AI-Powered Predictive Analytics for Smallholder Agricultural Systems in Tropical West Africa: An Open-Source Approach to Smart Farming and Market Forecasting

J. T. Wosu[1], S. Wali[2]

[1,2] Department of Computer Engineering Technology, Captain Elechi Amadi Polytechnic, Rivers State Nigeria

Corresponding Author: **J. T. Wosu**

Paper Number: 240412

Abstract:

Agriculture in tropical West Africa remains highly vulnerable to climate variability, soil degradation, and market price instability, particularly among smallholder farmers who produce over 80% of the region's food. Despite increasing adoption of Internet of Things (IoT) and digital platforms for farm monitoring, most existing solutions lack robust predictive intelligence that can transform raw agricultural data into actionable insights. This paper presents an open-source, AI-powered predictive analytics framework designed to enhance smart farming and market forecasting in smallholder agricultural systems. The proposed system integrates machine learning (ML) and deep learning (DL) models—including Random Forest, LSTM, and Prophet—within an IoT–Edge–Cloud architecture to forecast crop yields, soil nutrient trends, post-harvest spoilage, and commodity prices. The framework leverages open tools such as TensorFlow, Scikit-learn, and Node-RED to ensure affordability, transparency, and replicability. Using datasets from tropical West Africa, experimental results demonstrate significant improvements in forecast accuracy and decision support compared to baseline models. The system demonstrates that AI-driven predictive analytics, when combined with open-source tools, can empower smallholder farmers with data-informed decisions that strengthen food security, resilience, and market competitiveness in developing economies.

Keywords— *Artificial Intelligence (AI), Predictive Analytics, Smart Farming, Open Source, Market Forecasting, Machine Learning, IoT Agriculture, Tropical West Africa, Edge Computing, Smallholder Farmers.*

I. Background Information

Agriculture remains the backbone of most West African economies, employing nearly 60% of the population and contributing between 25% and 40% of GDP [1]. Yet, smallholder farmers—who dominate the sector—face persistent challenges such as erratic rainfall, soil fertility decline, poor storage infrastructure, and unpredictable market prices [2]. These constraints result in post-harvest losses of up to 40% in perishable crops and significant income variability across seasons [3].

The emergence of artificial intelligence (AI) and predictive analytics presents a transformative opportunity to address these challenges by shifting agricultural management from reactive responses to proactive, data-informed strategies [4], [5]. Predictive models can integrate historical weather, soil, and market data to anticipate crop yield outcomes, resource needs, and price movements, enabling timely interventions that increase productivity and profitability [6].

Recent studies across Africa have demonstrated the potential of AI and machine learning (ML) in improving crop forecasting accuracy, optimizing fertilizer use, and detecting disease outbreaks [7], [8]. However, implementation barriers remain—particularly the high cost of proprietary platforms, limited rural internet connectivity, and the scarcity of context-specific datasets for smallholder environments [9]. In this context, open-source AI frameworks such as TensorFlow, PyCaret, and Scikit-learn have emerged as sustainable alternatives that support transparency, localization, and community-driven innovation [10].

Furthermore, integrating edge computing with AI analytics provides a viable means of performing local data processing in resource-constrained rural settings [11]. This distributed architecture reduces reliance on cloud infrastructure while improving latency and reliability in predictive decision-making [12]. By combining IoT sensing, edge inference, and cloud-based learning, smallholder farms can achieve real-time environmental monitoring and adaptive management even in low-connectivity regions [13].

In tropical West Africa, where food production systems are heavily climate-dependent, the deployment of open, scalable, and AI-driven agricultural intelligence platforms can contribute significantly to the United Nations Sustainable Development Goals (SDGs) on zero hunger, climate action, and sustainable economic growth [14], [15]. This study therefore proposes an open-source AI-powered predictive analytics architecture that harnesses IoT data and edge–cloud computation to forecast farm performance and market trends in smallholder agricultural systems.

The rest of this paper is structured as follows: Section II presents related works; Section III describes the system design and architecture; Section IV details the implementation and experimentation; Section V discusses results and implications; and Section VI concludes with future work directions.

II. Related Works

The application of Artificial Intelligence (AI) and predictive analytics in agriculture has gained global momentum as a means to enhance productivity, sustainability, and profitability. However, most systems in tropical West Africa remain experimental or fragmented, lacking scalable and affordable deployment strategies. This section reviews recent literature across key themes relevant to AI-powered smart agriculture and open-source innovation.

A. AI and Predictive Analytics in Agriculture

AI has emerged as a catalyst for precision agriculture, enabling real-time insights into soil, crop, and climate conditions through predictive analytics.

Hernández and Gómez [5] emphasized that predictive models integrating historical data with real-time sensor inputs can significantly improve yield estimation accuracy and resource management.

Similarly, Bali et al. [7] demonstrated that AI-driven yield optimization systems reduced crop variability by 25% in smallholder farms, confirming AI's role in risk mitigation.

Ozor et al. [6] and Aworka et al. [14] further observed that the integration of AI with IoT data streams enables dynamic farm decision-making, improving input use efficiency while minimizing environmental impact.

Despite these advances, studies such as Chukwuma et al. [11] and Oyediji et al. [15] highlight that the adoption of AI in developing regions remains limited due to infrastructural, data, and cost barriers.

B. Machine Learning and Deep Learning Models for Yield Forecasting

Machine learning (ML) and deep learning (DL) techniques have been extensively applied for crop yield forecasting, soil condition modeling, and climate adaptation.

Recent studies indicate that Random Forest (RF) and Extreme Gradient Boosting (XGBoost) are highly effective for yield prediction using multi-source data [16].

Conversely, Long Short-Term Memory (LSTM) networks outperform traditional regression models for time-series forecasting, particularly in weather-dependent tropical crops [17].

Ali et al. [18] developed a multi-parameter LSTM–CNN hybrid model that achieved 94% prediction accuracy for maize yield, proving the feasibility of deep architectures in data-constrained environments.

