Unilag Journal of Mathematics and Applications,
Volume 6, Issue 1 (2026), Pages 107–129.
ISSN: 2805 3966. URL: http://lagjma.unilag.edu.ng

ENSEMBLE-BASED FRAUD DETECTION IN NIGERIAN BANKING: A SYNTHETIC DATA BENCHMARK AND COST-SENSITIVE ANALYSIS

GBOLAHAN ADENIRAN IDOWU* AND JOSIAH ENDURANCE OWOLABI

Abstract. Fraud detection in Nigerian banking is critically hampered by a scarcity of authentic transaction data for model development, a challenge exacerbated by the rapid growth of digital payments. To address this foundational data gap, this study introduces a novel, high-fidelity synthetic benchmark dataset of 1,000,000 financial transactions, meticulously calibrated to reflect the fraud patterns reported by the Nigeria Inter-Bank Settlement System (NIBSS). Using this dataset, we develop a comprehensive analytical framework to evaluate the economic efficacy of advanced machine learning models. The results demonstrate a substantial potential for fraud loss reduction: optimized Random Forest models achieved a 69.1% decrease in simulated fraud-related costs (from ₦81.8M to ₦25.2M) while maintaining perfect precision. Alternatively, XGBoost delivered superior recall (74.6%) with an F1-score of 0.854, providing a strategic option for institutions prioritizing fraud detection rates. A SHAP analysis identified transaction amount and associated behavioral features as the strongest fraud indicators and highlighted Web and Mobile channels as requiring enhanced monitoring. This paper makes three principal contributions: the first publicly available, NIBSS-calibrated fraud detection dataset for Nigeria, addressing a pivotal data scarcity in African financial research; empirically validated evidence that ensemble methods, combined with threshold optimization, can reduce fraud costs by up to 69%; and actionable implementation guidelines for Nigerian banks operating within existing regulatory compliance frameworks. The synthetic data methodology offers a replicable and privacy-preserving blueprint for other emerging markets facing similar constraints on data access and availability.

1. Introduction

Nigeria's electronic payment ecosystem has undergone remarkable transformation, with transaction volumes reaching ₦600 trillion in 2023; a 55% increase from the previous year according to the Nigeria Inter-Bank Settlement System

2010 *Mathematics Subject Classification.* Primary: 90-10. Secondary: 62P05.
Key words and phrases. Cost-sensitive Learning; Fraud Detection; Nigerian Banking; Random Forest; SHAP Interpretability; Synthetic dataset; XGBoost.

Submitted: January 19, 2026. Revised: February 20, 2026. Accepted: February 23, 2026.
* Correspondence.

(NIBSS) [1]. This growth trajectory was largely attributed to cash scarcity during this period. While this demonstrated the digitalization of financial transaction, it also created a gap in combating fraudulent activities. Traditional rule-based detection systems, which were originally designed for lower transaction volumes, now struggle with the velocity and sophistication of modern fraud schemes.

The challenge facing Nigerian financial institutions extends beyond mere scale. Contemporary fraudsters employ adaptive strategies that exploit vulnerabilities across multiple channels, with mobile banking and web platforms emerging as primary targets. NIBSS data reveals that mobile channels account for 49.75% of documented fraud cases, while web and Point-of-Sale systems contribute 22.91% and 18.38% respectively. These statistics underscore the urgent need for data-driven, adaptive fraud detection mechanisms capable of operating across Nigeria's diverse banking infrastructure. [2]

Current fraud detection approaches in Nigerian banking suffer from several critical limitations. Rule-based systems generate excessive false positives, disrupting legitimate customer transactions and eroding trust. More fundamentally, these systems lack the adaptive capacity to evolve with emerging fraud patterns, creating persistent vulnerabilities that sophisticated actors readily exploit. The economic implications are substantialbeyond direct financial losses, institutions face regulatory penalties, reputational damage, and customer attrition.

Despite the critical need for adaptive fraud detection in Nigeria's rapidly expanding digital payment ecosystem, financial institutions face a fundamental barrier: limited access to authentic transaction data due to privacy regulations (NDPR), institutional data-sharing constraints, and competitive considerations. This data scarcity prevents the development and validation of machine learning models that could address the inadequacies of rule-based systems. The core research question this work addresses is: Can synthetic data, calibrated to documented Nigerian fraud patterns, enable the development of cost-effective ensemble learning models that significantly outperform traditional approaches while providing actionable insights for real-world deployment?

This study aims to develop and validate a synthetic Nigerian financial transaction dataset calibrated to NIBSS-documented fraud distributions; systematically evaluate machine learning algorithms under cost-sensitive frameworks to minimize economic loss; and quantify the economic impact of ensemble methods through comprehensive cost-benefit analysis and SHAP-based interpretability.

This work addresses these challenges through three primary contributions. First, we develop a comprehensive synthetic dataset of Nigerian financial transactions, calibrated to documented fraud patterns from NIBSS reports, providing the first publicly available benchmark for Nigerian fraud detection research. Second, we conduct systematic evaluation of supervised learning algorithms under cost-sensitive frameworks, demonstrating that ensemble methods significantly outperform traditional approaches when appropriately calibrated. Finally, we provide detailed economic analysis showing that optimized machine learning models can reduce fraud-related costs by up to 69.1% compared to baseline approaches.

Our approach recognizes the practical constraints facing Nigerian financial institutions, particularly limited access to authentic fraud data due to privacy regulations and competitive considerations. The synthetic data framework we propose offers a viable pathway for developing and validating fraud detection systems while maintaining full ethical compliance and enabling reproducible research.

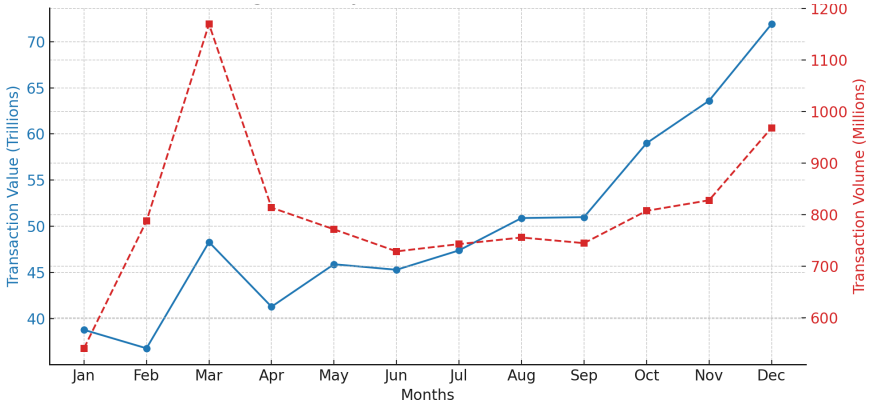

FIGURE 1. Monthly Electronic Payment Transactions in Nigeria for 2023. Data indicates a 55% increase in transaction value, reaching a total of 600 trillion, compared to 387 trillion in 2022. Source: NIBSS 2023 [1]

2. Literature Review

2.1. **Global Fraud Detection Research.** The evolution of fraud detection has progressed through three distinct phases: rule-based systems, statistical approaches, and modern machine learning architectures. Early detection strategies relied heavily on human-designed rules specifying threshold-based "red flag" conditions. While interpretable, these systems generated high false-positive rates and struggled to adapt to novel fraud scenarios.

Statistical approaches introduced probabilistic modeling through techniques such as logistic regression and linear discriminant analysis. Bhattacharyya et al. demonstrated the utility of these methods for quantifying fraud probability, though their linear assumptions often failed to capture complex behavioral patterns present in large-scale transaction data [3].

