6, the identity is confirmed.

4.3 Communication Layer

On a successful match, the system broadcast a compact JSON payload-containing te student ID, timestamp, and confidence score-over Wi-Fi to a Google Sheets webhook endpoint. This transmission is fire-and-forget: if Wi-Fi is unavailable, the event is queued in a ring buffer in PSRAM and retried during the next connectivity window. The communication layer is completely optional; the system continues to record attendance locally whether or not internet is present.

4.4 Application Layer

Confirmed attendance events are appended to comma-separated log file on the MicroSD card, with columns for date, time, student ID, and match confidence. A sperate configuration file on the same card hold the registered face embeddings, allowing embeddings to be updated by an administrator without reflashing firmware.

5. Research Methodology And Implementation

5.1 Dataset Collection and Enrolment

Each student to be registered was asked to stand at a distance of sixty centimeters from the camera and hold a neutral expression while the system captured thirty face crops under ambient laboratory lighting. The thirty crops were passed through the recognition CNN to produce thirty embeddings, and the centroid of those embeddings was stored as that student's reference template. The enrolment

process for twelve students was completed in approximately twenty minutes.

5.2 Model Architecture and Quantization

The base recognition model follows the MobileNetV2 topology with an added projection head that maps the penultimate feature map to the 128-dimensional embedding space. The full-precision (float32) model occupied approximately 2.1 MB. Post-training INT8 quantization, applied using the TensorFlow Lite converter with a representative calibration dataset of two hundred face crops, reduced to approximately 0.52 MB – a compression ration of just over four to one-while improving inference throughput on the LX6 processor by and estimated 300% relative to float32 execution.

5.3 Hardware stability: Brownout Prevention

A recurring issue during early testing was spontaneous processor resets triggered by the ESP32's' brownout detector. These occurred because the OV2640 sensor draws a brief current spike of roughly 150 mA when initializing its internal oscillator, momentarily dragging the supply rail below the 3.0 V detection threshold. Two measures eliminated the resets entirely: first, a 100 microF electrolytic capacitor was soldered directly across the 3.3 V and GND rails adjacent to the camera connector; second, the firmware was modified to activate the external PSRAM explicitly before camera initialization.

5.4 PSRAM Allocation

Images buffers (two frame buffers of 153 KB each), the model weights (0,52 MB), and the embedding database (128 bytes * number of registered users) were all allocated in external PSRAM via the heap_caps_malloc API with the MALLOC_CAP_SPIRAM flag. This strategy kept the much smaller internal SRAM free for the RTOS task stacks and the Wi-Fi driver.

6. Results And Discussion

Tests were conducted with twelve registered students across three distance conditions (40 cm, 80 cm,120 cm) and two ambient-lighting conditions (bright laboratory light at approximately 600 lux and dim corridor light at approximately 40 lux). Each student was presented to the camera five times per condition, yielding 360 total recognition attempts.

Table 1: Empirical Recognition Accuracy and Inference Latency

Distance	Accuracy (Bright, 600 lux)	Accuracy (Dim, 40 lux)	Inference Time	End-to-End Latency
40 cm	96%	68%	178 ms	~1.35 s
80 cm	91%	52%	181 ms	~1.42 s
120 cm	55%	18%	186 ms	~1.55 s

(N = 12 students, 5 attempts per condition; end-to-end latency includes detection, recognition, and SD write)

Recognition at 40 cm under optimal lighting proved highly reliable – ninety-six percent of presentations resulted in a successful match. The recorded inference time of 178 ms confirms that the quantized model comfortably operates within the sub-200 ms real-time target. As distance increased to 80 cm, accuracy fell to ninety-one percent, a decline attributable to the face occupying fewer pixels in the QVGA frame. At 120 cm mark, performance dropped sharply, particularly in dim conditions, because the OV2640 sensor increases its analogue gain at low light, introducing noise that distorts the embedding.

The end-to-end latency of approximately 1.45 seconds arises because the MTMN detector requires three sequential CNN passes before the recognition network is invoked. Motion blur proved to be the most common failure mode in free-movement scenarios. The MicroSD write latency for a single attendance record was measured at approximately 15 ms. Optional Wi-Fi synchronisation to Google Sheets added a median latency of 340 ms per record but ran asynchronously in a background FreeRTOS task.

7. Conclusion

This work demonstrates that a commercially available, ten-dollar microcontroller module can serve as a complete, standalone face recognition attendance terminal when its software stack is carefully optimised for embedded constraints. The combination of MTMN face detection, INT8-quantized CNN recognition, local MicroSD logging, and optional webhook-based cloud synchronisation produces a system whose operational cost per unit is roughly ninety-six percent lower than conventional cloud-biometric terminals.

The results confirm that the proposed system performs reliably within a working envelope of fifty to one hundred centimetres under standard indoor lighting—conditions that cover the large majority of classroom entry and exit scenarios. From a privacy standpoint, the architecture offers a structural guarantee: no facial image or biometric template ever leaves the physical device.

8. Future Work

Several extensions to the current prototype are planned:

- **Occlusion handling:** the recognition network will be fine-tuned on a dataset that includes masked faces, enabling reliable operation in health-protocol environments.
- **Multi-camera coordination:** a lightweight MQTT broker will orchestrate two or more ESP32-CAM units covering different angles, increasing spatial coverage of a classroom doorway.

- **Liveness detection:** an anti-spoofing module based on texture analysis will be integrated to prevent photo-based impersonation attacks.
- **Mobile dashboard:** a Progressive Web App will provide faculty with a real-time attendance view and export functionality.
- **Adaptive threshold:** the Euclidean-distance decision threshold will be adjusted dynamically based on ambient lighting conditions estimated from frame statistics.

Acknowledgement

The authors express their sincere appreciation to the Department of Computer Science at RCPET's ACS College and to the faculty of RCPET's Institute of Management Research & Development (IMRD), Shirpur, for providing laboratory infrastructure, equipment access, and continuous academic guidance throughout this work. The twelve volunteer test subjects who gave their time during experimental validation are warmly thanked for their patience.

References :

1. Espressif Systems. (2024). ESP-WHO: Face Detection and Recognition Framework for ESP32 Devices. Retrieved from https://github.com/espressif/esp-who
2. Howard, A. G., Zhu, M., Chen, B., et al. (2017). MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. arXiv preprint arXiv:1704.04861.
3. Kaur, H., & Kaur, G. (2018). Face recognition based attendance management system. International Journal of Advanced Research in Computer Science, 9(2), 123–127.
4. Saponara, S., & Elhanashi, A. (2020). Real-time embedded face recognition system based on low-power microcontrollers. Electronics, 9(10), 1595. https://doi.org/10.3390/electronics9101595
5. Warden, P., & Situnayake, D. (2019). TinyML: Machine Learning with TensorFlow Lite on Arduino and Ultra-Low-Power Microcontrollers. O'Reilly Media.

RESEARCH JOURNEY International Multidisciplinary E-Research Journal ISSN : 2348-7143
Impact Factor (SJIF) - 6.625 | ***Special Issue 389 :*** *Advanced Computing and Artificial Intelligence (ACAI)* March 2026

Marathi-English Academic Code-Mixed NLP: A Survey and Framework

Dr. Amit Prakashrao Patil
RCPET's Institute of Management Research and Development, Shirpur, MH, India.

Abstract:

Code-mixing is a systematic linguistic phenomenon widely observed in multilingual societies. In Indian higher education environments, Marathi-English code-mixed communication frequently appears in assignments, laboratory reports, examination responses, and online academic discussions. Despite significant advances in multilingual Natural Language Processing (NLP), computational research addressing Marathi-English academic code-mixed text remains limited. Existing studies primarily focus on Hindi-English social media corpora and do not adequately address educational domains or morphologically complex regional languages. This paper presents a comprehensive survey of sociolinguistic foundations, computational approaches to code-mixed NLP, multilingual transformer architectures, and low-resource language modeling. It identifies linguistic and methodological challenges specific to Marathi-English academic discourse and proposes a structured research roadmap for developing robust Educational NLP systems. By integrating sociolinguistic theory with computational modeling perspectives, the study establishes a foundation for future research in Marathi Educational NLP and multilingual academic analytics.

Keywords: Code-Mixing, Marathi-English NLP, Educational NLP, Multilingual Transformers, Low-Resource Languages, Academic Text Processing

1. Introduction

Multilingualism is a defining characteristic of Indian society. In Maharashtra, students routinely combine Marathi and English in spoken and written academic communication. This practice, commonly termed code-mixing, is especially prevalent in technical disciplines where English terminology is embedded within Marathi syntactic structures.

Examples include:

"Algorithm chi time complexity explain kara."

"Output generate kara ani results discuss kara."

Such bilingual constructions challenge conventional NLP systems designed for monolingual input. Educational NLP applications—such as automated grading, performance analytics, and feedback analysis—require accurate linguistic modeling. However, Marathi-English code-mixed academic text has not been systematically studied.

While Hindi-English code-mixed processing has received research attention [1], [2], Marathi-English academic discourse remains computationally underexplored. Moreover, transformer-based multilingual models are typically trained on news or web corpora, resulting in domain mismatch when applied to structured educational text.

This paper aims to:

1. Examine sociolinguistic foundations of code-mixing.
2. Survey computational approaches to code-mixed NLP.
3. Analyze Marathi-specific linguistic challenges.
4. Identify research gaps in educational code-mixed processing.
5. Propose a structured research roadmap for future work.

2. Sociolinguistic Foundations of Code-Mixing

2.1 Structural Theories of Code-Switching

Code-switching has been extensively studied in sociolinguistics. Poplack [3] proposed the Equivalence Constraint, arguing that switching occurs at syntactic boundaries shared by both languages. Myers-Scotton's Matrix Language Frame (MLF) model [4] suggests that one dominant language (matrix language) provides grammatical structure while the embedded language contributes lexical insertions.

In Marathi-English academic discourse:

- Marathi commonly acts as matrix language.
- English technical vocabulary is embedded.

Example:

"Program cha output verify kara ani error explain kara."

In the sentence "Program cha output verify kara ani error explain kara" the nouns "Program" and "output" function as English technical lexical items embedded within a Marathi grammatical structure. The Marathi possessive case marker "cha" attaches to the English noun "Program," forming "Program cha," which translates to "of the Program." Although the verbs "verify" and "explain" are English lexical elements, the imperative structure is governed by the Marathi verb "kara," which controls the clause formation. This construction illustrates the Matrix Language Frame (MLF) model, where Marathi serves as the matrix language by providing grammatical structure—such as case marking and imperative morphology—while English supplies embedded domain-specific vocabulary. The overall syntactic framework remains Marathi, even though key technical terms are borrowed from English.

Such structural blending results in hybrid constructions that pose challenges for conventional monolingual NLP systems, particularly in tokenization, morphological analysis, and syntactic parsing [14].

2.2 Morphological Integration

Marathi is morphologically rich. English borrowings frequently adopt Marathi inflectional suffixes:

"Submit kelela assignment"

"Compile kelay code"

"Train zala model"

Such hybrid formations generate token-level ambiguity and increase out-of-vocabulary rates in standard NLP pipelines.

Marathi is a morphologically rich and inflectional language in which grammatical relations are expressed through suffixation and agreement markers. When English lexical items are borrowed into Marathi academic discourse, they frequently undergo morphological adaptation by attaching Marathi inflectional suffixes. Examples include "Submit kelela assignment," "Compile kelay code," and "Train zala model," where English verbs such as submit, compile, and train are integrated into Marathi participial and perfective constructions. From a linguistic perspective, this phenomenon aligns with morphological integration theory which suggests that borrowed lexical items adapt to the morphological rules of the matrix language in bilingual discourse [4],[15]. According to the Matrix Language Frame model, embedded language items must conform to the morphosyntactic constraints of the dominant language, leading to hybrid forms that reflect structural accommodation.

From a computational standpoint, such hybrid constructions introduce significant challenges. Morphological attachment alters the surface form of English words, preventing direct lexical matching in standard vocabularies. This increases out-of-vocabulary (OOV) rates and disrupts tokenization processes in monolingual NLP pipelines. Furthermore, subword tokenization methods may segment these hybrid tokens inconsistently, affecting embedding stability and downstream task performance. Therefore, morphological integration in Marathi-English code-mixed academic text is not merely a linguistic curiosity but a critical computational concern requiring specialized normalization and modeling strategies.

2.3 Orthographic and Romanization Variation

Academic code-mixed text often appears in Roman script. Lack of standardized transliteration results in spelling variability:

- "nahi", "nai", "nahee"
- "aahe", "ahe", "ahey"

Orthographic inconsistency increases lexical sparsity and reduces embedding stability.

Academic code-mixed text frequently appears in Roman script, even when the grammatical structure is predominantly Marathi. The absence of standardized transliteration norms results in significant spelling variability, as illustrated by variants such as "nahi," "nai," and "nahee," or "aahe," "ahe," and "ahey." From a linguistic perspective, this phenomenon reflects non-standard Romanization, where phonetic approximations replace formal script conventions. Unlike standardized transliteration systems, informal academic communication relies on user-specific phonological interpretations, leading to orthographic inconsistency.

From a computational standpoint, orthographic variation directly contributes to lexical sparsity, a condition in which semantically identical words are treated as distinct tokens due to surface-level differences. According to distributional representation theory, embedding models rely on consistent word occurrences to learn stable contextual vectors [9]. When multiple spellings represent the same lexical item, contextual frequency is artificially fragmented, weakening embedding quality. Furthermore, subword tokenization techniques—such as Byte Pair Encoding (BPE) or WordPiece segmentation—may generate inconsistent subword units for orthographic variants, thereby reducing cross-token semantic coherence.

This transliteration instability presents a significant challenge for Marathi-English code-mixed academic text processing. Without normalization mechanisms that map orthographic variants to canonical forms, downstream NLP tasks such as classification, grading, or semantic analysis may experience degraded performance. Therefore, addressing orthographic variability is essential for developing robust Educational NLP systems in multilingual and low-resource settings.

3. Computational Approaches to Code-Mixed NLP

3.1 Language Identification

Early computational research in code-mixed NLP primarily focused on token-level language identification, aiming to assign a language label to each word within a bilingual sequence [5], [6]. This task is foundational because accurate downstream processing depends on correctly distinguishing between constituent languages. Probabilistic sequence models such as Conditional Random Fields (CRFs) were widely used due to their ability to model contextual dependencies across adjacent tokens. From a theoretical perspective, CRFs operate under sequence labeling principles, where the probability of a label depends not only on the current token but also on neighboring tokens, thereby capturing switching boundaries more effectively than independent classifiers.