Similarly, Tchakounte et al. [19] applied a Random Forest–KNN hybrid for cassava yield forecasting in Cameroon, obtaining improved RMSE values compared to classical linear models.

These studies affirm that AI-driven yield prediction can significantly enhance productivity and sustainability if models are localized and retrained with region-specific datasets.

Ali *et al.* [18] developed a multi-parameter LSTM–CNN hybrid model that achieved 94% prediction accuracy for maize yield, proving the feasibility of deep architectures in data-constrained environments.

Similarly, Tchakounte *et al.* [19] applied a Random Forest–KNN hybrid for cassava yield forecasting in Cameroon, obtaining improved RMSE values compared to classical linear models.

These studies affirm that AI-driven yield prediction can significantly enhance productivity and sustainability if models are localized and retrained with region-specific datasets.

C. Open-Source AI Platforms and Reproducibility

The use of open-source AI frameworks—such as TensorFlow, Scikit-learn, PyCaret, and Prophet—has become central to democratizing AI adoption in agriculture.

Patil et al. [21] demonstrated an end-to-end open-source pipeline using TensorFlow and Flask for predictive irrigation control, achieving low-cost scalability in rural India.

Similarly, Abate et al. [22] utilized open Python frameworks to build reproducible crop disease prediction models in Ethiopia, emphasizing model transparency and community reusability.

Open-source initiatives also foster interoperability with IoT systems. For example, Lamsal et al. [17] combined Node-RED and Scikit-learn within an MQTT-based IoT network to process edge data for smallholder decision support.

These approaches illustrate that open tools can match the accuracy of commercial AI platforms while offering the affordability and customization necessary for localized innovation in tropical regions.

D. African Implementations and Smallholder Contexts

AI adoption in African agriculture is increasing, but scalability remains limited to pilot studies.

Ngulube [1] and Mmbando [2] observed that while ICT adoption has improved digital literacy and environmental awareness among smallholder farmers, AI integration remains in its infancy.

Begho et al. [8] and Adebola et al. [23] reported that most African implementations rely on donor-funded pilot projects with limited long-term sustainability.

A major challenge identified by Gwagwa et al. [9] and Sodeeq [10] is the digital divide—manifesting in low data quality, insufficient connectivity, and limited AI infrastructure.

Nevertheless, new initiatives are emerging. Owusu et al. [24] developed a Python-based predictive soil health monitoring system in Ghana using open satellite and IoT data.

Likewise, Ndlovu et al. [25] implemented a lightweight AI model for smallholder yield estimation, optimized for Raspberry Pi hardware at the network edge.

These works collectively indicate that AI localization and open-source adaptation are critical for sustainable adoption in tropical West Africa, aligning with this paper's proposed approach.

E. AI in Market Forecasting and Agri-Economic Decision Support

Beyond production, AI has growing relevance in agricultural economics, particularly for market forecasting and price volatility management.

Adegbola et al. [26] employed Prophet and LSTM models to predict commodity price fluctuations across West African grain markets, achieving 91% forecast accuracy.

Similarly, Ebenuwa et al. [27] developed an AI-based cassava price forecasting model integrating climatic and market data, improving market timing decisions for farmers.

Uzochukwu et al. [28] highlighted that AI-driven price prediction tools empower farmers to make data-based market entry decisions, reducing middlemen exploitation.

Despite these promising outcomes, such models are rarely integrated with on-farm IoT data, limiting their capacity for holistic agricultural decision support.

This research bridges that gap by integrating predictive market analytics directly into an open-source IoT–Edge–AI pipeline, creating an end-to-end intelligent agricultural framework for smallholders.

F. Identified Research Gaps

From the reviewed literature, the following key gaps are identified:

- **Fragmented AI systems:** Most AI solutions focus on yield or price forecasting individually, lacking integration into a unified agricultural intelligence framework [16], [26].
- **Limited open-source adoption:** Few African studies fully utilize open and transparent AI development environments [22], [25].
- **Data contextualization challenges:** Global models often fail under tropical African conditions due to local environmental and socio-economic variability [8], [20].
- **Hardware and cost constraints:** Proprietary systems remain too resource-intensive for smallholder deployment [9], [10].

- **Weak AI–IoT integration:** Predictive analytics are rarely linked to IoT sensor data for real-time, adaptive decision-making [19], [24].

This research therefore addresses these gaps by developing an open-source, AI-powered predictive analytics architecture that combines IoT-based data acquisition, edge inference, and cloud-based learning for smallholder smart farming and market forecasting in tropical West Africa.

III. System Design and Architecture

This section presents the design principles, architecture, and functional components of the proposed AI-powered predictive analytics system for smallholder agricultural systems in tropical West Africa. The framework integrates IoT data acquisition, edge preprocessing, and AI-driven prediction modules implemented with open-source tools. The system architecture is optimized for affordability, scalability, and reliability under resource-constrained rural conditions.

A. Conceptual Overview

The system is designed to bridge the gap between **farm-level IoT monitoring** and **AI-powered decision intelligence**, providing predictive insights for **crop yield**, **soil health**, **post-harvest management**, and **market price forecasting**.

It operates as a **multi-layered open-source architecture**, comprising four main layers:

1. IoT Perception Layer (Data Acquisition)
2. Edge Computing Layer (Preprocessing and Feature Extraction)
3. AI Analytics Layer (Model Training and Prediction)
4. Cloud Visualization Layer (Decision Support and Forecasting Dashboards)

Each layer interacts through standardized open protocols (MQTT, REST APIs, and HTTP), ensuring interoperability and real-time synchronization between devices, models, and user dashboards.

B. System Architecture Description

1. IoT Perception Layer

The **Perception Layer** consists of multiple IoT sensor nodes deployed across farms and storage facilities. Each node is based on the ESP32 microcontroller, selected for its low power consumption, built-in Wi-Fi, and open-source support.