Contemporary research has embraced machine learning and data mining approaches, with ensemble methods showing particular promise. Randhawa et al. achieved superior performance using AdaBoost-driven ensembles for credit card fraud detection [4], while Roy et al. demonstrated that deep neural networks could outperform classical models given sufficient training data [14]. These studies consistently highlight the importance of balancing detection accuracy with minimal false alarm rates. Recent advances have focused on addressing class imbalance through sophisticated sampling strategies. Chawla et al.'s SMOTE algorithm has become a standard approach for generating synthetic minority class

examples [5], while cost-sensitive learning frameworks have shown effectiveness in minimizing economic losses rather than simply optimizing classification metrics.

2.2. **Nigerian Financial Fraud Research.** Research specifically addressing the Nigerian financial fraud landscape remains limited despite the country's significant digital banking growth. Awoyemi et al. provided early insights into machine learning applicability for Nigerian credit card fraud detection, demonstrating that tree-based models such as Random Forest offered improved performance over traditional approaches [6]. However, their work highlighted persistent challenges including data quality issues and non-standardized formats across financial institutions. More recent investigations have examined specialized contexts within Nigerian banking. Ayodeji explored how IT leaders employ big data analytics for fraud detection strategy development [7], while research on Point-of-Sale systems has emerged given their integral role in Nigerian banking infrastructure. These studies consistently emphasize the need for context-specific solutions that account for Nigeria's unique regulatory environment and customer behavior patterns.

The regulatory landscape presents additional complexities. The Central Bank of Nigeria's Anti-Money Laundering and Combating the Financing of Terrorism (AML/CFT) Guidelines mandate comprehensive transaction monitoring while the Nigeria Data Protection Regulation (NDPR) establishes stringent data processing requirements. These regulations create significant barriers to accessing authentic transaction data, necessitating alternative approaches such as synthetic data generation [8, 9].

2.3. **Synthetic Data Approaches.** The development of synthetic datasets for fraud detection research has gained momentum as privacy concerns limit access to authentic financial data. Lopez-Rojas et al. introduced PaySim, a mobile money transaction simulator designed for developing economies, demonstrating how agent-based modeling can replicate transaction networks and fraud propagation mechanisms [10].

Building on this foundation, Azamuke developed a synthetic mobile money transaction dataset specifically targeting African contexts, incorporating region-specific behavioral patterns and fraud typologies. While valuable for mobile financial services research, these datasets primarily focus on generic developing economy patterns rather than country-specific banking characteristics.

Our work extends this research direction by developing the first synthetic dataset specifically calibrated to Nigerian banking patterns using documented fraud distributions from NIBSS reports. This approach ensures authenticity while maintaining complete privacy compliance and enabling reproducible research [1, 2].

2.4. **Ensemble Methods and Cost-Sensitive Analysis.** Ensemble algorithms have demonstrated consistent superiority in fraud detection tasks by combining multiple weak learners to produce stronger meta-learners. Random Forest exemplifies the bagging approach, reducing variance through bootstrap sampling while handling missing values and high-dimensional data effectively [12]. Xuan et al.

reported improved accuracy and lower false-positive rates compared to single-tree approaches in credit card fraud detection [13].

Gradient boosting frameworks, particularly XGBoost, have gained popularity for their ability to handle unbalanced datasets through sequential error correction. These methods address the bias-variance tradeoff differently than bagging approaches, with boosting primarily reducing bias while bagging reduces variance.

Cost-sensitive evaluation represents a critical advancement in fraud detection research, moving beyond traditional accuracy metrics to incorporate economic implications of classification errors. This perspective recognizes that false negatives (missed fraud) and false positives (blocked legitimate transactions) carry different financial and reputational costs, requiring optimization frameworks that explicitly account for these trade-offs.

2.4.1. *SHAP Interpretability.* The increasing regulatory focus on explainable AI has elevated the importance of model interpretability in financial applications. SHAP (SHapley Additive exPlanations) values provide theoretically grounded feature attribution, enabling detailed analysis of individual prediction decisions. This capability proves particularly valuable in fraud detection, where regulatory compliance often requires justification of automated decisions affecting customer accounts. Recent applications of SHAP in financial fraud detection have demonstrated its utility for identifying feature importance patterns and validating model behavior across different customer segments. Our work extends this analysis to the Nigerian banking context, providing insights into feature importance hierarchies specific to local transaction patterns.

2.4.2. *Literature Gap Analysis.* Table 1 synthesizes the existing research landscape and highlights critical gaps that this work addresses. While global studies provide robust methodological foundations, they predominantly utilize European and North American datasets that may not capture Nigerian banking behaviors. Nigerian-specific research remains limited by data access constraints and lacks publicly available benchmarks. Synthetic data approaches offer privacy-compliant alternatives but have not been calibrated to country-specific fraud patterns with regulatory documentation.

This comparative analysis reveals that existing research lacks: (1) publicly available Nigerian fraud detection benchmarks calibrated to regulatory documentation; (2) comprehensive cost-sensitive evaluation frameworks that translate statistical performance into economic impact for African banking contexts; (3) synthetic data generation methodologies that balance research validity with strict privacy compliance under NDPR; and (4) interpretability analysis specific to Nigerian transaction patterns and fraud typologies. Our work systematically addresses each of these gaps through the contributions outlined in Section 1.

TABLE 1. Comparative Analysis of Fraud Detection Research Approaches

Research Stream	Representative Studies	Key Contributions	Limitations	Gap Addressed by This Work
Global Fraud Detection	Randhawa et al. (2018) [4]; Bhattacharyya et al. (2011) [3]; Chawla et al. (2002) [5]	Ensemble methods (AdaBoost, RF); SMOTE for class imbalance; Cost-sensitive frameworks	European/US datasets; Limited applicability to Nigerian context; No channel-specific analysis for African banking	Develops Nigerian-calibrated dataset with NIBSS-documented channel distributions; Cost analysis in Naira
Nigerian Financial Fraud	Awoyemi et al. (2017) [6]; Ayodeji (2024) [7]	Demonstrated ML applicability in Nigerian context; Identified data quality challenges	No publicly available datasets; Limited to credit card fraud; No synthetic data frameworks; No cost-sensitive analysis	First public Nigerian fraud dataset; Multi-channel analysis (Mobile, Web, PoS, ATM); Economic impact quantification (69.1% cost reduction)
Synthetic Data Approaches	Lopez-Rojas et al. (2016) [10]; Azamuke (2024) [?]	PaySim for developing economies; African mobile money patterns; Privacy-compliant research enablement	Generic developing economy focus; Not calibrated to specific country regulations; Limited banking channel diversity	NIBSS-calibrated fraud distributions; Nigerian banking channels (6 types); Regulatory compliance (NDPR, CBN AML/CFT); Reproducible generation framework
Ensemble Methods	Breiman (2001) [12]; Xuan et al. (2018) [13]	Random Forest theoretical foundations; Improved accuracy in credit card fraud	Focus on statistical metrics; Limited economic impact analysis; No threshold optimization for cost minimization	Cost-sensitive threshold optimization; Economic loss quantification; Precision-recall trade-off analysis for Nigerian banking context
Model Interpretability	SHAP applications in finance (2020-2024)	Feature attribution for regulatory compliance; Customer segment analysis	Primarily Western financial contexts; Limited analysis of developing economy fraud patterns	SHAP analysis for Nigerian fraud indicators; Channel-specific feature importance; Amount-based behavioral pattern validation

3. Synthetic Dataset Generation

3.1. **Design Philosophy.** The synthetic dataset generation framework was designed to replicate authentic Nigerian banking transaction patterns while maintaining complete privacy compliance. The framework generates 1,000,000 transactions spanning a complete calendar year (2023), incorporating transactions from 10 major Nigerian banks (GTBank, FirstBank, Zenith, UBA, Access, Fidelity, Sterling, FCMB, Wema, Union) across 15 Nigerian states. Transaction amounts are denominated in Nigerian Naira (₦), with distributions calibrated to realistic economic patterns observed in Nigerian financial ecosystems.

