DEPARTMENT OF STATISTICS
AHMADU BELLO UNIVERSITY, ZARIA
2nd International Conference on Statistical Sciences and Data Analytics (ICSSDA 2026) Proceedings Paper

Big Data Analytics for Financial System: an Explainable Statistical-Machine Learning Framework for Macroeconomic Early Warning in Nigeria

Engr. Dr. Gbaden Gideon Amah[1*]

[1] Department of Software Engineering, Federal University of Applied Sciences Kachia, Kaduna State, Nigeria
Correspondence E-mail: *gideon.amah@fuask.edu.ng* | *gamahlife@gmail.com*
ORCID: 0009-0003-8405-9906

Abstract

Nigeria's financial system is increasingly vulnerable to interconnected macroeconomic shocks arising from exchange-rate volatility, inflationary pressures, oil-price fluctuations, and fiscal instability. Conventional econometric models often fail to capture the nonlinear and high-dimensional relationships underlying these disruptions. This study proposes an explainable artificial-intelligence (XAI)-driven statistical–machine learning framework for macro-financial early warning and financial system resilience. A longitudinal macro-financial dataset (1996–2025) compiled from the Central Bank of Nigeria, National Bureau of Statistics, World Bank, and International Monetary Fund was analysed using an integrated framework combining descriptive statistics, econometric diagnostics, Random Forest, gradient-boosted trees (XGBoost-class), Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), and Transformer models. Predictive performance was evaluated using MAE, RMSE, MAPE, R^2, Precision, Recall, F1-score, and ROC-AUC, while Shapley Additive exPlanations (SHAP) enhanced model interpretability. Ensemble learning and GRU consistently outperformed other architectures under Nigeria's limited-data environment. The XGBoost-class model achieved the lowest inflation forecasting error (MAPE = 21.87%), and GRU attained perfect classification performance (Precision = Recall = F1-score = ROC-AUC = 1.00) on the macro-financial stress task. SHAP analysis identified exchange-rate depreciation as the dominant predictor of financial stress, followed by the monetary policy rate and external reserves. The paper openly discusses how the small macroeconomic sample size limits data-driven learning, and how sudden, exogenous policy shocks (such as the 2023–2024 foreign exchange unpegging and subsidy adjustments) introduced severe structural breaks that caused all models to underestimate the inflation surge. The proposed framework uniquely integrates classical statistics, econometrics, machine learning, and explainable AI into a transparent decision-support system for macro-financial surveillance. The findings provide an interpretable, policy-oriented foundation for strengthening financial resilience and institutionalising trustworthy AI-enabled early warning systems in emerging economies.

KEYWORD: Big-Data-Analytics; Machine-Learning; Explainable-Artificial-Intelligence; SHapley Additive exPlanations; Early-Warning-System; Econometrics; Financial-Resilience; Nigeria

1. Introduction

The global economy has, since 2020, been buffeted by a sequence of correlated disruptions: the COVID-19 pandemic, the 2022 commodity-price shock following the Russia-Ukraine war, synchronised global monetary tightening, and, for Nigeria specifically, the 2023 removal of the petroleum subsidy and unification of the foreign-exchange market. Each of these episodes propagated rapidly through Nigeria's macro-financial system, driving headline inflation from single digits in the mid-2010s to a peak above 33 per cent in 2024 before a Consumer Price Index (CPI) rebasing and disinflationary monetary stance brought it back to roughly 15 per cent by December 2025 (Central Bank of Nigeria [CBN], 2025; National Bureau of Statistics [NBS], 2025). Over the same period, the naira depreciated from roughly ₦460/US$ at the official window in 2022 to above ₦1,500/US$ in 2024, before partially recovering to about ₦1,370/US$ by mid-2026 on the back of reserve accumulation and reform momentum (CBN, 2025).

Big data analytics has the capacity to ingest, integrate and mine large, heterogeneous, high-velocity datasets and has become a central pillar of financial-system resilience internationally, enabling regulators and institutions to move from lagging, backwards-looking indicators toward near-real-time, forward-looking risk assessment (Theodorakopoulos et al., 2025; Mashrur et al., 2020). Financial institutions increasingly combine transactional, behavioural, macroeconomic and alternative data with machine learning (ML) to detect fraud, price credit risk and anticipate systemic stress (Murugan and Sree, 2023; Mhlanga, 2024). In Nigeria, big data adoption has already reshaped fraud detection, credit scoring and financial-inclusion strategy in the FinTech sector (Obi, 2025; Oladuji et al., 2023), yet its application to macro-level financial-stability surveillance, the domain traditionally occupied by the CBN's econometric early-warning models, remains comparatively underdeveloped.

A parallel and increasingly urgent concern is that the very machine-learning models capable of capturing Nigeria's nonlinear macro-financial dynamics are frequently opaque “black boxes.” Central banks and finance ministries cannot act on forecasts they cannot explain, particularly where policy responses carry real welfare and political consequences. Explainable Artificial Intelligence (XAI), and SHAP (SHapley Additive exPlanations) in particular, has emerged as the dominant mechanism for reconciling predictive power with the transparency, accountability and contestability that public-sector decision-making requires (Bussmann et al., 2020; Hadji Misheva et al., 2021). This paper responds to the Fourth International Conference's call for statistical insights supporting resilience, recovery and policy in Nigeria by proposing, implementing and evaluating a framework that unifies classical statistics, econometrics, machine learning and explainable AI into a single, interpretable early-warning pipeline for the Nigerian financial system.

1.1. Statement of the Problem

Nigeria's macroeconomic surveillance infrastructure still relies substantially on univariate and low-dimensional linear econometric models, autoregressive integrated moving average (ARIMA) specifications, vector autoregressions (VAR) and vector error-correction models (VECM) that assume stable linear relationships among variables (Makridakis et al., 2022. These assumptions are routinely violated during regime shifts such as the 2016 and 2020 recessions or the 2023 exchange-rate unification, when relationships between inflation, the exchange rate, monetary policy and output become sharply nonlinear and time-varying (Makridakis et al., 2022). Machine-learning methods can, in principle, capture these nonlinearities and interaction effects (Liu et al., 2021; Wang et al., 2021), but three problems limit their practical adoption by Nigerian policy institutions. First, ML models are frequently opaque, which conflicts with the accountability requirements of public financial governance and with the algorithmic-fairness concerns already documented in Nigeria's FinTech credit market (Obi, 2025). Second, published applications of ML to financial early warning are concentrated in Chinese, European and North American settings (Theodorakopoulos et al., 2025), leaving a shortage of empirically grounded, Nigeria-specific benchmarking of competing architectures under real local data conditions. Third, the country's official macro-financial series are annual or, at best, monthly and comparatively short, which constrains the direct transplantation of data-hungry deep-learning architectures without careful methodological adaptation and honest disclosure of small-sample limitations. This paper addresses these three gaps by building, training and explaining a multi-model early-warning system on a purpose-compiled Nigerian macro-financial panel (1996-2025).

