IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940



International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 10, October 2026

DOI: 10.17148/IJARCCE.2026.151061

AI-Based Fraud Detection in Digital Payment Systems

Puja Patil[1], Shruti Pingat[2], Prof. Nikita More[3]

Student, MCA Engineering, Samarth College of Engineering and Management, Belhe, Pune, India[1]

Student, MCA Engineering, Samarth College of Engineering and Management, Belhe, Pune, India[2]

Associate Professor, MCA Engineering, Samarth College of Engineering and Management, Belhe, Pune, India[3]

Abstract: Digital payment systems such as Unified Payments Interface (UPI), mobile wallets, internet banking and card-based payments have made transactions faster and more convenient. At the same time, digital transactions can be targeted by card-not-present fraud, account takeover, phishing, social engineering, suspicious beneficiary activity and other forms of payment abuse. Traditional rule-based systems are useful for identifying known patterns, but fixed rules may struggle when fraud strategies change or when legitimate customer behaviour varies. This paper proposes an AI-based fraud detection framework that combines transaction-risk indicators, machine-learning classification and rule-based checks to identify potentially fraudulent digital payments. The framework considers features such as transaction amount, frequency, time, device or session changes, beneficiary history and deviations from a customer's usual activity, subject to lawful access and privacy safeguards. A risk-scoring component classifies transactions as low, medium or high risk and recommends approval, additional verification or temporary review. The paper presents a conceptual system design and assessment methodology rather than claiming results from an experiment. Its practical value should be evaluated using representative, appropriately protected transaction data and metrics such as precision, recall, F1-score, false-positive rate and detection latency. The proposed approach aims to support timely fraud prevention while preserving customer convenience, explainability and responsible handling of financial data.

Keywords: Artificial Intelligence (AI), Digital Payments, Fraud Detection, Machine Learning, UPI, Anomaly Detection, Transaction Risk Scoring, Financial Cybersecurity, Data Privacy.

I. INTRODUCTION

Digital payment services have become an important part of everyday financial activity. Customers use UPI applications, debit and credit cards, mobile wallets and internet banking for person-to-person transfers, purchases, bill payments and merchant transactions. These services provide speed and accessibility, but the growth of digital transactions also creates opportunities for fraudsters to exploit stolen credentials, deceptive messages, fake payment requests, compromised accounts and unusual transaction patterns. The Reserve Bank of India has highlighted digital payments, including card and internet transactions, as a significant category in reported bank frauds by number [1]. The National Payments Corporation of India also warns users about common UPI-related scams, including fake cashback links, malicious QR-code schemes and social engineering [2].

Fraud detection is challenging because fraudulent transactions are relatively rare compared with legitimate transactions, fraud patterns evolve, and a genuine customer may sometimes make an unusual but valid payment. A system that blocks too aggressively can inconvenience customers, while a system that misses suspicious activity can expose customers and payment providers to financial loss. Artificial Intelligence (AI) and machine learning can help analyse multiple transaction characteristics and identify patterns that are difficult to capture using a small set of fixed rules. However, an AI score should support—not automatically replace—appropriate verification, investigation and customer-protection procedures.

This paper outlines a conceptual AI-based fraud detection system for digital payment systems. It describes the problem, objectives, related research, proposed architecture, transaction-risk assessment workflow, evaluation approach, applications, limitations and future scope. The framework is intended as a research design; performance claims require testing with real or carefully constructed data and are not asserted in this paper.

II. LITERATURE REVIEW

Research on financial fraud detection has increasingly examined machine-learning methods because they can learn patterns from historical transactions and adapt to complex relationships among features. Hernandez Aros et al. reviewed machine-learning approaches to financial fraud detection and discussed the datasets, methods and evaluation practices used across the literature [3]. Cherif et al. systematically reviewed credit-card fraud detection research and highlighted issues such as class imbalance, feature engineering and the use of traditional and deep-learning models [4].

DOI: 10.17148/IJARCCE.2026.151061

Btoush et al. also surveyed machine- and deep-learning methods for credit-card cyber-fraud detection, identifying continuing challenges in evaluating systems under changing fraud conditions [5].

