

DOI: <https://doi.org/10.64672/IJIFR/26.04.13.08.040>

PUBLISHED ON: APRIL 20, 2026

SECUREPAY SHIELD: AN ENSEMBLE MACHINE LEARNING ARCHITECTURE FOR REAL-TIME FRAUD DETECTION AND PROBABILISTIC RISK SCORING IN THE INDIAN DIGITAL PAYMENT ECOSYSTEM

Jeripiti Reddy Prasad ¹, V.Vijayalakshmi ²¹M.C.A. Student, ² Assistant Professor^{1,2} Department of Computer Applications,

Viswam Engineering College, Madanapalle, Andhra Pradesh, India

ABSTRACT

The exponential growth of digital financial transactions in India has precipitated a commensurate escalation in sophisticated payment fraud, necessitating intelligent, adaptive detection systems capable of operating at scale. This paper presents SecurePay Shield, a comprehensive, end-to-end machine learning pipeline engineered for real-time identification of fraudulent transactions within the Indian digital payment ecosystem. The proposed system addresses three principal challenges endemic to fraud detection: severe class imbalance, high-dimensional heterogeneous feature spaces, and the requirement for probabilistic, interpretable risk assessments amenable to regulatory scrutiny. The architecture employs an ensemble learning strategy integrating three complementary algorithms: a Random Forest Classifier (200 estimators), a Gradient Boosting Classifier (150 estimators), and an Isolation Forest anomaly detection model. A domain-specific feature engineering pipeline transforms 23 raw transaction attributes into a 35-dimensional feature space, computing composite risk indicators including the Risk_Composite score, IP_Risk_Score, and Velocity_Score, which collectively emerge as the most discriminative predictors of fraudulent activity. Class imbalance is mitigated through the application of the Synthetic Minority Over-sampling Technique (SMOTE), yielding a balanced training corpus of 22,560 instances. The Random Forest model, selected as the production deployment candidate, achieves an Area Under the Receiver Operating Characteristic Curve (AUC-ROC) of 1.0000 and an F1-Score of 1.0000 on the held-out test partition, with five-fold stratified cross-validation yielding a mean F1 of 0.9978 (± 0.0021), confirming model robustness and generalizability. Predictions and analytical artifacts are persisted in a structured MySQL database comprising five normalized tables, and operational insights are surfaced through a four-page Microsoft Power BI dashboard supporting real-time fraud monitoring. The system is demonstrated on a synthetic dataset of 15,000 Indian financial transactions and evaluated against 2,000 prospectively generated records, achieving a holistic prediction corpus of 17,000 transactions with four-tier risk stratification (Critical, High, Medium, Low).

KEYWORDS Fraud Detection; Ensemble Learning; SMOTE; Random Forest; Digital Payment Security; Risk Scoring; Feature Engineering

PAPER CITATION:

Prasad, R.J., Vijayalakshmi, V. :- " SecurePay Shield: An Ensemble Machine Learning Architecture for Real-Time Fraud Detection and Probabilistic Risk Scoring in the Indian Digital Payment Ecosystem", International Journal of Informative & Futuristic Research (IJIFR), Vol. (13) (8), April 2026, pp. 1206-1213 ,<https://doi.org/10.64672/IJIFR/26.04.13.08.040>



This article is an open access article published under the terms and conditions of the CC- BY –NC –SA 4.0 Creative Commons Attribution-Non Commercial- ShareAlike 4.0 International Public License. All copyrights reserved to the Authors & Journal Publisher. Copyright© Authors (IJIFR 2026).

1. INTRODUCTION

The proliferation of digital payment modalities in India — encompassing the Unified Payments Interface (UPI), Immediate Payment Service (IMPS), National Electronic Funds Transfer (NEFT), Real-Time Gross Settlement (RTGS), and card-based transactions — has fundamentally restructured the financial services landscape. Driven by the Government of India's Digital India initiative and facilitated by near-universal smartphone penetration, annual digital transaction volumes have surpassed billions, constituting a critical pillar of the national economy [7]. This dramatic expansion has, however, concomitantly enlarged the attack surface available to financially motivated adversaries, who exploit technological vulnerabilities, social engineering methodologies, and behavioral exploitation to perpetrate payment fraud at scale.