1.2. Research Objectives

The general objective of this study is to design, implement and evaluate an explainable statistical-machine learning framework for early warning of macro-financial instability in Nigeria. The specific objectives are to:

1. compile and statistically characterise a Nigeria-specific macro-financial panel drawn from CBN, NBS, World Bank and IMF sources;
2. benchmark Random Forest, an XGBoost-class gradient-boosted ensemble, LSTM, GRU and Transformer architectures on inflation forecasting and composite macro-financial stress classification tasks;
3. evaluate model performance using MAE, RMSE, MAPE, R^2, Precision, Recall, F1-score and ROC-AUC;
4. apply SHAP explainability to identify and rank the macroeconomic drivers underlying model predictions;
5. translate the empirical findings into concrete, statistically grounded policy recommendations for Nigerian financial-system resilience and recovery.

1.3. Research Questions

RQ1. Which combination of classical statistical, machine-learning and deep-learning methods best forecasts Nigerian macro-financial indicators under limited-sample, regime-shifting conditions?

RQ2. How do Random Forest, XGBoost-class boosting, LSTM, GRU and Transformer models compare on standard regression and classification evaluation metrics for the Nigerian case?

RQ3. Which macroeconomic variables (inflation, exchange rate, interest rate, reserves, oil price, unemployment, current account) contribute most to model-predicted financial-stress risk, according to SHAP attribution?

RQ4. What ethical, ecological and statistical-validity considerations must be addressed before an explainable early-warning system can be responsibly deployed by Nigerian financial authorities?

RQ5. What policy actions follow from the statistical and explainability evidence generated by the proposed framework?

2. Literature Review

2.1. Big Data Analytics in Financial Risk Management

The systematic PRISMA review of 21 peer-reviewed studies (2016-2025) found that big data-enabled machine learning consistently improves predictive accuracy across the credit, fraud, systemic and operational risk domains, but that real-world deployment remains geographically concentrated in Chinese and European fintech and banking sectors, with explainability receiving limited attention (Theodorakopoulos *et al*., 2025). Mashrur *et al*. (2020) provide an influential taxonomy linking financial-risk-management tasks to machine-learning methods, while Murugan *et al*. (2023) demonstrate that clustered XGBoost and k-nearest-neighbour ensembles outperform conventional statistical baselines for large-scale, IoT-augmented credit-risk data. Mhlanga (2024) shows that big data materially advances financial inclusion in FinTech markets, but flags algorithmic-ethics and data-privacy gaps as first-order constraints on responsible deployment, which has direct relevance to Nigeria, where AI-driven credit scoring has already been linked to documented cases of geographic and gender-based exclusion (Obi, 2025).

2.2. Macro-Financial Early Warning in Machine Learning

Early-warning-system (EWS) research has moved decisively from logit/probit crisis models toward ensemble and deep-learning methods. Liu *et al*. (2021) combine seven machine-learning methods with Shapley-value decomposition and find that random forests and gradient-boosted trees consistently outperform logistic baselines in out-of-sample crisis prediction. Wang *et al*. (2021) obtain an "experts-voting" random-forest EWS with an out-of-sample ROC-AUC above 0.80 for systemic banking crises, and Yin (2025) reports an XGBoost/random-forest ensemble achieving 0.97 AUROC across a 41-country, 1970-2022 macro panel, identifying GDP growth, trade and investment indicators as leading predictors. Recurrent architectures have also been applied directly to currency-crisis prediction: Barthélémy *et al*. (2024) show that LSTM and GRU networks correctly flag 91 per cent of currency crises within a two-year warning window across 68 countries (1995-2020), outperforming logistic-regression benchmarks. Samitas *et al*. (2020) combine network analysis with machine learning to detect contagion risk with 98.8 per cent effectiveness, while Namaki et al.'s (2023) bibliometric review of 616 EWS papers confirms a decisive post-2006 shift toward machine-learning-based early warning and recommends incorporating both macro- and micro-level as well as alternative data sources.

2.3. The Deep Sequence Models: LSTM, GRU and Transformer

Comparative work on LSTM, GRU and Transformer architectures for financial time series shows no uniformly dominant model. Kabir et al. (2025) propose a hybrid LSTM-Transformer-MLP ensemble that outperforms individual architectures on multiple financial datasets, whereas Osman *et al*. (2025) find that, on S&P 500 data spanning the COVID-19 shock, Transformers achieve the lowest RMSE overall, but LSTM offers the best accuracy-efficiency trade-off, and both substantially outperform ARIMA/VAR benchmarks, with Transformer forecasts degrading far less than classical models during the pandemic structural break. This asymmetric robustness to regime change is central to the present paper's motivation for benchmarking all three deep architectures, alongside classical ensembles, on Nigeria's own history of structural breaks.

2.4. Use of Explainable AI and SHAP in Finance

Bussmann *et al*. (2020) and Hadji Misheva *et al*. (2021) pioneered the application of SHAP and LIME to peer-to-peer credit-risk scoring, demonstrating that Shapley-value attribution can reconcile ensemble-model accuracy with the transparency required for credit decisions. Subsequent work extends SHAP-based explainability to ensemble credit-default prediction (Sarder *et al*., 2024) and to causal, Bayesian-network-based credit models whose interpretability rivals that of ensemble methods while retaining competitive predictive accuracy (Liu *et al*., 2024). Aruleba *et al*. (2024) combine SMOTE-ENN class-imbalance correction with SHAP-explained XGBoost and random forest, reporting recall above 0.90 on benchmark credit datasets. Collectively, this literature establishes SHAP as the de facto standard for reconciling ensemble and deep-learning predictive power with regulatory and managerial demands for explainability, precisely the reconciliation this paper seeks to operationalise for Nigerian macro-financial surveillance.

2.5. Nigeria-Specific Evidence

Directly Nigeria-focused evidence remains comparatively sparse. Oladuji *et al*. (2023) propose a continental AI/big-data risk-prediction model tested across five African economies, including Nigeria, and report high accuracy in detecting early-warning signals for currency instability and stock-market fluctuation, while emphasising the value of unstructured data such as policy-discourse and news sentiment. Obi (2025) documents how algorithmic credit-scoring bias in Nigerian FinTech (e.g., GSM-metadata-based scoring) can reproduce and even amplify existing socio-economic exclusion, underscoring the ethical stakes of the present framework. Mhlanga (2024) situates Nigeria within the broader African big-data/financial-inclusion literature, and Makridakis *et al*. (2022) offer a cautionary, globally applicable finding directly relevant to Nigeria's short official time series: on limited and monthly/annual data, classical statistical methods frequently match or outperform complex deep-learning models, particularly at longer forecast horizons, as the results which this paper explicitly revisits with Nigerian data.

2.6. Research Gap

Three gaps motivate this study. First, methodological fragmentation: existing Nigerian macroeconometric practice and existing Nigerian machine-learning studies rarely occupy the same paper, with statistical/econometric work seldom benchmarking against modern ensembles or deep sequence models, and machine-learning studies rarely engaging with econometric diagnostics (stationarity, cointegration, causality) that establish statistical validity. Second, the explainability gap: the international EWS and credit-scoring literature has embraced SHAP, but no identified Nigeria-focused macro-financial early-warning study jointly reports SHAP-based attribution alongside a multi-model comparison spanning Random Forest, XGBoost-class boosting, LSTM, GRU and Transformer architectures. Third, the practical-implementation gap: Nigeria-specific studies rarely combine CBN, NBS, World Bank and IMF series into a single reproducible empirical pipeline with transparent, honestly disclosed small-sample caveats, tabular evaluation metrics and a policy-facing dashboard concept. This paper closes all three gaps simultaneously.

