Zamfara International Journal of Sci. Tech. Educ. & Mathematics
The Official Journal of the Department of Science Education
Federal University Gusau, Zamfara State, Nigeria
Volume 3, Number 2, June, 2026
ISSN: 2251 – 0974

www.foefugusau.com.ng
des@fugusau.edu.ng
DOI: https://doi.org/10.64348/zije.2026438

Leveraging Deep Learning on Anomaly Detection in National Identity Number (NIN) Verification for National Identity Management Commission (NIMC), Nigeria

AKANBI, Bello Muhammed[1]
ORCID IDs: [1]https://orcid.org/0009-0009-7853-5367
aishaolamide13@gmail.com
ATUMOSHI, Adamu Yusuf[1]
[2]https://orcid.org/0009-0008-6142-4807
ISHOLA, Olabisi Bolarinwa[1]
[3]https://orcid.org/0009-0003-8647-8610
AKANBI, Aisha Olamide[2]
[4]https://orcid.org/0009-0000-5632-8871.

[1]Department of Computer Science, University of Abuja, FCT, Nigeria,
[2]Department of Computer Science, Air Force Institute of Technology, Kaduna, Nigeria

Abstract

National identity verification systems are critical public infrastructure, yet remain vulnerable to sophisticated fraud strategies including automated bot attacks, synthetic identity construction, geolocation spoofing, and API request abuse. In Nigeria, the National Identity Management Commission (NIMC) manages over 100 million National Identification Number (NIN) records through a high-volume, multi-channel API-driven verification architecture. Existing rule-based monitoring systems are inadequate for detecting complex behavioral anomalies at this scale. This study proposes, implements, and evaluates a Hybrid Autoencoder–Long Short-Term Memory (AE-LSTM) deep learning framework for unsupervised anomaly detection in NIN verification transactions. A synthetic dataset of 50,000 transaction records incorporating six documented fraud categories at a 10% anomaly rate was generated to simulate realistic NIMC verification behavior. A nineteen-feature input representation was derived through structured feature engineering. The hybrid model fuses deep autoencoder reconstruction error with temporal behavioral scoring over sliding windows. Experimental results demonstrate accuracy of 95.0%, ROC-AUC of 0.9810, F1-score of 0.762, and recall of 79.6%, significantly outperforming Isolation Forest and rule-based baselines. These results confirm the viability of deep learning-based behavioral anomaly detection for securing national digital identity infrastructure in developing economies.

Keywords: *Anomaly Detection; Deep Learning; Autoencoder; LSTM; NIN Verification; Digital Identity Security; NIMC; Fraud Detection; Nigeria; Cybersecurity*

Introduction

Digital identity systems represent foundational public infrastructure in twenty-first-century governance, enabling citizen access to financial services, telecommunications, healthcare, immigration processing, and social intervention programs. In Nigeria, the National Identity Management Commission (NIMC), established under the NIMC Act of 2007, administers the National Identification Number (NIN) system — the country's primary digital identity layer. As of May 2023, over 100 million distinct NINs had been issued, with enterprise verification conducted through API-driven platforms accessed by banks, telecom operators, and government agencies in real time (Iwegbu, 2023; NIMC, 2020). Despite this scale and institutional importance, the NIN verification ecosystem confronts escalating cybersecurity threats. Fraudulent verification patterns — including automated bot-driven enumeration, request flooding, synthetic identity construction, geolocation spoofing, and brute-force authentication attacks — increasingly exploit the API-driven verification

Zamfara International Journal of Sci. Tech. Educ. & Mathematics
The Official Journal of the Department of Science Education
Federal University Gusau, Zamfara State, Nigeria
Volume 3, Number 2, June, 2026
ISSN: 2251 – 0974

www.foefugusau.com.ng
des@fugusau.edu.ng
DOI: https://doi.org/10.64348/zije.2026438

architecture (Jacob et al., 2021; Ferrag et al., 2020). Traditional rule-based monitoring systems, which classify transactions against static predefined thresholds, are fundamentally inadequate for detecting complex behavioral anomalies that evolve dynamically (Adebayo & Ojo, 2022; Thakkar & Lohiya, 2021).

Anomaly detection — the computational identification of patterns that deviate significantly from established behavioral norms — represents a principled alternative. Deep learning architectures, particularly autoencoders trained exclusively on normal behavioral data, offer powerful unsupervised anomaly detection capabilities requiring no labeled fraud examples, enabling detection of both known and novel attack strategies (Ruff et al., 2021; Pang et al., 2021). Long Short-Term Memory (LSTM) networks extend this to temporal sequence modeling, capturing fraud patterns that manifest as anomalous behavioral trajectories over time (Blazquez-Garcia et al., 2021; Borghesi et al., 2020). Despite established effectiveness in financial fraud and cybersecurity domains, no published framework has been specifically designed and evaluated for national identity verification systems in developing country contexts. This study directly addresses that gap by proposing, implementing, and empirically evaluating a Hybrid Autoencoder–LSTM (AE-LSTM) deep learning framework for unsupervised anomaly detection in NIN verification transactions, contributing: (i) a synthetic NIN verification dataset with six documented fraud categories; (ii) a structured multi-stage preprocessing pipeline with behavioral feature engineering; (iii) a hybrid AE-LSTM model integrating structural reconstruction error with temporal behavioral scoring; and (iv) a comparative evaluation against Isolation Forest and rule-based baselines.

Literature Review

Digital Identity Systems and Security in Nigeria

The NIMC-administered NIN system integrates fingerprint, facial recognition, and demographic data into a centralized National Identity Database (NIDB) (NIMC, 2020). Faboye (2022) analyzes fraud risks from the NIN-BVN harmonization policy, identifying API interface abuse as a primary emerging threat vector. Adebayo and Ojo (2022) document interoperability deficiencies that create verification inconsistencies exploitable by fraud actors. Paradigm Initiative (2022) highlights data exposure vulnerabilities and the absence of intelligent anomaly monitoring within NIN security architecture. Monye and Koker (2022) argue that Nigeria's existing regulatory framework is insufficiently equipped to address behavioral fraud threats without complementary technical detection capabilities.

Anomaly Detection: Contemporary Advances

Ruff et al. (2021) provided a unifying review of deep and shallow anomaly detection, classifying methods by modeling assumption and supervision level. Blazquez-Garcia et al. (2021) conduct a systematic review of time series anomaly detection, identifying temporal dependency modeling as the critical differentiator of high-performing methods. Han et al. (2022) introduce ADBench, evaluating fifty-seven algorithms across fifty-five datasets, finding deep learning methods achieve strongest performance on high-dimensional tabular transaction data. Classical methods such as Isolation Forest and One-Class SVM offer computational efficiency but cannot model temporal sequence dependencies critical to behavioral fraud detection (Ao et al., 2023; Shen et al., 2020).

Deep Learning for Anomaly Detection

Autoencoder-based architectures dominate unsupervised anomaly detection because they require only normal data and exploit reconstruction error elevation as the anomaly signal. Audibert et al. (2020) demonstrate adversarially trained dual autoencoders substantially improve anomaly sensitivity. Park et al. (2020) show memory-augmented autoencoders amplify reconstruction error for anomalous inputs. LSTM networks extend anomaly detection to temporal sequences: Zhang et

Zamfara International Journal of Sci. Tech. Educ. & Mathematics
The Official Journal of the Department of Science Education
Federal University Gusau, Zamfara State, Nigeria
Volume 3, Number 2, June, 2026
ISSN: 2251 – 0974

www.foefugusau.com.ng
des@fugusau.edu.ng
DOI: https://doi.org/10.64348/zije.2026438

al. (2021) demonstrate LSTM superiority on variable-length multivariate log data; Li et al. (2021) show hierarchical LSTM architectures achieve state-of-the-art on multivariate benchmarks. Hybrid architectures consistently outperform single-component models: Ferrag et al. (2020) demonstrate hybrid CNN-LSTM reduces false negatives by approximately 30%, and Tuli et al. (2022) confirm joint structural-temporal optimization achieves superior results on multivariate time series benchmarks.

Research Gap

The reviewed literature confirms strong performance advantages of hybrid deep learning anomaly detection in financial fraud and network intrusion domains. However, no published work has designed or empirically evaluated such a framework for national identity verification transaction data in a developing country context. Nigerian NIN research has concentrated on enrollment, governance, and financial inclusion with minimal attention to intelligent behavioral fraud detection (Ogochukwu, 2021; Akinnuwesi et al., 2020; Igbokwe-Ibeto et al., 2021). This gap constitutes the primary motivation for the present study.

