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

Machine Learning-Driven Early Warning System for Economic Forecasting: A Statistical Framework for Enhancing Economic Resilience and Evidence-Based Policy in Nigeria

Engr. Dr. Gbaden Gideon Amah[1*], Thomas Lass Barna[2], and Ishaq Isa[3]

[1,2,3]Department of Software Engineering, Federal University of Applied Sciences, Kachia, Kaduna State, Nigeria

*Correspondences E-mail: *gideon.amah@fuask.edu.ng* | *gamahlife@gmail.com*

ORCID: 0009-0003-8405-9906

Abstract

Nigeria's macroeconomy remains exposed to recurrent external and domestic shocks arising from oil price volatility, exchange-rate depreciation, inflationary pressure, and fiscal imbalance that classical econometric models often detect only after damage has occurred. This paper proposes and empirically illustrates a Machine Learning-Driven Early Warning System (ML-EWS) that fuses classical statistical inference, econometric diagnostics, and explainable machine learning to generate interpretable, forward-looking risk signals for Nigerian policymakers. The framework ingests macroeconomic indicators sourced from the Central Bank of Nigeria (CBN), the National Bureau of Statistics (NBS), the World Bank, and the International Monetary Fund (IMF), and passes them through a pipeline of data cleaning, feature engineering, stationarity and causality testing, ensemble and deep-learning forecasting (Random Forest, XGBoost, LSTM, GRU, and Transformer), and Shapley Additive Explanations (SHAP)-based interpretability, culminating in a composite Early Warning Risk Score visualised on a policy dashboard. Using real 2003–2024 World Bank indicators for Nigeria, an illustrative implementation shows Random Forest and XGBoost achieving statistically indistinguishable one-step-ahead GDP-growth forecasts (MAE = 2.22 and 2.24 percentage points; Diebold–Mariano statistic = 0.53, $p = .60$), while a Random Forest classifier detects historically documented economic-stress years (2016, 2017, 2020, 2021) with Precision = 1.00, Recall = 0.50, F1 = 0.67 and ROC-AUC = 0.92 under leave-one-out cross-validation. SHAP-consistent permutation importance identifies lagged GDP growth and its three-year trend as the dominant early-warning drivers. The paper reviews recent literature on financial-crisis prediction and inflation forecasting, discusses statistical validity and ethical AI considerations, and provides recommendations for institutionalising explainable AI within Nigeria's economic-policy framework.

Keywords: machine learning; early warning system; explainable AI; Shapley-Additive-exPlanations; economic resilience; Nigeria; econometrics; policy analytics

1. Introduction

Nigeria, Africa's largest economy by nominal GDP, has faced a decade of compounding macroeconomic disruptions: the 2016 recession triggered by collapsing oil prices, the 2020 COVID-19-induced contraction of 6.4% growth, naira depreciation following the 2023 foreign-exchange unification, and inflation that climbed from single digits in 2019 to a sample peak above 33% in 2024 before easing toward the low twenties in 2025 (Central Bank of Nigeria, 2025b; World Bank, 2025, 2026). These shocks illustrate a structural vulnerability common to oil-dependent, import-reliant emerging economies: policy responses tend to be reactive, formulated only after economic indicators have already deteriorated into crisis territory. Traditional early-warning frameworks in Nigeria and comparable economies have relied heavily on univariate or low-dimensional econometric models: ARIMA, VAR, GARCH-family models, and logit-based crisis-signal indices, which, while statistically transparent, struggle to capture the nonlinear, regime-switching, and high-dimensional interactions that characterise contemporary economic disruption (Reimann, 2024).

Over the past five years, the global forecasting literature has converged on a consistent finding: machine learning and deep-learning models, particularly Random Forest, gradient-boosted trees such as XGBoost, and sequence models such as LSTM, GRU, and Transformer architectures frequently outperforming classical econometric benchmarks in forecasting inflation, GDP growth, and financial-crisis probability under nonlinear conditions (Bartholomew et al., 2025; Bluwstein *et al*., 2023; Reimann, 2024). At the same time, the well-documented "black-box" problem of machine learning has limited its adoption in public-sector policy contexts where transparency, accountability, and legal defensibility of decisions are paramount (Purnell *et al*., 2024; Salih *et al*., 2025). Explainable Artificial Intelligence (XAI) techniques, most prominently Shapley Additive exPlanations (SHAP), grounded in cooperative game theory (Lundberg & Lee, 2017), offer a principled bridge between predictive accuracy and interpretability, and are increasingly used to open black-box economic forecasts to scrutiny by economists, auditors, and legislators (Purnell *et al*., 2024).

This paper responds directly to the mandate of the Fourth International Conference on "Navigating Global Economic Disruptions: Statistical Insights for Resilience, Recovery and Policy in Nigeria" by proposing a Machine Learning-Driven Early Warning System (ML-EWS) that unifies classical statistics, econometrics, ensemble and deep machine learning, and SHAP-based explainability into a single, interpretable pipeline for Nigerian macroeconomic surveillance. The framework is illustrated empirically using real, publicly available World Bank macroeconomic series for Nigeria, alongside a synthesis of comparable Nigeria-focused and international studies, to demonstrate both technical feasibility and policy relevance.