The economic ramifications of payment fraud are substantial, with losses accruing to thousands of crores of Indian Rupees annually, imposing direct financial burdens upon consumers, merchants, and financial institutions, and eroding systemic trust in digital payment infrastructure [18]. Contemporary fraud typologies — including card-not-present (CNP) fraud, account takeover (ATO) attacks, synthetic identity fraud, and phishing-mediated credential theft — have grown increasingly sophisticated, leveraging automation, distributed bot networks, and adaptive evasion strategies that render static, rule-based detection mechanisms progressively inadequate [5].

Machine learning methodologies offer compelling advantages over rule-based antecedents by autonomously discovering latent, non-linear relational structures within high-dimensional transaction data, adapting to emergent fraud paradigms through model retraining, and furnishing probabilistic risk assessments that support nuanced, risk-stratified decision making [12]. Ensemble methods, in particular, demonstrate superior robustness to noise, reduced variance through model aggregation, and the capacity to leverage complementary inductive biases across constituent learners.

1.1 Contribution of the Paper

This paper makes the following primary contributions to the domain of computational fraud detection:

- A domain-specific feature engineering framework that systematically constructs eight composite behavioral risk indicators from 23 raw transactional attributes, expanding the feature space to 35 dimensions and providing the most discriminative predictors identified through permutation importance analysis.
- An end-to-end ensemble machine learning pipeline integrating Random Forest, Gradient Boosting, and Isolation Forest algorithms, with rigorous comparative evaluation across five performance dimensions (AUC-ROC, F1-Score, Precision, Recall, Accuracy).
- A SMOTE-augmented class balancing protocol demonstrably effective in improving minority-class (fraudulent transaction) recall within the severely imbalanced Indian payment transaction context.
- A normalized, five-table MySQL analytical schema and a four-page Power BI operational dashboard enabling end-to-end fraud monitoring from raw transaction ingestion to executive-level business intelligence.
- A four-tier probabilistic risk stratification model (Critical, High, Medium, Low) that transcends binary fraud classification to provide actionable, graduated risk signals for operational fraud management teams.

2. LITERATURE SURVEY

The academic literature on computational fraud detection is extensive, spanning rule-based heuristics, statistical anomaly detection, and, most recently, machine learning and deep learning paradigms. Phua et al. [5] provide a seminal taxonomy of data mining-based fraud detection techniques,

categorizing approaches along the axes of supervision, scalability, and interpretability, and establishing the foundational benchmarks against which subsequent methodological advances are evaluated.

Breiman's seminal formalization of the Random Forest algorithm [1] established the theoretical foundation for ensemble-based classification, demonstrating that the aggregation of decorrelated decision trees via bagging and random feature subspace selection yields classifiers with markedly reduced generalization error compared to individual tree models. Friedman's gradient boosting framework [4] extended ensemble methodology through sequential residual correction, providing a complementary inductive mechanism with demonstrated efficacy in structured, tabular classification tasks.

The challenge of class imbalance in fraud detection datasets — wherein fraudulent transactions constitute a small minority of total transaction volume — was systematically addressed by Chawla et al. [2] through the introduction of SMOTE, which generates synthetic minority class instances through k-nearest neighbor interpolation in feature space, substantially improving minority class recall without the information loss associated with undersampling. Japkowicz and Stephen [19] and Krawczyk [20] subsequently provided comprehensive analyses of class imbalance handling strategies, affirming the superior empirical performance of oversampling techniques relative to cost-sensitive learning and threshold optimization in high-imbalance regimes.

Liu et al.'s Isolation Forest [3] introduced an unsupervised anomaly detection paradigm predicated on the observation that anomalous instances are more susceptible to isolation through random recursive partitioning than normal instances, providing a computationally efficient complement to supervised classification in contexts where labeled fraud annotations are scarce or unreliable. Dal Pozzolo et al. [6] subsequently demonstrated the importance of calibrated probability estimation in imbalanced fraud classification, establishing that probability calibration substantially improves decision threshold selection and operational precision-recall trade-off management.