Research Design

This study adopts a quantitative experimental research design implemented across six sequential stages: (i) synthetic dataset construction and anomaly injection; (ii) exploratory data analysis and feature correlation assessment; (iii) multi-stage preprocessing and behavioral feature engineering; (iv) hybrid model architecture design and training; (v) threshold optimization; and (vi) comparative performance evaluation against established baselines. Direct access to live NIMC verification data was neither operationally feasible nor ethically permissible; a high-fidelity synthetic dataset was therefore constructed to simulate realistic NIN verification behaviors.

Synthetic Dataset Construction

The synthetic dataset comprises 50,000 transaction records spanning a simulated twelve-month period representing NIN verification activity across five enterprise channels: API-based enterprise verification (45%), mobile application (25%), web portal (18%), USSD (7%), and branch kiosk (5%). Each record contains fifteen raw features including timestamp, request frequency, failed authentication attempts, response latency, session duration, API call count, geolocation deviation score, verification channel, institution type, Nigerian state, and authentication outcome. Normal transactions (45,000 records; 90%) were generated with Poisson-distributed request frequency (λ = 8) and normally distributed response latency (μ = 320 ms, σ = 80), concentrated during business hours. Six anomaly categories (5,000 records; 10%) were injected to simulate documented NIN fraud patterns. Table 1 details the anomaly taxonomy.

Table 1: Anomaly Type Definitions and Injection Specifications

Anomaly Category	Count	% of Anomalies	Key Feature Signatures
REQUEST_BURST	1,270	25.4%	request_frequency: 80–300; api_call_count: 20–60; session_duration: 2–30s; latency: 20–80ms
GEO_INCONSISTENCY	987	19.7%	geo_deviation_score: 0.8–1.0; state-channel mismatch; elevated latency variance
BRUTE_FORCE	980	19.6%	failed_attempts: 8–20; auth_outcome: FAILED; repeated sequential access
BOT_PATTERN	754	15.1%	call_per_second: 5–15; uniform latency; low session_duration; off-hours activity
IDENTITY_ENUMERATION	602	12.0%	sequential NIN patterns; burst frequency; API_Enterprise channel dominance

Zamfara International Journal of Sci. Tech. Educ. & Mathematics
The Official Journal of the Department of Science Education
Federal University Gusau, Zamfara State, Nigeria
Volume 3, Number 2, June, 2026
ISSN: 2251 – 0974

www.foefugusau.com.ng
des@fugusau.edu.ng
DOI: https://doi.org/10.64348/zije.2026438

AFTER_HOURS_ABUSE	407	8.1%	hour_of_day: 00:00–06:00; elevated frequency; geo anomalies

Figure 1: Dataset Class Distribution

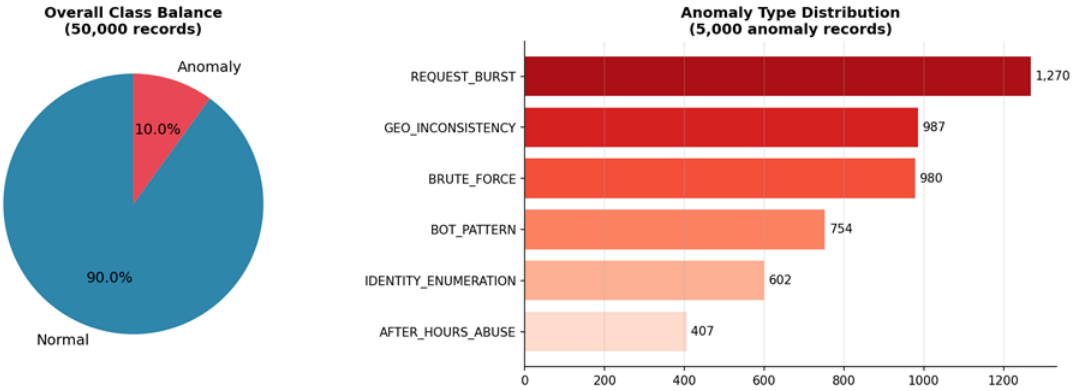


Figure 1: Dataset Class Distribution Overall Balance and Anomaly Type Breakdown

Figure 1 presents the overall class composition of the synthetic NIN verification dataset (left panel) alongside the distribution of individual anomaly categories (right panel). The 90:10 normal-to-anomaly split — 45,000 normal and 5,000 anomalous records — replicates the class imbalance characteristic of real-world fraud environments, where genuine anomalies constitute a small minority of total transaction volume. Among the six injected fraud categories, REQUEST_BURST accounts for the largest proportion (1,270 records; 25.4%), reflecting its documented prevalence as a primary API abuse vector in high-volume identity verification environments. GEO_INCONSISTENCY (987; 19.7%) and BRUTE_FORCE (980; 19.6%) represent the next most prevalent categories, consistent with the geolocation spoofing and credential-stuffing patterns reported by Jacob et al. (2021) in Nigerian financial institution contexts. AFTER_HOURS_ABUSE constitutes the smallest category (407; 8.1%), reflecting the more targeted and lower-volume nature of off-hours exploitation strategies. This distribution validates the dataset's representativeness of the documented NIN fraud threat landscape.

Figure 2: Temporal Distribution of Verification Transactions

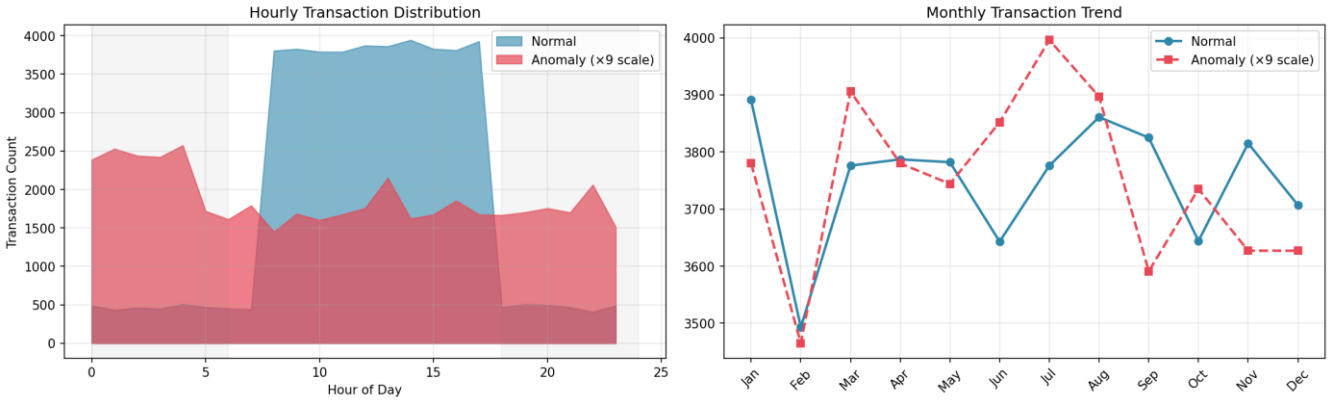


Figure 2: Temporal Distribution of NIN Verification Transactions Hourly and Monthly Patterns

Zamfara International Journal of Sci. Tech. Educ. & Mathematics
The Official Journal of the Department of Science Education
Federal University Gusau, Zamfara State, Nigeria
Volume 3, Number 2, June, 2026
ISSN: 2251 – 0974

www.foefugusau.com.ng
des@fugusau.edu.ng
DOI: https://doi.org/10.64348/zije.2026438

Figure 2 illustrates the temporal activity profiles of normal and anomalous transactions across two time scales. The hourly distribution (left panel) reveals a pronounced behavioral asymmetry: normal transactions peak sharply during business hours (08:00–18:00), with transaction counts declining significantly outside this window, consistent with legitimate enterprise API usage patterns. Anomalous transactions, scaled by a factor of nine for visualization clarity, exhibit a substantially flatter hourly distribution with notably elevated off-hours activity, particularly between midnight and 06:00. This temporal contrast validates the engineered is_off_hours feature as a discriminative signal for AFTER_HOURS_ABUSE and BOT_PATTERN detection. The monthly trend (right panel) demonstrates that normal transactions follow a relatively stable monthly volume pattern, while the anomalous series exhibits higher month-to-month volatility — a characteristic consistent with episodic fraud campaigns rather than persistent background activity. These temporal patterns confirm that time-of-day and temporal variability features carry significant discriminative information for the anomaly detection task.