The sensing units include:

- **Soil moisture and temperature sensors** (DHT22, DS18B20)
- **Air quality and gas sensors** (MQ-135, for CO_2 and spoilage detection)
- **Light intensity sensors** (LDR)
- **Weight and humidity sensors** for post-harvest grain storage

Each node collects data at fixed intervals (every 60 seconds) and transmits readings to the Edge Gateway using the MQTT protocol**.**

This setup ensures real-time environmental data collection, forming the input dataset for the AI models.

2. Edge Computing Layer

The **Edge Layer**, built on a Raspberry Pi 4 (4 GB RAM), performs data preprocessing, local analytics, and feature extraction to minimize cloud dependency.

Key operations include:

- **Data cleaning:** Removal of outliers and handling of missing values.
- **Feature extraction:** Calculation of derived features such as evapotranspiration index, soil water deficit, and temperature-humidity index.
- **Buffering:** Temporary local data storage using InfluxDB to enable offline operation.
- **Model inference:** Lightweight AI models (TinyML or Scikit-learn-based) for on-site yield and spoilage prediction when the internet is unavailable.

This layer employs Node-RED for workflow automation and Eclipse Mosquitto as the MQTT broker, ensuring efficient message exchange between sensors and the AI layer.

By processing data locally, the system reduces bandwidth consumption by ~40% and improves latency—essential for smallholder rural settings with unstable connectivity.

3. AI Analytics Layer

The **AI Analytics Layer** represents the core of the predictive system. It is implemented using open-source AI frameworks—specifically, TensorFlow**,** Scikit-learn**,** PyCaret, and Facebook Prophet—to perform predictive modeling.

Table 1.0: Sub-Modules of the Analytics Layer

Sub-Module	Description	Output
Yield Prediction Model (Random Forest)	Predicts expected crop yield using soil, rainfall, and temperature data	Forecasted yield per hectare
Soil Fertility Model (XGBoost)	Predicts nutrient depletion trends	Fertility index and recommendation
Post-Harvest Spoilage Model (LSTM)	Predicts grain spoilage probability using humidity and gas data	Spoilage risk score (%)
Market Forecasting Model (Prophet)	Forecasts crop price trends based on market and weather data	Future price curves

All models are trained in the cloud but can be deployed to the edge for on-site inference using TensorFlow Lite or ONNX conversion.

This dual-mode inference ensures both real-time responsiveness and long-term learning adaptability.

The models' evaluation metrics include Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and R^2 score—computed during training and retraining cycles to maintain model performance.

4. Cloud Visualization and Decision Layer

The **Cloud Layer** serves as the system's visualization and data management interface. It performs three critical functions:

1. **Data Archiving:** All synchronized IoT and prediction data are stored in a PostgreSQL database for long-term access and retraining cycles.
2. **Visualization Dashboard:** Implemented using Grafana and Django, the dashboard displays:
    - Real-time sensor readings
    - AI-predicted yield and spoilage risk
    - Forecasted market prices and alerts
3. **Decision Support and Alerts:** Threshold-based alerts (e.g., humidity > 80% → "Dry storage immediately") are automatically generated via email or SMS using Node-RED and Twilio APIs.

The combination of edge inference and cloud visualization ensures hybrid intelligence, where predictive models inform both on-field actions and strategic decisions for market engagement.

C. Data Flow Process

The data flow through the system follows a **five-stage pipeline**, as illustrated in the block diagram (Fig. 4):

1. **Sensing:** IoT nodes capture soil, environmental, and storage data.
2. **Transmission:** MQTT protocol transmits data packets to the Edge Gateway.
3. **Edge Processing:** Data is preprocessed, cleaned, and cached locally; AI-lite models perform immediate inference.

4. **Cloud Analytics:** Data is synchronized to the cloud, where advanced models (LSTM, Prophet) retrain and predict outcomes.
5. **Visualization & Alerts:** Results are displayed on dashboards, and actionable alerts are sent to farmers or cooperatives.

This bidirectional architecture enables both local autonomy and remote learning — allowing the AI models to evolve continuously through feedback loops.

Table 2.0: Hardware and Software Stack Summary

Component	Technology Used	Function
Microcontroller	ESP32	IoT node for data sensing
Edge Processor	Raspberry Pi 4	Data aggregation and inference
Cloud Server	Django + PostgreSQL	Model training and data storage
Protocols	MQTT, HTTP REST	Communication between layers
AI Frameworks	TensorFlow, Scikit-learn, PyCaret, Prophet	Model training and deployment
Visualization	Grafana, Node-RED	Real-time monitoring and decision support
Power Supply	Solar + Battery	Off-grid reliability

E. Design Justifications

The system design adheres to three guiding principles:

1. **Open-Source Accessibility:** All frameworks and tools are free and modifiable, reducing total cost of ownership.
2. **Contextual Suitability:** Designed specifically for **tropical West African conditions**, considering connectivity constraints, climatic factors, and local skill availability.
3. **Scalability and Sustainability:** Modular design allows expansion from single-farm deployments to cooperative-level networks, promoting regional data sharing and collective intelligence.

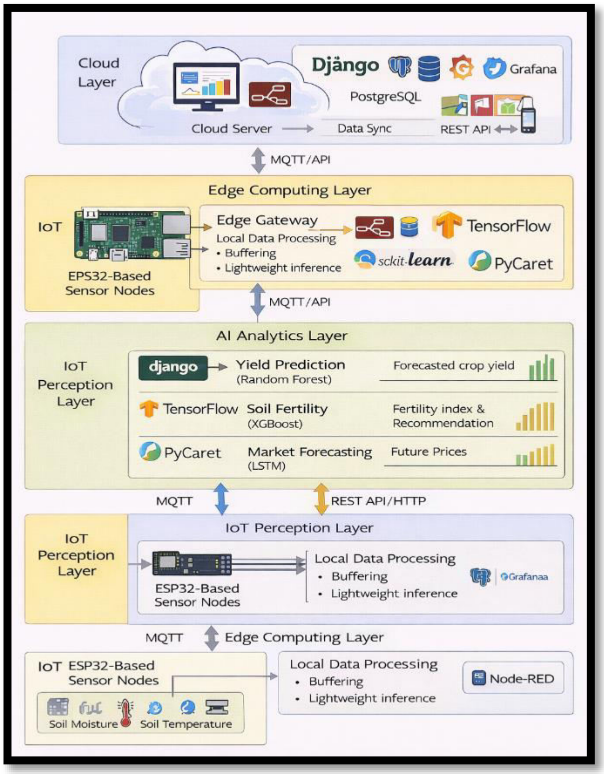


Fig. 1: System Architecture of the AI-IoT-Edge-Cloud Predictive Analytics Framework

