



FRAUD TRANSACTION DETECTION

Mrs.Vhadhankar P.S.¹, Muktabai Nagrale²

¹ Faculty Guide & Supervisor | ² Dattakala Group Of Institutions Swami - Chincholi tal. Daund dist. Pune, India

Corresponding author: muktainagrale@gmail.com

ARTICLE ID IJREK-2026-00156	ARTICLE TYPE Research Article	RESEARCH AREA Computer Science	PUBLICATION Volume 1, Issue Issue 1 (2026)
RECEIVED 23 October 2026	REVISED 29 October 2026	ACCEPTED 26 October 2026	PUBLISHED 29 October 2026

Abstract

Digital payments, online shopping, mobile banking, and card-not-present transactions have increased the speed and convenience of financial services. These channels also create opportunities for fraudulent activity, including stolen-card use, account takeover, identity misuse, and suspicious transaction patterns. This paper presents a proposed fraud transaction detection framework that combines transaction-data preparation, rule-based checks, machine-learning risk scoring, and human review. Transaction attributes such as amount, time, merchant category, location consistency, and recent activity can be transformed into features for a classification model. The system assigns a risk score and routes high-risk transactions for additional verification instead of treating every unusual transaction as confirmed fraud. The paper outlines the architecture, workflow, candidate algorithms, evaluation measures, privacy safeguards, and deployment considerations. Since no institution-specific dataset or completed experiment was supplied, performance values are not claimed; the model must be evaluated on representative, appropriately protected data before operational use.

Keywords: Fraud Detection; Transaction Monitoring; Machine Learning; Anomaly Detection; Financial Security; Classification; Risk Scoring

2. PROBLEM STATEMENT AND OBJECTIVES

The central problem is to identify potentially fraudulent transactions from a large stream of legitimate and suspicious activity while keeping false alerts manageable. A useful solution should support early risk assessment, provide reasons for a flag where possible, and allow analysts to review uncertain cases. 2.1 Objectives • Design a modular workflow for ingesting and validating transaction records. • Identify useful transaction-level and behaviour-based features. • Describe a hybrid approach combining business rules and machine-learning scores. • Define an evaluation plan suitable for imbalanced and time-dependent data. • Address privacy, fairness, security, and human oversight. 2.2 Scope The proposed framework is intended as a decision-support component for banks, payment gateways, or e-commerce platforms. It is not a complete banking system and does not replace authentication, payment authorization, regulatory controls, or fraud-investigation procedures. Model outputs should be treated as risk signals rather than proof of criminal activity.

3. BACKGROUND AND RELATED WORK

Traditional fraud controls commonly use deterministic rules, such as unusually large transactions, repeated declines, rapid purchases, or inconsistent device and location signals. Rules are straightforward to audit but need continuous maintenance and may be evaded when attackers change behaviour. Statistical anomaly detection identifies records that differ from established patterns, but legitimate unusual activity can also appear anomalous. Supervised classification learns from transactions labelled as fraud or legitimate. Logistic regression can provide a relatively interpretable baseline; decision trees and random forests capture nonlinear relationships; gradient-boosted trees often perform well on structured tabular data. Unsupervised methods may help surface emerging patterns when reliable labels are limited. No single method is universally best; performance depends on data quality, label delay, fraud prevalence, feature availability, and operating constraints. Prior research on credit-card fraud detection emphasizes class imbalance, the need for appropriate metrics, and the risks of evaluating models with unrealistic random splits. For transaction streams,

chronological validation is often more informative because it tests whether a model trained on earlier records generalizes to later behaviour.

4. PROPOSED SYSTEM OVERVIEW

The system accepts transaction details, validates and transforms them, applies deterministic controls and a trained risk model, then routes the decision to an appropriate action. Low-risk activity may proceed under existing authorization rules; medium-risk activity may require step-up verification; high-risk activity may be held for review, subject to the payment provider's policy.

5. SYSTEM ARCHITECTURE

Figure 1 shows the proposed high-level architecture. It separates ingestion, preprocessing, rules, machine-learning inference, decisioning, and analyst feedback so that each module can be tested and monitored independently.

