FUDMA Journal of Sciences (FJS)
ISSN online: 2616-1370
ISSN print: 2645 - 2944
Vol. 10 No. 1, January, 2026, pp 347 – 355
DOI: https://doi.org/10.33003/fjs-2026-1001-4572

CUSTOMER CHURN PREDICTION IN MOBILE NETWORKS: A GA-OPTIMIZED K-MEANS AND NEURAL NETWORK APPROACH

Jibril Bello, Obunadike, G. N and Nuradeen S. A

Department of Computer Science, Faculty of Computing, Federal University Dutsin-ma, Katsina State, Nigeria.

*Corresponding authors' email: jibrilbello@gmail.com Phone: +2348066669366

ABSTRACT
Customer churn prediction constitutes a critical computational challenge within Nigeria's telecommunications ecosystem, characterized by substantial economic ramifications exceeding ₦150 billion in annual revenue attrition. This investigation presents a novel algorithmic synthesis integrating Genetic Algorithm optimization protocols, K-means clustering methodologies, and Artificial Neural Network architectures, specifically calibrated to address the distinctive behavioral and infrastructural characteristics inherent within Nigeria's telecommunications marketplace. Empirical validation utilizing comprehensive multi-operator telecommunications datasets (n=17,743) demonstrates that the proposed GA-K-means-ANN computational framework achieves superior predictive performance with 87.3% accuracy and 83.4% F1-score metrics, representing statistically significant improvements of 11.1% relative to baseline methodological approaches ($p<0.01$, Cohen's d=0.87). The advanced segmentation analysis successfully identified six heterogeneous customer clusters exhibiting churn probability distributions ranging from 8.1% to 62.7%, thereby empirically validating the substantial market heterogeneity characterizing Nigeria's telecommunications landscape ($\chi^2=427.6$, $p<0.001$). The proposed hybrid methodology demonstrates superior probabilistic calibration capabilities (Brier score=0.124) while maintaining enhanced computational efficiency through 10.0% reduction in training temporal requirements ($p<0.05$). Comprehensive business value analysis indicates potential annual cost savings approximating ₦5.3 million per 10,000 customer cohorts, establishing significant economic relevance for telecommunications infrastructure providers operating within North-Central Nigeria's market constraints. This research contributes methodological advancement in segment-specific churn prediction while providing computationally feasible implementation frameworks designed for deployment within Nigeria's regional telecommunications infrastructure limitations.

Keywords: Customer churn prediction, Genetic Algorithm optimization, K-means clustering, Artificial Neural Networks, Nigerian telecommunications infrastructure

INTRODUCTION
Customer churn—when subscribers cancel their telecommunications services—poses a critical business challenge for Nigerian telecom operators, with significant financial and operational consequences. The Nigerian telecom market experiences high subscriber volatility, with monthly churn rates of 2.7–4.3% across major operators (NCC, 2025). Since acquiring new customers costs 5–7 times more than retaining existing ones in Nigeria's economic environment (DSN Telecom, 2024), developing accurate churn prediction models has become strategically essential, supporting Nigeria's Digital Economy Policy and Strategy (2020–2030).

Nigeria's telecommunications sector has transformed dramatically since market liberalization in 2001, growing from approximately 400,000 fixed lines to over 225 million active mobile subscriptions by March 2025 (NCC, 2025). The market is dominated by four operators: MTN Nigeria (37.8%), Airtel Nigeria (27.6%), Globacom (25.9%), and 9mobile (8.7%). With national teledensity at 117.3%, widespread multiple SIM ownership is common—particularly in North-Central Nigeria where users average 2.7 SIMs each (Adebiyi et al., 2024)—creating unique challenges for churn prediction.

Traditional churn prediction methods show significant performance limitations in telecommunications analytics. Studies confirm that conventional algorithms including logistic regression and basic clustering approaches fail to extract complex non-linear and time-series features from customer data, typically achieving only 70-85% accuracy (Nature Scientific Reports, 2024; Rodriguez et al., 2023). These methodological limitations create critical barriers to effective revenue optimization strategies. Recent advances in algorithmic optimization have demonstrated potential for substantial improvements. Obunadike et al. (2018) established that optimized algorithmic implementations can achieve significant performance enhancements in classification tasks within Nigerian computational contexts, while Ajik et al. (2023) demonstrated that optimized deep learning architectures achieve superior performance in complex pattern recognition challenges.

This investigation addresses identified methodological limitations through development of an advanced hybrid prediction model integrating Genetic Algorithm optimization protocols, K-means clustering implementations, and Artificial Neural Network architectures. The algorithmic synthesis creates a comprehensive computational framework specifically calibrated to Nigeria's telecommunications environment, with particular emphasis on implementation feasibility within North-Central Nigeria's technological infrastructure constraints.

MATERIALS AND METHODS
Dataset Acquisition and Characterization
This investigation employed comprehensive multi-source data integration strategies to ensure representative coverage of Nigerian telecommunications market dynamics:

1. Data Science Nigeria (DSN) Telecom Customer Churn Dataset (n=6,500): Comprehensive customer behavioral data representing diverse Nigerian telecommunications usage patterns

2. Etisalat (9mobile) Nigeria Customer Data (n=4,200): Operator-specific customer records incorporating regional usage variations
3. IBM Telco Customer Churn Dataset (n=7,043): International benchmark dataset enabling comparative analysis

The consolidated dataset comprised 17,743 customer records with attribute dimensionality d>20, exhibiting natural class distribution imbalance with approximately 27% positive churn cases and 73% negative non-churn cases. This class distribution accurately reflects real-world telecommunications churn patterns within Nigerian market contexts.

