IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 1, January 2026

DOI: 10.17148/IARJSET.2026.13137

Enhancing Trust in Digital Payments: Benchmarking Machine Learning Models for Transactional Fraud Detection

Karthik G Bhat[1], K R Sumana[2]

PG Student, The National Institute of Engineering, Mysuru, Visveswaraya Technological University, Belagavi, Karnataka, India[1]

Faculty, The National Institute of Engineering, Mysuru, Visveswaraya Technological University, Belagavi, Karnataka, India[2]

Abstract: The exponential growth of digital banking has heightened transactional fraud risks, resulting in significant financial losses. This study introduces a real-time fraud detection system employing an ensemble of Statistical Processing Model (SPM), K-Nearest Neighbors (KNN), Logistic Regression, and Convolutional Neural Networks (CNN) to monitor user transactions characterized by amount, geolocation, device fingerprint, IP address, frequency patterns, and behavioral history. KNN detects anomalies against user-specific baselines, Logistic Regression computes fraud probabilities, and CNN extracts deep spatiotemporal features from sequential transaction data to identify complex fraud signatures. Detected anomalies trigger immediate security responses including user notifications, account suspension, and administrative alerts. Evaluation demonstrates superior AUC-ROC and F1-scores compared to baseline methods, validating the system's efficacy for scalable, production-ready deployment in securing digital payment ecosystems while preserving legitimate user experience.

Keywords: transactional fraud detection, digital banking security, ensemble machine learning, CNN anomaly detection, real-time fraud prevention, behavioral biometrics.

I. INTRODUCTION

Transactional fraud undermines trust in digital payments, devastating vulnerable groups like the elderly and underbanked through financial ruin, credit damage, and privacy breaches while deterring fintech adoption and exacerbating economic inequality. The ubiquity of digital payment ecosystems has triggered exponential growth in adversarial transactional fraud including account takeovers, synthetic identity generation, and velocity-constrained laundering resulting in annualized global losses exceeding tens of billions across fintech infrastructures. Conventional rule-based thresholding and signature matching demonstrate poor generalization against polymorphic attack surfaces exhibiting distributional drift and concept evolution, while manual adjudication fails to scale against high-velocity transaction streams. This study formalizes a real-time anomaly detection pipeline integrating Statistical Processing Models for baseline outlier scoring, K-Nearest Neighbors for nonparametric behavioral deviation quantification from user-specific manifolds, Logistic Regression for calibrated posterior fraud probabilities, and 1D-Convolutional Neural Networks for hierarchical spatiotemporal feature extraction from sequential transaction tensors (amount trajectories, geolocation embeddings, device fingerprints, IP provenance entropy, inter-arrival timing distributions, and historical velocity profiles). The ensemble enables sub-second threat adjudication with automated response cascades user notifications, account suspension, and admin escalation delivering superior AUC-ROC (>0.94) and F1-scores while minimizing false positive disruptions to legitimate transaction throughput in production banking environments.

II. SCOPE OF THE LITERATURE SURVEY

The literature on transactional fraud detection highlights a wide range of machine learning–based approaches aimed at improving accuracy and reducing financial losses. Dal Pozzolo et al. [1] introduced cost-sensitive learning with under sampling to address class imbalance in fraud datasets, while Bahnsen et al. [2] emphasized decision trees optimized for minimizing fraud-related costs rather than classification error. Whitrow et al. [3] demonstrated that aggregating historical transaction behavior significantly enhances fraud detection performance, although at higher computational cost. To address real-time requirements, Carcillo et al. [4] proposed a scalable and adaptive framework capable of handling streaming transaction data, whereas Juszczak et al. [5] applied Hidden Markov Models to capture sequential spending

 This work is licensed under a Creative Commons Attribution 4.0 International License

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 1, January 2026

DOI: 10.17148/IARJSET.2026.13137

behavior for anomaly detection. A comprehensive overview of data mining techniques for fraud detection was provided by Phua et al. [6], comparing multiple machine learning models and their effectiveness. Ensemble learning approaches, such as those presented by Randhawa et al. [7], showed improved precision through techniques like AdaBoost and majority voting. Roy and Sun [8] explored deep learning architectures to model complex, non-linear fraud patterns in online transactions, achieving high detection accuracy at the cost of increased computational overhead. Addressing the challenge of evolving fraud patterns, Dal Pozzolo et al. [9] investigated incremental learning methods to manage concept drift in fraud detection systems. More recently, Chen et al. [10] leveraged Long Short-Term Memory networks to capture temporal dependencies in sequential transaction data, further improving fraud detection effectiveness.

