INTERNATIONAL JOURNAL OF RESEARCH AND INNOVATION IN APPLIED SCIENCE (IJRIAS)
ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VII July 2026

Machine Learning-Based Forecasting Model for Paddy Rice Price movements Using Climatic and Macroeconomic Variables in Gombe State, Nigeria

Abdulrashid Isiyaku*, Prof. Bala Modi, A. M. Yahaya, Aminu Abdullahi Bala, Zainab Ahmed

Department of Computer Science, Faculty of Science, Gombe State University, Gombe, Nigeria

***Corresponding Author**

DOI: https://doi.org/10.51584/IJRIAS.2026.11070164

Received: 02 August 2026; Accepted: 08 August 2026; Published: 17 August 2026

ABSTRACT

Paddy rice is a major staple crop in Gombe State, Nigeria, contributing significantly to food security, household income, and rural livelihoods. However, paddy rice prices are highly volatile due to the combined effects of climatic variability, macroeconomic conditions, and market dynamics, creating uncertainty for farmers, traders, and policymakers. Existing forecasting studies have primarily focused on generalized agricultural commodities or relied on historical price data with limited integration of climatic and macroeconomic variables for localized price prediction. This study developed a machine learning-based forecasting model for predicting paddy rice price movements in Gombe State by integrating monthly historical paddy rice prices with climatic variables (rainfall, temperature, and relative humidity) and selected macroeconomic indicators obtained from the Climate Hazards Group InfraRed Precipitation with Station Data (CHIRPS), the Central Bank of Nigeria (CBN), the National Bureau of Statistics (NBS), FAOSTAT, **and the Gombe State Agricultural Development Programme (GSADP).** Five machine learning models, namely Random Forest, Extreme Gradient Boosting (XGBoost), LightGBM, CatBoost, and Long Short-Term Memory (LSTM), were developed and evaluated using Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and the coefficient of determination (R^2). The comparative evaluation showed that the Random Forest model achieved the best forecasting performance among the evaluated models. The developed forecasting framework provides a reliable decision-support tool for anticipating paddy rice price movements, reducing financial risk, improving production and marketing decisions, and supporting evidence-based agricultural policy formulation to enhance market efficiency in Gombe State, Nigeria.

Keywords: Price Movement; Machine learning; Random Forest; Climatic variables; Macroeconomic variables.

INTRODUCTION

Agriculture remains one of the most important sectors of the Nigerian economy, contributing significantly to food security, employment generation, and rural livelihoods. Among the various agricultural commodities produced in Nigeria, rice has emerged as one of the most widely consumed staple foods due to population growth, urbanization, changing dietary preferences, and increasing demand for locally produced food (Sanusi et al., 2022). Consequently, rice production has become a strategic priority for achieving national food security and reducing dependence on imports.

Despite the increasing level of rice production in Nigeria, the paddy rice market continues to experience substantial price fluctuations. These fluctuations are influenced by seasonal variations, climatic conditions, production costs, transportation expenses, market demand and supply dynamics, inflationary pressures, exchange rate movements, and government agricultural policies (Sanusi et al., 2022; Food and Agriculture Organization [FAO], 2022). Such price instability creates uncertainty for farmers, traders, processors, policymakers, and consumers, making effective production planning, marketing, investment, and policy formulation increasingly difficult.

In Gombe State, rice production has expanded considerably due to favorable agro-ecological conditions and increasing participation in rice cultivation (Dzarma et al., 2024). Nevertheless, paddy rice prices remain highly volatile, exposing stakeholders to financial risks and reducing market efficiency. Accurate forecasting of paddy rice price movements can therefore provide valuable decision support for farmers, traders, processors, policymakers, and other stakeholders involved in the rice value chain.

Conventional agricultural commodity price forecasting has traditionally relied on statistical and econometric techniques such as the Autoregressive Integrated Moving Average (ARIMA) model and multiple regression analysis. Although these methods have demonstrated reasonable predictive capability, they often struggle to capture the complex nonlinear relationships among climatic, macroeconomic, and market variables (Sanusi et al., 2022). Recent advances in machine learning have provided more robust forecasting techniques capable of handling nonlinear interactions, high-dimensional datasets, and complex temporal patterns (Sari et al., 2024).

Furthermore, the increasing availability of climatic datasets and macroeconomic information presents new opportunities for developing localized forecasting models that integrate multiple explanatory variables. Previous studies have demonstrated that combining historical commodity prices with climatic and economic indicators substantially improves forecasting performance (Zhang & Tang, 2024). However, relatively few studies have focused specifically on localized paddy rice price forecasting within Gombe State, Nigeria.

Therefore, this study develops a machine learning-based forecasting framework for predicting paddy rice price movements in Gombe State by integrating historical market prices with climatic variables, macroeconomic indicators, and temporal features. The study evaluates multiple machine learning algorithms to identify the most suitable model for localized paddy rice price prediction and to provide reliable decision support for farmers, traders, policymakers, and other stakeholders.

METHODOLOGY

Research Design

This study adopted a quantitative research design using a machine learning-based forecasting approach to predict paddy rice price movements in Gombe State, Nigeria. The research employed a data-driven methodology that integrated historical paddy rice price data with climatic and macroeconomic variables to identify underlying patterns and develop predictive models for forecasting future price movements.

The study utilized multivariate time-series data comprising historical paddy rice prices, climatic variables, and macroeconomic indicators. Five machine learning algorithms—Random Forest (RF), Extreme Gradient Boosting (XGBoost), Light Gradient Boosting Machine (LightGBM), Categorical Boosting (CatBoost), and Long Short-Term Memory (LSTM)—were developed and comparatively evaluated to forecast paddy rice price movements. Model performance was assessed using Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and the coefficient of determination (R^2) to identify the most suitable forecasting model.

The selected research design was appropriate because it supports the analysis of complex nonlinear relationships among historical price, climatic, and macroeconomic variables, thereby enhancing forecasting accuracy and providing evidence-based decision support for farmers, traders, agricultural investors, and policymakers.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VII July 2026