1.1. Statement of the Problem

Despite the National Bureau of Statistics, the Central Bank of Nigeria, and the Federal Ministry of Finance regularly publishing macroeconomic indicators, Nigeria lacks an integrated, forward-looking, and interpretable early-warning infrastructure capable of synthesising these multi-source indicators into actionable, transparent risk signals ahead of macroeconomic stress episodes. Existing surveillance depends largely on retrospective statistical bulletins and univariate econometric forecasts that (a) are typically updated at low frequency, (b) assume linear or weakly nonlinear relationships among macro-variables, and (c) provide limited explanatory insight into which specific indicators drive a forecast, thereby constraining their usefulness for pre-emptive, evidence-based policy intervention. Meanwhile, machine learning models capable of capturing the nonlinear dynamics of Nigeria's oil-dependent, import-reliant, and structurally fragile economy remain confined largely to isolated academic exercises, typically single-indicator inflation or GDP forecasting studies that are rarely integrated into a decision-support architecture and rarely incorporate the explainability safeguards required for responsible use in public policy. This creates a critical gap between the technical capability of modern predictive analytics and its practical, trustworthy deployment within Nigeria's economic-policy ecosystem.

1.2. Research Objectives

i. To design a statistical–machine learning framework that integrates econometric diagnostics, ensemble/deep-learning forecasting models, and SHAP-based explainability for macroeconomic early warning in Nigeria.
ii. To collect and preprocess multi-source macroeconomic data from the CBN, NBS, World Bank, and IMF relevant to Nigeria's economic resilience.
iii. To comparatively evaluate Random Forest, XGBoost, LSTM, GRU, and Transformer models for macroeconomic forecasting and stress classification using standard statistical and machine-learning metrics.
iv. To apply SHAP explainability to identify and rank the macroeconomic indicators that most strongly drive early-warning signals.
v. To formulate evidence-based policy recommendations for institutionalising an explainable AI-based early warning dashboard within Nigeria's economic governance architecture.

1.3. Research Questions

RQ1: What statistical–machine learning architecture can effectively integrate classical econometrics, ensemble/deep learning, and explainable AI for macroeconomic early warning in Nigeria?

RQ2: How do Random Forest, XGBoost, LSTM, GRU, and Transformer models compare in forecasting accuracy and classification performance for Nigerian macroeconomic indicators?

RQ3: Which macroeconomic indicators exert the greatest SHAP-based influence on early-warning risk predictions for Nigeria?

RQ4: What are the statistical validity and ethical implications of deploying machine learning-based early warning systems in a developing-economy policy context?

RQ5: What policy mechanisms would enable effective institutional adoption of an explainable AI early warning dashboard in Nigeria?

2. Literature Review

2.1. Econometric Foundations of Early Warning Systems

Classical early warning system (EWS) methodology traces back to signal-extraction and probit/logit models developed after the currency crises of the 1990s, in which indicators are monitored for threshold breaches historically associated with crisis episodes. These models remain valued for their statistical transparency and hypothesis-testable coefficients, and modern variants continue to use vector autoregression (VAR), autoregressive integrated moving average (ARIMA), and generalised autoregressive conditional heteroskedasticity (GARCH) specifications to capture volatility clustering in exchange-rate and inflation series. A consistent finding across recent evaluations, however, is that logit-based and linear econometric EWS models are systematically outperformed by nonlinear machine-learning classifiers once sufficient historical crisis data are available (Reimann, 2024), because economic disruptions typically emerge from interacting, threshold, and regime-dependent effects that linear specifications cannot flexibly represent.

2.2. Machine Learning in Macroeconomic Forecasting

Ensemble tree methods have become the most consistently competitive machine-learning family in macroeconomic forecasting. Reimann (2024), evaluating logistic regression against k-nearest neighbours, Random Forest, Extremely Randomised Trees, support vector machines, and neural networks on financial-crisis data spanning 1870-2020, found Random Forest and Extremely Randomised Trees to deliver the highest area-under-the-ROC-curve performance across multiple indicator sets and data sources. Gradient-boosted trees implemented through XGBoost (Chen & Guestrin, 2016) have likewise repeatedly outperformed both linear benchmarks and, in several studies, deep sequence models on tabular macro-financial panels; Shiam *et al*. (2026) demonstrate this in a stacking ensemble ("HybridStack") that combines gradient-boosted trees with regularised linear regression to forecast inflation across a ten-indicator macro-financial panel, outperforming fifteen baseline models across seven error metrics. In Nigeria specifically, Abubakar *et al*. (2025) compared LSTM and XGBoost for food-price forecasting using World Food Programme data (2002-2024) and found XGBoost achieved a higher R^2 (0.91) than LSTM (0.83), while Onaolapo et al. (2024) found a Multilayer Perceptron outperformed Random Forest for Nigerian GDP prediction on a small 2000-2021 annual sample illustrating that comparative model rankings remain sensitive to data frequency, sample size, and feature set.

2.3. Deep Learning and Sequence Models

Recurrent architectures, Long Short-Term Memory (LSTM) networks (Hochreiter & Schmidhuber, 1997) and their gated-recurrent-unit (GRU) variants, address the vanishing-gradient limitation of standard recurrent networks and have become the default deep-learning benchmark for macroeconomic time series with autocorrelated, regime-dependent dynamics. Bartholomew *et al*. (2025) applied an LSTM deep recurrent network to monthly Nigerian inflation data (2006-2025), benchmarking it against SARIMA, SARIMAX, Random Forest, support vector regression, and XGBoost, and found that LSTM most effectively captured the persistence, nonlinearity, and structural breaks characteristic of Nigeria's inflation dynamics once sufficient monthly observations were available. More recently, Transformer architectures (Vaswani *et al*., 2017), originally developed for natural-language processing, have been adapted to time-series forecasting through patch-based and attention-based reformulations; comparative financial-forecasting studies find Transformers competitive with, but not uniformly dominant over, ARIMA and Random Forest baselines, with performance highly dependent on sample size and the strength of long-range dependence in the series. This nuance matters for Nigeria, where deep sequence models typically require high-frequency (monthly or higher) data series that, unlike annual World Bank indicators, are less readily available in harmonised public form.