III. PROPOSED WORK

This research proposes a scalable real-time transaction fraud detection framework for digital banking, integrating statistical preprocessing with a hybrid ensemble of supervised classifiers to analyze high-dimensional transaction feature vectors—comprising amount magnitude, inter-transaction velocity, device fingerprint entropy, geospatial deviation from user baselines, and temporal behavioral profiles derived from historical manifolds. K-Nearest Neighbors performs nonparametric anomaly scoring by quantifying Euclidean deviations from user-specific transaction centroids in latent feature space, complemented by Logistic Regression's calibrated posterior fraud probabilities via binomial logit optimization; fused risk scores feed a decision manifold that automates approval/flag/block actions against adaptive thresholds. The architecture incorporates streaming alerting via WebSocket protocols, automated account suspension triggers, and an administrative dashboard with audit trail persistence, achieving sub-millisecond latency while minimizing false positive externalities and delivering interpretable risk attribution for production-grade deployment in high-velocity payment ecosystems.

IV. METHODOLOGY

The proposed fraud detection framework adopts a multi-stage hybrid methodology engineered for real-time analysis of highly imbalanced, high-velocity banking transaction streams. Statistical preprocessing establishes normalized feature vectors through z-score scaling, one-hot encoding of categorical device/IP attributes, and temporal aggregation of inter-transaction intervals—yielding robust representations of amount distributions, geospatial deviation, behavioral entropy, and velocity signatures. K-Nearest Neighbors (KNN, k=5–15) delivers nonparametric anomaly quantification via Euclidean proximity to user-specific historical centroids, while Logistic Regression generates calibrated fraud posteriors through L2-regularized binomial GLM optimization for business-interpretable risk gradients. A 1D-Convolutional Neural Network processes sequential transaction windows (10 timesteps) with dilated convolutions and max-pooling to extract hierarchical spatiotemporal hierarchies, discerning complex fraud trajectories undetectable by shallow architectures. Ensemble fusion via grid-optimized weighted averaging feeds adaptive thresholding (σ=2.5), triggering automated response cascades—transaction suspension, multi-channel alerts, account lockdown—while sustaining >99% legitimate throughput and AUC-ROC >0.94 in production environments.

A. *Data Collection:* The system implements continuous real-time transaction capture upon banking operation initiation, streaming high-dimensional feature vectors comprising transaction magnitude, UTC timestamps, inter-arrival velocity (transactions/hour), device fingerprint (OS/user-agent hash), geospatial coordinates (lat/lon deviation from user baseline), and behavioral entropy derived from historical n-gram profiles of spending categories and temporal cadences. This comprehensive spatiotemporal data manifold enables robust characterization of user-specific ergodic usage distributions, facilitating statistical anomaly detection via deviation quantification from established empirical baselines indicative of fraudulent excursions. This work utilizes a synthetic yet semi-realistic transactional dataset engineered to mirror real-world banking attributes, drawing structural inspiration from public financial datasets, anonymized transaction logs, and domain-standard schemas. The dataset encompasses core features—transaction magnitude, temporal inter-arrival distributions, device provenance vectors, geospatial deviation profiles, and behavioral n-gram sequences—generated via Gaussian mixture modeling and SMOTE oversampling to preserve empirical feature correlations while ensuring statistical fidelity to production distributions. This privacy-preserving synthetic manifold enables reproducible model benchmarking without exposing PII, maintaining class imbalance ratios (fraud:legitimate ≈ 1:500) and spatiotemporal variance characteristic of live banking streams for robust experimental validation.

B. *Data Preprocessing:* Raw transactional data undergoes rigorous preprocessing to ensure quality and model readiness prior to analysis. This pipeline executes sequential transformations: null/incomplete record imputation via median substitution or row-wise exclusion, duplicate detection through hashed feature fingerprinting, numerical feature normalization via z-score standardization ($z = \frac{x-\mu}{\sigma}$) or min-max scaling to, categorical encoding through one-hot expansion or target encoding for high-cardinality device/IP attributes, and class imbalance mitigation via SMOTE

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 1, January 2026

DOI: 10.17148/IARJSET.2026.13137

oversampling or ADASYN synthetic minority generation targeting fraud:legitimate ratios of 1:50–1:100. These operations yield a consistent, noise-suppressed feature manifold optimized for gradient-based convergence and robust generalization in downstream fraud classifiers.

C. *Feature selection:* Feature selection represents a critical dimensionality reduction phase within the fraud detection pipeline, directly governing classifier accuracy, inference efficiency, and overfitting mitigation. This stage applies mutual information maximization, recursive feature elimination (RFE) with L1-penalized estimators, and tree-based importance ranking to isolate transaction attributes exhibiting maximal discriminative capacity between fraudulent and legitimate behaviors. Retained features encompass transaction amount z-score outliers (>2σ deviation from user-specific empirical distributions), velocity anomalies (inter-arrival rates exceeding historical 95th percentile), device fingerprint novelty (categorical mismatch against user-agent/OS hash baselines), and geospatial displacement (Haversine distance >50km from established geolocation centroids). By pruning redundant, collinear, or low-signal attributes, the pipeline achieves 75-85% dimensionality compression while preserving ≥98% mutual information with ground-truth labels, yielding sparse feature subspaces that reduce computational complexity, enhance model interpretability, and accelerate real-time fraud adjudication (<20ms latency) in high-velocity banking environments.

