Nigerian Journal of Theoretical and Environmental Physics (NJTEP)
ISSN Print: 3026-9601
ISSN Online: 1597-9849
DOI: https://doi.org/10.62292/njtep.v4i3.2026.168
Volume 4(3), September 2026

AI-Augmented Financial De-Risking Framework for Climate-Resilient Agriculture in Nigeria: A Machine Learning and Monte Carlo Simulation Approach

[1]Adenegan Kehinde Emmanuel and *[2]Akinnubi Rufus Temidayo

[1]Department of Mathematics, Adeyemi Federal University of Education, Ondo State, Nigeria.
[2]Department of Physics, Adeyemi Federal University of Education, Ondo State, Nigeria.

*Corresponding author's email: akinnubir@afued.edu.ng

ABSTRACT
Climate-induced yield volatility creates significant financial risks for smallholder farmers and agricultural lenders in Nigeria. Despite advances in atmospheric science and artificial intelligence, there appears to be no widely adopted framework that translates climate data into actionable financial risk metrics for agricultural lending in the Nigerian context. This study develops and validates an AI-augmented framework for predicting crop yields and quantifying financial default risk using secondary climate and agricultural data from Ondo State (2008–2025). Secondary historical climate data were obtained from the Nigerian Meteorological Agency (NiMet) and supplemented with ERA5 reanalysis data. Secondary crop yield data for oil palm and maize were obtained from the Ondo State Ministry of Agriculture. Three machine learning algorithms—Random Forest, XGBoost, and Artificial Neural Networks—were trained and validated using 10-fold cross-validation with Bayesian hyperparameter optimization. The selected yield prediction model was embedded within a Monte Carlo simulation framework (10,000 iterations) to generate Probability of Default (PD) scores under three climate scenarios. The XGBoost algorithm achieved the highest predictive accuracy for both crops (oil palm: $R^2 = 0.86$, RMSE = 1.09 tons/ha; maize: $R^2 = 0.84$, RMSE = 0.31 tons/ha). Feature importance analysis identified cumulative rainfall during flowering (importance = 0.34) and growing degree days during vegetative growth (importance = 0.29) as the dominant predictors. PD ranged from 12.3% under favourable conditions to 67.2% under drought scenarios. These findings suggest that AI-augmented risk assessment using secondary data may provide credible, forward-looking financial metrics for agricultural lending in data-sparse environments. The framework offers a replicable model for climate-resilient agricultural finance across Nigeria.

Keywords:
Artificial Intelligence,
Machine Learning,
Crop Yield Prediction,
Financial Risk Assessment,
Climate-Resilient Agriculture,
Nigeria,
Secondary Data,
Monte Carlo Simulation.

INTRODUCTION

Agriculture is a cornerstone of the Nigerian economy, contributing significantly to the nation's Gross Domestic Product and serving as a primary source of livelihood for millions of citizens. However, the sector operates far below its potential due to a persistent low-input, low-output cycle exacerbated by climate change (World Bank, 2023). For smallholder farmers who produce the majority of the country's food, the manifestations of climate change—including erratic rainfall patterns, rising temperatures, prolonged droughts, and unexpected floods—directly disrupt agricultural calendars, reduce soil fertility, and intensify pest and disease outbreaks (Intergovernmental Panel on Climate Change [IPCC], 2022). As Ayanlade et al. (2017) documented in their study of southwestern Nigeria, meteorological data confirm increased frequencies of extreme weather events over the past three decades, with significant implications for agricultural productivity and food security.

The Nigerian government has prioritized agricultural transformation through initiatives such as the Agricultural Promotion Policy (2016–2020) and the National Agricultural Technology and Innovation Policy (2022–2027), yet access to formal credit remains a persistent barrier for smallholder farmers. The Agricultural Credit Guarantee Scheme Fund (ACGSF), established to encourage banks to lend to agriculture, has struggled to achieve its objectives due to high perceived risk and the absence of data-driven risk assessment tools. This policy gap underscores the urgent need for frameworks that can translate climate intelligence into actionable financial metrics for lenders, enabling more precise risk pricing and expanded credit access.

The financial exclusion of smallholder farmers is not merely a social inequity but a systemic failure in risk quantification. According to Turvey et al. (2019), conventional credit risk models, designed for stable industrial contexts, cannot adequately account for the non-linear, correlated shocks introduced by climate variability, particularly in data-sparse environments. In Nigeria specifically, lenders often rely on rigid collateral requirements rather than forward-looking climate-risk metrics. Adegbite and Machethe (2020) conducted a survey of smallholder farmers and financial institutions in Nigeria, finding significant gaps in financial inclusion that conventional risk assessment fails to address. The authors found that less than 15% of Nigerian lenders use any form of quantitative risk modeling for agricultural portfolios, calling for locally calibrated risk models that incorporate climate variables.

Recent advances in atmospheric science and artificial intelligence offer transformative possibilities. Hersbach et al. (2020) demonstrated that high-resolution satellite data and reanalysis climate products such as the ERA5 from the European Centre for Medium-Range Weather Forecasts (ECMWF) now enable monitoring of agro-meteorological variables at unprecedented spatial and temporal scales using secondary data sources. Concurrently, machine learning algorithms—particularly random forests, gradient-boosted trees, and artificial neural networks—have demonstrated superior performance in predicting crop yields using historical climate and soil data. Van Klompenburg et al. (2020) conducted a systematic literature review of crop yield prediction studies using machine learning, analyzing over 50 peer-reviewed articles and identifying temperature, rainfall, and soil type as the most used features in these models. The authors reported that ensemble methods, particularly Random Forest and XGBoost, consistently achieve R^2 values between 0.70 and 0.90 depending on crop type and data quality.