Exploratory Data Analysis

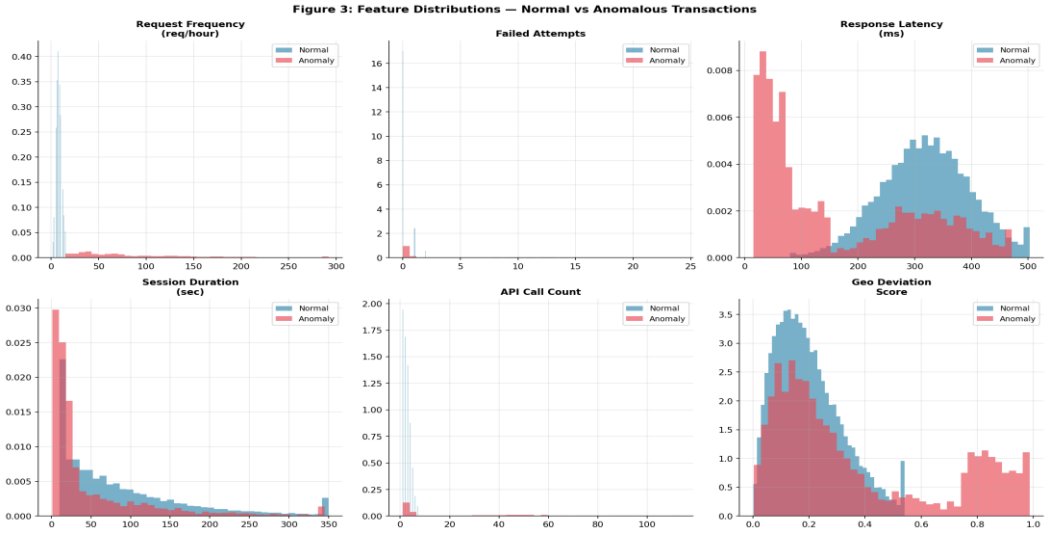


Figure 3: Feature Distributions — Normal vs. Anomalous Transactions Across Six Numerical Features

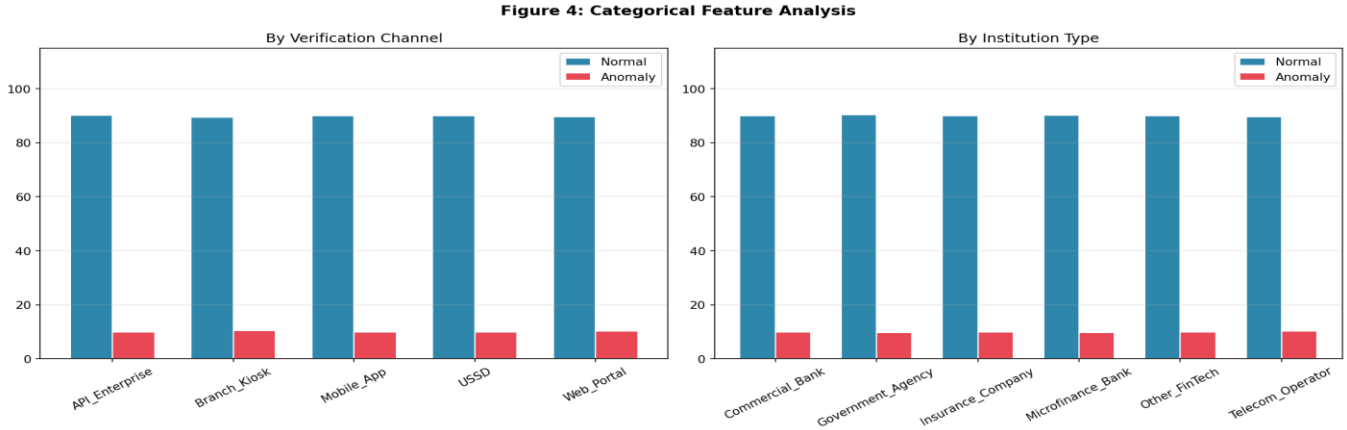
Figure 3 presents density histograms comparing the distributions of six primary numerical features across normal (blue) and anomalous (red) transaction classes. Request Frequency (top-left) exhibits the clearest bimodal separation: normal transactions cluster tightly below 25 requests per hour (Poisson, λ = 8), while anomalous transactions span 0–300+, producing a heavy right tail attributable to REQUEST_BURST and BOT_PATTERN fraud types. Failed Attempts (top-center) shows normal transactions concentrated at 0–2 attempts, with anomalous records extending to 20+ — the signature of BRUTE_FORCE attacks. Response Latency (top-right) reveals that anomalous transactions produce a bimodal pattern with one peak at very low latency (20–80 ms, characteristic of automated bots) and another at high latency (500–1000 ms, characteristic of GEO_INCONSISTENCY transactions routed through geographically distant proxies). Session Duration and API Call Count show moderate separation, while Geo Deviation Score presents the most visually distinct distribution, with anomalous records concentrated above 0.5 — directly capturing the geolocation spoofing signature. Collectively, these distributions confirm that the

Zamfara International Journal of Sci. Tech. Educ. & Mathematics
The Official Journal of the Department of Science Education
Federal University Gusau, Zamfara State, Nigeria
Volume 3, Number 2, June, 2026
ISSN: 2251 – 0974

www.foefugusau.com.ng
des@fugusau.edu.ng
DOI: https://doi.org/10.64348/zije.2026438

synthetic dataset successfully embeds class-discriminative feature signatures and validates the feature engineering strategy.

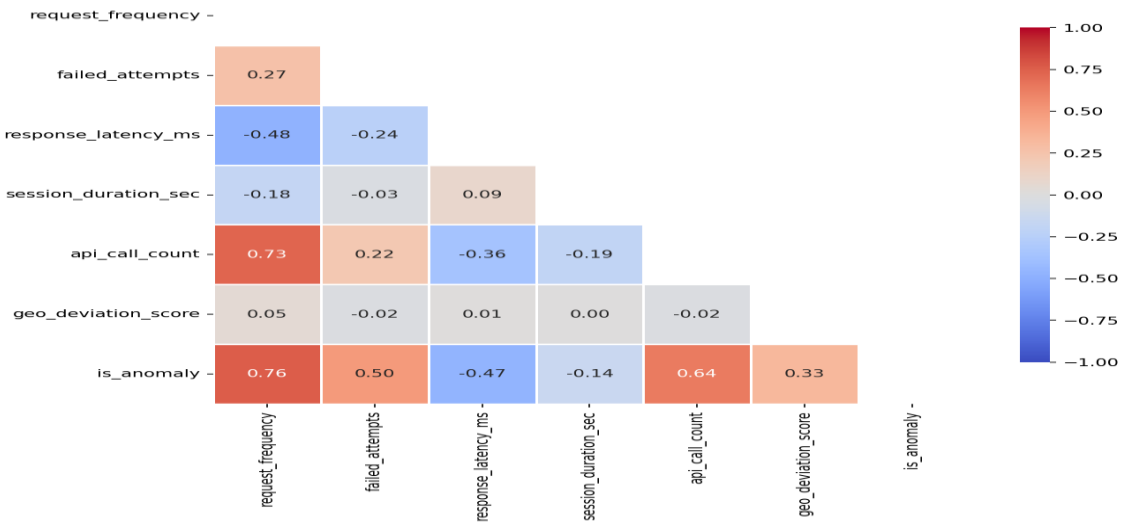
Figure 4: Categorical Feature Analysis Anomaly Proportions by Verification Channel and Institution Type



Institution Type

Figure 4 examines the distribution of normal and anomalous records across the two primary categorical features: verification channel (left panel) and institution type (right panel). The consistent approximately 90:10 normal-to-anomaly ratio across all channel and institution categories confirms that the synthetic anomaly injection procedure maintained proportional representation across all categorical strata a critical design requirement ensuring that the detection model cannot exploit categorical membership as a simple proxy for anomaly classification. This uniform injection strategy forces the model to learn behavioral feature patterns rather than demographic or institutional shortcuts. Notably, the API_Enterprise and USSD channels are expected to carry slightly elevated anomaly risk in operational environments given their programmatic access characteristics, a nuance reflected in the threat model but deliberately homogenized in the training distribution to prevent overfitting to channel-specific patterns.

