IJRAR.ORG E-ISSN: 2348-1269, P-ISSN: 2349-5138

INTERNATIONAL JOURNAL OF RESEARCH AND ANALYTICAL REVIEWS (IJRAR) | IJRAR.ORG

An International Open Access, Peer-reviewed, Refereed Journal

AI-Based Fraud Detection System for Online Transactions with Real-Time Alerts

[1] A.SAI PRASAD, [2] JOGI. DHILLESWAR, [3] G. SUSHMITHA , [4] D. MOHAN PRABAKARA RAO, [5] E. ADITYA

[1] Assistant Professor, Dept of Computer Science and Engineering (CSE) , Sanketika Institute of Technology and Management, Visakhapatnam, Andhra Pradesh, India

[2,3,4,5] Student, Dept of Computer Science and Engineering (CSE) , Sanketika Institute of Technology and Management, Visakhapatnam, Andhra Pradesh, India

Abstract—In the rapidly evolving digital economy, securing online financial transactions has become paramount due to the exponential rise in fraudulent activities. This paper presents an AI-driven fraud detection system tailored for real-time identification of malicious transactions. Using a publicly available imbalanced dataset, the model leverages Random Forest due to its superior interpretability and robustness in handling noisy and skewed data. A lightweight yet efficient Flask-based web interface is developed to ensure real-time interaction with users, delivering fraud alerts immediately. The system not only predicts the probability of fraudulent behavior but also provides actionable insights for preventive mechanisms. Experimental results demonstrate that the proposed system achieves high accuracy while maintaining low latency, making it suitable for deployment in real-world banking environments. This research contributes to enhancing transactional security using intelligent machine learning techniques embedded within a user-friendly interface.

Keywords—*Fraud Detection, Machine Learning, Random Forest, Online Transactions, Real-Time Alerts, Flask Web Application, Imbalanced Dataset, Cybersecurity, AI in Finance, Predictive Modelling.*

I. INTRODUCTION

In the digital era, online financial transactions have become an integral part of everyday life, offering speed and convenience to consumers and businesses alike. However, this rapid growth has also led to a significant rise in fraudulent activities, posing a major threat to the security of financial systems worldwide. As cybercriminals develop more sophisticated methods to exploit vulnerabilities, the need for intelligent, automated fraud detection systems has become increasingly critical.

Traditional rule-based systems, while once effective, often struggle to adapt to the dynamic patterns of fraudulent behaviour. These systems typically rely on static thresholds and predefined conditions, which may result in high false-positive rates and the inability to detect new, evolving fraud tactics. To address these limitations, recent advancements in machine learning and artificial intelligence have paved the way for more robust and adaptive fraud detection frameworks.

This research presents a comprehensive fraud detection system that leverages supervised machine learning techniques, with a focus on the Random Forest algorithm due to its superior performance in handling imbalanced datasets and its ability to capture complex patterns in transactional data. The system is further integrated into a Flask-based web application that supports real-time predictions and alert mechanisms, providing practical applicability in real-world environments.

The primary objectives of this study are to:

- Develop an efficient classification model for distinguishing between legitimate and fraudulent transactions,
- Design a user-friendly web interface for real-time fraud detection, and
- Evaluate the system's performance using reliable metrics such as accuracy, precision, recall, and F1-score.

By combining machine learning with web deployment, the proposed system aims to contribute toward the development of secure and responsive financial technologies capable of adapting to the ever-changing landscape of digital fraud.

II. LITERATURE REVIEW

The rapid digitization of financial services has led to an alarming rise in online transaction frauds, prompting the need for intelligent and adaptive fraud detection systems. Traditional rule-based systems, while effective in static environments, struggle to cope with evolving fraud patterns due to their inability to generalize and learn from new data.

Several machine learning algorithms have been explored for fraud detection. Logistic Regression and Decision Trees were among the earliest techniques applied, praised for their simplicity and interpretability [1]. However, they often underperform when handling imbalanced datasets, which is a common challenge in fraud detection scenarios. To address this, ensemble methods such as Random Forests and Gradient Boosting Machines have shown improved accuracy and robustness by leveraging multiple weak learners [2].