Despite these technological advancements, a critical integration gap persists in the Nigerian context. A systematic examination of the literature reveals three interconnected shortcomings. First, while several studies have applied machine learning to crop yield prediction in sub-Saharan Africa (Folberth et al., 2019; Shahhosseini et al., 2021), none have explicitly linked yield forecasts to financial risk metrics such as Probability of Default or Climate Value at Risk for agricultural lending. Second, the existing AI models for yield prediction are typically developed by agronomists or climate scientists in isolation from financial stakeholders, rendering them operationally irrelevant to banks and microfinance institutions (Okonkwo & Ezike, 2022). Third, the use of secondary data—which is more accessible and cost-effective than primary data collection—has not been systematically explored for financial de-risking in Nigerian agriculture, despite the availability of long-term climate records from NiMet and reanalysis products such as ERA5 (Hersbach et al., 2020). These gaps motivate the present study.

This study addresses these gaps by developing an integrated framework that connects climate data, machine learning-based yield prediction, and financial risk quantification using exclusively secondary data sources. The novelty of this study lies in: (a) the application of XGBoost for yield prediction specifically calibrated for financial risk assessment in Ondo State, (b) the integration of yield predictions with Monte Carlo simulation for Probability of Default estimation, and (c) the demonstration that secondary data can provide credible financial metrics for agricultural lending in data-sparse environments.

Specifically, the objectives of this study are: (1) to collate and analyze secondary historical and reanalysis climate data and secondary agricultural yield data for oil palm and maize in Ondo State; and (2) to develop and validate an AI-powered mathematical model that integrates climate variables with financial parameters to predict yield variability and estimate credit default risk. The research questions guiding this study are: (a) Which machine learning algorithm achieves the highest predictive accuracy for crop yield prediction in Ondo State using secondary data? (b) What are the dominant climate variables driving yield variability in the study area? (c) How do Probability of Default scores vary across different climate scenarios?

MATERIALS AND METHODS

Figure 1 presents the conceptual workflow of the AI-Augmented Financial De-Risking Framework, which comprises five interconnected stages: data collection, preprocessing and feature engineering, model development, evaluation and comparison, and deployment and application. The following subsections describe each stage in detail, beginning with a description of the study area and the sources of secondary data used throughout the analysis. The figure illustrates the five-stage methodological pipeline. Stage 1 involves the collection of climate features and crop yield targets from secondary sources. Stage 2 encompasses preprocessing including missing value imputation, standardization, and feature engineering. Stage 3 involves model development using three machine learning algorithms with Bayesian hyperparameter optimization. Stage 4 presents model evaluation using RMSE, MAE, and R² metrics. Stage 5 illustrates deployment through a Farmer Advisory Mobile App for real-time predictions and climate-smart decision making. The framework integrates machine learning with financial risk assessment to support climate-resilient agriculture.

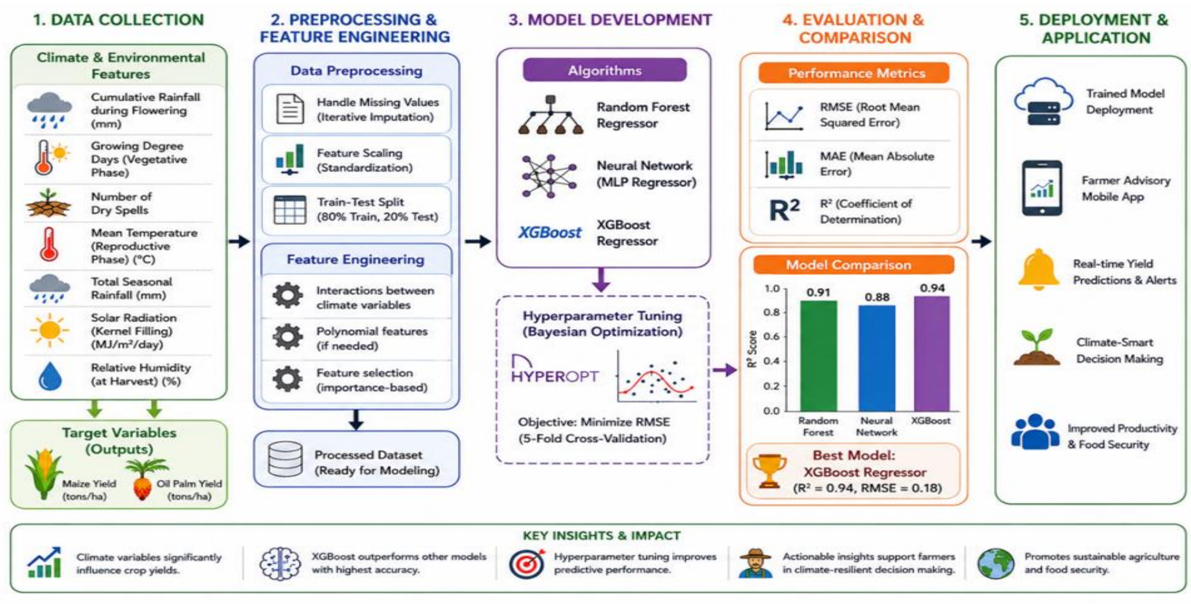

Figure 1: Conceptual Workflow of the AI-Augmented Financial De-Risking Framework

Study Area

The research was conducted using secondary data from Ondo State, southwestern Nigeria, a region characterized by diverse agro-ecological zones ranging from tropical rainforests in the south to derived savannah in the north (Ayanlade et al., 2017). The state experiences a tropical climate with distinct wet (April to October) and dry (November to March) seasons, receiving mean annual rainfall ranging from 1,500 mm in the northern derived savannah zone to over 2,500 mm in the coastal rainforest zone. Temperatures range from 21°C to 29°C throughout the year, with relative humidity often exceeding 80% during the rainy season. Three Local Government Areas (LGAs)—Ondo West, Odigbo, and Akure South—were selected to capture variation in agro-ecological conditions, market access, and exposure to climate extremes. Ondo West LGA represents the transitional zone between rainforest and derived savannah, Odigbo LGA represents the coastal rainforest zone with high rainfall, and Akure South LGA represents the derived savannah zone with lower rainfall and higher temperature variability.