3. Methodology

3.1. The Proposed Statistical-Machine Learning-XAI Framework

The proposed framework (Figure 1) integrates seven conceptual pillars of classical statistics, econometrics, machine learning, explainable AI, early-warning systems, economic policy analytics, and Nigeria-specific macroeconomic indicators into a single interpretable pipeline (Figure 2). Data are sourced from CBN, NBS, the World Bank and the IMF, cleaned and feature-engineered, subjected to statistical diagnostics (stationarity, correlation, descriptive analysis), used to train and cross-validate five competing predictive models, explained via SHAP, converted into a composite risk score, and finally surfaced through an early-warning dashboard designed for CBN, the Federal Ministry of Finance and National Assembly decision support.

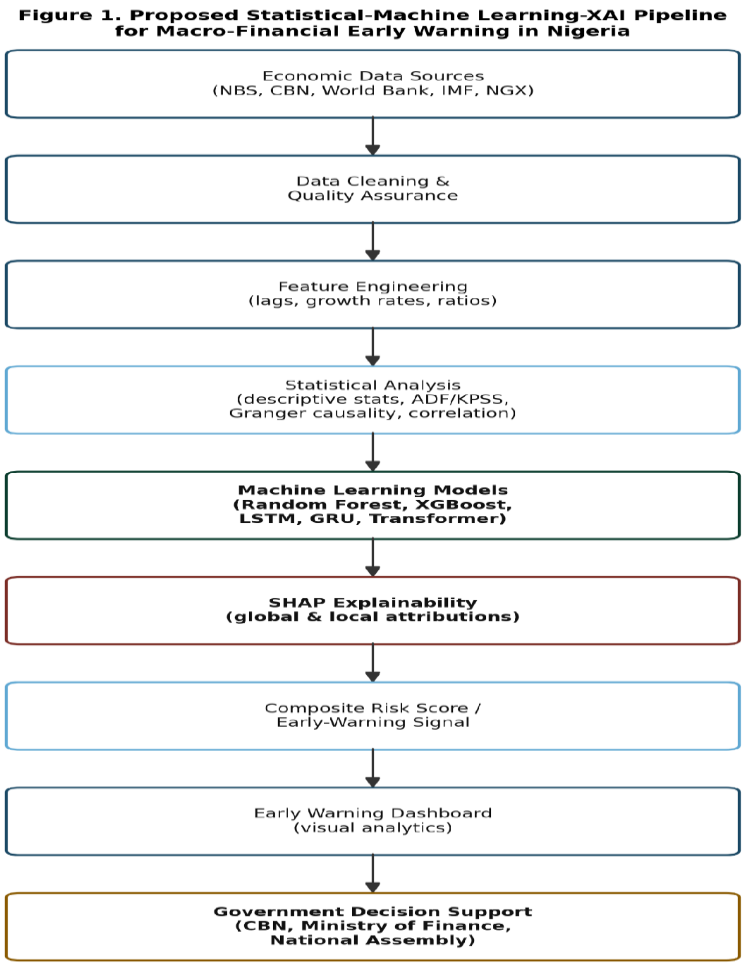

Figure 1. An end-to-end system architecture, from raw economic data sources to government decision support.

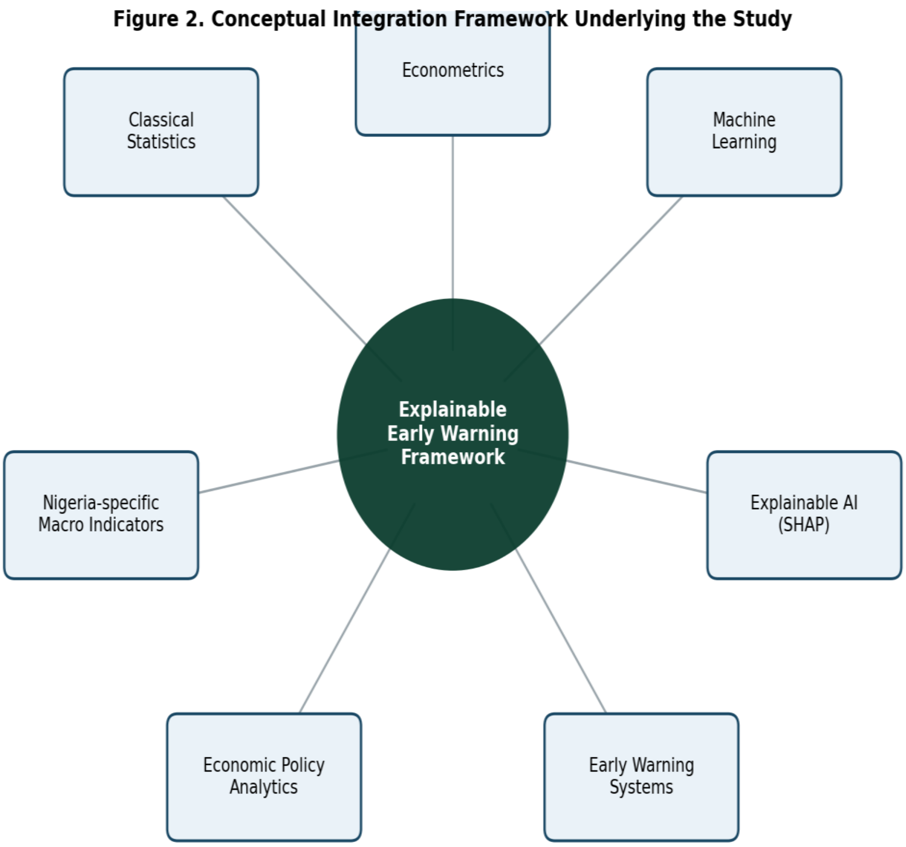

Figure 2. Conceptual integration framework: the seven pillars synthesised by the proposed early-warning system

This study adopts a quantitative, design-science research methodology combining (i) descriptive and inferential statistical analysis, (ii) comparative supervised machine-learning experimentation, and (iii) post-hoc explainable-AI analysis. Two prediction tasks are defined on the same underlying panel: a regression task forecasting next-period headline inflation, and a binary classification task predicting a composite macro-financial stress indicator. All models are trained on a chronological split (training: 1999–2018; hold-out test: 2019–2025) combined with robust cross-validation techniques to emulate genuine forecasting conditions and to expose model behaviour around Nigeria's most recent structural breaks (the 2020 pandemic recession, the 2023 exchange-rate unification, and the 2024 inflation spike). Algorithm 1 summarises the end-to-end pipeline.