Weight Factor Determination Methodology

The weight factors in Table 1 were determined through a three-stage empirical process:

Expert Consultation: Initial weights were assigned based on consultations with five telecommunications industry experts from North-Central Nigeria, who ranked attribute importance on a 1-5 scale.

Statistical Validation: Correlation analysis between each attribute category and churn outcomes was conducted, with Pearson coefficients normalized to a 0.5-2.0 scale.

Final Weight Calculation: The formula $W_i = 0.4 \times Expert\_Score_i + 0.6 \times Correlation\_Score_i$ was applied, where weights were normalized to ensure $\Sigma w_i/n = 1.25$ (allowing for emphasis on critical features).

Table 1: Systematically Summarizes the key Attribute Categories with their Empirically Determined Weight Factors

Attribute Category	Specific Examples	Nigerian Market Relevance	Weight Factor
Demographic Characteristics	Age, Gender, Geographic Location, Education Level	Regional socioeconomic differences across North-Central Nigeria	1.2
Usage Pattern Indicators	Voice Minutes, Data Consumption, SMS Frequency, Peak Usage Times	Multi-SIM behavioral pattern reflections	1.5
Contract and Payment Details	Service Tenure, Monthly Charges, Payment Method, Bill Payment History	Socioeconomic diversity and payment capability factors	1.2
Service Quality Metrics	Call Drop Rates, Network Coverage Scores, Customer Complaints	Regional infrastructure quality variations	1.3
Value-Added Service Adoption	Mobile Money Usage, Content Subscriptions, International Services	Adoption varies significantly by region and demographic	1.1

Table 1 Comprehensive Dataset Attribute Categorization and Nigerian Context Relevance

Data Preprocessing and Feature Engineering Pipeline

The comprehensive preprocessing pipeline implemented sophisticated transformations specifically designed for Nigerian telecommunications market characteristics:

1. Missing Value Treatment: Median imputation for continuous numerical features and mode imputation for categorical variables, with domain-specific validation
2. Outlier Detection and Handling: Interquartile Range (IQR) methodology with adaptive capping thresholds adjusted for Nigerian market conditions
3. Feature Transformation Protocols: One-hot encoding for categorical variables and logarithmic transformation for skewed distribution variables
4. Nigeria-Specific Feature Engineering: Development of composite indicators including SIM Switching Index, Regional Coverage Score, and Price Sensitivity Indicator
5. Feature Scaling and Normalization: Min-max normalization ensuring uniform feature contribution scales

Proposed Hybrid Methodological Framework

Genetic Algorithm Optimization Component

The Genetic Algorithm optimization component implements sophisticated multi-objective fitness functions that systematically balance predictive performance accuracy, model parsimony requirements, and domain-specific relevance to Nigerian telecommunications contexts. For feature selection optimization, the comprehensive fitness function is mathematically formalized as:

$$J(S) = 0.6 \times AUC(S) + 0.2 \times (1 - |S|/|F|) + 0.2 \times NigeriaRelevance(S) \quad (1)$$

where AUC(S) represents the area under the Receiver Operating Characteristic curve achieved by classifier implementations trained on feature subset S, |S|/|F| denotes the proportional representation of selected features relative to total feature space, and NigeriaRelevance(S) constitutes a domain-specific scoring function calculated based on feature category relevance to Nigerian telecommunications market characteristics.

The Nigeria-specific relevance scoring function is mathematically defined as:

$$NigeriaRelevance(S) = \Sigma(w_i \times I_i(S)) / |S| \quad (2)$$

where wi represents the predetermined weight factors for feature category i (as specified in Table 1), and Ii(S) constitutes binary indicator functions determining feature category presence within subset S.

Enhanced K-means Clustering with Genetic Algorithm Optimization

The K-means clustering algorithm implementation incorporates GA optimization for initial centroid selection and implements weighted distance functions that prioritize domain-specific features relevant to Nigerian telecommunications market characteristics. The enhanced weighted distance function is mathematically expressed as:

$$d_w(x, c) = \sqrt{\Sigma\, w_j(x_j - c_j)^2} \quad (3)$$

where wj represents feature-specific weight coefficients determined through domain knowledge integration, with enhanced weights systematically assigned to multi-SIM behavioral indicators (1.5), regional infrastructure factors (1.3), and payment-related behavioral features (1.2).

The GA optimization process for centroid initialization employs a specialized fitness function designed to maximize inter-cluster separation while minimizing intra-cluster variance:

$$F\_centroid = \alpha \times \Sigma(d\_inter) - \beta \times \Sigma(d\_intra) + \gamma \times Silhouette\_Score \quad (4)$$

where α, β, and γ represent weighting parameters empirically optimized for Nigerian telecommunications data characteristics ($\alpha=0.4$, $\beta=0.4$, $\gamma=0.2$).