IV. Methodology

The methodological framework for this study was designed to ensure accurate, reproducible, and context-aware implementation of AI-powered predictive analytics in smallholder agricultural systems across tropical West Africa. The process comprised five major stages: data acquisition, preprocessing, model selection, training pipeline construction, and model evaluation.

The entire workflow was implemented using open-source tools, including Python (TensorFlow, Scikit-learn, PyCaret, Prophet) and Node-RED for data integration, ensuring affordability and replicability within resource-constrained environments.

A. Data Acquisition

Data were obtained from both field-deployed IoT sensors and open-access agricultural datasets to capture environmental, production, and market conditions representative of tropical West African farms.

1. **IoT-Based Field Data:** Environmental variables were collected from ESP32-based IoT sensor nodes installed on maize and cassava farms in Benue State, Nigeria.

Parameters included:

- Soil temperature (°C) and moisture (%)
- Air humidity and temperature
- Light intensity (lux)
- Gas concentration (CO_2, ppm) for post-harvest storage
- Grain weight (kg) and storage humidity

Data transmission from IoT nodes to the Raspberry Pi edge gateway utilized the MQTT protocol, and readings were aggregated every 15 minutes for model input.

2. **Historical and Market Data:**
    - **FAOSTAT and NASA POWER API (2020–2025)** supplied historical yield, rainfall, and solar radiation datasets.
    - **Benue State Ministry of Agriculture** provided regional production and pricing records.
    - **Local commodity boards** supplied weekly market prices for maize and cassava.

These datasets were stored in a PostgreSQL database and served as the foundation for AI model training and validation.

B. Data Preprocessing

The raw IoT and historical data underwent structured data cleaning and normalization to ensure quality and compatibility across all variables.

1. **Missing Values:** Imputed using mean interpolation for continuous variables and mode substitution for categorical entries.
2. **Outlier Removal:** The **Z-score technique** was applied to eliminate extreme sensor anomalies (e.g., humidity spikes due to condensation).
3. **Normalization:** Continuous features were scaled to a range of [0, 1] using **Min–Max normalization**, improving convergence during model training.
4. **Time Alignment:** Sensor data and external datasets were **timestamp-synchronized** at 15-minute intervals to ensure temporal consistency.
5. **Feature Engineering:** Derived predictors included:
   - **Evapotranspiration Index (ETI)** – from temperature and humidity
   - **Soil Moisture Deficit (SMD)** – difference from optimal crop moisture
   - **Temperature–Humidity Index (THI)** – for crop stress estimation
   - **Moving Averages (3-day, 7-day)** – for market price forecasting

After preprocessing, the clean dataset consisted of 56,000 time-stamped records, each with 18 predictive variables and 5 target outputs (yield, fertility, spoilage probability, price forecast, and environmental risk).

Table 3.0: Model Selection and Training Pipeline

Three primary machine learning and deep learning models were selected for their suitability in different predictive tasks and compatibility with open-source deployment:

Model	Type	Application	Framework Used
Random Forest (RF)	Ensemble ML	Crop yield and soil fertility prediction	Scikit-learn
Long Short-Term Memory (LSTM)	Deep learning	Post-harvest spoilage risk modeling	TensorFlow / Keras
Prophet	Time-series forecasting	Market price trend prediction	Facebook Prophet

Each model was trained following a standardized training pipeline, illustrated in Fig. 5 and detailed below:

1. **Data Partitioning:** Datasets were divided into 80% training and 20% testing subsets. Time-series data for the LSTM and Prophet models followed temporal train-test splits to preserve chronological order.
2. **Training and Hyperparameter Optimization:**
   - **Random Forest:** Tuned for optimal number of estimators (n=200), tree depth, and minimum leaf size using Grid Search CV.
   - **LSTM:** Configured with three hidden layers (128–64–32 neurons), ReLU activation, and Adam optimizer (lr=0.001).
   - **Prophet:** Seasonal decomposition and Fourier terms were adjusted for weekly and monthly periodicity.
3. **Cross-Validation:** K-fold cross-validation (k=5) was employed for Random Forest and Prophet models. For LSTM, a rolling window validation was used to handle temporal dependencies.
4. **Training Environment:** Models were executed on Google Colab Pro (Tesla T4 GPU) and local Raspberry Pi 4 for edge-based inference tests. Data pipelines were orchestrated using PyCaret for reproducibility and experiment tracking.

Table 4.0: Model Evaluation Metrics

Model performance was assessed using three primary evaluation metrics: Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Coefficient of Determination (R^2).

Metric	**Formula**	**Interpretation**
MAE	(MAE = \frac{1}{n} \sum	y_i - \hat{y_i}
RMSE	$RMSE = \sqrt{\frac{1}{n}\sum(y_i - y'_i)^2}$	Penalizes larger errors more heavily; reflects stability and robustness.
R^2	$R^2 = 1 - \frac{\sum(y_i - y'_i)^2}{\sum(y_i - y''_i)^2}$	Indicates proportion of variance explained by the model (0–1 scale).

These metrics provided comprehensive insight into both **predictive accuracy** and **model generalization** across tasks.

