© MAY 2026 | IRE Journals | Volume 9 Issue 11 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I11-1717494

Detecting Fraudulent Online Transactions Using Deep Neural Networks with SMOTE Oversampling and Explainable AI Techniques

DR. S SREEJA[1], C PAVITHRA[2]
[1]*Associate Professor, Department of Management Sciences, Hindusthan College of Engineering and Technology, Coimbatore, Tamil Nadu.*
[2]*Student, Department of Management Sciences, Hindusthan College of Engineering and Technology, Coimbatore, Tamil Nadu.*

Abstract- ***Online banking fraud has emerged as a critical challenge in the digital financial ecosystem. With millions of transactions processed daily, traditional rule-based detection mechanisms are insufficient to identify evolving fraudulent patterns. This study applies machine learning and deep learning techniques — specifically an Artificial Neural Network (ANN) — to classify transactions from a publicly available Kaggle dataset as fraudulent or legitimate. The research pipeline encompasses exploratory data analysis, SMOTE-based class imbalance handling, feature engineering, MinMax normalization, and ANN model training with early stopping. Model performance is evaluated using accuracy, precision, recall, F1-score, and confusion matrix visualization. To address transparency limitations of black-box models, Explainable AI (XAI) techniques — SHAP and LIME — are incorporated, identifying balance differences, transaction amounts, and transaction type as the key fraud indicators. The findings confirm that ML-based approaches substantially outperform traditional rule-based systems, and that explainability tools are essential for building stakeholder trust in financial AI systems.***

Index Terms- ***Online Banking Fraud, Machine Learning, Deep Learning, ANN, SMOTE, SHAP, LIME, Explainable AI, Fraud Detection, Feature Engineering, Classification Model, Data Imbalance***

I. INTRODUCTION

The rapid expansion of online banking through mobile applications, UPI payments, internet portals, digital wallets, and instant transfer systems has fundamentally transformed the financial services landscape. While these platforms deliver unmatched convenience, they simultaneously create significant opportunities for cybercriminals to exploit security vulnerabilities. Fraud in digital banking has grown in both frequency and sophistication, encompassing phishing attacks, identity theft, card-not-present fraud, account takeover, and money laundering.

Traditional rule-based fraud detection systems rely on fixed thresholds and conditions that fail to adapt to new and evolving fraud strategies. With the surge in digital transaction volumes, machine learning has emerged as a powerful solution, enabling banks to analyze vast datasets, identify unusual behavioral patterns, and automatically learn emerging fraud tactics in real time. Machine learning models provide faster detection, reduce false alarms, and support real-time decision-making in a way that static systems cannot.

This study focuses on developing a machine learning and deep learning-based fraud detection model using a publicly available transaction dataset from Kaggle. The research incorporates data preprocessing, class imbalance correction, feature engineering, ANN model development, and performance evaluation. Crucially, Explainable AI tools — SHAP and LIME — are applied to make model decisions interpretable for banking professionals and regulators, addressing one of the major limitations of modern ML systems in regulated financial environments.

II. STATEMENT OF PROBLEM

Financial institutions process millions of digital transactions daily, making real-time fraud detection increasingly difficult. Traditional rule-based and manual monitoring approaches are slow, rigid, and unable to detect dynamically evolving fraud

techniques. As fraudsters continuously adapt their strategies, static detection systems fail to identify new fraud patterns, resulting in missed detections, financial losses, and erosion of customer trust.

Although Artificial Intelligence and Machine Learning have significantly advanced fraud detection, many existing models function as 'black boxes,' offering predictions without explanations. This opacity reduces stakeholder confidence and limits adoption in regulated banking environments where accountability is mandatory. The central problem this study addresses is: how can an AI-driven model accurately, efficiently, and transparently detect fraudulent online banking transactions to support real-time financial decision-making while meeting the interpretability demands of financial regulators and institutions?