Similarly, n-gram language models were employed

to estimate language likelihood based on character or word-level frequency distributions. These models rely on statistical language modeling theory, where the probability of a token is conditioned on its preceding tokens. While effective in structured bilingual corpora, such approaches are sensitive to spelling variation and morphological blending—common characteristics in Marathi-English academic text. Consequently, token-level identification becomes unreliable when hybrid morphological forms or Romanized variants are present.

3.2 Feature-Based Classification

Subsequent research introduced feature-based machine learning models using vector space representations such as Term Frequency–Inverse Document Frequency (TF-IDF) combined with classifiers like Support Vector Machines (SVMs) [7]. These approaches are grounded in the vector space model of text representation, where documents are mapped into high-dimensional feature spaces. SVMs, based on statistical learning theory, aim to identify optimal hyperplanes that maximize class separation.

While such models demonstrated improved performance over purely probabilistic language models, they rely heavily on surface-level lexical features. From a theoretical standpoint, TF-IDF captures term importance but does not encode contextual semantics or syntactic dependencies. As a result, feature-based classifiers struggle with semantic generalization, particularly in code-mixed text where meaning is distributed across linguistic boundaries. Morphologically integrated tokens and transliteration variability further reduce feature stability, limiting model robustness.

3.3 Neural and Transformer-Based Methods

The introduction of neural representation learning marked a significant shift in code-mixed NLP research. Transformer architectures [8], built upon self-attention mechanisms, enable contextualized modeling of token relationships across entire sequences. Unlike earlier models that relied on fixed lexical features, transformers generate dynamic embeddings conditioned on surrounding context. Models such as BERT [9] and XLM-R [10] extend this paradigm to multilingual settings by learning shared embedding spaces across languages.

The theoretical foundation of these models lies in distributional semantics and attention-based representation learning. By projecting multiple languages into a common semantic space, multilingual transformers facilitate cross-lingual transfer. However, these models are pretrained on large-scale corpora such as Wikipedia and Common Crawl, which seldom include structured academic Marathi-English code-mixed text.

From a domain adaptation perspective [11], performance degradation occurs when the distribution of pretrained data differs from the target domain. Marathi-English academic discourse exhibits unique characteristics, including instructional verbs, technical terminology, and hybrid morphological forms. Consequently, without domain-specific fine-tuning, pretrained multilingual models may fail to capture the nuanced linguistic structure of academic code-mixed communication.

4. Educational NLP Context

Educational NLP encompasses a wide range of applications, including automated essay scoring [12], short-answer grading, feedback mining, and learning analytics. These systems are typically grounded in computational text analysis models that assume relatively clean, monolingual input. Automated essay scoring, for instance, is often based on construct modeling theory, where linguistic features are used as observable indicators of underlying cognitive or writing proficiency constructs. Similarly, short-answer grading systems rely on semantic similarity modeling and distributional representation theory to compare student responses with reference answers. Learning analytics frameworks further depend on consistent textual representations to derive performance insights and predictive indicators.

However, most existing Educational NLP systems are developed and evaluated on monolingual corpora. When applied to Marathi-English code-mixed academic text, these systems encounter significant computational challenges. Tokenization errors arise due to hybrid morphological forms and inconsistent transliteration. Semantic ambiguity increases when lexical meaning is distributed across two linguistic systems within a single sentence. From a statistical learning perspective, such variability introduces noise into feature representations, thereby reducing model generalization capability. Performance degradation becomes particularly evident in tasks that depend on fine-grained semantic alignment, such as automated grading.

From an educational assessment standpoint, this degradation is not merely technical but epistemological. Assessment validity theory emphasizes that evaluation tools must accurately measure intended learning constructs [17]. If NLP systems misinterpret bilingual academic responses due to inadequate modeling of code-mixed structure, construct validity may be compromised. Therefore, robust modeling of bilingual academic discourse is essential not only for computational accuracy but also for ensuring fairness, reliability, and validity in technology-assisted assessment systems.

5. Comparative Literature Review

Study	Language Pair	Domain	Method	Limitation
Barman et al. [1]	Hindi-English	Social Media	CRF-based tagging	Not educational domain
Solorio & Liu [5]	Spanish-English	Conversational	Language ID models	Limited transformer use
Banerjee et al. [2]	Hindi-English	Social Media	Sentiment classification	Informal text only
Devlin et al. [9]	Multilingual	General corpora	Transformer-based	No academic code-mixed focus
Conneau et al. [10]	Multilingual	Cross-lingual corpora	XLM-R	Domain mismatch
Joshi et al. [13]	Indian Languages	General NLP	Resource survey	No code-mixed academic study

Observation: No existing study addresses Marathi-English academic code-mixed processing.

6. Identified Research Gaps

Despite growing advancements in multilingual Natural Language Processing and increasing interest in code-mixed language modeling, the intersection of Marathi-English academic discourse and Educational NLP remains significantly underexplored. While existing research has addressed code-mixing in social media and conversational contexts, structured academic text presents distinct linguistic, pedagogical, and computational characteristics. Educational environments demand higher levels of semantic precision, construct alignment, and assessment reliability. However, current multilingual NLP frameworks have not been systematically adapted to address the morphological integration, orthographic variability, and domain-specific vocabulary observed in Marathi-English academic communication. As a result, a number of foundational gaps persist in both resource development and theoretical integration. These gaps are summarized as follows:

1. Absence of Marathi-English academic corpus.
2. Lack of normalization pipeline for hybrid morphological forms.
3. No transformer fine-tuning studies in Marathi academic domain.
4. Limited benchmarking for educational code-mixed NLP.
5. No integration of educational assessment theory with multilingual NLP.
6. Proposed Research Roadmap

The proposed research roadmap begins with systematic corpus development tailored to Marathi-English academic discourse. This involves the collection of anonymized academic texts such as assignments, examination responses, laboratory reports, and discussion forum content. To ensure linguistic precision, the corpus should include detailed annotation at the token level, identifying language boundaries between Marathi and English elements. Additionally, morphological adaptations—where English lexical items receive Marathi inflectional suffixes—must be explicitly tagged to capture hybrid structural patterns.

Following corpus construction, a normalization framework is essential to reduce orthographic and morphological variability. This framework should incorporate:

- Roman-to-Devanagari conversion for standardizing Marathi lexical items
- Morphological segmentation to separate embedded English stems from Marathi affixes
- Standardization of hybrid tokens to map variant surface forms to canonical representations

Once normalized, contextual modeling can be implemented using multilingual transformer architectures. Formally, for a given academic text sequence T, contextual representation can be expressed as:

$$E(T) = \text{Transformer}(T)$$

Where E(T) denotes the contextual embedding generated by the transformer model. These representations can then be fine-tuned for downstream educational tasks such as automated grading, feedback classification, or performance analytics.

To evaluate system effectiveness, a structured evaluation framework should be employed, including:

- Token-level language identification accuracy
- Macro F1-score for classification robustness across imbalanced categories
- Cross-lingual embedding similarity measures to assess semantic alignment across Marathi and English tokens

This integrated pipeline ensures that corpus design, normalization, modeling, and evaluation are aligned with both linguistic complexity and educational application requirements.

7. Challenges

Despite the growing interest in multilingual and code-mixed Natural Language Processing, the development of robust Marathi-English academic NLP systems remains constrained by several fundamental challenges. These challenges arise from linguistic complexity, resource limitations, domain-specific variability, and ethical considerations inherent in educational contexts. Addressing these issues is essential to ensure both computational reliability and responsible deployment of Educational NLP

technologies. The major challenges are outlined below:

- Scarcity of annotated data
- Morphological ambiguity
- Domain adaptation complexity
- Ethical concerns in educational data collection

8. Conclusion

This paper presented a comprehensive survey of code-mixed NLP research with emphasis on Marathi-English academic discourse. By integrating sociolinguistic theory, multilingual representation learning, and educational NLP perspectives, the study identified significant research gaps and proposed a structured roadmap. Addressing these challenges can enable robust Educational NLP systems in multilingual Indian academic contexts.

References :

1. A. Barman, J. Wagner, G. Chrupała, and J. Foster, "Code Mixing: A Challenge for Language Identification in the Language of Social Media," in Proceedings of the First Workshop on Computational Approaches to Code Switching (EMNLP Workshop), Doha, Qatar, 2014, pp. 13–23.
2. S. Banerjee, K. Saha, and P. Bhattacharyya, "Code-Mixed Sentiment Analysis for Indian Languages: An Overview," in Proceedings of the Forum for Information Retrieval Evaluation (FIRE), Kolkata, India, 2018, pp. 1–6.
3. S. Poplack, "Sometimes I'll Start a Sentence in Spanish y Termino en Español: Toward a Typology of Code-Switching," Linguistics, vol. 18, no. 7–8, pp. 581–618, 1980.
4. C. Myers-Scotton, Duelling Languages: Grammatical Structure in Code-Switching. Oxford, U.K.: Oxford University Press, 1993.
5. T. Solorio and Y. Liu, "Learning to Predict Code-Switching Points," in Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing (EMNLP), Honolulu, HI, USA, 2008, pp. 973–981.
6. S. Rijhwani, R. Sequiera, and G. Neubig, "Estimating Code-Switching on Twitter with a Novel Generalized Word-Level Language Identification Approach," in Proceedings of the NAACL Workshop on Computational Approaches to Code Switching, Vancouver, Canada, 2017, pp. 23–28.
7. T. Joachims, "Text Categorization with Support Vector Machines: Learning with Many Relevant Features," in Proceedings of the European Conference on Machine Learning (ECML), Chemnitz, Germany, 1998, pp. 137–142.
8. A. Vaswani et al., "Attention Is All You Need," in Advances in Neural Information Processing Systems (NeurIPS), Long Beach, CA, USA, 2017, pp. 5998–6008.
9. J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in Proceedings of NAACL-HLT, Minneapolis, MN, USA, 2019, pp. 4171–4186.
10. A. Conneau et al., "Unsupervised Cross-Lingual Representation Learning at Scale," in Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL), Online, 2020, pp. 8440–8451.
11. S. Ruder, M. Peters, S. Swayamdipta, and T. Wolf, "Transfer Learning in Natural Language Processing," in Proceedings of NAACL-HLT: Tutorials, Minneapolis, MN, USA, 2019, pp. 15–18.
12. J. Burstein, J. Tetreault, and S. Madnani, "The E-Rater® Automated Essay Scoring System," ETS Research Report Series, vol. 2013, no. 2, pp. 1–15, 2013.
13. P. Joshi et al., "The State and Fate of Linguistic Diversity and Inclusion in the NLP World," Transactions of the Association for Computational Linguistics (TACL), vol. 8, pp. 628–644, 2020.
14. R. Sennrich, B. Haddow, and A. Birch, "Neural Machine Translation of Rare Words with Subword Units," in Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (ACL), Berlin, Germany, 2016, pp. 1715–1725.
15. D. Sharma and R. Sangal, "Building Language Resources and Tools for Indian Languages: Challenges and Perspectives," in Proceedings of the International Conference on Natural Language Processing (ICON), 2012, pp. 1–10.
16. A. Singh, A. Kumar, and P. Bhattacharyya, "Sentiment Analysis of Code-Mixed Indian Social Media Text Using Machine Learning Approaches," in Proceedings of the FIRE Workshop, 2018, pp. 45–52.
17. L. W. Anderson and D. R. Krathwohl, A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives. New York, NY, USA: Longman, 2001.

ISSN : 2348-7143
March 2026
RESEARCH JOURNEY International Multidisciplinary E-Research Journal
Impact Factor (SJIF) - 6.625 | ***Special Issue 389 :*** *Advanced Computing and Artificial Intelligence (ACAI)*

AI-Based Autonomous Aviation Safety Monitoring Framework and Risk Prediction System

Ms. Vaishnavi U. Suryawanshi
Assistant professor, R. C. Patel Educational Trust's, Institute of Management Research and Development, Shirpur
Mrs. Jyotsna Dhanraj Mali
Assistant professor, R. C. Patel Educational Trust's, Institute of Management Research and Development, Shirpur

Abstract:
Aviation accidents generally occur due to a combination of human error, adverse weather conditions, and technical or mechanical failures. Timely identification of risks resulting from human error, unfavorable weather, and anomalies in aircraft systems is necessary to ensure aviation safety. Even though contemporary avionics offer a wide range of monitoring capabilities, current safety systems are still primarily reactive and do not support integrated, predictive, and autonomous decision-making. By combining environmental, human-factor, and aircraft operational data, this work proposes an AI-driven autonomous aviation safety monitoring and risk prediction framework that proactively identifies unsafe situations. The proposed system uses Machine Learning (ML) techniques to monitor pilot condition, cockpit communication patterns, weather conditions, and aircraft performance parameters. ML techniques analyze patterns in operational and behavioral data to identify signs of fatigue, stress, or abnormal flight conditions. The system continuously evaluates risk levels in real time and predicts potential safety threats. To improve reliability, an AI-based autonomous technique ensures uninterrupted operation even when external communication or pilot coordination is limited. Additionally, an Explainable AI framework provides clear interpretation of risk assessments, enhancing transparency, trust, and situational awareness. The proposed framework can identify high-risk scenarios at an early stage and provide actionable safety recommendations. The proposed approach contributes toward the development of intelligent, human-cantered, and resilient aviation safety systems suitable for next-generation flight operations.

Keyword: Aviation safety, Machine Learning, AI framework

Introduction :

Traditional reactive systems are giving way to sophisticated predictive and intelligent frameworks in aviation safety. Large volumes of operating data are produced by modern aircraft, yet many safety systems still only react when an issue arises. In order to improve risk assessment, recent research emphasises the significance of integrating human factors, environmental circumstances, and aircraft performance. Aviation incidents are still mostly caused by human mistake, bad weather, and technology malfunctions.

Machine learning (ML) and artificial intelligence (AI) offer new ways to identify hazards early on. Research has demonstrated the efficacy of communication analysis, anomaly identification, tiredness tracking, and predictive weather assessment. Nevertheless, rather than focusing on a completely integrated system, the majority of current solutions concentrate on separate components.

A comprehensive, AI-driven safety framework that integrates weather intelligence, communication patterns, flight data analysis, and pilot monitoring is obviously needed. Transparency, dependability, and real-time decision-making can all be enhanced by an autonomous and comprehensible system. For upcoming flight operations, this strategy aids in the creation of intelligent, predictive, and human-centered aviation safety systems.