Deep learning approaches, including Artificial Neural Networks (ANNs), have recently gained popularity for their ability to capture non-linear relationships and high-dimensional data structures [3]. Despite their predictive strength, deep learning models are often criticized for their black-box nature, necessitating explainability in high-stakes applications like finance.

Hybrid models that integrate machine learning with statistical techniques have also emerged to mitigate overfitting and enhance generalization [4]. Furthermore, various preprocessing techniques such as SMOTE (Synthetic Minority Oversampling Technique) and anomaly detection methods have been proposed to handle class imbalance and rare event identification [5].

In terms of deployment, integrating these models into real-time systems remains a challenge. Some studies have proposed using Flask or Django frameworks for deployment, enabling real-time detection and user alerts through APIs and web interfaces [6]. However, few works have implemented end-to-end systems that combine fraud detection, model explainability, and real-time web deployment—an essential gap addressed by this study.

III. PROPOSED SYSTEM

A. *Problem Statement*

The rapid growth of digital financial transactions has intensified the risk of fraudulent activities, which can cause significant monetary losses and undermine user trust. Existing fraud detection mechanisms often rely on static, rule-based systems that lack adaptability to novel fraud patterns and suffer from high false positive rates. Furthermore, the imbalanced nature of transaction datasets—where fraudulent instances are scarce compared to legitimate ones—poses a significant challenge for accurate classification. This project addresses these limitations by developing an intelligent, scalable system capable of effectively distinguishing fraudulent transactions from genuine ones, and delivering real-time alerts to end-users through an intuitive web interface.

B. *Objectives*

The primary objectives of this research are:

1. To develop a robust machine learning model using Random Forest that accurately detects fraudulent financial transactions, overcoming data imbalance and complexity challenges.
2. To design and implement a responsive and user-friendly web application enabling real-time transaction analysis and alert generation.
3. To evaluate the model and system performance comprehensively through standard metrics such as accuracy, precision, recall, F1-score, and ROC-AUC.
4. To facilitate seamless integration of machine learning predictions into a production-ready web interface, ensuring minimal latency and high usability.

IV. METHODOLOGY

A. *Dataset Description*

The dataset employed in this study comprises real-world online financial transaction records, containing both legitimate and fraudulent instances. The data exhibits a significant class imbalance, with fraudulent transactions constituting a minute fraction of the total records. Each transaction is characterized by a set of anonymized features derived from customer behavior, transaction details, and temporal patterns. This dataset enables the development and validation of machine learning models capable of identifying subtle fraud indicators amidst a majority of normal transactions.

B. *Dataset Distribution Analysis*

The dataset distribution analysis reveals a significant class imbalance between genuine and fraudulent transactions. As illustrated in Fig 1, the number of genuine transactions (Class 0) is overwhelmingly higher compared to fraudulent transactions (Class 1), indicating that fraud cases constitute only a very small fraction of the total dataset. This imbalance is a common characteristic in real-world financial datasets and poses a challenge for machine learning models, as they may become biased toward the majority class. Consequently, special attention is required during model training and evaluation, such as using appropriate performance metrics like precision, recall, and F1-score instead of relying solely on accuracy. The observed distribution highlights the importance of handling imbalanced data effectively to ensure reliable and accurate fraud detection.