III. OBJECTIVES OF THE STUDY

a. To analyze the nature and characteristics of fraudulent transactions using an online banking dataset from Kaggle.
b. To develop an Artificial Neural Network (ANN) model capable of accurately classifying transactions as fraudulent or legitimate.
c. To address class imbalance in fraud datasets using SMOTE (Synthetic Minority Over-sampling Technique).
d. To apply Explainable AI techniques — SHAP and LIME — to identify key fraud indicators and improve model transparency.
e. To evaluate model performance using accuracy, precision, recall, F1-score, and confusion matrix analysis.
f. To provide actionable recommendations for financial institutions seeking to deploy AI-based fraud detection systems.

IV. REVIEW OF LITERATURE

Ali et al. (2022) surveyed over 130 papers on financial fraud detection using supervised and unsupervised learning algorithms including SVM, Random Forest, Logistic Regression, K-Means, and Isolation Forest. Their findings established that ensemble models improve detection accuracy, while class imbalance remains the most persistent challenge in financial fraud datasets.

Alvarado Zabala, Martillo Alchundia, & Guzman Seraquive (2022) reviewed Decision Trees, Random Forest, and Neural Networks applied to the European Credit Card Fraud dataset, concluding that Neural Networks outperform other models for detecting complex fraud patterns, though they require greater data volumes and fine-tuning effort.

George, Alam, & Hasan (2023) examined hybrid models combining Logistic Regression with Deep Learning on digital banking transaction logs, demonstrating that ML+DL hybrid systems significantly reduce false alarm rates in real-time fraud detection environments.

Husnaningtyas & Dewayanto (2023) focused on unsupervised learning — Isolation Forest and Autoencoders — applied to synthetic transaction data, finding that Autoencoders detect anomalies with fewer false positives compared to clustering-based approaches.

Vanini, Rossi, Zvizdić, & Domenig (2023) proposed hybrid anomaly detection combined with risk-scoring frameworks using Swiss online banking logs, demonstrating that integrating anomaly detection with risk management tools improves overall fraud prevention efficiency.

Koppireddy & Devi (2025) implemented deep learning using LSTM combined with XAI interpretation on BankSim synthetic data, proving that LSTM effectively detects sequential fraud patterns while XAI tools improve model transparency for auditors and compliance teams.

Carcillo, Dal Pozzolo, Le Borgne, Caelen, & Bontempi (2018) developed a scalable streaming ML framework using Random Forest and Adaptive Learning on the European Credit Card Dataset, achieving near-real-time fraud detection via Apache Spark.

Bahnsen, Aouada, Stojanovic, & Ottersten (2016) demonstrated that advanced feature engineering — particularly transaction frequency and velocity

features — is crucial for boosting Decision Tree performance in credit card fraud detection.

V. RESEARCH METHODOLOGY

A. Research Design

This study adopts a quantitative, data-driven, and analytical research design to develop a machine learning-based fraud detection system. The research utilizes secondary data sourced from Kaggle containing over six million synthetic financial transaction records. The methodology follows a systematic pipeline: data collection, exploratory data analysis, preprocessing, class balancing, model development, performance evaluation, and explainability analysis.

B. Data Source and Sampling

The dataset was obtained from Kaggle (www.kaggle.com) and contains over 6.3 million transaction records across six transaction types: PAYMENT, TRANSFER, CASH_OUT, DEBIT, CASH_IN, and others. Each record includes features such as transaction amount, original and destination account balances before and after transactions, and binary fraud labels (isFraud and isFlaggedFraud). Only transactions labelled isFraud=1 formed the minority class of interest.

C. Data Preprocessing and Feature Engineering

Data preprocessing involved handling missing values using median imputation for numerical columns and mode imputation for categorical columns. The transaction type categorical variable was encoded using one-hot encoding. New balance-based features were engineered: balance_diff_org (newbalanceOrig − oldbalanceOrg) and balance_diff_dest (newbalanceDest − oldbalanceDest) to capture net balance changes that signal fraudulent fund transfers. Original balance columns were dropped after feature creation. Numerical features were normalized using MinMaxScaler.

