International Journal of Scientific Research & Engineering Trends
Volume 12, Issue 5, Sep-Oct-2026, ISSN (Online): 2395-566X

Artificial Intelligence in Fraud Detection: A Review of Machine Learning Methods for Banking, Digital Payments and Financial Services

Chapekar Ved Sudhir[1]; Prof. M. A. Auti[2]; Prof. M. C. Kolse[3]; Prof. S. K. Said[4]
Department Of Artificial Intelligence And Data Science Jaihind College Of Engineering Pune, Maharashtra, India

Abstract—Fraud in banking, digital payments, online shopping and insurance causes large financial losses and erodes customer trust. Traditional fraud detection relies on predefined rules and manual verification, which struggle with complex and constantly changing fraud patterns. Artificial Intelligence (AI) enables systems that analyse large volumes of transaction data in real time, recognize unusual behaviour and generate alerts. This paper reviews AI-based fraud detection. It explains the characteristics of fraud data, including extreme class imbalance, concept drift and delayed labels, and surveys statistical approaches, supervised classifiers such as logistic regression, random forests and gradient boosting, unsupervised anomaly detection, sequence models that learn from customers' transaction histories, and graph neural networks that expose fraud rings. Hybrid systems that combine supervised and unsupervised learning, benchmark datasets and evaluation metrics suited to imbalanced problems are discussed. A five-layer real-time fraud detection architecture is proposed, along with research gaps, objectives, a comparison of methods and challenges including false positives, explainability and adversarial adaptation by fraudsters. The review concludes that layered AI systems with human investigators in the loop provide the most effective defence.

Index Terms—Fraud detection, machine learning, credit card fraud, anomaly detection, class imbalance, concept drift, graph neural networks, digital payments

I. INTRODUCTION

India's digital payments have grown rapidly through cards, net banking and UPI. As transaction volumes rise, so do opportunities for fraud: stolen card details, account takeovers, fake merchants, phishing-based payment requests and insurance claims for events that never happened. Every fraudulent transaction that slips through causes direct loss, and every genuine transaction wrongly blocked frustrates a customer.

Bolton and Hand [1] reviewed statistical fraud detection and noted that fraud detection is difficult because fraudsters adapt, legitimate behaviour varies widely, and labelled fraud is rare. Rule-based systems capture known patterns but miss new ones.

This paper reviews AI techniques for fraud detection, proposes a real-time architecture and discusses practical challenges.

II. THEORETICAL BACKGROUND

2.1 Types Of Financial Fraud

Fig. 1 classifies common fraud types and the AI methods used against them, including card-not-present fraud, account takeover, money laundering, insurance fraud and merchant fraud.

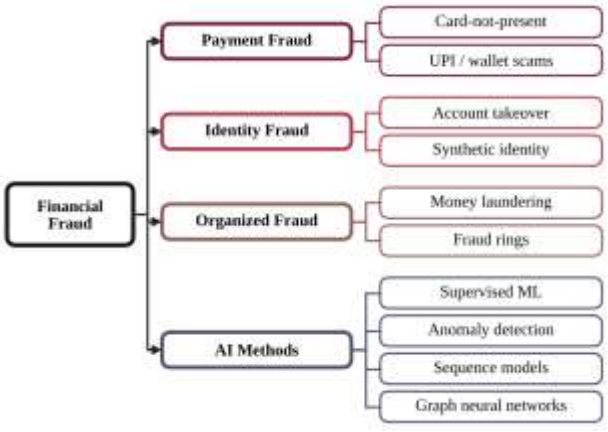


Fig. 1. Common fraud types and AI methods used against them.

2.2 Characteristics Of Fraud Data

Fraud typically makes up a tiny fraction of transactions, so a model that labels everything genuine can appear highly accurate while catching nothing. Patterns drift as fraudsters

© 2026 IJSRET

change tactics, and confirmed labels often arrive weeks later when customers dispute charges.

III. SURVEY OF EXISTING APPROACHES

Ngai *et al.* [2] classified data mining applications in financial fraud detection and found classification and anomaly detection to be the most common techniques. Bhattacharyya *et al.* [3] compared logistic regression, support vector machines and random forests on real credit card data and found random forests performed best overall.

Dal Pozzolo *et al.* [4] described a realistic fraud detection setting in which investigators can check only a limited number of alerts and feedback is delayed, and proposed a learning strategy that separately handles investigator feedback and delayed labels. Jurgovsky *et al.* [5] modelled each customer's sequence of transactions with LSTM networks, capturing behavioural context that single-transaction features miss.