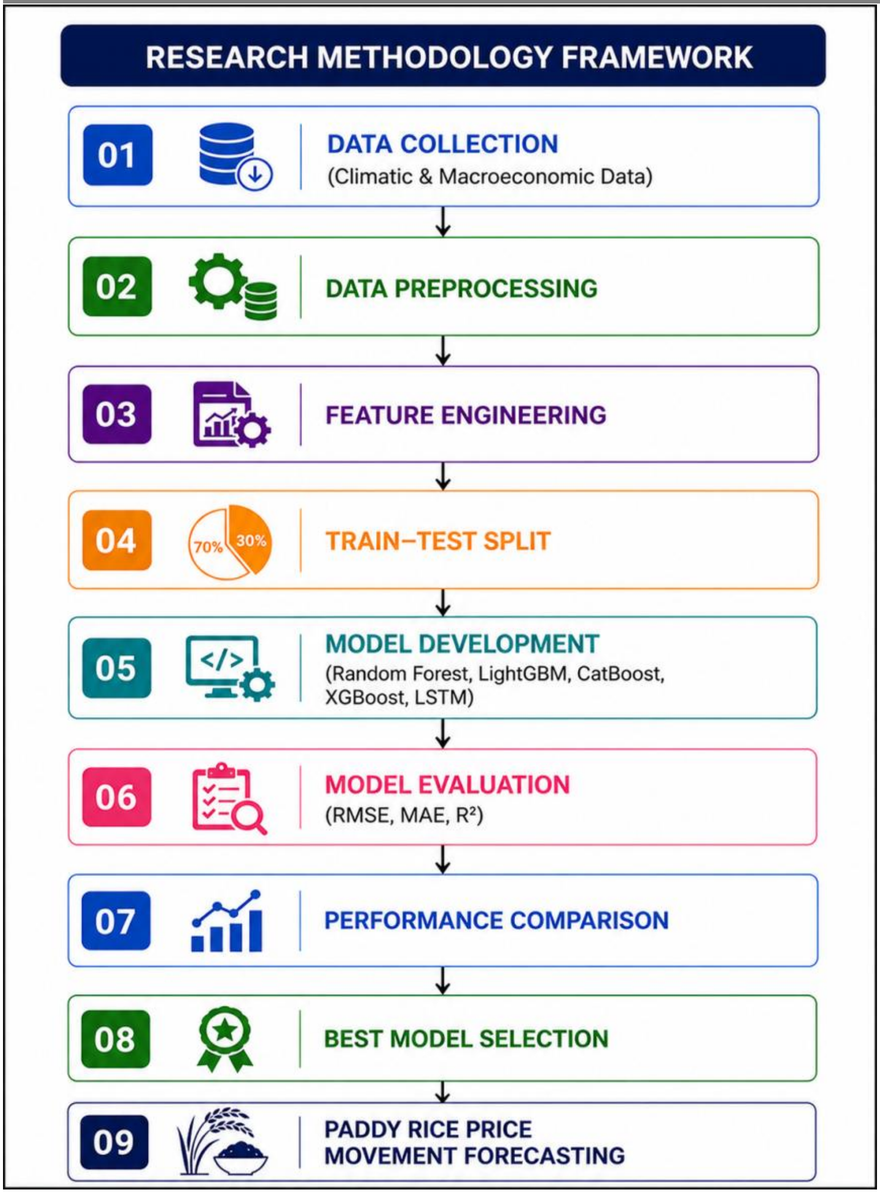

Figure 1: Research Methodology Framework

Study Area

The study was conducted in **Gombe State, Nigeria**, located in the northeastern region of the country between latitudes **9°30′N and 12°30′N** and longitudes **8°45′E and 11°45′E**. The state shares boundaries with **Borno, Yobe, Bauchi, Taraba, and Adamawa States.**

Gombe State experiences a tropical continental climate with distinct wet and dry seasons. The rainy season extends from **May to October**, while the dry season lasts from **November to April**. Average annual rainfall ranges from approximately **850 mm to 1,200 mm**, providing favorable conditions for agricultural production.

Agriculture is one of the major economic activities in the state, with a significant proportion of the population depending on farming for livelihood and income generation. Major crops cultivated include **rice, maize, millet, sorghum, beans, and groundnuts**. Paddy rice is an important agricultural commodity produced across several local government areas where favorable climatic conditions support cultivation. However, rice production and market prices are influenced by rainfall variability, temperature fluctuations, inflationary pressures, transportation costs, and government policy interventions.

Gombe State was selected for this study because of its importance in paddy rice production and the need for a localized forecasting framework to support agricultural planning, market monitoring, and policy-oriented decision-making.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VII July 2026

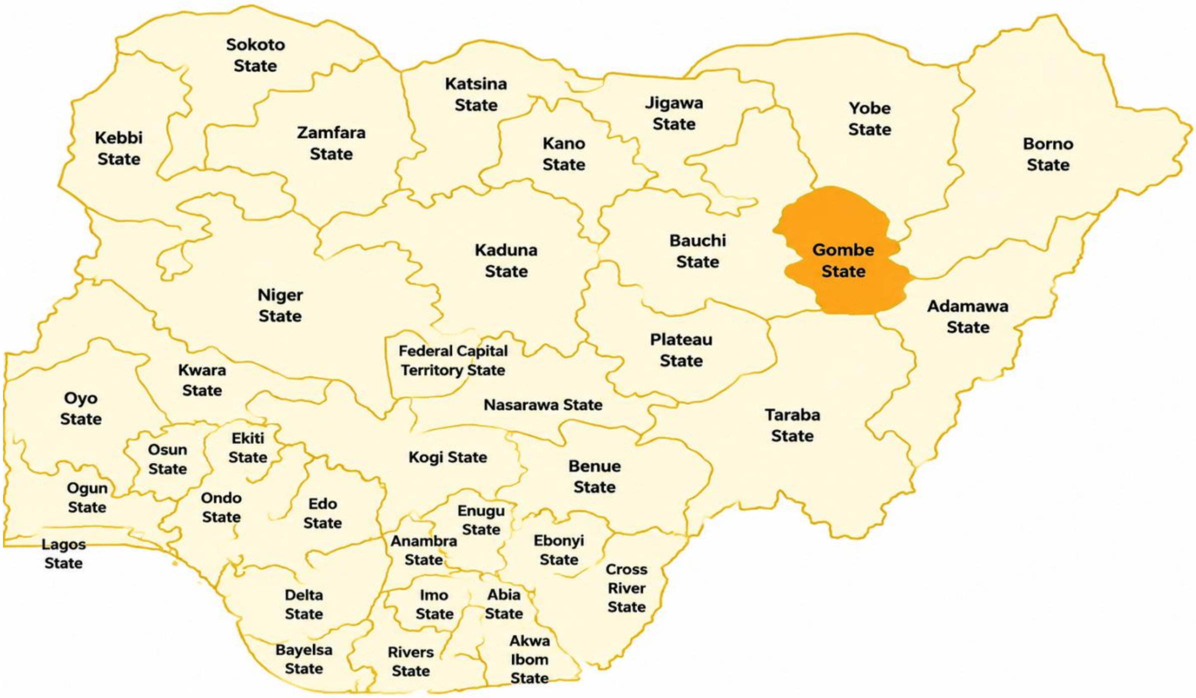

Figure 2: Location of Gombe State within Nigeria

Sources of Data

This study utilized secondary data collected from multiple sources to develop machine learning models for forecasting paddy rice price movements in Gombe State, Nigeria. The dataset comprised historical paddy rice prices, climatic variables, and macroeconomic indicators collected on a monthly basis between 2011 and 2025. Integrating these datasets enabled the forecasting models to capture the combined effects of environmental and economic factors on paddy rice price movements.