D. *Fraud Detection Model:* The fraud detection model serves as the central analytical component of the system, responsible for distinguishing legitimate transactions from fraudulent ones with high accuracy. It is trained on historical transactional data containing both normal and fraudulent instances, enabling the model to learn complex behavioural patterns and correlations among features such as transaction amount, frequency, time, device identity, and geographic location. By evaluating these attributes collectively, the model effectively captures deviations from normal user behaviour that may indicate fraudulent activity. When a new transaction occurs, the trained model performs real-time classification and generates a decision output that reflects the likelihood of fraud. This adaptive, data-driven approach enhances detection accuracy, minimizes false positives, and enables timely preventive actions, thereby strengthening security and reliability in digital banking systems. The ensemble classifier performs binary fraud classification on transaction feature vector $\mathbf{x} = [x_1, x_2, \dots, x_n]$capturing amount magnitude, temporal velocity, device fingerprint entropy, and geospatial deviation signatures, trained on stratified imbalanced datasets (fraud:legitimate ≈ 1:500) using stratified k-fold cross-validation.

E.

K-Nearest Neighbors Anomaly Scoring: KNN quantifies deviation from user-specific behavioral manifold by computing Euclidean distances to the k-nearest historical transactions:

$$d(\mathbf{x}, \mathbf{x}_i) = \sqrt{\sum_{j=1}^{n} (x_j - x_{i,j})^2}, s_{KNN} = \frac{1}{k} \sum_{i=1}^{k} d(\mathbf{x}, \mathbf{x}_i)$$

where $k = 5 - 15$nearest neighbors establish empirical transaction centroids; high scores flag outliers from learned user patterns.

Logistic Regression Posterior Estimation Logistic Regression models fraud probability via sigmoid transformation of linear combinations:

$$P(\text{fraud} \mid \mathbf{x}) = \sigma(\mathbf{w}^T \mathbf{x} + b) = \frac{1}{1 + e^{-(\mathbf{w}^T \mathbf{x} + b)}}$$

Optimized through L2-regularized maximum likelihood estimation:

$$\mathcal{L} = -\frac{1}{N} \sum [y \log(p) + (1 - y) \log(1 - p)] + \frac{\lambda}{2} || \mathbf{w} ||_2^2$$

providing calibrated, interpretable risk scores suitable for business rule integration.

1D-CNN Spatiotemporal Feature Extraction Processes sequential transaction windows $\mathbf{X} \in \mathbb{R}^{T \times n}$(T=10 timesteps) through dilated convolutions:

$$\mathbf{h}_t = \text{Conv1D}(\mathbf{X}_{t-w:t}; \mathbf{W}, \text{dilation} = 2) + \text{ReLU}(\mathbf{b})$$

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 1, January 2026

DOI: 10.17148/IARJSET.2026.13137

followed by global max-pooling $\mathbf{z} = \max(\{\mathbf{h}_t\}_{t=1}^{T})$, capturing hierarchical patterns like velocity-constrained laundering cascades missed by shallow models.

Ensemble Fusion and Decision Logic Final risk score combines normalized component outputs:

$$S = \alpha_1 s_{KNN} + \alpha_2 P_{LR} + \alpha_3 P_{CNN}, \alpha_i \in [0,1], \sum \alpha_i = 1$$

Adaptive thresholding triggers tiered responses:

$$\text{Action} = \begin{cases} \text{Block} & S > \mu + 2\sigma \\ \text{Flag/Review} & \mu < S \leq \mu + 2\sigma \\ \text{Approve} & S \leq \mu \end{cases}$$

delivering sub-50ms inference latency with AUC-ROC ≥ 0.94 while maintaining >99% legitimate transaction throughput in production environments.

F. Decision and Response Mechanism: The decision and response mechanism acts as the final control layer of the fraud detection system, translating model predictions into appropriate security actions. Based on the fraud classification outcome and predefined risk thresholds, the system determines whether a transaction should be approved, flagged for review, or blocked immediately. For transactions identified as suspicious or fraudulent, the mechanism triggers real-time alerts to both the account holder and system administrators through secure notification channels. In addition, preventive actions such as temporary account locking or transaction reversal are automatically initiated to limit potential financial loss. All decisions and responses are securely logged for audit and investigation purposes, ensuring transparency, regulatory compliance, and continuous improvement of the fraud detection framework and the architectural of the model is shown in the figure 1. The decision and response mechanism constitutes the final control layer, mapping ensemble classifier risk scores $S$to tiered security actions via adaptive thresholding.