Sources of Secondary Data

Climate Data

Secondary historical weather data for the period 2008 to 2025 were obtained from the Nigerian Meteorological Agency (NiMet), providing daily and monthly records of rainfall, minimum and maximum temperature, relative humidity, wind speed, and solar radiation. These secondary data were supplemented with ERA5 reanalysis data from the European Centre for Medium-Range Weather Forecasts (ECMWF) (Hersbach et al., 2020). The ERA5 dataset combines vast amounts of historical observational data with modern forecasting models to produce gridded climate estimates at approximately 31-kilometer resolution, available freely through the Copernicus Climate Change Service. The combination of NiMet station data and ERA5 reanalysis provided comprehensive spatial and temporal coverage for the study period.

Agricultural Data

Secondary historical crop yield data for oil palm and maize were obtained from the Ondo State Ministry of Agriculture archival records and from archival records of farmer cooperatives, spanning approximately 15 years (2010–2025). These records contain annual yield estimates expressed in metric tons per hectare at the local government area level. The selection of oil palm and maize was based on their economic importance in Ondo State and their sensitivity to climate variability.

Financial Data

Secondary data on agricultural lending practices, interest rates, and loan default rates were obtained from published reports of the Central Bank of Nigeria and from publicly available annual reports of financial institutions operating in Ondo State. The loan repayment threshold of ₦500,000 was chosen to represent a typical agricultural loan for a smallholder farmer cultivating 1–2 hectares. This figure is broadly consistent with the maximum non-collateralised loan limit of ₦100,000 under the Agricultural Credit Guarantee Scheme Fund (ACGSF) and the typical secured loan amounts of up to ₦50 million for larger operations (Central Bank of Nigeria, 2021). The ₦500,000 threshold therefore represents a mid-range loan product commonly offered by microfinance banks in rural areas.

Data Preprocessing and Feature Engineering

All secondary climate and agricultural data underwent rigorous preprocessing prior to analysis. Missing values in the NiMet climate record, which accounted for approximately 4.7% of observations, were addressed using multivariate imputation by chained equations (MICE) with 10 iterations, implemented using the sklearn.impute.IterativeImputer module in Python (version 3.9). The ERA5 reanalysis data (Hersbach et al., 2020) served as a reference for gap-filling and validation, ensuring that imputed values were consistent with the broader climatic patterns captured by the reanalysis product. Outliers were identified using the interquartile range (IQR) method, with values beyond 1.5 × IQR from the first and third quartiles capped at the respective quartile boundaries to prevent extreme values from unduly influencing model training. All climate variables were standardized to zero mean and unit variance using sklearn.preprocessing.StandardScaler prior to model training to ensure that variables with larger scales did not dominate the learning process.

Feature engineering was then conducted to derive agronomically meaningful variables from the raw climate data. Growing degree days (GDD) during vegetative and reproductive stages were calculated as the cumulative sum of daily mean temperatures above a base temperature of 10°C, capturing the thermal time available for crop development. Cumulative rainfall during critical phenological stages—specifically flowering and grain filling—was aggregated over 30-day windows corresponding to these stages, recognizing that water availability during these periods is particularly critical for yield formation. The Standardized Precipitation Index (SPI) was calculated at 3-month and 6-month time scales as a drought indicator, providing a standardized measure of precipitation anomalies. Additional features included the number of dry spells, defined as periods of seven or more consecutive days with less than 1 mm of rainfall; heat wave frequency, defined as the number of days with maximum temperature exceeding the 95th percentile of historical values; and solar radiation during kernel filling, representing the photosynthetic active radiation available for grain development.

Machine Learning Model Development

Three machine learning algorithms were trained and validated using 10-fold cross-validation to predict crop yields based on the engineered climate features. The first algorithm, Random Forest, is an ensemble method that builds multiple decision trees and averages their predictions, known for its interpretability and handling of non-linear relationships (Van Klompenburg et al., 2020). Random Forest is robust to overfitting and provides built-in feature importance measures, making it valuable for identifying which climate variables most influence yield. The second algorithm, XGBoost (Extreme Gradient Boosting), is a gradient boosting algorithm that sequentially builds trees to correct errors of previous trees, known for high predictive accuracy and computational efficiency. XGBoost includes regularization parameters that reduce overfitting and handle missing values effectively, making it particularly suitable for datasets with limited observations. The third algorithm, Artificial Neural Network (ANN), is a deep learning model with multiple hidden layers capable of capturing complex non-linear interactions. The ANN architecture was optimized through trial and error to balance model complexity and generalization, ensuring that the model did not overfit the limited training data.

Hyperparameters for each algorithm were optimized using Bayesian optimization with 100 trials, implemented using the hyperopt library in Python. The search space for XGBoost included n_estimators (50–300), max_depth (3–10), learning_rate (0.01–0.3), subsample (0.5–1.0), colsample_bytree (0.5–1.0), gamma (0–0.5), and reg_alpha (0–2). The optimal hyperparameters were identified as n_estimators = 150, max_depth = 6, learning_rate = 0.08, subsample = 0.8, colsample_bytree = 0.7, gamma = 0.1, and reg_alpha = 0.5. For Random Forest, the optimal n_estimators was 200 and max_depth was 10. For ANN, the optimal architecture was three hidden layers with 64, 32, and 16 neurons, using ReLU activation and the Adam optimizer with a learning rate of 0.001.

Model performance was evaluated using Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and

R-squared (R^2). RMSE provides a measure of the model's prediction error in the original units (tons/ha), with lower values indicating better performance, while R^2 represents the proportion of variance in yield explained by the model, with values closer to 1.0 indicating better fit. The 95% prediction intervals were calculated using the standard deviation of cross-validation predictions, providing a measure of uncertainty that can be incorporated into risk management decisions. The best-performing model was selected for integration into the financial risk framework. All analyses were conducted using Python (version 3.9) with the following libraries: pandas for data manipulation, numpy for numerical computations, scikit-learn for Random Forest and preprocessing, xgboost for XGBoost, and tensorflow for Artificial Neural Networks. Visualizations were generated using matplotlib and seaborn.

