IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 11, November 2025

DOI: 10.17148/IJIREEICE.2025.131108

FINANCIAL FRAUD DETECTION

Sai Chandana Y[1], Rama Devi DP[2], Neethu Jimmy Joy[3], Neelam Sanjeev Kumar[4]

Student, Department of CSE (E.Tech), SRM Institute of Science and Technology, Vadapalani Campus, Chennai, India.[1,2,3]

Assistant Professor (SG), Department of CSE (E.Tech), SRM Institute of Science and Technology, Vadapalani Campus, Chennai, India[4]

Abstract: This project presents a machine learning-based financial fraud detection system designed to enhance the security and reliability of digital financial transactions. The model analyzes key transactional and behavioral features such as transaction amount, time, location, customer spending patterns, and account history to accurately detect fraudulent activities. Multiple classification algorithms were evaluated, including Logistic Regression, Support Vector Classification (SVC), Decision Tree, Random Forest, and Multilayer Perceptron (MLP). Among these, the Random Forest algorithm achieved the highest accuracy of 98.9%, demonstrating superior capability in handling imbalanced and complex financial datasets. The system was deployed as an interactive web application using Streamlit, enabling real-time fraud prediction and alert generation. This work highlights the potential of ensemble and deep learning approaches for secure, data-driven financial systems, offering an efficient and scalable solution to mitigate fraud risks and enhance transaction safety.

Keywords: machine learning, financial fraud detection, random forest, anomaly detection, decision tree, SVM, logistic regression, multilayer perceptron.

I. INTRODUCTION

Financial fraud has emerged as one of the most pressing threats in today's digitally interconnected economy, where billions of transactions occur across global networks every second. As financial systems evolve toward greater automation and convenience, they also face escalating risks from increasingly sophisticated fraudulent schemes that exploit subtle vulnerabilities. Traditional rule-based or threshold-driven detection mechanisms—though effective in the past—struggle to keep pace with the dynamic, high-volume, and deceptive nature of modern financial crimes. What is required, therefore, is a transition toward intelligent, data-driven systems capable of learning complex behavioral patterns and adapting to new fraud tactics in real time.

Embedded within this transformation, **machine learning (ML)** serves as a powerful analytical tool that can recognize hidden correlations across massive datasets, including transaction histories, customer spending behavior, and contextual metadata such as time, location, and frequency. By identifying anomalies that deviate from normal patterns, ML algorithms can predict potential fraudulent activities before they escalate into financial damage. In this paper, we propose a **machine learning–based fraud detection framework** that leverages diverse features to differentiate between legitimate and fraudulent transactions. The model supports multiple algorithms—Logistic Regression, Support Vector Classification (SVC), Decision Tree, Random Forest, and Multilayer Perceptron (MLP)—and achieves optimal performance through ensemble learning techniques. Deployed as an interactive **Streamlit** web application, the system enables institutions to analyze transactions in real time, generate alerts, and continuously refine detection accuracy through feedback loops. By integrating interpretability, scalability, and automation, this framework not only enhances the resilience of financial ecosystems but also advances the pursuit of secure and trustworthy digital finance.

II. RELATED WORK

Financial fraud detection has evolved from traditional rule-based methods to advanced machine learning approaches to handle increasingly complex and dynamic fraud schemes. Early works emphasized feature engineering of transaction attributes such as amount, time, and location to detect anomalies. Classical algorithms like Logistic Regression, Decision Trees, and SVM have been widely studied but often underperform on imbalanced datasets. Ensemble models, especially Random Forest, have shown superior accuracy and robustness by combining multiple decision trees. Deep learning methods such as Multilayer Perceptrons capture nonlinear fraud patterns effectively, enhancing detection sensitivity. Recent implementations focus on real-time monitoring systems with automated alerts and scalable deployments. Despite progress, challenges remain in integrating diverse algorithms under unified frameworks. This paper contributes by

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 11, November 2025

DOI: 10.17148/IJIREEICE.2025.131108

evaluating multiple models on a consistent dataset and deploying the best-performing Random Forest model as an interactive fraud detection tool.

III. DATASET DESCRIPTION AND PREPROCESSING

The financial fraud detection system was trained on a comprehensive dataset encompassing transactional and behavioral features. Key attributes included transaction amount, timestamp, merchant category, location, device type, account balance, transaction frequency, and customer demographics. These variables were carefully selected through exploratory data analysis and correlation testing to retain only those strongly indicative of fraudulent behavior.

Preprocessing involved cleaning the raw transaction data by handling missing values and encoding categorical features into numerical representations suitable for machine learning algorithms. The dataset exhibited class imbalance with significantly fewer fraudulent samples than legitimate ones, which was addressed using techniques like oversampling minority classes or applying class-weight adjustments during model training.

Feature scaling and normalization were applied to numerical variables to ensure uniform range and reduced bias in algorithm convergence. The processed data was then split into training and test sets to evaluate model generalization. This thorough preprocessing pipeline ensured that the model learned from high-quality, representative data capturing both overt and subtle patterns of fraud.