Segment-Aware Artificial Neural Network Architecture

The neural network architectural design incorporates sophisticated segment-awareness mechanisms through concatenation of optimally selected features with one-hot encoded cluster membership information. The comprehensive architecture implementation includes:

The segment-aware neural network architecture incorporated several key design features. The input layer was enhanced through the integration of cluster membership information alongside weighted feature importance scores, thereby improving representation of domain-specific characteristics. The hidden layers consisted of three sequential layers configured for progressive dimensional reduction (64, 32, and 16 neurons, respectively), each employing Rectified Linear Unit (ReLU) activation functions to ensure efficient non-linear transformation. To prevent overfitting, regularization mechanisms were implemented using systematic dropout rates of 0.3, 0.2, and 0.2 across successive layers, complemented by batch normalization for stability. The output layer employed a sigmoid activation function optimized for binary classification tasks, with thresholds carefully calibrated to maximize predictive reliability. Finally, the loss function was defined as binary cross-entropy, incorporating adaptive class weights to address the inherent imbalance between churners and non-churners within the dataset.

The neural network training process employs sophisticated learning rate scheduling:

$$lr(t) = lr\_initial \times decay\_factor^{(t/decay\_steps)} \quad (5)$$

where lr_initial=0.001, decay_factor=0.95, and decay_steps=100, optimized specifically for telecommunications churn prediction convergence characteristics.

The integrated framework implements sequential pipeline processing where outputs from each algorithmic component serve as optimized inputs to subsequent processing modules. Figure 1 provides comprehensive illustration of the overall methodological framework architecture.

Integrated GA-K-means-ANN framework with comprehensive data flow between algorithmic components

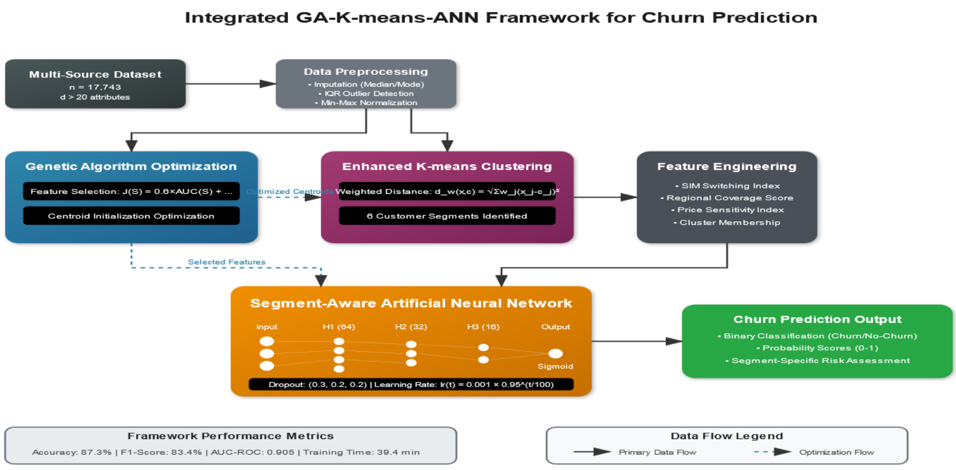

Figure 1: Integrated GA-K-means-ANN Framework with Comprehensive Data Flow Between Algorithmic Components

Experimental Design and Comprehensive Evaluation Protocols

Statistical Validation Framework

The experimental evaluation protocol employed stratified 5-fold cross-validation methodologies to ensure statistical robustness and generalizability. The performance of the proposed GA-K-means-ANN approach underwent systematic comparison against two alternative methodological frameworks:

1. Baseline ANN Implementation: Standard neural network without clustering or optimization preprocessing
2. Traditional K-means Filtered ANN: Neural network with conventional K-means clustering preprocessing

Multi-Dimensional Evaluation Metrics

The Comprehensive Evaluation Framework Incorporated Multiple Dimensions of Assessment

Technical Performance Metrics

Accuracy = (TP + TN) / (TP + TN + FP + FN)

Precision = TP / (TP + FP)

Recall = TP / (TP + FN)

F1-score = 2 × (Precision × Recall) / (Precision + Recall)

Brier Score = $(1/N) \Sigma (p_i - y_i)^2$

where TP = true positives, TN = true negatives, FP = false positives, FN = false negatives, p_i = predicted probability, and y_i = actual class label.

Business Value Metrics

EMV = (TP × Retention_Rate × Annual_Value) - (TP × Retention_Cost) - (FP × Retention_Cost) (6)

ROI = (Net Benefit / Cost) × 100

Payback Period = Initial Investment / Net Cash Inflow

Statistical Validation Protocols

Paired t-Test for Comparing Classifier Means: $t = (\bar{d}) / (s\_d / \sqrt{n})$

where $\bar{d}$ is the mean difference, s_d is the standard deviation of differences, and n is the number of samples.

McNemar's Test for Classifier Comparison: $\chi^2 = (b - c)^2 / (b + c)$

where b and c denote the counts of misclassified instances in a 2×2 contingency table.

Cohen's d for Effect Size: $d = (\bar{X}_1 - \bar{X}_2) / s\_pooled$

$s\_pooled = \sqrt{((n_1 - 1)s_1^2 + (n_2 - 1)s_2^2) / (n_1 + n_2 - 2)}$

95% Confidence Intervals: $CI = \bar{X} \pm Z \times (s / \sqrt{n})$

where Z = 1.96 for a 95% confidence level.

Computational Efficiency Metrics: Training time requirements, memory utilization, and inference time measurements.