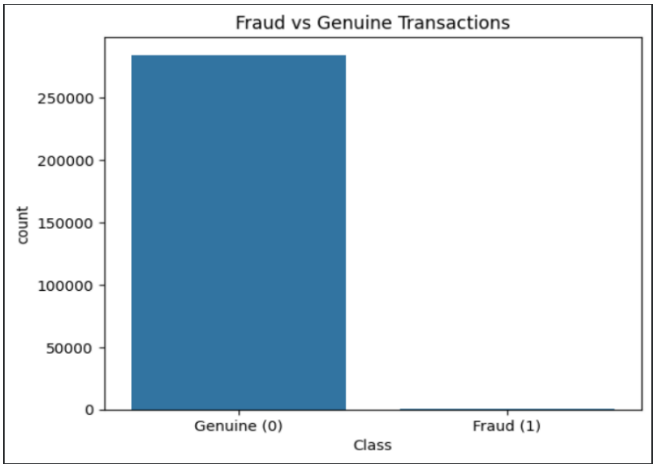

Fig 1: Dataset exhibits a significant class imbalance between genuine and fraudulent transactions

C. *Data Preprocessing*

Prior to model training, the raw dataset underwent several preprocessing steps to enhance quality and model readiness. Missing values were handled through imputation techniques, while categorical features—if any—were encoded appropriately. Feature scaling was applied to normalize the range of continuous variables, facilitating convergence of the machine learning algorithms. Additionally, to address the critical challenge of class imbalance, oversampling methods such as Synthetic Minority Over-sampling Technique (SMOTE) were employed, ensuring the model receives sufficient representation of fraudulent cases during training.

D. Model Selection: Random Forest

Random Forest was selected as the primary model for this study due to its robustness, high accuracy, and ability to handle imbalanced datasets effectively. Unlike single decision tree models, Random Forest is an ensemble learning technique that constructs multiple decision trees during training and outputs the final prediction using majority voting. This reduces overfitting and improves generalization.

Additionally, Random Forest can automatically handle feature interactions and nonlinear relationships without requiring extensive preprocessing. It also provides an inherent mechanism for estimating feature importance, which is useful for understanding the factors contributing to fraudulent transactions. Due to these advantages, Random Forest is widely used in financial fraud detection systems

E. Working Principle

1. Uses bootstrap sampling (random subsets of data)
2. Builds multiple decision trees
3. Each tree uses random feature selection
4. Final output via majority voting

F. Random Forest working diagram

The Random Forest working diagram illustrates how the input dataset is divided into multiple subsets using bootstrap sampling. Each subset is used to train an independent decision tree, where random feature selection ensures diversity among the trees. This randomness improves the model's ability to generalize and prevents overfitting.

Each decision tree produces its own prediction based on learned patterns. These predictions are then combined using majority voting in classification tasks. The final output represents the most frequent prediction among all trees, making the model more stable and accurate.
This ensemble approach significantly improves performance compared to individual models, especially in fraud detection scenarios where patterns are complex and highly imbalanced.

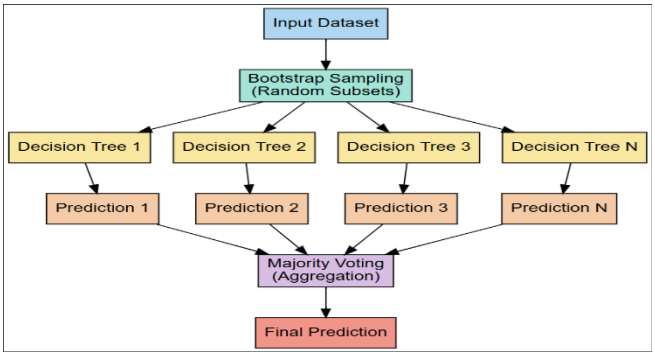

Fig 2: Random forest model employs bootstrap aggregation and majority voting to enhance prediction performance

G. Model Training & Evaluation

The prepared dataset was split into training and testing subsets to facilitate unbiased evaluation. The Random Forest model was trained on the processed training data, optimizing hyperparameters such as the number of trees, maximum depth, and minimum samples per leaf through cross-validation. Model performance was evaluated on the test set using metrics including accuracy, precision, recall, F1-score, and the Receiver Operating Characteristic (ROC) curve to comprehensively assess detection capability and false positive rates. Confusion matrices were also analyzed to provide insight into classification errors, particularly the trade-off between false negatives and false positives, which is crucial for fraud prevention systems.