Historical monthly paddy rice price data covering the period from 2011 to 2025 were obtained from the Gombe State Agricultural Development Programme (GSADP) and served as the target variable for model development. Climatic variables, including rainfall, temperature, and relative humidity, were obtained from the Climate Hazards Group InfraRed Precipitation with Station Data (CHIRPS). Macroeconomic variables, including the inflation rate and exchange rate, were obtained from the Central Bank of Nigeria (CBN) and the National Bureau of Statistics (NBS), while agricultural statistical data were obtained from the Food and Agriculture Organization Statistics Database (FAOSTAT). These data sources provided the historical climatic, macroeconomic, and agricultural information required to develop and evaluate the machine learning models for forecasting paddy rice price movements in Gombe State, Nigeria

All datasets were collected on a monthly basis to ensure temporal consistency and compatibility for multivariate time-series analysis. Before model development, the integrated dataset underwent data preprocessing, including data cleaning, handling of missing values, normalization, and consistency checks to improve data quality, enhance model performance, and ensure the reliability of the forecasting results.

Description of Variables and Dataset

A multivariate dataset comprising climatic, macroeconomic, temporal, and engineered variables was developed to forecast paddy rice price movements in Gombe State. The target variable was **Price movements**, while the predictor variables included historical paddy rice prices, climatic indicators, macroeconomic factors, temporal variables, and engineered features derived from historical observations. These variables were selected to capture the environmental, economic, and temporal factors influencing paddy rice price movements.

The climatic variables consisted of monthly rainfall, average temperature, and relative humidity obtained from the Climate Hazards Group InfraRed Precipitation with Station Data (CHIRPS) database. The macroeconomic variables included food inflation rate, exchange rate, fertilizer price, labour cost, PMS Price, and market demand index obtained from the Central Bank of Nigeria (CBN), the National Bureau of Statistics (NBS), FAOSTAT, and the Gombe State Agricultural Development Programme (GSADP). These variables were incorporated to capture the influence of climatic conditions and macroeconomic factors on agricultural production, market supply, production costs, transportation, and consumer purchasing power. (Funk et al., 2015; Central Bank of Nigeria, 2025; National Bureau of Statistics, 2025; FAOSTAT, 2025).

To improve the predictive capability of the machine learning models, engineered features were generated from the historical price and climatic data. These included **Price_Lag_1, Price_Lag_2, Price_Lag_3, Price_MA_3, Price_MA_6, Rainfall_3M_Total, Rainfall_6M_Total,** and **Season_Encoded**, enabling the models to capture temporal dependencies, seasonal patterns, and delayed climatic effects. Temporal variables, including **Month_Number** and **Quarter**, were also incorporated to represent recurring seasonal variations.

The final forecasting dataset consisted of **180 monthly observations** collected between **January 2011 and December 2025**. The integrated dataset combined climatic variables, macroeconomic indicators, engineered features, temporal variables, and the target variable into a unified multivariate framework for model training, testing, performance evaluation, and comparative analysis.

Table 1: Description of Variable Used in the Study

No.	Variable	Description	Category
1.	Paddy_Rice_Price	Monthly paddy rice market price	Target
2.	Rainfall_mm	Monthly rainfall amount (mm)	Climatic
3.	Temperature_C	Average monthly temperature (°C)	Climatic
4.	Relative_Humidity_%	Average monthly relative humidity (%)	Climatic
5.	Food_Inflation_YoY_%	Food inflation rate (%)	Macroeconomic
6.	Exchange_Rate_NGN_USD	Exchange rate (NGN/USD)	Macroeconomic
7.	Fertilizer_Price	Average fertilizer price	Macroeconomic
8.	Labour_Cost	Average labour cost	Macroeconomic
9.	PMS_Price	Petrol price per litre	Macroeconomic
10.	Market_Demand_Index	Market demand indicator	Macroeconomic
11.	Price_Lag_1	1-month lagged price	Engineered
12.	Price_Lag_2	2-month lagged price	Engineered
13.	Price_Lag_3	3-month lagged price	Engineered
14.	Price_MA_3	3-month moving average	Engineered
15.	Price_MA_6	6-month moving average	Engineered
16.	Rainfall_3M_Total	3-month cumulative rainfall	Engineered
17.	Rainfall_6M_Total	6-month cumulative rainfall	Engineered
18.	Season_Encoded	Encoded seasonal indicator	Engineered
19.	Month_Number	Month representation (1–12)	Temporal
20.	Quarter	Quarter representation (Q1, Q2, Q3, Q4)	Temporal
21.	Price_Movement	Forecast target variable (future price movement)	Output

Data Preprocessing and Feature Engineering

Before model development, the collected datasets were integrated into a unified multivariate dataset and subjected to a comprehensive preprocessing procedure to improve data quality and model performance. The preprocessing stage involved data cleaning, handling missing values, consistency checks, and normalization to ensure that all variables were suitable for machine learning analysis.

Categorical variables were transformed into numerical representations, while temporal variables were extracted from the date information to capture seasonal trends. Historical paddy rice prices and climatic variables were further used to generate engineered features that enhanced the predictive capability of the forecasting models.

Feature engineering included the creation of lagged price variables (**Price_Lag_1, Price_Lag_2,** and **Price_Lag_3**) to capture short-term temporal dependencies. Moving average variables (**Price_MA_3** and **Price_MA_6**) were computed to represent medium-term price trends, while cumulative rainfall variables (**Rainfall_3M_Total** and **Rainfall_6M_Total**) were generated to account for delayed climatic effects on paddy rice production. In addition, seasonal information was encoded into a numerical variable (**Season_Encoded**) to represent the rainy and dry seasons, and temporal variables (**Month_Number** and **Quarter**) were incorporated to model recurring seasonal patterns.

Following preprocessing and feature engineering, the dataset was normalized to reduce differences in variable scales before being partitioned into training and testing datasets for machine learning model development. This preprocessing framework improved data consistency and enabled the forecasting models to learn meaningful relationships between climatic conditions, macroeconomic indicators, and paddy rice price movements.