Studies of fraud detection commonly examine supervised methods, including logistic regression, decision trees, random forests and gradient-boosting models, when labelled historical transactions are available. Unsupervised or semi-supervised anomaly-detection methods may be useful when labels are incomplete or emerging fraud patterns are not yet well represented. No single algorithm is best for every payment environment: model performance depends on the quality of the data, transaction types, fraud prevalence, cost of errors and operational constraints. Accuracy alone can be misleading when the dataset contains many more legitimate transactions than fraudulent ones; precision, recall, F1-score and false-positive rate are also important.

In India, fraud prevention also depends on user awareness and secure payment behaviour. NPCI's public guidance describes common deception techniques and reminds users to be cautious with unknown links, QR codes and requests for sensitive credentials [2]. This indicates that technical detection should be combined with customer education, secure authentication and clear reporting channels. Based on these themes, the proposed framework combines risk-based machine learning with rule checks and human review rather than treating an AI prediction as a final decision in isolation.

A continuing practical challenge is balancing detection effectiveness with low friction for legitimate users. The proposed study therefore emphasises risk-tiered responses, privacy-conscious feature selection and evaluation that includes both fraud detection and customer-impact measures.

III. PROBLEM STATEMENT

Digital payment providers need to identify suspicious transactions quickly while limiting unnecessary interruption to genuine customers. Fixed rules may not detect previously unseen patterns, while highly sensitive automated systems can generate excessive false alarms. The problem addressed in this paper is to design a conceptual AI-based framework that analyses transaction behaviour, assigns a risk level and recommends a proportionate response, while maintaining privacy, explainability and the possibility of human review.

IV. OBJECTIVES

The main objectives of the proposed system are:

- To examine how AI and machine-learning techniques can support fraud detection in digital payment systems.
- To identify useful transaction-level indicators, such as amount, frequency, timing, beneficiary history and changes in device or session context where lawfully available.
- To combine model-based risk scoring with predefined security rules.
- To classify transactions into low-, medium- and high-risk categories for proportionate action.
- To reduce missed fraud while monitoring false positives and unnecessary customer friction.
- To define an evaluation approach using precision, recall, F1-score, false-positive rate and detection latency.
- To consider data privacy, model explainability, bias, security and human oversight in the design.

V. PROPOSED SYSTEM

The proposed system uses a layered workflow to assess each payment before or during processing, depending on the payment provider's architecture and permitted response time. Transaction information is first validated and transformed into a set of relevant features. A rule engine checks known high-risk conditions, while a machine-learning model estimates the probability that the transaction is fraudulent. These signals are combined into a risk score. The score is mapped to an action policy: low-risk transactions proceed through the normal payment process; medium-risk transactions may trigger additional authentication or confirmation; and high-risk transactions may be held for review or declined according to the provider's policy and applicable requirements.

The framework should not rely on a single feature or automatically assume that an unusual transaction is fraudulent. For example, a large payment may be legitimate, and a fraudster may make a small transaction to test a compromised account. Decisions should consider multiple signals, the cost of errors, the available evidence and customer-protection procedures. The system should provide a reason summary for authorised reviewers, such as an unusual transaction frequency or a new beneficiary combined with other risk indicators, without exposing sensitive detection rules to unauthorised users.

The proposed system does not itself transfer money, investigate a customer, or guarantee that all fraud will be detected. It is a decision-support framework that requires secure integration, testing, monitoring and oversight by the payment service provider.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940



International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 10, October 2026

DOI: 10.17148/IJARCCE.2026.151061

5.1 Main Components

A. Transaction Data Input: Receives permitted transaction attributes such as amount, timestamp, payment channel, transaction frequency and beneficiary-related indicators.

B. Data Validation and Pre-processing: Checks missing or inconsistent values, normalises fields and prevents duplicate or malformed records from distorting analysis. Sensitive data should be minimised, masked or tokenised where appropriate.

C. Feature Extraction: Derives indicators such as transaction velocity over a time window, deviation from a customer's usual amount range, repeated failed attempts, a new beneficiary or a change in device/session context. Only lawful and necessary data should be used.