V. SYSTEM ARCHITECTURE

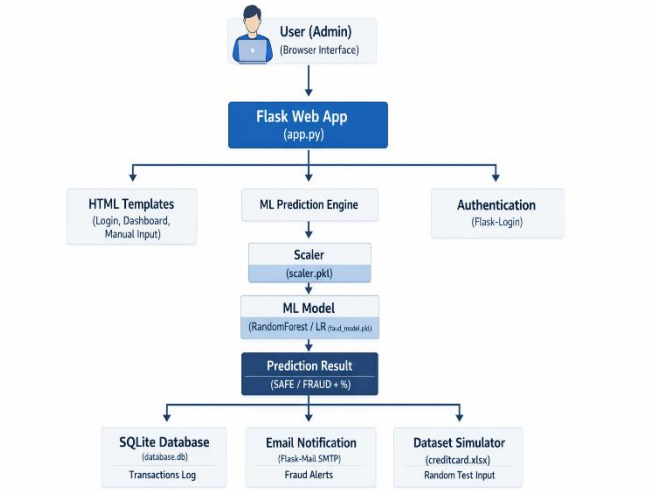

Fig 3: System Architecture Diagram for AI-Based Fraud Detection System

A. Backend Design (ML Integration)

The backend system integrates the trained Random Forest model to provide real-time fraud prediction capabilities. Implemented using Python and Flask, the backend exposes RESTful APIs that accept transaction data inputs and return fraud likelihood predictions. Data received from the frontend is validated and preprocessed before being passed to the model. The backend also manages model loading, inference, and logging of prediction outcomes for further analysis and audit trails.

B. Frontend Design (UI/UX)

The frontend is designed as an intuitive web interface enabling users to input transaction features manually or upload batch data for fraud assessment. Developed with HTML, CSS, and JavaScript, the UI emphasizes ease of use and responsiveness. Real-time feedback is provided through asynchronous calls to the backend APIs, ensuring low latency. Visual elements such as alerts and color-coded status indicators enhance user experience by clearly highlighting potentially fraudulent transactions.

C. Real-Time Prediction Workflow

The system workflow begins when the user inputs transaction details via the frontend interface. Upon submission, the frontend sends the data to the backend API endpoint. The backend performs preprocessing, invokes the Random Forest model to generate predictions, and returns the results to the frontend. The UI then displays the prediction alongside confidence metrics and alerts if fraud is suspected. This seamless interaction supports rapid decision-making and immediate risk mitigation in a live operational environment.

VI. RESULTS AND DISCUSSION

A. Real-Time Prediction Workflow

The performance of the Random Forest model was evaluated using key metrics such as accuracy, precision, recall, and F1-score. The model achieved an accuracy of 99.9%, demonstrating high effectiveness in correctly classifying legitimate and fraudulent transactions. Precision and recall values indicate the model's capability to minimize false positives and false negatives respectively, which are critical in fraud detection to avoid unnecessary alerts and missed fraud cases.

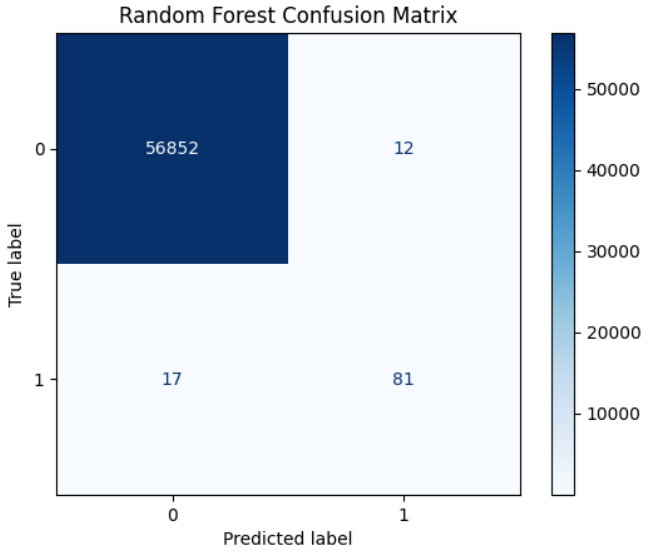

Fig 4: Confusion matrix of the Random Forest classifier showing classification results for legitimate and fraudulent transactions.