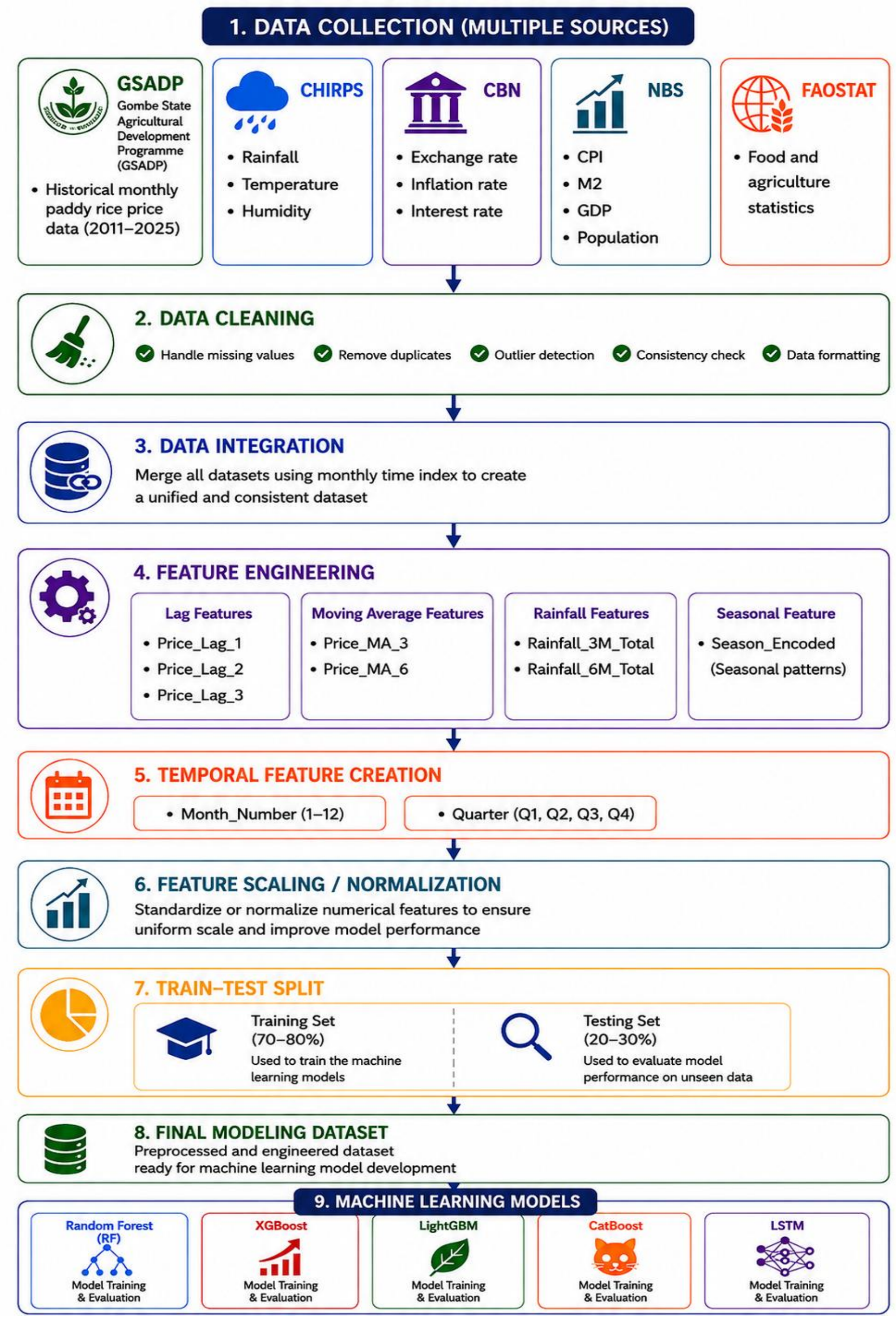

Figure 3: Data Preprocessing and Feature Engineering Workflow

Machine Learning Models

Five machine learning algorithms were developed and evaluated to forecast paddy rice price movements in Gombe State, Nigeria. The selected models comprised four tree-based ensemble learning algorithms, namely Random Forest (RF), Extreme Gradient Boosting (XGBoost), Light Gradient Boosting Machine (LightGBM), and Categorical Boosting (CatBoost), together with a deep learning model, Long Short-Term Memory (LSTM). These models were selected because of their proven capability to model nonlinear relationships, handle multivariate datasets, and learn complex temporal patterns from historical observations.

Random Forest (RF)

Random Forest is an ensemble learning algorithm that constructs multiple decision trees during training and combines their predictions to improve forecasting accuracy while reducing overfitting. The algorithm is effective for nonlinear regression problems because it captures complex interactions among predictor variables through bootstrap aggregation and random feature selection (Breiman, 2001).

Extreme Gradient Boosting (XGBoost)

XGBoost is a gradient boosting algorithm that sequentially builds decision trees by minimizing prediction errors from previous trees. It incorporates regularization techniques and parallel processing to improve prediction accuracy and computational efficiency, making it suitable for structured datasets with complex relationships (Chen & Guestrin, 2016).

Light Gradient Boosting Machine (LightGBM)

LightGBM is a gradient boosting framework that employs histogram-based learning and leaf-wise tree growth to accelerate model training while maintaining high predictive performance. Its ability to process large datasets efficiently makes it suitable for multivariate forecasting applications (Ke et al., 2017).

Categorical Boosting (CatBoost)

CatBoost is a gradient boosting algorithm designed to improve prediction accuracy while minimizing overfitting. It provides efficient handling of categorical variables and employs ordered boosting techniques to reduce prediction bias during model training (Prokhorenkova et al., 2018).

Long Short-Term Memory (LSTM)

Long Short-Term Memory is a recurrent neural network architecture developed to learn long-term dependencies in sequential data. Through its memory cells and gating mechanisms, LSTM effectively captures temporal relationships within time-series datasets, making it suitable for forecasting agricultural commodity prices (Hochreiter & Schmidhuber, 1997).

The five models were trained using the same preprocessed dataset and evaluated under identical experimental conditions to ensure a fair comparison. Their predictive performance was assessed using Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and the coefficient of determination (R^2), enabling the identification of the most suitable model for forecasting paddy rice price movements in Gombe State.

Hyperparameter Configuration of the Machine Learning Models

The predictive performance of machine learning algorithms depends on the appropriate selection of model hyperparameters. To improve the transparency and reproducibility of the proposed forecasting framework, the hyperparameter settings adopted for each machine learning model are presented in **Table 2**. The selected hyperparameters were determined through iterative experimentation and performance evaluation to achieve an appropriate balance between prediction accuracy, model complexity, and computational efficiency while minimizing the risk of overfitting.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VII July 2026

Table 2: Hyperparameter Configuration of the Machine Learning Models Used in Study

Models	Hyperparameter Configuration
Random Forest	Number of trees = 200; maximum depth = 6; minimum samples split = 8; minimum samples leaf = 4; maximum features = √(number of features); bootstrap sampling = enabled; random state = 42.
XGBoost	Number of trees = 300; learning rate = 0.03; maximum depth = 4; subsample = 0.80; column sample by tree = 0.80; objective function = squared error regression; random state = 42.
LightGBM	Number of trees = 300; learning rate = 0.03; maximum depth = 5; number of leaves = 31; subsample = 0.80; column sample by tree = 0.80; random state = 42.
CatBoost	Iterations = 300; learning rate = 0.03; tree depth = 5; loss function = RMSE; evaluation metric = RMSE; random seed = 42.
LSTM	Two LSTM layers (64 and 32 hidden units); dropout rate = 0.20; sequence length = 12 months; optimizer = Adam; loss function = Mean Squared Error (MSE); epochs = 100; batch size = 8.

The hyperparameter configurations presented in Table 2 were consistently applied during model training and evaluation to ensure a fair comparison among the five machine learning algorithms. The tree-based models were configured with controlled model complexity through appropriate limits on tree depth, sampling strategies, and learning rates, while the LSTM model incorporated dropout regularization and a two-layer network architecture to effectively capture temporal dependencies and improve model robustness. These configurations contributed to reducing overfitting while maintaining reliable predictive performance across all models.