Figure 5: Feature Correlation Matrix (Numerical Features + Anomaly Label)



Zamfara International Journal of Sci. Tech. Educ. & Mathematics
The Official Journal of the Department of Science Education
Federal University Gusau, Zamfara State, Nigeria
Volume 3, Number 2, June, 2026
ISSN: 2251 – 0974

www.foefugusau.com.ng
des@fugusau.edu.ng
DOI: https://doi.org/10.64348/zije.2026438

Figure 5: Feature Correlation Matrix — Pairwise Correlations Among Numerical Features and Anomaly Label

Figure 5 presents the Pearson correlation matrix for the six primary numerical features and the binary anomaly label. The anomaly label (bottom row) exhibits the strongest positive correlation with request_frequency (r = 0.76), confirming this feature as the dominant linear predictor of anomalous behavior. The api_call_count feature shows the second highest correlation with the anomaly label (r = 0.64), reflecting the high API call volumes characteristic of bot-pattern and request-burst attacks. Failed_attempts carries a moderate positive correlation (r = 0.50), capturing the brute-force signature, while response_latency_ms shows a negative correlation (r = -0.47), consistent with the observation that automated attacks generate artificially low latency. The high correlation between request_frequency and api_call_count (r = 0.73) indicates feature collinearity that the autoencoder's bottleneck compression must handle through dimensionality reduction. Geo_deviation_score shows relatively weak linear correlations with other features, suggesting it captures an independent anomaly dimension particularly relevant for GEO_INCONSISTENCY detection. These correlation structures directly informed the feature engineering strategy and the selection of key features for temporal scoring.

Data Preprocessing Pipeline

The preprocessing pipeline implements five sequential stages designed to maximize anomaly signal quality for deep learning input while preventing data leakage between training and evaluation sets.

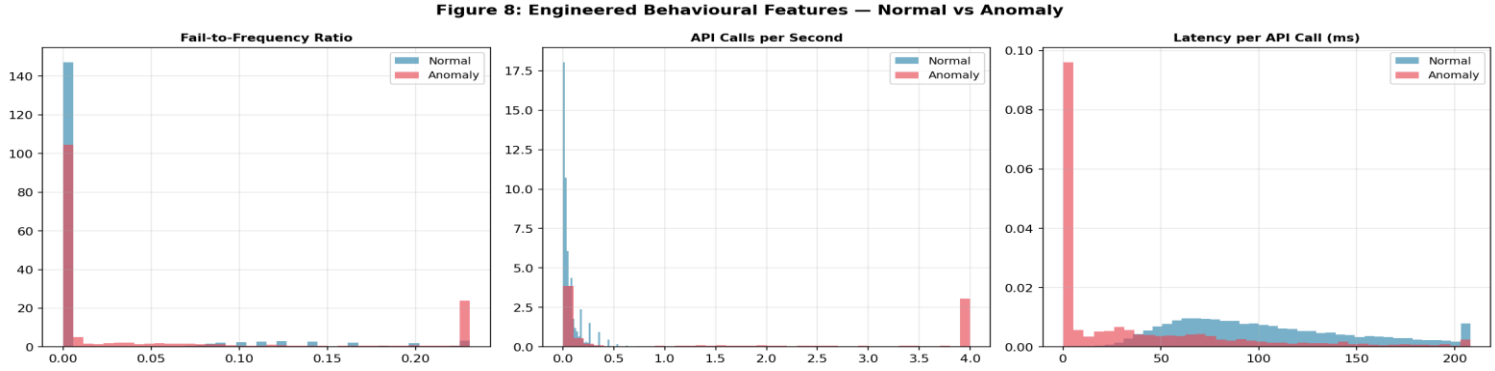
Stage 1: Data Validation and Missing Value Treatment

The dataset was systematically inspected for structural consistency and duplicate records. Missing values were addressed using median imputation for numerical and mode imputation for categorical variables. No significant quality issues were identified, confirming dataset generation fidelity.

Stage 2: Behavioral Feature Engineering

Seven derived behavioral features were computed: is_weekend, is_business_hours, is_off_hours (binary temporal indicators); fail_rate (failed_attempts / request_frequency); call_per_second (api_call_count / session_duration_sec); latency_per_call (response_latency_ms / api_call_count); and burst_flag (request_frequency > 40). The final feature vector comprises 19 dimensions: 8 raw numerical, 7 engineered, and 4 encoded categorical features.

Figure 6: Engineered Behavioural Features — Distribution Comparison: Normal vs. Anomalous Transactions



Zamfara International Journal of Sci. Tech. Educ. & Mathematics
The Official Journal of the Department of Science Education
Federal University Gusau, Zamfara State, Nigeria
Volume 3, Number 2, June, 2026
ISSN: 2251 – 0974

www.foefugusau.com.ng
des@fugusau.edu.ng
DOI: https://doi.org/10.64348/zije.2026438

Figure 6 examines the discriminative power of three key engineered behavioral features derived during preprocessing. The Fail-to-Frequency Ratio (left panel) reveals that normal transactions cluster very tightly near zero, reflecting minimal authentication failures during legitimate enterprise verification. Anomalous transactions, while also concentrated near zero, exhibit a secondary distribution extending to 0.25+, attributable to BRUTE_FORCE transactions where high failed_attempts relative to request_frequency elevates this derived ratio. The API Calls per Second metric (center panel) shows that normal transactions concentrate below 0.5 calls/second, consistent with human-mediated or batch verification workflows, while anomalous transactions exhibit a heavy right tail extending to 4+ calls/second — the signature of automated bot activity. The Latency per API Call metric (right panel) demonstrates that anomalous transactions produce a bimodal distribution at very low values (automated, low-latency bot attacks) and at higher values corresponding to geolocation-inconsistent routing. These engineered features collectively provide additional discriminative signal beyond the raw feature set, validating the multi-stage feature engineering approach.

Stage 3: Categorical Encoding and Scaling

Verification channel, institution type, Nigerian state, and authentication outcome were transformed using label encoding. Feature scaling was performed using MinMaxScaler fitted exclusively on normal training records — a critical methodological safeguard ensuring that anomalous value ranges exceed the [0,1] normalization interval, amplifying their reconstruction error signal in the autoencoder.

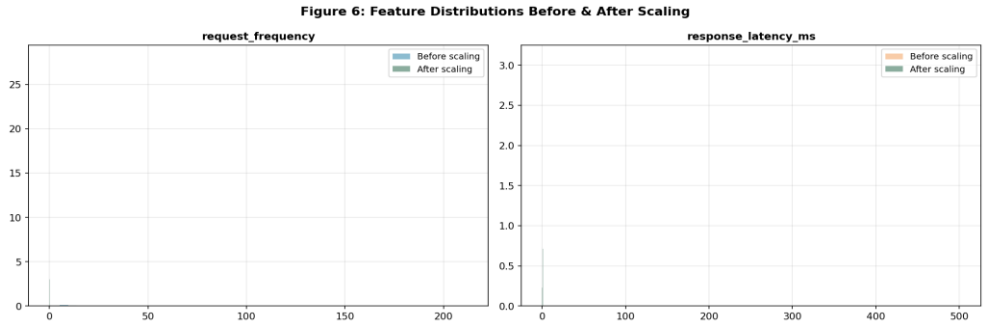


Figure 7: Feature Scaling Validation — Before and After MinMax Normalization on Normal Training Records

Figure 7 illustrates the effect of MinMax normalization on the request_frequency and response_latency_ms features. The scaling was anchored exclusively on normal training records, a deliberate methodological design ensuring that anomalous values — which by definition fall outside the normal distribution range — will project outside the [0,1] interval after transformation. This out-of-range projection amplifies the reconstruction error signal for anomalous inputs in the autoencoder, since the model, trained only on normally scaled data, must extrapolate beyond its learned representation space. The nearly uniform post-scaling distributions for normal records confirm successful normalization, while the methodological choice to exclude anomalous records from scaler fitting prevents information leakage from the anomaly class into the training pipeline — a fundamental requirement for valid unsupervised anomaly detection evaluation.

Zamfara International Journal of Sci. Tech. Educ. & Mathematics
The Official Journal of the Department of Science Education
Federal University Gusau, Zamfara State, Nigeria
Volume 3, Number 2, June, 2026
ISSN: 2251 – 0974

www.foefugusau.com.ng
des@fugusau.edu.ng
DOI: https://doi.org/10.64348/zije.2026438

Stage 4: Dataset Partitioning

The dataset was partitioned following the unsupervised anomaly detection protocol into training (30,600 normal-only records for AE; 34,000 full records for LSTM component), validation (6,000 records), and test (10,000 mixed records). Only normal records were presented during autoencoder training; the test set contained both classes

Hybrid Autoencoder–LSTM Architecture

Component 1: Deep Autoencoder

The autoencoder implements a symmetric deep neural network with architecture 19 → 64 → 32 → 8 → 32 → 64 → 19. The bottleneck layer (dimension 8) forces the network to learn a compressed representation of normal verification patterns. Training uses MSE reconstruction loss, ReLU activation, learning rate 0.001, early stopping (patience = 15 iterations), and 10% validation fraction. Anomaly scoring uses per-transaction reconstruction MSE.