The criticality of feature engineering in fraud detection was rigorously established by Bahnsen et al. [12], who demonstrated that domain-specific feature construction — including transaction velocity metrics, amount deviation indicators, and temporal behavioral signals — yields substantially greater predictive uplift than algorithmic sophistication alone. Bolton and Hand [15] provide a comprehensive review of statistical fraud detection methodologies, articulating the fundamental tension between model sensitivity (recall of fraudulent transactions) and operational specificity (minimization of false positives that disrupt legitimate customer activity).

Interpretability constraints in financial machine learning were addressed by Lundberg and Lee [13] through the introduction of SHAP (SHapley Additive exPlanations) values, which provide theoretically grounded, instance-level attribution of model predictions to individual feature contributions, satisfying the explainability requirements of financial regulators and enabling transparent fraud reasoning. Ahmed et al. [11] extended anomaly detection methodology to network-level transaction graph structures, anticipating the Graph Neural Network approaches subsequently applied to ring fraud detection. The preceding body of work collectively establishes the theoretical and empirical foundations upon which the SecurePay Shield architecture is constructed.

3. Proposed Work

The proposed architecture operationalizes a nine-stage, modular machine learning pipeline that encompasses synthetic data generation, feature engineering, class balancing, ensemble model training, probabilistic risk scoring, analytical database persistence, and business intelligence visualization. Each stage is designed to be independently testable, configurable, and extensible, supporting future integration of streaming data sources and additional model architectures.

3.1 System Architecture Description

The SecurePay Shield system is organized into five hierarchically ordered architectural layers, as enumerated below. Each layer maintains a well-defined interface with adjacent layers, ensuring separation of concerns and enabling independent scaling:

Layer 1: Data Acquisition & Generation Layer

Responsible for ingestion of historical transaction records from Excel-formatted data stores and programmatic generation of synthetic Indian financial transaction datasets using domain-calibrated statistical distributions. Configuration parameters (NUM_RECORDS = 15,000, FRAUD_RATE = 0.06) govern the generation of 14,100 legitimate and 900 fraudulent transaction records. Input: Raw Excel file (SecurePay_Transaction_Data.xlsx, 23 attributes). Output: Pandas DataFrame with 15,000 records and injected fraud patterns.

Layer 2: Preprocessing & Feature Engineering Engine

Executes temporal feature extraction (Day_of_Week, Is_Weekend, Is_Night_Transaction), amount-based feature derivation (Amount_to_Balance_Ratio, Amount_Deviation, Amount_Log), and composite behavioral risk score computation (Velocity_Score, Risk_Composite). Label encoding is applied to seven categorical variables; StandardScaler normalization is applied to the complete 35-dimensional feature matrix. Output: Normalized feature matrix X (15,000 $\times$ 35), target vector y (binary).

Layer 3: Intelligence & Model Training Layer

Implements stratified 80/20 train-test splitting, SMOTE oversampling to yield 22,560 balanced training instances, and parallel training of three models: Random Forest (200 estimators, max_depth=15, class_weight='balanced'), Gradient Boosting (150 estimators, max_depth=8, learning_rate=0.1), and Isolation Forest (contamination=0.06). Five-fold stratified cross-validation governs model selection. The Random Forest model is designated the production deployment candidate based on AUC-ROC superiority.

Layer 4: Database Persistence Layer

A normalized MySQL schema (database: securepay_shield) comprising five analytical tables: transaction_predictions (17,000 records), model_performance (3 model entries), feature_importance (35 feature entries), fraud_summary (aggregated risk statistics), and risk_distribution (four-tier risk level counts). Batch insertion (2,000 records per commit) optimizes write throughput.

Layer 5: Presentation & Visualization Layer

A four-page Power BI dashboard consuming the MySQL analytical tables through DirectQuery or scheduled import, surfacing: (i) Transaction Overview — aggregate fraud rate, total at-risk value; (ii) Risk Analysis — risk tier distribution, geographic heat maps; (iii) Model Performance — ROC curves, confusion matrices, cross-validation trends; (iv) Geographic-Behavioral Analysis — city-level fraud concentration, merchant category risk profiles.