D. Rule-Based Risk Checks: Applies explicit checks for known high-risk conditions and required security controls. Rules are complementary to—not a replacement for—the learned model.

E. AI/ML Detection Engine: Uses a trained and validated classifier, such as a random forest or gradient-boosting model, to estimate transaction risk. An anomaly-detection method may supplement the classifier when appropriate.

F. Risk Scoring and Decision Layer: Combines model output and rule signals into a risk category and applies an approved response policy.

G. Alert and Review Module: Sends flagged cases to authorised fraud analysts or appropriate verification flows and records the reason codes, action and outcome.

H. Feedback and Model Monitoring: Uses confirmed outcomes and reviewed alerts to monitor model drift and support controlled retraining. Changes must be tested and versioned before deployment.

5.2 Proposed Workflow

Transaction Initiation → Data Validation → Feature Extraction → Rule-Based Checks + AI/ML Model → Risk Score Calculation → Low / Medium / High Risk Classification → Approve / Additional Verification / Hold for Review → Outcome Logging and Monitoring.

VI. AI-BASED FRAUD RISK ASSESSMENT MODEL

The proposed risk assessment model converts transaction information into a risk category that can guide the next action. It contains five stages: (1) input assessment, where permitted transaction attributes are collected; (2) feature analysis, where behavioural and contextual indicators are calculated; (3) risk estimation, where the trained model produces a fraud-risk score; (4) policy mapping, where thresholds and rules map the score to a risk category; and (5) action and feedback, where the payment is handled and the confirmed outcome is recorded.

Illustrative risk categories:

• Low risk: The available signals are consistent with expected activity. Continue normal processing, subject to standard controls.
• Medium risk: Some indicators are unusual but do not independently establish fraud. Request step-up authentication or customer confirmation where appropriate.
• High risk: Multiple strong indicators suggest possible fraud. Apply the provider's approved hold, decline or manual-review procedure.

Thresholds must be calibrated on validation data and reviewed against the cost of false negatives and false positives. They should not be selected arbitrarily or presented as universal values. The model should produce an interpretable reason summary for authorised staff, and the provider should retain an appeal or correction path for legitimate customers affected by a mistaken flag.

VII. APPLICATIONS

1. UPI and Instant Payments: Detects unusual payment frequency, suspicious recipient patterns or activity inconsistent with a customer's normal use.

2. Debit and Credit Card Transactions: Supports identification of unusual purchase locations, rapid repeated transactions and deviations from historical spending patterns.

3. Mobile Wallets: Helps flag suspicious account access, abnormal wallet transfers and unusual top-up or withdrawal behaviour.

4. Internet Banking: Supports risk analysis for new beneficiaries, unusual transfer amounts and repeated authentication failures.



International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 10, October 2026
DOI: 10.17148/IJARCCE.2026.151061

5. Merchant Payment Monitoring: Helps payment providers identify potentially abnormal merchant transaction patterns and suspicious refund activity.

6. Fraud Analyst Support: Prioritises alerts and presents reason codes so investigators can review higher-risk cases more efficiently. The system should support analyst decisions rather than claim to replace them.

VIII. EVALUATION METHODOLOGY

A valid evaluation should use a representative dataset with transaction labels and an appropriate separation of training, validation and test data. Where possible, time-based splitting should be used so that future transactions are not inadvertently used to train a model evaluated on earlier transactions. Duplicate records and information that would not have been available at decision time should be excluded to prevent data leakage.

The model should be compared with a simple rule-based baseline. Recommended measures include: (i) precision, the proportion of flagged transactions that are confirmed fraud; (ii) recall, the proportion of known fraudulent transactions detected; (iii) F1-score, a combined measure of precision and recall; (iv) false-positive rate, the proportion of legitimate transactions incorrectly flagged; (v) PR-AUC where class imbalance is substantial; and (vi) detection latency and operational workload. Results should also be examined across payment channels and relevant user groups, subject to privacy and fairness requirements.