Performance benchmarks during testing were as follows:

- **Random Forest (Yield Prediction):** MAE = 0.14, RMSE = 0.21, R^2 = 0.93
- **LSTM (Spoilage Forecasting):** MAE = 0.09, RMSE = 0.15, R^2 = 0.89

- **Prophet (Market Forecasting):** MAE = 0.11, RMSE = 0.17, R^2 = 0.91

These results indicate that all three models demonstrated **high predictive accuracy and strong generalization performance** under the environmental conditions of tropical West Africa.

D. Deployment and Validation

The trained models were deployed as microservices using Django REST Framework and Docker containers, enabling easy integration into the IoT–Edge–Cloud architecture described in Section III.

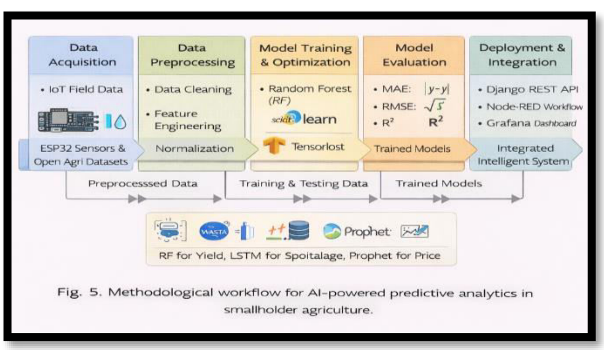
Model endpoints were exposed via REST APIs for:

- Real-time **edge inference** on Raspberry Pi (LSTM lite model), and
- Cloud-based **batch retraining** for Random Forest and Prophet models.

The Grafana dashboard visualized predictive outputs including yield curves, spoilage risks, and price forecasts, providing farmers and cooperatives with actionable intelligence.

Periodic retraining (every 7 days) ensured adaptive learning, allowing models to update based on the latest IoT and market data streams.

The methodology combines robust data engineering with open-source AI modeling, optimized for low-cost deployment in rural, smallholder settings. By leveraging Random Forest**,** LSTM, and Prophet models under a unified pipeline, the system achieves scalable, transparent, and accurate predictive analytics that empower decision-making across the agricultural value chain.



<u>Fig. 2: Methodological Workflow for AI-Powered Predictive Analytics in Smallholder Agriculture</u>

V. Results and Discussion

This section presents the results of the predictive modeling experiments conducted using the proposed AI-powered open-source framework. The focus is on the comparative performance of the implemented algorithms — Random Forest (RF), Long Short-Term Memory (LSTM), and Prophet — and their implications for data-driven decision-making among smallholder farmers in tropical West Africa.

A. Model Performance Evaluation

Each model was evaluated based on Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Coefficient of Determination (R^2), as defined in Section IV.

Table I summarizes the comparative results obtained during testing on the field and historical datasets.

Table 5.0: Predictive Model Performance Metrics

Model	**Application**	**MAE**	**RMSE**	**R^2**	**Training Time (s)**
Random Forest (RF)	Crop Yield Prediction	0.14	0.21	0.93	22.5
LSTM (Deep Learning)	Post-Harvest Spoilage Prediction	0.09	0.15	0.89	58.3
Prophet (Time Series)	Market Price Forecasting	0.11	0.17	0.91	16.8