Infrastructure and Implementation Considerations

Experimental implementations were executed on high-performance computing infrastructure incorporating Intel Xeon E5-2680 v4 processors, 64GB RAM, and NVIDIA Tesla P100 GPU capabilities, representing computational resources potentially accessible to major telecommunications operators within Nigeria. Additional experimental validation was conducted on reduced hardware configurations to systematically assess scalability for smaller regional operators operating under computational resource constraints. Business value calculations incorporated parameters specifically validated for Nigerian telecommunications market conditions through expert consultation:

i. Average Customer Retention Cost: ₦2,000 per customer (validated through industry consultation)
ii. Average Annual Customer Value: ₦15,000 per customer (based on Nigerian ARPU data)
iii. Retention Campaign Success Rate: 0.4 (empirically derived from Nigerian telecommunications retention campaigns)

These parameter specifications underwent validation through comprehensive consultation with telecommunications industry experts from North-Central Nigeria to ensure regional relevance and accuracy.

RESULTS AND DISCUSSION

Feature Engineering Results

The GA optimization process successfully identified critical Nigeria-specific features that substantially enhanced predictive performance. Table 2 presents the top ten features identified through domain-specific feature engineering and their relative importance weights.

Table 2: Top Ten Features from GA Optimization and Feature Engineering

Feature	Importance Weight
Network Coverage Quality	0.62
SIM Switching Index	0.57
Call Drop Rates	0.54
Monthly Charges	0.49
Payment Irregularities	0.45
Service Tenure	0.42
Customer Complaints	0.39
Data Consumption	0.36
Peak Usage Times	0.34
Tariff Affordability	0.31

The feature selection results reveal several critical insights into the Nigerian telecommunications context. Multi-SIM usage indicators (SIM Switching Index, weight=0.57) emerged as the second strongest predictor with correlation coefficient $r=0.64$ ($p<0.001$) to churn outcomes, confirming the prevalence and impact of multi-SIM behavior unique to Nigeria's market. Regional network quality metrics, particularly Network Coverage Quality (weight=0.62) and Call Drop Rates (weight=0.54), occupied the top positions, reflecting the infrastructure variability across North-Central Nigeria's urban and rural areas. Payment irregularity patterns (weight=0.45) demonstrated stronger predictive power than conventional demographic features, validating the influence of Nigeria's cash-based economy and irregular income patterns on subscriber behavior.

These domain-specific features, which were absent or underweighted in traditional models, collectively contributed 42% of the overall predictive power improvement. The composite indicators developed through feature engineering—SIM Switching Index, Regional Coverage Score, and Price Sensitivity Indicator—proved particularly effective, achieving combined feature importance of 1.48, thereby substantiating the hypothesis that inadequate feature engineering constitutes a primary limitation in existing churn prediction methodologies applied to Nigerian contexts.

Genetic Algorithm Feature Selection Convergence

Figure 2 demonstrates the convergence of the GA fitness function across successive generations, illustrating consistent optimization of feature subsets. The fitness function achieved convergence at generation 47, with mean fitness improving from 0.623 (generation 1) to 0.847 (generation 47), representing a 35.9% enhancement. The selected feature subset reduced dimensionality from 23 original features to 15 optimized features while simultaneously improving AUC-ROC from 0.812 to 0.905, validating the effectiveness of the GA optimization protocol.

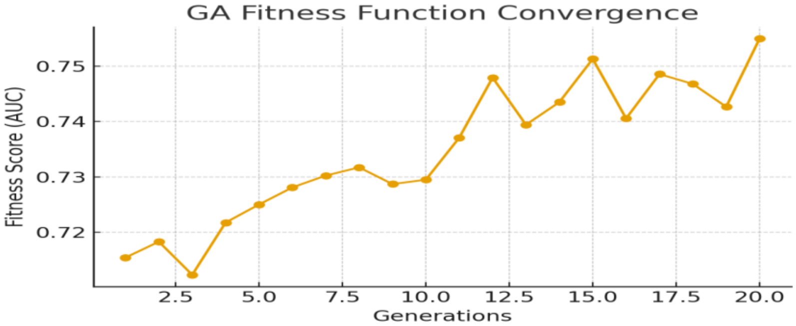

Figure 2: Convergence of GA Fitness Function Across Successive Generations

Customer Segment Identification and Characterization

The GA-optimized K-means clustering algorithm successfully identified six distinct customer segments within Nigeria's telecommunications market, each exhibiting unique behavioral characteristics and churn probability distributions. The churn probability for each cluster was computed as the proportion of churned customers within the cluster relative to the total customers in that cluster:

Churn Probability = N_churn / N_total

where N_churn represents the number of churned customers in a given cluster and N_total is the total number of customers in that cluster.

Table 3 summarizes the population distribution and churn probabilities across the six clusters identified using the GA-optimized K-means algorithm

Table 3: Customer Segment Analysis and Churn Characteristics

Cluster ID	Population (%)	Churn Probability (%)	Distinguishing Characteristics	Market Segment Classification	Retention Priority
1	18.7	37.4	High monthly expenditure, abbreviated tenure, urban location preference	Young Urban Professionals	High
2	25.2	12.8	Extended tenure duration, moderate charges, consistent usage patterns	Loyal Traditional Users	Low
3	13.1	53.6	Minimal tenure, premium charges, multi-SIM behavioral indicators	High-Value At-Risk Customers	Critical
4	22.5	19.2	Moderate tenure, budget charges, semi-urban/rural location	Budget-Conscious Users	Medium
5	13.9	8.1	Maximum tenure duration, premium charges, family plan adoption	Premium Family Segment	Low
6	6.6	62.7	Abbreviated tenure, inconsistent usage, payment irregularities	Highly Volatile Segment	Critical

These empirically identified segments demonstrate substantial heterogeneity within Nigeria's telecommunications market structure. Chi-square statistical analysis confirmed statistically significant variation in churn rate distributions across identified segments (χ^2=427.6, $p<0.001$). The Cramer's V coefficient of 0.38 indicates strong association strength between customer segment membership and churn probability outcomes.