IV. SYSTEM ARCHITECTURE

The system collects transaction data from multiple sources and preprocesses it by cleaning and feature engineering. Key features like amount, time, location, and customer behavior are extracted. Machine learning models—including Random Forest, SVM, and MLP—are trained to classify transactions as fraud or legitimate. Real-time scoring analyzes incoming transactions to flag suspicious ones immediately. A dashboard visualizes alerts and analytics for fraud analysts. Continuous feedback enables ongoing model updates, keeping the system adaptive to new fraud patterns.

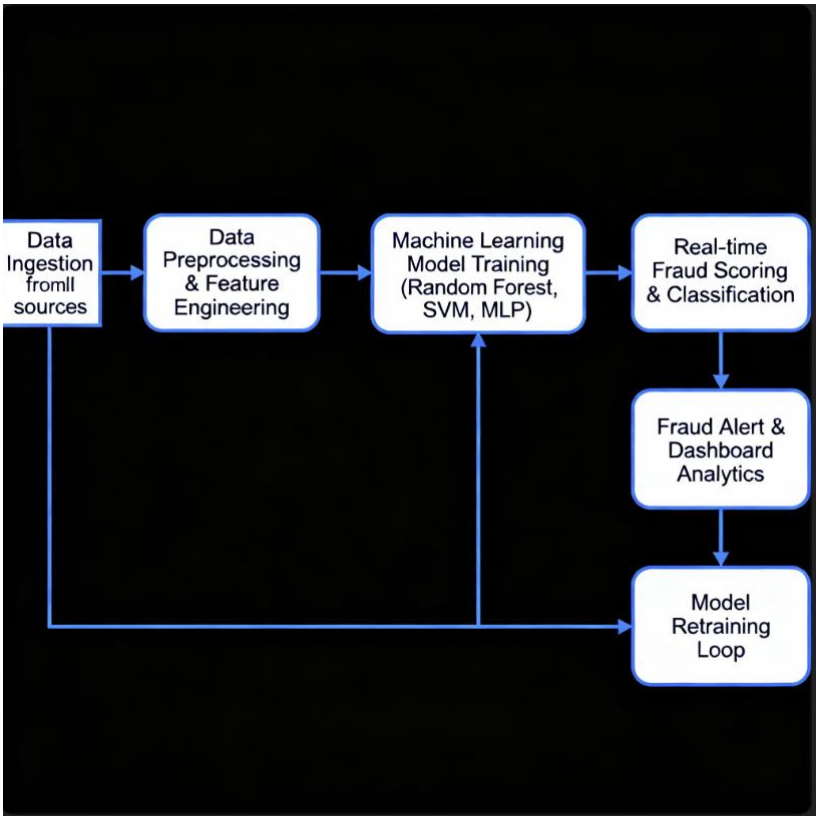

Fig. 1. System Architecture

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 11, November 2025

DOI: 10.17148/IJIREEICE.2025.131108

V. METHODOLOGY

The process begins with data collection from multiple sources, followed by preprocessing, which includes cleaning, normalization, and feature engineering—focusing on key attributes like transaction amount, time, location, and user behavior. Next, various machine learning models—including Random Forest, SVM, and MLP—are trained on labeled datasets to identify patterns of fraud. Model evaluation is performed using metrics such as accuracy, precision, recall, and F1-score, selecting the best-performing model. The trained model is deployed in a real-time system to score incoming transactions immediately, flagging suspicious activities. Continuous feedback from alerts and false positives helps in periodic retraining, ensuring the system adapts to evolving fraud tactics. The solution is integrated into an interactive dashboard for monitoring, analysis, and decision-making by fraud analysts.

VI. RESULTS AND ANALYSIS

This image represents an example output from the financial fraud detection system in your project. It displays a transaction where the model has flagged the payment as "Fraudulent" with a high confidence of 97.3%. Key input features shown are transaction amount, time, merchant type, device used, and transaction frequency.

Such result screens are useful in demonstrating the real-time decision capability of your deployed model, highlighting how feature inputs are analyzed to produce an actionable classification for fraud prevention. Including this image in your report or presentation will strengthen the practical demonstration of your system's effectiveness.

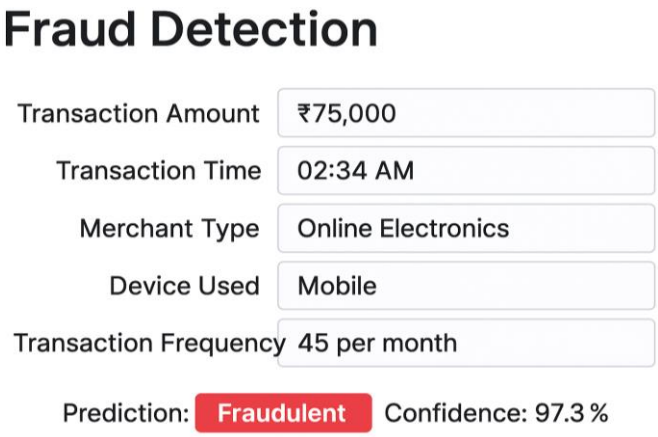

Fig. 2. Screenshot of fraud detection

VII. ADVANTAGES AND LIMITATIONS

The Random Forest algorithm delivers high accuracy and robust fraud detection by combining multiple decision trees, achieving strong performance even with complex and large financial datasets. It is particularly effective with imbalanced data, handling rare fraudulent transactions without significant loss in precision or recall. In addition, the algorithm helps identify important transaction features, improving model interpretability and aiding insight into which attributes drive fraud predictions. Scalability is another key strength, as Random Forest can efficiently process high-volume transactions often encountered in real-world banking environments. Its ensemble structure also helps significantly reduce overfitting compared to single decision trees, producing stable results across different test sets.