No experimental dataset or measured performance results were supplied for this draft. Therefore, no accuracy percentage or claim of superior performance is reported. The framework should be implemented and evaluated before conclusions about effectiveness are drawn.

IX. SECURITY, PRIVACY AND LIMITATIONS

The system may process sensitive financial and behavioural information. Data collection should be limited to what is necessary, protected in transit and at rest, and accessible only to authorised personnel. Retention periods, audit logs, access controls and incident-response procedures should be defined. Model explanations should be useful to authorised reviewers without revealing exploitable security details to attackers.

Important limitations include incomplete or delayed fraud labels, evolving fraud strategies, class imbalance, false alerts, data quality issues and possible bias caused by historical data. A model trained on one provider or payment channel may not transfer reliably to another. AI can also be manipulated or affected by adversarial behaviour. Human oversight, monitoring, periodic testing and a clear process for customers to challenge incorrect decisions remain necessary.

X. CONCLUSION

This paper presented a conceptual framework for AI-based fraud detection in digital payment systems. The proposed approach combines transaction feature analysis, rule-based checks, machine-learning risk estimation and risk-tiered responses. It is designed to support timely identification of suspicious activity while considering the impact of false alerts on genuine customers. The review of related work suggests that machine-learning methods are widely studied for financial fraud detection, but their performance depends on data quality, fraud prevalence, model choice and evaluation design. The framework therefore recommends testing against a rule-based baseline and reporting precision, recall, F1-score, false-positive rate and detection latency rather than relying on accuracy alone. Since no empirical transaction data were analysed for this draft, the proposal should be treated as a design framework, not as a validated production system. Responsible deployment requires data protection, explainability, customer safeguards and ongoing human supervision.

XI. FUTURE SCOPE

Future work may implement a prototype using a suitable anonymised transaction dataset and compare algorithms such as logistic regression, random forest and gradient boosting. Further research could examine streaming detection, graph-based analysis of connected accounts, privacy-preserving learning, explainable AI, concept-drift monitoring and cost-sensitive threshold selection. A practical evaluation could also measure customer friction, analyst workload and the time between a suspicious transaction and detection. Any deployment should follow applicable law, payment-provider requirements and responsible data-governance practices.

ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to Prof. P. J. Gopale, research guide at Samarth Engineering, Management and Polytechnic College, Belhe, for valuable guidance, encouragement and support during the preparation of this research paper. The authors are also thankful to the college for providing the opportunity and support to carry out this work.



International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 10, October 2026

DOI: 10.17148/IJARCCE.2026.151061

REFERENCES

[1] Reserve Bank of India, "Annual Report 2024–25," Chapter VI: Regulation, Supervision and Financial Stability, May 29, 2025.

[2] National Payments Corporation of India (NPCI), "Fraud Awareness," NPCI official website.

[3] L. Hernandez Aros, L. X. Bustamante Molano, F. Gutierrez-Portela, J. J. Moreno Hernandez, et al., "Financial fraud detection through the application of machine learning techniques: a literature review," Humanities and Social Sciences Communications, vol. 11, article 1130, 2024, doi: 10.1057/s41599-024-03606-0.

[4] A. Cherif, A. Badhib, H. Ammar, S. Alshehri, M. Kalkatawi, and A. Imine, "Credit card fraud detection in the era of disruptive technologies: A systematic review," Journal of King Saud University—Computer and Information Sciences, vol. 35, no. 1, pp. 145–174, 2023, doi: 10.1016/j.jksuci.2022.11.008.

[5] E. A. L. M. Btoush, X. Zhou, R. Gururajan, K. C. Chan, R. Genrich, and P. Sankaran, "A systematic review of literature on credit card cyber fraud detection using machine and deep learning," PeerJ Computer Science, vol. 9, e1278, 2023, doi: 10.7717/peerj-cs.1278.

[6] A.-S. Yussiff et al., "The best machine learning model for fraud detection on e-platforms: a systematic literature review," Computer Science and Information Technologies, vol. 5, no. 2, pp. 195–204, 2024, doi: 10.11591/csit.v5i2.pp195-204.