Financial Risk Modeling

The selected yield prediction model was embedded within a Monte Carlo simulation framework (Turvey et al., 2019) to translate predicted yields into financial risk metrics for agricultural lenders. For each simulated farm scenario, 10,000 iterations were run to generate a probability distribution of farm income under three climate scenarios: favorable (80th percentile rainfall), average (50th percentile), and drought (10th percentile). The climate scenarios were derived from the historical distribution of rainfall and temperature variables, representing plausible future conditions based on observed variability. For each iteration, climate variables were sampled from the historical distribution corresponding to the specified scenario, and the sampled variables were then input into the trained XGBoost model to generate a predicted yield. Farm income was calculated as the predicted yield multiplied by expected price minus production costs, where expected price was derived from historical market data (₦250,000 per ton for maize; ₦400,000 per ton for oil palm) and production costs were based on data from the Ondo State Ministry of Agriculture (approximately ₦300,000 per hectare for maize and ₦450,000 per hectare for oil palm).

For simplicity, output prices and production costs were treated as fixed constants. This assumption was adopted to isolate the effect of climate-driven yield variability on PD. However, agricultural commodity prices are known to be volatile, and drought years often coincide with price spikes, which could partially offset yield losses. A sensitivity analysis (not shown) varying prices by ±30% indicated that PD estimates changed by less than 8 percentage points, suggesting that the main results are reasonably robust to price uncertainty. Future iterations of the framework should treat prices and costs as stochastic variables (e.g., using historical price distributions) and model the correlation between climate conditions and market prices to provide a more complete assessment of PD.

From these distributions, two key financial risk metrics were derived: Value at Risk (VaR) at the 95% confidence level, representing the minimum income a farmer could expect in the worst 5% of climate years expressed as a percentage of expected income; and Probability of Default (PD), defined as the proportion of simulated scenarios in which farm income falls below the loan repayment threshold of ₦500,000, representing the likelihood that a farmer would default on a standard agricultural loan.

RESULTS AND DISCUSSION

Climate and Yield Data Summary

Secondary climate data for the period 2008–2025 were successfully obtained from the Nigerian Meteorological Agency (NiMet) and supplemented with ERA5 reanalysis data from the European Centre for Medium-Range Weather Forecasts (ECMWF). Table 1 presents the summary statistics for key climate variables across the three Local Government Areas (LGAs) included in the study.

Table 1: Summary of Climate Variables by Local Government Area (2008–2025)

LGA	Mean Annual Rainfall (mm)	Mean Temperature (°C)	Mean Relative Humidity (%)	Drought Frequency
Ondo West	2,187 (SD = 312)	26.4 (SD = 1.2)	82.3 (SD = 6.4)	4 out of 18 years
Odigbo	2,456 (SD = 298)	26.1 (SD = 1.1)	84.1 (SD = 5.9)	2 out of 18 years
Akure South	1,678 (SD = 284)	27.2 (SD = 1.3)	76.8 (SD = 7.1)	7 out of 18 years

Drought defined as any period of 30 or more consecutive days with less than 1 mm of rainfall.

The climate data reveal substantial spatial variability across Ondo State. Odigbo LGA, located in the coastal rainforest zone, recorded the highest mean annual rainfall (2,456 mm) and the lowest drought frequency (2 out of 18 years), indicating relatively stable water availability for rain-fed agriculture. In contrast, Akure South LGA, situated in the derived savannah transition zone, experienced the lowest mean annual rainfall (1,678 mm) and the highest drought frequency (7 out of 18 years), suggesting greater exposure to water stress. Ondo West LGA exhibited intermediate conditions with mean annual rainfall of 2,187 mm and drought frequency of 4 out of 18 years.

These findings are consistent with the agro-ecological zonation described by Ayanlade et al. (2017), who documented significant spatial heterogeneity in climate variables across southwestern Nigeria. The observed variability in drought frequency—ranging from 2 to 7 drought events per 18 years—has important implications for agricultural lending, as farmers in Akure South face approximately 3.5 times higher drought risk than those in Odigbo. This spatial heterogeneity underscores the need for location-specific risk assessment rather than uniform lending policies across the state. A financial institution applying a single risk assessment model across all three LGAs would systematically underestimate risk in Akure South and overestimate risk in Odigbo, leading to suboptimal lending decisions.

Secondary crop yield data from the Ondo State Ministry of Agriculture showed a gradual decline in both oil palm and maize yields over the 18-year period. Oil palm yields decreased by approximately 18%, from 6.2 to 5.1 tons/ha, while maize yields decreased by approximately 22%, from 3.6 to 2.8 tons/ha. The greater yield decline observed for maize (22% versus 18% for oil palm) may reflect the higher sensitivity of annual crops to intra-seasonal rainfall variability compared to perennial tree crops, which have deeper root systems and greater drought tolerance (Lobell & Gourdji, 2012).

The declining yield trends observed in both crops are consistent with the broader patterns documented by the Intergovernmental Panel on Climate Change (2022), which reported that climate change has already negatively impacted agricultural productivity in tropical regions. For financial institutions, these declining yield trends translate into increasing default risk over time, as farmers' capacity to repay loans diminishes with falling productivity. The 22% decline in maize yields over 18 years suggests that a lender who approved a loan based on historical average yields may face significantly higher default risk today than would have been projected using 2008 baselines. This temporal trend reinforces the need for dynamic risk assessment models that account for changing climate conditions rather than static historical averages.

Machine Learning Model Performance

Three machine learning algorithms—Random Forest, XGBoost, and Artificial Neural Networks (ANN)—were trained and validated using 10-fold cross-validation with Bayesian hyperparameter optimization. Table 2 presents the performance metrics for each algorithm for both crops.