3.2 Dataset Specification

The SecurePay Shield dataset comprises 15,000 synthetically generated Indian financial transaction records reflecting realistic payment patterns. The 23 base attributes span demographic characteristics (Customer_Age, Customer_Gender), transactional parameters (Transaction_Amount_INR, Account_Balance_INR, Transaction_Type), behavioral signals (Txn_Frequency_Last_7Days, Time_Since_Last_Txn_Hours, Login_Attempts), and network risk indicators (IP_Risk_Score, Is_International). Transaction amounts are drawn from a log-normal distribution parameterized as $\mu=9.5$, $\sigma=1.8$, clipped to the INR [100, 200,000] range, reflecting the heavy-tailed distribution characteristic of real Indian payment data. Customer ages follow a normal distribution (loc=38, scale=12), clipped to [18, 80] years. Geographic coverage spans 48 Indian cities across all major metropolitan and tier-2 markets.

Table 1: Dataset Attribute Specification and Statistical Properties

Attribute	Data Type	Range / Categories	Description
Transaction_Amount_INR	Float	INR 100 – 2,00,000	Transaction value; log-normal distributed

IP_Risk_Score	Float	0 – 100	Network-level risk score of originating IP
Login_Attempts	Integer	1 – 10	Consecutive login attempts pre-transaction
Txn_Frequency_Last_7Days	Integer	0 – 50	Transaction count in preceding 7-day window
Transaction_Type	Categorical	UPI / NEFT / RTGS / IMPS / POS	Payment channel classification
Is_International	Binary	0 / 1	Domestic vs. cross-border transaction indicator
Is_Fraud	Binary	0 / 1	Ground truth fraud label (6% prevalence)

3.3 Feature Engineering Pipeline

The feature engineering module constitutes the most consequential determinant of predictive performance within the SecurePay Shield pipeline. Eight derived features are systematically computed from the base 23 attributes, organized into three functional categories:

(i) Temporal Risk Features: Day_of_Week (integer 0–6), Is_Weekend (binary), Is_Night_Transaction (binary; transactions occurring between 22:00–5:00 are flagged, reflecting elevated fraud prevalence during off-hours).

(ii) Amount Deviation Features: Amount_to_Balance_Ratio = Transaction_Amount_INR / (Account_Balance_INR + 1); Amount_Deviation = |Transaction_Amount_INR – Previous_Txn_Amount_INR|; Amount_Log = $\log_2(1 + \text{Transaction_Amount_INR})$.

(iii) Behavioral Composite Features: Velocity_Score = Txn_Frequency_Last_7Days / (Time_Since_Last_Txn_Hours + 0.1); Risk_Composite = (IP_Risk_Score × 0.3) + (Login_Attempts × 10) + (Is_International × 20) + (Is_Night_Transaction × 15) + (Velocity_Score × 2).

The Risk_Composite score emerges as the single most predictive feature (importance = 0.2599), followed by IP_Risk_Score (0.2176) and Velocity_Score (0.1556), collectively accounting for 63.1% of total model predictive capacity. These composite indicators capture multi-dimensional behavioral risk signals that are substantially more difficult for adversaries to systematically manipulate than individual raw attributes.

4. RESULTS AND DISCUSSION

The experimental evaluation of SecurePay Shield encompasses comparative model benchmarking, cross-validation stability analysis, feature importance ranking, and risk distribution characterization. All experiments are conducted on the 15,000-record synthetic Indian transaction dataset, with model selection and hyperparameter optimization performed on the 80% training partition and final performance metrics reported on the 20% held-out test set (3,000 transactions: 2,820 legitimate, 180 fraudulent).

4.1 Comparative Model Performance

Table 2 presents the comprehensive performance evaluation of the three constituent models across five standard evaluation metrics. The Random Forest classifier achieves near-perfect discrimination on all metrics, attributable to the richness of the engineered feature space and the effectiveness of SMOTE-based class balancing in ensuring adequate minority-class representation during training.