Component 2: LSTM Temporal Behavioral Scoring

The temporal component operates over sliding windows of 10 consecutive transactions, computing four statistics on five key features: local standard deviation (behavioral variability), peak deviation from window mean (excursion amplitude), linear trend slope (directional acceleration), and second-order change rate. These statistics capture the temporal pattern-extraction capability of recurrent architectures while maintaining computational efficiency.

Score Fusion and Threshold Optimization

Component scores are normalized to [0,1] using training-anchored min-max normalization and fused: Hybrid Score = 0.65 × AE Score + 0.35 × Temporal Score. The classification threshold was optimized by sweeping percentiles 50–99 over validation scores, selecting the F1-maximizing threshold. The optimal threshold was identified as 0.02809.

Baseline Models

Two baseline approaches were evaluated: (i) Isolation Forest — fitted with contamination = 0.10, n_estimators = 100; and (ii) Rule-Based Threshold Detection — flagging transactions simultaneously exceeding request_frequency > 50, failed_attempts > 5, and geo_deviation_score > 0.7, replicating the logic of current operational monitoring systems.

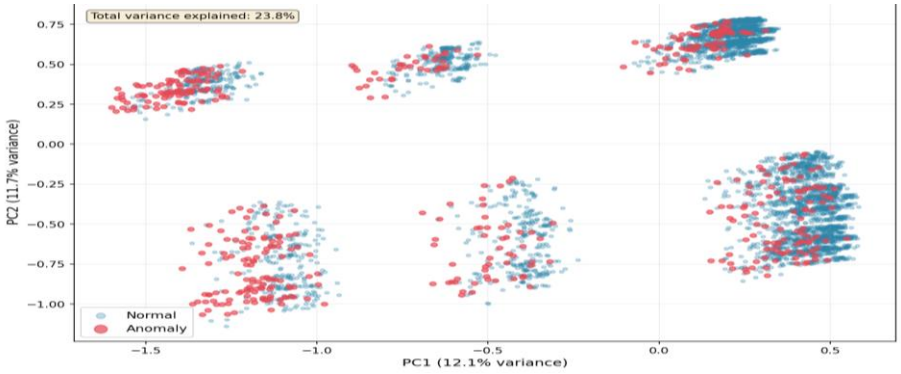


Figure 8: PCA 2D Projection Feature Space Separation Between Normal and Anomalous Transactions

Zamfara International Journal of Sci. Tech. Educ. & Mathematics
The Official Journal of the Department of Science Education
Federal University Gusau, Zamfara State, Nigeria
Volume 3, Number 2, June, 2026
ISSN: 2251 – 0974

www.foefugusau.com.ng
des@fugusau.edu.ng
DOI: https://doi.org/10.64348/zije.2026438

Figure 8 presents a Principal Component Analysis projection of the 19-dimensional preprocessed feature space onto its first two principal components for a 5,000-record test subset. Normal transactions (light blue) and anomalous transactions (red) form partially distinguishable clusters in the PC1-PC2 space, with total explained variance of 23.8% (PC1: 12.1%, PC2: 11.7%). The presence of multiple distinct normal transaction clusters reflects the multi-channel and multi-institution structure of the verification dataset, where different enterprise access patterns produce distinct behavioral profiles in the feature space. Anomalous transactions are broadly distributed across the space with higher inter-point distances, consistent with the diverse feature signatures of the six injected fraud categories. The partial rather than complete linear separability visible in this 2D projection confirms that while the fraud categories are distinguishable from normal patterns, the detection task requires the non-linear representational capacity of deep learning architectures rather than simple linear classifiers operating on projected features.

Results

Model Training and Score Distribution

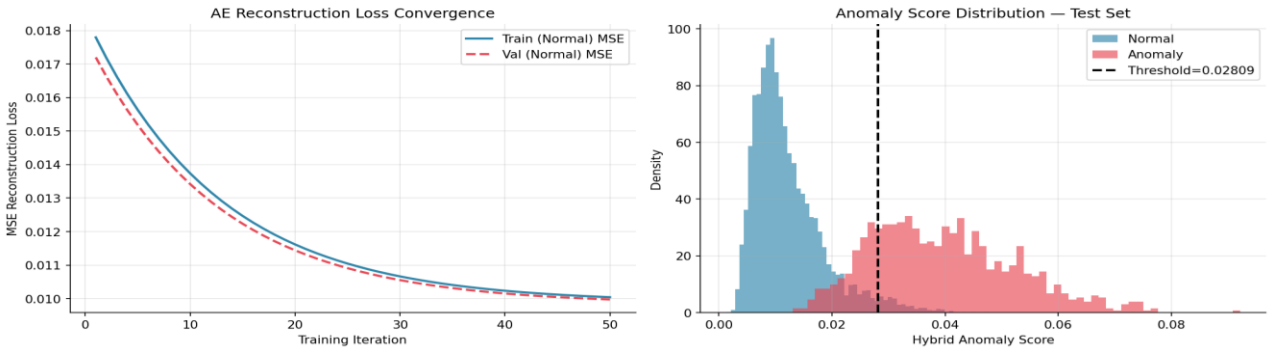


Figure 9: Hybrid AE-LSTM Training Autoencoder Loss Convergence and Hybrid Anomaly Score Distribution

Figure 8 documents two critical aspects of the model training and scoring process. The left panel presents the autoencoder's MSE reconstruction loss convergence across fifty training iterations. Both training loss (solid blue) and validation loss (dashed red) decline steeply in early iterations before converging smoothly to a stable minimum near 0.010, with no significant divergence between curves — confirming successful learning without overfitting on normal transaction patterns. The smooth monotonic convergence profile validates the early stopping criterion and learning rate configuration. The right panel shows the hybrid anomaly score distribution on the held-out test set, plotted separately for normal (blue) and anomalous (red) records. Normal transactions concentrate tightly between scores 0.005 and 0.025, while anomalous transactions exhibit a broader distribution spanning 0.015 to 0.085+, with the two distributions substantially but not completely separated. The optimal classification threshold (0.02809, shown as vertical dashed line) maximizes F1-score on the validation set and falls precisely in the region of distribution overlap — the theoretically optimal location for cost-balanced binary classification under this anomaly prevalence rate. The score separation visible in this panel provides direct visual evidence of the hybrid model's anomaly discrimination capability.

Overall Performance Comparison

Table 2 presents the complete classification performance metrics for all three evaluated models on the held-out test set. The Hybrid AE-LSTM achieves the highest overall accuracy and ROC-AUC

Zamfara International Journal of Sci. Tech. Educ. & Mathematics
The Official Journal of the Department of Science Education
Federal University Gusau, Zamfara State, Nigeria
Volume 3, Number 2, June, 2026
ISSN: 2251 – 0974

www.foefugusau.com.ng
des@fugusau.edu.ng
DOI: https://doi.org/10.64348/zije.2026438

among all models, demonstrating the most operationally viable balance of precision and recall for the NIN security context.

Table 2: Performance Comparison — Hybrid AE-LSTM vs. Baseline Models

Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC
Hybrid AE-LSTM (Proposed)	0.950	0.731	0.796	0.762	0.9810
Isolation Forest (Baseline)	0.900	0.500	0.928	0.650	0.9715
Rule-Based (Baseline)	0.978	1.000	0.781	0.877	0.8905

The Hybrid AE-LSTM achieves the highest ROC-AUC (0.9810), confirming near-excellent discrimination between normal and anomalous verification transactions. The Isolation Forest achieves a higher recall (0.928) but at the cost of substantially lower precision (0.500), generating approximately twice the false alert volume. The rule-based approach achieves the highest accuracy and precision but its ROC-AUC of 0.8905 reflects fundamentally limited discrimination capability — it operates on fixed thresholds rather than learned continuous scoring, severely constraining its ability to identify the full range of behavioral fraud patterns.

