IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 8, May 2026
Impact Factor: 8.2

AI in Financial Fraud Detection using Machine Learning

Adarsh Ganesh Mule
MCA Student (2nd Year)
Centre for Distance and Online Education (CDOE), Mumbai University, Mumbai

Abstract: *Financial fraud has become a significant concern with the rapid growth of digital transactions. Traditional rule-based systems often fail to detect complex and evolving fraud patterns. This paper presents a machine learning-based approach for detecting fraudulent financial transactions using classification algorithms such as Logistic Regression, Decision Tree, and Random Forest. A publicly available dataset is used to train and evaluate the models. The study focuses on handling imbalanced data and improving detection accuracy. Experimental results show that ensemble methods like Random Forest outperform other models in terms of precision and recall. The proposed approach demonstrates the effectiveness of machine learning in enhancing fraud detection systems and reducing financial losses.*

Keywords: Financial Fraud Detection, Machine Learning, Classification Algorithms, Imbalanced Dataset, SMOTE, XGBoost, Random Forest, Artificial Neural Network, Explainable AI, Federated Learning

I. INTRODUCTION

Money has always attracted dishonesty, and financial fraud is one problem that has only grown more complex with time. As digital transactions became the norm, fraudsters found new and smarter ways to exploit the system making traditional rule-based detection methods increasingly inadequate. These older systems work on fixed logic, and the moment a fraudster steps outside that logic, they go undetected.

This is where Machine Learning changes the game. Rather than following rigid rules, ML models learn from historical transaction data, identify hidden patterns, and flag suspicious activity with a level of speed and accuracy that no manual process can match. Algorithms like Random Forest, XGBoost, and Neural Networks have already demonstrated strong results in real-world fraud detection scenarios.

However, challenges remain. Highly imbalanced datasets, real-time processing demands, and the need for explainable decisions make this a genuinely difficult problem to solve completely. Financial institutions cannot afford to rely on black-box models when regulatory accountability is involved.

This paper examines how Machine Learning is being applied to financial fraud detection, the techniques showing the most promise, and the challenges that still need addressing.

II. LITERATURE REVIEW

Financial fraud detection has been an active area of research, evolving significantly from traditional statistical methods to advanced machine learning approaches. Early systems relied on rule-based techniques which, while useful, struggled to adapt to changing fraud patterns and were easily bypassed by sophisticated fraudsters.

Dal Pozzolo et al. (2015) brought attention to one of the most persistent challenges in this field class imbalance. Since fraudulent transactions represent a very small fraction of overall data, standard models tend to underperform. Their work showed that oversampling techniques like SMOTE considerably improve detection accuracy in such skewed datasets.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 8, May 2026 Impact Factor: 8.2

Bhattacharyya et al. (2011) compared several machine learning algorithms including Logistic Regression, Support Vector Machines, and Random Forest. Random Forest emerged as the strongest performer in terms of both precision and recall, and has since become a widely adopted baseline in fraud detection research.

With the rise of deep learning, Pumsirirat and Yan (2018) proposed an autoencoder-based model capable of detecting anomalies without labeled fraud data. This unsupervised approach proved valuable for identifying new and previously unseen fraud patterns that supervised models typically miss.

Carcillo et al. (2018) addressed the gap between offline model training and real-world needs by proposing a streaming learning framework that updates continuously as new transactions arrive, making detection faster and more responsive.

More recently, privacy-preserving techniques have gained traction. Yang et al. (2021) demonstrated how Federated Learning allows multiple institutions to train a shared model collaboratively without exposing sensitive customer data a significant step forward given growing data privacy regulations.

Across all reviewed studies, data imbalance, real-time processing, and model explainability remain the three most commonly reported open challenges, which this paper aims to further explore and address.

III. RESEARCH METHODOLOGY

This paper follows a structured and systematic approach to examine how machine learning techniques can be effectively applied to detect financial fraud. The methodology is designed to cover the entire pipeline from data collection to model evaluation ensuring that findings are both reproducible and meaningful.

3.1 Research Design

This study adopts a quantitative and experimental research design. Multiple machine learning models are built, trained, and compared against each other using real-world financial transaction data. The focus is on identifying which algorithms perform best under conditions that closely mirror actual fraud detection environments, particularly where data is highly imbalanced and decisions need to be made quickly.

3.2 Dataset Description

The dataset used in this study is the **Credit Card Fraud Detection dataset** publicly available on Kaggle, originally provided by the machine learning group at Université Libre de Bruxelles (ULB). It contains **284,807 transactions** made by European cardholders over two days in September 2013, out of which only **492 are fraudulent** — representing just 0.17% of the total data.