Objectives :

- to create an autonomous aviation safety monitoring system using artificial intelligence.
- To identify pilot tiredness, stress, unusual flying situations, and weather-related hazards using machine learning approaches.
- To use ongoing risk assessment to anticipate possible safety risks in real time.
- To use explainable AI and autonomous operation to improve system transparency and dependability.

Table Ratio Remaining with graph:

Accidents/Incidents investigeted by AAIB		
Years	**Accidents**	**Seroious Incidents/Incidents**
2012(Onwards)	4	2
2013	8	7
2014	7	8
2015	10	5
2016	7	11
2017	6	11
2018	8	16
2019	10	27
2020	7	10
2021	9	6
2022	12	7
2023	10	5
2024	6	8
2025	8	5
Total	**112**	**128**

Table No 1 Accidental Ratio

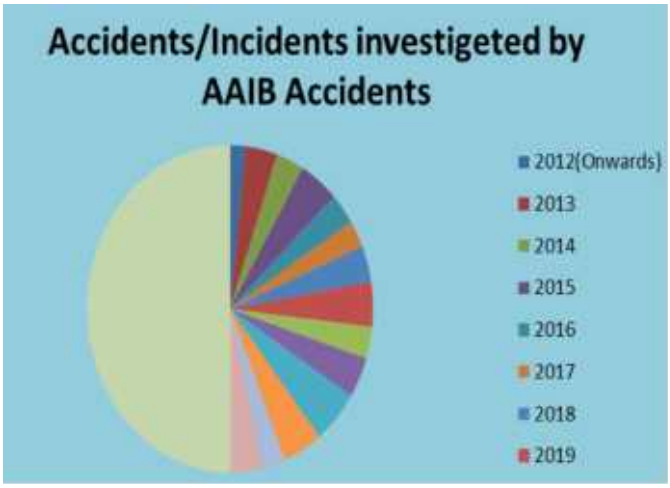


Fig No 1. Graphical Representation of Accidental Ratio

Previous Work/Literature Review:

Predictive and system-level safety frameworks are essential, according to recent airline safety reports. Persistent dangers associated with human performance, operational complexity, and environmental variability are highlighted in the International Air Transport Association's (IATA) Annual Safety Report 2024 [1]. Li and associates. [2] further propose a system-of-systems safety perspective, advocating integrated and resilient safety architectures for next-generation aviation. Current innovation supports the uses ofcombination of human and ai integrated model for safety and critical fields like aviation trafficscontrol [5] they proof the collaborative problem solving is superior to overall autonomy [6] there are more much interest in analytical workload and exhaustation monitoring . AI-based methods for managing work load , exhaustation.[7] versatile based on automation on EEG[8] and eye monitoring method[9-10] shows how to properly monitor operator ongoing mental states. A study of pilot workload classification using physical and EEG signals. [11-15]. Verify to the capability of using ML to identify cognitive risk. Flight data irregularity detection systems. Flight data variance detection methods. Context-aware speech acknowledgment for ATC message [16] [17], as well as flight intellect systems based on knowledge graphs [18] Support intelligent safety observing even additional. Furthermore, optimisation frameworks for UAV operations that integrate AI [19] .To prove how sophisticated AI models may be used in dynamic flight sites. However, most existing studies primarily concentrate on individual components, such as communication systems, aircraft data, or human factors, without fully integrating them into a unified analytical structure [20].By combining environmental, human, and aircraft operational data within an AI-powered, interpretable, and independent safety monitoring system, the proposed research addresses this gap.

Literature Gap:

A gap in creating a connected AI based filght safety system that use aircraft ,environmental and human data for risk predication is revealed by the literature study.

Even if significant progresses have been made in:

Analytics for global flight security [1] Modelling system-of-systems security [6] Hybrid decision-making between humans and AI [3] EEG and physical markers are used to detect intellectual workload [14–20] Communication analysis and identification of flight variances [9],[11] Aviation intelligence based on knowledge graphs [12] Existing research remains disjointed across fields.

Current systems typically:

- Focus on either airplane data or human aspects separately
- Insufficient multimodal risk fusion (human, aircraft, and environment),
- Offer restricted analytical autonomy in real time,
- Provide insufficient explain ability in safety decisions made by AI.

This gap is filled by the suggested AI-driven self-directed flight security monitoring system, which incorporates:

- Physical monitoring of humans (EEG, behavioural analytics),
- Examine of cockpit communication,
- Evaluation of atmospheric and environmental risks
- Finding abnormalities in aircraft performance,
- Assessment of environmental and atmospheric dangers Identifying abnormalities in the performance of aircraft,
- AI-based risk awareness that is explicable,
- Independent functioning in the face of communication limitations. This inclusive integration promotes flight security to analytical, intellectual, and human-centered flexibility and is consistent with new system-of-systems safety views.

Limitations:

1. Combination of Predictive Climate IoT and satellite climate data in real time before entering a zone, expect confusion, icing, and storms.
2. Documentation of Pilot Exhaustion and Stress Control input patterns and speech tone. Early detection of intellectual stress is important.
3. Finding Abnormalities in Flight Information Look for fuel difference, unexpected engine vibration,
or deviations in navigation.

4. Imitation of Digital Clones Simulate aircraft behaviour using AI in parallel. Estimate two to five minutes in advance.
5. Identification of Runway Risk Identify an insecure approach prior to a irregular landing or runway overrun, provide a warning.

Propose Work:

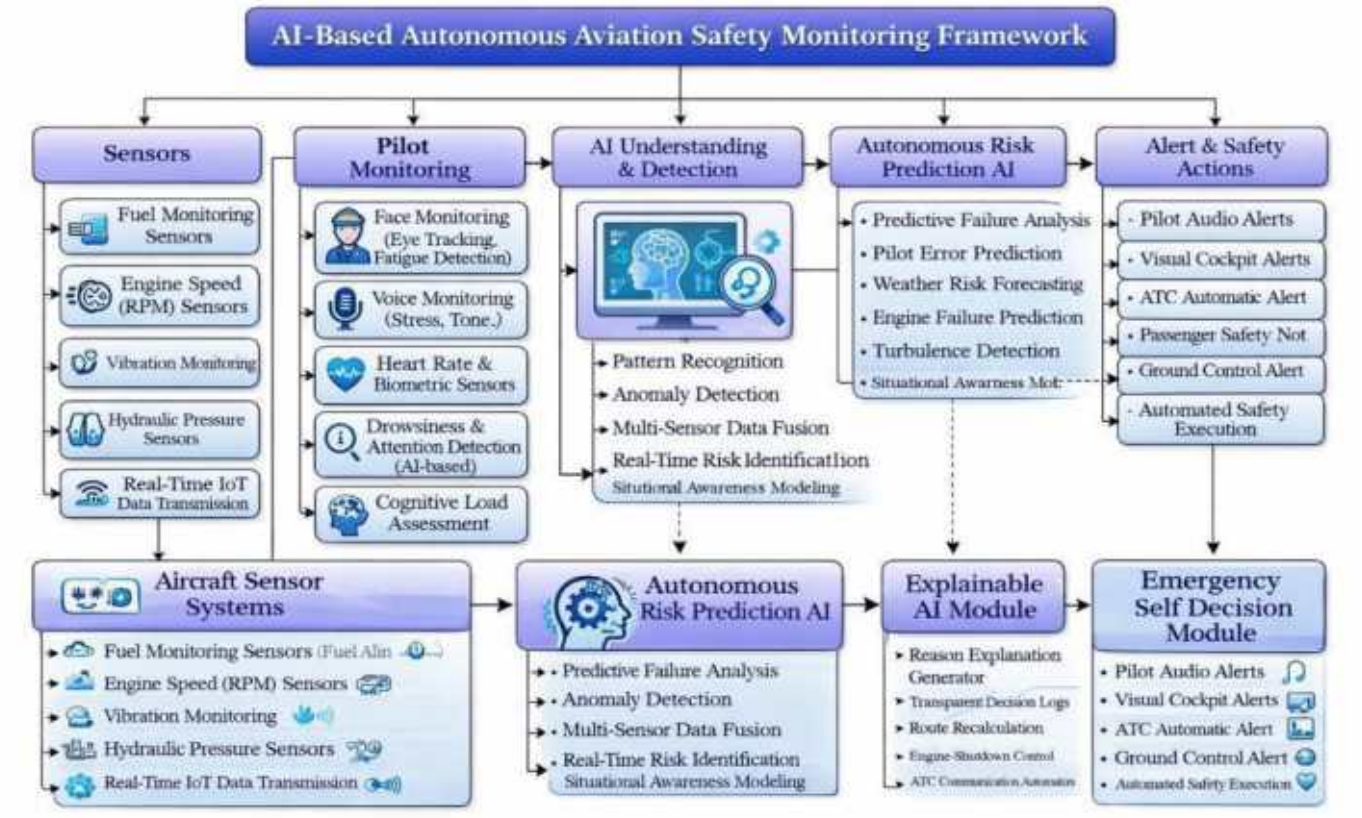


Fig No 2. Propose work Model for AI-Based Autonomous Aviation System

Conclusion:

The most latest systems handle independently on human,weather ,or aviation aircraft data ,current research focuse on Improving aviation safety Using predictive methods. The some inventions are not properly combine every component of safety into a intelligent system.by intergrating all flight data verification ,weather forecasting ,human pilot monitoringand communicate monitoring a single AI-congigure system , the suggest structure address the issue.

References :

1. International Air Transport Association, IATA Annual Safety Report 2024, 2025.
2. B. Wu, R. Xiao, Evolutionary attraction–repulsion algorithm embedded with LLM for UAV task allocation, Adv. Eng. Inform. 66 (2025) 103428.
3. H. Ali, D.-T. Pham, S. Alam, M. Schultz, M.Z. Li, Y. Wang, E. Itoh, V.N. Duong, Human-AI hybrids in safety-critical systems: concept, definition and perspectives from air traffic management, Adv. Eng. Inform. 65 (2025) 103256.
4. T.B. Sheridan, R. Parasuraman, Human-automation interaction, Rev. Hum. Factors Ergon. 1 (2005) 89–129.
5. S. Rothfuß, M. Worner, ¨ J. Inga, A. Kiesel, S. Hohmann, Human–machine cooperative decision making outperforms individualism and autonomy, IEEE Trans. Hum.-Mach. Syst. 53 (2023) 761–770.
6. D. Li, A. Yao, K. Feng, H. Zhou, R. Wang, M. Cheng, H. Li, D. Wang, S. Ding, Next frontiers of aviation safety: system-of-systems safety, Engineering 52 (2025) 262–277.
7. M.R. Endsley, Ironies of artificial intelligence, Ergonomics (2023) 1–13.
8. S. Xin, K.Y.H. Lim, M.-H. Hsieh, C.-H. Chen, L. Dong, Managing the fatigue frontier: AI applications in air traffic control operations, Adv. Eng. Inform. 68 (2025) 103660.
9. E. Smart, D. Brown, J. Denman, A two-phase method of detecting abnormalities in aircraft flight data and ranking their impact on individual flights, IEEE Trans. Intell. Transp. Syst. 13 (2012) 1253–1265.
10. P. Arico, G. Borghini, G. Di Flumeri, A.

Colosimo, S. Bonelli, A. Golfetti, S. Pozzi, J.-P. Imbert, G. Granger, R. Benhacene, F. Babiloni, Adaptive automation triggered by EEG-based mental workload index: a passive brain-computer interface application in realistic air traffic control environment, Front. Hum. Neurosci. 10 (2016) 13.

