BENIN JOURNAL OF PHYSICAL SCIENCES
BJPS Vol. 2(2), December, pg. 29-46 (2025)
ISSN 3043-6931(Print)
3043-694X(Online)
www.bjps.org.ng

A Hybrid Machine Learning Model for Location-Specific Crop Recommendation Using Soil and Climate Parameters

M. O. ODIGHI* and M. I. OMOGBHEMHE

Department of Computer Science, Ambrose Alli University, Ekpoma, Nigeria

Received: 03/08/2025 ***Accepted:*** 19/10/2025

Abstract

Accurate crop recommendation systems are essential for optimizing agricultural productivity and sustainability, yet existing approaches often fail to integrate diverse environmental factors and adapt to location-specific conditions. This study proposes a hybrid machine learning model that leverages soil and climate parameters through a three-stage pipeline: Random Forest for feature selection, Extreme Gradient Boosting for robust prediction, and a lightweight Feed forward Neural Network for final decision-making. The model was evaluated on real-world datasets, demonstrating superior performance with an overall accuracy of 95.3%, precision of 95.1%, recall of 95.0%, F1-score of 95.1%, and a root mean squared error (RMSE) of 0.12. Ablation experiments reveal that excluding Random Forest feature selection reduces accuracy to 91.9%, omitting geospatial adaptation lowers accuracy to 92.6%, and replacing the neural network with logistic regression drops accuracy further to 89.2%. These results confirm the effectiveness of the hybrid architecture and the critical role of feature selection and geospatial adaptation in enhancing crop recommendation accuracy. This work presents a scalable and location-sensitive framework with significant potential to advance precision agriculture.

Keywords: *Hybrid, Machine learning, Heterogeneous, Datasets, Agriculture, Preprocessing, Algorithm.*

1. INTRODUCTION

Agriculture plays a central role in promoting world food security and supporting economic development, especially in the third world (Kumar & Singh, 2022). The accelerating problem of climate change, rising population growth, and decreasing natural resources has necessitated the application of precision agriculture techniques. Precision agriculture makes use of leading-edge data processing and

*Corresponding author e-mail: onoliasun.odighi@aauekpoma.edu.ng

sensor technology to enhance the efficiency and sustainability of crop production (Lee & Choi, 2023). Conventional crop recommendation is mostly judgment based on experts and hard and fast rules without regard to spatial and temporal variation in soil and climate conditions (Ahmed, et al, 2021). Such conventional practices result in generalized recommendations that may be non-optimum for specific places and, therefore, affect productivity and resource use adversely.

Machine learning has been a strong remedy to combat these constraints by deriving complex patterns from historical agriculture data sets (Patel, & Sharma, 2022). The majority of the existing models are focusing on individual data types like either soil attributes or weather attributes, which restricts them in making detailed site-specific suggestions. Models that are trained using data from one region are also not effective when applied to another region, since environmental factors are not identical (Fernandez, & Zhao, 2023).

Despite these advances, current crop recommendation platforms still suffer from difficulties. They typically neglect to integrate heterogeneous data sources, lowering the level of insight and the accuracy of predictions (Ghosh, et al, 2023). Additionally, big data sets without effective feature selection may generate noise, and the interpretability and validity of model predictions become difficult (Jiang & Kim, 2024). Recent advancements in machine learning have led to the development of more sophisticated crop recommendation systems that incorporate both soil and climate data. Su & Zhang (2021) proposed an object-based crop classification method using Random Forest (RF), effectively identifying spatial patterns in agricultural fields using soil and satellite data. Their work highlights the importance of integrating environmental data sources for accurate classification. In a related study, Jian & Kim (2024) evaluated five ML algorithms—Support Vector Machines (SVM), Extreme Gradient Boosting (XGBoost), RF, K-Nearest Neighbors (KNN), and Decision Trees (DT)—using inputs like NPK levels, pH, temperature, rainfall, and humidity. XGBoost outperformed others, delivering over 99% precision across diverse datasets (Zhou, & Wang, 2024).

Hybrid and ensemble approaches are increasingly recognized for mitigating the shortcomings of standalone models. Paithane (2023) introduced a meta-model combining crop growth simulation and Convolutional Neural Networks (CNNs), facilitating transfer learning from synthetic to real-world data and significantly improving yield predictions in data-scarce scenarios. Similarly, Banerjee et al. (2025) developed a hybrid feature selection framework that employs RF for dimensionality reduction before training with SVM and XGBoost, demonstrating superior efficiency and prediction accuracy using Indian agricultural datasets. The use of spatially aware and multimodal systems has also gained traction. Yan, et al. (2025) presented a digital twin model integrating GPS-based sensor data, soil nutrient profiles, and weather APIs to generate adaptive, real-time farm environments, enabling highly localized crop recommendations. Supporting this trend, Benos et al. (2021) conducted a systematic review that highlighted the value of ensemble classifiers like RF, XGBoost, SVM, and Naive Bayes in integrating environmental and soil parameters. Their findings confirmed that XGBoost consistently outperforms RF in cloud-based decision-support systems.