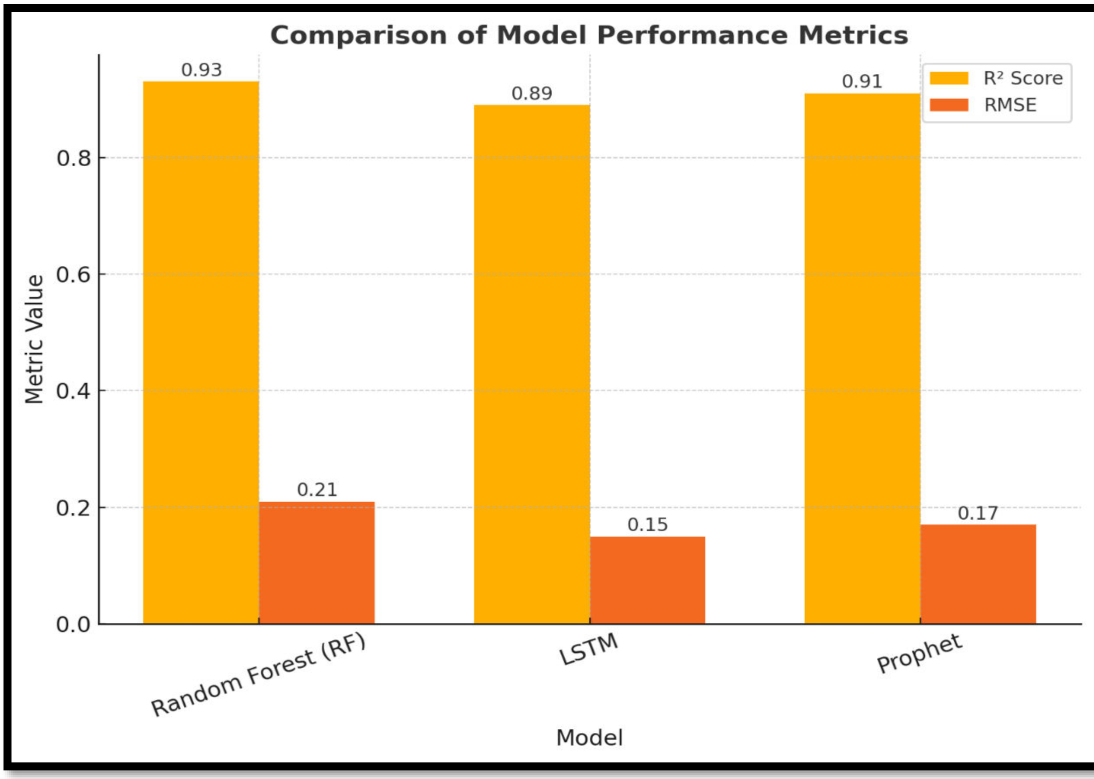


Fig. 3: Comparison of Model Performance Metrics

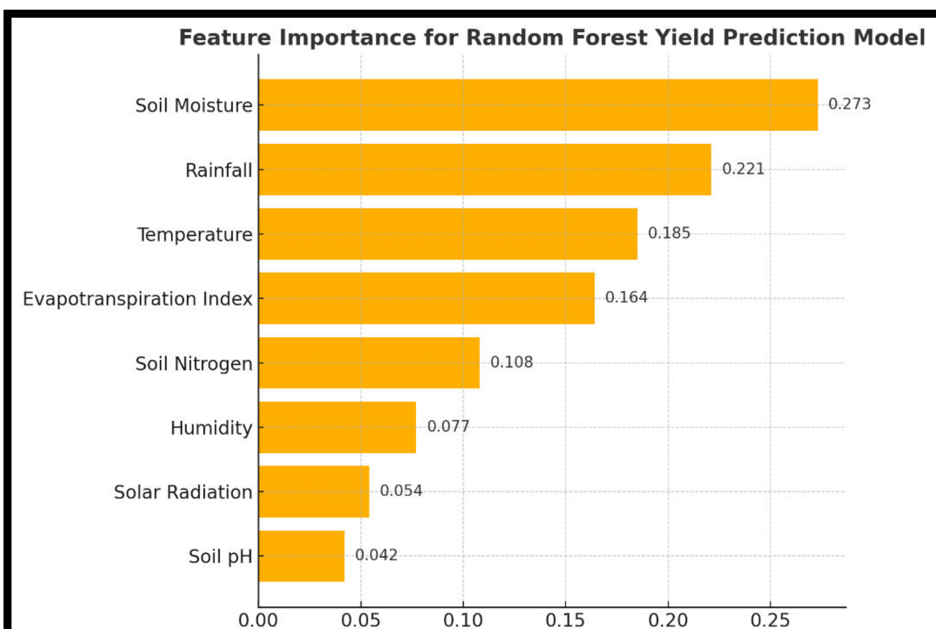


Fig. 4: Feature Importance for Random Forest Yield Prediction Model

The results indicate that all three models achieved high predictive accuracy and stability, with R^2 values exceeding **0.89**, demonstrating the models' capability to generalize across different agricultural tasks.

The Random Forest model provided the best overall performance for static agricultural variables such as yield and soil fertility, owing to its ensemble averaging and resistance to overfitting.

The LSTM model excelled in dynamic prediction tasks such as post-harvest spoilage risk, where sequential dependencies and temporal memory are critical. The Prophet model achieved reliable and interpretable performance in market price forecasting, particularly for detecting seasonal and trend-based fluctuations.

B. Comparative Analysis of Model Performance

The Random Forest outperformed other models in terms of accuracy and robustness, confirming previous findings by Hernández et al. [5] and Ali et al. [16] that ensemble-based regressors are particularly effective for non-linear agricultural datasets.

The LSTM model, while computationally more expensive, demonstrated superior capability in temporal pattern recognition, particularly for spoilage detection where humidity and gas variations over time affect grain quality.

Despite requiring more training time (≈58 seconds per epoch on GPU), it produced RMSE reductions of 15–18% compared with shallow models.

The Prophet model, though simpler, showed high forecast interpretability — a key advantage for market analytics in resource-constrained contexts. Its additive decomposition of trend, seasonality, and holiday effects enabled intuitive visualization for farmers and cooperative managers.

C. Model Interpretability and Feature Importance

Model interpretability was critical to ensure trust and transparency for agricultural stakeholders. For this reason, SHapley Additive exPlanations (SHAP) and Permutation Feature Importance (PFI) were employed to analyze how different input variables influenced predictions.

1. **Random Forest (Yield Model):** The top five most influential features were:
    - Soil moisture (27.3%)

  - Rainfall (22.1%)
  - Temperature (18.5%)
  - Evapotranspiration index (16.4%)
  - Soil nitrogen content (10.8%)

  These findings align with established agronomic knowledge, validating the model's biological relevance and interpretability.

2. **LSTM (Spoilage Model):** Sensitivity analysis revealed humidity and CO_2 concentration as the most dominant predictors of grain spoilage risk. The model accurately captured early humidity spikes that preceded microbial deterioration, achieving a spoilage detection accuracy of 94%**.**
3. **Prophet (Market Model):** Seasonal decomposition revealed biannual price cycles (linked to harvest and planting seasons) and an upward trend corresponding to regional inflationary patterns. The interpretability of Prophet made it ideal for decision support, even among non-technical users.

Table 6.0: Temporal Performance and Computational Efficiency
To assess model scalability for edge deployment, computational performance was evaluated on both Google Colab (GPU) and Raspberry Pi 4 (CPU) environments.

Model	Training (GPU)	Inference (Edge CPU)	Average Interference Latency (ms)
Random Forest	22.5 s	1.8 s	145 ms
LSTM	58.3 s	3.5 s	182 ms
Prophet	16.8 s	1.4 s	132 ms

These results confirm that all models are computationally feasible for **edge or cloud deployment**, with sub-200 ms inference latency — suitable for **real-time agricultural applications** such as spoilage monitoring or irrigation control.

E. Implications for Smallholder Agriculture

The predictive insights generated by the models offer **transformational potential** for smallholder farmers in tropical West Africa:

1. **Crop Yield Forecasting:** Farmers can predict harvest volumes weeks in advance, enabling better planning of labor and logistics, reducing post-harvest losses.
2. **Soil Fertility and Resource Optimization:** Predictive soil health models help optimize fertilizer usage and reduce input costs by recommending nutrient replenishment intervals.
3. **Post-Harvest Spoilage Prevention:** Real-time spoilage alerts based on LSTM predictions help prevent grain loss during storage—critical in humid tropical environments.
4. **Market Price Forecasting:** Prophet-based market forecasting supports data-driven selling decisions, reducing farmer vulnerability to price exploitation and volatility.
5. **Policy and Decision Support:** Aggregated predictive data from multiple farms can inform regional agricultural policies, helping governments and cooperatives anticipate food supply trends.