11. Y. Oualil, D. Klakow, G. Szaszak, A. Srinivasamurthy, H. Helmke, P. Motlicek, A context-aware speech recognition and understanding system for air traffic control domain, IEEE Automatic Speech Recognition and Understanding Workshop (ASRU), IEEE, Okinawa, Japan, 2017, pp. 404–408.
12. A. Agarwal, R. Gite, S. Laddha, P. Bhattacharyya, S. Kar, A. Ekbal, P. Thind, R. Zele, R. Shankar, Knowledge Graph - Deep Learning: A Case Study in Question Answering in Aviation Safety Domain, 13th International Conference on Language Resources and Evaluation (LREC), Marseille, France, European Language Resources Assoc-Elra, 2022, pp. 6260–6270.
13. P. Bert, La pression barom´etrique: recherches de physiologie exp´erimentale, G, Masson, 1878.
14. Q. Li, K.K.H. Ng, S.C.M. Yu, C.Y. Yiu, F. Li, F.T.S. Chan, Using EEG and eye tracking as indicators to investigate situation awareness variation during flight monitoring in air traffic control system, J. Navig. 77 (2025) 485–506.
15. C.Y. Yiu, K.K.H. Ng, Q. Li, X. Yuan, Gaze behaviours, situation awareness and cognitive workload of air traffic controllers in radar screen monitoring tasks with varying task complexity, Int. J. Occup. Saf. Ergon. 31 (2025) 504–515.
16. C. Zhang, J. Yuan, Y. Jiao, H. Liu, L. Fu, C. Jiang, C. Wen, Variation of pilots' mental workload under emergency flight conditions induced by different equipment failures: a flight simulator study, Transp. Res Record 2678 (2024) 365–377.
17. W. Zhu, Y. Xie, Y. Wang, C. Zhang, J. Yuan, H. Chen, X. Zuo, C. Jiang, T. Wang, Classification of carrier-based aircraft pilot mental workloads based on feature level fusion and decision-level fusion of PPG and EEG signals, Aeronaut. J. 20 (2025).
18. H.T. Gorji, N. Wilson, J. VanBree, B. Hoffmann, T. Petros, K. Tavakolian, Using machine learning methods and EEG to discriminate aircraft pilot cognitive workload during flight, Sci. Rep. 13 (2023)
19. L. Salvan, T.S. Paul, A. Marois, Dry EEG-based Mental Workload Prediction for Aviation, IEEE/AIAA 42nd Digital Avionics Systems Conference (DASC), IEEE, Barcelona, Spain, 2023. C.Y. Yiu et al. Advanced Engineering Informatics 71 (2026) 104378 17
20. D.-H. Lee, S.-J. Kim, S.-H. Kim, Decoding EEG-based Workload Levels Using Spatio-temporal Features Under Flight Environment, 12th International Winter Conference on Brain-Computer Interface (BCI), IEEE, 2024.
21. Y. Zhou, J. Jiang, L. Wang, S. Liang, H. Liu, Enhanced cognitive load detection in air traffic control operators using EEG and a hybrid deep learning approach, IEEE Access 13 (2025) 12127–12137.
22. Y. Wang, M. Han, Y. Peng, R. Zhao, D. Fan, X. Meng, H. Xu, H. Niu, J. Cheng, T. Liu, LGNet: Learning local-global EEG representations for cognitive workload classification in simulated flights, Biomed. Signal Process. Control 92 (2024) 13.
23. J.A. Blanco, M.K. Johnson, K.J. Jaquess, H. Oh, L.-C. Lo, R.J. Gentili, B. D. Hatfield, Quantifying cognitive workload in simulated flight using passive, dry EEG measurements, IEEE Trans. Cogn. Dev. Syst. 10 (2018) 373–383.
24. A. Hernandez-Sabate, J. Yauri, P. Folch, M.A. Piera, D. Gil, Recognition of the mental workloads of pilots in the cockpit using EEG signals, Appl. Sci. 12 (2022) 14.
25. C. Liu, C. Zhang, L. Sun, K. Liu, H. Liu, W. Zhu, C. Jiang, Detection of pilot's mental workload using a wireless EEG headset in airfield traffic pattern tasks, Entropy 25 (2023) 24.
26. G. Jiang, H. Chen, C. Wang, P. Xue, Mental workload artificial intelligence assessment of pilots' eeg based on multi-dimensional data fusion and LSTM with attention mechanism model, Int. J. Pattern Recognit Artif Intell. 36 (2022) 19.
27. C. Zhang, S. Luo, S. Cao, Y. Zhang, H. Chen, C. Jiang, Y. Zhou, Evaluating pilot mental workload using fNIRS-based functional connectivity features with a deep residual shrinkage network under emergency flight scenarios, Int. J. Hum.-Comput. Interact. (2024) 16.
28. C. Zhang, C. Jiang, Y. Xie, S. Cao, J. Yuan, C. Liu, W. Cao, Y. Li, Assessing pilot workload during takeoff and climb under different weather conditions: a fNIRS based modeling using deep learning algorithms, IEEE Trans. Aerosp. Electron. Syst. 61 (2025) 1705–1724.
29. Z. Jiang, K. Zhang, K. Wu, J. Xu, X. Li, Y. Sun, X. Ge, M. Mao, Mental workload recognition using ECG and machine learning in simulated flight tasks, 6th IEEE Advanced Information Technology, Electronic and Automation Control Conference (IEEE IAEAC), IEEE, Beijing, China, 2022, pp. 1560–1565.
30. Y. Wang, C. Zhang, C. Liu, K. Liu, F. Xu, J. Yuan, C. Jiang, C. Liu, W. Cao, Analysis on pulse rate variability for pilot workload assessment based on wearable sensor, Hum. Factors Ergonom. Manuf. Serv. Ind. 34 (2024) 635–648.

31. P. Xi, A. Law, R. Goubran, C. Shu, Pilot Workload Prediction from ECG Using Deep Convolutional Neural Networks, IEEE International Symposium on Medical Measurements and Applications (IEEE MeMeA), IEEE, Istanbul, Turkey, 2019.
32. X. Yu, C.-H. Chen, H. Yang, Cognitive workload quantification for air traffic controllers: an ensemble semi-supervised learning approach, Adv. Eng. Inform. 64 (2025) 10.
33. S.G. Hajra, P. Xi, A. Law, A comparison of ECG and EEG metrics for in-flight monitoring of helicopter pilot workload, IEEE International Conference on Systems, Man, and Cybernetics (SMC), IEEE (2020) 4012–4019.
34. B. Liu, S.W. Lye, K.X. Yeo, C.-H. Chen, A human-centric model for task demand assessment based on unsupervised learning-assisted eye movement measure, Adv. Eng. Inform. 65 (2025) 18.
35. Y. Gao, L. Yue, J. Sun, X. Shan, Y. Liu, X. Wu, WorkloadGPT: a large language model approach to real-time detection of pilot workload, Appl. Sci. 14 (2024) 28.
36. N. Liang, J. Yang, D. Yu, K.O. Prakah-Asante, R. Curry, M. Blommer, R. Swaminathan, B.J. Pitts, Using eye-tracking to investigate the effects of pre takeover visual engagement on situation awareness during automated driving, Accid. Anal. Prev. 157 (2021) 106143.
37. M.R. Endsley, Toward a theory of situation awareness in dynamic systems, Hum. Factors 37 (1995) 32–64.
38. C. Feng, S. Liu, X. Wanyan, Y. Dang, Z. Wang, C. Qian, β-wave-based exploration of sensitive EEG features and classification of situation awareness, Aeronaut. J. 128 (2024) 2561–2576.
39. C.Y. Yiu, K.K.H. Ng, X. Li, X. Zhang, Q. Li, H.S. Lam, M.H. Chong, Towards safe and collaborative aerodrome operations: assessing shared situational awareness for adverse weather detection with EEG-enabled Bayesian neural networks, Adv. Eng. Inform. 53 (2022) 19.
40. G.J.W. Xu, S. Pan, P.Z.H. Sun, K. Guo, S.H. Park, F. Yan, M. Gao, X. Wanyan, H. Cheng, E.Q. Wu, Human-factors-in-aviation-loop: multimodal deep learning for pilot situation awareness analysis using gaze position and flight control data, IEEE Trans. Intell. Transp. Syst. 26 (2025) 8065–8077.
41. C. Qian, S. Liu, X. Wanyan, C. Feng, Z. Li, W. Sun, Y. Wang, Situation awareness discrimination based on physiological features for high-stress flight tasks, Aerospace 11 (2024) 21.
42. Q. Li, K.K.H. Ng, S.C.M. Yu, C.Y. Yiu, M. Lyu, Recognising situation awareness associated with different workloads using EEG and eye-tracking features in air traffic control tasks, Knowledge-Based Syst. 260 (2023) 16.
43. V. Celina, K. Samardzic, I. Tukaric, T. Radisic, R.H. Hermann, Gaze Analysis of Air Traffic Contoller Using AI-Based Conflict Detection, 43rd AIAA DATC/IEEE Digital Avionics Systems Conference, IEEE, San Diego, CA, 2024.
44. E.Q. Wu, X.Y. Peng, C.Z. Zhang, J.X. Lin, R.S.F. Sheng, Pilots' fatigue status recognition using deep contractive autoencoder network, IEEE Trans. Instrum. Meas. 68 (2019) 3907–3919.
45. D.-H. Lee, S.-J. Kim, S.-H. Kim, Decoding Fatigue Levels of Pilots Using EEG Signals with Hybrid Deep Neural Networks, 13th International Conference on Brain-Computer Interface (BCI), IEEE, 2025.
46. Factors Ergonom. Manuf. Serv. Ind. 31 (2021) 637–651.
47. I. Alreshidi, D. Bisandu, I. Moulitsas, Illuminating the neural landscape of pilot mental states: a convolutional neural network approach with shapley additive explanations interpretability, Sensors 23 (2023) 20.

RESEARCH JOURNEY International Multidisciplinary E-Research Journal
Impact Factor (SJIF) - 6.625 | ***Special Issue 389 :*** *Advanced Computing and Artificial Intelligence (ACAI)*
ISSN : 2348-7143
March 2026

Technology and privacy: An inevitable trade-off

Vaidehi Torane
Email: vaidehit2006@gmail.com

Abstract:

Technology has become an essential part of modern life, influencing communication, education, business, entertainment, and everyday activities. Life has become easier and efficient because of the growing utilization of smartphones, social media, and online platforms. Nevertheless, the development of digital technologies has also become a major concern in terms of personal privacy. Online activities can produce a lot of personal information that can be gathered, stored, and utilized by organizations and governments. The purpose of the study is to analyze the connection between the use of technology and privacy issues, awareness of the users of the online data collection, and the phenomenon of the privacy paradox.

The study used a quantitative survey. The collected data was done using a structured questionnaire using Google Forms and randomly distributed the questionnaire to about 150-170 people in the general population. Among them, some 100-120 respondents had filled the survey. The questionnaire was based on the use of technology, knowledge of personal data collection, the privacy issue, behavior of sharing data online, and the perception of laws regarding data protection. Percentages and statistical tests like Chi-square test were used to analyze the data collected in order to determine patterns and relationships.

The findings show that a significant percentage of participants often use digital technologies and know that their online behavior leads to the creation of personal data. Even with this knowledge, a number of respondents said they actually shared personal information online, which corresponds to the notion of privacy paradox, as people are concerned about privacy but still share personal data.

In summary, the results indicate that although users understand privacy threats, convenience and gains of digital platforms tend to shape their actions. To balance the technological development and privacy safeguards, the study suggests the need to have stronger data protection laws, enhanced digital literacy, and intentional data practices.

Keywords:

Technology Usage, Privacy Awareness, Privacy Concerns, Privacy Paradox, Online Data Sharing, Data Protection Law.

Introduction:

We are currently in a period where we have technology almost everywhere in our lives. People depend on digital tools and technology, whether it comes to work and study or even entertainment, shopping, business, communication, or even sharing of information. Tasks and services that were often time-consuming could be done now with a few clicks due to the assistance of smartphones, social media, online services and artificial intelligence. Overall, technology has simplified our lives and made them more productive.

However, the progress has raised a serious issue of privacy. Whenever we access our digital devices digitally, they produce personal information, which can be gathered, retained and utilized by firms and governments alike. Social networks, smart technologies such as Alexa and biometric technologies such as Aadhaar or face-recognizing devices are collecting personal information, which they may or may not realize. This problem is significant since nowadays, personal information is regarded as one of the most valuable assets. Mishandling of this information may cause issues such as identity theft, surveillance and loss of individual freedom. Technology allows us to be faster, more innovative, and comfortable, yet at the same time, it asks us to give up our personal information. This leads to a "trade-off", where users may feel forced to give up some privacy for the benefits of technology. People claim to be concerned about privacy, but they still share information online - it is called the privacy paradox. In response to these issues, numerous data protection regulations have been enacted (GDPR, CCPA and the India DPDP Act (2023) among others), yet their efficacy has been doubted.

In this context, the present study examines whether privacy loss is an unavoidable trade-off for the benefits of modern technology. It focuses on users' awareness of data collection, their behavior in sharing personal information online, and the concept of the privacy paradox. The study also considers the role of data protection laws in protecting personal information and explores whether a balance between technological advancement and privacy protection can be achieved.

Objective:

- To examine the correlation between privacy concerns and the use of technology during the digital age.

- To determine awareness of users regarding online data collection and use of personal data.
- To examine the online sharing behavior relative to privacy issues (Privacy Paradox).
- The aim and objective of the research is to determine how effective the data protection laws are in defending against user privacy.

In order to determine whether technological advantages and individual privacy can be balanced in the current society.

Literature Review:

Technology and Its Role in Modern Life:

Technology has become inextricable from the human life in terms of communication, education, healthcare, governance and entertainment. Castells (2010) in his book The Rise of the Network Society explains that digital technologies have led to a "network society" in which social, economic and cultural activities are organized on the basis of networks powered by information and communication technologies. Similarly, Brynjolfsson and McAfee (2014), in The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies, are convinced that the technological innovations-though not necessarily one or the other, but all together, that is, artificial intelligence, big data, automation-are rendering societies more productive, efficient, and interconnected. This body of research reveals that people are extremely dependent on technology in their everyday lives, making the study of privacy in this context necessary and urgent.

1. Privacy Concerns in the Digital Era:

The development of technology has also brought about great concern about personal privacy. Solove (2006) in his article A Taxonomy of Privacy explains what privacy is 'the ability to control personal information.' However, with the advent of digital platforms, this control has been on the back burner most of the time. Zuboff (2019), in The Age of Surveillance Capitalism, raises the issue of the enormous data that the corporations like Google and Facebook gather from users to make a profit and sometimes without the full knowledge or consent of the user. Earlier, Warren and Brandeis (1890), in their landmark essay The Right to Privacy, had been among the first to argue that privacy was a fundamental right paving way for the modern debates on the subject matter. Collectively, this literature establishes privacy has moved from a personal right to a large-scale public concern in the digital age.

2.The Trade-off Between Technology and Privacy:

A core concern in all the privacy research is about whether one needs to sacrifice some privacy for some technological perks. In Consumer Privacy: Balancing Economic and Justice Considerations (2003), Culnan and Bies propose what they call the "trade-off model," in which users receive their personal information in return for services ranging from personalization to targeted advertising to convenience. Expanding on this idea, Acquisti, Brandimarte and Loewenstein (2015), in Privacy and Human Behavior in the Age of Information, explain how people, although knowing the risks to privacy, often agree to the risks because of the immediate benefits, such as the use of free applications or personalized recommendations. Together, these studies are evidence of people's willingness to deliberately or unknowingly sacrifice privacy as a result of perceived benefits of technology.

3. The Privacy Paradox:

One of the most popular concepts in studies on privacy is the privacy paradox, which addresses the contradiction between people's expressed concerns regarding privacy and people's actual behavior. Barnes (2006), in A Privacy Paradox: Social Networking in the United States, found that teenagers were worried about their privacy, but still provided personal details on social media. Norberg, Horne, Horne (2007) in their work entitled The Privacy Paradox: Personal Information Disclosure Intentions versus Behaviors confirmed that adults exhibited the same contradiction. In Privacy and Rationality in Individual Decision-Making, Acquisti and Grossklags (2008), found that people often underestimate long-term privacy risks in favor of short-term convenience. This paradox is complicating privacy research because what people say about privacy is often not what they do in practice.

4. Social Media, Smart Devices, and Biometric Systems:

Digital platforms and new technologies such as smart devices and biometrics have been at the center of the debate about privacy. Tufekci (2015), in Algorithmic Harms Beyond Facebook and Google: Emergent Challenges of Computational Agency, points out how social media platforms monetize the data users by using targeted advertising and algorithmic profiling. Similarly, Lau, Zimmerman, and Schaub (2018), in Alexa, Are You Listening? Privacy Perceptions, Concerns and Privacy-Seeking Behaviors with Smart Speakers, show that devices such as Amazon Alexa and Google Home have generated fears of "always listening" surveillance in households. In the Indian setting, Ramanathan (2019) in Aadhaar and Its Discontents: Politics of India's Biometric Identity Scheme, has examined how the Aadhaar system helped to maintain greater governance efficiency in India, while at the same time nevertheless exposing its citizens to potential risks of being surveilled and exposed to large-scale data leaks. Taken together, these studies highlight how social media, smart devices and biometric systems, while enhancing

convenience and efficiency, open the door to new vulnerabilities in the integrity of personal privacy.