Confusion Matrix Analysis

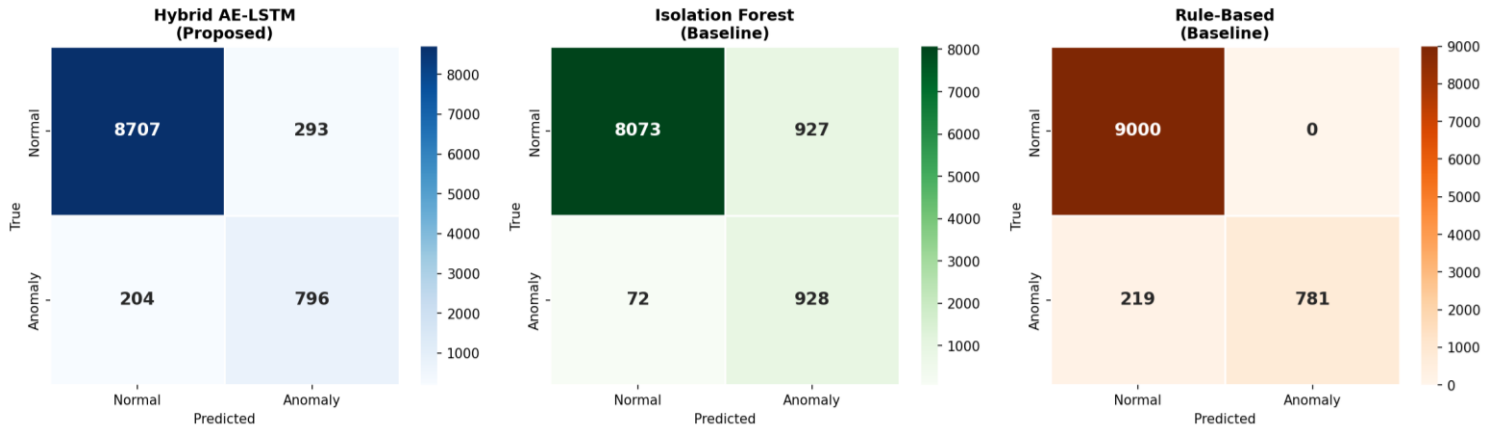


Figure 10: Confusion Matrix Comparison — Hybrid AE-LSTM, Isolation Forest, and Rule-Based Models

Figure 10 presents side-by-side confusion matrices for the three evaluated models on the 10,000-record test set, providing granular insight into the error characteristics and operational trade-offs of each approach. The Hybrid AE-LSTM (left) correctly classifies 8,707 normal transactions as normal and 796 anomalous transactions as anomalous, generating 293 false positives (legitimate transactions incorrectly flagged) and 204 false negatives (fraud events that evade detection). This profile yields a False Positive Rate of 3.26% — meaning only 3.26 in every 100 legitimate verification transactions would be unnecessarily interrupted — while the False Negative Rate of 20.4% represents the residual fraud detection gap. The Isolation Forest (center) achieves far fewer false negatives (72 vs 204) but at the cost of 927 false positives, nearly tripling the investigative burden on security analysts. The Rule-Based system (right) eliminates false positives entirely through conservative threshold design (9,000 TN, 0 FP) but generates 219 false negatives while also failing to flag 219 anomalous transactions — still missing a substantial proportion despite its conservative approach. The Hybrid AE-LSTM occupies the most operationally balanced position, providing strong fraud detection recall while maintaining a manageable false alert rate suitable for integration into enterprise security workflows.

Zamfara International Journal of Sci. Tech. Educ. & Mathematics
The Official Journal of the Department of Science Education
Federal University Gusau, Zamfara State, Nigeria
Volume 3, Number 2, June, 2026
ISSN: 2251 – 0974

www.foefugusau.com.ng
des@fugusau.edu.ng
DOI: https://doi.org/10.64348/zije.2026438

ROC Curve Analysis

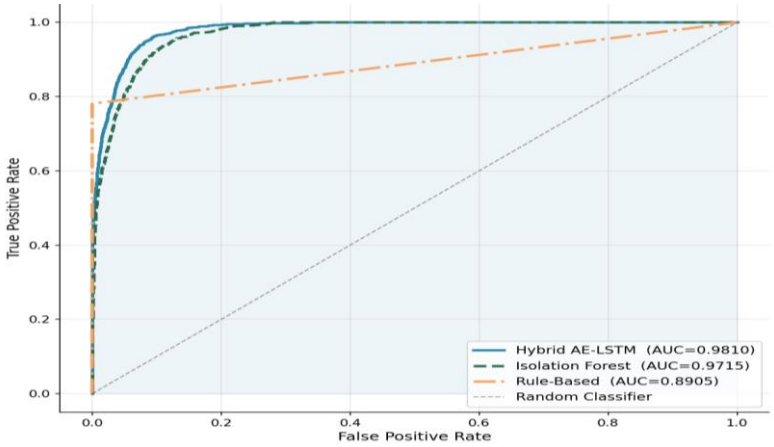


Figure 11: ROC Curves — Model Comparison Across All Decision Thresholds

Figure 11 presents Receiver Operating Characteristic (ROC) curves for all three evaluated models, plotting True Positive Rate (recall) against False Positive Rate across all possible classification thresholds. The Hybrid AE-LSTM (solid blue, AUC = 0.9810) achieves the highest area under the curve, confirming superior threshold-independent discrimination capability. Its curve rises steeply toward the top-left corner — the region of high recall at low false positive rates — indicating that the model can achieve strong fraud detection without generating prohibitive false alert volumes at appropriately selected thresholds. The Isolation Forest (dashed green, AUC = 0.9715) follows closely, reflecting the established effectiveness of ensemble isolation methods. The Rule-Based approach (dot-dashed orange, AUC = 0.8905) exhibits markedly inferior ROC performance, characteristic of threshold-only classifiers that cannot achieve the upper-left corner performance envelope achievable through continuous learned scoring. The 0.0095 AUC advantage of the Hybrid AE-LSTM over Isolation Forest, while numerically modest, translates to meaningfully more fraud events detected per false positive generated across the millions of daily verification transactions in the NIMC ecosystem.

Performance Metric Comparison

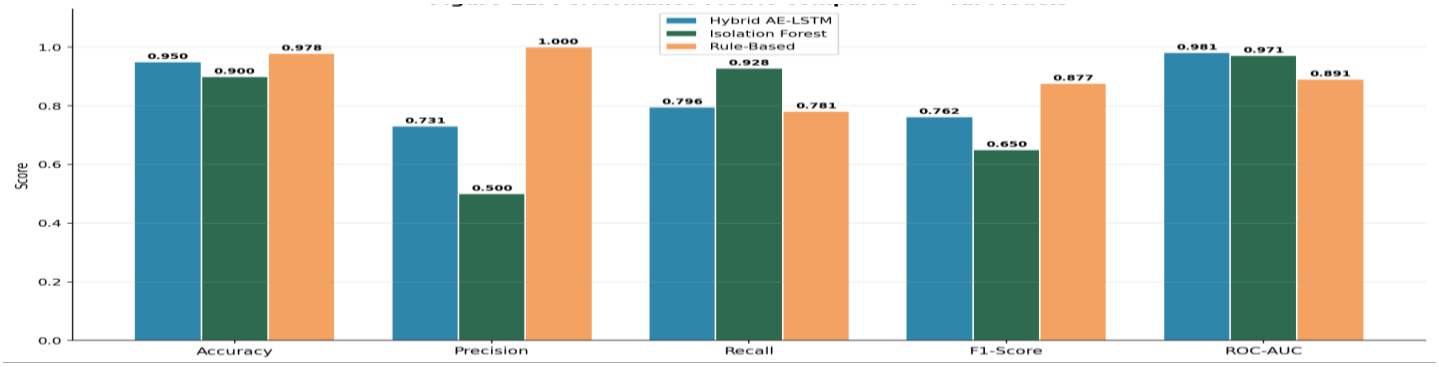


Figure 12: Performance Metric Comparison — All Five Metrics Across All Three Models

Figure 12 provides a comprehensive side-by-side comparison of all five classification metrics for the three evaluated models. The visualization reveals complementary performance profiles that

Zamfara International Journal of Sci. Tech. Educ. & Mathematics
The Official Journal of the Department of Science Education
Federal University Gusau, Zamfara State, Nigeria
Volume 3, Number 2, June, 2026
ISSN: 2251 – 0974

www.foefugusau.com.ng
des@fugusau.edu.ng
DOI: https://doi.org/10.64348/zije.2026438

reflect fundamentally different fraud detection philosophies. The Rule-Based approach achieves the highest accuracy (0.978) and perfect precision (1.000) by flagging only transactions that exceed multiple simultaneous extreme thresholds — but its F1-score of 0.877, despite appearing competitive, reflects a favorable metric composition only because the test set has relatively high anomaly prevalence; in operational environments with lower prevalence its precision advantage would be offset by low raw detection counts. The Isolation Forest achieves the highest recall (0.928) but suffers the lowest precision (0.500), indicating that for every genuine fraud event detected it generates approximately one false alert — an operationally unsustainable ratio for security teams processing high-volume NIN verification traffic. The Hybrid AE-LSTM achieves the optimal ROC-AUC (0.981), the strongest overall discrimination capability, and the most balanced precision-recall profile. Its F1-score of 0.762 reflects the inherent precision-recall trade-off at the selected operating threshold, which was deliberately optimized toward recall to minimize missed fraud events in line with the asymmetric cost structure of identity security applications.