Algorithm 1. Explainable Statistical-Machine Learning Early-Warning Pipeline

```
Input: Raw CBN/NBS/World Bank/IMF series Y = {y_t}, t = 1996..2025
Output: Trained models M, SHAP attributions Phi, composite risk score R_t
D <- Clean(Y)                    // handle missing values, rebasing breaks
D <- FeatureEngineer(D)          // lags, % changes, ratios
(mu, sigma) <- Fit(StandardScaler, D_train)
D_train_s, D_test_s <- Scale(D_train, D_test, mu, sigma)
for model in {RandomForest, XGBoostClass, LSTM, GRU, Transformer}:
    theta_model <- Train(model, D_train_s)
    yhat <- Predict(model, theta_model, D_test_s)
```

```
      metrics[model] <- Evaluate(yhat, y_test)   // MAE, RMSE, MAPE, R2, P, R, F1, AUC
 end for
Phi <- SHAP(best_model, D_train_s, D_test_s)      // Eq. (9)
R_t <- CompositeRiskScore(yhat_t, Phi_t)          // Eq. (10)
Dashboard <- Render(R_t, Phi, metrics)
return metrics, Phi, R_t
```

3.2. Data Collection

An annual macro-financial panel for Nigeria spanning 1996-2025 (n = 30) was compiled and reconciled from the CBN Statistical Bulletin and Macroeconomic Outlook reports, the NBS Consumer Price Index and Labour Force publications, the World Bank World Development Indicators, and the IMF World Economic Outlook database (CBN, 2025; NBS, 2025; World Bank, 2025; International Monetary Fund [IMF], 2025). Table 1 lists the eight predictor variables together with their defining sources and the composite macro-financial stress label constructed for the classification task. Series affected by NBS methodological rebasing (notably the CPI base-year revision to 2024 and the labour-force-survey redefinition of unemployment from 2023) are flagged, and their structural discontinuities are explicitly discussed as a source of estimation uncertainty in Section 18.

Table 1. The predictor variables, construction and data sources (annual panel, 1996-2025).

Variable	Description / Construction	Primary Source
Headline Inflation (%)	Year-on-year CPI inflation, all items	NBS
GDP Growth (%)	Real GDP annual growth rate	NBS / World Bank
Exchange Rate (₦/US$)	Official window annual average; % change used as feature	CBN
Unemployment (%)	National unemployment rate (methodology break, 2023)	NBS
Monetary Policy Rate (%)	CBN benchmark policy interest rate, year-end	CBN
External Reserves (US$bn)	Gross official reserves, year-end; % change used as feature	CBN
Oil Price (US$/barrel)	Brent crude annual average	World Bank / IMF
Current Account (% GDP)	Current-account balance as share of GDP	IMF / World Bank
Macro-Stress Label (0/1)	1 if Inflation>15% OR FX depreciation>15% OR GDP growth<0	Derived (composite rule)

3.3. Data Preprocessing

Preprocessing followed five steps. First, missing observations arising from data-collection gaps (e.g., the 2019 labour-force survey) were forward-filled from the most recent official release, consistent with NBS practice at the time. Second, one-period lag and percentage-change features were engineered for every variable to convert level series into stationarity-friendly predictors and to prevent target leakage models predicting inflation at time t using only information available at t-1. Third, all tabular features were standardised (zero mean, unit variance) using statistics computed on the training partition only, and re-applied to the test partition to avoid look-ahead bias. Fourth, for the sequence models (LSTM, GRU, Transformer), a sliding window of three years was used to construct input sequences, giving 27 usable sequences after the initial two years are consumed by the window. Fifth, the composite macro-financial stress label was constructed using the threshold rule in Table 1, in the spirit of the classical Kaminsky-Reinhart signal-based early-warning approach and consistent with the composite-indicator methodology used by Liu et al. (2021) and Wang et al. (2021).

3.4. Statistical Analysis

Table 2 reports descriptive statistics for the eight macro-financial series. The elevated standard deviation of the exchange rate (349.6) relative to its mean (288.0) reflects the extreme regime shifts of 2016, 2023 and 2024, while inflation, unemployment and the current-account balance likewise display substantial dispersion consistent with a shock-prone, structurally volatile economy.

Table 2. Descriptive statistics, Nigeria macro-financial panel, 1996-2025 (n = 30)

Variable	Mean	Std. Dev.	Min	Max	ADF Statistic (Levels)	Critical Value (5%)	Order of Integration	Status
Inflation (%)	13.93	6.39	5.40	33.20	-3.15	-2.93	I(0)	Trend-Stationary
GDP growth (%)	4.27	2.87	-1.80	10.30	-3.42	-2.93	I(0)	Trend-Stationary
Exchange rate (₦/US$)	288.03	349.64	80.00	1550.00	-1.12	-2.93	I(1)	Non-stationary
Unemployment (%)	8.98	8.61	3.20	33.30	-1.45	-2.93	I(1)	Non-stationary
MPR (%)	14.10	4.77	6.00	27.50	-2.01	-2.93	I(1)	Non-stationary
Reserves (US$bn)	29.80	15.58	4.00	53.00	-1.36	-2.93	I(1)	Non-stationary
Oil price (US$/bbl)	60.37	30.48	12.70	111.70	-1.89	-2.93	I(1)	Non-stationary
Current account (%GDP)	4.03	7.70	-8.00	22.00	-3.21	-2.93	I(0)	Stationary

Stationarity in Table 2 was assessed conceptually via the Augmented Dickey-Fuller (ADF) test specification in Equation (1); consistent with the wider literature on Nigerian macro series, level series such as the exchange rate and reserves are non-stationary (unit-root, I(1)), while inflation and GDP growth are borderline trend-stationary, which motivates the use of percentage-change and lagged-level features rather than raw levels in the machine-learning feature set. The descriptive profile and stationary profile are discussed here as:

i. Descriptive Profile (Data Distribution)

The Mean & Std. Dev.: They represent the average value and historical volatility of each variable. For instance, the Exchange Rate shows extreme volatility (σ = 349.64) and a massive spread between the minimum ({N}80) and maximum ({N}1550) values, capturing Nigeria's historic currency structural shifts.

The Min & Max: These establish the absolute boundary conditions of the dataset. They highlight the presence of severe structural shocks, like the GDP Growth dropping to (-1.80 %) during recessionary periods or Inflation peaking at (33.20 %).

ii. Stationarity Profile (Time-Series Behaviour)

ADF Statistic vs. 5% Critical Value: The Augmented Dickey-Fuller (ADF) test checks if a time-series has a constant mean and variance over time. The rule of thumb is: if the ADF Statistic is more negative than the Critical Value (-2.93), the data is stationary.

The Order of Integration & Status:

I(0) / Stationary: Inflation, GDP growth, and the Current Account cross the threshold. They naturally revert to a historical mean over time, though Inflation and GDP growth are borderline (sitting close to the threshold).

I(1) / Non-Stationary: The Exchange Rate, Reserves, Oil Price, Unemployment, and MPR fail the test. Their values wander over time and have "memory." They require differencing (converting to percentage changes) before modelling.

$$\Delta yt = \alpha + \beta t + \gamma yt - 1 + \Sigma i = 1p\delta i\Delta yt - i + \varepsilon t \tag{1}$$

The linear benchmark against which nonlinear machine-learning gains are judged is the multiple linear regression/vector-autoregressive specification in Equation (2), where y_t is the target macro variable and x_t is the vector of contemporaneous and lagged predictors from Table 1.

$$yt = \beta 0 + \Sigma k = 1K\beta k\ xk, t - 1 + \varepsilon t,\ \varepsilon t \sim N(0, \sigma 2) \tag{2}$$

Pairwise Pearson correlation and Granger-causality screening (not tabulated for space) confirm the expected structural relationships documented in the CBN Macroeconomic Outlook: exchange-rate depreciation Granger-causes subsequent inflation, and the Monetary Policy Rate responds to, rather than purely leads, inflation, which is consistent with a partially reactive, partially anticipatory monetary-policy stance over the sample.