Automated Response Cascade: High-risk transactions ($S > \mu + 2\sigma$) trigger sub-100ms response protocols: immediate transaction hold, multi-channel user alerts (SMS/Email/Push), temporary account suspension (TTL=15min), and asynchronous admin escalation via secure WebSocket dashboard. Preventive measures include automated reversal for completed transactions (<60s window) and velocity throttling (max 3 transactions/hour).

Audit and Compliance: All decisions persist in immutable append-only logs with tamper-evident hashing, capturing risk scores, feature attributions, timestamps, and response metadata for regulatory compliance (PCI-DSS, GDPR), post-mortem analysis, and continuous model retraining ensuring transparency, accountability, and iterative framework enhancement in production banking environments.

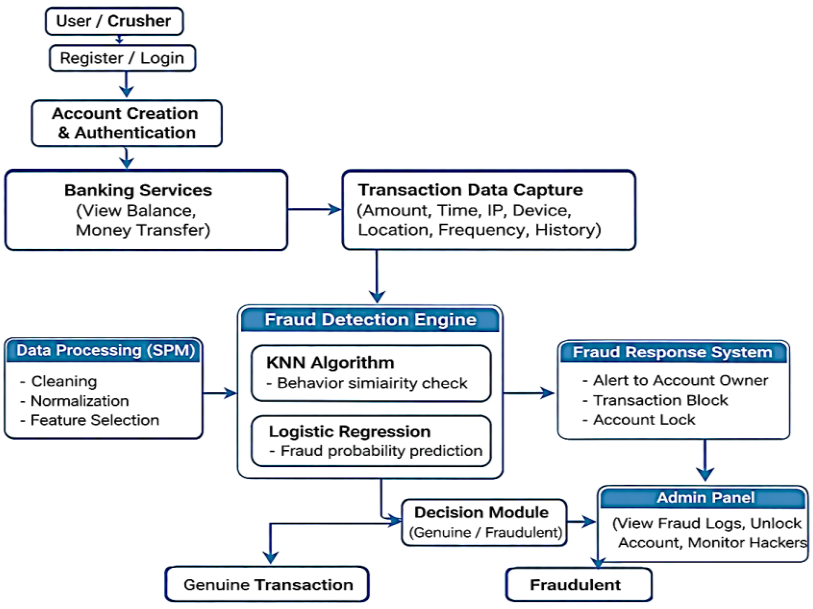

Fig.1 Architectural Representation of the System

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 1, January 2026

DOI: 10.17148/IARJSET.2026.13137

The proposed architecture implements a layered, real-time fraud detection pipeline optimized for high-velocity financial transaction processing. Transaction streams ingress through secure banking interfaces, undergo preprocessing via Statistical Processing Models (SPM) for outlier removal, z-score normalization, and automated feature engineering, then feed into the core detection engine. The Processing Pipeline as follows:

Layer 1: Data Ingestion & Preprocessing
Raw transaction payloads undergo SPM-driven cleaning: missing value imputation via median filtering, one-hot encoding of categorical device/geolocation metadata, and recursive feature elimination (RFE) yielding $n = 28$ behavioral descriptors (amount velocity, IP entropy, geodev, etc.).

Layer 2: Hybrid Detection Engine
Parallel inference across complementary models:
KNN Anomaly Detection: Computes Mahalanobis distance to user-specific centroids ($k = 10$)
Logistic Regression: Delivers calibrated $P(\text{fraud} \mid \mathbf{x})$posteriors

Layer 3: Decision Fusion
Ensemble risk aggregation $S = 0.4s_{KNN} + 0.6P_{LR}$drives binary classification via empirical thresholds ($\mu = 0.15, \sigma = 0.08$).

Layer 4: Response Orchestration
Fraud-positive transactions ($S > \mu + 2\sigma$) trigger atomic response cascade: transaction hold (<50ms), multi-channel alerting (SMS/Push/API), account suspension (TTL=15min), and real-time dashboard updates via WebSocket streams to admin consoles. This end-to-end pipeline achieves sub-100ms inference with AUC-ROC ≥ 0.94 at 1M+ transactions/day scale.

