

AI-Based Fraud Detection in Digital Payment Systems

Junaid Akhtar¹, Bibek Mohanty², Himadri Deep³, Nrupati Keshari Beher⁴, Sucharita Nathsharma⁵

^{1,2,3,4}UG, Computer Science Engineering, Radhakrishna Institute of Technology and Engineering, Bhubaneswar, Odisha, India.

⁵Associate Professor, Computer Science Engineering, Radhakrishna Institute of Technology and Engineering, Bhubaneswar, Odisha, India.

Email Id: akhtarjunaid82099@gmail.com¹, bibekmohanty326@gmail.com², himadrideep47@gmail.com³, nrupatikesharib@gmail.com⁴, sucharita.nathsharma73@gmail.com⁵

Abstract

Recent developments in research highlight the importance of electronic payment methods in contemporary financial practice. Nevertheless, fraudulent transactions, phishing, identity theft, takeovers, card not present, and other irregularities undermine the system's efficiency. Notably, rule-based systems for detecting fraud are inappropriate in the present circumstances, given the complexity of fraudulent activities. As such, the paper proposes an AI-based method to detecting fraud in the electronic payment system. The technique relies on machine learning, mining, and behavioural analytics algorithms to detect and avert fraudulent transactions. The concept described in the paper can examine the transaction's features, including amount, frequency, locations, devices, merchants, time, and behavioural patterns, to detect fraud. The framework can be implemented to enhance the performance of digital payment fraud detection and response systems. Sure enough, in practice, the designed framework would assess the risk score of the transaction and based on the results, approve, notify, request additional information, defer or reject the payment.

Keywords: AI, Fraud Detection, Digital Payment Systems, Machine Learning, Anomaly Detection, Behavioral Analytics.

1. Introduction

Digital payments have changed the way value traditionally used to be transferred: India's UPI processed tens of billions of payments for banks and other payment service providers (NPCI, 2026). Consequently, analysing payment operations to identify and neutralize fraudulent transactions has become one of the most pressing concerns for the banking sector, where dozens of fraudulent transactions on average lead to the unauthorized transfer of hundreds of millions of rupees. Digital payments fraud, in particular card/internet fraud is the largest category of bank fraud by volume for 2024-25 according to RBI (RBI, 2025). Therefore, the paper will propose a conceptual framework for detecting fraud in digital payments. The framework will utilise transaction features, behavioural patterns, supervised machine learning, detection rules, and risk scoring to define a responsive framework that would either allow the transaction to proceed or decline it with various options in between [1].

2. Background and Related Work

Fraud detection analytics typically involve either supervised classification (for labelled datasets) or unsupervised anomaly detection algorithms (for unlabelled datasets) (Hernandez Aros et al., 2024). Random Forest applies randomly pruned trees to define transactions' characteristics interactions (scikit-learn developers, 2026a), whereas Isolation Forest employs randomly partitioned data to identify uncommon transactions isolated by partitions, with more splits indicating anomalies (scikit-learn developers, 2026b). Behavioural analytics offer a set of user-specific expected characteristics; the BIS project Hertha highlights the ability to combine them (to some extent) with traditional transaction fraud detection analytics while emphasizing the need for labelled data, feedback loops, explainability, and privacy (BIS, 2025). Notably, multiple algorithms mentioned allow interpreting the results using SHAP and LIME (Gotadke, 2025).

3. Problem Statement

Digital payments systems require extremely high transaction processing rates with the need to separate legitimate transactions from fraudulent ones with minimum disruption to customers. Fraud-detection systems cannot use traditional binary rules due to the extreme scale of transactions, and fraud cases keep changing over time and with different users. What is needed is a system which takes into account both traditional fraudulent-transaction characteristics and behavioural anomalies and implements graduated response options from simple approval to transaction suspension.

4. Proposed AI-Based Fraud Detection Framework

Figure 1 below shows a conceptual system framework which can be used for fraud detection. In it, a request for a transaction is analyzed in terms of features relevant to fraud detection. The features are then used by supervised classification, anomaly detection, and behavioural analytics components to generate signals which can be combined into a conceptual risk score to define an appropriate action.