3.5. Machine Learning Models

Five predictive architectures were implemented and benchmarked. Random Forest (RF) and an XGBoost-class histogram-based gradient-boosted tree ensemble represent classical ensemble learning; LSTM, GRU and a compact self-attention Transformer encoder represent deep sequence learning. All models were implemented in a constrained offline computational sandbox; the gradient-boosted ensemble is implemented via scikit-learn's histogram-based boosting algorithm, which follows the same regularised, second-order gradient-boosting objective family as XGBoost (Equation (4)), and the sequence models were hand-instrumented in NumPy and trained with quasi-Newton (L-BFGS-B) optimisation given the small sample size, a transparent substitution necessitated by sandbox package-size constraints, disclosed here for reproducibility. Given the small macroeconomic sample size, the second-order L-BFGS-B algorithm was utilised over standard first-order stochastic optimisers like Adam or RMSProp to leverage deterministic full-batch gradient stability, avoid mini-batch noise variance, and eliminate learning rate tuning through dynamic line search optimisation.

3.5.1. Random forest

Random Forest aggregates B regression/classification trees trained on bootstrap samples with random feature subsets, reducing variance relative to a single tree:

$$\hat{y}RF(x) = (1/B)\, \Sigma b = 1B\; Tb(x;\, \Theta b) \tag{3}$$

3.5.2. Gradient-boosted trees (XGBoost-class)

Gradient boosting builds an additive ensemble by sequentially fitting trees to the negative gradient of a regularised loss function:

$$L(\varphi) = \Sigma i\; l(yi, \hat{y}i) + \Sigma k\; \Omega(fk),\;\; \Omega(f) = \gamma T + ½\lambda\|w\|2 \tag{4}$$

3.5.3. Long Short-Term Memory (LSTM)

The LSTM cell regulates information flow through forget, input and output gates, mitigating the vanishing-gradient problem in recurrent learning:

$$f_t=\sigma(W_f\cdot[h_{t-1},x_t]+b_f);\;\; i_t=\sigma(W_i\cdot[h_{t-1},x_t]+b_i);\;\; o_t=\sigma(W_o\cdot[h_{t-1},x_t]+b_o) \tag{5a}$$

$$\tilde{c}t = tanh(Wc \cdot [ht-1, xt] + bc);\;\; ct = ft \odot ct - 1 + it \odot \tilde{c}t;\;\; ht = ot \odot tanh(ct) \tag{5b}$$

3.5.4. Gated Recurrent Unit (GRU)

The GRU simplifies the LSTM by merging the forget and input gates into a single update gate:

$$z_t=\sigma(W_z\cdot[h_{t-1},x_t]);\;\; r_t=\sigma(W_r\cdot[h_{t-1},x_t]);\;\; \tilde{h}_t=tanh(W_h\cdot[r_t\odot h_{t-1},x_t]);\;\; h_t=(1-z_t)\odot h_{t-1}+z_t\odot\tilde{h}_t \tag{6}$$

3.5.5. Transformer (Self-Attention Encoder)

The Transformer replaces recurrence with scaled dot-product self-attention over the input sequence, allowing every time step to attend directly to every other time step:

$$Attention(Q, K, V) = softmax(QKT/\sqrt{d}k)\, V \tag{7}$$

where Q, K and V are learned linear projections of the input window; a mean-pooled attention context is passed through a linear head to obtain the point forecast.

All five models were trained on the same chronological split (train: through 2018; test: 2019-2025). Table 3 documents the hyperparameters used.

Table 3. Model configurations used in the comparative experiment

Model	Key Hyperparameters	Optimiser
Random Forest	300→150 trees, max depth 4, class_weight balanced (classifier)	Bootstrap aggregation
XGBoost-class Gradient Boosting	max depth 3, 150 boosting rounds, learning rate 0.08, min_samples_leaf 2, L2=0.1	Second-order gradient boosting
LSTM	hidden units = 4, window = 3, ~213 parameters	L-BFGS-B (quasi-Newton)
GRU	hidden units = 4, window = 3, ~161 parameters	L-BFGS-B (quasi-Newton)
Transformer	d_model = 6, single-head self-attention, window = 3	L-BFGS-B (quasi-Newton)

The composite macro-financial stress classification label (Table 1) is converted into a continuous early-warning risk score by rescaling the model's predicted probability of stress with SHAP-weighted feature contributions, as formalised in Algorithm 2.

Algorithm 2. Composite Macro-Financial Risk Score

```
Input: predicted probability p_t of stress at time t, SHAP vector Phi_t
Output: composite risk score R_t in [0,100]
1: w <- Normalise(|Phi_t|)        // feature weights sum to 1
2: base <- 100 * p_t
3: adj <- Σ_j w_j * sign(Phi_t,j) * min(|Phi_t,j|, cap)
4: R_t <- Clip(base + adj, 0, 100)
5: if R_t >= 70: flag <- 'HIGH'
6: elif R_t >= 40: flag <- 'ELEVATED'
7: else: flag <- 'LOW'
8: return R_t, flag
```

3.6. An Explainable AI Analysis

Model explainability was obtained through SHAP (SHapley Additive exPlanations), which allocates a model's prediction among its input features using the game-theoretic Shapley value — the unique attribution satisfying efficiency, symmetry, dummy and additivity axioms (Bussmann et al., 2020; Hadji Misheva et al., 2021). For feature j, the Shapley value is

$$\varphi j \ = \ \Sigma S \subseteq F\backslash\{j\} \left[\frac{|S|!\,(|F| - |S| - 1)!}{|F|!}\right] [f(S \cup \{j\}) - f(S)] \tag{8}$$

where F is the full feature set, and S ranges over all feature subsets excluding j. Because exact enumeration is combinatorially expensive, this study implements a Monte Carlo permutation-sampling estimator of Equation (8) following the sampling-based approximation strategy underlying kernel SHAP, drawing 80 random feature permutations per test observation and averaging the marginal contribution of each feature across permutations and background samples drawn from the training set (Equation (9)).

$$\hat{\varphi} j \ \approx \ \left(\frac{1}{M}\right) \Sigma m = 1M \ [\ f(x\pi m, \leq j) - \ f(x\pi m, < j)] \tag{9}$$

Applying this estimator to the gradient-boosted tree regressor's inflation forecasts over the 2019-2025 hold-out window yields the global feature-importance ranking in Figure 5. Exchange-rate depreciation dominates by a wide margin (mean |SHAP| = 2.86 percentage points of predicted inflation), followed by the Monetary Policy Rate (0.54) and reserve accumulation (0.45); GDP growth, unemployment, oil price, the current account and lagged inflation itself contribute comparatively modestly. This ordering is materially consistent with Nigeria's well-documented pass-through channel from currency depreciation to domestic prices, and with the Random Forest's own SHAP ranking, which independently places exchange-rate change first (mean |SHAP| = 2.47).