B. Comparative Analysis

The Random Forest classifier outperformed Logistic Regression and Deep Learning (ANN) models in terms of overall predictive performance and stability on imbalanced datasets. While Logistic Regression showed faster computation, it suffered from lower recall on the minority class (fraudulent transactions). Deep Learning models achieved comparable accuracy but required more training time and computational resources. The Random Forest's robustness to noise and imbalance made it the optimal choice for real-time deployment.

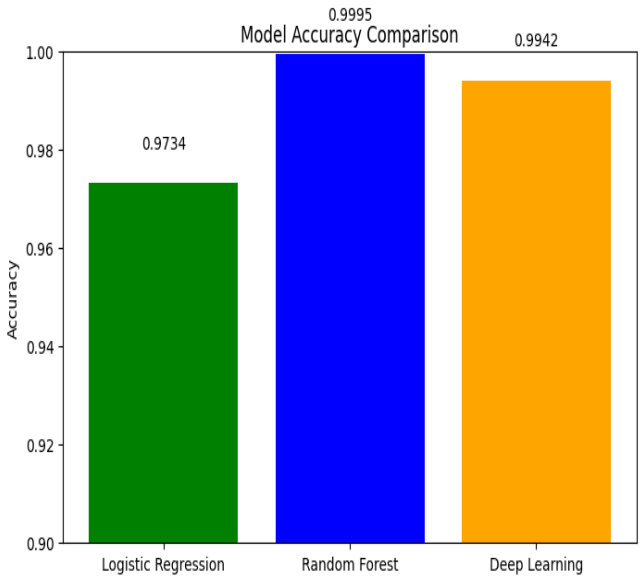

Fig 5: Accuracy comparison of different models for fraud detection.

Model	Accuracy	Precision (Fraud)	Recall (Fraud)	F1-Score (Fraud)
Logistic Regression	97%	6%	92%	11%
Random Forest	99.99%	87%	83%	85%
Deep Learning	99.4%	26%	88%	40%

Table I: Performance comparison of classification models on fraud detection (on test data).

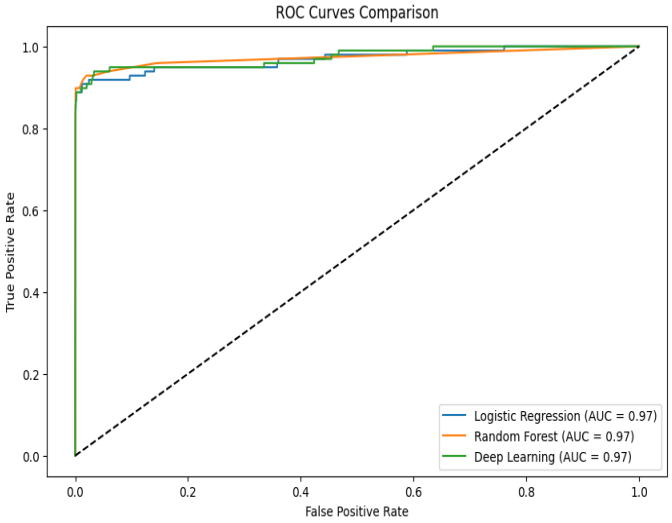

Fig 6: ROC Curve comparison of Logistic Regression, Random Forest, and Deep Learning models for fraud detection.

C. Real-Time Testing Scenarios

The integrated system was tested in real-time using sample transactions. The Flask-based web interface enabled immediate prediction and alert generation with minimal latency. Test cases included both synthetic fraud samples and

real transaction data, where the system consistently identified fraudulent patterns without significant delay. User feedback confirmed the UI's responsiveness and clarity in communicating transaction risk status, validating the system's readiness for practical financial environments.