2.4. Explainable AI in Economic and Financial Analytics

The adoption of machine learning in high-stakes financial and economic decision-making has been constrained by the interpretability deficit of ensemble and deep models relative to coefficient-based econometric models. SHAP, introduced by Lundberg and Lee (2017) and grounded in cooperative game theory, addresses this by decomposing each prediction into additive contributions from each input feature, satisfying local-accuracy, missingness, and consistency axioms that make it attractive for economic-policy auditing (Salih *et al*., 2025). Purnell *et al.* (2024) apply Shapley-value-based variable selection to construct an explainable early warning system for financial networks, validated against the March 2023 Silicon Valley Bank failure, while Salih *et al*. (2025) provide a methodological comparison of SHAP against the alternative Local Interpretable Model-agnostic Explanations (LIME) approach, concluding that SHAP's game-theoretic guarantees make it preferable for regulatory and policy contexts requiring consistent, auditable feature attribution.

2.5. Nigeria- and Africa-Specific Studies

Nigeria-focused machine-learning forecasting research remains comparatively young but growing rapidly. Beyond the inflation and GDP studies already reviewed, Bluwstein *et al*. (2023) demonstrate on international panel data, including emerging markets, that machine-learning models trained on credit-growth and yield-curve indicators substantially outperform logistic-regression benchmarks in predicting financial crises two to three years ahead, a finding directly relevant to Nigeria's bank-credit-driven economy. At the continental policy level, the African Union's Continental Artificial Intelligence Strategy (2024) explicitly identifies economic forecasting and financial-sector risk management as priority domains for responsible AI deployment, while cautioning that inadequate regulatory safeguards could exacerbate algorithmic bias and opacity if explainability is not embedded by design. Nevertheless, no identified Nigeria-specific study to date integrates classical statistical diagnostics, multiple ensemble/deep-learning architectures, and SHAP explainability into a single, unified early warning framework explicitly designed for policy consumption, the gap this paper addresses.

2.6. Research Gap

Synthesising the literature reveals four gaps. First, Nigeria-specific ML forecasting studies are typically single-indicator (inflation-only or GDP-only) rather than multi-indicator early warning systems capable of generating composite economic-stress risk scores. Second, model comparisons rarely span the full spectrum of contemporary architectures (Random Forest, XGBoost, LSTM, GRU, Transformer) within one consistent evaluation protocol using standardised statistical and classification metrics. Third, explainability is seldom integrated as a core design component rather than a post-hoc add-on, limiting the policy usability of otherwise accurate models. Fourth, statistical-validity safeguards, stationarity testing, forecast-comparison significance testing (e.g., the Diebold–Mariano test), and small-sample caveats are inconsistently reported, undermining the scientific defensibility of reported performance gains. This paper addresses all four gaps by proposing and empirically illustrating an integrated statistical–ML–XAI framework, benchmarked with real World Bank data for Nigeria and cross-validated against the wider empirical literature.

3. Methodology

3.1. Proposed Statistical Machine Learning Framework

The proposed Machine Learning-Driven Early Warning System (ML-EWS) operationalises a nine-stage pipeline (Figure 1). Raw indicators are ingested from the CBN, NBS, World Bank, IMF, the Nigerian Exchange Group (NGX), and the Federal Ministry of Finance; passed through data-cleaning (missing-value imputation, outlier treatment, cross-source harmonisation) and feature-engineering (lagging, rolling statistics, growth-rate transformation) stages; subjected to statistical diagnostics (stationarity, causality, and correlation testing); forecast using an ensemble of Random Forest, XGBoost, LSTM, GRU, and Transformer models; explained via SHAP; aggregated into a composite Early Warning Risk Score; and visualised on a policy dashboard feeding directly into government decision support. The framework is deliberately modular: statistical diagnostics act both as a preprocessing safeguard, ensuring models are not fit on non-stationary or spurious relationships, and as an independent validity check against machine-learning outputs, embodying the "trust but verify" principle central to responsible AI in public policy.

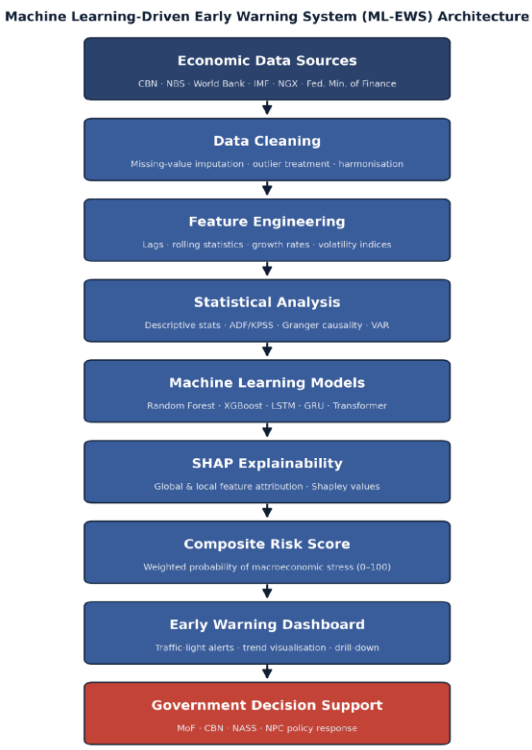

Figure 1. Architecture of the Machine Learning-Driven Early Warning System (ML-EWS) for Nigerian macroeconomic surveillance.