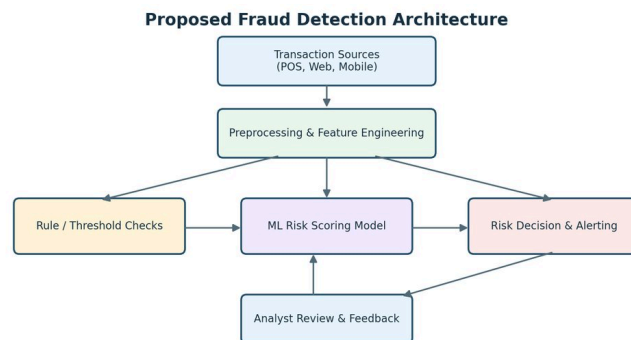


Figure 1. Proposed modular architecture for fraud transaction detection.

5.1 MAIN COMPONENTS

Transaction ingestion: Receives transaction events from an authorized payment source and assigns a unique event identifier. Preprocessing: Checks required fields, handles missing values, normalizes formats, and derives time-window features.

Rules engine: Applies transparent policy checks and known indicators. Risk model: Produces a score using features available at decision time. Decision service: Maps score bands and rules to allow, verify, hold, or review actions. Feedback loop: Stores investigator outcomes and confirmed fraud labels for later evaluation and retraining.

5.2 EXAMPLE FEATURE GROUPS

Potential features include transaction amount relative to the customer’s normal range; transaction count within recent time windows; time since the previous transaction; merchant category; device or account age; payment channel; country or region mismatch; and prior confirmed chargebacks. Features should be selected only when lawfully available and operationally justified. Sensitive personal attributes should not be included simply because they are available.

5.3 DECISION PRINCIPLES

A risk score is not a final legal or financial determination. Thresholds should be selected using validation data and the cost of each action. The system should provide an explanation code, preserve an audit trail, and support an appeal or manual review path where appropriate.

6. DETECTION WORKFLOW

The workflow begins when a transaction is received and ends with a recorded decision. The diagram illustrates a simplified decision path; production deployments must integrate with the institution’s actual authorization and incident-response procedures.

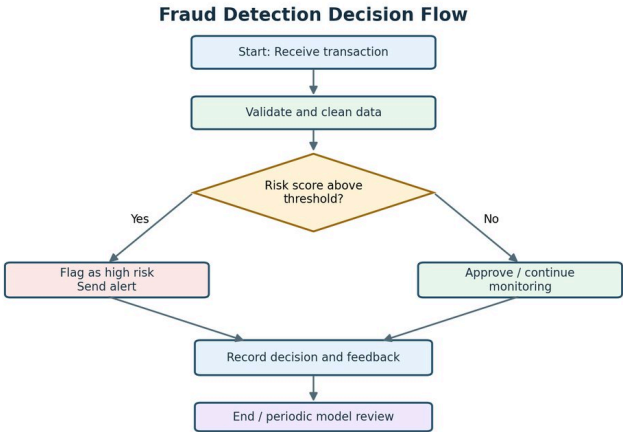


Figure 2. Flowchart of transaction validation, risk assessment, and action routing.

6.1 STEP-BY-STEP OPERATION

- Receive the transaction and verify schema, timestamp, and required identifiers.
- Transform raw fields into model-ready features without using information unavailable at decision time.
- Run known fraud rules and obtain a model-generated risk score.
- Combine signals according to a documented policy and select an action band.
- Record the decision, explanation code, model version, and later confirmed outcome.

7.1 DATA PREPARATION

The dataset should contain timestamped transactions and trustworthy outcome labels, with a documented definition of fraud and the time at which labels become available. Duplicate events, missing values, inconsistent currencies, and impossible timestamps should be handled systematically. Personally identifiable fields should be removed, tokenized, or access-restricted when not essential. Any public or synthetic dataset must be described accurately and should not be presented as a live financial institution dataset.

7.2 FEATURE ENGINEERING

Features can summarize transaction behaviour over fixed windows, such as count and total value over the last few minutes or hours. Aggregations must be calculated using past data only. Customer-level histories should be built in a way that prevents

information leakage from future transactions or from labels that would not have been known at the time of scoring.