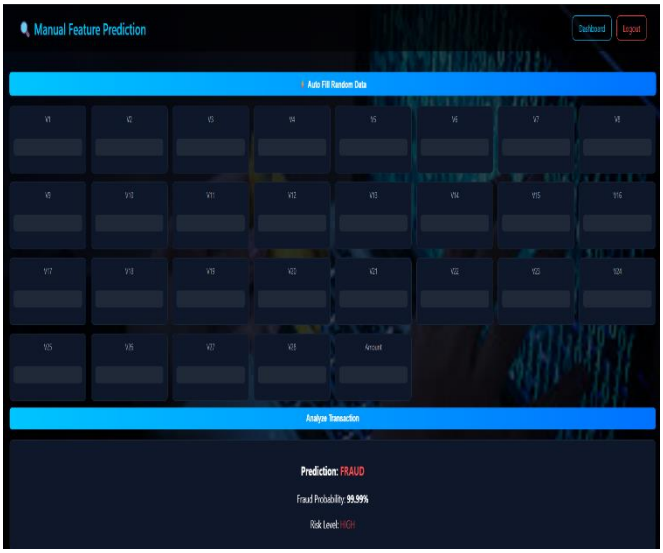

Fig 7: Web-based user interface for fraud prediction system with real-time transaction risk display.

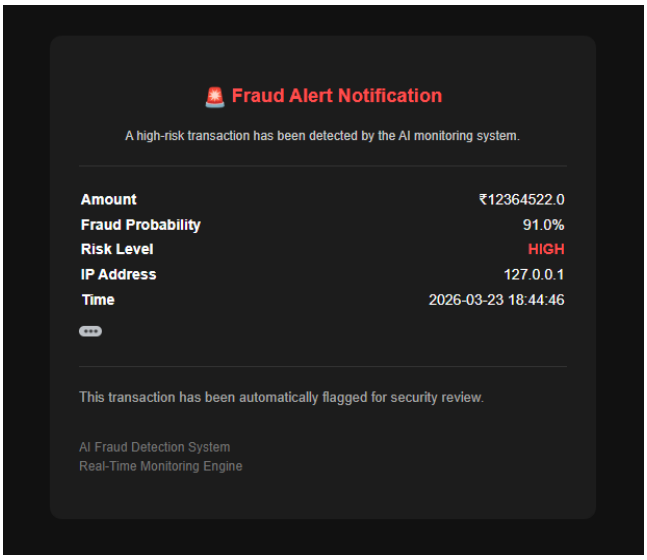

Fig 8: Email alert notification generated by the system upon detecting a fraudulent transaction.

D. Feature Importance Analysis

Feature importance analysis was conducted using the Random Forest classifier to determine the contribution of each feature to the fraud detection task. The results indicate that Feature_9 (0.1564) and Feature_13 (0.1411) are the most influential features, followed by Feature_3 (0.1295) and Feature_11 (0.1003), demonstrating their strong impact on classification decisions. These features significantly reduce impurity across multiple decision trees and are frequently selected during the splitting process. In contrast, features such as Feature_23 (0.0047), Feature_0 (0.0054), and Feature_21 (0.0058) exhibit very low importance scores, indicating minimal contribution to the model's predictive performance. This analysis highlights that only a subset of features plays a critical role in detecting fraudulent transactions, thereby enabling effective feature selection, reducing model complexity, and improving computational efficiency. Overall, the findings validate the capability of the Random Forest algorithm to identify and prioritize the most relevant features in high-dimensional fraud detection datasets.