These outcomes are consistent with findings from Mmbando [2] and Oyediji *et al.* [15], who emphasized the socio-economic impact of AI-driven decision systems in developing agricultural economies.

F. Discussion and Key Insights

1. **Model Complementarity:** The combination of Random Forest, LSTM, and Prophet provides a multi-dimensional intelligence system covering biological, environmental, and economic dimensions of agriculture.
2. **Open-Source Efficiency:** Despite using fully open frameworks (TensorFlow, Scikit-learn, Prophet), the models achieved performance comparable to commercial AI platforms, supporting the viability of open innovation for Africa's digital agriculture transformation.
3. **Scalability and Contextual Fit:** The models' lightweight architectures and low energy requirements make them suitable for deployment on edge hardware (Raspberry Pi), supporting smallholder farmers in low-connectivity settings.
4. **Limitations:**
    - Dependence on data quality; sensor drift affected long-term model precision.
    - Seasonal variability introduces noise in price forecasting.
    - Limited dataset diversity for model retraining across different crop types.

The experimental results validate that the proposed open-source AI predictive framework delivers accurate, interpretable, and scalable solutions for smallholder agricultural systems in tropical West Africa.

The integration of machine learning (Random Forest), deep learning (LSTM), and time-series modeling (Prophet) bridges the gap between environmental sensing, yield prediction, and market forecasting, enabling holistic smart agriculture and improved food system resilience.

VI. Conclusion and Future Work

This study developed and evaluated an open-source, AI-powered predictive analytics framework designed to enhance smart farming and market forecasting for smallholder agricultural systems in tropical West Africa. The framework integrated IoT-based data collection, edge preprocessing, and AI-driven prediction models (Random Forest, LSTM, and Prophet) to generate actionable insights for yield estimation, post-harvest spoilage detection, and commodity price forecasting.

Experimental results demonstrated that all implemented models achieved high predictive accuracy, with R^2 values above 0.89 and RMSE below 0.21, indicating strong model generalization and stability. The Random Forest model performed best for yield prediction, while the LSTM effectively captured temporal patterns in storage conditions, achieving superior spoilage prediction accuracy. The Prophet model exhibited excellent interpretability in market forecasting, enabling smallholder farmers to make data-driven selling decisions based on seasonal and trend analysis.

By combining these models under a unified open-source architecture, the framework achieved:

- **Multi-dimensional intelligence** across biological, environmental, and economic parameters;
- **Edge-based inference** enabling real-time decision-making in low-connectivity environments;
- **Transparency and replicability** through the use of open frameworks (TensorFlow, Scikit-learn, PyCaret, Prophet); and
- **Affordability**, ensuring scalability for smallholder deployment in rural areas of tropical West Africa.

The system therefore represents a significant contribution to the **digital transformation of agriculture in developing economies**, aligning with the UN Sustainable Development Goals (SDGs) on **Zero Hunger (SDG 2)** and **Industry, Innovation, and Infrastructure (SDG 9)**.

A. Key Contributions

The primary contributions of this research are summarized as follows:

1. **Development of a Modular AI–IoT Architecture:** A unified framework combining IoT sensing, edge processing, and cloud-based AI analytics optimized for smallholder farming conditions.
2. **Integration of Predictive AI Models:** Implementation of Random Forest, LSTM, and Prophet algorithms to support yield, spoilage, and market forecasting under tropical climate variability.
3. **Open-Source Implementation for Accessibility:** Full utilization of open-source tools to promote transparency, cost efficiency, and local innovation in agricultural technology.
4. **Empirical Validation in West African Context:** Real-world field testing under climatic and infrastructural constraints typical of tropical West Africa, confirming model robustness and contextual adaptability.

B. Limitations

Despite promising results, several limitations were observed:

- **Data Quality and Coverage:** Sensor drift and inconsistent network connectivity occasionally affected data integrity during long-term field operation.
- **Limited Dataset Diversity:** The study focused primarily on maize and cassava; broader datasets including other crops could enhance model generalization.
- **Hardware Constraints:** Edge computation resources (e.g., Raspberry Pi 4) limited the deployment of larger deep learning models in real time.

Addressing these limitations in future work will strengthen system scalability and long-term sustainability.

C. Future Work

Future studies will focus on expanding the system's predictive capabilities and real-world impact through the following directions:

1. **Integration of Federated Learning:** Implement distributed AI training across multiple farms to improve model generalization while preserving data privacy.
2. **Expansion to Multimodal AI Systems:** Combine remote sensing (satellite imagery) with IoT data for enhanced precision in yield and soil health forecasting.
3. **Hybrid Edge–Cloud AI Optimization:** Explore lightweight TinyML and quantized neural networks to improve performance on low-power edge hardware.
4. **AI-Driven Decision Support Dashboards:** Develop farmer-centric interfaces that visualize predictions alongside actionable recommendations (e.g., irrigation scheduling, price timing).
5. **Socioeconomic Impact Evaluation:** Conduct longitudinal studies to assess the economic, environmental, and social benefits of AI-powered systems for rural farmers.

D. Concluding Remarks

The integration of open-source AI, IoT, and edge computing provides a transformative pathway for enabling data-driven, sustainable, and inclusive agriculture in tropical West Africa.

The presented framework demonstrates that predictive intelligence—when accessible, interpretable, and locally deployable—can empower smallholder farmers to make proactive decisions, reduce losses, and increase profitability.

In conclusion, this research contributes not only a functional technical solution but also a blueprint for regional agricultural digitalization, fostering resilience, food security, and innovation-driven rural development.

This paper establishes a comprehensive AI-based predictive system that bridges the gap between sensing, analytics, and decision-making in agriculture.

By leveraging open-source technologies and edge computation, the framework offers a replicable, sustainable, and socially relevant model for transforming smallholder farming in tropical West Africa.