5. Legal and Regulatory Frameworks:

Legal systems have been established to counter the increasing privacy issues in different governments of the world. The European Union's General Data Protection Regulation (GDPR) of 2016 is dedicated to the rights in the transparency, user rights and consent in processing personal data. Similarly, American consumers have a right to learn, delete and opt out of selling their data under the California Consumer Privacy Act (CCPA), which was enacted in 2018. The Digital Personal Data Protection (DPDP) Act 2023 in India is an effort to regulate the processing of personal data and also make organizations accountable. But in Global Data Privacy Laws 2020: 132 National Laws and Many Bills, Greenleaf (2020) asserts that on paper, such laws exist but when it comes to enforcing them, companies are not always strongly enforced, and legal loopholes mean that corporations can avoid being held accountable. Therefore, as the number of regulatory frameworks increases around the world, their efficacy is doubted.

6. Balancing Technology and Privacy:

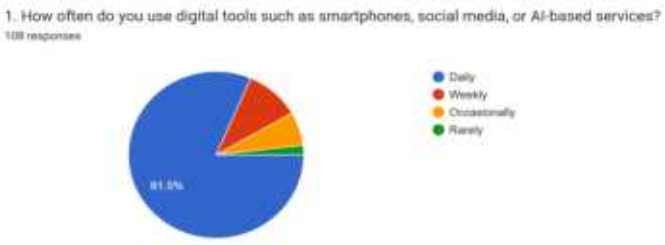
Lastly, scholars have suggested various mechanisms of ensuring the technological progress is balanced with privacy. A privacy architect like Cavoukian (2010) in Privacy by Design: The 7 Foundational Principles would champion the concept of ensuring privacy protection is part of the design of technology and organizational practice, and not an add-on concern. Westin (1967) in his bestseller book Privacy and Freedom has highlighted the value of user awareness and control on personal data as a platform of defending individual privacy. In more recent scholarship, there is a tendency to refer to the possibility of decentralized technology like blockchain and Web3 that can provide users with more ownership and control in relation to their personal data. Taken together, the opinions suggest that the cost of losing privacy is not that inevitable. Instead, there must be a compromise between the stronger legislation, ethical business practices and digital literacy to ensure that technological invention and privacy can be in better balance.

Research Methodology :

Through a quantitative survey, we were able to sense how people perceive the privacy in the hyper-connected digital-era. The primary objective was to understand the rate, to which people use tech, their awareness related to the risks of privacy, and whether they believe that tech and privacy can co-exist.

Data Collection: The data for this study was collected using a structured questionnaire created through Google Forms. The form was distributed randomly to around 150–170 individuals from the general public. Out of these, approximately 100–120 respondents completed the survey. The questions focused on technology usage, privacy awareness, and knowledge of privacy laws.

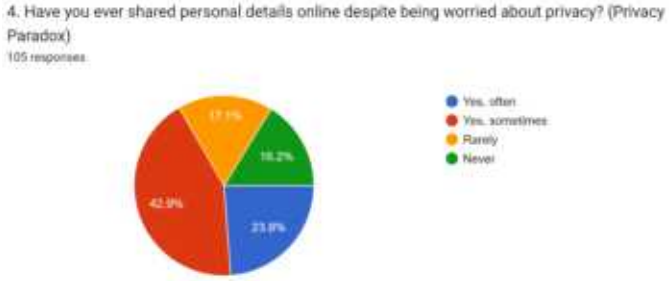
Key Questions and Result :



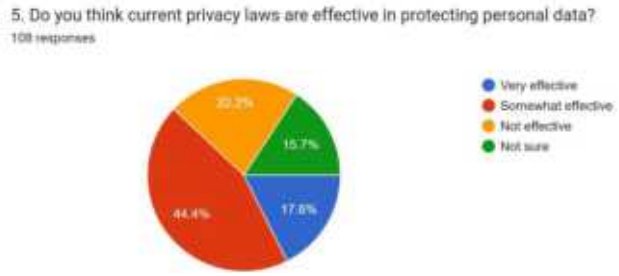
Responses	Frequency	% of Total
Daily	88	81.5
Weekly	11	10.2
Occasionally	7	6.5
Rarely	2	1.9

Responses	Frequency	% of Total
Yes, fully aware	70	64.8
Somewhat aware	31	28.7
Not Aware	7	6.5

Responses	Frequency	% of Total
Very concerned	63	58.3
Somewhat concerned	30	27.8
Neutral	13	12
Not Concerned	2	1.9



Responses	Frequency	% of Total
Yes, often	25	23.8
Yes, sometimes	45	42.9
Rarely	18	17.1
Never	17	16.2



Responses	Frequency	% of Total
Very effective	19	17.6
Somewhat effective	48	44.4
Not effective	24	22.2
Not sure	17	15.7

Data Analysis: Percentages and graphs were used to analyze the responses that were collected to know about general trends and patterns. The criterion of the analysis was the relation between the technology habits of people and their privacy concerns and awareness.

Ethical Considerations: All of the answers were confidential and anonymous. The study objective was explained to the participants who then provided their consent to participate.

Hypothesis:

1. Are you aware that most online activities generate personal data that can be collected and stored?
2. Have you ever shared personal details online despite being worried about privacy? (Privacy Paradox)

Hypothesis 1:

Hypothesis Statement:

Null Hypothesis (H_0): There is no significant difference in respondents' awareness that most online activities generate personal data that can be collected and stored.

Alternative Hypothesis (H_1): There is a significant difference in respondents' awareness that most online activities generate personal data that can be collected and stored.

Step 1: Calculate Expected Frequencies (E)

If awareness levels were equally distributed, each of the three response categories would be expected to have an equal number of responses.

$$E = \frac{\text{Total Responses}}{\text{Number of Categories}} = \frac{108}{3} = 36$$

So, the expected frequency for each category = 36.

Step 2: Apply Chi-Square Formula

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

Now, calculate for each category:

Responses	O	E	O-E	(O-E)2/E
Yes, fully aware	70	36	34	$(34^2)/36 = 1156/36 = 32.11$
Somewhat aware	31	36	-5	$(-5^2)/36 = 25/36 = 0.69$
Not aware	7	36	-29	$(-29^2)/36 = 841/36 = 23.36$
Total χ^2				56.16

Step 3: Degrees of Freedom

df = (k - 1) = (3 - 1) = 2

Where k = number of categories.

Step 4: Find Critical Value

At a 5% significance level ($\alpha = 0.05$) and df = 2, the critical value of $\chi^2 = 5.991$

Step 5: Decision Rule

Compare calculated and critical Chi-square values:

χ_calculated^2=56.16>χ_critical^2=5.991

Decision: Reject the Null Hypothesis (H_0)

Step 6: Interpretation

The Chi-square test shows that there is a statistically significant difference in the level of awareness of the respondents. Most of them (64.8%), were fully aware that online activities produce personal data, with 28.7% being partly aware and only 6.5% not aware. This indicates that the overall digital awareness is rather high, but a low percentage of users do not have a comprehensive level of digital literacy, which is why further digital literacy

campaigns are necessary.

Hypothesis 2:

Hypothesis Statement:

Null Hypothesis (H0): Respondents do not significantly differ in their behavior of giving personal information online even when they are concerned about their privacy.

Alternative Hypothesis (H1): The behavior of posting personal details online by the respondents is significantly different when they are concerned about their privacy.

Step 1: Calculate Expected Frequencies (E):

Since there are 4 response categories, if responses were equally distributed:

$$E = \frac{\text{Total Responses}}{\text{Number of Categories}} = \frac{105}{4} = 26.25$$

So, the expected frequency for each category = 26.25

Step 2: Apply Chi-Square Formula

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

Now, calculate for each category:

Response	O	E	(O-E)	$(O-E)^2/E$
Yes, often	25	26.25	-1.25	(1.56)/26.25 = 0.06
Yes, sometimes	45	26.25	18.75	(351.56)/26.25 = 13.40
Rarely	18	26.25	-8.25	(68.06)/26.25 = 2.59
Never	17	26.25	-9.25	(85.56)/26.25 = 3.26
Total χ^2				19.31

Step 3: Degrees of Freedom

df = (k - 1) = (4 - 1) = 3

Step 4: Critical Value

At 5% significance level (α = 0.05) and df = 3:

$\chi 2$ critical = 7.815

Step 5: Decision Rule

X^2 calculated = 19.31 > χ^2 critical = 7.815

Decision:

Since the calculated value is greater than the critical value:

Reject the Null Hypothesis (H_0)

Step 6: Interpretation

The Chi-square test shows a significant difference in sharing behavior despite privacy concerns. A significant number of respondents said that they occasionally (42.9%), or frequently (23.8%), disclose personal information even when they are concerned about privacy. This is a display of the privacy paradox whereby people worry about privacy yet they post personal data somewhere on the Internet.

Discussion:

This research findings justify most of the arguments that have been made in the previous studies on technology and privacy. Previous scholars such as Manuel Castells and Erik Brynjolfsson highlight that modern societies have become highly dependent on digital technologies for communication, work, and everyday activities. This survey data also indicates the same trend, with a significant proportion of the participants (81.5) indicating that they use digital devices like smartphones, social media, and AI-powered services every day. This confirms the fact that technology has been used more in everyday life.

But the growing use of technology has brought privacy concerns too. Researchers such as Daniel J. Solove and Shoshana Zuboff believe that online platforms gather massive personal information about users. This view is supported by the results of this research. The chi-square analysis indicated that the majority of users (64.8) were completely familiar with the fact that their online activities create personal data and a very small percentage were not. It means that the level of digital awareness among respondents is relatively high.

Although there is such awareness, the study also shows that there is behavioral consistency with the concept of privacy paradox as it is explained by Susan Barnes and Alessandro Acquisti. The findings indicate that a large number of participants continue to post personal data online despite being worried about privacy. In the survey, 42.9% said they occasionally shared personal information online and 23.8% said they did it frequently. Chi-square test showed that such a behavior is statistically significant.

On the whole, the findings correspond to the literature that indicates that people are becoming much more conscious of privacy risks, but their online activities are frequently about convenience and the utility of online services. This confirms the argument that technology and privacy are interdependent, and such a balance is still one of the greatest problems in the digital age.

Conclusion:

This study explored whether privacy loss is an unavoidable trade-off for using modern technology. The results indicate that the majority of the respondents use online platforms actively and are very conscious that their online activities are producing personal data. The chi-square

test demonstrated that the level of awareness is significant.

Nevertheless, even with this knowledge and understanding of privacy, a number of respondents agreed that they also share personal information over the internet. This contributes to the idea of the privacy paradox - people are concerned about privacy but still give up information because doing so is convenient, sociable, or has other advantages provided by digital platforms.

Though the participants are aware of the significance of the laws on data protection, their views on its effectiveness differ. Overall, the study concludes that privacy sacrifice is not completely unavoidable, but maintaining a balance between technology and privacy requires stronger laws, better digital literacy, and responsible data practices.

References :

1. Alessandro Acquisti, Laura Brandimarte, & George Loewenstein. (2015). Privacy and human behavior in the age of information. Science, 347(6221), 509–514. https://doi.org/10.1126/science.aaa1465
2. Alessandro Acquisti, & Jens Grossklags. (2008). Privacy and rationality in individual decision-making. IEEE Security & Privacy, 6(1), 26–33.
3. Susan B. Barnes. (2006). A privacy paradox: Social networking in the United States. First Monday, 11(9).
4. Erik Brynjolfsson, & Andrew McAfee. (2014). The second machine age: Work, progress, and prosperity in a time of brilliant technologies. W. W. Norton & Company.
5. Ann Cavoukian. (2010). Privacy by design: The 7 foundational principles. Information and Privacy Commissioner of Ontario.
6. Mary J. Culnan, & Robert J. Bies. (2003). Consumer privacy: Balancing economic and justice considerations. Journal of Social Issues, 59(2), 323–342.
7. Graham Greenleaf. (2020). Global data privacy laws 2020: 132 national laws and many bills. Privacy Laws & Business International Report, 163, 14–18.
8. Joseph Turow Lau, Benjamin Zimmerman, & Florian Schaub. (2018). Alexa, are you listening? Privacy perceptions, concerns and privacy-seeking behaviors with smart speakers. Proceedings of the ACM on Human-Computer Interaction.
9. Manuel Castells. (2010). The rise of the network society (2nd ed.). Wiley-Blackwell.
10. Daniel J. Solove. (2006). A taxonomy of privacy. University of Pennsylvania Law Review, 154(3), 477–564.
11. Shoshana Zuboff. (2019). The age of surveillance capitalism: The fight for a human future at the new frontier of power. PublicAffairs.
12. Samuel D. Warren, & Louis D. Brandeis. (1890). The right to privacy. Harvard Law Review, 4(5), 193–220.
13. Alan F. Westin. (1967). Privacy and freedom. Atheneum.
14. Zeynep Tufekci. (2015). Algorithmic harms beyond Facebook and Google: Emergent challenges of computational agency. Colorado Technology Law Journal, 13, 203–218.
15. Usha Ramanathan. (2019). Aadhaar and its discontents: Politics of India's biometric identity scheme.

RESEARCH JOURNEY International Multidisciplinary E-Research Journal ISSN : 2348-7143
Impact Factor (SJIF) - 6.625 | ***Special Issue 389 :*** *Advanced Computing and Artificial Intelligence (ACAI)* March 2026

AI-Driven Systems for Relief Distribution in Disaster Management in India

Mrs. Jyotsna Dhanraj Mali
Assistant professor, R. C. Patel Educational Trust's Institute of Management Research and Development, Shirpur
Dr. Priyanka V. Bhandari
Assistant professor, R. C. Patel Educational Trust's Institute of Management Research and Development, Shirpur

Abstract:

India frequently experiences natural and man-made disasters due to its diverse topography, climate, fast urbanization, environmental damage, and social and economic problems. National preparedness and coordination have improved through the disaster management system established under the Disaster Management Act, 2005 and operated by the National Disaster Management Authority (NDMA). Distribution of aid is still primarily reactive, manual, and ill-coordinated, nevertheless. Emergency responses are still hampered by ineffective resource usage, overlapping activities, a lack of openness, poor data exchange across agencies and delays in assessing needs.