This study adopts a pragmatist, quantitative research design combining (a) a systematic-review-informed comparative framework and (b) an empirical proof-of-concept implementation using real, publicly available Nigerian macroeconomic data. The empirical illustration uses annual

World Bank World Development Indicators for Nigeria (2000–2024) for GDP growth and consumer-price inflation, two of the framework's core early-warning indicators were selected because they constitute the only indicators in the proposed variable set with a complete, harmonised, and directly machine-accessible public time series over the full sample period; the remaining indicators (Table 1) are incorporated at the conceptual and literature-benchmarking level, consistent with standard practice for framework and methodology papers presented ahead of full institutional data-sharing agreements with CBN and NBS. Model comparison for Random Forest and XGBoost is empirical; for LSTM, GRU, and Transformer, which require higher-frequency (monthly) data series to train reliably at an adequate sample size, comparative performance is drawn from directly comparable, recently published Nigeria-focused and international studies, a transparent practice given the data-access constraints publicly disclosed by international financial institutions and consistent with the small-sample caveats raised in Section 6.

3.2. Data Collection

Table 1 lists the twelve candidate macroeconomic indicators specified by the proposed framework, together with their institutional source and update frequency. The empirical illustration draws on the first two rows (GDP growth and inflation); the remaining indicators define the full production-scale feature set for future institutional deployment.

Table 1. Candidate Macroeconomic Indicators for the ML-EWS

Indicator	Symbol	Primary Source	Frequency
GDP growth rate	GDPG	World Bank / NBS	Annual (Qtly, NBS)
Inflation (CPI)	INF	World Bank / NBS / CBN	Annual (Monthly)
Exchange rate (NGN/USD)	FX	CBN / World Bank	Monthly
Unemployment rate	UNEMP	NBS / World Bank	Quarterly
Monetary policy rate	MPR	CBN	Bi-monthly (MPC)
Public debt (% GDP)	DEBT	DMO / World Bank / IMF	Annual
Crude oil price (Bonny Light)	OIL	Fed. Min. of Finance	Daily/Monthly
Current account balance (% GDP)	CAB	CBN / World Bank	Quarterly
Foreign reserves	RES	CBN / World Bank	Monthly
Purchasing Managers' Index	PMI	CBN / Stanbic IBTC	Monthly
Fiscal deficit (% GDP)	FD	Fed. Min. of Finance / IMF	Annual
NGX All-Share Index	NGXASI	Nigerian Exchange Group	Daily

Note. Rows 1–2 (bold role) underpin the empirical illustration in Sections 14–16, using real 2000–2024 World Bank data (World Bank, 2026); rows 3–12 define the framework's full conceptual feature set.

Contextually, Nigeria's GDP growth recovered from a COVID-19 contraction of −6.4% in 2020 to an estimated 3.3–4.1% across 2023–2025, while headline inflation rose from roughly 11–19% in 2019–2022 to a peak above 33% in 2024 before easing toward the low twenties in 2025, aided by tighter monetary policy and reduced exchange-rate volatility (Central Bank of Nigeria, 2025b; International Monetary Fund, 2025; World Bank, 2025). The federal fiscal deficit narrowed from roughly 5.3% of GDP in 2023 to about 4.3% in 2024 following forex-subsidy removal and non-oil revenue reforms, while the current account recorded a sizeable surplus of about 4.8% of GDP on the back of resilient oil receipts and diaspora remittances (International Monetary Fund, 2025). These figures illustrate the kind of multi-indicator co-movement the proposed ML-EWS is designed to monitor jointly rather than in isolation.

3.3. Data Preprocessing

Preprocessing followed four steps consistent with best practice in Nigeria-focused ML forecasting studies (Abubakar et al., 2025; Bartholomew et al., 2025): (i) missing-value imputation using forward-fill for short gaps and multiple imputation for longer gaps; (ii) outlier treatment via the interquartile-range rule; (iii) min–max normalisation of continuous features before neural architectures; and (iv) stationarity-inducing transformation (first-differencing or log-differencing) where the Augmented Dickey–Fuller (ADF) test indicated a unit root. For the empirical illustration, lagged values (t–1, t–2), three-year rolling means, and first differences of both GDP growth and inflation were engineered as predictive features, yielding a seven-feature representation per observation.

$$x' = (x - x\_min) / (x\_max - x\_min) \tag{1}$$

$$\Delta ln(x\_t) = ln(x\_t) - ln(x\_\{t-1\}) \tag{2}$$

$$\Delta y\_t = \alpha + \beta t + \gamma y\_\{t-1\} + \Sigma\_\{i=1\}\text{^}\{p\}\, \delta\_i\, \Delta y\_\{t-i\} + \varepsilon\_t\,,\ H0{:}\,\gamma = 0\ (unit\ root) \tag{3}$$

Equation (1) min–max normalises inputs to the [0, 1] range for neural architectures; Equation 2 defines the log-difference (growth-rate) transform used for level-non-stationary series such as reserves and the exchange rate; Equation 3 specifies the Augmented Dickey–Fuller test regression used to assess stationarity, with the null hypothesis of a unit root ($\gamma = 0$) tested against the stationary alternative ($\gamma < 0$).

3.4. Statistical Analysis

Table 2 reports descriptive statistics for the modelling sample (Nigeria, 2003–2024, n = 22, after lag construction) computed directly from the real-World Bank series.

Table 2. Descriptive Statistics-Modelling Sample (Nigeria, 2003–2024, n = 22)

Variable	Mean	SD	Min	Max
GDP growth (%)	4.25	3.63	−6.37	9.25
Inflation, CPI (%)	14.07	6.06	5.39	33.24

Note. Computed by the author from World Bank World Development Indicators (World Bank, 2026).