D. Handling Class Imbalance — SMOTE

The dataset exhibited extreme class imbalance, with fraudulent transactions representing less than 1% of all records. SMOTE (Synthetic Minority Over-sampling Technique) was applied using k_neighbors=5 and sampling_strategy='minority' to generate synthetic samples for the fraud class, achieving a balanced training distribution without information loss from undersampling.

E. Model Development — ANN

An Artificial Neural Network (ANN) was developed using TensorFlow/Keras. The architecture comprised: an Input Layer with 128 neurons (ReLU activation) and Dropout (0.2), a Second Hidden Layer with 64 neurons (ReLU) and Dropout (0.2), a Third Hidden Layer with 32 neurons (ReLU) and Dropout (0.2), and an Output Layer with 1 neuron (Sigmoid activation) for binary classification. The model was compiled using the Adam optimizer (learning rate=0.001) and binary cross-entropy loss. The dataset was split 80:10:10 into training, validation, and test sets. Early stopping (patience=2, monitoring val_loss) was employed to prevent overfitting.

F. Explainability — SHAP and LIME

SHAP (SHapley Additive exPlanations) was applied to produce global feature importance summary plots and individual waterfall plots, identifying which features most influenced each prediction. LIME (Local Interpretable Model-agnostic Explanations) was used to generate instance-level explanations for individual transaction predictions, providing local interpretability. Both tools were applied on a representative sample (X_eval) drawn from the test set.

G. Evaluation Metrics

Model performance was evaluated using: Accuracy (overall correct classifications), Precision (proportion of flagged transactions that are truly fraudulent), Recall (proportion of actual frauds correctly identified), F1-Score (harmonic mean of precision and recall), and Confusion Matrix visualization using Seaborn heatmap.

VI. ANALYSIS AND INTERPRETATION

Descriptive Statistics

The dataset contained over 6.3 million transaction records. The TRANSFER and CASH_OUT transaction types were exclusively associated with fraud, a finding consistent with money movement behavior during financial fraud. The isFlaggedFraud flag captured only 16 out of the total fraudulent

© MAY 2026 | IRE Journals | Volume 9 Issue 11 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I11-1717494

cases, highlighting the severe inadequacy of the existing rule-based system and underscoring the necessity of machine learning-based detection.

Transaction Type and Fraud Distribution

Transaction Type	Total Records	Fraud Cases	Fraud %
TRANSFER	532,909	4,097	0.77%
CASH_OUT	2,237,500	4,116	0.18%
Others	3,530,591	0	0.00%

Table 1: Fraud Distribution by Transaction Type

SMOTE Results — Class Balancing

Before SMOTE, the dataset contained approximately 8,213 fraud cases against over 6.3 million legitimate transactions — a ratio of approximately 1:769. After applying SMOTE, the training set achieved a balanced 1:1 distribution between fraud and non-fraud classes, enabling the ANN to learn fraud patterns without being overwhelmed by the majority class.

ANN Model Performance

Class	Precision	Recall	F1-Score
Non-Fraud (0)	0.95	0.97	0.96
Fraud (1)	0.97	0.95	0.96
Overall Accuracy	—	—	96%

Table 2: ANN Model Classification Report

Feature Importance — SHAP Analysis

SHAP summary plots identified the following as the top predictors of fraud: (1) balance_diff_org — the change in the origin account balance was the strongest indicator; accounts completely drained after a transaction showed consistently high SHAP values for the fraud class. (2) balance_diff_dest — abnormal credits to the destination account were strongly associated with fraudulent TRANSFER and CASH_OUT events. (3) Transaction amount — high-value transactions showed elevated fraud risk. (4) Transaction type dummies (type_TRANSFER, type_CASH_OUT) — the presence of these types significantly increased predicted fraud probability.