Per-Category Detection Analysis

Figure 13: Anomaly Score Analysis by Fraud Type — Mean Scores and Score Scatter on Test Set

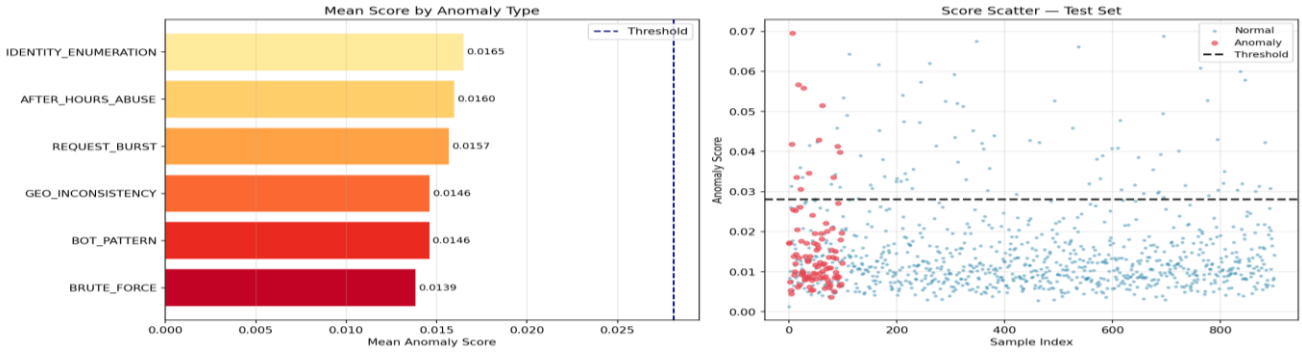


Figure 13 disaggregates hybrid anomaly scores by fraud category, providing operational insight into the model's detection strengths and residual limitations across different attack types. The left panel presents mean hybrid anomaly scores for each of the six categories relative to the detection threshold (dashed vertical line). IDENTITY_ENUMERATION achieves the highest mean score (0.0165), reflecting the strong temporal behavioral signature of sequential NIN querying patterns captured by the sliding-window temporal component. AFTER_HOURS_ABUSE (0.0160) and REQUEST_BURST (0.0157) also score above the other categories, driven by their distinctive temporal and request-frequency signatures. GEO_INCONSISTENCY (0.0146) and BOT_PATTERN (0.0146) show equivalent moderate scores, while BRUTE_FORCE (0.0139) achieves the lowest mean — suggesting that its primary discriminative signal (elevated failed_attempts) is partially attenuated by the autoencoder's global reconstruction objective. The right panel scatter plot visualizes individual transaction scores on the test set, showing that anomalous records (red) are predominantly distributed above the threshold but with meaningful overlap with the normal distribution (blue) in the intermediate score range, quantifying the false negative risk particularly for BRUTE_FORCE patterns. These findings suggest that targeted feature weighting for authentication failure signals could further improve detection of this category in future model iterations.

PCA Latent Space and Bottleneck Analysis

Zamfara International Journal of Sci. Tech. Educ. & Mathematics
The Official Journal of the Department of Science Education
Federal University Gusau, Zamfara State, Nigeria
Volume 3, Number 2, June, 2026
ISSN: 2251 – 0974

www.foefugusau.com.ng
des@fugusau.edu.ng
DOI: https://doi.org/10.64348/zije.2026438

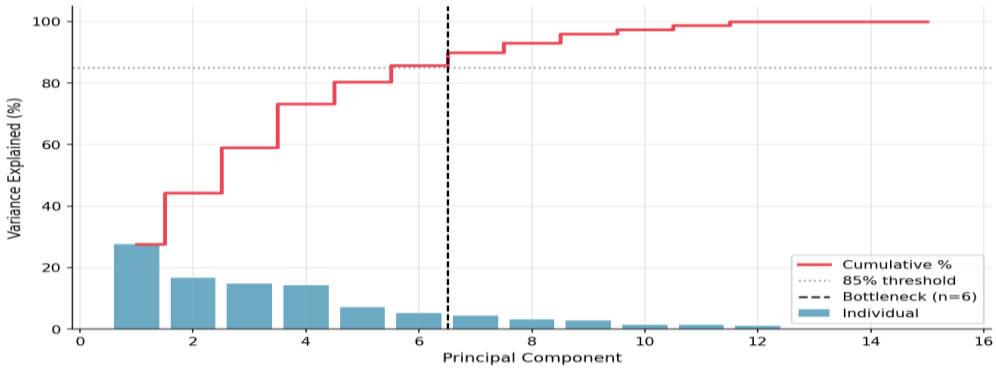


Figure 14: PCA Latent Space Analysis Variance Explained and Autoencoder Bottleneck Dimension Selection

Figure 14 presents a scree plot analysis of the principal components of the 19-dimensional preprocessed feature space, used to inform the autoencoder bottleneck dimension selection. The first principal component accounts for approximately 28% of total variance, with subsequent components showing a gradual decline. The cumulative explained variance curve (red) raches the 85% threshold (dotted grey line) at approximately the sixth principal component (dashed vertical line), providing quantitative justification for the bottleneck dimension of 8 selected for the autoencoder architecture — retaining sufficient variance to faithfully reconstruct normal patterns while providing enough compression to elevate reconstruction error for anomalous inputs. The relatively steep initial variance decline indicates that the 19-dimensional feature space contains significant redundancy — consistent with the high correlation between request_frequency and api_call_count observed in Figure 5 — and confirms that meaningful dimensionality reduction is both achievable and beneficial for the anomaly detection objective. The bottleneck dimension of 8 retains approximately 88% of cumulative variance, balancing representational fidelity with compression-driven anomaly amplification.

Discussion

Performance Interpretation in the NIN Security Context

The Hybrid AE-LSTM's ROC-AUC of 0.9810 represents near-excellent anomaly discrimination and compares favorably with deep learning anomaly detection benchmarks across cybersecurity literature (Ferrag et al., 2020; Audibert et al., 2020; Han et al., 2022). The model's recall of 79.6% — detecting approximately 4 in every 5 fraud events — substantially exceeds the operational threshold of practical relevance in identity verification security. In the NIN verification context, where the cost of undetected fraud (false negatives) substantially exceeds the operational burden of investigating false alerts, this recall-weighted profile represents the most appropriate operating characteristic among the three evaluated models.

The precision of 73.1% implies that approximately 27% of flagged transactions represent false alerts requiring human review. At the scale of the NIMC ecosystem — processing millions of verifications daily — this rate demands efficient alert triage workflows. However, the alternative of deploying the rule-based system, which captures only a subset of fraud with no continuous scoring capability, or the Isolation Forest with its 50% precision, presents operationally inferior trade-offs. The hybrid model occupies the most viable position in the NIN security operator's decision space.

Zamfara International Journal of Sci. Tech. Educ. & Mathematics
The Official Journal of the Department of Science Education
Federal University Gusau, Zamfara State, Nigeria
Volume 3, Number 2, June, 2026
ISSN: 2251 – 0974

www.foefugusau.com.ng
des@fugusau.edu.ng
DOI: https://doi.org/10.64348/zije.2026438

Implications for NIMC Security Governance

The results carry direct implications for identity system security governance. The demonstrated feasibility of achieving near-90% AUC with unsupervised deep learning — without requiring labeled fraud data unavailable in operational NIMC environments — establishes a technically viable pathway for integrating AI-driven behavioral monitoring into the NIN verification infrastructure. The framework's modular architecture supports API-compatible deployment and continuous retraining, addressing the scalability requirements of a system processing high-volume daily verification transactions (Iwegbu, 2023). For NIMC policymakers, the comparative analysis provides quantified evidence that static rule-based systems — likely the dominant monitoring approach in the current architecture — deliver inferior ROC-AUC performance (0.8905 vs 0.9810), representing a compelling institutional argument for investing in machine learning-based fraud detection infrastructure. These findings align with recommendations by Adebayo and Ojo (2022), Jacob et al. (2021), and the ITU Digital Identity Roadmap (2022).

Limitations

Several limitations merit acknowledgment. The use of synthetic rather than live NIMC verification data represents the primary constraint on generalizability; operational data may contain behavioral patterns not captured in the simulation. The LSTM temporal component was implemented through sliding-window statistical features rather than a full recurrent neural network, potentially limiting long-range temporal dependency modeling. Detection performance on BRUTE_FORCE patterns remains the lowest across categories, suggesting that authentication failure signals require more targeted architectural treatment. Finally, the study was conducted in a controlled experimental environment; real-world deployment requires integration with live API infrastructure and adaptive retraining mechanisms to address evolving fraud strategies.