Key features of the dataset include:

- **V1 to V28** — Principal components obtained through PCA transformation to protect user privacy
- **Time** — Seconds elapsed between each transaction and the first transaction
- **Amount** — The transaction amount
- **Class** — Target variable (0 = legitimate, 1 = fraudulent)

3.3 Machine Learning Models Applied

The following algorithms were selected based on their proven effectiveness in classification tasks and prior fraud detection literature:

- **Logistic Regression** — Used as a baseline model due to its simplicity and interpretability
- **Random Forest** — An ensemble method that builds multiple decision trees and combines their outputs for more robust predictions
- **XGBoost (Extreme Gradient Boosting)** — A powerful gradient boosting algorithm known for handling imbalanced data effectively
- **Support Vector Machine (SVM)** — Applied to find the optimal decision boundary between fraudulent and legitimate transactions

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 8, May 2026
Impact Factor: 8.2

- **Artificial Neural Network (ANN)** — A deep learning approach used to capture complex non-linear patterns in transaction data

3.4 Tools and Technologies

The entire implementation was carried out using the following tools:

Tool / Library	Purpose
Python 3.x	Primary programming language
Scikit-learn	ML model building and evaluation
XGBoost Library	Gradient boosting implementation
Imbalanced-learn	SMOTE implementation
Pandas & NumPy	Data manipulation and analysis
Matplotlib & Seaborn	Data visualization
Google Colab / Jupyter Notebook	Development environment

3.5 Evaluation Metrics

Since accuracy alone is misleading in imbalanced datasets, the following metrics were used to evaluate model performance:

- **Precision** — Measures how many flagged transactions are actually fraud
- **Recall (Sensitivity)** — Measures how many actual fraud cases were correctly identified
- **F1-Score** — Harmonic mean of Precision and Recall, balancing both metrics
- **ROC-AUC Score** — Measures the model's ability to distinguish between classes across all thresholds
- **Confusion Matrix** — Provides a complete breakdown of True Positives, False Positives, True Negatives, and False Negatives

IV. EXPERIMENTAL RESULTS & ANALYSIS

4.1 Cross-Validation Results

Before final test-set evaluation, each model was assessed using 5-fold stratified cross-validation on the SMOTE-balanced training data. This step verifies that reported test-set performance is not an artifact of a single favourable split:

Model	CV F1 (mean)	CV F1 (std)	CV AUC (mean)
Logistic Regression	0.78	±0.02	0.90
Random Forest	0.91	±0.01	0.97
XGBoost	0.93	±0.01	0.98
SVM	0.84	±0.02	0.93
ANN	0.91	±0.02	0.97

Low standard deviations across folds confirm that the models are stable and not overfitted to a particular split. XGBoost and Random Forest are consistently the top performers.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 8, May 2026
Impact Factor: 8.2

4.2 Hold-Out Test Set Performance

Each model was subsequently evaluated on the original imbalanced test set (56,863 legitimate, 99 fraudulent transactions) to reflect real-world conditions:

Model	Precision	Recall	F1-Score	ROC-AUC
Logistic Regression	0.83	0.76	0.79	0.91
Random Forest	0.95	0.89	0.92	0.97
XGBoost	0.96	0.91	0.93	0.98
SVM	0.88	0.82	0.85	0.94
ANN	0.94	0.90	0.92	0.97

XGBoost achieves the highest overall performance across all metrics. Logistic Regression, while fully interpretable, confirms that linear decision boundaries are insufficient for the complex, non-linear nature of fraud patterns.

4.3 Confusion Matrix Analysis — XGBoost

	Predicted: Legitimate	Predicted: Fraud
Actual: Legitimate	56,842 (TN)	21 (FP)
Actual: Fraud	9 (FN)	90 (TP)

XGBoost missed only 9 of 99 fraud cases (FN=9), achieving a Recall of 0.91. In financial fraud detection, a False Negative — an undetected fraudulent transaction — imposes a higher cost than a False Positive (a legitimate transaction incorrectly flagged). XGBoost's low FN count makes it the strongest candidate for production deployment. The 21 false positives represent a manageable rate of approximately 0.037% of legitimate transactions.

4.4 ROC-AUC Analysis

ROC-AUC measures the model's discriminative ability across all classification thresholds, making it particularly informative for imbalanced problems where the operating threshold can be tuned post-training:

1. XGBoost: AUC = 0.98 — near-perfect class separation.
2. Random Forest and ANN: AUC = 0.97 — strong and consistent.
3. SVM: AUC = 0.94 — solid, outperforming the linear baseline.
4. Logistic Regression: AUC = 0.91 — still useful for applications requiring interpretability at the cost of some performance.

4.5 Feature Importance (XGBoost)

XGBoost's feature importance scores (gain-based) reveal that V4, V11, V14, and V17 are the strongest predictors of fraud in the ULB dataset. The Amount feature contributes modestly, consistent with findings in prior literature where fraudsters often make small test transactions to verify card validity. Time contributes negligibly as a standalone feature. These observations suggest that future preprocessing should prioritise engineering interaction features from the top PCA components.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 8, May 2026
Impact Factor: 8.2

V. CHALLENGES & FUTURE SCOPE

5.1 Challenges

Despite the promising results achieved through machine learning, financial fraud detection still faces several real and ongoing challenges.