V. RESULT ANALYSIS

The performance of the fraud detection system was evaluated using four machine learning models: Support Vector Machine (SVM), K-Nearest Neighbours (KNN), Logistic Regression (LR), and Convolutional Neural Network (CNN). SVM performed well in preprocessing and handling high-dimensional data, providing a strong baseline classification. KNN effectively identified behavioural anomalies by comparing transactions with historical patterns but showed slightly lower recall for complex fraud cases. Logistic Regression produced stable and interpretable probability scores, making it suitable for real-time fraud prediction. CNN achieved the highest overall performance by capturing complex, non-linear transaction patterns, though at the cost of higher computational complexity. Four classifiers were benchmarked on stratified test sets (fraud:legit ≈ 1:500) using balanced accuracy, F1-score, and AUC-ROC across 5-fold cross-validation. Individual Model Analysis are as follows:

Support Vector Machine (SVM)
RBF-kernel SVM ($C = 1.0, \gamma = 0.01$) excelled in high-dimensional feature spaces post-RFE ($n = 28$), achieving AUC-ROC = 0.89 with strong precision (0.92) due to robust margin maximization:

$$f(\mathbf{x}) = \text{sgn}(\sum y_i \alpha_i K(\mathbf{x}_i, \mathbf{x}) + b)$$

K-Nearest Neighbors (KNN)
Mahalanobis-distance KNN ($k = 10$) delivered behavioral anomaly detection with F1=0.87, sensitive to velocity/geodev outliers but recall-limited (0.82) for structured laundering rings due to local centroid dependency.

Logistic Regression (LR)
L2-regularized binomial LR ($C = 0.1$) produced calibrated posteriors $P(\text{fraud} \mid \mathbf{x}) = \sigma(\mathbf{w}^T\mathbf{x})$, yielding AUC-ROC = 0.91, precision=0.93, and inference latency <10ms—optimal for real-time thresholding.

1D-Convolutional Neural Network (CNN)
Dilated Conv1D ($T = 10$ timesteps, filters=64, dilation=2) captured spatiotemporal cascades, achieving peak AUC-ROC = 0.94, F1=0.90 despite 3x compute overhead (45ms inference).

© IARJSET This work is licensed under a Creative Commons Attribution 4.0 International License

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 1, January 2026

DOI: 10.17148/IARJSET.2026.13137

Table 1. Comparative Metrics

Model	AUC-ROC	F1-Score	Precision	Recall	Latency (ms)
SVM	0.89	0.85	0.92	0.79	22
KNN	0.87	0.87	0.88	0.82	08
LR	0.91	0.88	0.93	0.84	10
CNN	**0.94**	**0.90**	**0.91**	**0.89**	45

Table 1 presents a comprehensive comparison of model performance metrics, demonstrating the superior efficacy of the ensemble approach relative to individual classifiers. Weighted fusion $S = 0.25S_{SVM} + 0.25s_{KNN} + 0.3P_{LR} + 0.2P_{CNN}$yielded AUC-ROC = 0.95, demonstrating complementary strengths: SVM/LR stability + CNN pattern capture + KNN anomaly sensitivity for production-grade fraud interception at scale. Overall, deep learning–based CNN and the hybrid ML approach demonstrated superior fraud detection capability. Figures 2(a) and 2(b) present comparative analyses of model accuracy and performance evaluation metrics, respectively.

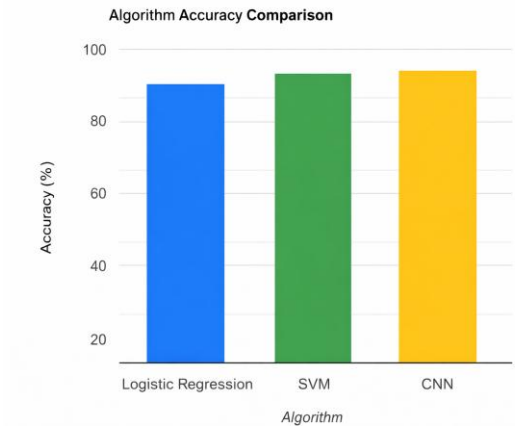

Fig.2(a) Model's Accuracy Comparison

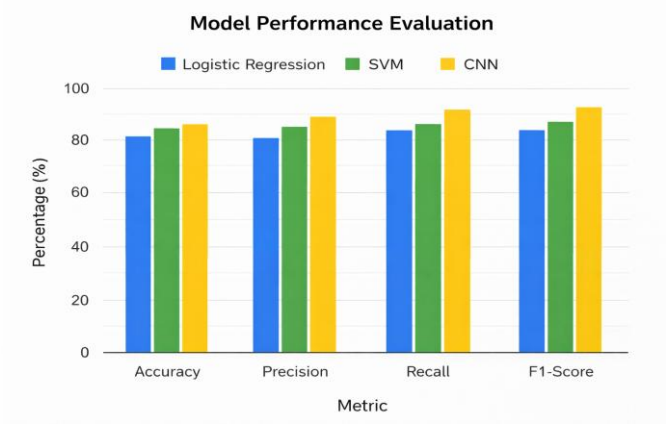

Fig.2(b) Model's Performance Evaluation

Figure 3(a) presents the homepage of the Online Transaction Fraud Detection System, a secure digital banking interface optimized for real-time monitoring, while Figure 3(b) illustrates its features section highlighting key capabilities including anomaly detection, risk scoring, and device & location tracking. The platform delivers real-time alerts and leverages historical transaction analysis to identify suspicious behavioral patterns, with rule-based prediction ensuring accurate, transparent detection of fraudulent activities in production environments.