7.3 MODEL DEVELOPMENT

Begin with a transparent rule-based baseline and a simple supervised classifier such as logistic regression. Compare with a tree-based model if the dataset and compute budget permit. Use class weighting or carefully designed sampling only within training data. Calibrate predicted probabilities if they are used as risk estimates. Select the operating threshold on validation data based on an explicit trade-off between missed fraud, false alerts, customer friction, and review capacity.

7.4 VALIDATION DESIGN

Split data chronologically into training, validation, and test periods. Keep records from the same transaction chain or duplicate event from leaking across splits. Report metrics on an untouched test period and, where possible, repeat evaluation across multiple time windows. If labels arrive late, define a maturity window so that transactions are not incorrectly treated as legitimate merely because a fraud report has not yet arrived.

7.5 IMPLEMENTATION STACK

Table 1 summarizes example features and why they may be useful. Actual feature availability varies by payment

Feature	Example signal	Purpose
Amount deviation	Amount far above recent typical spend	Identify unusual transaction value
Velocity	Many attempts in a short window	Detect rapid repeated activity
Time pattern	Activity at an unusual hour	Add contextual risk signal
Device / channel	New device or channel change	Support account-takeover screening
Location consistency	Unexpected geographic change	Flag for additional verification
Historical outcomes	Prior confirmed fraud or chargeback indicators	Use past evidence when lawful and reliable

Table 1. Candidate feature groups for transaction risk assessment.

Risk band	Illustrative response	Safeguard
Low	Continue through normal authorization checks	Continue monitoring; do not bypass baseline controls
Medium	Request step-up authentication or additional confirmation	Offer accessible verification alternatives
High	Hold or route for manual review under policy	Avoid automatic irreversible action from model score alone

8.1 ACTION MAPPING

Risk band Illustrative response Safeguard
Low Continue through normal authorization checks Continue monitoring; do not bypass baseline controls
Medium Request step-up authentication or additional confirmation Offer accessible verification alternatives
High Hold or route for manual review under policy Avoid automatic irreversible action from model score alone

Table 2. Example risk-response mapping; thresholds and actions require institution-specific approval.

Metric	Meaning	Why it matters
Precision	Share of flagged cases that are fraud	Estimates analyst workload and false alarms
Recall	Share of known fraud detected	Measures missed-fraud risk
PR-AUC	Precision-recall performance across thresholds	Useful for highly imbalanced labels
False-positive rate	Legitimate transactions incorrectly flagged	Tracks customer friction
Latency	Time required to score a transaction	Checks suitability for real-time use

8.2 EVALUATION METRICS

Metric Meaning Why it matters Precision Share of flagged cases that are fraud Estimates analyst workload and false alarms Recall Share of known fraud detected Measures missed-fraud risk PR-AUC Precision-recall performance across thresholds Useful for highly imbalanced labels False-positive rate Legitimate transactions incorrectly flagged Tracks customer friction Latency Time required to score a transaction Checks suitability for real-time use Table 3. Recommended evaluation measures. Accuracy alone is insufficient when fraud cases are rare.

9.1 DATA PROTECTION

Transaction records can reveal sensitive financial behaviour. Collect only necessary fields, restrict access by role, encrypt data in transit and at rest, rotate credentials, and define retention and deletion policies. Development datasets should be de-identified where possible. Logs must avoid storing full card numbers, authentication secrets, or unnecessary personal details.

9.2 MODEL AND SYSTEM SECURITY

Protect scoring endpoints against unauthorized access, replay, malformed inputs, and denial-of-service attacks. Validate requests, rate-limit appropriately, sign or authenticate internal service calls, and maintain versioned model artifacts. Monitor for data drift, changes in fraud patterns, unusual score distributions, and increases in false alerts. A safe fallback should exist if the model or feature service becomes unavailable.

9.3 FAIRNESS AND EXPLAINABILITY

Fraud controls can affect customers differently if proxies reflect uneven data coverage or historical investigation bias. Review error rates across relevant groups only where lawful and ethically justified. Prefer explainable reason codes for customer support and analyst review. Do not use protected characteristics as shortcuts for risk. Provide a mechanism for correcting data errors and reviewing disputed decisions.