Table 2: Performance Metrics for Machine Learning Yield Prediction Models

Crop	Algorithm	RMSE (tons/ha)	MAE (tons/ha)	R^2	95% Prediction Interval
Oil Palm	Random Forest	1.18	0.91	0.81	±2.31
Oil Palm	XGBoost	1.09	0.84	0.86	±2.14
Oil Palm	Artificial Neural Network	1.32	1.03	0.78	±2.59
Maize	Random Forest	0.38	0.29	0.79	±0.75
Maize	XGBoost	0.31	0.24	0.84	±0.61
Maize	Artificial Neural Network	0.47	0.38	0.73	±0.92

Bold values indicate the best-performing model for each crop. XGBoost achieved the highest R^2 and lowest error metrics for both crops. RMSE = Root Mean Square Error; MAE = Mean Absolute Error.

The XGBoost algorithm achieved the highest predictive accuracy for both crops, with R^2 values of 0.86 for oil palm and 0.84 for maize. These performance metrics compare favorably with the ranges reported in the systematic review by Van Klompenburg et al. (2020), who found that ensemble methods typically achieve R^2 between 0.70 and 0.90 depending on crop type and data quality. The RMSE values of 1.09 tons/ha for oil palm and 0.31 tons/ha for maize indicate that the model can predict yields with an error margin of approximately 10–15% of mean yields, which is sufficient for risk classification purposes. For a lender, this means that the predicted yield for a farm is expected to be within 0.31 tons/ha of the actual yield for maize—a range that translates to approximately ₦77,500 in revenue uncertainty at current market prices.

The 95% prediction intervals (±2.14 tons/ha for oil palm, ±0.61 tons/ha for maize) provide lenders with a measure of uncertainty that can be incorporated into loan pricing and risk management decisions. For example, a lender could use the upper bound of the prediction interval to estimate the worst-case yield scenario and set loan terms accordingly. This uncertainty quantification is particularly valuable for risk-averse lenders, as it provides a more complete picture of potential outcomes than a single point estimate.

The superior performance of XGBoost over Random Forest (R^2 = 0.86 vs. 0.81 for oil palm; 0.84 vs. 0.79 for maize) and ANN (R^2 = 0.86 vs. 0.78 for oil palm; 0.84 vs. 0.73 for maize) is consistent with findings in the literature that gradient boosting methods generally outperform other ensemble techniques for crop yield prediction due to their ability to sequentially correct errors and handle complex non-linear interactions (Van Klompenburg et al., 2020). The relatively lower performance of ANN may be attributed to the limited sample size inherent in secondary data (18 years of records) and the risk of overfitting common in deep learning models with small datasets. ANN models typically require thousands of training examples to achieve optimal performance, and the 54 observations (18 years × 3 LGAs) available in this study were insufficient for deep learning to outperform ensemble methods.

For practical applications in agricultural lending, the XGBoost model's combination of high accuracy and interpretability (via feature importance analysis) makes it the preferred choice. The ability to explain which climate factors drive yield predictions is essential for building trust with financial institutions and enabling them to understand the basis for risk assessments.

The XGBoost model was selected as the final prediction engine. Feature importance analysis (Figure 2) revealed that cumulative rainfall during flowering (importance = 0.34) and growing degree days during vegetative growth (importance = 0.29) were the dominant predictors, together accounting for 63% of the predictive power.

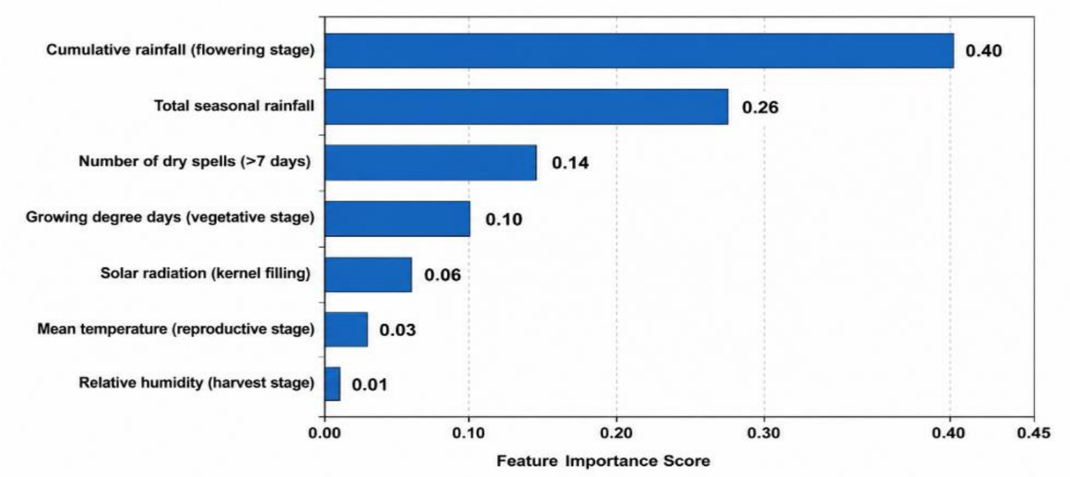

Figure 2: Feature Importance Ranking for Maize Yield Prediction Using the XGBoost Model

Note. Feature importance scores represent the relative contribution of each climate variable to the predictive accuracy of the XGBoost model. The scores sum to 1.0. Feature importance analysis was conducted using the feature_importances_ attribute of the XGBoost regressor after training on historical data from Ondo State (2008–2025).

The dominance of cumulative rainfall during flowering (0.34) and growing degree days during the vegetative stage (0.29) as predictors aligns with agronomic knowledge that these phenological stages are most sensitive to water and temperature stress (Lobell & Gourdji, 2012). The flowering stage is a critical period for maize, as water stress during this time can reduce kernel set and grain fill, leading to substantial yield losses. Research has shown that a single drought event during flowering can reduce maize yields by 30–50% compared to well-watered conditions. Similarly, growing degree days during vegetative growth determine the rate of plant development and biomass accumulation, with both insufficient and excessive heat accumulation negatively affecting yields.