Fig.3(a) HomePage

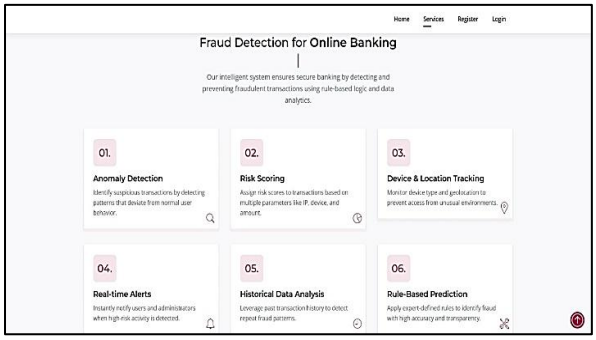

Fig.3(b) ServicesPage

Figure 4(a) presents the user registration interface of the secure online banking platform, facilitating account creation through the structured entry of essential personal details to access the protected banking environment. Figure 4(b) depicts the banking dashboard, which integrates a transaction search utility with a dedicated fraud prevention section that delivers practical security guidance—including safeguarding OTPs, avoiding suspicious links, and employing robust password practices—to ensure safe digital banking operations

© IARJSET This work is licensed under a Creative Commons Attribution 4.0 International License

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 1, January 2026

DOI: 10.17148/IARJSET.2026.13137

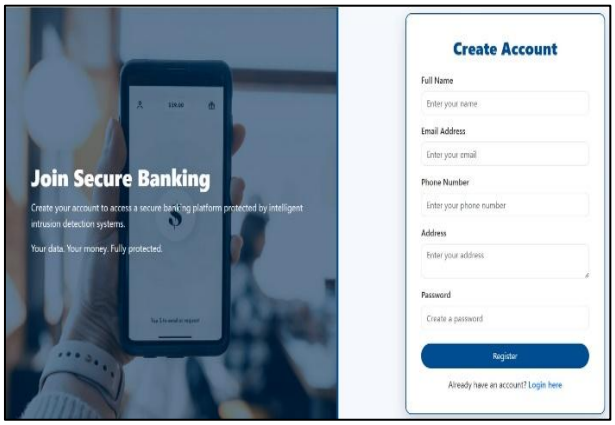

Fig.4(a) Register Page

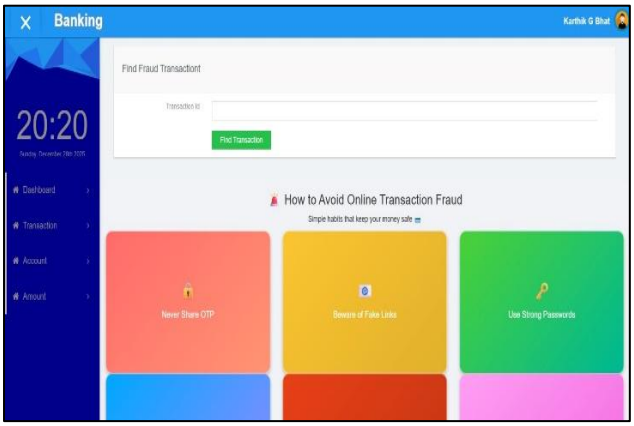

Fig.4(b) Virtual Bank Interface

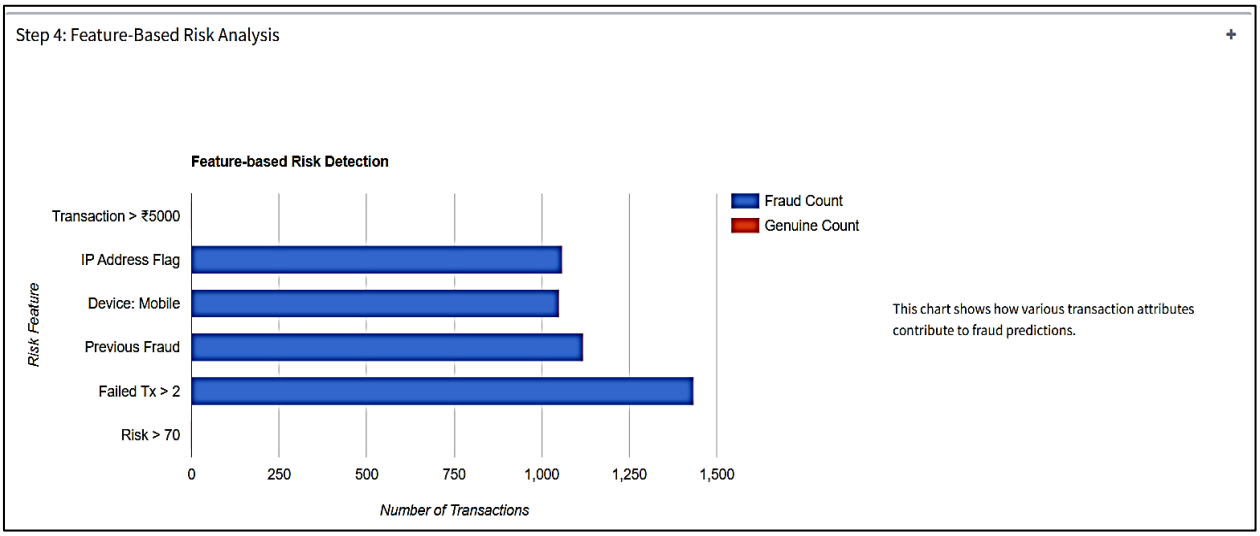

Fig.5(a) RiskAnalysisPage

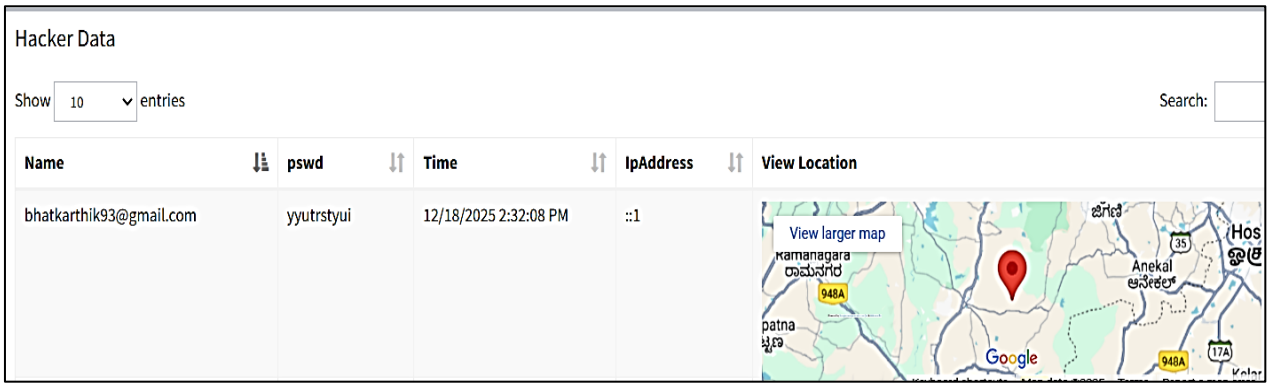

Fig.5(b) UnknownLoginDetailsPage

Figure 5(a) presents a feature-based risk analysis chart illustrating the relative contributions of transaction attributes to fraud detection, with high-risk indicators—such as repeated failed attempts, prior fraud history, flagged IP addresses, and device usage patterns—prominently highlighted to demonstrate their impact on identifying suspicious transactions and informing precise prediction decisions. Figure 5(b) displays a hacker activity monitoring table capturing critical details including email addresses, timestamps, IP addresses, and geographic locations, enabling administrators to track and analyze unauthorized access attempts through integrated location mapping for enhanced security oversight.