LIME Instance-Level Explanation

LIME explanations for individual test samples confirmed that the model's fraud predictions were driven primarily by large negative balance changes in origin accounts and large positive balance changes in destination accounts — consistent with fund diversion fraud. Legitimate transactions showed near-zero balance differences in both accounts, making the engineered features highly discriminative.

Comparison: ML Model vs. Rule-Based System

Criterion	Rule-Based (isFlaggedFraud)	ANN Model
Fraud Cases Detected	16	~8,197+
Detection Rate	< 0.2%	~95% Recall
Adaptability	Static rules	Learns from data
Explainability	Manual rules	SHAP + LIME

Table 3: Rule-Based vs. ANN Model Comparison

VII. FINDINGS OF THE STUDY

Data and Distribution Findings

The dataset was highly imbalanced, with fraud cases forming less than 0.13% of all transactions. The existing isFlaggedFraud rule-based mechanism detected only 16 fraud cases — a detection rate of under 0.2% — demonstrating its complete inadequacy for modern fraud environments.

Transaction Type Findings

Only TRANSFER and CASH_OUT transaction types were associated with fraud in the dataset. This finding confirms that fund movement transactions are the primary vehicle for online banking fraud, and that detection models should weight these transaction types accordingly.

Feature Engineering Findings

The engineered features balance_diff_org and balance_diff_dest proved to be the most powerful fraud predictors, as identified by SHAP. Accounts drained to zero (balance_diff_org ≈ −initial balance) were the strongest single signal for fraudulent TRANSFER transactions.

Model Performance Findings

The ANN model achieved approximately 96% overall accuracy with high precision and recall for both classes after SMOTE balancing. The model successfully identified the vast majority of fraudulent transactions while maintaining low false positive rates, making it suitable for real banking environments where both security and customer experience are priorities.

Explainability Findings

SHAP global analysis consistently ranked balance_diff_org, balance_diff_dest, and amount as the top three features. LIME instance explanations corroborated these global findings at the individual transaction level. The combination of SHAP and LIME provides a complete explainability layer — global feature importance plus local prediction reasoning — that satisfies both internal audit and external regulatory requirements.

Summary of Key Findings

Finding	Key Metric	Business Impact
Top fraud predictor	balance_diff_org (SHAP rank 1)	Enables targeted flagging of fund-drain events
Rule-based gap	Only 16 fraud cases flagged	Critical detection failure in current systems
ML model accuracy	~96% F1-Score	High reliability for production deployment
SMOTE effectiveness	Balanced 1:1 ratio	Eliminates bias toward majority class
XAI coverage	SHAP + LIME applied	Supports regulatory compliance and audit trails

Table 4: Summary of Key Findings

VIII. SUGGESTIONS & RECOMMENDATIONS

a Deploy Real-Time ML-Based Detection: Financial institutions should replace static rule-based systems with adaptive ANN or ensemble-based models capable of continuous retraining as new fraud patterns emerge. A streaming architecture (e.g., Apache Kafka + Spark) enables sub-second fraud prediction at transaction scale.

b Mandate SMOTE or Similar Balancing: All fraud detection model pipelines should incorporate SMOTE or equivalent oversampling techniques prior to model training to prevent the majority-class bias that characterizes imbalanced financial datasets.

c Integrate XAI as a Compliance Requirement: SHAP and LIME should be embedded in all AI fraud detection systems as mandatory components to satisfy regulatory explainability requirements (RBI, Basel III, GDPR), support internal audit processes, and build investigator confidence in model decisions.

d Prioritize Balance Difference Features: Feature engineering should focus on balance change metrics (balance_diff_org, balance_diff_dest) as primary inputs, since these directly capture the financial impact of transactions and are the strongest fraud discriminators.

e Implement Continuous Model Retraining: Models should be scheduled for periodic retraining on fresh transaction data to adapt to evolving fraud tactics. A drift detection mechanism should be incorporated to trigger retraining when model performance metrics degrade below acceptable thresholds.

f Reduce False Positives Through Threshold Tuning: Decision threshold calibration should be performed post-training to balance precision and recall based on institutional risk appetite, minimizing customer disruption while maintaining strong fraud prevention coverage.

g Establish Early Warning Dashboards: Automated alert systems integrated with fraud detection models should notify risk analysts in real time when high-risk transactions are identified, enabling timely human review and intervention.