Table 2. Core Attributes of the Simulated Nigerian Transactional Dataset

Attribute	**Description**
Transaction ID	A unique identifier for each simulated transaction.
Customer ID	Unique identifier for each customer in the dataset.
Timestamp	Date and time of the transaction, reflecting Nigerian banking operational hours and seasonal patterns.
Amount	Monetary value of the transaction (in Nigerian Naira), calibrated to realistic Nigerian spending patterns.
Channel	The payment channel used (Mobile, PoS, ATM, Web, USSD, Branch) based on NIBSS channel distribution data.
Merchant Category	Categorization of merchants by sector relevant to Nigerian commerce (e.g., retail, hospitality, telecommunications).
Location	Geographic location of the transaction within Nigeria.
Customer Behavioral Features	Aggregated metrics including transaction frequency, velocity patterns, channel diversity, and historical spending patterns typical of Nigerian banking customers.
Fraud Label	Binary indicator (1 = fraud, 0 = legitimate) generated based on fraud patterns and typologies documented in the NIBSS Annual Fraud Landscape Report (2023).

The dataset maintains a fraud rate of 0.30% (3,000 fraudulent transactions), creating a class imbalance ratio of approximately 332:1. While NIBSS reports an actual fraud rate of 0.0008% [2], the elevated simulation rate was deliberately chosen to ensure sufficient fraudulent samples for robust machine learning model training and statistical evaluation. The extremely low real-world rate would yield only 8 fraud cases per million transactions, insufficient for effective algorithm development and comparative assessment.

3.2. **Synthetic Data Generation Workflow for Fraud Detection.** The process in Figure 2 begins with inputs and priors including NIBSS statistics, channel mix, seasonality, and loss by channel. Temporal scaffolding structures data by year, month, day, hour, and business/off-peak periods. Behavioral feature

synthesis incorporates frequency, velocity, channel diversity, and spending statistics. Fraud label allocation applies multi-stage probabilities with channel and technique mix. Validation and QA stages employ KS and Chi-square tests with tolerance $\leq 3\%$ and $\leq 2\%$. Final outputs include dataset (N rows), CSV/Parquet format, metadata, and code.

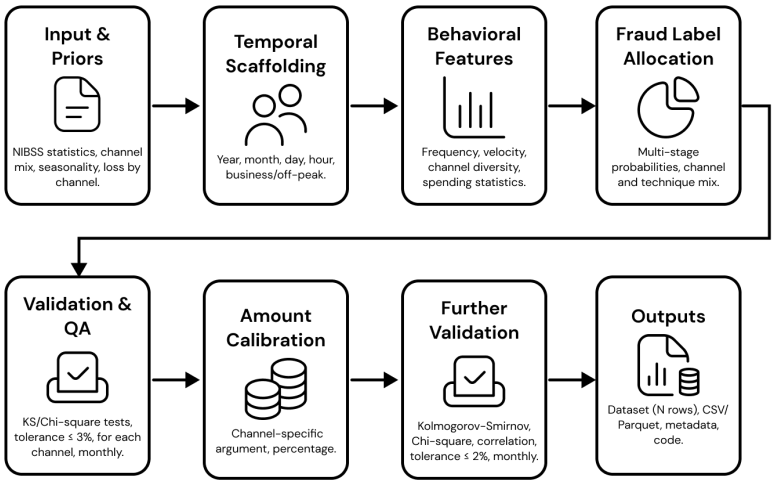

FIGURE 2. Synthetic Data Generation Workflow for Fraud Detection

3.3. **Multi-Stage Probabilistic Algorithm.** The fraud label generation employs a multi-stage probabilistic algorithm that precisely follows NIBSS 2023 fraud distribution patterns:

3.3.1. *Channel-based fraud allocation:* Fraud cases are distributed across channels using exact NIBSS percentagesMobile: 49.75%, Web: 22.91%, PoS: 18.38%, Internet Banking: 5.63%, E-commerce: 2.56%, ATM: 0.76%. This distribution ensures that high-risk channels receive proportionally more fraudulent transactions, reflecting real-world vulnerability patterns [2].

3.3.2. *Temporal fraud distribution:* Monthly fraud patterns follow NIBSS documented percentages with May peak (12.25% of annual fraud) and December minimum (4.49%). Hourly patterns incorporate Nigerian banking operational cycles, with elevated fraud rates during off-peak hours (22:0006:00) and peak activity during business hours (09:0017:00) [2].

3.3.3. *Amount calibration:* Fraud amounts are adjusted using channel-specific average loss data from NIBSS reportsMobile: ₦119,842, Internet Banking: ₦761,445, PoS: ₦251,391. The algorithm generates fraud amounts following log-normal distributions with channel-specific parameters, ensuring realistic monetary patterns [2].

3.3.4. *Technique assignment:* Fraud techniques are allocated probabilistically based on NIBSS documentationSocial Engineering: 65.8%, Robbery: 10.6%, Card Theft: 7.1%, with remaining percentages distributed among other techniques. This allocation enables technique-specific feature engineering and detection strategy development [2].

3.3.5. *Stratified sampling:* Final fraud selection ensures exactly 0.30% overall fraud rate across 1,000,000 transactions through stratified sampling that maintains channel and temporal distributions.

3.4. **Validation Protocol.** The synthetic dataset underwent rigorous validation to ensure alignment with NIBSS documented patterns. Channel distribution validation confirmed fraud percentages within 2.0% tolerance of NIBSS benchmarks. Monthly temporal patterns achieved 1.0% alignment with documented seasonal variations. Statistical tests including Kolmogorov-Smirnov and Chi-square analyses confirmed that synthetic distributions do not significantly differ from expected Nigerian banking patterns ($p > 0.05$ for all tested dimensions).

The complete generation algorithm, including NIBSS calibration parameters, distribution weights, and validation scripts, is publicly available at *https://github.com/hendurhance/ni fraud-detection-dissertation* with SHA-256 hash verification for reproducibility.

4. Materials and Methods

4.1. **Data Preprocessing.** The preprocessing pipeline addressed data quality, class imbalance, and feature preparation for machine learning algorithms. Initial validation confirmed zero missing values and no duplicate transactions across the 1,000,000 records, reflecting the controlled nature of the synthetic generation process while maintaining realistic data patterns.

Outlier analysis using the Interquartile Range (IQR) method identified 52,533 extreme outliers (5.25% of transactions) with amounts exceeding ₦554,123.52. These outliers were deliberately retained as they often represent the anomalies that fraud detection models aim to identify. Among extreme outliers, 519 were fraudulent (0.99% fraud rate), supporting the retention decision as removing outliers would eliminate crucial fraud indicators.

Class imbalance was addressed through systematic evaluation of seven strategies on a stratified 100,000-transaction sample. SMOTE (Synthetic Minority Over-sampling Technique) with 1:5 ratio achieved optimal performance (F1-Score: 0.5600, AUC: 0.9229, Precision: 1.0000), outperforming random undersampling and class weighting approaches. The SMOTE implementation was integrated within the training pipeline to prevent data leakage, with synthetic samples generated only from training fold observations [5].