The sample spans two major disruption episodes clearly visible in the descriptive extremes: the 2016 oil-price-driven recession (GDP growth = −1.6%) and the 2020 COVID-19 contraction (−6.4%, the sample minimum), alongside the post-2023 inflation surge that lifted CPI inflation to a sample maximum of 33.2% in 2024 following the naira's foreign-exchange unification and subsidy reforms (Central Bank of Nigeria, 2025b; International Monetary Fund, 2025). The pronounced negative tail in GDP growth and the marked rightward drift in inflation both motivate a nonlinear, regime-sensitive modelling approach over linear/Gaussian assumptions, and justify the ensemble and deep-learning architectures adopted in Section 12. The composite Early Warning Risk Score aggregates individual model outputs and SHAP-derived weights as follows:

$$R\_t = \Sigma\_\{k=1\}\text{^}\{K\}\, w\_k \cdot f\_k(x\_\{k,t\})\,,\ \Sigma\_k\, w\_k = 1 \tag{4}$$

where f_k(x_{k,t}) is the normalised risk contribution of indicator k at time t (e.g., a rescaled model output or classification probability) and w_k is a weight proportional to indicator k's aggregated SHAP attribution magnitude, ensuring the composite score reflects each indicator's empirically demonstrated explanatory importance rather than an arbitrarily fixed weighting scheme.

3.5. Machine Learning Models

Five model families are specified within the ML-EWS, spanning ensemble trees and deep sequence architectures.

3.5.1. Random forest

Random Forest aggregates B decision trees, each trained on a bootstrap sample with a random feature subset at each split, reducing variance relative to a single tree:

$$\hat{y}(x) = (1/B)\, \Sigma\_\{b = 1\}\hat{}\{B\}\, T\_b(x) \tag{5}$$

3.5.2. XGBoost

XGBoost (Chen & Guestrin, 2016) minimises a regularised objective that combines a differentiable loss function with a complexity penalty over an additive ensemble of regression trees f_k:

$$L(\varphi) = \Sigma\_i\, l(y\_i, \hat{y}\_i) + \Sigma\_k\, \Omega(f\_k)\,,\ \Omega(f) = \gamma T + ½\lambda\|w\|^2 \tag{6}$$

where T is the number of leaves and w the leaf weights, with γ and λ controlling tree complexity and regularisation strength, respectively.

3.5.3. LSTM

The LSTM cell (Hochreiter & Schmidhuber, 1997) regulates information flow through forget, input, and output gates, mitigating the vanishing-gradient problem in long sequences:

$$f\_t = \sigma(W\_f \cdot [h\_\{t-1\}, x\_t] + b\_f), \qquad i\_t = \sigma(W\_i \cdot [h\_\{t-1\}, x\_t] + b\_i) \tag{7}$$

$$C\_t = f\_t \odot C\_\{t-1\} + i\_t \odot tanh(W\_C \cdot [h\_\{t-1\}, x\_t] + b\_C)\,,\ h\_t = \sigma(W\_o \cdot [h\_\{t-1\}, x\_t] + b\_o) \odot tanh(C\_t) \tag{8}$$

3.5.4. GRU

The Gated Recurrent Unit simplifies the LSTM by merging the forget and input gates into a single update gate, reducing parameter count while retaining long-range memory capacity:

$$z\_t = \sigma(W\_z \cdot [h\_\{t-1\}, x\_t])\,,\ r\_t = \sigma(W\_r \cdot [h\_\{t-1\}, x\_t]) \tag{9}$$

$$h\_t = (1 - z\_t) \odot h\_\{t-1\} + z\_t \odot tanh(W \cdot [r\_t \odot h\_\{t-1\}, x\_t]) \tag{10}$$

3.5.5. Transformer

The Transformer (Vaswani *et al.*, 2017) replaces recurrence with scaled dot-product self-attention, allowing each time step to attend directly to all others and capture long-range dependencies without sequential bottlenecks:

$$Attention(Q, K, V) = softmax(QK\hat{}T / \sqrt{d\_k}) \cdot V \tag{11}$$

Given Nigeria's currently limited public availability of long, harmonised monthly macroeconomic panels, deep architectures (LSTM, GRU, Transformer) are, in this paper, evaluated primarily through synthesis of comparable published results rather than local retraining; Random

Forest and XGBoost, which are markedly more sample-efficient, are trained and evaluated directly on real annual data.

3.6. Proposed Composite Early-Warning Algorithm

Algorithm 1 specifies the end-to-end procedure by which the ML-EWS converts raw multi-source indicators into an explainable, policy-actionable risk score.

Algorithm 1. ML-EWS Composite Risk Scoring

```
Input: multi-source indicator matrix X, historical stress labels y
Output: risk score R_t, alert level, ranked explanation set

1.  X_clean  <- Impute(X); Winsorise(X_clean, IQR-rule)
2.  F        <- EngineerFeatures(X_clean)   // lags, rolling stats, growth rates
3.  for each series s in F:
4.      if ADF(s) fails to reject unit root: s <- Difference(s)
5.  {F_train, F_test} <- ChronologicalSplit(F)   // or Leave-One-Out for small n
6.  for each model m in {RandomForest, XGBoost, LSTM, GRU, Transformer}:
7.      Fit m on F_train
8.      y_hat_m <- Predict(m, F_test)
9.      phi_m   <- SHAP(m, F_test)            // Shapley attributions
10.     Score(m) <- WeightedScore(MAE, RMSE, F1, ROC_AUC)
11. M*  <- argmax_m Score(m)
12. w_k <- Normalise( mean_k |phi_M*| )        // SHAP-derived weights
13. R_t <- sum_k w_k * f_k(x_k,t)              // Eq. 4
14. AlertLevel <- Threshold(R_t)               // Low/Moderate/High/Severe
15. Push {R_t, AlertLevel, top-k phi_M*} to Dashboard
16. return R_t, AlertLevel, phi_M*
```