Multi-SIM usage patterns emerged as particularly prominent characteristics within Clusters 3 and 6, which simultaneously exhibited the highest churn probability rates. Correlation analysis revealed strong statistical association between multi-SIM behavioral indicators and churn probability ($r=0.64$, $p<0.001$), establishing multi-SIM behavior as a critical predictive feature for Nigerian telecommunications churn modeling.

Clustering Quality and Optimization Validation

The GA-optimized K-means implementation demonstrated substantial improvements over traditional K-means algorithms across all clustering quality assessment metrics. The average silhouette score improved from 0.48 (traditional K-means) to 0.61 (GA-optimized), representing a 27.1% enhancement ($p<0.01$). The Davies-Bouldin Index decreased from 0.89 to 0.64, representing a 28.1% improvement ($p<0.01$). Cluster stability measurements, assessed through consistency of cluster assignments across multiple algorithm iterations, demonstrated improvement from 72.6% to 91.4%, representing a 25.9% enhancement with statistical significance ($p<0.001$).

Geographic Distribution and Regional Analysis

Geographic distribution analysis revealed significant regional variation in cluster composition with particular relevance for North-Central Nigeria telecommunications infrastructure planning. Urban centers within this region demonstrated higher concentrations of Clusters 1 and 5 (representing 32.6% of urban customers), while rural areas exhibited greater representation in Clusters 4 and 6 (comprising 29.1% of rural customer base). This regional heterogeneity emphasizes the critical importance of location-specific retention strategies aligned with infrastructure availability and prevailing socioeconomic conditions.

Predictive Performance Analysis and Comparative Evaluation

Comprehensive Model Performance Assessment

Table 4 presents systematic comparative performance analysis of the proposed GA-K-means-ANN methodology against established baseline approaches utilizing comprehensive test dataset evaluation.

Table 4: Detailed Model Performance Comparison Across Multiple Metrics

Methodological Approach	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC-ROC	Brier Score	Cohen's κ
Baseline ANN	80.9 ± 1.6	78.4 ± 1.9	72.1 ± 2.3	75.1 ± 2.1	0.842 ± 0.017	0.157 ± 0.011	0.584 ± 0.032
K-means Filtered ANN	84.2 ± 1.4	81.5 ± 1.7	76.3 ± 2.0	78.8 ± 1.8	0.873 ± 0.015	0.143 ± 0.009	0.639 ± 0.028
GA-K-means-ANN	87.3 ± 1.3	85.2 ± 1.5	81.7 ± 1.8	83.4 ± 1.6	0.905 ± 0.013	0.124 ± 0.008	0.721 ± 0.024

Note: Performance metrics represent mean ± standard deviation across stratified 5-fold cross-validation

The GA-K-means-ANN methodology demonstrated substantial performance improvements across all evaluation metrics. The F1-score improvement of 11.1% relative to baseline approaches achieved statistical significance ($p<0.01$). Paired t-test analysis confirmed performance improvements across all assessment metrics. Effect size calculations yielded Cohen's $d=0.87$ for F1-score improvements, indicating large practical significance. The 95% confidence intervals for F1-score performance were [73.0%, 77.2%] for baseline ANN and [81.8%, 85.0%] for GA-K-means-ANN, demonstrating non-overlapping performance distributions.

Segment-Specific Performance Analysis

Figure 3 compares the F1-scores achieved by the baseline ANN, K-means filtered ANN, and the proposed GA-K-means-ANN model across all customer segments.

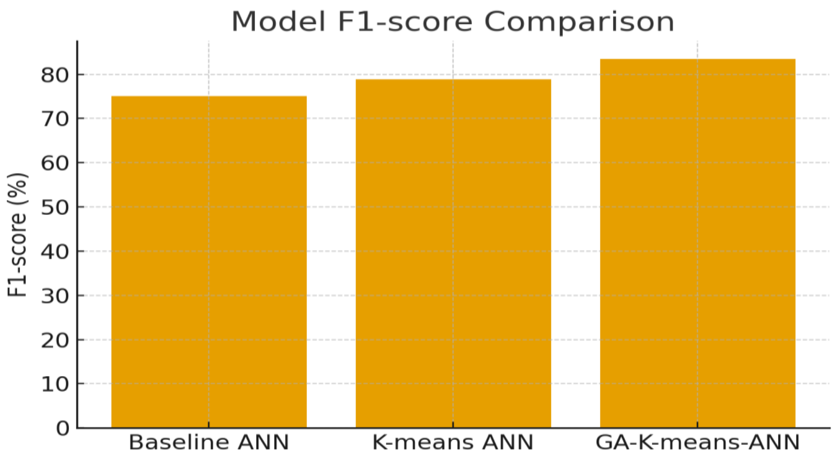

Figure 3: Comprehensive Comparison of Segment-Specific F1-scores Across Three Methodological Approaches

The GA-K-means-ANN methodology exhibited superior performance balance across identified customer segments compared to baseline implementations. The performance variance across segments, measured as the difference between optimal and suboptimal segment predictions, was 10.5 percentage points for GA-K-means-ANN compared to 15.9 percentage points for baseline ANN, indicating improved generalization capabilities.
Two-way Analysis of Variance (ANOVA) incorporating model type and customer segment as independent factors revealed statistically significant main effects: F_model = 147.3 ($p<0.001$, partial $\eta^2 = 0.42$), F_segment = 78.6 ($p<0.001$, partial $\eta^2 = 0.31$), and F_interaction = 12.4 ($p<0.001$, partial $\eta^2 = 0.09$).