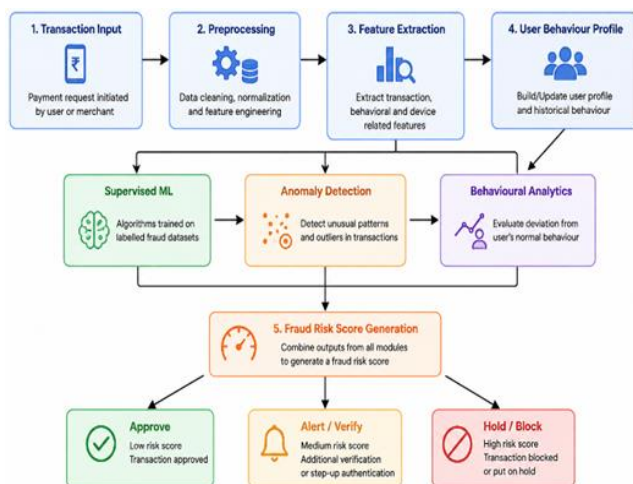


Figure 1 Proposed AI-Based Fraud Detection Architecture

Table 1 below shows general categories of digital-payment fraud which can be considered by the system as well as general indicators which can be used for their detection.

- Table 1. Types of Digital Payment Fraud and Detection Indicators.
- Fraud type Possible indicators

- Unauthorized transfer Uncharacteristic amount, recipient, date, frequency
- Phishing-assisted fraud New device, location, context
- Identity theft Inconsistent profile, device, location
- Account takeover Significant behavioural changes, device, location
- Card-not-present fraud Unusual merchant, amount, purchase pattern [2]

Table 1 Types of Digital Payment Fraud and Detection Indicators

Fraud type	Possible indicators
Unauthorized transfer	Unusual amount, recipient, timing, frequency
Phishing-assisted fraud	New device, unusual location, unexpected context
Identity theft	Profile inconsistency, unfamiliar device/location
Account takeover	Rapid changes in behaviour, device, location
Card-not-present fraud	Unusual merchant, amount, purchase pattern

5. System Architecture and Data Flow

The system follows a general data-flow architecture, with Figure 2 below showing a conceptual input of a request for a transaction and an output action. It is assumed that no single indicator is sufficient for fraud detection, with individual features being combined into a feature vector and analyzed by supervised, unsupervised, and behavioural analytics components.

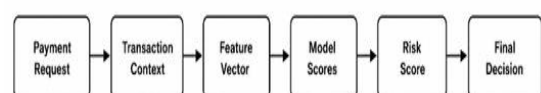


Figure 2 Transaction Data Flow from Input to Final Decision

6. Methodology

An experimental validation of the proposed system would require a labelled transaction data set with details on whether a transaction was fraudulent and

relevant statistics for frequency, amount, location, devices, time of transaction, and user behavioural patterns, as described by Table 2 below [3].

Table 2 Important Transaction Features

Feature	Purpose in detection
Amount	Identifies unusually large/atypical values
Frequency	Captures sudden increases in activity
Location / Device	Highlights unusual geography or device
Merchant type	Detects unusual merchant-category use
User behaviour	Measures deviation from history

Labelled data would need to be split into training and test sets with a chronological order to avoid leakage. Note that fraud detection analytics typically operate with imbalanced data, so class-weighting and resampling techniques would need to be described and experimentally validated. Figure 3 below shows a workflow which would be used for the experiments, while Figure 4 below shows a general machine-learning training workflow, including model selection and validation.

- Table 2. Important Transaction Features.
- Feature Purpose in detection
- Amount Allow identification of unusual values
- Frequency Detect sudden value increases
- Location/device Detect unusual location/device
- Merchant type Detect unusual merchant categories
- User behaviour Allow identification of deviations