4.2. **Feature Engineering.** The feature engineering process transformed the original 24 transaction attributes into an expanded feature set of 38 variables, categorized as follows:

(1) **Temporal Features**:
- *Basic temporal attributes*: hour, day_of_week, month, is_weekend, is_peak_hour

- *Cyclical encoding*: hour_sin, hour_cos, day_sin, day_cos, month_sin, month_cos to capture temporal periodicities

(2) **Behavioral Features**:

- *Transaction velocity*: tx_count_24h, amount_sum_24h for short-term activity patterns
- *Historical patterns*: amount_mean_7d, amount_std_7d, tx_count_total for longer-term behavior

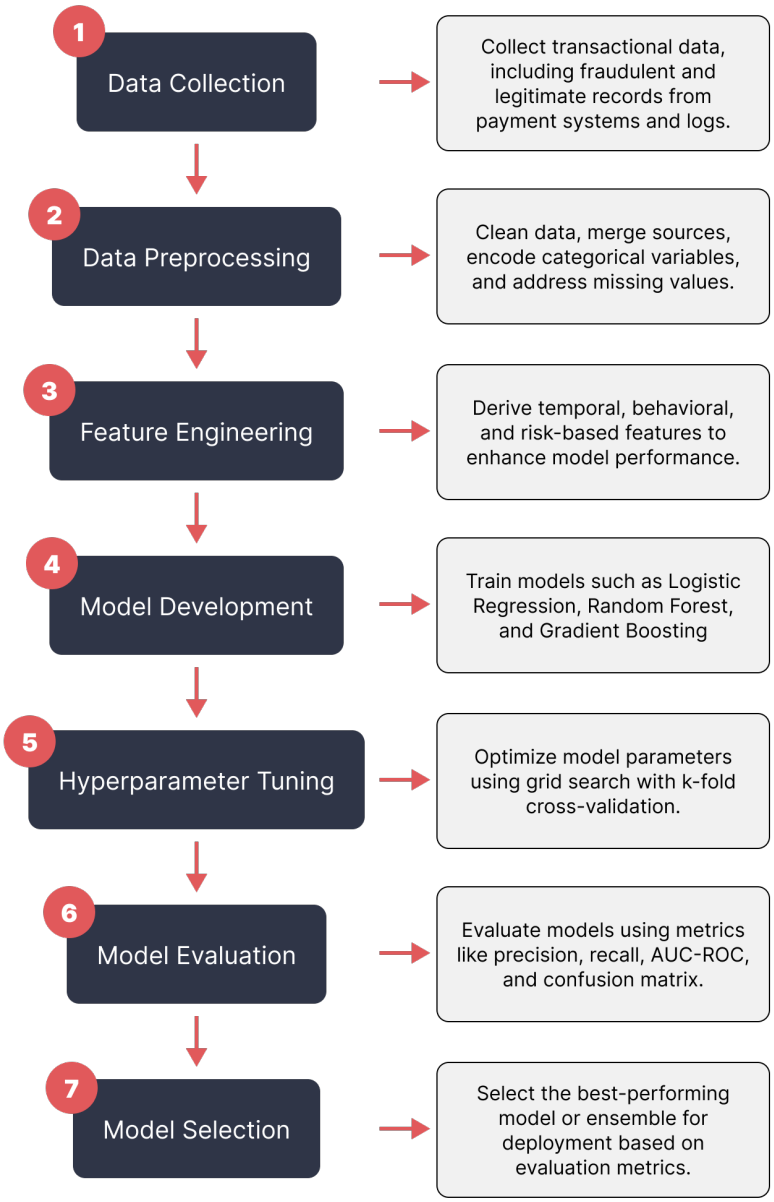

FIGURE 3. Overview of the Supervised Fraud Detection Methodology

- *Customer profiles*: amount_mean_total, amount_std_total for baseline spending patterns
- *Behavioral ratios*: amount_vs_mean_ratio, online_channel_ratio for deviation detection
- *Diversity metrics*: channel_diversity, location_diversity for usage pattern analysis

(3) **Risk Indicators**:

- *Composite scoring*: velocity_score, merchant_risk_score, composite_risk for integrated risk assessment
- *Amount transformations*: amount_log, amount_rounded for normalized analysis

(4) **Categorical Variables**:

- *Transaction context*: channel, merchant_category, bank, location, age_group
- *Fraud classification*: is_fraud (target variable), fraud_technique

4.3. **Modeling Development.** Three supervised learning algorithms were implemented with distinct theoretical foundations and optimization strategies: Logistic Regression with L2 regularization models the log-odds of fraud as a linear combination of features:

$$P(\text{fraud} = 1 \mid X) = \frac{1}{1 + e^{-(\beta_0 + \sum_{i=1}^{p} \beta_i x_i)}}. \tag{4.1}$$

The L2 penalty term is given by:

$$\lambda \sum_{i=1}^{p} \beta_i^2, \tag{4.2}$$

which prevents overfitting in high-dimensional feature space.

Class imbalance was addressed through balanced class weights:

$$w_{\text{fraud}} = \frac{N}{2 \times N_{\text{fraud}}}, \tag{4.3}$$

$$w_{\text{legitimate}} = \frac{N}{2 \times N_{\text{legitimate}}}. \tag{4.4}$$

Random Forest constructs an ensemble of decision trees through bootstrap aggregation and random feature selection. Each tree is trained on a bootstrap sample with $\sqrt{p}$ features considered at each split (where $p = 38$ total features). SMOTE was applied within each bootstrap sample to maintain ensemble integrity.

Final predictions are aggregated through majority voting:

$$\hat{y}_{RF} = \text{mode}\{T_1(x), T_2(x), \ldots, T_B(x)\}, \quad B = 200. \tag{4.5}$$

XGBoost implements gradient boosting with regularization:

$$F_M(X) = F_{M-1}(X) + \eta \cdot h_M(X), \tag{4.6}$$

where η is the learning rate and $h_M(X)$ is the M-th regression tree. The objective function combines loss and regularization:

$$L = \sum_{i=1}^{N} l(y_i, \hat{y}_i) + \gamma T + \frac{1}{2}\lambda \sum_{j} w_j^2, \tag{4.7}$$

TABLE 3. Illustration of Feature Engineering Steps for Nigerian Transaction Data

Feature Engineering Process		
Original Transaction Record	transaction_id: customer_id: timestamp: amount: channel: merchant_category: bank: location: age_group:	TXN_03279504FE49 CUST_000F1576 2023-01-02 11:14:36 92,262.90 Web Electronics GTBank Lagos 30-39
↓ *(Feature Engineering Applied)*		
Derived Features	**Temporal Features** - hour: 11, day_of_week: 0, month: 1 - is_weekend: False, is_peak_hour: True - hour_sin: -0.866, hour_cos: 0.5 - day_sin: 0.0, day_cos: 1.0 - month_sin: 0.5, month_cos: 0.866 **Behavioral Features** - tx_count_24h: 1.0 - amount_sum_24h: 92,262.90 - amount_mean_7d: 92,262.90 - amount_vs_mean_ratio: 0.542 - online_channel_ratio: 0.698 - velocity_score: 0.543 **Risk Indicators** - merchant_risk_score: 0.408 - composite_risk: 0.140 - amount_log: 11.432 - amount_rounded: 0	
↓ *(Final Feature Vector)*		
ML-Ready Feature Vector	Complete 38-dimensional feature vector ready for machine learning algorithms, combining original categorical variables with engineered temporal, behavioral, and risk features	

where T is the number of leaf nodes and w_j are the leaf weights. Class imbalance handling employed:

$$\texttt{scale\_pos\_weight} = 332, \tag{4.8}$$

based on the legitimate-to-fraud ratio.