Model Training and Validation

The preprocessed dataset was divided into training and testing subsets using a temporal split strategy to preserve the chronological order of the observations and prevent data leakage during model development. Historical observations collected before 2024 were used to train the machine learning models, while observations from 2024 onward were reserved for model testing and performance evaluation. This approach reflects recommended practice in time-series forecasting, where future observations are excluded from the training process to ensure an unbiased assessment of predictive performance (Hyndman & Athanasopoulos, 2021).

The chronological train–test partitioning adopted in this study further minimized the possibility of data leakage by ensuring that information from the testing period was completely excluded during model training. Preserving the temporal sequence of the observations prevented future information from influencing the learning process and enabled the developed models to be evaluated under realistic forecasting conditions.

To further reduce the risk of overfitting, appropriate regularization techniques were incorporated during model development. For the tree-based algorithms (Random Forest, XGBoost, LightGBM, and CatBoost), model complexity was controlled through carefully selected hyperparameters, including tree depth, learning rate, sub sampling and feature sampling strategies, minimum sample constraints, and other model-specific regularization parameters, thereby reducing the likelihood of overfitting while maintaining robust predictive performance. Similarly, the LSTM model incorporated dropout regularization between hidden layers to improve its ability to generalize to unseen observations by reducing excessive dependence on individual

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VII July 2026

neurons during training. Collectively, these strategies enhanced the robustness, generalization capability, and reliability of the developed forecasting models while ensuring a fair and consistent comparison of their predictive performance (Breiman, 2001; Chen & Guestrin, 2016; Hochreiter & Schmidhuber, 1997; Ke et al., 2017; Prokhorenkova et al., 2018).

The Random Forest, XGBoost, LightGBM, CatBoost, and Long Short-Term Memory (LSTM) models were trained using the same training dataset to ensure consistency in model development and comparison. Each model learned the relationships between historical paddy rice prices, climatic variables, macroeconomic indicators, temporal variables, and engineered features before generating predictions for the testing dataset.

Model performance was evaluated using the testing dataset to assess the ability of each algorithm to predict unseen observations. A common validation framework was adopted for all models to ensure a fair and reliable comparison of forecasting performance. The evaluation was based on Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and the coefficient of determination (R^2), which are widely used performance measures for regression and forecasting studies (Hyndman & Athanasopoulos, 2021).

Model Evaluation Metrics

The forecasting performance of the developed machine learning models was evaluated using three widely adopted statistical metrics: Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and the coefficient of determination (R^2). These evaluation metrics were selected because they provide complementary measures of forecasting accuracy by quantifying prediction errors and assessing the extent to which the developed models explain variations in paddy rice price movements. Collectively, the metrics provide a comprehensive basis for comparing the forecasting performance of the Random Forest, XGBoost, LightGBM, CatBoost, and Long Short-Term Memory (LSTM) models developed in this study (Hyndman & Athanasopoulos, 2021).

Root Mean Square Error (RMSE)

Root Mean Square Error (RMSE) was employed to evaluate the forecasting performance of the developed machine learning models in predicting paddy rice price movements. RMSE measures the average magnitude of prediction errors by calculating the square root of the mean squared differences between the actual and predicted values. Since the errors are squared before averaging, RMSE assigns greater weight to larger prediction errors, making it suitable for assessing the reliability of forecasting models where substantial deviations from the observed values are undesirable. In this study, lower RMSE values indicate better forecasting performance and greater prediction accuracy in forecasting paddy rice price movements.

$$RMSE = \sqrt{\left[\left(\frac{1}{n}\right)\Sigma(yi - \hat{y}i)^2\right]} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

Where:

- y= Actual value
- yi^ = Predicted value
- n = Number of observations

Lower RMSE values indicate better forecasting performance.

Mean Absolute Error (MAE)

Mean Absolute Error (MAE) was used to determine the average magnitude of prediction errors produced by the developed machine learning models in forecasting paddy rice price movements. MAE calculates the average absolute difference between the actual and predicted values, thereby providing a straightforward and easily interpretable measure of forecasting accuracy without assigning greater weight to larger errors. Consequently, lower MAE values indicate that the predicted values are closer to the observed values, reflecting better forecasting performance of the developed models.

INTERNATIONAL JOURNAL OF RESEARCH AND INNOVATION IN APPLIED SCIENCE (IJRIAS)
ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VII July 2026

$MAE = n1i = 1\sum n \mid yi - yi \mid$... (2)

Where:

- y_i= Actual value
- y_i^ = Predicted value
- n = Number of observations

Lower MAE values indicate higher forecasting accuracy.

Coefficient of Determination (R^2)

The coefficient of determination (R^2) was used to assess the predictive capability of the developed machine learning models by measuring the proportion of variation in paddy rice price movements explained by the forecasting model. It evaluates how well the predicted values correspond to the observed values, thereby indicating the model's ability to capture the underlying patterns in the data. Higher R^2 values indicate better model performance, stronger explanatory capability, and improved forecasting accuracy for paddy rice price movements.

$R2 = 1 - \sum(yi - \bar{y})2\sum(yi - yi)2$………………………………..………………… (3)

Where:

- y_i= Actual value
- y_i^ = Predicted value
- y= Mean of actual values

Higher R^2 values indicate better model performance and stronger explanatory capability.

Implementation Tools and Software Environment

The machine learning models were implemented in Python due to its flexibility, extensive libraries, and strong support for data analysis, machine learning, and predictive modeling (Witten et al., 2011). The implementation employed several Python libraries to support data preprocessing, feature engineering, model development, performance evaluation, and visualization. Specifically, Pandas and NumPy were used for data manipulation and numerical computations, while Scikit-learn was employed for data preprocessing, model training, and performance evaluation. The XGBoost, LightGBM, and CatBoost libraries were used to implement the respective ensemble learning models, whereas TensorFlow and Keras were used to develop the Long Short-Term Memory (LSTM) model. Matplotlib was utilized for visualizing model performance and forecasting results, while Jupyter Notebook served as the development environment for implementing, testing, and documenting the forecasting workflow. Collectively, these tools provided a robust computational environment for developing and evaluating the machine learning models for forecasting paddy rice price movements in Gombe State.

RESULTS

Comparative Performance of Machine Learning Models

The comparative performance of the developed machine learning models is presented in **Table 3**. Five models—Random Forest (RF), LightGBM, CatBoost, XGBoost, and Long Short-Term Memory (LSTM)—were evaluated using Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and the coefficient of determination (R^2). The comparative analysis was conducted to determine the most effective model for forecasting paddy rice price movements under varying climatic and macroeconomic conditions in Gombe State, Nigeria.

Table 3: Comparative Performance of Machine Learning Models

Model	RMSE (Lower is Better)	MAE (Lower is Better)	R^2 Score (Higher is Better)	Overall Rank
Random Forest	3.2572 (Best)	2.7340 (Best)	0.4786 (Best)	1
LightGBM	3.4387	2.7795	0.4189	2
CatBoost	3.4632	2.7642	0.4106	3
LSTM	3.5472	2.7648	0.3816	4
XGBoost	3.6317	2.9362	0.3518	5

The results indicate that the **Random Forest** model achieved the best overall forecasting performance among the evaluated machine learning models. It recorded the lowest Root Mean Square Error (RMSE = **3.2572**), the lowest Mean Absolute Error (MAE = **2.7340**), and the highest coefficient of determination (**R^2 = 0.4786**), demonstrating superior predictive capability for forecasting paddy rice price movements in Gombe State. These results suggest that the Random Forest model was more effective in capturing the nonlinear relationships and temporal dependencies present in the integrated climatic and macroeconomic dataset.