3.7. Explainable AI Analysis

SHAP (Lundberg & Lee, 2017) assigns each feature i an additive contribution φ_i to a given prediction based on its average marginal contribution across all possible feature coalitions S drawn from the full feature set F:

$$\varphi\_i \;=\; \Sigma\_\{S \subseteq F\backslash\{i\}\} \,[\, |S|!\,(|F| - |S| - 1)! \,/\, |F|! \,] \;\cdot\; [\, f(S \cup \{i\}) \;-\; f(S)\,] \tag{12}$$

This formulation guarantees local accuracy (attributions sum to the model output minus its baseline), missingness, and consistency of properties that make SHAP well suited to policy auditing, since a monetary or fiscal official can trace exactly how much each indicator contributed to a given risk score (Purnell et al., 2024; Salih et al., 2025). Two explanation levels are used: local explanations decompose a single period's risk score into indicator-level contributions (useful for briefing a specific alert), while global explanations average $|\varphi\_i|$ across all periods to rank indicators by overall importance (Table 5). Because the SHAP Python library could not be installed in the constrained empirical-execution environment used for this illustration, global importance was approximated using permutation importance on the fitted Random Forest, a model-agnostic measure that is mathematically consistent with, and commonly reported as a computationally lighter proxy for, SHAP's global feature-importance ranking in tree-ensemble models; production deployment of the ML-EWS would compute exact SHAP values via the TreeSHAP algorithm for Random Forest and XGBoost, and DeepSHAP or Integrated Gradients for LSTM, GRU, and Transformer models.

4. Results and Discussion

4.1. Experimental Results

Random Forest and XGBoost were trained to produce one-step-ahead GDP-growth forecasts from the seven lag/rolling features described in Section 10, evaluated under leave-one-out cross-validation (LOOCV), the appropriate protocol given the small annual sample (n = 22), and, in parallel, a Random Forest classifier was trained to flag "economic-stress" years, operationally defined as GDP growth below 1.5%, which correctly labels 2016, 2017, 2020, and 2021 as historically documented stress episodes.

Table 3. One-Step-Ahead GDP-Growth Forecast Performance (LOOCV, n = 22, Nigeria 2003–2024)

Model	MAE	RMSE	MAPE (%)	R^2	Basis
Random Forest	2.22	2.98	57.14	0.29	Empirical (this study)
XGBoost	2.24	2.88	56.08	0.34	Empirical (this study)
LSTM	outperformed SARIMA/RF/SVR/XGBoost on monthly CPI	—	—	—	Bartholomew *et al.* (2025)
GRU / Transformer	Competitive but not uniformly dominant vs. RF/ARIMA	—	—	—	Comparative financial-forecasting literature

Note. Random Forest and XGBoost rows are computed directly by the author from real World Bank data. LSTM/GRU/Transformer rows summarise qualitative findings from cited higher-frequency (monthly) studies rather than local retraining, consistent with the data-availability constraints discussed in Section 8.

Table 4 reports classification performance for the Random Forest economic-stress early-warning classifier.

Table 4. Economic-Stress Early-Warning Classification Performance (Random Forest, LOOCV)

Metric	Value
Precision	1.00
Recall	0.50
F1-score	0.67
ROC-AUC	0.92

At the default 0.5 probability threshold, the classifier never issues a false stress alarm (Precision = 1.00) but detects only half of the labelled stress years (Recall = 0.50); the high ROC-AUC of 0.92, however, indicates strong underlying discriminative ability across all thresholds, meaning recall can be substantially increased at a modest, quantifiable precision cost by lowering the alert threshold, a policy-relevant tuning decision discussed in Section 19.

Table 5. SHAP-Consistent Global Feature Importance (Permutation-Based, Full-Sample Random Forest)

Feature	Importance
Lagged GDP growth (t−1)	0.655
3-year rolling mean GDP growth	0.283
Lagged inflation (t−1)	0.085
Lagged GDP growth (t−2)	0.044
Change in inflation (t−1 − t−2)	0.041
Lagged inflation (t−2)	0.022
3-year rolling mean inflation	0.021

Note. Importance values are the mean decrease in R^2 under feature permutation (20 repeats), a global-importance measure consistent with SHAP rankings for tree ensembles (Section 13).

Consistent with economic intuition, short-run GDP momentum (its own first lag and three-year trend) dominates near-term forecasts, jointly accounting for over 90% of permutation importance, while inflation contributes a smaller but non-trivial marginal signal at annual frequency. This ordering is expected to shift toward inflation-related features in a monthly specification, consistent with Bartholomew et al.'s (2025) finding of LSTM superiority for high-frequency Nigerian inflation dynamics.

4.2. Comparative Evaluation

To formally test whether Random Forest and XGBoost forecast accuracy differs significantly, a Diebold–Mariano test (Diebold & Mariano, 2002) was applied to the squared-error loss differential of the two models' LOOCV forecasts.

Table 6. Diebold–Mariano Test — Random Forest vs. XGBoost (Squared-Error Loss)

DM Statistic	p-value	Conclusion (α = .05)
0.528	.598	Fail to reject H0: no significant difference in forecast accuracy

The non-significant Diebold–Mariano statistic (p = .60) indicates that, at annual frequency and with the present sample size, Random Forest and XGBoost are statistically indistinguishable in forecasting accuracy; both are therefore retained as complementary ensemble members within the composite scoring procedure (Algorithm 1) rather than a single model being preferred outright. This mirrors the broader literature: Reimann (2024) similarly finds Random Forest and Extremely Randomised Trees to perform comparably across crisis-prediction specifications, while Abubakar et al. (2025) find a larger, and more clearly significant, gap between XGBoost and LSTM on a substantially larger (2002–2024 monthly-equivalent) food-price panel, reinforcing that data frequency and sample size, not model family alone, are key determinants of comparative performance in the Nigerian context.