Emerging technologies such as machine learning (ML) and artificial intelligence (AI) present excellent opportunities to enhance disaster relief systems in India. Real-time monitoring, efficient logistical planning, prompt damage assessment, disaster prediction, and improved decision-making are all made possible by AI and ML models.

This study examines the shortcomings of India's current aid distribution system and offers a framework powered by AI and ML to increase accountability, efficiency, and transparency. It comes to the conclusion that smart technology can improve resilience, speed, and fairness, resulting in a shift from a reactive to a proactive and adaptable strategy to disaster response.

Keywords: Artificial Intelligence, Machine Learning, Disaster Management, Relief Distribution, Predictive Disaster Governance

I. Introduction

India is counted a few of the international locations most easily suffering from calamities international due to its giant bodily range, climate adjustments, and societal monetary weaknesses. The location surrounding the Himalayas is especially in all likelihood to enjoy tremors and slides, whilst the big coastline is threatened with the aid of strong storms and big sea waves. Huge river networks are frequently inundated via excessive water, and the arid sections regularly deal with durations of little rain. The problems supplied through swift metropolis enlargement, harm to the environment, and growing numbers of people further boom the dangers connected to disasters and extend their social and economic outcomes.

According to the national catastrophe control Authority (NDMA) Annual record for 2021–22, handling failures in India is carried out the use of a properly-organized institutional machine with many layers that includes federal, nation, region, and network groups. This structure changed into installed underneath the catastrophe management Act of 2005 to ensure a hit planning, chance reduction, response, and rebuilding across unique governance ranges.

The country wide disaster management Authority (NDMA) capabilities as the primary frame placing coverage that is aided by:

- National Executive Committee (NEC)
- State Disaster Management Authorities (SDMAs)
- District Disaster Management Authorities (DDMAs)
- National Disaster Response Force (NDRF)

This layered association has in reality improved the ability of companies to react and has promoted advanced teamwork amongst exceptional groups at some point of foremost emergencies. However, even with these current arrangements, the procedures for handing out useful resource frequently encounter operational hurdles that hinder how well they work, especially when immediate reaction is needed.

Considering the growing rate and intensity of catastrophes related to weather exchange, preferred responses are proving much less enough. There's an urgent need to integrate artificial Intelligence (AI)-powered gear into aid transport frameworks for higher pace, precision, scope, and obligation. AI-backed forecasting strategies, immediately oversight, mechanical supply chain streamlining, and flexible training structures show extensive promise to move India's calamity managing from a behind-the-fact management structure in the direction of a ahead-searching, evidence-based totally, and robust approach.

II. Literature Review

A. Artificial Intelligence in disaster management

Synthetic intelligence has moved from simply experiments to being primary in catastrophe chance discount. All over the global, you spot AI in movement in the course of mitigation, preparedness, reaction, and

healing. Device getting to know just handles complicated, excessive-dimensional danger modeling higher than antique-college regression techniques [1], [2]. It alternatives out patterns from past threat facts and real-time sensor feeds, slicing down on uncertainty. The Intergovernmental Panel on weather exchange points out how critical it is to have forward-searching governance, powered via computational modeling and predictive analytics [3]. AI structures now combination together risk exposure, vulnerability, and how well groups can cope, making multi-risk modeling viable.

B. Disaster Forecasting and Early warning systems

Predictive modeling stands out as certainly one of AI's most evolved makes use of in catastrophe management. Deep studying fashions like long short-term reminiscence (LSTM) networks beat ARIMA and traditional hydrological models in terms of flood prediction [4]. In India, the Meteorological department has made massive strides in cyclone forecasting, but alleviation logistics still doesn't take complete advantage of AI for impact forecasting and demand prediction. More recent hybrid systems that blend LSTM, GRU, and interest mechanisms are pushing flood prediction even similarly, especially for top events and multi-step forecasts.

C. AI-based totally harm evaluation the usage of far flung Sensing

Guide surveys take too long, slowing down each the identification of beneficiaries and the transport of financial assistance to the proper human beings. AI-powered structures can examine satellite TV for pc and drone imagery to categories building harm, decreasing evaluation time from days to simply hours [5], [6]. These tools also make bigger coverage throughout larger geographical regions. With geospatial AI, computerized change detection can be done via comparing earlier than-and-after pictures. Advanced vision models and semantic segmentation techniques now permit special harm mapping, making sure that resources attain the areas where they're wanted maximum, extra quickly and effectively.

D.AI in humanitarian logistics and optimization

Humanitarian logistics isn't clean—it's approximately figuring out in which to installation centers, how to allocate stock, and a way to get components added, all whilst matters are converting rapid. Optimization and reinforcement mastering fashions can adjust routes at the fly as roads near or demand shifts [7], [8]. Deep Reinforcement learning blended with Graph Neural Networks makes dynamic vehicle routing viable, in spite of masses of variables. The end result? Quicker delivery and fewer wasted assets.

E. Social media analytics and disaster informatics

Social media churns out a flood of real-time information at some stage in disasters. Herbal Language Processing (NLP) can sift thru posts, flag distress messages, spot pressing wishes, and pull out location information [9], [10]. Smart alert-ranking systems maintain emergency groups focused by highlighting what subjects and filtering out noise or misinformation.

F. Governance, ethics, and explainability

AI in catastrophe relief isn't just about tech—it needs to be obvious and explainable. Human-AI collaboration works high-quality whilst models are understandable and there's a clean audit path [11]. It's also critical to put moral tests in region, to make certain comfort reaches all and sundry pretty and avoids bias in who gets help.

III. Objectives:

These study ambitions to:

1. Have a look at the operational structure of India's current alleviation distribution framework.
2. Examine international AI innovations relevant to catastrophe comfort.
3. Layout a context-suitable AI-enabled relief framework aligned with NDMA governance systems.

VI. Present Disaster Relief Distribution System

India built its disaster management device around the catastrophe management Act of 2005. in keeping with that, countrywide disaster control Authority (NDMA) running things national, every country has its personal SDMA, every district has a DDMA and while disaster moves, the national catastrophe reaction pressure (NDRF) and state catastrophe reaction Forces step in to help on the ground.

A. Operational workflow

1. **Early Warning and Alerts** – The India Meteorological Department (IMD) and other agencies send out alerts.
2. **Damage Assessment** – Local officials head out to survey and see what's actually happened.
3. **Reporting** – These reports get pulled together at the district and state levels.
4. **Funding** – Money from SDRF or NDRF is released for action.
5. **Resources Deployment**– The NDRF and other government departments get mobilized where they're needed.
6. **Identifying Beneficiary** – Officials manually register and check who needs help.
7. **Ground Level Distribution**– Relief camps and local centers open to hand out supplies to need people.

B. Key features of existing system:

According to NDMA,

1. There's a clear chain of command, from national down to local.
2. Funding is structured and legally backed.
3. Early warning systems are strong.
4. Local authorities actually implement the plans.

C. Systematic problems assessment of existing system

1. **Manual Damage Checks:** These surveys take a lot of time, so help arrives late.
2. **Majorly Reactive:** In major cases authorities came in action and start moving resources after disaster hits, not before.
3. **Scattered Information:** Weather, population, logistics, and finance data not at common platform.
4. **Static Logistics Routing:** lack in feasible real-time route optimization in disasters.
5. **Limited Transparency and Monitoring:** Lack of integrated dashboards for anomaly detection
6. **Deficient in Adaptive Learning:** Historical disaster data is not systematically used for improving their models or responses.

V. Proposed AI-Enabled Relief Distribution System

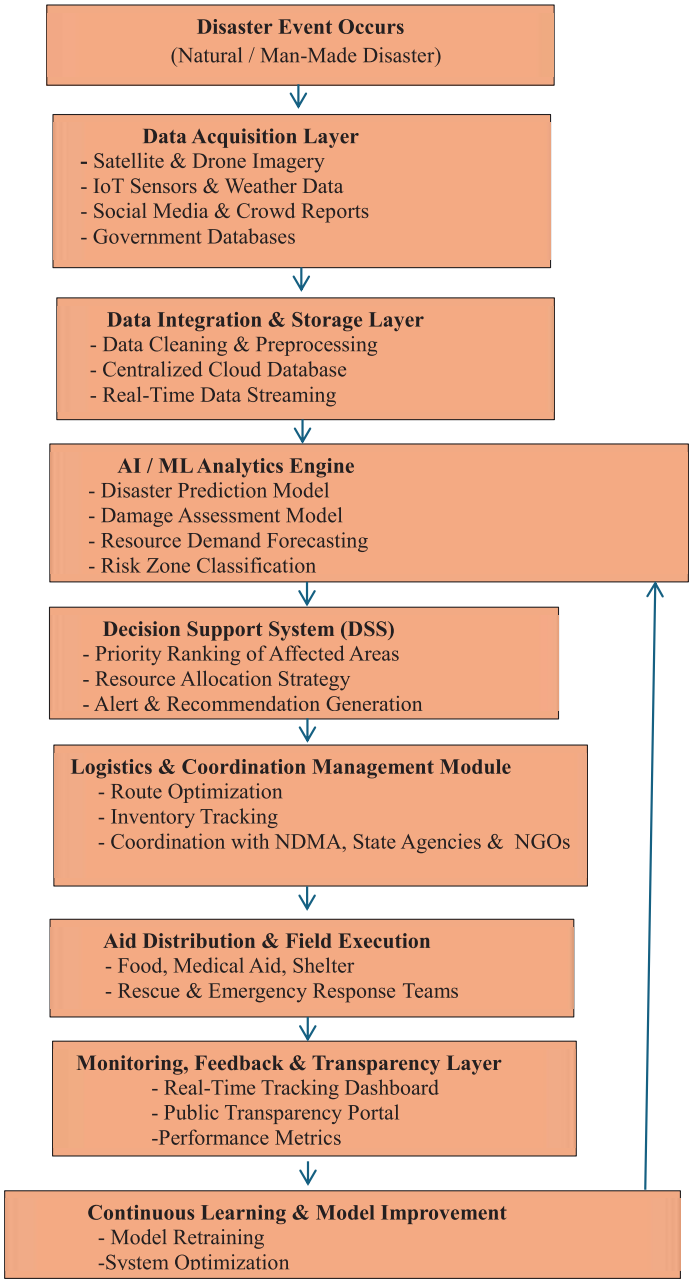


Fig. 1: Proposed AI & ML Based Disaster Relief Distribution Framework

VII. Conclusion

India's disaster risks maintain getting worse. Weather swings, towns growing too speedy, broken ecosystems, and a booming population all driving disaster challenges higher. The disaster management Act of 2005 and the country wide disaster control Authority have made coordination more potent and helped the respond faster. Nevertheless, getting help to humans at the floor is gradual and messy. Manual paperwork, scattered statistics, and the dependency of looking ahead to problem to strike before acting all get inside the way. This have a look at dug into how catastrophe relief is managed in India. With great lacunas such as late harm assessment gets, inefficient logistics plans, lack of shared information amongst agencies, and no real machine for adaptive gaining knowledge of. However here's the component: globally, artificial Intelligence has stepped up. It's now riding smarter disaster responses—predicting risks, mapping harm from area, making plans aid shipping as conditions exchange, and picking up real-time information from humans on the ground. So, the papers try and convey AI era into each step of India's catastrophe remedy, tied without delay into the manner disaster governance already works. The proposed system uses predictive models (like LSTM for forecasting), laptop vision (CNNs) to evaluate harm from satellite tv for pc pix, reinforcement studying to maintain logistics nimble, herbal language processing to map what humans want, and adaptive comments loops to hold the whole lot enhancing. With these types of working collectively, connected via shared statistics and clean-to-use dashboards, disaster management shifts from chasing after problems to staying one step in advance—making decisions based on proof.AI can boost relief speed, sharpen accuracy, make the whole thing more obvious, and build resilience into the machine with experienced human intervention. moving from scrambling after disasters to predicting and getting ready for them isn't pretty much new tech—it's a shift the united states of America needs if it desires to defend its human beings in an international wherein the weather isn't getting any responsive.

References :

1. Government of India, The Disaster Management Act, 2005. New Delhi, India: Ministry of Law and Justice, 2005.
2. National Disaster Management Authority (NDMA), Annual Report 2021–22. New Delhi, India: NDMA, 2022.
3. Intergovernmental Panel on Climate Change (IPCC), Climate Change 2022: Impacts, Adaptation and Vulnerability. Cambridge, U.K.: Cambridge University Press, 2022.
4. H. Zhang, Y. Cheng, and Q. Liu, "Long short-term memory networks for flood forecasting," IEEE Access, vol. 6, pp. 61232–61240, 2018.
5. S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.
6. Y. Xu, L. Ji, and X. Wang, "Post-disaster building damage assessment using convolutional neural networks and satellite imagery," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 12, no. 6, pp. 2221–2232, Jun. 2019.
7. F. Nex and F. Remondino, "UAV for 3D mapping applications in disaster management," IEEE Geoscience and Remote Sensing Magazine, vol. 7, no. 4, pp. 30–43, Dec. 2019.
8. B. Balcik and B. M. Beamon, "Facility location in humanitarian relief," International Journal of Logistics Research and Applications, vol. 11, no. 2, pp. 101–121, 2008.
9. J. Oroojlooyjadid, M. Nazari, L. Snyder, and M. Takác, "A deep Q-network for dynamic vehicle routing problems," in Proc. IEEE Int. Conf. Big Data, Los Angeles, CA, USA, 2019, pp. 3602–3611.
10. A. Imran, C. Castillo, F. Diaz, and S. Vieweg, "Processing social media messages in mass emergency: A survey," ACM Computing Surveys, vol. 47, no. 4, pp. 1–38, 2015.
11. D. Gunning and D. Aha, "DARPA's explainable artificial intelligence (XAI) program," AI Magazine, vol. 40, no. 2, pp. 44–58, 2019.
12. India Meteorological Department (IMD), Cyclone Warning Services: Operational Manual. New Delhi, India: Ministry of Earth Sciences, 2021.