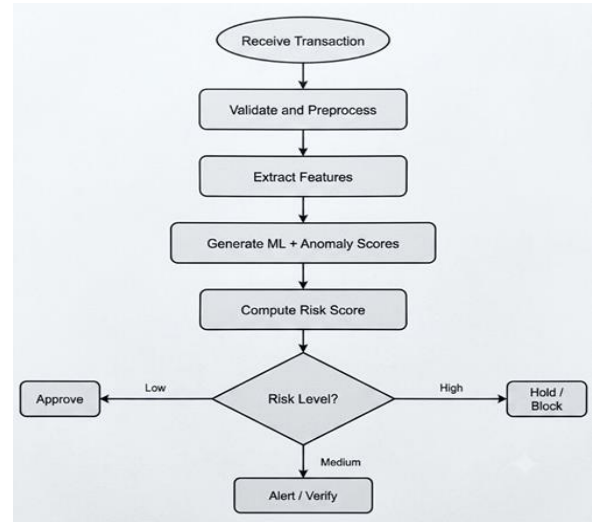


Figure 3 Flowchart of the Proposed Fraud-Detection Processing

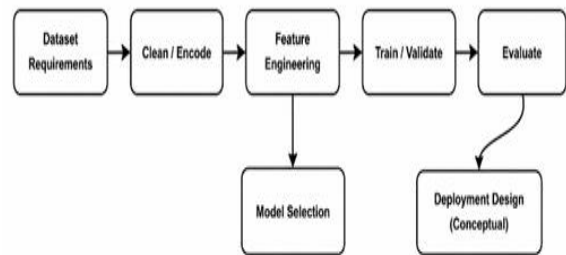


Figure 4 Machine-Learning Workflow for the Proposed Framework

7. Algorithms and Models

Logistic Regression can be used as a simple and interpretable supervised model, while Random Forest can be used for analysis of nonlinear interactions between features (scikit-learn developers, 2026a). Isolation Forest is another option for an anomaly detection model which can isolate uncommon transactions by using randomly chosen partitions (scikit-learn developers, 2026b). Table 3 below compares possible supervised, anomaly detection, and unsupervised models for fraud detection purposes. The risk score described in the conceptual framework can take the form of $R = w_1P_f + w_2A + w_3B$, with weights defining the relative importance of each score, serving as the basis for actions taken. Figure 5 below shows how such a score can be used to select the action based on a risk threshold, while Table 4 below describes evaluation metrics which

can be used for model performance assessment.

- Table 3. Comparison of Possible Machine-Learning Models.
- Model Type Role
- Logistic Regression Supervised Baseline
- Random Forest Supervised Primary candidate
- Isolation Forest Anomaly Anomaly component
- Autoencoder Unsupervised Alternative

Table 3 Comparison of Possible Machine-Learning Models

Model	Type	Role
Logistic Regression	Supervised	Baseline
Random Forest	Supervised	Primary candidate
Isolation Forest	Anomaly	Anomaly component
Autoencoder	Unsupervised	Alternative

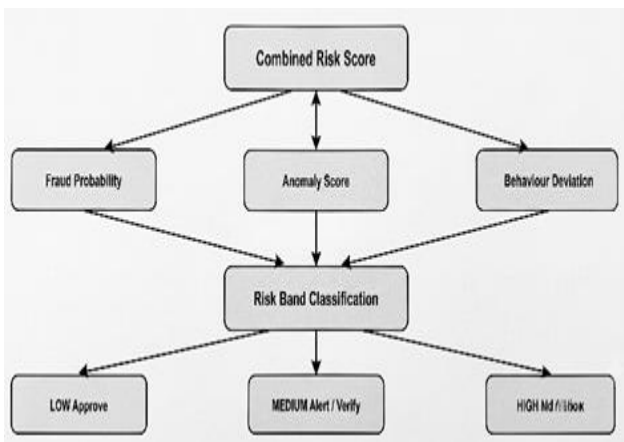


Figure 5 Conceptual Fraud Risk-Score Decision Mechanism

Table 4 Evaluation Metrics and Their Meaning

Metric	Meaning
Precision	Correctness of fraud alerts
Recall	Share of fraud cases detected
F1-score	Balance of precision & recall
False-positive rate	Legitimate cases wrongly flagged

8. Workflow of the Proposed System

The workflow would start with a request for a payment followed by validation and conversion into features used for fraud detection. A supervised model would estimate the probability of fraud, while anomaly detection and behavioural analytics score deviations from normal parameters, with the signals combined into a risk score defining the response to be taken. Low-risk transactions would be authorized, while medium-risk transactions could be subjected to additional checks, and high-risk transactions could be suspended with further actions being taken based on the situation. The options correspond to the need for a graduated response as discussed in the conceptual framework and the approved abstract [4].

9. Implementation Design

The following design considers a future prototype implementation and thus reflects a potential design rather than an actual production system. A frontend component would allow the user to view the status of their transaction, while a backend API would accept transaction requests with relevant information and a database storing transaction requests, user data, device fingerprints, and alerts can be considered, as shown by Figure 7 below. The proposed system will be implemented using different software languages such as Python, software application development frameworks such as Flask and FastAPI, databases such as PostgreSQL and MySQL, and machine learning algorithms such as SKLearn among others. The figure below is an optional use-case diagram of the proposed system shown in figure 6.