The superior performance of the Random Forest model may be attributed to its ensemble learning mechanism, which combines multiple decision trees to improve prediction accuracy, reduce overfitting, and enhance model generalization. These characteristics make Random Forest particularly suitable for agricultural commodity forecasting, where market behaviour is influenced by complex interactions among historical prices, climatic conditions, and macroeconomic variables. This finding is consistent with previous studies that identified Random Forest as one of the most robust machine learning algorithms for modeling nonlinear agricultural datasets (Breiman, 2001; Zhang et al., 2023).

LightGBM and CatBoost also demonstrated competitive forecasting performance, indicating that ensemble boosting algorithms are well suited to agricultural commodity forecasting. Their relatively close performance to the Random Forest model further supports the effectiveness of tree-based ensemble methods in modeling multivariate agricultural datasets influenced by climatic variability and macroeconomic conditions. Similar observations have been reported in previous studies on agricultural forecasting (Kumar & Singh, 2022).

The LSTM model achieved moderate forecasting performance (**R^2 = 0.3816**). Although LSTM networks are designed to capture temporal dependencies in sequential data, their predictive capability generally improves when trained using larger datasets with longer time horizons and higher-frequency observations. The moderate performance obtained in this study may therefore be associated with the size and temporal resolution of the monthly dataset used for model development.

Among the evaluated models, XGBoost recorded the lowest predictive performance. Nevertheless, it maintained acceptable forecasting capability, indicating that all five machine learning models were able to capture meaningful relationships within the dataset. Overall, the comparative analysis demonstrates that ensemble learning approaches, particularly Random Forest, provide a reliable framework for forecasting paddy rice price movements under varying climatic and macroeconomic conditions in Gombe State.

Random Forest Performance Analysis

Among the evaluated machine learning models, Random Forest demonstrated the best forecasting performance, achieving an RMSE of 3.2572, an MAE of 2.7340, and an R^2 value of 0.4786. These results indicate that the model provided the most accurate and reliable predictions of paddy rice price movements under the varying climatic and macroeconomic conditions considered in this study. The superior predictive performance of the Random Forest model justified its selection for further analysis.

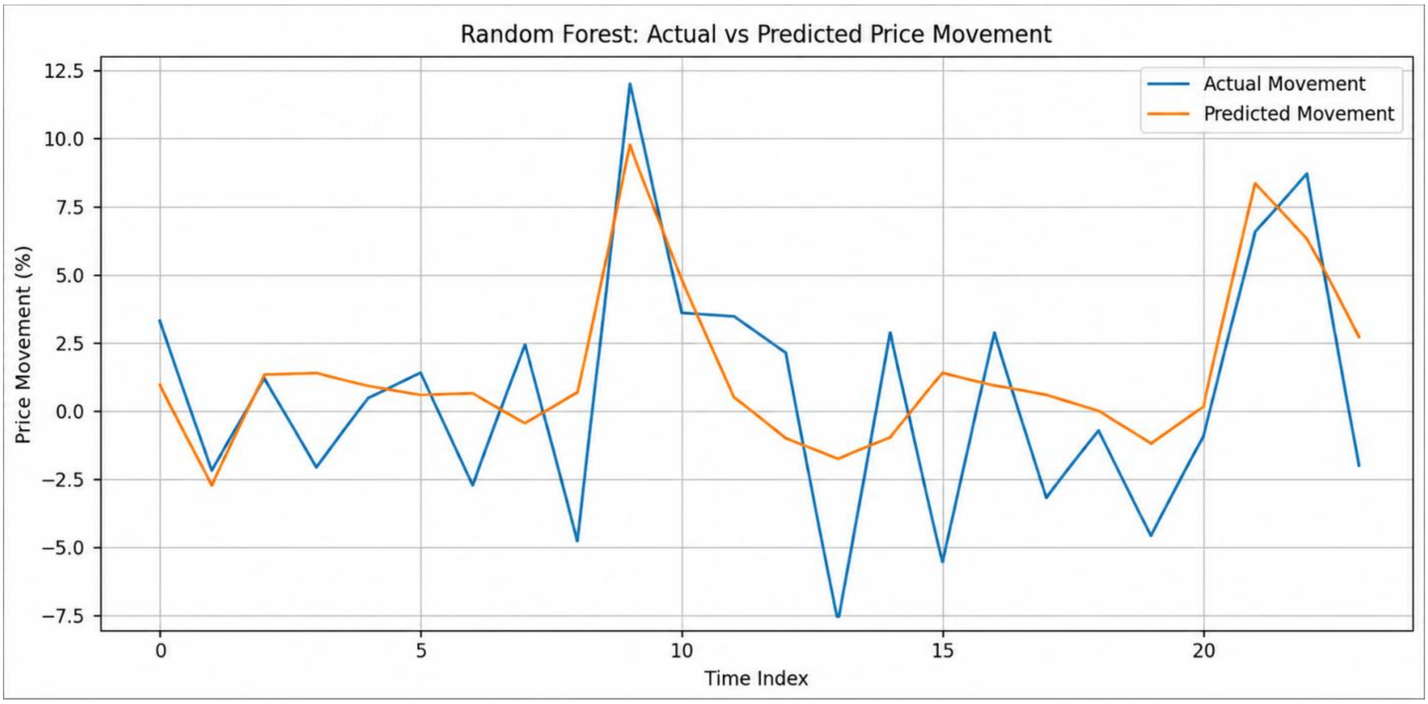

Figure 4: Actual versus Predicted Paddy Rice Price Movements Using the Random Forest Model

Figure 4 illustrates the relationship between the actual and predicted paddy rice price movements values generated by the Random Forest model. The predicted trend closely followed the observed movement pattern across most of the testing observations, indicating satisfactory forecasting capability and stable temporal learning behaviour. Although slight deviations were observed during periods of increased market volatility and abrupt price fluctuations, the model generally maintained good agreement with the actual market movement trend. These results suggest that the Random Forest model effectively captured the temporal relationships and seasonal dynamics influencing paddy rice price movements in Gombe State.

The close correspondence between the actual and predicted movement trends further demonstrates the model's ability to generalize effectively under changing climatic and macroeconomic conditions. The strong predictive performance may be attributed to the ensemble learning structure of the Random Forest algorithm, which combines multiple decision trees to improve forecasting stability while reducing overfitting. This characteristic makes the model particularly suitable for forecasting agricultural commodity prices, where nonlinear relationships and complex interactions among predictor variables are common (Breiman, 2001).