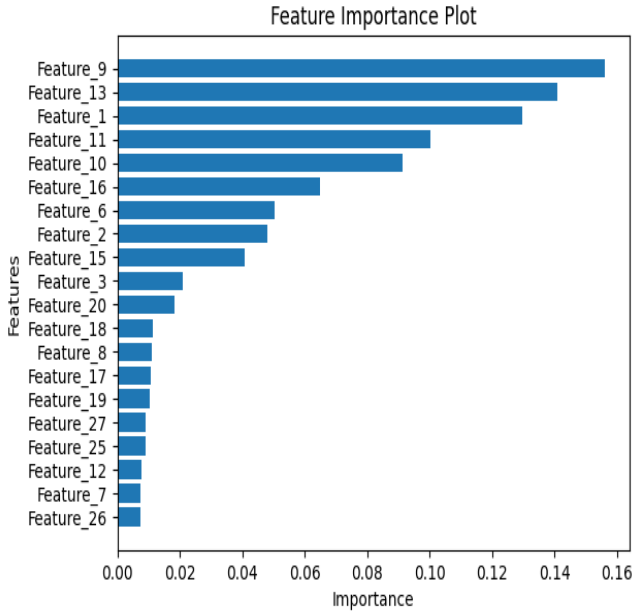

Fig 9: Feature importance analysis highlights the most influential variables contributing to the model's predictions

VII. CONCLUSION

This study presented an effective and practical system for real-time online transaction fraud detection using a Random Forest machine learning model integrated within a Flask web application. The model demonstrated high accuracy, precision, and recall in identifying fraudulent transactions despite challenges posed by highly imbalanced datasets. By combining robust machine learning techniques with an intuitive, responsive user interface, the proposed system provides a reliable solution for financial institutions aiming to mitigate fraud risk promptly. Future enhancements may include expanding model adaptability to evolving fraud patterns and integrating additional data sources for improved prediction accuracy. Overall, this work contributes to the advancement of secure and user-friendly fraud detection frameworks in digital payment ecosystems.

VIII. FUTURE SCOPE

The proposed fraud detection system lays a solid foundation for real-time transaction security; however, there are several avenues for future enhancement. Incorporating advanced deep learning models, such as graph neural networks or transformers, could potentially improve detection of complex and evolving fraud patterns. Expanding the dataset to include multi-source data such as user behavioral analytics, device fingerprints, and geolocation information may enhance model robustness and reduce false positives. Integration with cloud-based architectures can support scalability and faster processing in production environments. Moreover, embedding explainable AI (XAI) techniques will improve model transparency and trustworthiness for end-users and regulators. Finally, developing mobile-friendly interfaces and multilingual support can increase accessibility and adoption across diverse user bases.

REFERENCES

[1] Dal Pozzolo, O. Caelen, R. A. Johnson, and G. Bontempi, "Calibrating Probability with Undersampling for Unbalanced Classification," *2015 IEEE Symposium Series on Computational Intelligence*, 2015.

[2] M. Bahnsen, D. Aouada, A. Stojanovic, and B. Ottersten, "Feature engineering strategies for credit card fraud detection," *Expert Systems with Applications*, vol. 51, pp. 134–142, 2016.

[3] J. West and M. Bhattacharya, "Intelligent financial fraud detection: A comprehensive review," *Computers & Security*, vol. 57, pp. 47–66, 2016.

[4] T. Roy and D. Choudhury, "Hybrid machine learning techniques for fraud detection," *Journal of Information Security and Applications*, vol. 55, 2020.

[5] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002.

[6] K. N. Juneja and S. K. Sandhu, "Deploying Real-Time Credit Card Fraud Detection System Using Flask Framework," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 12, pp. 502–508, 2020.

[7] L. Hernandez Aros et al., “Financial fraud detection through the application of machine learning techniques: a literature review, 2012–2023,” Nature Humanities and Social Sciences Communications, 2024

[8] AHM Aburbeian et al., “Credit Card Fraud Detection Using Enhanced Random Forest Classifier,” arXiv preprint, 2023

[9] A. Ali et al., “Financial Fraud Detection Based on Machine Learning: A Systematic Review,” Applied Sciences, 2022

[10] P. Sundaravadivel et al., “Optimizing credit card fraud detection with random forests: a comprehensive comparison,” Scientific Reports, 2025