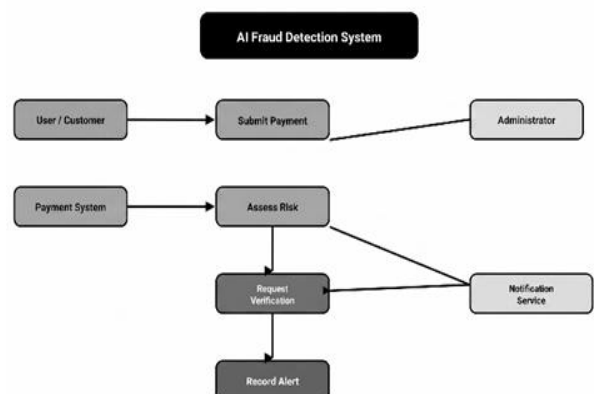


Figure 6 Optional Use-Case View of the Proposed System

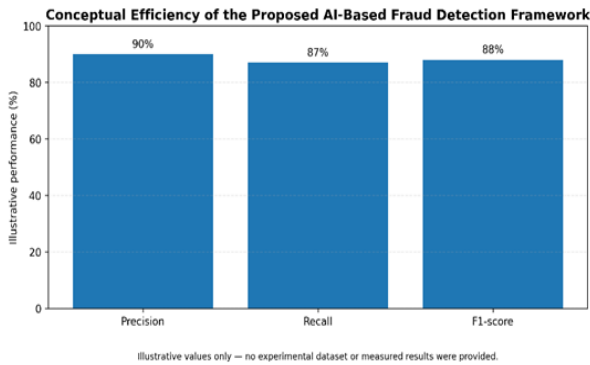


Figure 7 Conceptual Data Entities for the Proposed Framework

10. Results and Discussion

No results are presented because no labelled data, machine-learning models, or evaluation were provided with the approved abstract; thus, the discussion only considers expected results based on the general methodology and model evaluation. The framework would allow for more flexible fraud detection than a traditional binary approach suggested by related works. The evaluation would show if the combination of approaches would be beneficial, with precision, recall, F1-score, and false-positive rate reported for baseline and combined approaches. Table 5 below shows some of the main advantages and limitations of the proposed design.

- Table 5. Advantages and Limitations of the Proposed Framework [5].
- Advantages Limitations
- Combine multiple signals for fraud detection Require calibration on real data
- Allow graduated response options May require manual intervention
- Can complement traditional rules May involve false positives

Table 5 Advantages and Limitations of the Proposed Framework

Advantages	Limitations
Combines multiple fraud signals	No experimental validation here
Supports risk-based actions	Requires calibration on real data
Can complement fixed rules	May generate false positives

11. Security, Privacy, And Ethical Considerations

Fraud detection system would require processing of sensitive data. Therefore, the data must be minimized, protected and encrypted. BIS Project Hertha notes that legal, privacy, and regulatory aspects in practice should be taken into consideration before such a system is deployed in production (BIS, 2025). Fraud-detection systems should be designed to avoid false positives, as they disrupt legitimate transactions. When automated decisions are made for transactions, they should follow proportional response options and be subject to auditing to avoid discrimination based on protected factors [6].

12. Future Scope

The future scope includes development of system for stream processing which would process the data in real-time, applying techniques such as federated learning to preserve the data, providing explanation of the model using SHAP or LIME. The model would require regular training and retraining to adapt to concept drift. Use of blockchain could be considered to provide audit trail. Future work can also consider alternative approaches, such as graph-based analytics, cost-sensitive learning, and private machine learning, while reporting experimental results on well-defined benchmarks can be used to support any claims.

Conclusion

The paper proposed a conceptual framework for an AI-based fraud-detection system for digital payments which combines transaction features, behavioural analytics, supervised learning, and anomaly detection methods to implement a graduated response option for suspicious transactions beyond simple binary fraud/approval rules. The paper is primarily conceptual in nature, as no labelled data set was provided for experimentation; thus, future work should focus on implementation of the proposed machine-learning pipeline, its comparison to traditional fraud-detection rules, and evaluation in terms of precision, recall, F1-score, and false-positive rate.

Acknowledgements

We would like to thank Prof. Sucharita Nathsharma for her guidance in this proposed research work. The authors declare no external funding for this study.

Acknowledgements

We would like to thank Prof. Sucharita Nathsharma for her guidance in this proposed research work. The authors declare no external funding for this study.

References

- [1]. Bank for International Settlements. (2025, June 5). Project Hertha: Identifying financial crime patterns in real-time retail payment systems. BIS Innovation Hub.
- [2]. Gotadke, T. (2025). A machine learning framework for financial fraud detection using explainable AI techniques. *Int. J. Computer Sciences and Engineering*, 13 (5), 1725-1734.
- [3]. Hernandez Aros, L. et al. (2024). Financial fraud detection through machine learning: A literature review. *Humanities and Social Sciences Communications*, 11, 1130.
- [4]. National Payments Corporation of India. (2026). UPI product statistics. npci.org.in
- [5]. Reserve Bank of India. (2025). Annual report 2024-25: Regulation, supervision and financial stability.
- [6]. scikitlearn developers. (2026a). Random forests. *scikitlearn user guide*.