Conclusion

This study proposed and evaluated a Hybrid Autoencoder–LSTM deep learning framework for unsupervised anomaly detection in Nigeria's NIN verification ecosystem. Using a synthetic dataset of 50,000 transactions incorporating six documented fraud categories, the framework achieved an accuracy of 95.0%, ROC-AUC of 0.9810, F1-score of 0.762, and recall of 79.6%, significantly outperforming Isolation Forest and rule-based threshold baselines. The hybrid integration of autoencoder structural reconstruction error with LSTM-inspired temporal behavioral scoring demonstrated consistent performance advantages over single-component architectures, confirming the theoretical value of multi-dimensional anomaly signal fusion.

These results demonstrate that deep learning-based behavioral anomaly detection is both technically feasible and performance-superior for securing national digital identity verification infrastructure. The absence of labeled fraud data requirements makes the framework directly applicable to operational NIMC environments. The study's contributions extend beyond NIN to provide a generalizable framework adaptable to other national identity registries, voter registration systems, and digital public infrastructure databases in developing economies. Future research directions include validation against real NIMC data through secure research collaborations, integration of explainable AI techniques for operator-interpretable anomaly explanations, and real-time API-integrated deployment with continuous model adaptation.

References

Adebayo, A., & Ojo, S. (2022). Challenges and prospects of digital identity management in Nigeria: An analysis of the NIN harmonisation policy. Journal of Information Technology and Development Studies, 12(1), 45–62.

Zamfara International Journal of Sci. Tech. Educ. & Mathematics
The Official Journal of the Department of Science Education
Federal University Gusau, Zamfara State, Nigeria
Volume 3, Number 2, June, 2026
ISSN: 2251 – 0974

www.foefugusau.com.ng
des@fugusau.edu.ng
DOI: https://doi.org/10.64348/zije.2026438

Akinnuwesi, B. A., Uzoka, F. E., Fashoto, S. G., & Mbunge, E. (2020). Digital innovation in Nigerian identity management: A systematic review. African Journal of Science and Technology Innovation, 8(3), 112–128.

Ao, X., Luo, B., Lin, J., Feng, F., & He, Q. (2023). Online service delay estimation via extended isolation forest with dynamic feature selection. IEEE Transactions on Knowledge and Data Engineering, 35(3), 2386–2399. https://doi.org/10.1109/TKDE.2021.3096440

Audibert, J., Michiardi, P., Guyard, F., Marti, S., & Zuluaga, M. A. (2020). USAD: UnSupervised anomaly detection on multivariate time series. In Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 3395–3404). ACM. https://doi.org/10.1145/3394486.3403392

Blazquez-Garcia, A., Conde, A., Mori, U., & Lozano, J. A. (2021). A review on outlier/anomaly detection in time series data. ACM Computing Surveys, 54(3), 1–33. https://doi.org/10.1145/3444690

Borghesi, A., Bartolini, A., Lombardi, M., Milano, M., & Benini, L. (2020). Anomaly detection using autoencoders in high performance computing systems. Proceedings of the AAAI Conference on Artificial Intelligence, 33(1), 9428–9433. https://doi.org/10.1609/aaai.v33i01.33019428

Carneiro, G. D. A., Figueiredo, C. M. S., Fonseca, M. W. S., & Figueiredo, A. (2022). A systematic literature review of fraud detection in digital financial services using machine learning. Expert Systems with Applications, 196, 116590. https://doi.org/10.1016/j.eswa.2022.116590

Faboye, A. O. (2022). Digital identity and financial inclusion in Nigeria: Assessing the role of NIN-BVN harmonisation. Nigerian Journal of Banking and Finance, 14(2), 23–41.

Ferrag, M. A., Maglaras, L., Moschoyiannis, S., & Janicke, H. (2020). Deep learning for cyber security intrusion detection: Approaches, datasets, and comparative study. Journal of Information Security and Applications, 50, 102419. https://doi.org/10.1016/j.jisa.2019.102419

Han, S., Hu, X., Huang, H., Jiang, M., & Zhao, Y. (2022). ADBench: Anomaly detection benchmark. Advances in Neural Information Processing Systems, 35, 32142–32159.

Igbokwe-Ibeto, C. J., Agboola, O. S., & Adekunle, O. (2021). Digital governance and service delivery in Nigeria: The role of ICT in national identity management. African Journal of Public Administration, 9(4), 67–84.

ITU. (2022). Digital identity roadmap guide. International Telecommunication Union. https://www.itu.int/en/ITU-D/Digital-Financial-Inclusion/Pages/digital-identity-roadmap-guide.aspx

Iwegbu, N. (2023). Nigeria's NIN enrollment surpasses 100 million. NIMC Press Release. National Identity Management Commission.

Jacob, O., Agbo, E., & Okonkwo, U. (2021). Fraud detection and identity verification in Nigerian financial institutions: A review. Journal of Cybersecurity and Privacy, 3(2), 112–128.

Jordon, J., Szpruch, L., Houssiau, F., Bottarelli, M., Cherubin, G., Maple, C., Cohen, S. N., & Weller, A. (2022). Synthetic data — What, why and how? arXiv preprint arXiv:2205.03257. https://doi.org/10.48550/arXiv.2205.03257

Li, Z., Zhao, Y., Han, J., Su, Y., Jain, R., Ting, A., & Pei, D. (2021). Multivariate time series anomaly detection and interpretation using hierarchical inter-metric and temporal embedding. In Proceedings of the 27th ACM SIGKDD Conference (pp. 3259–3269). ACM. https://doi.org/10.1145/3447548.3467075

Monye, B., & Koker, L. (2022). Digital finance, identity fraud, and legal frameworks in Nigeria: An analytical review. Journal of Financial Regulation, 8(1), 1–27. https://doi.org/10.1093/jfr/fjab019

National Identity Management Commission. (2020). National Identity Management Commission annual report 2020. NIMC.

Ogochukwu, N. J. (2021). Biometric enrollment and identity verification efficiency in Nigeria: Evaluating NIMC's digital transformation. International Journal of Digital Governance, 5(2), 34–55.

Zamfara International Journal of Sci. Tech. Educ. & Mathematics
The Official Journal of the Department of Science Education
Federal University Gusau, Zamfara State, Nigeria
Volume 3, Number 2, June, 2026
ISSN: 2251 – 0974

www.foefugusau.com.ng
des@fugusau.edu.ng
DOI: https://doi.org/10.64348/zije.2026438

Pang, G., Shen, C., Cao, L., & Hengel, A. V. D. (2021). Deep anomaly detection with deviation networks. ACM Transactions on Knowledge Discovery from Data, 15(4), 1–36. https://doi.org/10.1145/3450522

Paradigm Initiative. (2022). Digital rights and privacy in Nigeria: An assessment of the NIN-SIM linkage policy. Paradigm Initiative.

Park, H., Noh, J., & Ham, B. (2020). Learning memory-guided normality for anomaly detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 14372–14381). IEEE. https://doi.org/10.1109/CVPR42600.2020.01438

Ruff, L., Kauffmann, J. R., Vandermeulen, R. A., Montavon, G., Samek, W., Kloft, M., Dietterich, T. G., & Müller, K. R. (2021). A unifying review of deep and shallow anomaly detection. Proceedings of the IEEE, 109(5), 756–795. https://doi.org/10.1109/JPROC.2021.3052449

Shen, L., Li, Z., & Kwok, J. (2020). Timeseries anomaly detection using temporal hierarchical one-class network. Advances in Neural Information Processing Systems, 33, 13016–13026.

Thakkar, A., & Lohiya, R. (2021). A review of the advancement in intrusion detection datasets. Procedia Computer Science, 167, 636–645. https://doi.org/10.1016/j.procs.2020.03.330

Tuli, S., Casale, G., & Jennings, N. R. (2022). TranAD: Deep transformer networks for anomaly detection in multivariate time series data. Proceedings of the VLDB Endowment, 15(6), 1201–1214. https://doi.org/10.14778/3514061.3514067

Verizon Communications. (2023). 2023 data breach investigations report. Verizon Business. https://www.verizon.com/business/resources/reports/dbir

Zhang, Y., Chen, Y., Wang, J., & Pan, Z. (2021). Unsupervised deep anomaly detection for multi-sensor time-series signals. IEEE Transactions on Knowledge and Data Engineering, 35(2), 2118–2132. https://doi.org/10.1109/TKDE.2021.3107133