Table 2: Comparative Model Performance Evaluation on 20% Held-Out Test Set

Model	AUC-ROC	F1-Score	Precision	Recall	Accuracy
Random Forest (RF)	1.0000	1.0000	1.0000	1.0000	1.0000
Gradient Boosting (GB)	0.9998	0.9987	0.9991	0.9983	0.9994

Isolation Forest (IF)	N/A	N/A	N/A	N/A	0.9791
-----------------------	-----	-----	-----	-----	--------

The Gradient Boosting classifier demonstrates competitive performance (AUC-ROC = 0.9998, F1 = 0.9987), validating the complementary utility of sequential boosting as a secondary model for ensemble stacking in future system iterations. The Isolation Forest, operating in an unsupervised mode without access to fraud labels during training, achieves an overall accuracy of 97.91%, which is notable given the absence of supervised signal, though its fraud-specific recall is lower than supervised counterparts, rendering it more suitable as an anomaly score supplement than a primary classifier.

4.2 Cross-Validation Stability Analysis

Table 3: Five-Fold Stratified Cross-Validation F1-Score Results — Random Forest

Fold 1	Fold 2	Fold 3	Fold 4	Fold 5
0.9981	0.9963	0.9987	0.9974	0.9984

Mean F1 = 0.9978; Standard Deviation = 0.0021

The marginal variance across folds (SD = 0.0021) confirms that the model’s exceptional performance is not attributable to overfitting on a specific train-test partition but reflects genuine learning of discriminative fraud signals generalizable across diverse transaction subpopulations. This stability is particularly significant in the context of the severe class imbalance (6% fraud rate), as stratified folding ensures that each fold maintains a representative fraud prevalence.

4.3 Feature Importance Analysis

Table 4: Top-10 Predictive Features by Permutation Importance — Random Forest Model

Ran k	Feature	Importance Score	Feature Category
1	Risk_Composite	0.2599	Behavioral Composite
2	IP_Risk_Score	0.2176	Network-Level Signal
3	Velocity_Score	0.1556	Behavioral Composite
4	Login_Attempts	0.0842	Behavioral Signal
5	Amount to Balance Ratio	0.0631	Amount Deviation
6	Is International	0.0514	Contextual Flag
7	Amount Deviation	0.0407	Amount Deviation
8	Is Night Transaction	0.0318	Temporal Risk
9	Txn_Frequency_Last_7Days	0.0276	Behavioral Signal
10	Amount Log	0.0241	Amount Transformation

4.4 Risk Distribution Analysis

Table 5: Predicted Risk Level Distribution Across 17,000-Record Production Corpus

Risk Level	Transaction Count	% of Total	Fraud Probability Threshold	Action Recommended
Critical	~312	~1.84%	≥ 0.80	Immediate Block
High	~578	~3.40%	≥ 0.60	Manual Review
Medium	~892	~5.25%	≥ 0.40	Step-Up Auth
Low	~15,218	~89.51%	< 0.40	Approve

5. CONCLUSION

SecurePay Shield demonstrates the feasibility and efficacy of an integrated, open-source machine learning architecture for real-time financial fraud detection within the Indian digital payment ecosystem. The system achieves exceptional predictive performance — AUC-ROC of 1.0000 and F1-Score of 1.0000 for the production-selected Random Forest model, with five-fold cross-validation confirming robust generalizability (mean F1 = 0.9978, SD = 0.0021) — through the synergistic combination of

domain-specific feature engineering, SMOTE-augmented class balancing, and ensemble learning methodology.

The architecture's modular, layered design provides clear separation between data ingestion, preprocessing, model intelligence, database persistence, and visualization, enabling independent optimization and extension of each component. The five-table MySQL analytical schema and four-page Power BI operational dashboard collectively deliver end-to-end observability from raw transaction ingestion through executive fraud monitoring, establishing a production-viable platform foundation for Indian financial institution deployment.

The four-tier probabilistic risk stratification framework transcends binary classification to provide graduated, actionable risk signals that support differentiated operational responses — immediate blocking for Critical-risk transactions, manual review for High-risk, step-up authentication for Medium-risk, and automated approval for Low-risk — optimizing the precision-recall trade-off in accordance with the institution's fraud appetite and customer experience objectives.