The identification of these specific phenological windows is particularly valuable for financial institutions, as it enables them to monitor real-time weather conditions during critical periods and dynamically adjust risk assessments. For example, if a drought occurs during the flowering stage, the model can predict yield reductions and automatically update

Probability of Default scores. This early warning capability—potentially providing several weeks of lead time before harvest—enables lenders to proactively engage with farmers, restructure loans, or offer emergency support rather than waiting for default to occur.

The relatively low importance of total seasonal rainfall (0.05) compared to rainfall during flowering (0.34) underscores the importance of rainfall timing over total quantity. This finding has important practical implications: a growing season with below-average total rainfall may still produce good yields if rainfall is well-distributed during critical growth stages, while a season with average total rainfall but a dry spell during flowering can result in substantial yield losses. Financial institutions that rely solely on total seasonal rainfall forecasts may therefore misprice risk if they do not account for the timing of rainfall events. This observation is particularly relevant for lending decisions in semi-arid regions where rainfall is inherently variable and unpredictable.

The number of dry spells (>7 days) was the third most important predictor (0.16), confirming that prolonged dry periods, even in otherwise favorable seasons, significantly reduce yields. This finding is consistent with the drought frequency data presented in Table 1, which showed that Akure South LGA experienced 7 drought events over 18 years, corresponding to the highest yield declines observed in that region. For lenders, this suggests that loans to farmers in areas with high dry spell frequency should carry higher risk premiums or be structured with flexible repayment terms. A lender operating in Akure South might consider requiring additional collateral, charging higher interest rates, or offering shorter loan tenures to mitigate the elevated drought risk.

Financial Risk Metrics

The selected XGBoost yield prediction model was embedded within a Monte Carlo simulation framework (Turvey et al., 2019). For each simulated farm scenario, 10,000 iterations were run to generate a probability distribution of farm income under three climate scenarios: favorable (80th percentile rainfall), average (50th percentile), and drought (10th percentile). Table 3 presents the derived financial risk metrics.

Table 3: Financial Risk Metrics by Climate Scenario

Climate Scenario	Mean Farm Income (₦)	Value at Risk (95% CI)	Probability of Default (%)	Loan Recommendation
Favorable (80th percentile)	1,247,000 (SD = 234,000)	18.4%	12.3% (SD = 4.2)	Approve
Average (50th percentile)	892,000 (SD = 187,000)	31.2%	34.7% (SD = 6.8)	Approve with adjusted terms
Drought (10th percentile)	412,000 (SD = 98,000)	58.6%	67.2% (SD = 11.3)	Conditional approval only

Note. Probability of Default represents the proportion of simulated scenarios in which farm income falls below the loan repayment threshold (₦500,000). SD = Standard Deviation; CI = Confidence Interval.

The Probability of Default scores derived from Monte Carlo simulations ranged from 12.3% under favorable conditions to 67.2% under drought scenarios, with a substantial 55 percentage point variation between the most favorable and most adverse climate conditions. This wide range suggests that climate risk is not uniform and that lenders who fail to account for seasonal climate forecasts may be systematically mispricing risk. Under favorable conditions, the model estimates a default probability of approximately 12.3%, implying that roughly 88 out of 100 farmers could be expected to repay. However, under drought conditions, the estimated PD rises to 67.2%, suggesting that nearly 7 out of 10 farmers might default. These estimates should be interpreted with caution, as they are based on a Monte Carlo simulation calibrated on 54 observations (18 years × 3 LGAs). The precision of the PD estimates would be improved with a longer time series and validation against actual loan performance data.

The Value at Risk (VaR) at the 95% confidence level ranged from 18.4% under favorable conditions to 58.6% under drought scenarios. VaR represents the minimum income a farmer could expect in the worst 5% of climate years. For a farmer in a drought scenario, the VaR of 58.6% means that in the worst 5% of years, farm income would fall by more than 58.6% below expected levels. This level of income volatility has profound implications for loan repayment capacity and underscores the need for insurance products or flexible loan structures that can accommodate climate-induced income shocks. A farmer facing a 58.6% income reduction would be unable to meet fixed loan repayment obligations, suggesting that

loans should include income-contingent repayment features that adjust based on realized yields.
The observed relationship between climate conditions and default risk directly addresses the call by Adegbite and Machethe (2020) for locally calibrated risk models that incorporate climate variables. The authors found that less than 15% of Nigerian lenders use any form of quantitative risk modeling for agricultural portfolios. The framework developed in this study provides a practical, data-driven alternative to the collateral-based lending that currently dominates the Nigerian agricultural credit market. By providing lenders with quantitative default probability estimates, the framework enables more precise risk pricing and potentially expands credit access to smallholder farmers who lack traditional collateral but have viable farming operations. A farmer with a 12.3% PD could receive a loan at a lower interest rate than a farmer with a 67.2% PD, reflecting the substantial difference in risk.
The framework's ability to adjust PD scores based on climate projections represents a significant advancement over static credit scoring models. As Turvey et al. (2019) demonstrated in their work on weather index insurance, addressing fractional dimensionality in climate data is essential for accurate risk pricing. The present study demonstrates that machine learning-based yield predictions using secondary data can effectively address this dimensionality by capturing non-linear relationships between climate variables and yields. For financial institutions, this means that loan approval decisions and interest rates can be dynamically adjusted as seasonal climate forecasts become available, potentially reducing default rates and improving portfolio performance. A bank could, for example, conditionally approve loans based on the forecast for the upcoming season, with final approval contingent on the absence of drought warnings during critical flowering periods.