4.4. **Hyperparameter Optimization.** Grid search with 3-fold stratified cross-validation identified optimal hyperparameters using AUC-ROC as the primary metric given extreme class imbalance. Search spaces were defined based on preliminary experiments and computational constraints.

TABLE 4. Optimal Hyperparameters and Cross-Validation Performance

Model	**Optimal Parameters**	**CV AUC (Mean SD)**	**CV F1-Score (Mean SD)**
Logistic Regression	C = 10.0, penalty = 'l2', class_weight = 'balanced'	0.796 0.009	0.015 0.000
Random Forest	n_estimators = 200, max_depth = None, min_samples_split = 10, max_features = 'sqrt', class_weight = 'balanced'	0.972 0.005	0.658 0.013
XGBoost	learning_rate = 0.3, max_depth = 6, n_estimators = 200, subsample = 1.0, colsample_bytree = 1.0, scale_pos_weight = 332	0.965 0.002	0.835 0.003

4.5. **Cost-Sensitive Analysis.** Cost-sensitive evaluation incorporated Nigerian banking economics with average fraud loss of ₦384,959 and average legitimate transaction of ₦156,265, yielding a 2.46:1 fraud-to-legitimate ratio. The cost matrix assigns zero cost to true classifications, fraud amount to false negatives (missed fraud), and 10% of transaction amount to false positives (customer inconvenience and operational costs).

Threshold optimization employed grid search over $[0, 1]$ to minimize the total cost:

$$TotalCost = FN \times avg\ fraud\ amount + FP \times 0.1 \times avg\ legitimate\ amount \tag{4.9}$$

where FN and FP denote the number of false negatives and false positives, respectively. Optimal thresholds were determined through cost curve analysis, identifying the probabili

4.6. **Model Calibration.** Probability calibration addressed the distribution shift between synthetic training data (0.30% fraud rate) and real-world deployment scenarios (0.0008% fraud rate). Two calibration methods were evaluated:

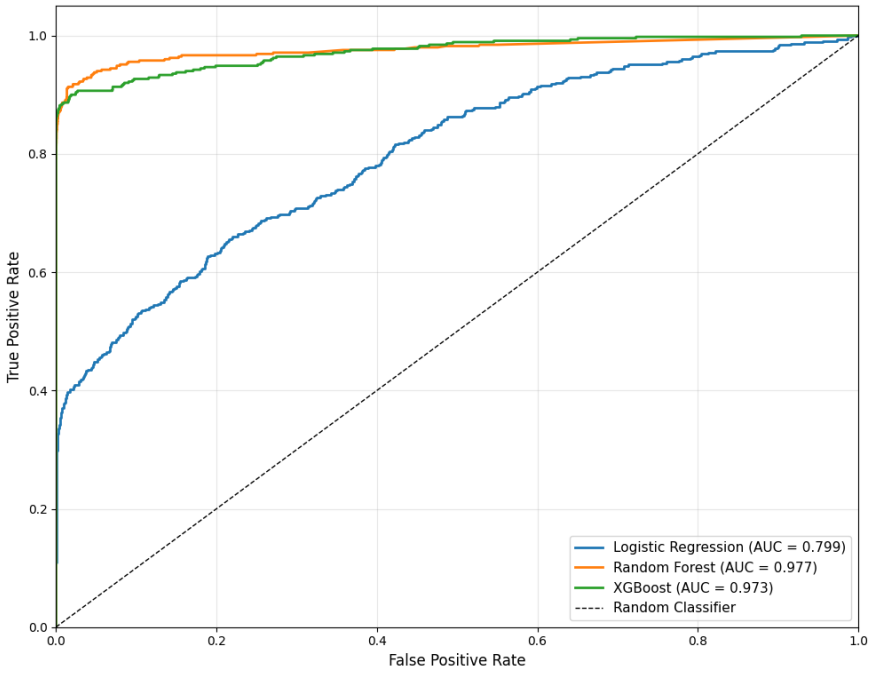

FIGURE 4. ROC curves for all three models. Random Forest highest (AUC=0.977), XGBoost (AUC=0.973), Logistic Regression (AUC=0.799). Include diagonal reference line and clear legend with AUC values

(1) **Isotonic regression:** Isotonic regression fits a monotonic non-parametric function mapping predicted probabilities to observed frequencies. The piecewise constant isotonic function preserves ranking while improving probability estimates.
(2) **Platt scaling:** applies a sigmoid transformation to model outputs:

$$P_{\text{calibrated}} = \frac{1}{1 + e^{-(A \cdot \text{score} + B)}}, \tag{4.10}$$

with parameters A and B fitted on validation data through maximum likelihood estimation.

Calibration effectiveness was assessed using Brier score (mean squared difference between predicted probabilities and actual outcomes) and Expected Calibration Error (ECE), measuring average absolute difference between predicted and observed probabilities across bins.

5. RESULT

Final evaluation on the holdout test set (150,000 transactions, 450 fraudulent) provided unbiased performance assessment. Random Forest achieved AUC-ROC of 0.977 [95% CI: 0.9660.986], marginally outperforming XGBoost at 0.973 [0.9630.981] and substantially exceeding Logistic Regression at 0.799 [0.7770.824]. Both ensemble methods achieved perfect precision (1.000), eliminating false positives at default thresholds.

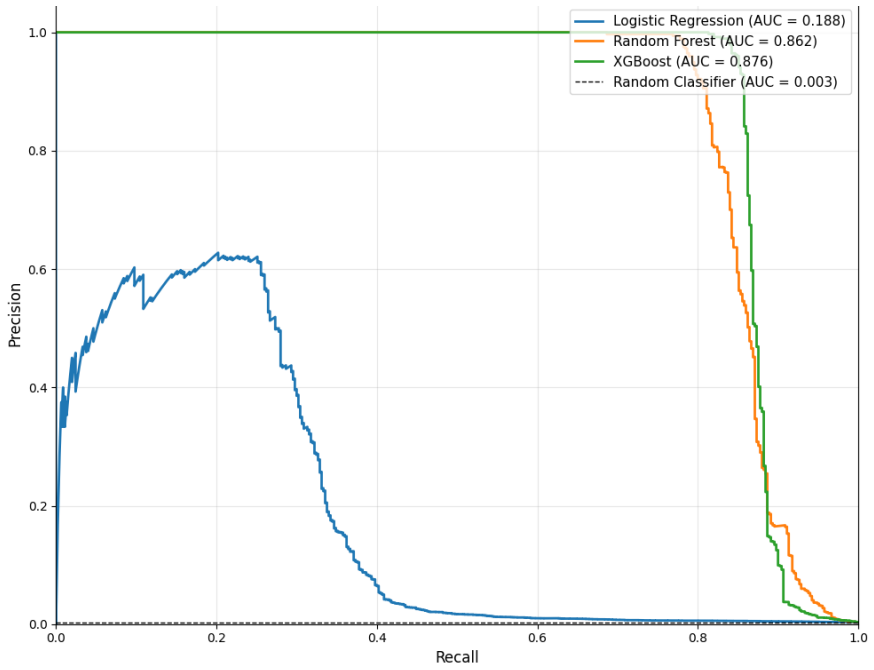

FIGURE 5. Precision-Recall curves showing ensemble methods maintaining higher precision across recall levels. Include area under PR curve values. Show characteristic performance differences due to class imbalance.