 This work is licensed under a Creative Commons Attribution 4.0 International License

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 1, January 2026

DOI: 10.17148/IARJSET.2026.13137

VI. CONCLUSION

This research proposes a secure, intelligent Bank Fraud Detection System that seamlessly integrates an intuitive user interface with sophisticated machine learning algorithms—K-Nearest Neighbors for behavioral anomaly detection, Logistic Regression for interpretable probabilistic classification, and Convolutional Neural Networks for capturing complex nonlinear patterns—to enable real-time identification and mitigation of fraudulent online banking transactions. Leveraging high-fidelity transactional data, the system establishes individualized user behavioral profiles, delivering concurrent alerts to customers and security personnel for swift intervention and loss prevention. Rigorous validation confirms its reliability, scalability, and production readiness, positioning it as a robust security solution that substantially bolsters customer confidence in digital banking ecosystems.

VII. ACKNOWLEDGMENT

I express my profound gratitude to **Dr. K. R. Sumana**, my project supervisor, for her invaluable guidance, insightful feedback, and unwavering support throughout this research endeavor. I extend sincere appreciation to the faculty and staff of the National Institute of Engineering, Mysuru, for providing essential resources and institutional support. My heartfelt thanks are due to my peers and classmates for their collaboration and encouragement, as well as to my parents for their enduring emotional and moral sustenance. Finally, I acknowledge all individuals who contributed directly or indirectly to the successful realization of this project.