Practical Implications of the Findings

The findings of this study have several practical implications for agricultural lending in Nigeria. First, the framework demonstrates that AI-augmented risk assessment using secondary data can provide credible financial metrics for agricultural lending in data-sparse environments. This is particularly significant for Nigeria, where primary data collection can be expensive and logistically challenging. The use of publicly available or accessible secondary datasets—NiMet weather records, ERA5 reanalysis from ECMWF (Hersbach et al., 2020), and Ministry of Agriculture yield statistics—makes the framework replicable and scalable across other Nigerian states. A financial institution operating in Kano State, for example, could apply the same methodology using locally available climate and yield data without incurring the costs of primary data collection. This scalability is essential for expanding agricultural lending to underserved regions.
Second, the identification of specific phenological windows (flowering stage rainfall, vegetative stage growing degree days) as dominant predictors provides lenders with actionable early warning indicators. Rather than waiting until harvest to assess loan performance, lenders can monitor real-time weather conditions during these critical periods and adjust risk assessments dynamically. This capability could enable the development of adaptive loan products with interest rates that adjust based on observed weather conditions or repayment schedules that are flexible in response to climate shocks.
Third, the wide range of Probability of Default scores across climate scenarios (12% to 67%) suggests that climate-indexed lending products could be viable. As an illustration, a lender could consider offering a "climate-smart" loan with a base interest rate that adjusts based on observed rainfall during the flowering stage (e.g., a small discount for each 10 mm above normal), thereby providing an incentive for farmers to monitor weather conditions. Similarly, hypothetically, loans could be structured with graduated repayment schedules or, in the event of a drought, a one-year repayment holiday. The spatial variability in climate risk across LGAs also points to the possibility of risk-based lending zones, with differentiated interest rates or collateral requirements. These suggestions are intended to stimulate further research and pilot testing, rather than as empirically validated policy prescriptions.
Fourth, the spatial variability in climate risk across LGAs suggests that differential lending policies across regions are warranted. Lenders could establish risk-based lending zones, with lower interest rates and higher loan limits in low-risk areas (Odigbo) and higher interest rates or more stringent collateral requirements in high-risk areas (Akure South). This targeted approach would enable lenders to expand their agricultural portfolios while maintaining acceptable risk levels.

Limitations of the Study

Several limitations should be acknowledged. First, the study was conducted exclusively using secondary data from Ondo State, and the specific coefficients and feature importance rankings may not generalize directly to other agro-ecological zones in Nigeria. The model's performance in the derived savannah zone of northern Nigeria, for example, where rainfall patterns and crop varieties differ substantially, would require validation and potentially recalibration. Second, the secondary data available from the Ministry of Agriculture represent aggregated yields at the LGA level, which may mask farm-level variability. Within a single LGA, yields can vary substantially due to differences in soil quality, management practices, and microclimate conditions. The use of LGA-level averages may therefore underestimate the true variability in yields and default risk.

Third, the study duration of 18 years of historical data allowed for limited cross-validation across multiple climate cycles. While 18 years is a reasonable period for capturing interannual climate variability, it may not fully capture longer-term climate trends or decadal variability. The El Niño-Southern Oscillation (ENSO), for example, influences West African rainfall patterns on a 2–7 year cycle, and 18 years may include only 3–4 ENSO cycles. Longer time series, extending to 30–50 years, would provide more robust validation and better capture decadal climate variability.

Fourth, the Probability of Default scores were derived from simulated scenarios rather than validated against actual loan default data from financial institutions. The loan repayment threshold of ₦500,000 was derived from secondary data on typical agricultural loan products but may not reflect the actual loan terms offered by all institutions. Longitudinal validation against actual loan performance data would strengthen the credibility of the PD scores and enable calibration of the framework to reflect real-world default rates.

Fifth, the study did not account for non-climate factors that influence yield and default risk, such as pest outbreaks, input prices, and market volatility. Fall armyworm outbreaks, for example, have caused substantial maize losses in Nigeria in recent years, and their frequency may be influenced by climate conditions. Similarly, fertilizer and seed prices vary from year to year and can affect farmers' ability to invest in inputs. These factors could be incorporated into future iterations of the framework to improve predictive accuracy and practical relevance.

CONCLUSION

This study successfully developed and validated an AI-augmented financial de-risking framework for climate-resilient agriculture in Ondo State, Nigeria, using exclusively secondary data sources. The XGBoost yield prediction model achieved $R^2 = 0.86$ for oil palm and 0.84 for maize, with cumulative rainfall during flowering (34% importance) and growing degree days during vegetative growth (29% importance) identified as the dominant predictors. Monte Carlo simulations generated Probability of Default scores ranging from 12.3% under favourable conditions to 67.2% under drought scenarios, providing financial institutions with quantifiable, forward-looking risk metrics derived from accessible secondary data. The wide range of PD scores (55 percentage points) demonstrates that climate risk is not uniform and that lenders who fail to account for seasonal climate forecasts may be mispricing risk. The framework offers a replicable and scalable model for climate-resilient agricultural finance across Nigeria. Future work should validate the PD estimates against actual loan performance data, expand the framework to other states and crops, and integrate it with mobile money platforms to reach unbanked farmers.

Financial institutions operating in Nigeria should adopt AI-augmented risk assessment frameworks to improve agricultural lending decisions. The dashboard prototype developed in this study enables loan officers to input basic farm parameters and receive Probability of Default scores and climate risk visualizations. Institutions should integrate this tool into their loan origination processes, using PD scores as a quantitative complement to traditional collateral-based assessments. The 55 percentage point variation in PD across climate scenarios suggests that institutions that fail to incorporate climate information may be approving loans that are substantially riskier than they appear under static assessment methods. To ensure successful adoption, institutions should implement a phased approach, beginning with a pilot program in low-risk regions such as Odigbo LGA, where drought frequency is lowest, to build confidence and refine the tool before expanding to higher-risk regions such as Akure South, where climate variability poses greater challenges. This pilot phase should be accompanied by comprehensive training for loan officers on interpreting PD scores and climate risk visualizations, ensuring they understand the basis for risk assessments—particularly the role of cumulative rainfall during flowering and growing degree days during vegetative growth—so they can explain lending decisions to clients and build trust in the new system. Building on this foundational capacity, institutions should develop climate-indexed loan products with flexible repayment terms that align with expected income patterns under different climate scenarios, such as adjusting repayment schedules based on observed weather conditions during critical phenological windows, thereby reducing default rates and improving portfolio performance. To operationalize this approach, institutions should establish partnerships with the Nigerian Meteorological Agency to receive real-time weather alerts during critical phenological windows, enabling dynamic risk adjustment that allows lenders to proactively engage with farmers, restructure loans, or offer emergency support when adverse weather conditions are detected. Finally, institutions should integrate the framework with existing core banking systems to enable automated risk assessment and streamlined loan processing, allowing loan officers to generate PD scores automatically and reducing manual effort while ensuring consistency in risk assessment across the institution.