Carcillo *et al.* [6] combined unsupervised outlier scores with supervised classifiers and showed improved detection, especially for new fraud patterns. Liu *et al.* [7] proposed a graph neural network approach for imbalanced fraud detection on graphs linking users and transactions, exposing coordinated fraud. Lopez-Rojas *et al.* [8] created PaySim, a simulator of mobile money transactions that provides realistic data for research where real data cannot be shared. West and Bhattacharya [9] reviewed intelligent financial fraud detection and emphasized cost-sensitive evaluation. Table I summarizes these works.

TABLE I Summary of Literature Survey

Ref.	Method	Key Findings	Limitations
[1]	Review of statistical fraud detection	Defined core challenges	Pre-modern ML
[2]	Classification framework for fraud mining	Mapped techniques to fraud types	Literature review
[3]	LR vs. SVM vs. random forest	Random forest best overall	Single dataset
[4]	Realistic modelling with feedback	Handles alert limits and delayed labels	Card data only
[5]	LSTM on transaction sequences	Captures customer behaviour	Needs long histories
[6]	Unsupervised + supervised hybrid	Better detection of new patterns	More complex pipeline
[7]	Imbalanced graph neural network	Detects coordinated fraud	Graph construction cost
[8]	PaySim mobile money simulator	Shareable realistic data	Synthetic

IV. COMPARATIVE STUDY

Table II compares methods. Rules are transparent but brittle; gradient boosting is accurate and fast on tabular data; isolation forests catch novel anomalies; sequence models capture behaviour; and graph neural networks reveal organized fraud rings.

Method	Detects	Speed	Explainability
Rules	Known patterns	Very fast	High
Gradient boosting	Learned patterns	Fast	Medium (SHAP)
Isolation forest	Novel anomalies	Fast	Medium
LSTM sequences	Behaviour changes	Medium	Low
Graph neural network	Fraud rings	Medium	Low–Medium

TABLE II Comparison of Fraud Detection Methods

V. RESEARCH GAPS

- Real transaction data is rarely shared, so many studies use small or synthetic datasets.
- Models degrade as fraud patterns drift, but adaptive retraining is seldom evaluated.
- High false-positive rates frustrate customers and overload investigators.
- Explanations for blocked transactions are often missing, which matters for customers and regulators.
- UPI and mobile wallet fraud in India is under-represented in published research.

VI. AIMS AND OBJECTIVES

1. 1) To review fraud types and the challenges of fraud data.
2. 2) To examine supervised, unsupervised, sequence and graph-based AI methods.
3. 3) To propose a real-time, layered fraud detection architecture.
4. 4) To compare methods in accuracy, speed and interpretability.

© 2026 IJSRET

5. 5) To identify future research directions.

VII. PROPOSED ARCHITECTURE

The proposed architecture is shown in Fig. 2.

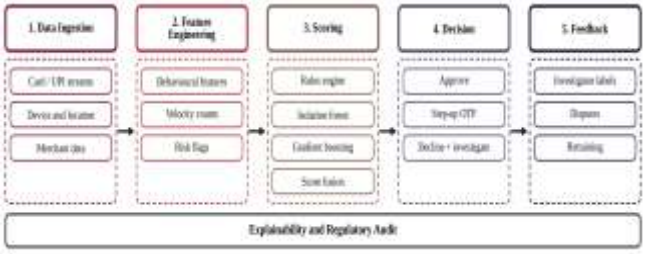


Fig. 2. Proposed real-time AI fraud detection architecture.

7.1 Data Ingestion Layer

Transaction events stream from card networks, UPI switches and online channels, enriched with device, location and merchant information.

B. Feature Engineering Layer

Features include transaction amount relative to the customer's average, time since last transaction, number of transactions in the past hour, new device or location flags and merchant risk.

7.2 Scoring Layer

A rules engine blocks known patterns, an isolation forest [10] scores unusual behaviour and a gradient-boosted model [11] estimates fraud probability. Scores are combined into a final risk score within milliseconds.

7.3 Decision Layer

Low-risk transactions are approved, medium-risk ones trigger step-up authentication such as an OTP, and high-risk ones are declined and queued for investigation. SHAP explanations [12] accompany each alert.

7.4 Feedback Layer

Investigator decisions and customer disputes are fed back as labels for regular retraining and drift monitoring.

VIII. PROCESS FLOW AND EVALUATION

Fig. 3 outlines the decision flow. Because fraud is rare, performance is evaluated with precision–recall metrics and cost-based measures. Expected cost for a threshold τ is

$$Cost(\tau) = C_FN \cdot FN(\tau) + C_FP \cdot FP(\tau) \qquad (1)$$

where C_FN is the loss from a missed fraud and C_FP the cost of reviewing or blocking a genuine transaction. The threshold is chosen to minimize this cost subject to investigator capacity.