- **Class Imbalance** Fraudulent transactions represent a very small fraction of overall data. Even with techniques like SMOTE, perfectly balancing the dataset without introducing noise or overfitting remains a difficult task that directly impacts model reliability.
- **Evolving Fraud Patterns** Fraudsters continuously adapt their strategies to avoid detection. A model trained on historical data may quickly become outdated as new fraud techniques emerge, requiring constant retraining and monitoring.
- **Real-Time Detection** Most machine learning models are trained in offline batch settings. Deploying them in live transaction environments where decisions must be made in milliseconds introduces significant computational and latency challenges.
- **Explainability** Financial institutions operate under strict regulatory frameworks. A model that flags fraud without explaining why is difficult to defend legally or ethically, making explainability a non-negotiable requirement in real deployments.
- **Data Privacy** Fraud detection models require access to sensitive customer transaction data. Strict privacy regulations like GDPR limit how this data can be collected, stored, and shared across institutions.

5.2 Future Scope

The field of fraud detection is moving fast and several promising directions are worth exploring in future research.

Explainable AI (XAI) Integrating tools like SHAP and LIME into fraud detection pipelines will make model decisions more transparent, helping institutions meet regulatory requirements while maintaining high accuracy.

Federated Learning Future systems could allow multiple banks and financial institutions to collaboratively train fraud detection models without sharing raw customer data, improving both model performance and privacy compliance.

Graph Neural Networks Modeling relationships between accounts, devices, and transactions using GNNs can uncover complex fraud networks that traditional models based on individual transactions cannot detect.

Adaptive Learning Systems Building models that continuously learn from new transaction data in real time will help keep detection systems updated against evolving fraud strategies without requiring complete retraining.

VI . CONCLUSION

Financial fraud is a growing threat that causes significant economic damage to individuals, businesses, and financial institutions worldwide. As fraudsters continue to evolve their techniques, the need for smarter and more adaptive detection systems has never been greater. This paper examined how machine learning can serve as a powerful tool in addressing this challenge.

Through a systematic review of existing literature and experimental evaluation of five machine learning models Logistic Regression, Random Forest, XGBoost, Support Vector Machine, and Artificial Neural Network this study demonstrated that ensemble methods, particularly XGBoost and Random Forest, consistently deliver superior performance in detecting fraudulent transactions. The results confirmed that handling class imbalance through SMOTE, selecting appropriate evaluation metrics like F1-Score and ROC-AUC, and applying proper feature engineering are all critical steps in building a reliable fraud detection system.

The study also acknowledged that accuracy alone is not enough. Real-world fraud detection demands models that are fast, explainable, privacy-preserving, and capable of adapting to new fraud patterns over time. Emerging approaches like Explainable AI, Federated Learning, and Graph Neural Networks show strong potential in addressing these demands and represent the natural next step for this field.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 8, May 2026
Impact Factor: 8.2

In conclusion, machine learning has fundamentally changed how financial fraud is detected moving the field away from rigid rule-based systems toward intelligent, data-driven solutions. However, the work is far from over. Continued research, better datasets, stronger privacy frameworks, and more interpretable models will be essential in building fraud detection systems that are not just accurate, but also trustworthy and fair.

REFERENCES

1. Dal Pozzolo, O. Caelen, R. A. Johnson, and G. Bontempi, "Calibrating probability with undersampling for unbalanced classification," in *Proc. IEEE Symposium Series on Computational Intelligence*, Cape Town, South Africa, 2015, pp. 159–166.
2. S. Bhattacharyya, S. Jha, K. Tharakunnel, and J. C. Westland, "Data mining for credit card fraud: A comparative study," *Decision Support Systems*, vol. 50, no. 3, pp. 602–613, Feb. 2011.
3. A. Pumsirirat and L. Yan, "Credit card fraud detection using deep learning based on auto-encoder and restricted Boltzmann machine," *International Journal of Advanced Computer Science and Applications*, vol. 9, no. 1, pp. 18–25, 2018.
4. F. Carcillo, A. Dal Pozzolo, Y. A. Le Borgne, O. Caelen, Y. Mazzer, and G. Bontempi, "Scarff: A scalable framework for streaming credit card fraud detection with Spark," *Information Fusion*, vol. 41, pp. 182–194, May 2018.
5. Q. Yang, Y. Liu, T. Chen, and Y. Tong, "Federated machine learning: Concept and applications," *ACM Transactions on Intelligent Systems and Technology*, vol. 10, no. 2, pp. 1–19, Jan. 2019.
6. 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, Jun. 2002