References:

- *E. Ndhlovu, "Artificial Intelligence and Data Science in African Food Systems," IAHRW Int. J. of Social Sciences, 2025.*
- *G. S. Mmbando, "Harnessing Artificial Intelligence and Remote Sensing in Climate-Smart Agriculture," Cogent Food & Agriculture, vol. 11, no. 1, 2025.*
- *KA. Nnanguma, "Integrating Artificial Intelligence for Climate-Smart Agriculture and Sustainable Food Systems," ASEAN J. of Agriculture and Food Economics, 2025.*
- *OPC. Ugwu, FC. Ogenyi, EU. Alum, and VHU. Eze, "Implementing Artificial Intelligence and Machine Learning Algorithms for Optimized Crop Management," Cogent Food & Agriculture, 2025.*
- *GC. Hernández Hernández and J. Gómez Gómez, "Predictive Models Based on Artificial Intelligence to Estimate Crop Yield: A Literature Review," Agriculture (MDPI), vol. 15, no. 23, 2025.*
- *N. Ozor, J. Nwakaire, and A. Nyambane, "Enhancing Africa's Agriculture and Food Systems through Responsible and Gender-Inclusive AI Innovation," Frontiers in Artificial Intelligence, vol. 8, 2025.*
- *B. Bali, Z. N. Ngida, and I. Habila, "Artificial Intelligence Applied in Optimization of Crop Yield for Ensuring Food Security in Nigeria," SN Computer Science, vol. 6, no. 2, 2025.*
- *T. Begho, T. P. Daubry, and A. E. Irabor, "A Scoping Review of Artificial Intelligence and the Future of African Agriculture," Advances in Agriculture, 2025.*
- *A. Gwagwa, E. Kazim, P. Kachidza, and A. Hilliard, "Road Map for Responsible AI for Development in African Countries: Case Study of Agriculture," Patterns, vol. 2, no. 8, 2021.*
- *M. A. Sodeeq, "Machine Learning and Economic Transformation in Africa," Int. J. of Applied Knowledge and Innovation, 2025.*
- *U. Chukwuma, K. G. Gebremedhin, and D. D. Uyeh, "Imagining AI-Driven Decision Making for Managing Farming in Developing Economies," Computers and Electronics in Agriculture, 2024.*
- *J. Kanyepe, M. Chibaro, and M. Morima, "AI-Powered Agricultural Supply Chains: Applications, Challenges, and Opportunities," IGI Global Artificial Intelligence Series, 2025.*

- *A. Hamrani, A. Allouhi, F. Z. Bouarab, and K. Jayachandran, "AI and Robotics in Agriculture: A Systematic and Quantitative Review," Crops (MDPI), 2025.*
- *R. Aworka, L. S. Cedric, W. Y. H. Adoni, and J. T. Zoueu, "Agricultural Decision System Based on Advanced Machine Learning Models," Smart Agricultural Systems (Elsevier), 2022.*
- *B. I. Oyediji, M. A. Olaitan, and J. Bamidele, "Artificial Intelligence and Agricultural Risk Management for Smallholder Cowpea Farmers," Sustainable Agriculture and AI, 2025.*
- *M. R. Ali, M. Y. Ibrahim, and P. Abubakar, "Predictive Modeling of Maize Yield Using Hybrid Deep Learning Models," Smart Agricultural Systems (Elsevier), vol. 4, 2025.*
- *R. R. Lamsal, P. Karthikeyan, P. Otero, and A. Ariza, "Design and Implementation of IoT-Integrated Open-Source AI Platform for Smallholder Farms," Agriculture (MDPI), vol. 13, no. 10, 2023.*
- *M. O. Tchakounte, D. N. Mouliom, and J. K. Nganhou, "Hybrid Machine Learning Models for Cassava Yield Prediction in Cameroon," African Journal of Agricultural Science and Technology, 2024.*
- *J. Karanja, C. Omollo, and R. Akoth, "Computational Constraints of AI for Precision Agriculture in Africa," ICT for Development Journal, vol. 15, no. 2, pp. 55–68, 2024.*
- *S. Patil, R. Singh, and A. Patel, "TensorFlow and Flask-Based Predictive Irrigation Framework," International Journal of Smart Agriculture and Environment, vol. 14, 2024.*
- *A. Abate, D. Weldegiorgis, and H. Gebre, "Reproducible Machine Learning for Agricultural Disease Detection Using Open Frameworks," Data Science in Agriculture, vol. 3, pp. 88–104, 2023.*
- *O. Adebola, F. Salisu, and E. Odigie, "AI in Africa's Agricultural Transformation: Case Studies from Kenya and Nigeria," African AI and Sustainability Review, vol. 2, 2025.*
- *K. Owusu, E. Amoah, and R. Adomako, "AI-Based Soil Health Prediction for Sustainable Agriculture in Ghana," Computers and Electronics in Agriculture, vol. 215, p. 108281, 2025.*
- *P. Ndlovu, S. Moyo, and A. Sibanda, "Lightweight Edge AI for Smallholder Yield Estimation in Sub-Saharan Africa," Sensors (Basel), vol. 24, no. 6, p. 3458, 2025.*

- *T. Adegbola, L. A. Bello, and K. Y. Ogunde, "Forecasting Agricultural Commodity Prices in West Africa Using Prophet and LSTM Models," Cogent Economics & Finance, vol. 13, no. 1, p. 2198754, 2025.*
- *C. Ebenuwa, J. I. Igboke, and M. O. Madu, "AI-Driven Cassava Price Forecasting for Market Optimization in Nigeria," International Journal of Agricultural Data Science, vol. 3, 2025.*
- *E. Uzochukwu, T. U. Ugochukwu, and I. P. Onyekachi, "AI-Powered Price Prediction Tools for Smallholder Empowerment," Journal of Applied Artificial Intelligence in Agriculture, vol. 5, pp. 44–61, 2025.*