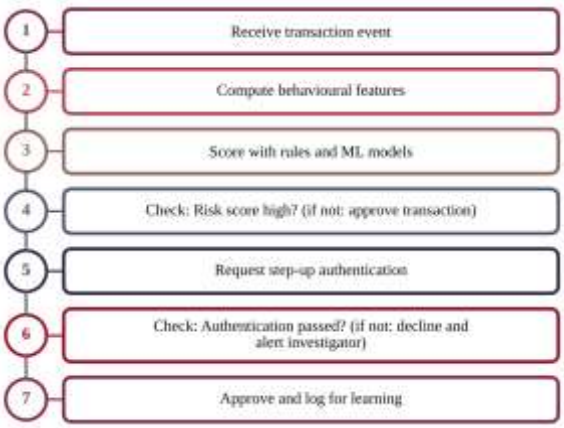


Fig. 3. Real-time fraud decision workflow.

IX. LIMITATIONS AND FUTURE SCOPE

Table III summarizes key challenges. Future work includes federated learning between banks to share fraud intelligence without sharing customer data, real-time graph analytics on UPI networks, and detection of AI-generated identity documents and voice scams.

Challenge	Mitigation
Class imbalance	Cost-sensitive learning; resampling
Concept drift	Drift monitoring; frequent retraining
False positives	Step-up authentication instead of blocking
Explainability	SHAP reasons for each alert
Data sharing limits	Federated learning; simulators

TABLE III Challenges and Mitigation

X. CONCLUSION

AI has become indispensable for detecting fraud in modern financial systems. The reviewed literature shows that ensembles such as random forests and gradient boosting perform strongly, that unsupervised and sequence models catch new and behavioural fraud, and that graph methods expose coordinated attacks. Effective systems combine these models in layers, handle imbalance and drift, explain their decisions and keep human investigators in the loop.

10.1 Acknowledgment

The authors express their sincere gratitude to the Department of Artificial Intelligence and Data Science, Jaihind College of Engineering, Pune, for providing the facilities, guidance and encouragement needed to carry out this work. The authors also thank the faculty members and

© 2026 IJSRET

colleagues whose valuable suggestions helped improve the quality of this review.

References

1. R. J. Bolton and D. J. Hand, “Statistical fraud detection: A review,” Stat. Sci., vol. 17, no. 3, pp. 235–255, 2002.
2. E. W. T. Ngai, Y. Hu, Y. H. Wong, Y. Chen, and X. Sun, “The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature,” Decis. Support Syst., vol. 50, no. 3, pp. 559–569, 2011.
3. S. Bhattacharyya, S. Jha, K. Tharakunnel, and J. C. Westland, “Data mining for credit card fraud: A comparative study,” Decis. Support Syst., vol. 50, no. 3, pp. 602–613, 2011.
4. A. Dal Pozzolo, G. Boracchi, O. Caelen, C. Alippi, and G. Bontempi, “Credit card fraud detection: A realistic modeling and a novel learning strategy,” IEEE Trans. Neural Netw. Learn. Syst., vol. 29, no. 8, pp. 3784–3797, 2018.
5. J. Jurgovsky et al., “Sequence classification for credit-card fraud detection,” Expert Syst. Appl., vol. 100, pp. 234–245, 2018.
6. F. Carcillo, Y.-A. Le Borgne, O. Caelen, Y. Kessaci, F. Oblé, and G. Bontempi, “Combining unsupervised and supervised learning in credit card fraud detection,” Inf. Sci., vol. 557, pp. 317–331, 2021.
7. Y. Liu et al., “Pick and choose: A GNN-based imbalanced learning approach for fraud detection,” in Proc. Web Conf. (WWW), 2021, pp. 3168–3177.
8. E. A. Lopez-Rojas, A. Elmir, and S. Axelsson, “PaySim: A financial mobile money simulator for fraud detection,” in Proc. 28th Eur. Model. Simul. Symp. (EMSS), 2016, pp. 249–255.
9. J. West and M. Bhattacharya, “Intelligent financial fraud detection: A comprehensive review,” Comput. Secur., vol. 57, pp. 47–66, 2016.
10. F. T. Liu, K. M. Ting, and Z.-H. Zhou, “Isolation forest,” in Proc. 8th IEEE Int. Conf. Data Min. (ICDM), 2008, pp. 413–422.
11. T. Chen and C. Guestrin, “XGBoost: A scalable tree boosting system,” in Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min., 2016, pp. 785–794.
12. S. M. Lundberg and S.-I. Lee, “A unified approach to interpreting model predictions,” in Proc. NeurIPS, vol. 30, 2017.

© 2026 IJSRET