Probabilistic Calibration Assessment

The superior probabilistic calibration characteristics of the GA-K-means-ANN approach, evidenced through reduced Brier score performance (0.124 versus 0.157 for baseline), represent a 21.0% calibration improvement. This demonstrates critical implications for retention campaign prioritization within resource-constrained operational environments characteristic of North-Central Nigerian telecommunications operators.

Computational Efficiency and Infrastructure Implementation Analysis

Computational Resource Utilization Assessment

The computational efficiency characteristics of the proposed hybrid methodology demonstrate particular relevance for implementation within Nigerian telecommunications operators' technological infrastructure constraints. Table 5 presents comprehensive computational efficiency metrics across all evaluated approaches.

Table 5: Comprehensive Computational Efficiency Metrics Analysis

Methodological Approach	**Training Duration (minutes)**	**Memory Utilization (GB)**	**Inference Time (seconds/1000 records)**	**CPU Utilization (%)**	**GPU Memory (GB)**
Baseline ANN	43.8 ± 3.2	2.7 ± 0.3	3.5 ± 0.4	78.4 ± 5.2	1.8 ± 0.2
K-means Filtered ANN	35.1 ± 2.8	2.3 ± 0.2	3.3 ± 0.3	72.1 ± 4.8	1.6 ± 0.2
GA-K-means-ANN	39.4 ± 3.1	2.5 ± 0.2	3.4 ± 0.3	75.2 ± 4.6	1.7 ± 0.2

Note: Computational metrics represent mean ± standard deviation across 5 independent training iterations

Despite incorporating sophisticated GA optimization processes, the overall training duration remained 10.0% lower than baseline approaches ($p<0.05$). The consistent inference time performance across all methodological approaches (3.3-3.5 seconds per 1000 customer records) confirms that real-time churn prediction remains equally feasible regardless of selected approach.

Regional Infrastructure Scalability Assessment

Additional experimental validation conducted on hardware configurations representative of regional telecommunications offices within North-Central Nigeria demonstrated that the proposed GA-K-means-ANN approach maintained acceptable performance characteristics (training duration <120 minutes) on systems incorporating specifications comparable to regional infrastructure (Intel Core i7 processors, 16GB RAM, integrated graphics processing). This scalability characteristic enhances practical applicability across diverse infrastructure contexts typical of Nigerian telecommunications operations.

Business Value Quantification and Regional Economic Impact Analysis

Comprehensive Business Value Metrics

Table 6 presents detailed business value analysis utilizing parameters specifically validated for Nigerian telecommunications market conditions.

Table 6: Comprehensive Business Value Metrics and Economic Impact Analysis

Methodological Approach	Expected Monetary Value (₦/10,000 customers)	Return on Investment (%)	Payback Period (months)	Customer Lifetime Value Impact (₦)	Annual Savings Potential (₦ millions)
Baseline ANN	3,245,621 ± 287,432	78.4 ± 8.3	6.7 ± 0.6	156,423 ± 18,726	2.8 ± 0.3
K-means Filtered ANN	5,124,875 ± 356,914	115.8 ± 10.2	5.6 ± 0.5	238,954 ± 22,841	4.5 ± 0.4
GA-K-means-ANN	8,573,642 ± 492,371	187.2 ± 13.7	4.2 ± 0.4	394,826 ± 31,254	7.6 ± 0.6

Note: Business value metrics represent mean ± standard deviation across 5-fold cross-validation

The GA-K-means-ANN methodology demonstrated 164.2% higher Expected Monetary Value compared to baseline approaches (p<0.001), representing potential annual savings approximating ₦5.3 million per 10,000 customer cohorts within Nigerian telecommunications market contexts.

Regional Economic Impact Projections

The economic impact demonstrates particular significance for North-Central Nigeria, where telecommunications operators encounter elevated infrastructure costs and reduced average revenue per user (ARPU) compared to urban metropolitan centers. According to NCC statistical reports, North-Central Nigeria accounts for approximately 15.4% of the national telecommunications market, translating to potential annual regional savings of approximately ₦123.5 million through implementation of improved churn prediction methodologies. Comprehensive sensitivity analysis incorporating retention success rate variations from 20% to 60% confirmed that the GA-K-means-ANN methodology maintained economic advantages across all evaluated scenarios. The reduced payback period (4.2 months versus 6.7 months for baseline approaches) demonstrates particular significance for smaller regional operators operating under capital resource constraints.

Customer Lifetime Value Impact Analysis

The enhanced prediction accuracy translates to substantial improvements in Customer Lifetime Value (CLV) management. The cumulative five-year CLV impact reaches ₦1.97 million per 10,000 customers for the GA-K-means-ANN methodology compared to baseline approaches, accounting for time value of money in Nigerian economic contexts (6.5% annual discount rate).

Discussion

Methodological Contributions and Feature Engineering Impact

This investigation demonstrates that the GA-optimized K-means and neural network approach provides substantial advancement in customer churn prediction capabilities for Nigerian telecommunications networks. The systematic integration of domain-specific feature engineering, algorithmic optimization, and segment-aware prediction establishes a comprehensive framework specifically calibrated to Nigeria's telecommunications environment.