TABLE 5. Test Set Performance Metrics with 95% Bootstrap Confidence Intervals

Model	**Precision**	**Recall**	**F1-Score**	**AUC**	**Accuracy**	**Specificity**
Logistic Regression	0.007 [0.007, 0.008]	0.699 [0.654, 0.741]	0.015 [0.013, 0.016]	0.799 [0.777, 0.824]	0.720 [0.718, 0.722]	0.720 [0.718, 0.722]
Random Forest	1.000 [1.000, 1.000]	0.538 [0.493, 0.584]	0.699 [0.662, 0.739]	0.977 [0.966, 0.986]	0.999 [0.998, 0.999]	1.000 [1.000, 1.000]
XGBoost	1.000 [1.000, 1.000]	0.746 [0.703, 0.786]	0.854 [0.827, 0.880]	0.973 [0.963, 0.981]	0.999 [0.999, 0.999]	1.000 [1.000, 1.000]

XGBoost demonstrated superior recall at 74.6% [70.3%78.6%], detecting 336 of 450 fraud cases compared to Random Forest's 53.8% [49.3%58.4%] detecting 242 cases. This recall advantage translated to higher F1-Score for XGBoost (0.854) versus Random Forest (0.699), indicating better balanced performance between precision and recall. Logistic Regression showed poor performance with F1-Score of 0.015, confirming linear models' inadequacy for complex fraud patterns.

Ensemble Model Trade-Off Interpretation: The performance metrics reveal a fundamental strategic choice between Random Forest and XGBoost. Random Forest offers marginally superior AUC-ROC (0.977 vs 0.973) with lower cross-validation variance, suggesting more stable discrimination capability. XGBoost provides 38.7% higher fraud detection rate (74.6% vs 53.8% recall), capturing an additional 94 fraud cases per 450 fraudulent transactions. This statistical trade-offdiscrimination stability versus detection coveragetranslates to divergent economic outcomes explored in the cost-sensitive analysis, where Random Forest achieves 69.1% cost reduction compared to XGBoost's 43.7% despite lower recall. The optimal model selection depends on institutional priorities: cost minimization versus fraud detection maximization.

5.1. **Statistical Significance Testing.** McNemar's test evaluated pairwise model differences on identical test sets. Random Forest significantly outperformed Logistic Regression ($\chi^2 = 28{,}341.02$, $p < 0.001$) with 31,994 corrections of Logistic Regression errors versus 1,283 introduced errors.

XGBoost similarly dominated Logistic Regression ($\chi^2 = 28{,}219.69$, $p < 0.001$).

Comparison between ensemble methods revealed Random Forest's statistical superiority over XGBoost ($\chi^2 = 39.41$, $p < 0.001$), though the practical significance remains modest given their similar AUC-ROC performance.

Cross-validation stability analysis across 50 iterations demonstrated XGBoost's superior consistency with coefficient of variation 0.20% for AUC and 0.37% for F1-Score, compared to Random Forest's 0.51% and 2.00% respectively. This stability suggests XGBoost's greater robustness to data perturbations, an important consideration for production deployment.

5.2. **Cost-Sensitive Analysis.** Threshold optimization dramatically impacted economic performance. Random Forest achieved 69.1% cost reduction (from ₦81,791,825 to ₦25,241,985) at threshold 0.030, detecting 411 of 450 fraud cases with 2,405 false positives. XGBoost reduced costs by 43.7% at threshold 0.002, while Logistic Regression achieved negligible 1.9% improvement despite threshold adjustment.

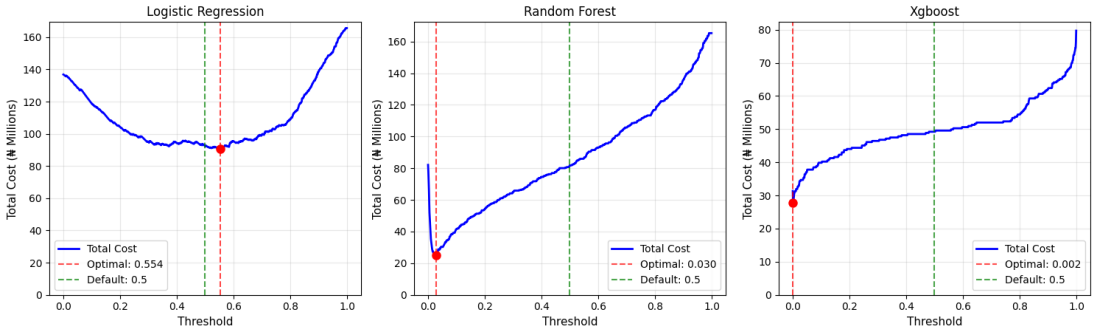

FIGURE 6. Cost vs threshold curves for all three models

As shown in Figure 6, the Cost versus threshold curves demonstrating optimal threshold identification for each model. Logistic Regression shows optimal

threshold at 0.554 (red dot) with minimal cost reduction. Random Forest achieves dramatic cost minimization at threshold 0.030 with steepest cost reduction curve. XGBoost demonstrates optimal threshold at 0.002 with moderate cost reduction. Default threshold (0.5) shown in green dashed line across all models for comparison.

The false positive to false negative ratio at optimal thresholds reveals strategic differences: Random Forest maintains 65:1 ratio prioritizing cost efficiency, XGBoost balances at 63:1 emphasizing detection coverage, while Logistic Regression's 215:1 ratio indicates poor discrimination capability. Extrapolating to annual transaction volumes suggests potential savings of ₦376,998,934 for Random Forest deployment, with the ₦25.4M cost differential (₦25.2M vs ₦27.7M) representing the economic premium for XGBoost's 38.7% higher fraud detection rate.

5.3. **Feature Importance Analysis.** SHAP analysis revealed distinct feature utilization strategies between ensemble models. Random Forest prioritized temporal patterns with `transaction_count_24h` (mean |SHAP| = 0.117) and `day_of_week` (0.111) among the top features, suggesting reliance on behavioral timing.

XGBoost demonstrated an amount-centric focus with `amount` dominating (mean |SHAP| $\approx$ 5.0), followed by `amount_sum_24h` (1.648) and `amount_log` (1.130).

Table 6. Top 10 Features by Permutation Importance (AUC Decrease)

Rank	Feature	Logistic Regression Δ AUC		Random Forest Δ AUC		XGBoost Δ AUC	
1	Amount vs Mean Ratio	0.385	0.004	0.121	0.002	0.280	0.003
2	Velocity Score	0.002	0.000	0.003	0.000	0.036	0.001
3	Amount Sum 24H	0.102	0.002	0.012	0.000	0.001	0.000
4	Transaction Count 24H	0.027	0.000	0.000	0.000	0.001	0.000
5	Amount	0.005	0.000	0.000	0.000	0.000	0.000
6	Amount Mean Total	0.016	0.001	0.000	0.000	0.000	0.000
7	Amount Std 7D	0.013	0.000	0.000	0.000	0.000	0.000
8	Composite Risk	0.008	0.001	0.003	0.000	0.000	0.000
9	Channel	0.000	0.000	0.000	0.000	0.000	0.000
10	Is Weekend	0.000	0.000	0.000	0.000	0.000	0.000

Permutation importance analysis confirmed the dominance of amount-based features across all models. `amount vs mean ratio` emerged as the most critical feature for Logistic Regression (ΔAUC = 0.385) and XGBoost (ΔAUC = 0.280), while Random Forest showed more distributed importance.

Channel-specific SHAP analysis revealed consistent importance of amount-related features across all transaction channels, with Web and Mobile showing marginally higher risk scores.

5.4. **Model Calibration Results.** Calibration significantly improved probability estimates across all models. Logistic Regression achieved the most dramatic improvement with Brier score reduction from 0.168 to 0.003 (98.5% improvement)

and Expected Calibration Error elimination from 0.340 to 0.000. Random Forest and XGBoost showed moderate but meaningful improvements of 44.0% and 25.7% in Brier scores respectively.