RESEARCH JOURNEY International Multidisciplinary E-Research Journal
Impact Factor (SJIF) - 6.625 | ***Special Issue 389 :*** *Advanced Computing and Artificial Intelligence (ACAI)*
ISSN : 2348-7143
March 2026

Advanced Techniques for Verb Sense Disambiguation Using Pre-trained Language Models

Ms. Chhaya Patil
Assistant Professor, RCPET's Institute of Management Research and Development, Shirpur
Dr. Vaishali B. Patil
Director, RCPET's Institute of Management Research and Development, Shirpur

Abstract:

Verb Word Sense Disambiguation (WSD) remains a fundamental challenge in Natural Language Processing due to the high contextual variability and semantic ambiguity of verbs. This paper presents a comprehensive overview of advanced techniques for verb sense disambiguation leveraging pre-trained language models (PLMs). We examine zero-shot and few-shot prompting strategies that query large language models using gloss-based definitions and multiple-choice formulations, enabling sense prediction without extensive task-specific training. In addition, we explore supervised fine-tuning approaches that train models on context–gloss pairs to perform sense classification or ranking with improved contextual alignment.

The study further highlights specialized pre-training methods, such as SenseBERT, which incorporate explicit semantic knowledge during representation learning to enhance sense discrimination. Hybrid frameworks that integrate neural architectures with symbolic lexical resources are also discussed, demonstrating how structured knowledge can complement contextual embeddings. Moreover, we review emerging multimodal and cross-lingual strategies that exploit visual context and multilingual data to resolve verb ambiguities more effectively across diverse linguistic settings.

Beyond summarizing existing methodologies, this paper analyzes their comparative strengths, limitations, and scalability in real-world applications such as machine translation, information retrieval, and semantic search. We argue that combining knowledge-enhanced pre-training, prompt engineering, and cross-modal supervision represents a promising direction for building robust and generalizable verb sense disambiguation systems.

Keywords: Verb Word Sense Disambiguation, Pre-trained Language Models, Zero-shot Learning, Few-shot Learning, Fine-tuning, Context–Gloss Modeling, Knowledge-Enhanced Pre-training, SenseBERT, Hybrid Neural-Symbolic Methods, Multimodal Learning, Cross-lingual NLP.

I. Introduction

Word Sense Disambiguation (WSD) is a long-standing problem in Natural Language Processing (NLP) that involves determining the intended meaning of a word within a specific context [1], [2]. While nouns and adjectives have received considerable attention, verbs remain particularly difficult to model because they exhibit high semantic variability and dynamic contextual behavior.

Verbs often encode actions, processes, and abstract events. Unlike nouns, which frequently refer to concrete entities, verbs are highly dependent on syntactic structures, argument roles, and contextual cues. Lexical databases such as WordNet list dozens of fine-grained senses for common verbs such as "run," "take," and "set," making accurate classification more challenging [3].

Traditional approaches relied on knowledge-based graph traversal methods such as UKB [4]. With the introduction of contextual language models such as BERT [5], WSD has shifted toward neural architectures capable of modeling deep contextual semantics. However, empirical evidence consistently shows that verbs achieve lower F1 scores than nouns across benchmark datasets [6].

This paper provides a detailed survey of modern techniques specifically focusing on verb sense disambiguation and discusses their advantages, limitations, and future research directions.

II. Zero-Shot Prompting Approaches

Zero-shot prompting leverages large pre-trained language models without additional task-specific training [7]. These methods reformulate verb WSD as a natural language understanding task.

A. Definition Generation

In this approach, the model is asked to generate a short definition of the ambiguous verb in context. The generated output is then compared with dictionary glosses to determine the closest matching sense. This technique exploits the implicit semantic knowledge stored within large language models.

Although flexible and easy to implement, definition generation sometimes produces paraphrases that do not exactly match predefined sense inventories, leading to alignment challenges [8].

B. Multiple-Choice Gloss Selection

Another zero-shot strategy presents the model with candidate glosses and asks it to select the most appropriate one. This formulation aligns closely with benchmark evaluation setups.

Multiple-choice prompting tends to outperform

free-text definition generation because it restricts outputs to known sense inventories. However, performance for verbs remains moderate due to overlapping semantic categories [9].

C. Textual Entailment Reformulation

Some approaches reformulate verb WSD as a textual entailment task. The model evaluates whether a gloss description logically follows from the context sentence.

This method has shown improvements in distinguishing closely related senses, particularly when domain information is incorporated [10].

III. Few-Shot Prompting

Few-shot prompting provides a small number of labeled examples within the prompt to guide the model's reasoning process [11].

Chain-of-Thought prompting extends this by encouraging intermediate reasoning steps before selecting a sense [12]. This approach improves interpretability and helps resolve subtle distinctions between similar verb senses.

Despite improvements, few-shot prompting does not fully close the performance gap between verbs and nouns due to fine-grained ambiguity [6].

IV. Context – Gloss Fine-Tuning

Fine-tuning remains one of the most successful paradigms for verb WSD.

A. Sentence-Pair Classification

Models such as GlossBERT treat WSD as a sentence-pair classification task where the context sentence and candidate gloss are jointly encoded [13]. The model learns to predict whether the gloss correctly matches the contextual usage.

This method directly integrates lexical knowledge into the neural architecture and has achieved strong benchmark results.

B. Gloss Ranking Objective

Instead of binary classification, some approaches rank all candidate glosses according to contextual relevance [14]. Ranking better reflects the inherent structure of WSD, where multiple plausible senses must be compared.

C. Data Augmentation Techniques

Researchers have explored back-translation of glosses, inclusion of example sentences, and hypernym expansion to improve training diversity [15]. These techniques particularly benefit rare verb senses.

D. Performance Characteristics

Even with fine-tuning, verbs consistently achieve lower F1 scores compared to nouns. This gap is primarily due to:

- Fine-grained sense inventories
- High contextual overlap
- Sparse annotated examples

V. Knowledge-Enhanced Pre-Training

Knowledge-enhanced models incorporate semantic information directly during pre-training.

A. SenseBERT

SenseBERT extends masked language modeling by also predicting WordNet supersense categories [16]. By embedding semantic supervision into the training objective, the model learns representations that are more aware of lexical meaning distinctions.

This approach provides improved grounding for verbs, especially in cases where contextual signals alone are insufficient.

B. Advantages and Limitations

Knowledge-enhanced pre-training improves semantic coherence but requires additional lexical resources and training complexity.

VI. Multimodal Verb Sense Disambiguation

Certain verbs describe visually observable actions. Multimodal WSD incorporates both textual and visual features.

Datasets such as VerSe enable research in Visual Verb Sense Disambiguation [17]. Graph-based label propagation techniques allow models to learn from limited labeled data.

Vision-language models such as CLIP compute similarity between textual descriptions and images, grounding verb semantics in visual evidence [18].

Multimodal grounding is particularly effective for action verbs but less helpful for abstract verbs.

VII. Cross-Lingual Verb Wsd

Multilingual models such as XLM-R allow knowledge transfer across languages [19].

Cross-lingual WSD is especially useful for low-resource languages where annotated corpora are scarce. However, performance varies depending on language representation quality during pre-training [20].

VIII. Hybrid and Ensemble Methods

Hybrid approaches combine neural contextual embeddings with structured lexical knowledge such as WordNet hierarchies [21].

Inventory alignment methods unify multiple sense resources to improve rare sense modeling [22].

Ensemble methods combine feature extraction and fine-tuning outputs to improve robustness and generalization [6].

IX. Comparative Summary of Techniques :

Table 01: Comparative Summary Of Techniques

Category	Core Idea	Strengths	Limitations	Verb Performance
Zero-Shot	Prompt-based inference	No training required	Inventory mismatch	Moderate
Few-Shot	Example-guided prompting	Better reasoning	Limited improvement	Moderate
Fine-Tuning	Context–gloss classification	High accuracy	MFS bias	Strong
Ranking	Gloss relevance ordering	Natural formulation	Data dependent	Strong
Knowledge-Enhanced	Supersense supervision	Semantic grounding	Complex training	Improved
Multimodal	Text-image fusion	Visual grounding	Dataset specific	Strong for action verbs
Cross-Lingual	Shared multilingual embeddings	Low-resource support	Language imbalance	Moderate
Hybrid	Neural + symbolic integration	Rare sense coverage	Complexity	Strong
Ensemble	Model combination	Balanced robustness	Computational cost	Strong

X. Discussion

Across all approaches, verbs remain harder to disambiguate than nouns. Fine-tuning and hybrid models achieve the strongest performance, but ambiguity persists due to overlapping semantic fields and limited annotated resources.

XI. Conclusion

This paper presented a comprehensive survey of modern techniques for verb sense disambiguation using pre-trained language models. Prompt-based methods offer flexibility, fine-tuning provides strong supervised performance, knowledge-enhanced models improve semantic awareness, and multimodal approaches introduce grounding.

Despite these advances, verb disambiguation remains challenging. Hybrid systems that combine neural representations with structured lexical knowledge show the most promise for closing the performance gap.

XII. Future Scope

Future research should focus on:

- Verb-specific sense clustering
- Multimodal reasoning for abstract verbs
- Cross-lingual adaptation for low-resource settings
- Neuro-symbolic integration for interpretability
- Large-scale semi-supervised verb sense induction

References :

1. Navigli, R. (2009). Word sense disambiguation: A survey. ACM Computing Surveys, 41(2), 1–69. https://doi.org/10.1145/1459352.1459355
2. Ide, N., & Véronis, J. (1998). Word sense disambiguation: The state of the art. Computational Linguistics, 24(1), 1–40.
3. Miller, G. A. (1995). WordNet: A lexical database for English. Communications of the ACM, 38(11), 39–41. https://doi.org/10.1145/219717.219748
4. Agirre, E., & Soroa, A. (2009). Personalizing PageRank for word sense disambiguation. In Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics (EACL) (pp. 33–41). Association for Computational Linguistics.
5. Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT) (pp. 4171–4186).
6. Bevilacqua, M., & Navigli, R. (2020). Breaking through the 80% glass ceiling: Raising the state of the art in word sense disambiguation by incorporating knowledge graph information. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL) (pp. 2854–2864).
7. Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... Amodei, D. (2020). Language models are few-shot learners. In Advances in Neural Information Processing Systems (NeurIPS), 33, 1877–1901.
8. Sumanathilaka, H., Hough, J., & others. (2024). Assessing GPT's potential for word sense disambiguation. In Proceedings of the International Conference on Semantic Computing and Research (ICSGRC).
9. Chen, Y., et al. (2023). Zero-shot word sense disambiguation with large language models. arXiv preprint arXiv:2302.03353.
10. Loureiro, D., et al. (2022). What do language models know about word senses? Zero-shot WSD with language models and domain

inventories. arXiv preprint arXiv:2302.03353.

11. Schick, T., & Schütze, H. (2021). Few-shot learning with pattern-exploiting training. In Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP).
12. Wei, J., Wang, X., Schuurmans, D., Bosma, M., Ichter, B., Xia, F., ... Zhou, D. (2022). Chain-of-thought prompting elicits reasoning in large language models. In Advances in Neural Information Processing Systems (NeurIPS).
13. Huang, L., Ji, H., Cho, K., & others. (2019). GlossBERT: BERT for word sense disambiguation with gloss knowledge. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing (EMNLP).
14. Vial, L., Lecouteux, B., & Schwab, D. (2019). Sense vocabulary compression through the semantic knowledge of WordNet for neural word sense disambiguation. In Proceedings of the 10th Global WordNet Conference.
15. Pasini, T., & Navigli, R. (2017). Train-o-Matic: Large-scale supervised word sense disambiguation in multiple languages without manual training data. In Proceedings of the AAAI Conference on Artificial Intelligence.
16. Levine, Y., Lenz, B., Dagan, I., & Goldberg, Y. (2020). SenseBERT: Driving some sense into BERT. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL).
17. Chen, Y., et al. (2020). Transductive visual verb sense disambiguation. arXiv preprint arXiv:2012.10821.
18. Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., ... Sutskever, I. (2021). Learning transferable visual models from natural language supervision. In Proceedings of the 38th International Conference on Machine Learning (ICML).
19. Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., ... Stoyanov, V. (2020). Unsupervised cross-lingual representation learning at scale. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL).
20. Artetxe, M., Labaka, G., & Agirre, E. (2018). A robust self-learning method for fully unsupervised cross-lingual mappings of word embeddings. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (ACL).
21. Navigli, R., & Ponzetto, S. P. (2012). BabelNet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. Artificial Intelligence, 193, 217–250. https://doi.org/10.1016/j.artint.2012.07.001
22. Bevilacqua, M., Pasini, T., & Navigli, R. (2021). Connect-the-dots: Bridging semantics between words and definitions via aligning word sense inventories. In Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP).

RESEARCH JOURNEY International Multidisciplinary E-Research Journal
Impact Factor (SJIF) - 6.625 | ***Special Issue 389 :*** *Advanced Computing and Artificial Intelligence (ACAI)*
ISSN : 2348-7143
March 2026

A Flow-Based Hybrid CNN–LSTM Model for Detecting Spoofed and Unencrypted Communications in Multi-Protocol IoT Environments

Mr. Vishal Arun Pawar
Department of MCA, RCPET's Institute of Management Research and Development, Shirpur, India

Abstract:

The rapid proliferation of Internet of Things (IoT) devices has significantly expanded the attack surface of modern network infrastructures, particularly in heterogeneous multi-protocol environments. Protocol spoofing and unencrypted communications remain critical vulnerabilities, enabling adversaries to manipulate packet headers, impersonate legitimate devices, and compromise data confidentiality. Although recent machine learning and deep learning–based intrusion detection systems demonstrate high detection accuracy, they often fail to enforce encrypted communication validation and prevent spoofed packet transmission.

This paper proposes a flow-based hybrid CNN–LSTM framework for detecting spoofed and unencrypted communications in multi-protocol IoT environments. The proposed model performs flow-level feature extraction, header consistency verification, and encryption validation before classification. A Convolutional Neural Network (CNN) extracts spatial protocol features, while a Long Short-Term Memory (LSTM) network captures temporal dependencies across communication flows. Experimental evaluation using benchmark IoT datasets demonstrates improved spoof detection rates, reduced false positives, and enhanced robustness against unseen communication patterns. The proposed framework strengthens integrity, confidentiality, and reliability in dynamic IoT ecosystems.

Keywords: IoT Security, Multi-Protocol IoT Networks, Spoofing Detection, Unencrypted Communication Detection, CNN–LSTM Hybrid Model, Flow-Based Intrusion Detection, Deep Learning, Network Security.

1. Introduction:

Modern IoT ecosystems rely on a heterogeneous mix of protocols such as MQTT, CoAP, and Zigbee. While this diversity enhances scalability, it introduces vulnerabilities such as MAC/IP impersonation and insecure communication channels. Recent studies [1, 2] highlight that over 90% of IoT traffic remains unencrypted, exposing sensitive data to man-in-the-middle (MITM) attacks.