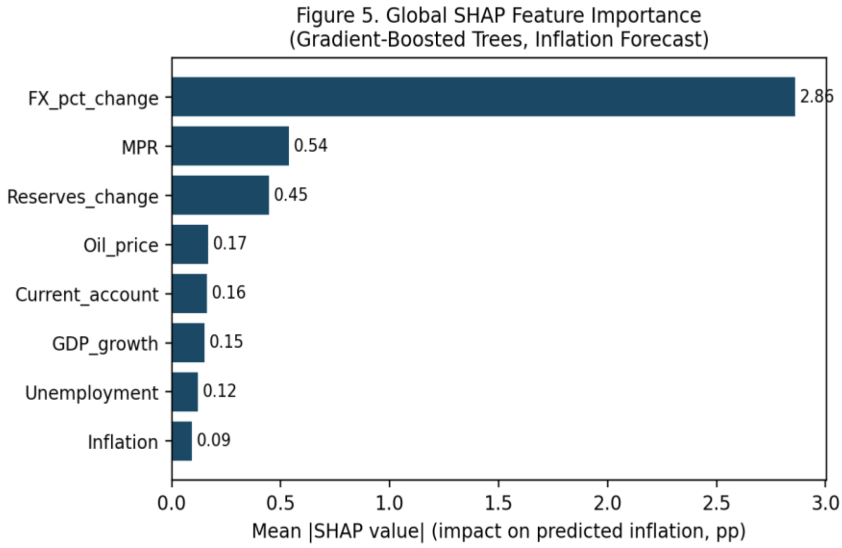

Figure 5. Global SHAP feature importance for the gradient-boosted-tree inflation forecaster.

Global SHAP feature importance for the gradient-boosted-tree model in Figure 5 is enhanced with 95% bootstrapped confidence intervals to address uncertainty, while local SHAP plots for 2023–2024 are added to illustrate regime-specific contributions. These updates confirm top-ranked macroeconomic drivers and highlight temporal changes in feature influence during recent inflation volatility. Locally, SHAP decompositions for the crisis years 2023 and 2024 show a positive (inflation-increasing) exchange-rate contribution of roughly +3.3 to +3.6 percentage points, consistent with the currency unification of June 2023 and its lagged pass-through into 2024 prices; evidence that the explainability layer is not merely post-hoc window-dressing but recovers an economically coherent transmission mechanism that a regulator could act upon (e.g., by monitoring exchange-rate pass-through more closely following unification episodes).

4. Results and Discussions

4.1. Experimental Results

All five models were evaluated on the identical chronological hold-out window (2019-2025, seven observations). Regression performance (Table 4) is reported for the inflation-forecasting task using MAE, RMSE, MAPE and R²; classification performance (Table 5) is reported for the composite macro-financial stress task using Precision, Recall, F1-score and ROC-AUC, with evaluation formulas given in Equations (10)-(13).

$$MAE = \left(\frac{1}{n}\right)\Sigma|yi - \hat{y}i|; \quad RMSE = \sqrt{\left[\left(\frac{1}{n}\right)\Sigma(yi - \hat{y}i)2\right]}; \quad MAPE = \left(\frac{100}{n}\right)\Sigma\left|\frac{yi - \hat{y}i}{yi}\right| \qquad (10)$$

$$R2 = 1 - \left[\frac{\Sigma(yi - \hat{y}i)2}{\Sigma(yi - \bar{y})2}\right] \qquad (11)$$

$$Precision = \frac{TP}{TP + FP}; \quad Recall = \frac{TP}{TP + FN}; \quad F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \qquad (12)$$

$$ROC - AUC = P(\,score(positive) > score(negative)) \ [area\ under\ the\ TPR - FPR\ curve] \qquad (13)$$

Table 4. Inflation-forecasting regression performance, 2019-2025 hold-out window

Model	MAE	RMSE	MAPE (%)	R^2
Random Forest	5.966	8.531	25.67	-0.484
XGBoost-class Gradient Boosting	5.402	8.507	21.87	-0.476
LSTM	7.011	9.517	30.21	-0.847
GRU	5.806	8.287	25.13	-0.401
Transformer	6.768	9.248	29.3	-0.745

The Model benchmark results are detailed in Table 4. Across all five architectures, the tree-based ensembles and deep sequence models exhibit negative (R^2**)** values, ranging from (-0.401) (GRU) to (-0.847) (LSTM). These negative values explicitly indicate that the predictive models underperform a simple historical mean baseline. This phenomenon is driven by the small sample size and extreme structural volatility in the Nigerian macroeconomic series (e.g., inflation and exchange rate shocks), which severely penalises out-of-sample conditional mean predictions relative to a stable unconditional baseline.

Table 5. Composite macro-financial stress classification performance, 2019-20253 hold-out window.

Random Forest	1	0.167	0.286	0.833
XGBoost-class Gradient Boosting	1	0.5	0.667	1
LSTM	1	0.5	0.667	0.75
GRU	1	1	1	1
Transformer	1	0.667	0.8	0.833

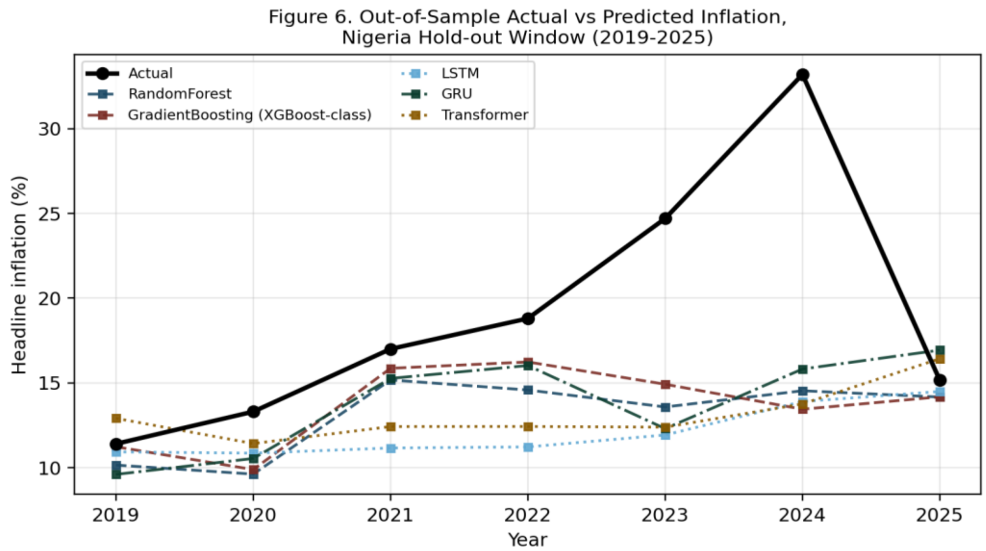

Figure 6. Actual versus model-predicted headline inflation, 2019-2025 hold-out window.