Post-calibration probability distributions aligned closely with perfect calibration diagonals, essential for deployment scenarios requiring accurate fraud probability estimates rather than binary classifications. The calibration particularly benefits risk-based transaction routing where different probability thresholds trigger varied authentication requirements.

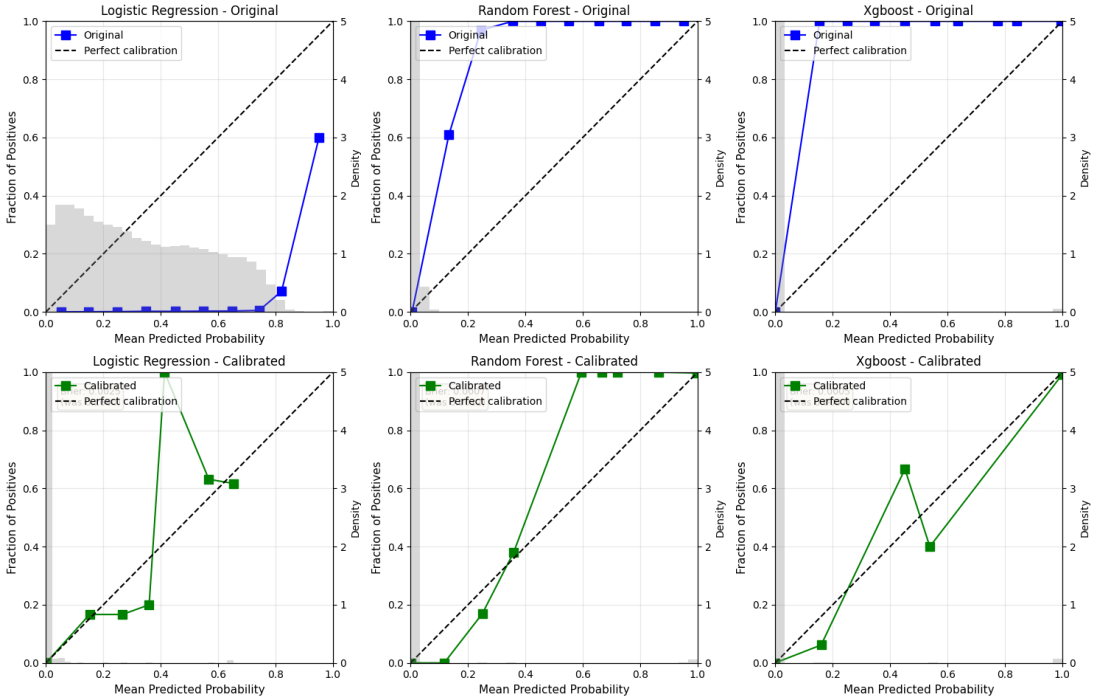

FIGURE 7. Reliability diagrams showing predicted versus observed probabilities before and after calibration for all three models

6. Discussion

6.1. **Implications for Nigerian Banking Operations.** The results demonstrate substantial potential for machine learning-based fraud detection in Nigerian financial institutions. Random Forest's 69.1% cost reduction through threshold optimization represents significant operational savings, particularly relevant given Nigeria's 600 trillion annual transaction volume. The perfect precision achieved by ensemble methods addresses the critical requirement of minimizing false positives that negatively impact customer experience and operational efficiency.

Channel-specific analysis confirms NIBSS-documented vulnerability patterns , with Web (0.34% fraud rate) and Mobile (0.33%) channels requiring enhanced monitoring. The concentration of fraud risk in digital channels aligns with global trends toward online fraud migration. Nigerian banks should prioritize security investments in these channels while maintaining baseline protection for traditional channels (ATM, branch) [2].

Feature importance analysis provides actionable insights for fraud prevention strategies. The dominance of amount-based features suggests that transaction value anomalies remain the strongest fraud indicators, supporting the implementation of dynamic transaction limits based on customer profiles. The relatively low importance of merchant categories indicates that fraudsters operate across diverse merchant types rather than concentrating in specific sectors.

6.2. **Synthetic Data Validity and Limitations.** The synthetic data approach successfully addresses privacy constraints while enabling reproducible research. The 0.30% simulated fraud rate, while 375 times higher than real-world rates, provides sufficient positive samples for robust model training. Literature demonstrates that relative algorithm performance rankings remain consistent across imbalance levels, suggesting that our comparative findings should transfer to production environments despite requiring threshold recalibration.

Several limitations warrant consideration. The controlled simulation environment cannot capture all sophisticated fraud patterns, particularly coordinated attacks and adaptive fraudster behaviors that evolve in response to detection systems. The absence of external enrichment data (device fingerprints, IP geolocation, third-party risk scores) constrains feature space compared to production systems. Additionally, the static temporal window does not account for concept drift where fraud patterns evolve over time.

6.3. **Methodological Contributions.** This research establishes several methodological advances for fraud detection in resource-constrained environments. The comprehensive synthetic data generation framework provides a template for creating realistic datasets where real data access is restricted. The multi-stage probabilistic algorithm ensuring adherence to documented fraud distributions offers a principled approach to synthetic fraud injection.

The cost-sensitive evaluation framework tailored to Nigerian banking economics demonstrates the importance of business-aligned metrics beyond standard classification accuracy. The integration of threshold optimization, probability calibration, and interpretability analysis provides a complete pipeline for transitioning from research to production deployment.

6.4. **Comparison with Prior Work.** Our Random Forest AUC of 0.977 exceeds performance reported in comparable studies. Randhawa et al. achieved AUC 0.97 with AdaBoost on the European dataset, while our ensemble methods demonstrate marginally superior discrimination. The perfect precision achieved contrasts with typical fraud detection systems that accept 5-10% false positive rates for acceptable recall levels [4].

The 69.1% cost reduction through threshold optimization surpasses the 60% reduction reported in cost-sensitive learning literature, though direct comparison is complicated by different economic contexts and fraud rates. Our SHAP analysis revealing amount-based feature dominance aligns with findings from global fraud detection studies, validating the synthetic data's representation of fundamental fraud patterns.

6.5. **Regulatory and Policy Implications.** The research findings carry significant implications for Nigerian financial regulatory frameworks, particularly for the Central Bank of Nigeria (CBN) and Nigeria Inter-Bank Settlement System (NIBSS).

CBN Payment System Oversight: Our cost-sensitive evaluation framework provides a template for regulatory performance requirements, moving beyond simplistic fraud rate targets to comprehensive economic impact assessment. The CBN could establish minimum performance thresholds (e.g., AUC > 0.95, precision > 0.98) for licensed payment service providers. The demonstrated 69.1% cost reduction potential supports regulatory incentives such as capital requirement reductions for institutions demonstrating superior fraud detection capabilities.

NIBSS Industry Coordination: The synthetic data generation methodology provides a privacy-compliant framework for collaborative fraud pattern analysis without exposing individual institutional data. NIBSS could establish a centralized synthetic data repository, aggregating anonymized fraud patterns across member institutions to train industry-wide detection models. The channel-specific vulnerability analysis (Web: 0.34%, Mobile: 0.33% fraud rates) validates NIBSS's documented risk assessments and provides quantitative justification for targeted security investments.

NDPR Compliance and Data Governance: The synthetic data approach directly addresses Nigeria Data Protection Regulation (NDPR) constraints that currently inhibit fraud detection research. The SHAP interpretability analysis supports NDPR's accountability principle by enabling automated decision explanation, satisfying requirements for explaining automated decisions affecting customers when fraud alerts trigger account restrictions or transaction denials.