Future research directions of significant potential impact include: (i) deployment as a RESTful API microservice using FastAPI, transforming the batch pipeline into a sub-millisecond real-time scoring engine; (ii) integration of Long Short-Term Memory (LSTM) or Transformer architectures to capture temporal sequence dependencies across customer transaction histories; (iii) Graph Neural Network (GNN) modeling of account-device-merchant interaction networks to detect ring fraud and money laundering typologies; (iv) federated learning implementation enabling collaborative model training across multiple financial institutions without raw data sharing, addressing privacy and regulatory constraints; (v) SHAP-based transaction-level explainability integration to satisfy emerging regulatory transparency mandates; and (vi) streaming pipeline replacement using Apache Kafka to achieve genuine real-time fraud interception latency.

6. REFERENCES

- [1] Breiman, L., "Random Forests," Machine Learning, Vol. 45, Issue 1, 2001, pp. 5–32.
- [2] Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P., "SMOTE: Synthetic Minority Over-sampling Technique," Journal of Artificial Intelligence Research, Vol. 16, 2002, pp. 321–357.
- [3] Liu, F. T., Ting, K. M., & Zhou, Z. H., "Isolation Forest," Proceedings of the IEEE International Conference on Data Mining (ICDM), 2008, pp. 413–422.
- [4] Friedman, J. H., "Greedy Function Approximation: A Gradient Boosting Machine," Annals of Statistics, Vol. 29, Issue 5, 2001, pp. 1189–1232.
- [5] Phua, C., Lee, V., Smith, K., & Gayler, R., "A Comprehensive Survey of Data Mining-based Fraud Detection Research," arXiv preprint arXiv:1009.6119, 2010.
- [6] Dal Pozzolo, A., Caelen, O., Johnson, R. A., & Bontempi, G., "Calibrating Probability with Undersampling for Unbalanced Classification," IEEE Symposium Series on Computational Intelligence, 2015.
- [7] Reserve Bank of India, "Annual Report on Payment and Settlement Systems," RBI Publications, Mumbai, 2024.
- [8] Pedregosa, F., et al., "Scikit-learn: Machine Learning in Python," Journal of Machine Learning Research, Vol. 12, 2011, pp. 2825–2830.
- [9] Lemaitre, G., Nogueira, F., & Aridas, C. K., "Imbalanced-learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine Learning," Journal of Machine Learning Research, Vol. 18, Issue 17, 2017, pp. 1–5.
- [10] McKinney, W., "Data Structures for Statistical Computing in Python," Proceedings of the 9th Python in Science Conference, 2010, pp. 56–61.
- [11] Ahmed, M., Mahmood, A. N., & Hu, J., "A Survey of Network Anomaly Detection Techniques," Journal of Network and Computer Applications, Vol. 60, 2016, pp. 19–31.
- [12] Bahnsen, A. C., Aouada, D., Stojanovic, A., & Ottersten, B., "Feature Engineering Strategies for Credit Card Fraud Detection," Expert Systems with Applications, Vol. 51, 2016, pp. 134–142.
- [13] Lundberg, S. M., & Lee, S. I., "A Unified Approach to Interpreting Model Predictions," Advances in Neural Information Processing Systems (NIPS), Vol. 30, 2017.
- [14] Hand, D. J., & Henley, W. E., "Statistical Classification Methods in Consumer Credit Scoring: A Review," Journal of the Royal Statistical Society: Series A, Vol. 160, Issue 3, 1997, pp. 523–541.

- [15] Bolton, R. J., & Hand, D. J., "Statistical Fraud Detection: A Review," Statistical Science, Vol. 17, Issue 3, 2002, pp. 235–249.
- [16] MySQL 8.0 Reference Manual, "Chapter 13: SQL Statements," Oracle Corporation Documentation, 2024.
- [17] Microsoft Power BI Documentation, "DAX Function Reference," Microsoft Learn Platform, 2024.
- [18] National Payments Corporation of India, "UPI Transaction Statistics Report Q3 2024," NPCI Publications, Mumbai, 2024.
- [19] Japkowicz, N., & Stephen, S., "The Class Imbalance Problem: A Systematic Study," Intelligent Data Analysis, Vol. 6, Issue 5, 2002, pp. 429–449.
- [20] Krawczyk, B., "Learning from Imbalanced Data: Open Challenges and Future Directions," Progress in Artificial Intelligence, Vol. 5, Issue 4, 2016, pp. 221–232.