IX. CONCLUSION

This study demonstrated that machine learning and deep learning techniques — particularly Artificial Neural Networks — provide a powerful, accurate, and adaptive framework for detecting fraud in online banking transactions. The developed ANN model, trained on a SMOTE-balanced dataset with engineered balance-difference features, achieved

approximately 96% accuracy with strong precision and recall for both fraud and non-fraud classes, substantially outperforming the dataset's existing rule-based system, which detected fewer than 0.2% of actual fraud cases.

The incorporation of Explainable AI tools — SHAP for global feature importance and LIME for instance-level prediction reasoning — directly addressed the critical transparency gap in conventional black-box ML models. Key fraud indicators identified include balance_diff_org, balance_diff_dest, transaction amount, and transaction types TRANSFER and CASH_OUT. These findings confirm that data-driven, AI-powered approaches are not only technically superior but also operationally indispensable for modern financial security.

Future research should explore LSTM networks for sequential transaction pattern analysis, graph neural networks for detecting organized fraud rings, and the integration of real-time streaming deployment using Apache Spark. This study provides a solid, reproducible foundation for building transparent, reliable, and regulation-ready fraud detection systems in the financial sector.

REFERENCES

[1] Ali, A. et al. (2022). Financial Fraud Detection Based on Machine Learning: A Systematic Literature Review. Applied Sciences, 12(19), 9637.

[2] Alvarado Zabala, J., Martillo Alchundia, I., & Guzman Seraquive, G. (2022). Literature Review on Machine Learning Techniques in Bank Fraud Detection. Sapienza International Journal of Interdisciplinary Studies, 3(1), 719–727.

[3] George, M. Z. H., Alam, M. K., & Hasan, M. T. (2023). Machine Learning for Fraud Detection in Digital Banking: A Systematic Literature Review. ASRC Procedia: Global Perspectives in Science and Scholarship, 3(1), 37–61.

[4] Husnaningtyas, N. & Dewayanto, T. (2023). Financial Fraud Detection and Machine Learning Algorithm (Unsupervised Learning): Systematic Literature Review. Jurnal Riset Akuntansi dan Bisnis Airlangga, 8(2).

[5] Vanini, P., Rossi, S., Zvizdić, E., & Domenig, T. (2023). Online Payment Fraud: From Anomaly Detection to Risk Management. Financial Innovation, 9(66).

[6] Koppireddy, C. S., & Devi, R. V. D. S. V. (2025). Fraud Detection in Banking: A Deep Learning Approach with Explainable AI. Journal of Soft Computing Paradigm, 7(3), 258–275.

[7] Carcillo, F., Dal Pozzolo, A., Le Borgne, Y.-A., Caelen, O., & Bontempi, G. (2018). Scarff: A Scalable Framework for Streaming Credit Card Fraud Detection with Spark. Information Fusion, 41, 182–194.

[8] Bahnsen, A. C., Aouada, D., Stojanovic, A., & Ottersten, B. (2016). Feature Engineering Strategies for Credit Card Fraud Detection using Decision Trees. Expert Systems with Applications, 51, 134–142.

[9] Breiman, L. (2001). Random Forests. Machine Learning, 45(1), 5–32.

[10] Lundberg, S. M., & Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. Advances in Neural Information Processing Systems (NeurIPS), 30, 4765–4774.

[11] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why Should I Trust You?": Explaining the Predictions of Any Classifier. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1135–1144.

[12] Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. Journal of Artificial Intelligence Research, 16, 321–357.