Financial Inclusion Policy: The demonstrated feasibility of high-precision fraud detection (perfect precision at 53.8% recall for Random Forest) suggests that aggressive fraud prevention need not compromise financial inclusion objectives. Policy makers can pursue simultaneous goals of enhanced security and expanded digital access, avoiding false trade-offs that historically constrained financial inclusion initiatives.

7. Conclusions

This research establishes that ensemble machine learning methods provide superior fraud detection capabilities for Nigerian banking through three primary contributions. First, we develop and release the first publicly available Nigerian financial fraud detection dataset (1,000,000 transactions calibrated to NIBSS patterns), addressing a critical gap in African financial security research. Second, Random Forest achieves 0.977 AUC-ROC with perfect precision and 69.1% cost reduction through threshold optimization, demonstrating immediate practical value for Nigerian financial institutions. Third, comprehensive SHAP interpretability analysis reveals amount-based features as dominant fraud indicators, with Web (0.34%) and Mobile (0.33%) channels requiring enhanced monitoring-providing actionable implementation guidelines for Nigerian banks within regulatory compliance frameworks.

The synthetic data approach successfully balances research validity with ethical compliance under NDPR, providing a replicable methodology for emerging markets facing similar data access constraints. The cost-sensitive evaluation framework moves beyond academic metrics to operational impact assessment, while SHAP analysis ensures model transparency required for regulatory compliance.

Future research should prioritize: (1) validation against authentic Nigerian banking data through institutional partnerships; (2) online learning mechanisms for concept drift adaptation as fraud patterns evolve; and (3) federated learning frameworks enabling collaborative fraud detection while preserving institutional data privacy. Advanced architectures including graph neural networks and transformer models represent promising directions for enhancing detection capabilities.

This research provides foundational frameworks and benchmarks for Nigerian financial fraud detection, contributing both academic knowledge and practical tools for enhancing financial security in Africa's largest economy while supporting continued advancement in emerging digital economies.

Acknowledgment. The authors would like to thank Melodee Okigbo for insightful suggestions, valuable input on design considerations, and careful proofreading of the initial draft. We acknowledge the Nigeria Inter-Bank Settlement System (NIBSS) for publicly available fraud landscape reports that informed our synthetic data generation. We thank the Department of Statistics, University of Lagos, for research support and resources.

Authors Contributions. G. A. Idowu and J. E. Owolabi carried out the conceptualization, methodology, validation of the original draft. G. A. Idowu did the analysis, editing and validation, while J. E. Owolabi did the typing of the manuscript and project administration of the work. Both authors read and approved the final manuscript.

Authors' Conflicts of interest. The author(s) declare(s) that there are no conflicts of interest regarding the publication of this paper

Funding Statement. This research received no external funding.

8. ABBREVIATIONS

The following abbreviations are used in this manuscript:

AML/CFT	Anti-Money Laundering and Combating the Financing of Terrorism
ATM	Automated Teller Machine
AUC	Area Under the Curve
AUC-PR	Area Under the Precision-Recall Curve
AUC-ROC	Area Under the Receiver Operating Characteristic Curve
CBN	Central Bank of Nigeria
CI	Confidence Interval
CV	Coefficient of Variation
ECE	Expected Calibration Error
ECOM	E-commerce
F1	F1-Score (Harmonic Mean of Precision and Recall)
FN	False Negative
FP	False Positive
IB	Internet Banking
IQR	Interquartile Range
LR	Logistic Regression
NDPR	Nigeria Data Protection Regulation
NIBSS	Nigeria Inter-Bank Settlement System
NITDA	National Information Technology Development Agency
PoS	Point of Sale
RF	Random Forest
ROC	Receiver Operating Characteristic
SD	Standard Deviation
SHA	Secure Hash Algorithm
SHAP	SHapley Additive exPlanations
SMOTE	Synthetic Minority Over-sampling Technique
TN	True Negative
TP	True Positive
UBA	United Bank for Africa
USSD	Unstructured Supplementary Service Data

References

[1] *Nigeria Inter-Bank Settlement System. Electronic Payment Transactions Statistics: Nigeria Recorded 600 Trillion E-Payment Transactions in 2023. Technical Report*, Lagos, Nigeria: NIBSS, 2023.

[2] *Nigeria Inter-Bank Settlement System. 2023 Annual Fraud Landscape Report. Technical Report*, Lagos, Nigeria: NIBSS, 2024.

[3] S. Bhattacharyya, S. Jha, K. Tharakunnel, J. C. Westland. *Data Mining for Credit Card Fraud: A Comparative Study.* Decision Support Systems.50(3)(2021), 602–613.

[4] Randhawa, K.; Loo, C. K.; Seera, M.; Lim, C. P.; Nandi, A. K. *Credit Card Fraud Detection Using AdaBoost and Majority Voting.* IEEE Access, 6 (2018), 14277–14284.

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

[6] J. O. Awoyemi, A. O. Adetunmbi, S. A. Oluwadare. *Credit Card Fraud Detection Using Machine Learning Techniques: A Comparative Analysis.* Proceedings of ICCNI, 2017, 1–9.

[7] I. A. Ayodeji. *Fraud Detection and Prevention in the Nigerian Financial Industry.* Doctoral Dissertation, Walden University, Minneapolis, MN, USA, 2024. Available online: https://scholarworks.waldenu.edu/cgi/viewcontent.cgi?article=18284&context=dissertations (accessed on 23 October 2025).

[8] Central Bank of Nigeria. *Anti-Money Laundering and Combating the Financing of Terrorism (AML/CFT) Guidelines for Financial Institutions.* Regulatory Guideline, Abuja, Nigeria: CBN, 2022.

[9] National Information Technology Development Agency. *Nigeria Data Protection Regulation (NDPR) 2019.* Regulatory Framework, Abuja, Nigeria: NITDA, 2019.

[10] E. A. Lopez-Rojas, S. Ellegaard, S. Axelsson. *PaySim: A Financial Mobile Money Simulator for Fraud Detection. Proceedings of the European Modeling and Simulation Symposium (EMSS)*, 2016, 249–255.

[11] D. Azamuke, D. Ssebuggwawo, S. Hoppenbrouwers. *Analyzing Mobile Money Transaction Data for Fraud Detection Using Machine Learning: A Case Study Approach.* In Proceedings of the 2022 International Conference on Information and Communication Technologies for Development (ICT4D); ACM: New York, NY, USA, 2022; pp. 1–10. DOI: 10.1145/3531056.3542774.

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

[13] S. Xuan, G. Liu, Z. Li, L. Zheng, S. Wang, C. Jiang. *Random Forest for Credit Card Fraud Detection.* 2018 IEEE International Conference on Networking, Sensing and Control (ICNSC), 1–6.

[14] E. G. Dada, C. G. Nwaorgu, D. O. Oyewola, A. O. Olajuyin*A Systematic Literature Review of Machine Learning Techniques in Financial Fraud Prevention and Detection.* International Journal of Statistical Sciences, *23*(3)(2023), 315–344. DOI: 10.1504/IJSSS.2023.10057287.

Gbolahan Adeniran Idowu*
Department of Mathematics, Lagos State university, Ojo, Lagos State, Nigeria.
Department of Statistics, University of Lagos, Akoka, Lagos State, Nigeria.
E-mail address: gbolahan.idowu@lasu.edu.ng

Josiah Endurance Owolabi
Department of Statistics, University of Lagos, Akoka, Lagos State, Nigeria.
E-mail address: 210806502@live.unilag.edu.ng