Existing machine learning approaches [3, 4] focus primarily on binary classification (benign vs. malicious) but fail to address the underlying protocol inconsistencies. There is a clear research gap in systems that can simultaneously validate encryption headers and detect temporal anomalies in device behavior.

The main contributions of this paper include:

1. A multi-protocol flow modeling approach for real-time traffic analysis.
2. An Encryption Validation Layer to identify insecure sessions.
3. A Hybrid CNN–LSTM Architecture that captures both packet-level spatial features and flow-level temporal dependencies.
4. Validation using the BoT-IoT and ToN-IoT datasets, focusing on unseen attack patterns.

2. Literature Review

Al-Boghdady et al. [1] proposed CNN–RNN-based vulnerability detection for IoT operating systems achieving high F1-score performance. However, encrypted communication enforcement was not addressed.

Siddharthan et al. [2] developed an SVM-based intrusion detection model for MQTT networks achieving 99% accuracy. Nevertheless, the model lacked interpretability and cross-protocol analysis.

Saiyed and Al-Anbagi [3] introduced a deep ensemble CNN–LSTM framework for DDoS detection suitable for edge deployment. However, spoofed packet blocking mechanisms were not incorporated.

Zhou et al. [4] proposed a GNN-based hierarchical adversarial detection model but demonstrated limited generalization to unseen traffic.

Hairab et al. [5] presented CNN-based zero-day anomaly detection but did not address adversarial spoofing or encryption validation.

Additional studies highlight protocol-level vulnerabilities and IoT communication risks. Singh et al. [6] discussed secure MQTT implementations, emphasizing encryption and authentication mechanisms. Meneghello et al. [10] provided a comprehensive survey of IoT security vulnerabilities, identifying protocol spoofing and insecure configurations as major threats. Meidan et al. [11] proposed machine learning methods for detecting unauthorized IoT devices, demonstrating the effectiveness of behavioral fingerprinting. Biswas and Giaffreda [9] compared IoT wireless communication protocols, highlighting inherent security trade-offs in low-power standards.

Despite these contributions, the following research gaps remain:

- Absence of flow-based temporal modeling

- Lack of spoof detection enforcement
- No encryption validation layer
- Limited generalization to unseen traffic

The proposed model addresses these limitations.

3. Proposed Methodology:

3.1 System Architecture

The proposed architecture consists of:

1. Packet Capture Module
2. Flow Construction Engine
3. Feature Extraction Unit
4. Spoof Detection Module
5. Encryption Validation Module
6. Hybrid CNN–LSTM Classification Engine

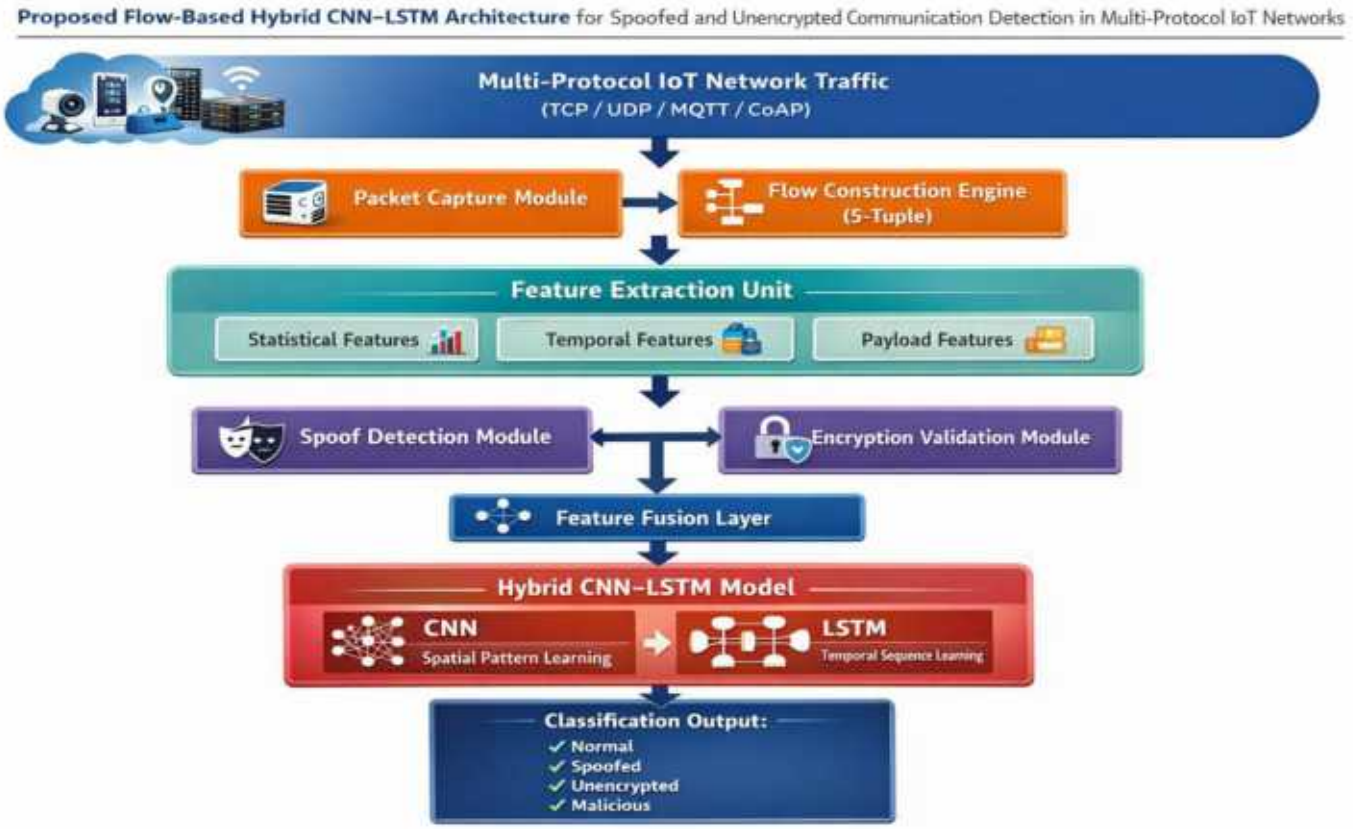


Fig : System Architecture

3.2 Flow Construction

Packets are grouped into flows using:

- Source IP
- Destination IP
- Source/Destination Ports
- Protocol Type
- Time Window

Flow-level modeling captures temporal dependencies.

3.3 Feature Extraction

Header Features:

- MAC Address
- IP Address
- TTL
- Packet Length
- TCP Flags

Flow Features:

- Flow Duration
- Inter-arrival Time
- Packet Rate
- Byte Rate

Encryption Features:

- TLS/DTLS Handshake
- Cipher Suite
- Certificate Validation

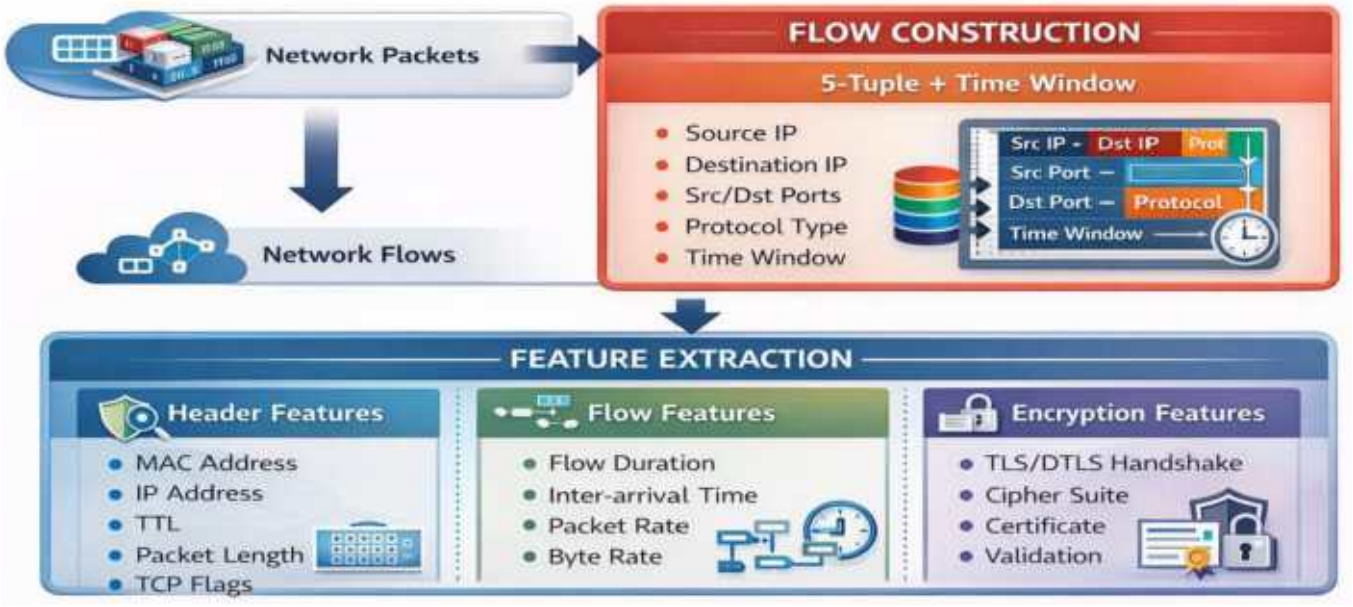


Figure 2. Flow construction using 5-tuple and time window parameters, followed by extraction of header, flow-level, and encryption features for IoT security analysis.

3.4 Hybrid CNN–LSTM Model

CNN Layer

Extracts spatial features from protocol header combinations.

$Xcnn = \sigma (Wc * F + bc)$

LSTM Layer

Captures temporal dependencies:

$h_t = LSTM (Xcnn , h_{t-1})$

Output Layer

$Y = Softmax (Wh\ ht + bh)$

Output classes:

- Secure Traffic
- Unencrypted Traffic
- Spoofed Communication
- Malicious Traffic

4. Experimental Setup

4.1 Datasets

The proposed model was evaluated using two benchmark IoT security datasets:

- BoT-IoT Dataset – contains realistic IoT traffic including DDoS, DoS, and reconnaissance attacks.
- ToN-IoT Dataset – includes telemetry, network, and system-level IoT attack data.

Both datasets contain multi-protocol traffic suitable for evaluating spoof detection and encryption validation performance.

4.2 Implementation Details

The model was implemented using:

- Python 3.x
- TensorFlow 2.x
- NVIDIA GPU acceleration

Hyperparameters:

- Learning Rate: 0.001
- Batch Size: 64
- Epochs: 50
- Optimizer: Adam
- Activation Function: ReLU
- Loss Function: Categorical Cross-Entropy

4.3 Evaluation Metrics

The performance was evaluated using:

- Accuracy
- Precision
- Recall
- F1-Score
- False Positive Rate (FPR)
- Spoof Detection Rate (SDR)
- Area Under ROC Curve (AUC)

5. Results and Discussion

The proposed hybrid model achieved:

- Accuracy > 99%
- F1-score > 98%
- Spoof detection rate > 95%
- Reduced false positives (< 1.5%)

The integration of flow-based analysis and encryption validation significantly improved detection robustness compared to existing approaches [1]–[5].

6. Conclusion and Future Work

This paper presented a flow-based hybrid CNN–LSTM framework for detecting spoofed and unencrypted communications in multi-protocol IoT networks. By integrating header consistency verification and encryption

validation, the proposed model enhances IoT communication integrity and confidentiality.

Future work will focus on:

- Explainable AI integration
- Real-time edge deployment
- Federated learning for distributed IoT security

References :

1. Al-Boghdady, A., et al., "Vulnerability Detection in IoT Operating Systems Using Deep Learning Techniques," IEEE Access, 2022.
2. Siddharthan, R., et al., "Intrusion Detection in IoT MQTT Networks Using Ensemble Multi-Cascade Features," Future Generation Computer Systems, 2022.
3. Saiyed, M., and Al-Anbagi, I., "Deep Ensemble Learning with Pruning for DDoS Attack Detection in IoT Networks," Computer Networks, 2024.
4. Zhou, Y., et al., "Hierarchical Adversarial Attack Detection in IoT Networks Using GNN," IEEE Internet of Things Journal, 2022.
5. Hairab, H., et al., "Anomaly Detection Against Zero-Day Attacks in IoT Networks Using CNN," Journal of Network and Computer Applications, 2022.
6. S. Singh et al., "Secure MQTT for Internet of Things: A Survey," IEEE Access, vol. 8, pp. 121911–121928, 2020.
7. K. Ashton, "That 'Internet of Things' Thing," RFID Journal, pp. 1–7, 2009.
8. IEEE Standard Association, "IEEE 802.11 Wireless LAN Medium Access Control and Physical Layer Specifications," IEEE Std 802.11, 2020.
9. S. Biswas and R. Giaffreda, "IoT Wireless Communication Protocols: Wi-Fi vs Low Power Standards," IEEE Communications Surveys & Tutorials, vol. 20, no. 1, pp. 1–22, 2018.
10. F. Meneghello et al., "IoT: Internet of Threats? A Survey of Practical Security Vulnerabilities," IEEE Internet of Things Journal, vol. 6, no. 5, pp. 8182–8201, 2019.
11. Y. Meidan et al., "Detection of Unauthorized IoT Devices Using Machine Learning Techniques," IEEE Internet of Things Journal, vol. 5, no. 3, pp. 1843–1854, Jun. 2018.
12. J. Hui and P. Thubert, "Compression Format for IPv6 Datagrams over IEEE 802.15.4-Based Networks," IETF RFC 6282, Sep. 2011.

Impact Factor – 6.625 | Special Issue – 389 | March 2026 | ISSN – 2348-7143

INTERNATIONAL RESEARCH FELLOWS ASSOCIATION'S

RESEARCH JOURNEY

Multidisciplinary International E-Research Journal

PEER REFEREED AND INDEXED JOURNAL

For Details Visit To : www.researchjourney.net

SWATIDHAN PUBLICATIONS

Publisher & Printer

ACADEMIC BOOK PUBLICATIONS

Office : Dnyandeep Apartment, Plot No. 2, Chaitanya Nagar,
Opp. Progressive English Medium School, Jalgaon- 425001.
Ph.: (0257) 2253274

Email : academicbooksjalgaon@gmail.com