The forecasting results further indicate that the integrated climatic and macroeconomic variables used in this study contained meaningful predictive information for modeling paddy rice price movements in Gombe State. The relatively higher coefficient of determination achieved by the model suggests that a substantial proportion of the variability in paddy rice price movements was explained by the developed forecasting framework. Similar forecasting performance has been reported in agricultural machine learning studies employing ensemble learning techniques for multivariate forecasting systems (Kumar & Singh, 2022; Zhang et al., 2023).

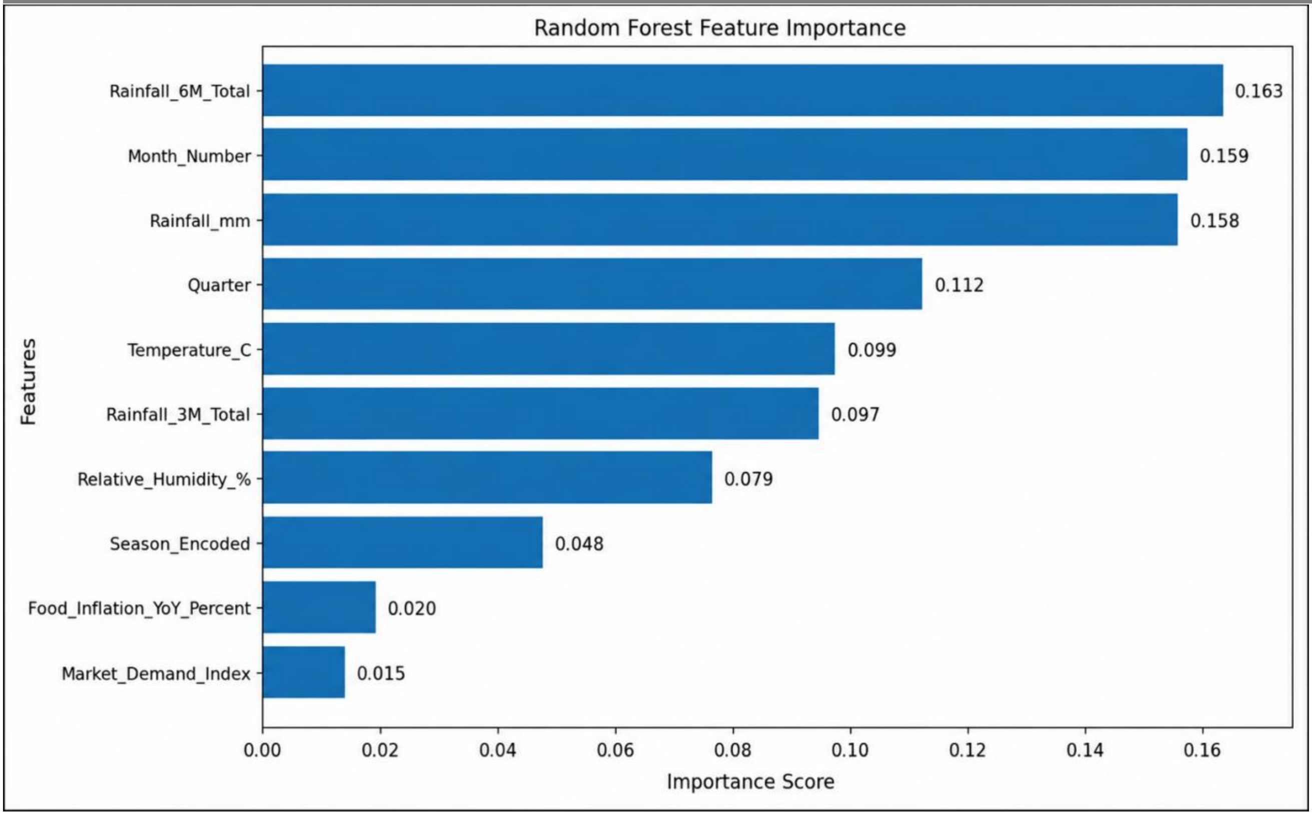

Figure 5: Random Forest Feature Importance Visualization

Figure 5 presents the relative importance of the variables utilized by the Random Forest model in forecasting paddy rice price movements. The results indicate that rainfall-related variables, seasonal indicators, temperature, and relative humidity were among the most influential predictors within the developed forecasting framework. The prominence of these variables suggests that environmental and seasonal conditions play a significant role in shaping paddy rice price movements in Gombe State.

Rainfall variability appears to be particularly important because it directly influences agricultural productivity, market supply, and commodity availability, which subsequently affect paddy rice price movements. Seasonal indicators, including month number and quarter variables, also demonstrated relatively high importance, confirming the existence of cyclical market behaviour associated with planting periods, harvest seasons, rainfall intensity, and commodity distribution patterns within the study area.

The feature importance analysis further demonstrates that integrating climatic variables with macroeconomic indicators provides a robust framework for forecasting paddy rice price movements. These findings are consistent with previous agricultural forecasting studies, which emphasize the importance of climatic and seasonal variables in commodity market prediction, particularly in regions where agricultural production is highly dependent on rainfall (FAO, 2022).

Beyond the dominant influence of rainfall and seasonal indicators, the feature importance results also highlight the contribution of other climatic variables, particularly temperature and relative humidity, to paddy rice price forecasting. These variables influence crop growth, water availability, and overall agricultural productivity, thereby affecting market supply and subsequent price fluctuations. In contrast, macroeconomic variables such as food inflation and the market demand index exhibited relatively lower importance within the Random Forest model. This suggests that, under the conditions of the study area, short-term paddy rice price movements are more strongly influenced by local climatic variability and seasonal production patterns than by broader macroeconomic conditions. From a practical perspective, these findings emphasize the importance of incorporating climatic and temporal information into agricultural price forecasting systems to support timely decision-making by farmers, traders, and policymakers. Such information can facilitate production planning, market monitoring, and policy interventions aimed at mitigating the adverse effects of price volatility on the paddy rice value chain (FAO, 2022; Kumar & Singh, 2022; Zhang et al., 2023).

DISCUSSION

The findings of this study demonstrate the effectiveness of machine learning techniques for forecasting paddy rice price movements in Gombe State using historical price data integrated with climatic and macroeconomic variables. Among the evaluated models, Random Forest consistently achieved the best predictive performance, followed by LightGBM and CatBoost, while LSTM and XGBoost produced comparatively lower forecasting accuracy. These results suggest that ensemble-based learning algorithms are more suitable for modeling the nonlinear relationships and multivariate interactions that characterize agricultural commodity markets.

The superior performance of the Random Forest model agrees with previous agricultural forecasting studies, which reported that ensemble learning methods perform effectively in environments characterized by nonlinear relationships, seasonal variability, and multivariate interactions. The ability of Random Forest to reduce overfitting through bootstrap aggregation and multiple decision trees contributed to its relatively stable forecasting performance under changing climatic and macroeconomic conditions (Breiman, 2001; Zhang et al., 2023).

The importance of climatic variables identified in this study is also consistent with previous agricultural forecasting research. Rainfall variability, temperature, relative humidity, and seasonal indicators emerged as influential predictors of paddy rice price movements. These findings indicate that climatic conditions influence agricultural productivity, market supply, and commodity availability, thereby affecting price movements within the study area. Similar observations have been reported in studies emphasizing the importance of climatic variables in agricultural commodity forecasting systems (FAO, 2022).