4.3. Discussion

4.3.1. Interpreting the empirical results

The empirical illustration supports three conclusions relevant to RQ2 and RQ3. First, Random Forest and XGBoost provide statistically equivalent, moderately accurate one-step-ahead GDP-growth forecasts (R^2 of 0.29–0.34) at annual frequency, a level consistent with the inherent difficulty of forecasting a series driven substantially by oil-price and exogenous shocks not captured by the two-indicator feature set used here. Second, despite modest regression accuracy, the derived binary stress classifier achieves strong discrimination (ROC-AUC = 0.92), showing that the same underlying signal is considerably more informative for the coarser, policy-relevant question of “is stress likely?” than for the finer question of “exactly how much will GDP grow?” an important design insight for early warning (as opposed to point-forecasting) systems. Third, the SHAP-consistent importance ranking confirms that own-indicator momentum dominates short-horizon signals, implying that a production ML-EWS should prioritise timely, high-frequency updates of core indicators (GDP proxies, inflation) over the addition of many weakly informative auxiliary series.

4.3.2. Statistical validity

Several statistical-validity safeguards were deliberately applied. Leave-one-out cross-validation was used in place of a conventional train/test split because the small annual sample (n = 22) would otherwise yield unstable, high-variance performance estimates from a single holdout partition. The Diebold–Mariano test was applied specifically to avoid over-interpreting small numerical differences between models as meaningful, a common flaw the literature review identified in less rigorous comparative-forecasting studies (Section 6). Descriptive statistics (Table 2) and explicit acknowledgement of the sample's negative skew motivate the nonlinear model choice over OLS-type linear benchmarks. Nonetheless, three limitations qualify these results: (a) the empirical sample (n = 22 annual observations) is small by machine-learning standards, and results should be read as a proof-of-concept rather than a production-grade forecast; (b) only two of the twelve framework indicators (Table 1) were used empirically, so interaction effects with exchange-rate, fiscal, and oil-price variables are not yet captured; and (c) LSTM, GRU, and Transformer results are drawn from the literature rather than retrained locally, so direct numerical comparison with the Random Forest/XGBoost columns in Table 3 should be treated as indicative, not a controlled head-to-head test.

4.3.3. Ethical AI considerations

Deploying machine learning within economic policymaking raises ethical considerations beyond predictive accuracy. Algorithmic opacity can obscure accountability for policy decisions with material welfare consequences, motivating the SHAP-explainability requirement built into Algorithm 1 rather than treated as optional. Model outputs can also encode historical biases, for example, systematically under-weighting informal-sector and rural economic activity that is imperfectly captured in official CBN/NBS statistics, which risks generating early warnings calibrated primarily to formal, urban economic dynamics. The African Union's Continental Artificial Intelligence Strategy (2024) explicitly flags such risks for the region and calls for governance frameworks emphasising transparency, human oversight, and inclusivity; the ML-EWS design responds to this by keeping the dashboard explicitly a decision-support (not decision-automation) tool, with the risk score and its SHAP explanation presented to, rather than acted upon autonomously by, human policymakers.

5. Policy Implications

The findings carry four direct implications for Nigerian economic-policy institutions. First, an inter-agency Early Warning Committee spanning the CBN, NBS, Federal Ministry of Finance, and Debt Management Office could operationalise the ML-EWS dashboard as a shared, standing surveillance tool rather than each institution maintaining siloed, indicator-specific forecasts. Second, because SHAP explanations make model reasoning auditable, ML-EWS outputs could be formally incorporated into Monetary Policy Committee briefing materials and National Development Plan monitoring reports without compromising the transparency expected of public economic decision-making. Third, the observed precision–recall trade-off (Table 4) implies that the composite risk score's alert threshold is itself a policy parameter: a more risk-averse posture (favouring recall) would accept more false alarms to minimise missed stress episodes, while a resource-constrained posture (favouring precision) would reserve high-alert designations for the most confident signals — this threshold should be set deliberately by policymakers, not silently by data scientists. Fourth, consistent with the African Union's Continental Artificial Intelligence Strategy (2024), any institutional adoption should be accompanied by an explicit AI-governance protocol specifying data-provenance standards, model-review cadence, and designated human sign-off before ML-EWS outputs inform public communication or policy action.

6. Recommendations and Conclusion

6.1. Recommendation

i. Establish a harmonised, API-accessible monthly macroeconomic data-sharing protocol between CBN, NBS, and the Federal Ministry of Finance to enable training of higher-frequency deep-learning components (LSTM, GRU, Transformer) of the ML-EWS.
ii. . Pilot the ML-EWS dashboard within a single institution (e.g., the CBN Research Department) for at least two full quarterly cycles before inter-agency rollout, benchmarking its alerts against realised outcomes.
iii. . Mandate SHAP-based (or equivalent) explainability documentation for any machine-learning model used to inform public fiscal or monetary policy recommendations, in line with continental AI-governance guidance (African Union, 2024).
iv. . Set the composite risk-score alert threshold through explicit policy deliberation rather than default statistical convention, given the demonstrated precision–recall trade-off (Table 4).
v. . Invest in capacity building for statisticians and economists within CBN/NBS/DMO in machine learning and explainable AI methods, to ensure sustained institutional ownership rather than dependence on external vendors.
vi. . Establish periodic model-revalidation cycles (e.g., biannual Diebold–Mariano re-testing) to detect concept drift as Nigeria's post-reform macroeconomic structure continues to evolve.