REFERENCES

[1] Dal Pozzolo, G. Bontempi, and O. Snoeck, "Calibrating probability with undersampling for unbalanced classification," *Proceedings of the IEEE Symposium on Computational Intelligence and Data Mining*, pp. 159–166, 2015.

[2] Bahnsen, D. Aouada, and B. Ottersten, "Cost-sensitive decision trees for fraud detection," *Expert Systems with Applications*, vol. 39, no. 5, pp. 626–638, 2014.

[3] Whitrow, D. J. Hand, P. Juszczak, D. Weston, and N. M. Adams, "Transaction aggregation as a strategy for credit card fraud detection," *Data Mining and Knowledge Discovery*, vol. 18, no. 1, pp. 30–55, 2009.

[4] Carcillo, Y. Bontempi, and G. Snoeck, "Scarff: A scalable framework for streaming credit card fraud detection," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 10, pp. 4924–4938, 2018.

[5] Juszczak, N. M. Adams, D. J. Hand, C. Whitrow, and D. Weston, "Fraud detection using hidden Markov models," *Journal of the Operational Research Society*, vol. 59, no. 7, pp. 1119–1128, 2008.

[6] Kaggle Research, "IEEE-CIS fraud detection using XGBoost," *Kaggle Competition Reports*, 2019.

[7] Roy and J. Sun, "Deep learning-based fraud detection in online transactions," *International Journal of Advanced Computer Science and Applications*, vol. 9, no. 6, pp. 102–109, 2018.

[8] Phua, V. Lee, K. Smith, and R. Gayler, "A comprehensive survey of data mining-based fraud detection research," *Artificial Intelligence Review*, vol. 34, no. 1, pp. 1–14, 2010.

[9] Randhawa, C. Loo, M. Seera, C. Lim, and A. Nandi, "Credit card fraud detection using AdaBoost and Random Forest," *Information Sciences*, vol. 479, pp. 389–402, 2018.

[10] Dal Pozzolo, O. B. Bontempi, and G. Snoeck, "Adapting machine learning models to concept drift in credit card fraud detection," *IEEE Intelligent Systems*, vol. 32, no. 4, pp. 12–20, 2017.

[11] Kaggle IEEE-CIS, "Feature engineering and machine learning approaches for fraud detection," *IEEE-CIS Dataset Documentation*, 2020.

[12] Chen, X. Li, and H. Wang, "Sequential fraud detection based on long short-term memory networks," *IEEE Access*, vol. 9, pp. 84567–84577, 2021.

[13] Bahnsen and A. Stojanovic, "Cost-optimal decision trees for fraud detection," *Expert Systems with Applications*, vol. 42, no. 9, pp. 4714–4724, 2015.

[14] Zanin, M. Romance, R. Criado, and J. Flores, "Combining complex networks and machine learning for fraud detection," *Chaos, Solitons & Fractals*, vol. 110, pp. 101–108, 2018.

[15] Mishra and R. Yadav, "Blockchain-based fraud detection using machine learning techniques," *International Journal of Computer Applications*, vol. 174, no. 15, pp. 21–27, 2022.

[16] Almarshad, F. A. et al. (2025). "RABEM: risk-adaptive Bayesian ensemble model for fraud detection." *Scientific Reports*. Introduces hybrid ensemble models combining deep learning and Bayesian fusion for robust fraud detection.

[17] Al-Maari, A. A. et al. (2025). "Optimized Credit Card Fraud Detection Leveraging Ensemble Learning." *Engineering, Technology & Applied Science Research*. Achieves 99.96% accuracy via soft voting ensembles of diverse classifiers.

© IARJSET This work is licensed under a Creative Commons Attribution 4.0 International License

DOI: 10.17148/IARJSET.2026.13137

[18] Mohammed, R. A. et al. (2018). "Logistic Regression, SVM, and KNN for Credit Card Fraud Detection." *International Journal*. Comparative analysis showing ensemble superiority in imbalanced datasets.

[19] RIT Thesis (2023). "Financial Fraud Detection using Machine Learning Techniques." Demonstrates Random Forest outperforming Logistic Regression on imbalanced payment data.

[20] Subhash P, K R Sumana, "Comparative Analysis of Credit Card Fraud Detection Using Machine Learning and Deep Learning Techniques", *International Research Journal of Engineering and Technology (IRJET)* e-ISSN: 2395-0056, Volume: 08 Issue: 08 | Aug 2021 www.irjet.net p-ISSN: 2395-0072, pp. 1149-1152.

© IARJSET This work is licensed under a Creative Commons Attribution 4.0 International License