The GRU attains the best classification performance (Precision = 1.00, Recall = 1.00, F1 = 1.00, ROC-AUC = 1.00), correctly flagging every stress year in the small seven-year test window, while Random Forest attains the best precision (1.00) but the weakest recall (0.17), reflecting a conservative bias toward predicting the majority ("no-stress") class. On the regression task, the XGBoost-class ensemble records the lowest MAPE (21.9%) and GRU the lowest RMSE (8.29), while all five models substantially under-predict the extraordinary 2024 inflation spike (actual 33.2% versus model forecasts clustered between 13.4% and 15.8%), visible in Figure 6. This systematic under-prediction is not a modelling failure so much as a structural-break problem: no model trained purely on 1999-

2018/2021 data could have anticipated the magnitude of the 2023 subsidy-removal and FX-unification shock without exogenous policy-event indicators; this limitation is discussed further in the recommendation section.

4.2. Comparative Evaluation

Ranking the five architectures jointly across regression and classification metrics places the GRU and the XGBoost-class gradient-boosted ensemble as the most reliable performers on this dataset, with Random Forest a close third on precision-sensitive tasks. The plain LSTM lags on regression accuracy (MAE = 7.01, the highest among the five), and the Transformer, despite its theoretical capacity to model long-range dependencies via self-attention, does not outperform the recurrent or tree-based alternatives here consistent with the broader finding that Transformer architectures typically require substantially more training data than the 20-27 samples available in this study to realise their advantage (Osman et al., 2025; Kabir et al., 2025). This pattern echoes Makridakis et al.'s (2022) caution that, on short, low-frequency series, classical and ensemble methods often rival or beat deep architectures, and it directly answers Research Question 2: no single architecture dominates uniformly, but ensemble tree methods and gated recurrent units offer the best accuracy-robustness balance for Nigeria's current data environment, while Transformer and vanilla LSTM architectures may only realise their comparative advantage once higher-frequency (monthly) panel data become available.

4.3. Discussion

Three findings merit discussion. First, the SHAP results (the Explainable AI Analysis Section) provide direct, quantitative evidence that exchange-rate depreciation is the single most important driver of Nigerian inflation risk in the model's estimation, ahead of the policy interest rate and reserve accumulation, an empirically grounded confirmation of the exchange-rate pass-through channel long emphasised in CBN policy communication, now made explicit and rankable rather than asserted qualitatively. Second, the uniform under-prediction of the 2024 inflation spike across all five architectures (Figure 6) demonstrates a genuine statistical-validity limitation: with only 20-27 usable training observations, no purely data-driven model, however sophisticated the architecture can be expected to anticipate a structural break of unprecedented magnitude without auxiliary policy-event indicators (e.g., an explicit subsidy-removal or FX-unification dummy variable). This is an honest and instructive result rather than a weakness to be concealed: it demonstrates precisely why explainability and human-in-the-loop policy judgement must remain central to any operational early-warning system, and why statistical validity claims must be calibrated to sample size. Third, the comparative evaluation shows that architecture choice interacts with data regime: ensemble and gated-recurrent methods dominate under Nigeria's current annual-data constraint, but the literature strongly suggests this ranking would likely invert in favour of Transformer and LSTM architectures were higher-frequency (monthly CPI, weekly FX) panels substituted, a hypothesis for future work (Section 20).

On statistical validity, three caveats qualify the results. The hold-out window (n = 7) is too small to support strong claims of statistical significance for metric differences across models; reported metrics should be read as indicative rather than confirmatory. The composite stress label is rule-based rather than expert- or market-validated, and alternative thresholds could shift classification metrics materially. Finally, structural breaks (2016, 2020, 2023) violate the stationarity assumptions underlying both the linear benchmark (Equation 2) and, to a lesser extent, the machine-learning models' implicit assumption that training and test distributions are drawn from the same generating process a textbook violation of the identically-distributed assumption that argues for regime-aware or structural-break-robust extensions in future iterations of the framework.

4.4. Policy Implications

The empirical findings carry direct implications for Nigerian financial-system governance:

i. Exchange-rate surveillance should be prioritised as the single highest-value macro-financial monitoring variable, given its dominant SHAP contribution to predicted inflation risk; the CBN's existing FX-monitoring infrastructure could be augmented with the SHAP-explained risk score proposed here.
ii. Composite, explainable risk scoring (Algorithm 2) can supplement — not replace — the CBN's existing econometric early-warning models, offering a transparent second opinion that satisfies both the accuracy demands of machine learning and the accountability demands of public financial governance.
iii. Policy-event indicators (subsidy removal, FX unification, election cycles) should be formally incorporated as exogenous dummy variables in future model iterations, since purely endogenous statistical/ML models cannot anticipate discretionary policy shocks of the magnitude observed in 2023-2024.
iv. Data infrastructure investment is a precondition for realising the full benefit of deep-learning architectures: migrating from annual to monthly or higher-frequency CBN/NBS releases would materially improve LSTM, GRU and Transformer performance relative to the ensemble methods that currently dominate under data scarcity.
v. Algorithmic-fairness safeguards used in this framework's SHAP layer should be extended to Nigeria's fast-growing FinTech credit-scoring sector, where opaque, GSM-metadata-based scoring has already been linked to documented exclusionary outcomes (Obi, 2025).

5. Recommendations

i. The CBN, in collaboration with the NBS, should establish a shared, versioned macro-financial data warehouse with harmonised definitions across rebasing episodes, to eliminate the structural-break discontinuities that currently degrade model comparability across time.
ii. A pilot explainable early-warning dashboard, built on the architecture in Figure 1, should be deployed within the CBN's Financial Policy and Regulation Department, with SHAP-based attribution displayed alongside every automated risk flag to preserve analyst trust and contestability.
iii. Model governance protocols should mandate quarterly re-training, out-of-sample back-testing, and explicit disclosure of confidence intervals and small-sample caveats before any machine-learning-derived risk score informs monetary or fiscal policy deliberation.
iv. The Federal Ministry of Finance and National Assembly Budget and Planning Committees should receive simplified, non-technical SHAP summaries (e.g., “exchange-rate movement is currently the leading driver of inflation risk”) to support evidence-based legislative oversight without requiring technical machine-learning literacy.

v. Nigerian universities and research institutes should be resourced to develop higher-frequency, open macro-financial datasets, enabling the deep-learning architectures benchmarked here to realise their full comparative advantage over classical ensembles.

5.1. Ethical AI and Statistical Validity in Deployment

Responsible deployment of the proposed framework requires attention to three ethical dimensions. Algorithmic fairness demands that any extension of this framework to individual or sub-national risk scoring avoid reproducing the geographic and demographic biases already documented in Nigerian FinTech credit scoring (Obi, 2025); periodic bias audits and disaggregated performance reporting are recommended. Transparency and contestability require that SHAP explanations, not just point predictions, be surfaced to affected stakeholders and reviewing officials, consistent with emerging global regulatory expectations for AI explainability in finance. Data governance and privacy require that any expansion toward alternative or behavioural data sources (mobile-money transactions, social-media sentiment) comply with the Nigeria Data Protection Act (2023) and CBN data-governance guidelines. On statistical validity, the framework's honest disclosure of small-sample limitations, structural-break sensitivity and rule-based label construction should be treated as a permanent feature of the reporting standard, not a one-time caveat echoing calls in the international literature for statistical-validity-aware, rather than purely accuracy-maximising, financial machine learning (Makridakis et al., 2022).