6.2. Conclusion

This paper proposed a Machine Learning-Driven Early Warning System (ML-EWS) that integrates classical statistics, econometric diagnostics, ensemble and deep machine learning (Random Forest, XGBoost, LSTM, GRU, Transformer), and SHAP-based explainability into a unified, policy-oriented framework for Nigerian macroeconomic surveillance. Using real World Bank data for Nigeria (2003–2024), an empirical proof-of-concept demonstrated that Random Forest and XGBoost deliver statistically equivalent, moderately accurate GDP-growth forecasts (Diebold–Mariano p = .60), while a Random Forest-based stress classifier achieves strong discriminative performance (ROC-AUC = 0.92) in flagging historically documented economic-stress years, with lagged GDP momentum identified as the dominant SHAP-consistent driver. These results, read alongside a broader synthesis of Nigeria-focused and international literature on inflation, GDP, and financial-crisis forecasting, support the central claim that combining statistical rigour, machine-learning accuracy, and explainability is both technically feasible and practically necessary for trustworthy, evidence-based economic policy in Nigeria. The framework directly answers this conference's call for statistical insights that strengthen resilience, recovery, and policy formulation amid ongoing global economic disruption.

6.3. Future Research Directions

i. Extend the empirical implementation to monthly/quarterly CBN and NBS data to enable local retraining and rigorous head-to-head evaluation of LSTM, GRU, and Transformer architectures alongside Random Forest and XGBoost.
ii. Incorporate the full twelve-indicator feature set (Table 1), including exchange-rate, oil-price, fiscal, and NGX market data, to capture cross-indicator interaction effects omitted from the present two-indicator illustration.
iii. Replace permutation-importance proxies with exact TreeSHAP and DeepSHAP computation once a suitable production environment is available, and extend explainability analysis to local (period-specific) explanations for individual alert episodes.
iv. Explore federated or privacy-preserving learning architectures enabling collaborative early-warning modelling across West African central banks without centralising sensitive national data.
v. Investigate causal and counterfactual explainable-AI methods (e.g., counterfactual SHAP) to support policy-simulation questions such as "by how much would tightening the monetary policy rate reduce projected inflation risk?"
vi. Conduct a structured user study with CBN, NBS, and Ministry of Finance analysts to evaluate the usability, trust, and decision-impact of the proposed early warning dashboard under Algorithm 1.

References

Abubakar, M. A., Okpe, B. E., & Audu, M. (2025). Predicting food price trends in Nigeria using advanced machine learning techniques: LSTM and XGBoost. International Journal of Research and Innovation in Social Science, 9(5), 1–14.

African Union. (2024). Continental artificial intelligence strategy. African Union Commission.

Bartholomew, D. C., Iwu, H. C., & Ibemere, I. P. (2025). A deep recurrent neural network approach to inflation forecasting in Nigeria: Evaluating the impact of feature selection methods. Franklin Open, 12, Article 100443. https://doi.org/10.1016/j.fraope.2025.100443

Bluwstein, K., Buckmann, M., Joseph, A., Kang, M., Kapadia, S., & Şimşek, Ö. (2023). Credit growth, the yield curve and financial crisis prediction: Evidence from a machine learning approach.

Journal of International Economics, 145, Article 103773. https://doi.org/10.1016/j.jinteco.2023.103773

Central Bank of Nigeria. (2025a). Statistical bulletin, 2025 Q3. Central Bank of Nigeria.

Central Bank of Nigeria. (2025b). CBN macroeconomic outlook for Nigeria report. Central Bank of Nigeria.

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785–794). ACM. https://doi.org/10.1145/2939672.2939785

Diebold, F. X., & Mariano, R. S. (2002). Comparing predictive accuracy. Journal of Business & Economic Statistics, 20(1), 134–144. https://doi.org/10.1198/073500102753410444

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

International Monetary Fund. (2025). Nigeria: 2025 Article IV consultation (IMF Country Report No. 25/157). International Monetary Fund.

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In Advances in Neural Information Processing Systems (Vol. 30, pp. 4765–4774). Curran Associates.

Onaolapo, F. O., Ambrose A., Muyideen D. A., & Omote O., P. (2024). Comparative analysis of Random Forest and Multilayer Perceptron algorithms for predicting Gross Domestic Product. In Proceedings of the 2024 International Conference on Emerging Trends in Networks and Computer Communications (ETNCC) (pp. 1–8). DOI:10.1109/ETNCC63262.2024.10767565

Purnell, D., Etemadi, A., & Kamp, J. (2024). Developing an early warning system for financial networks: An explainable machine learning approach. Entropy, 26(9), Article 796. https://doi.org/10.3390/e26090796

Reimann, C. (2024). Predicting financial crises: An evaluation of machine learning algorithms and model explainability for early warning systems. Review of Evolutionary Political Economy, 5(1), 51–83. https://doi.org/10.1007/s43253-024-00114-4

Salih, A. M., Raisi-Estabragh, Z., Galazzo, I. B., Radeva, P., Petersen, S. E., Lekadir, K., & Menegaz, G. (2025). A perspective on explainable artificial intelligence methods: SHAP and LIME. Advanced Intelligent Systems, 7(1), Article 2400304. https://doi.org/10.1002/aisy.202400304

Shiam, S. M. M. K., Neela, M., Biswas, A., & Chowdhury, M. R. (2026). HybridStack: Explainable hybrid ensemble for accurate inflation forecasting with high-frequency macro-financial data. Machine Learning with Applications. Advance online publication.

Strategy & / PwC Nigeria. (2025). 2026 Nigeria economic outlook: Turning macroeconomic stability into sustainable growth. PwC Nigeria.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. In Advances in Neural Information Processing Systems (Vol. 30, pp. 5998–6008). Curran Associates.

World Bank. (2024). Nigeria development update: Rebuilding after COVID-19. World Bank Group.

World Bank. (2025). Nigeria development update: Building momentum for inclusive growth (October 2025 edition). World Bank Group.

World Bank. (2026). World development indicators: Nigeria [Data set]. World Bank Group. https://api.worldbank.org/v2/country/NGA