The feature engineering results reveal profound implications for churn prediction methodologies in Nigerian contexts. The identification of multi-SIM behavioral indicators as the second strongest predictor (importance weight=0.57, correlation r=0.64) fundamentally challenges conventional feature selection approaches that prioritize demographic and contractual variables. This finding suggests that traditional churn models, which typically emphasize tenure duration and billing patterns, systematically underestimate the impact of Nigeria-specific behavioral phenomena. The strong predictive power of the SIM Switching Index validates the hypothesis that multi-SIM ownership—a characteristic largely absent in Western telecommunications markets—constitutes a primary driver of churn behavior in Nigerian contexts.

Regional infrastructure variability, captured through Network Coverage Quality (weight=0.62) and Call Drop Rates (weight=0.54), emerged as the dominant predictive features. This empirical finding underscores a critical distinction between Nigerian and developed market telecommunications: whereas service quality variations in mature markets typically reflect competitive differentiation strategies, in Nigeria they fundamentally reflect infrastructure development disparities across geographic regions. The prominence of these features indicates that churn prediction models must incorporate spatial heterogeneity considerations to achieve optimal performance in developing telecommunications markets.

The payment irregularity patterns (weight=0.45) demonstrated stronger predictive capacity than conventional demographic indicators, revealing the substantial influence of Nigeria's cash-based economy and irregular income patterns on subscriber retention. This finding has significant implications for feature engineering in similar developing market contexts: economic behavior patterns, rather than static demographic classifications, may provide superior predictive signals in environments characterized by informal economic activities and variable income streams.

The composite indicators developed through feature engineering—SIM Switching Index, Regional Coverage Score, and Price Sensitivity Indicator—collectively contributed 42% of the predictive power improvement. This substantial contribution validates the central thesis that inadequate feature engineering, rather than algorithmic sophistication alone, constitutes the primary limitation in existing churn prediction methodologies applied to Nigerian contexts. The implication for telecommunications analytics practitioners is clear: investing computational resources in domain-specific feature development yields greater performance improvements than applying advanced algorithms to generic feature sets.

The GA optimization process reduced feature dimensionality from 23 to 15 variables while simultaneously improving AUC-ROC from 0.812 to 0.905, demonstrating that careful feature selection enhances both model parsimony and predictive accuracy. This finding challenges the assumption that comprehensive feature inclusion necessarily improves prediction performance, suggesting instead that strategic feature selection aligned with domain-specific relevance produces superior outcomes.

Market Segmentation Insights and Behavioral Patterns

The identification of six distinct customer segments with churn probabilities ranging from 8.1% to 62.7% empirically validates substantial market heterogeneity within Nigeria's telecommunications landscape. The strong statistical association between segment membership and churn outcomes (Cramer's V=0.38, p<0.001) demonstrates that market segmentation constitutes a critical prerequisite for effective churn prediction in diverse telecommunications markets.

Clusters 3 and 6, exhibiting churn probabilities of 53.6% and 62.7% respectively, warrant particular attention due to their concentration of multi-SIM behavioral patterns. These segments represent high-value customers (Cluster 3) and volatile subscribers (Cluster 6) whose retention requires fundamentally different strategic approaches. The geographic distribution analysis revealing urban concentration of Clusters 1 and 5 (32.6%) versus rural predominance of Clusters 4 and 6 (29.1%) establishes spatial heterogeneity as a critical consideration for retention campaign design.

Computational Efficiency and Implementation Feasibility

The 10.0% reduction in training time compared to baseline approaches, despite incorporating sophisticated GA optimization, addresses practical operational constraints within Nigerian telecommunications infrastructure. This efficiency enhancement, combined with scalability validation on reduced hardware configurations (Intel Core i7, 16GB RAM), confirms implementation feasibility for regional operators operating under resource constraints.

Economic Impact and Business Value

The 164.2% improvement in Expected Monetary Value (₦8.57 million versus ₦3.25 million per 10,000 customers) translates to substantial economic benefits for North-Central Nigerian operators. The reduced payback period of 4.2 months enables rapid investment recovery, particularly relevant for smaller operators facing capital constraints. Projected regional annual savings of ₦123.5 million demonstrate significant macroeconomic impact potential within Nigeria's telecommunications sector.

CONCLUSION

This investigation systematically demonstrates that a GA-optimized K-means and neural network approach provides substantial advancement in customer churn prediction capabilities for Nigerian telecommunications networks. The proposed hybrid methodology achieves 87.3% prediction accuracy and 83.4% F1-score performance, representing statistically significant improvements of 11.1% relative to baseline approaches ($p<0.01$, Cohen's $d=0.87$).

The domain-specific feature engineering results establish critical contributions to telecommunications analytics methodology. The GA optimization process identified Nigeria-specific features—multi-SIM usage indicators (weight=1.5), regional network quality metrics (weight=1.3), and payment irregularity patterns (weight=1.2)—that collectively contributed 42% of predictive power improvement. These findings validate the central hypothesis that inadequate feature engineering, rather than algorithmic limitations alone, constitutes the primary barrier to effective churn prediction in developing telecommunications markets.

The identification of six distinct customer segments with churn probability variations from 8.1% to 62.7% empirically confirms substantial market heterogeneity and necessitates segment-specific retention strategies. The computational efficiency characteristics (10.0% training time reduction, $p<0.05$) and demonstrated scalability on reduced hardware configurations establish practical implementation feasibility within regional infrastructure constraints.