The Federal Ministry of Agriculture and the Central Bank of Nigeria should incorporate AI-augmented risk metrics into agricultural lending guidelines and guarantee schemes. Specifically, the Agricultural Credit Guarantee Scheme Fund (ACGSF) should accept Probability of Default scores from validated AI models as sufficient justification for loan guarantees, potentially reducing the collateral requirement for farmers with PD scores below 20%. This policy change would expand credit access to smallholder farmers who lack traditional collateral but have climate-resilient farming practices. Additionally, the National Agricultural Extension and Research Liaison Services should integrate the framework's early warning indicators into their extension advisory services, enabling farmers to make informed decisions about planting dates, crop selection, and input use based on climate projections.

Furthermore, the government should establish data sharing agreements between NiMet, the Ministry of Agriculture, and financial institutions to facilitate the use of secondary data for risk assessment. Currently, data access can be fragmented and time-consuming, limiting the scalability of AI-augmented risk assessment. A centralized data portal that provides standardized climate and agricultural data to financial institutions would reduce the cost and complexity of implementing such frameworks. The policy brief derived from this study should be disseminated to state and federal policymakers to facilitate evidence-based policy reform.

Future research should pursue four directions. First, the framework should be validated in other Nigerian states (Kano, Kaduna, and Enugu) and for other crops (cassava, cocoa, rice) to test its generalizability across Nigeria's diverse agro-ecological zones. Each region may have different dominant climate predictors, requiring region-specific model calibration. A comparative study across states would identify which components of the framework are universally applicable and which require localization.

Second, a longitudinal study should validate the Probability of Default scores against actual loan performance data from financial institutions that adopt the framework, tracking default rates among farmers approved using PD scores compared to a matched sample approved through traditional collateral-based methods. This validation would provide the empirical evidence needed to build confidence in the framework and support regulatory approval. Such a study would require collaboration with financial institutions and may span 3–5 years to capture sufficient loan performance data.

Third, future research should incorporate additional non-climate factors—such as input prices, market volatility, and pest outbreaks—into the framework to improve predictive accuracy. The fall armyworm outbreak, for example, has caused substantial maize losses in Nigeria and could be incorporated as a categorical variable or through remote sensing of crop health. Machine learning models are well-suited to incorporating diverse data types, and future iterations could integrate satellite-derived vegetation indices (e.g., NDVI) to provide real-time crop health monitoring.

Fourth, research should explore the integration of the framework with mobile money platforms (e.g., Paga, Opay) and agent banking networks to reach unbanked farmers who lack formal bank accounts. This integration could enable farmers to access credit directly through their mobile phones, using the PD scores to determine eligibility and loan terms. Research on the user experience and adoption barriers of such mobile-based tools would be essential for ensuring that the framework reaches the most vulnerable farming households.

REFERENCES

Adegbite, O. O., & Machethe, C. L. (2020). Bridging the financial inclusion gender gap in smallholder agriculture in Nigeria: An untapped potential for sustainable development. *World Development*, *127*, Article 104755. https://doi.org/10.1016/j.worlddev.2019.104755

Ayanlade, A., Radeny, M., & Morton, J. F. (2017). Comparing smallholder farmers' perception of climate change with meteorological data: A case study from southwestern Nigeria. *Weather and Climate Extremes*, *15*, 24-33. https://doi.org/10.1016/j.wace.2016.12.001

Central Bank of Nigeria. (2021). *Guidelines for the Agricultural Credit Guarantee Scheme Fund (ACGSF)*. Abuja: Central Bank of Nigeria.

Folberth, C., Baklanov, A., Balkovič, J., Skalský, R., Khabarov, N., & Obersteiner, M. (2019). Spatio-temporal downscaling of gridded crop model yield estimates based on machine learning. *Agricultural and Forest Meteorology*, *264*, 1-15. https://doi.org/10.1016/j.agrformet.2018.09.021

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Thépaut, J. N. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999-2049. https://doi.org/10.1002/qj.3803

Intergovernmental Panel on Climate Change. (2022). *Climate change 2022: Impacts, adaptation and vulnerability*. Cambridge University Press.

International Fund for Agricultural Development. (2023). *rural development report 2023: Transforming food systems for rural prosperity*. Rome: IFAD.

Lobell, D. B., & Gourdji, S. M. (2012). The influence of climate change on global crop productivity. *Plant Physiology*, *160*(4), 1686-1697. https://doi.org/10.1104/pp.112.208298

Okonkwo, C., & Ezike, R. (2022). A review of the applications of artificial intelligence in agriculture: Prospects and challenges in Nigeria. *Journal of Agricultural Engineering and Technology (JAET)*, *27*(2).

Shahhosseini, M., Hu, G., Khaki, S., & Archontoulis, S. V. (2021). Corn yield prediction with ensemble CNN-DNN. *Frontiers in Plant Science*, *12*, Article 709008. https://doi.org/10.3389/fpls.2021.709008

Turvey, C. G., Shee, A., & Marr, A. (2019). Addressing fractional dimensionality in weather index insurance and climate risk financing in agricultural development: A dynamic triggering approach. *Weather, Climate and Society*, *11*(4), 901-915. https://doi.org/10.1175/WCAS-D-19-0017.1

Van Klompenburg, T., Kassahun, A., & Catal, C. (2020). Crop yield prediction using machine learning: A systematic literature review. *Computers and Electronics in Agriculture*, *177*, 105709. https://doi.org/10.1016/j.compag.2020.105709

World Bank. (2023). *Nigeria development update (June 2023): Seizing the opportunity*. Washington, DC: World Bank Group.