9.4 HUMAN OVERSIGHT

Analysts should be able to inspect supporting signals, record dispositions, and escalate uncertain cases. Confirmed outcomes should be reviewed before they become training labels. The system should document who changed thresholds, when model versions were deployed, and why an alert was resolved. Automated actions should be proportionate to risk and reversible where possible.

10. EXPECTED OUTCOMES AND LIMITATIONS

The expected deliverable is a modular prototype and an evaluation plan that make fraud-risk assessment more systematic. Actual detection performance, throughput,

and business impact can only be established after testing with representative data. The system may miss new fraud patterns, generate false alerts, or perform poorly when labels are delayed or features change. An attacker may adapt to known controls, so monitoring and periodic review are necessary. This paper does not report experimental accuracy, financial savings, or a production deployment because no completed dataset-based experiment was provided. Future work should measure precision, recall, PR-AUC, false-positive rates, scoring latency, and analyst workload using a clearly documented evaluation protocol.

11. CONCLUSION AND FUTURE WORK

This paper proposes a hybrid framework for fraud transaction detection using transaction validation, feature engineering, transparent rules, machine-learning risk scoring, and human review. The architecture separates the scoring components from decision policy and feedback, making the system easier to test and monitor. A reliable evaluation must account for class imbalance, delayed labels, changing fraud behaviour, and the real cost of false positives and false negatives. Privacy, explainability, access control, and analyst oversight are essential parts of a responsible design. Future work should implement the prototype, evaluate it on a suitable chronological dataset, compare baseline and advanced models, and test whether the system meets operational latency and review-capacity requirements.

REFERENCES

- Bolton, R. J., and Hand, D. J. (2002). Statistical fraud detection: A review. *Statistical Science*, 17(3), 235–255. <https://doi.org/10.1214/ss/1042727940>
- Phua, C., Lee, V., Smith, K., and Gayler, R. (2010). A comprehensive survey of data mining-based fraud detection research. *arXiv:1009.6119*. <https://arxiv.org/abs/1009.6119>
- Whitrow, C., Hand, D. J., Juszczak, P., Weston, D., and Adams, N. M. (2009). Transaction aggregation as a strategy for credit card fraud detection. *Data Mining and Knowledge Discovery*, 18, 30–55. <https://doi.org/10.1007/s10618-008-011>
- Jurgovsky, J., Granitzer, M., Ziegler, K., Calabretto, S., Portier, P.-E., He-Guelton, L., and Caelen, O. (2018). Sequence classification for credit-card fraud detection. *Expert Systems with Applications*, 100, 234–245. <https://doi.org/10.1016/j.eswa.2018.01.037>
- Vesselinov, R., and others. (2021). Machine learning approaches for credit card fraud detection: A review. (Use as a search lead; verify full bibliographic details before formal submission).
- Carcillo, E., Le Borgne, Y.-A., Caelen, O., and Bontempi, G. (2019). Streaming active learning strategies for real-life credit card fraud detection: Assessment and visualization. *International Journal of Data Science and Analytics*, 8, 285–300.
- Dal Pozzolo, A., Boracchi, G., Caelen, O., Alippi, C., and Bontempi, G. (2015). Credit card fraud detection: A realistic modeling and a novel learning strategy. *IEEE Transactions on Neural Networks and Learning Systems*. <https://doi.org/10.1109/TNNLS.2014.2364934>
- European Banking Authority. (2024). Report on payment fraud. European Banking Authority. <https://www.eba.europa.eu/>
- National Institute of Standards and Technology. (2023). Artificial Intelligence Risk Management Framework (AI RMF 1.0). NIST AI 100-1. <https://doi.org/10.6028/NIST.AI.100-1>
- PCI Security Standards Council. (2024). Payment Card Industry Data Security Standard (PCI DSS), version 4.0.1. <https://www.pcisecuritystandards.org/>

Publication archival note: Official scholarly contribution published under CC BY 4.0 International License. Permanent record on <https://www.ijrek.org> | Article ID: IJREK-2026-00156