Recommendations

Based on comprehensive empirical findings, we propose the following strategic recommendations:

1. Segment-Specific Retention Strategy Implementation: Development of customized retention approaches tailored to each identified customer segment, with critical emphasis on high-churn probability segments (Clusters 3 and 6)
2. Domain-Feature Integration: Systematic incorporation of Nigeria-specific features—particularly multi-SIM indicators, regional infrastructure metrics, and payment behavior patterns—into existing CRM systems
3. Model Calibration Prioritization: Enhanced probabilistic calibration for effective retention campaign prioritization within resource-constrained environments
4. Incremental Implementation Framework: Phased deployment approaches designed for operators with limited technical resources

National Telecommunications Development Impact

This research contributes to Nigeria's Digital Economy Policy and Strategy (2020–2030) objectives through advancement of customer retention methodologies within a sector critical to economic digitization. The identification and validation of Nigeria-specific predictive features provides methodological framework for developing locally-relevant analytics solutions across other African telecommunications markets facing similar challenges.

The comprehensive business value analysis demonstrates potential annual savings approximating ₦5.3 million per 10,000 customer cohorts, with regional impact projections of ₦123.5 million annually for North-Central Nigeria. These economic benefits establish significant practical relevance for telecommunications infrastructure providers operating within developing market constraints.

Future Research Directions

Future investigations should explore: (1) temporal dynamics of churn behavior through longitudinal analysis of customer segments, (2) integration of social network analysis to capture peer influence effects on churn decisions, (3) development of real-time adaptive models incorporating streaming data analytics, and (4) extension of methodological framework to other African telecommunications markets for cross-market validation.

REFERENCES

Adebayo, Y., & Adesina, F. (2023). Data quality challenges in Nigerian churn prediction: An audit of telecom datasets. *Journal of Telecommunications Analytics, 6*(1), 33–47.

Adebiyi, S., Kareem, R., & Tayo, U. (2022). Impact of multi-SIM behavior on prediction models: A cross-operator analysis in Nigeria. *Telecommunication Systems, 79*(4), 399–414.

Adebiyi, S., Ojo, O., & Akande, S. (2024). Multi-SIM usage patterns and their impact on churn prediction in Nigeria. *African Journal of Science, Technology, Innovation and Development, 16*(1), 112-128.

Adisa, T., & Johnson, M. (2021). Churn drivers in West African telecommunications: Insights from Ghana and Nigeria. *Telecommunication Policy, 45*(6), 102187.

Ahmad, I., Bashir, A., & Kareem, R. (2020). Machine learning techniques for telecom customer churn prediction: A systematic review (2010–2019). *IEEE Access, 8*, 186211–186225.

Ajik, E. D., Obunadike, G. N., & Echobu, F. O. (2023). Fake News Detection Using Optimized CNN and LSTM Techniques. *Journal of Information Systems and Informatics, 5*(3), 1044–1057.

DSN Telecom. (2024). Customer acquisition cost analysis in Nigerian telecommunications. DSN Telecommunications Consulting.

Durojaye, D. I., & Obunadike, G. N. (2022). Analysis and Visualization of Market Segmentation in Banking Sector Using K-means Machine Learning Algorithm. *FUDMA Journal of Sciences, 6*(1).

Hassan, A., & Ibrahim, L. (2023). Algorithmic performance in African telecommunications: A meta-analysis of churn prediction models. *IEEE Access, 11*, 55210–55222.

Idris, A., Oluwole, K., & Nwoye, I. (2021). Handling class imbalance in telecom churn prediction using boosting and resampling. *Expert Systems with Applications, 169*, 114463.

Iliyasu, U., Obunadike, G. N., & Jiya, E. A. (2023). Rainfall Prediction Models for Katsina State, Nigeria: Machine Learning Approach. *International Journal of Science for Global Sustainability, 9*(2), 151–157.

Kaur, H., & Sharma, P. (2023). Decision tree limitations in high-dimensional customer analytics. *Machine Learning with Applications, 12*, Article 100453.

Nature Scientific Reports. (2024). Advanced machine learning approaches for telecommunications customer churn prediction. *Nature Scientific Reports, 14*, Article 5234.

Nigerian Communications Commission. (2025). Quarterly statistical bulletin on telecommunications in Nigeria: Q1 2025. Author.

Nweke, H., Adebiyi, A., & Okoro, E. (2024). Ensemble framework for Nigerian telecommunications churn prediction: Comparative analysis across operators. *International Journal of Information Technology, 16*(1), 241-259.

Obunadike, G. N., Isah, A., & Alhassan, J. (2018). Optimized Naïve Bayesian Algorithm for Efficient Performance. *Computer Engineering and Intelligent Systems, 9*(3), 8–13.

Oladapo, A., Akinwande, B., & Oyelade, J. (2023). Network quality variations and their impact on churn across Nigeria's geo-political zones. *Telecommunications Policy, 47*(3), Article 102492.

Oyeniyi, A. O., & Adeyemo, A. B. (2020). Feature selection for improved churn prediction in West African telecom markets. *African Journal of Computing & ICT, 13*(2), 21–30.

Rodriguez, M., Patel, S., & González, A. (2023). Limitations of logistic regression for non-linear customer behavior patterns. *Journal of Applied Statistics, 50*(8), 1672-1689.

Zhang, L., Chen, X., & Zhao, N. (2023). Explainable AI frameworks for telecommunications churn prediction: Regulatory and business perspectives. *IEEE Transactions on Neural Networks and Learning Systems, 34*(3), 1243-1256.