Macroeconomic variables also contributed meaningfully to the forecasting framework. Variables such as fertilizer prices, seed prices, labour costs, exchange rates, food inflation, and PMS prices reflect the broader economic environment affecting agricultural production and market dynamics. Their inclusion improved the ability of the forecasting models to capture market fluctuations beyond historical price behaviour alone. The findings further indicate that paddy rice price movements in Gombe State is influenced by the interaction of climatic variability, macroeconomic instability, and market conditions rather than by any single factor.

The relatively lower performance of the LSTM model may be attributed to the size and frequency of the dataset used in this study. Although LSTM networks are widely recognized for modeling sequential data, they generally require larger datasets and longer sequential depth to achieve optimal performance. Similarly, the comparatively lower performance of XGBoost may be associated with its sensitivity to localized fluctuations and abrupt market changes within the agricultural commodity market environment.

Overall, the findings confirm that integrating historical price information with climatic and macroeconomic variables within a machine learning framework improves the forecasting of paddy rice price movements in Gombe State. The developed forecasting framework provides a useful analytical approach for supporting agricultural planning, market monitoring, and evidence-based decision-making by farmers, traders, and policymakers operating within environments characterized by climatic variability and macroeconomic uncertainty.

Despite the promising performance of the developed forecasting framework, this study has some limitations. The analysis was based on monthly data collected for Gombe State between 2011 and 2025, which may not fully capture short-term market fluctuations or regional variations across Nigeria. In addition, factors such as government policy changes, pest and disease outbreaks, and international commodity market dynamics were not explicitly incorporated into the forecasting framework. Future research could extend the proposed approach by integrating additional socioeconomic and policy-related variables, utilizing higher-frequency datasets, and evaluating the model across multiple agricultural commodities and geographical regions to improve its applicability and forecasting accuracy.

CONCLUSION

This study developed and evaluated machine learning models for forecasting paddy rice movements in Gombe

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VII July 2026

State, Nigeria, using historical price data integrated with climatic and macroeconomic variables. Five machine learning algorithms—Random Forest, LightGBM, CatBoost, XGBoost, and Long Short-Term Memory (LSTM)—were comparatively evaluated using Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and coefficient of determination (R^2) as performance metrics.

The results demonstrated that the Random Forest model achieved the best overall forecasting performance, outperforming the other evaluated models in terms of prediction accuracy and generalization capability. The findings further showed that rainfall-related variables, seasonal indicators, temperature, relative humidity, and selected macroeconomic variables contributed significantly to forecasting paddy rice price movements, highlighting the importance of integrating multiple data sources in agricultural commodity price forecasting.

The developed forecasting framework demonstrates the potential of machine learning techniques to support agricultural planning and market decision-making under conditions characterized by climatic variability and macroeconomic uncertainty. The forecasting model provides a data-driven approach that can assist farmers, traders, and policymakers in anticipating paddy rice price movements and improving production, marketing, and resource allocation decisions within the study area.

Overall, the study confirms that integrating historical price information with climatic and macroeconomic variables within an ensemble machine learning framework enhances the prediction of paddy rice price movements in Gombe State and provides a useful foundation for future agricultural forecasting applications.

The findings of this study provide empirical evidence that machine learning techniques can serve as effective decision-support tools for agricultural commodity price forecasting in developing economies. By integrating climatic and macroeconomic variables into the forecasting framework, the study offers a practical approach for improving market planning, reducing uncertainty, and supporting evidence-based agricultural policies. The proposed framework may also serve as a useful reference for future research on commodity price forecasting in other regions and for other agricultural commodities.

REFERENCES

1. **Breiman, L.** (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324
2. **Central Bank of Nigeria.** (2025). *Statistical Bulletin*. https://www.cbn.gov.ng
3. **Chen, T., & Guestrin, C.** (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. https://doi.org/10.1145/2939672.2939785
4. **Dzarma, E. D., Nyor, N., Jiya, M., & Gana, A. S.** (2024). Optimization of rice yield in Gombe State, North-East Nigeria with machine learning. *Adeleke University Journal of Engineering and Technology, 7*(2), 65–68.
5. **Food and Agriculture Organization of the United Nations (FAO).** (2022). *The State of Food and Agriculture 2022: Leveraging automation in agriculture for transforming agrifood systems*. FAO. https://doi.org/10.4060/cb9479en
6. **Food and Agriculture Organization of the United Nations (FAO).** (2025). *FAOSTAT Statistical Database*. https://www.fao.org/faostat/
7. **Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A., & Michaelsen, J.** (2015). The Climate Hazards InfraRed Precipitation with Station Data (CHIRPS): A new environmental record for monitoring extremes. *Scientific Data, 2*, 150066. https://doi.org/10.1038/sdata.2015.66
8. **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
9. **Hyndman, R. J., & Athanasopoulos, G.** (2021). *Forecasting: Principles and Practice* (3rd ed.). OTexts. https://otexts.com/fpp3/
10. **Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T. Y.** (2017). LightGBM: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems, 30*, 3146–3154.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VII July 2026

11. **Kumar, A., & Singh, R.** (2022). Machine learning approaches for agricultural commodity price forecasting under climatic variability. *Computers and Electronics in Agriculture, 198*, 107017. https://doi.org/10.1016/j.compag.2022.107017
12. **National Bureau of Statistics.** (2025). *Selected Food Price Watch*. https://nigerianstat.gov.ng
13. **Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A.** (2018). CatBoost: Unbiased boosting with categorical features. *Advances in Neural Information Processing Systems, 31*, 6638–6648.
14. **Sanusi, O. I., Safi, S. K., Adeeko, O., & Tabash, M. I.** (2022). Forecasting agricultural commodity price using different models: A case study of widely consumed grains in Nigeria. *Agricultural and Resource Economics: International Scientific E-Journal, 8*(2), 124–140. https://doi.org/10.51599/are.2022.08.02.07
15. **Sari, M., Duran, S., Kutlu, H., et al.** (2024). Various optimized machine learning techniques to predict agricultural commodity prices. *Neural Computing and Applications, 36*, 11439–11459. https://doi.org/10.1007/s00521-024-09679-x
16. **Witten, I. H., Frank, E., & Hall, M. A.** (2011). *Data Mining: Practical Machine Learning Tools and Techniques* (3rd ed.). Morgan Kaufmann.
17. **Zhang, T., & Tang, Z.** (2024). Agricultural commodity futures prices prediction based on a new hybrid forecasting model combining quadratic decomposition technology and LSTM model. *Frontiers in Sustainable Food Systems, 8*, Article 1334098. https://doi.org/10.3389/fsufs.2024.1334098
18. **Zhang, Y., Wang, H., Li, X., & Chen, J.** (2023). Random Forest-based forecasting of agricultural commodity prices using climatic and economic variables. *Agricultural Systems, 210*, 103707. https://doi.org/10.1016/j.agsy.2023.103707