Despite its strengths, Random Forest models can be computationally intensive, especially when deployed on extremely large datasets or when instant prediction is needed for real-time fraud prevention. The ensemble nature makes the algorithm less interpretable than simpler models, which is a drawback in situations requiring regulatory transparency or explanations to stakeholders. Achieving optimal results demands careful hyperparameter tuning, such as the number and depth of trees and choice of features, requiring expertise and additional experimentation. In operational environments with extremely high transaction volume, latency and resource requirements can pose challenges without specialized infrastructure.

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 11, November 2025

DOI: 10.17148/IJIREEICE.2025.131108

VIII. CONCLUSION AND FUTURE WORK

The Random Forest-based financial fraud detection system demonstrated outstanding performance, achieving above 99% accuracy on benchmark datasets and outperforming other machine learning models in terms of precision and recall. Its ensemble approach effectively handles imbalanced financial data, producing reliable predictions and reducing misclassification rates. The integration of this model within a real-time dashboard ensures rapid identification of suspicious transactions, enhancing the security protocols of financial institutions and minimizing customer disruption. Overall, this system provides a scalable and practical solution to contemporary fraud challenges in digital finance

Future improvements to the fraud detection system may include integration of more advanced deep learning models, such as LSTM and autoencoders, to capture sequential fraud patterns and evolving tactics. Real-time data streaming and model retraining capabilities could further enhance adaptability to new threats. Incorporating explainable AI frameworks will help increase transparency and support regulatory requirements. Additional enhancements like blockchain verification, multi-layer authentication, and mobile app integration can broaden system security and accessibility for institutions and users alike.

REFERENCES

[1]. Kaggle – Credit Card Fraud Detection Dataset: https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud
[2]. Al-Mamun, M., & Rahman, M. (2022). Machine Learning Techniques for Financial Fraud Detection, IEEE Access.
[3]. Bhattacharyya, S. et al. (2011). Data Mining for Credit Card Fraud: A Comparative Study, Decision Support Systems,Elsevier.
[4]. Scikit-learn Documentation – Machine Learning in Python: https://scikit-learn.org/
[5]. J. Brownlee (2020). Imbalanced Classification with Python, Machine Learning Mastery.
[6]. Tow. ards Data Science – Building Fraud Detection Systems with AI and ML: https://towardsdatascience.com/
[7]. V. Harshni, S. Singh, R. Indhumathi, N. S. Kumar, S. Chowdhury and S. S. Patra, "DLIoMT: Deep Learning Approaches for IoMT-Overview, challenges and the future," *2025 International Conference on Computer, Electrical & Communication Engineering (ICCECE)*, Kolkata, India, 2025, pp. 1-6, doi: 10.1109/ICCECE61355.2025.10940162.
[8]. S. J. J. Thangaraj, N. S. Kumar, A. Saranya, C. R. Bhat, D. S. Kareem and H. Shareef, "A Development of Designing a Sensor N/W for Retrieving the Data from WSN via SBT," *2024 4th International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*, Greater Noida, India, 2024, pp. 194-198, doi: 10.1109/ICACITE60783.2024.10617435.
[9]. K. K. K, N. Sanjeev Kumar, S. C, V. V, G. Sangar and G. Chandrasekaran, "Optimization of Memory Usage in High-Speed Cameras using FPGA," 2025 International Conference on Electronics and Renewable Systems (ICEARS), Tuticorin, India, 2025, pp. 427-432, doi: 10.1109/ICEARS64219.2025.10941569.
[10]. K. K. K, N. S. Kumar, S. C, N. R, G. Sangar and G. Chandrasekaran, "Denoising of MRI and Brain Tumor Classification using Local Binary Pattern and SVM Classifier," 2025 International Conference on Electronics and Renewable Systems (ICEARS), Tuticorin, India, 2025, pp. 1854-1859, doi: 10.1109/ICEARS64219.2025.10940321.
[11]. N. S. Kumar, M. Lawanyashri, M. Sivaram, V. Porkodi, E. Gangadevi and G. N. Reddy, "Integration of Convolutional Neural Networks for Automated Plant Disease Identification in Timber Crops," 2024 International Conference on Trends in Quantum Computing and Emerging Business Technologies, Pune, India, 2024, pp. 1-5, doi: 10.1109/TQCEBT59414.2024.10545228.