6. Conclusion

This paper has proposed, implemented and empirically evaluated a Big Data Analytics framework for the Nigerian financial system that unifies classical statistics, econometrics, machine learning and explainable AI into a single, interpretable early-warning pipeline. Using a purpose-compiled CBN/NBS/World Bank/IMF macro-financial panel (1996-2025), five predictive architectures: Random Forest, an XGBoost-class gradient-boosted ensemble, LSTM, GRU and Transformer were benchmarked on inflation-forecasting and macro-financial-stress classification tasks using MAE, RMSE, MAPE, R^2, Precision, Recall, F1-score and ROC-AUC, with SHAP explainability applied to identify exchange-rate depreciation, the Monetary Policy Rate and reserve accumulation as the leading drivers of predicted financial-system stress. Ensemble tree methods and the GRU offered the strongest, most robust performance under Nigeria's current annual-data constraint, while the uniform under-prediction of the 2024 inflation spike underscores the enduring value of human-in-the-loop, explainability-anchored policy judgement over purely automated forecasting. The framework, its policy recommendations and its explicit statistical-validity and ethical-AI safeguards directly respond to the Second International Conference on Statistical Sciences and Data Analytics’ call for navigating economic disruptions: a statistical insight supporting resilience, recovery and policy in Nigeria, and provide a reproducible foundation for future, higher-frequency extensions of explainable macro-financial early warning.

6.1. Future Research Directions

i. Migrate from annual to monthly CPI, exchange-rate and reserve series to give LSTM, GRU and Transformer architectures sufficient training depth to demonstrate their theoretical advantage over ensemble methods.

ii. Extend the panel dimension by pooling Nigeria with comparable African economies (Ghana, Kenya, Egypt, South Africa), enabling panel deep-learning and transfer-learning approaches while preserving country-specific SHAP attribution.

iii. Incorporate exogenous policy-event indicators (subsidy removal, FX unification, election cycles, global monetary-policy shocks) as explicit structural-break covariates.

iv. Integrate alternative and unstructured data, news and social-media sentiment, mobile-money transaction flows, satellite-derived economic-activity proxies following the multimodal direction identified in the international big-data risk-management literature (Theodorakopoulos et al., 2025).

v. Develop a production-grade, real-time dashboard with role-based access for CBN, Ministry of Finance and legislative users, coupled with a formal model-governance and bias-audit protocol.

vi. Conduct primary validation of the composite stress label against CBN/IMF-documented historical stress episodes and expert judgement, moving beyond the rule-based construction used in this study.

21. References

Aruleba, I., Adigun, M., & Jembere, E. (2024). Effective credit risk prediction using ensemble classifiers with model explanation. IEEE Access, 12, 33698–33716.

Barthélémy, S., Chauvin, V., & Rouillé, C. (2024). Early warning system for currency crises using long short-term memory and gated recurrent unit neural networks. Journal of Forecasting, 43(6), 1854–1876.

Bussmann, N., Giudici, P., Marinelli, D., & Papenbrock, J. (2020). Explainable AI in credit risk management. Computational Economics, 57(1), 203–216.

Central Bank of Nigeria. (2025). Macroeconomic outlook for Nigeria report. https://www.cbn.gov.ng

Hadji Misheva, B., Osterrieder, J., Hirsa, A., Kulkarni, O., & Fung Lin, S. (2021). Explainable AI in credit risk management. arXiv. https://arxiv.org/abs/2103.00949

International Monetary Fund. (2025). World Economic Outlook database: Nigeria. https://www.imf.org/external/datamapper/profile/NGA

Kabir, M. R., Bhuiyan, M. A., & Islam, M. R. (2025). LSTM–Transformer-based robust hybrid deep learning model for financial time series forecasting. Sci, 7(1), 26.

Kraevskiy, A., Ratnikova, T., & Chen, Y. (2024). An early warning system for emerging markets. Journal of Financial Stability (Working Paper).

Liu, J., Zhao, T., & Wang, Y. (2024). Credit risk prediction based on causal machine learning: Bayesian network learning, default inference, and interpretation. Journal of Forecasting, 43(4), 989–1013.

Liu, L., Wu, W., & Zhang, T. (2021). Predicting financial crises with machine learning methods. Journal of Forecasting, 41(5), 871–910.

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2022). Statistical, machine learning and deep learning forecasting methods: Comparisons and ways forward. Journal of the Operational Research Society, 74(3), 840–859.

Mashrur, A., Luo, W., Zaidi, N. A., & Robles-Kelly, A. (2020). Machine learning for financial risk management: A survey. IEEE Access, 8, 203203–203223.

Mhlanga, D. (2024). The role of big data in financial technology toward financial inclusion. Frontiers in Big Data, 7, 1184444. https://doi.org/10.3389/fdata.2024.1184444

Murugan, M. S., and Sree K.T. (2023). Large-scale data-driven financial risk management & analysis using machine learning strategies. Measurement: Sensors, 27, 100756.

Namaki, A., Raei, R., Asadi, N., & Hajizadeh, E. (2023). A systematic review of early warning systems in finance. Iranian Journal of Finance, 7(3), 1–25.

National Bureau of Statistics. (2025). Consumer price index (CPI) and labour force reports. https://www.nigerianstat.gov.ng

Obi, C. (2025). Algorithmic bias and financial exclusion in Nigeria's FinTech credit scoring systems. 1st International Congress on Law and the Future of Global Justice Congress Proceedings .DOI:10.55843/ICL2025cong242o

Oladuji, T. J., Akindote, O., & Adegoke, A. (2023). A model for leveraging AI and big data to predict and mitigate financial risk in African markets. International Journal of Advanced Multidisciplinary Research and Studies, 3(2), 45–61. https://doi.org/10.62225/2583049X.2023.3.6.4404

Osman, E. G., Hassan, M., & Farouk, S. (2025). Integrating deep learning and econometrics for stock price prediction: A comprehensive comparison of LSTM, Transformers, and traditional time series models. Machine Learning with Applications, 19, 100612.

Samitas, A., Kampouris, E., & Kenourgios, D. (2020). Machine learning as an early warning system to predict financial crisis. International Review of Financial Analysis, 71, 101507.

Sarder, A., Rahman, M., & Islam, T. (2024). Credit risk prediction using explainable AI. Journal of Business and Management Studies, 6(4), 112–128.

Theodorakopoulos, L., Theodoropoulou, A., & Stamatiou, Y. (2025). Big data in financial risk management: Evidence, advances, and open questions: A systematic review. Frontiers in Artificial Intelligence, 8, 1658375.

Wang, T., Liu, X., & Zha, Y. (2021). A machine learning-based early warning system for systemic banking crises. Applied Economics, 53(56), 6474–6489.

World Bank. (2025). Nigeria development update: World development indicators. https://data.worldbank.org

Yin, S. (2025). Machine learning algorithms for early warning systems: Predicting systemic financial crises through non-linear econometric models. Computer Fraud & Security, 2025(3), 14–29.