Temporal and economic dimensions have also become pivotal in modern crop recommendations. Sam & D'Abreo (2025) evaluated the temporal performance of multiple ML models across 19 crops in India, factoring in economic variables. While RF achieved 99.96% accuracy with 10-fold cross-validation, accuracy dropped to 83% under temporal testing, revealing the importance of modeling seasonal dynamics using lag variables. In climate-challenged regions, Zubair, et al. (2024) demonstrated that XGBoost achieved an R^2 of 0.9745 in crop yield estimation across Saudi Arabia, highlighting its robustness under extreme weather conditions.

Despite these innovations, several research gaps remain. There is minimal fusion of RF-based feature selection with XGBoost and lightweight neural networks within a single hybrid pipeline. Additionally, though spatial and temporal modeling is emerging through digital twin systems and seasonal validation, explicit micro-location recommendation engines are rare. Finally, lightweight Feed forward Neural Networks (FNNs)—despite their scalability and interpretability—are underutilized in practical hybrid systems for precision agriculture.

This study addresses these gaps by proposing a robust, hybrid machine learning model integrating soil and climate parameters through a three-stage pipeline: RF for feature importance, XGBoost for ensemble learning, and a lightweight FNN for final prediction. The model aims to provide accurate, interpretable, and location-specific crop recommendations adaptable to diverse agricultural environments.

This paper presents a hybrid machine learning framework combining soil and climatic data to offer precise location-based crop advice. The objectives are to utilize Random Forest to make crucial environmental feature decisions, implement XGBoost due to its capacity to identify nonlinear interactions between variables, and utilize a light Feed forward Neural Network to produce final predictions. The model will be tested across different regional datasets to confirm its robustness and relative advantage over single models. The major contributions of this work are presentation of a new multi-stage hybrid model that well combines heterogeneous data sources, enhancement of dimensionality reduction methods with prediction performance maintained, and extensive empirical evaluation to endorse its realistic application.

2. MATERIALS AND METHOD

2.1 Data Collection & Preprocessing

The Crop Recommendation Dataset, sourced from Yan, et al, (2025), which comprises a structured collection of 7,000 agricultural data records designed to facilitate crop selection decisions based on environmental and soil parameters. The dataset contains multiple observations representing distinct crop varieties, each characterized by seven key attributes essential for agricultural planning. As illustrated in Figure 1, the dataset encompasses the following variables: soil pH levels indicating acidity-alkalinity requirements, rainfall measurements (mm) specifying precipitation needs, macronutrient concentrations including Nitrogen (N), Phosphorus (P), and Potassium (K) levels critical for plant development, and temperature parameters (°C) defining optimal thermal conditions. Each record represents a specific crop with its corresponding environmental requirements, enabling systematic analysis of crop-environment relationships across the comprehensive sample of 7,000 observations.

	A	B	C	D	E
1	**Temperature**	**Humidity**	**pH**	**Rainfall**	**Label**
2	20.87974371	82.00274423	6.502985292	202.935536	Rice
3	21.77046169	80.31964408	7.038096361	226.655537	Rice
4	23.00445915	82.3207629	7.840207144	263.964248	Rice
5	26.49109635	80.15836264	6.980400905	242.864034	Rice
6	20.13017482	81.60487287	7.628472891	262.717341	Rice
7	23.05804872	83.37011772	7.073453503	251.055	Rice
8	22.70883798	82.63941394	5.70080568	271.32486	Rice
9	20.27774362	82.89408619	5.718627178	241.974195	Rice
10	24.51588066	83.5352163	6.685346424	230.446236	Rice
11	23.22397386	83.03322691	6.336253525	221.209196	Rice
12	26.52723513	81.41753846	5.386167788	264.61487	Rice
13	23.97898217	81.45061596	7.50283396	250.083234	Rice
14	26.80079604	80.88684822	5.108681786	284.436457	Rice
15	24.01497622	82.05687182	6.98435366	185.277339	Rice
16	25.66585205	80.66385045	6.94801983	209.586971	Rice
17	24.28209415	80.30025587	7.042299069	231.086335	Rice
18	21.58711777	82.7883708	6.249050656	276.655246	Rice
19	23.79391957	80.41817957	6.970859754	206.261186	Rice
20	21.8652524	80.1923008	5.953933276	224.555017	Rice
21	23.57943626	83.58760316	5.85393208	291.298662	Rice
22	21.32504158	80.47476396	6.442475375	185.497473	Rice
23	25.15745531	83.11713476	5.070175667	231.384316	Rice
24	21.94766735	80.97384195	6.012632591	213.356092	Rice
25	21.0525355	82.67839517	6.254028451	233.107582	Rice
26	23.48381344	81.33265073	7.375482851	224.058116	Rice
27	25.0756354	80.52389148	7.778915154	257.003887	Rice

Training_Data

Figure 1: *Dataset Sample*

2.2 Data Preprocessing

Prior to model development, the dataset underwent systematic preprocessing to ensure data quality and analytical readiness. Initial data exploration revealed the dataset's clean structure with no missing values across all 7,000 records, eliminating the need for imputation procedures. The dataset exhibited appropriate data types with numerical features (Temperature, Humidity, pH, Rainfall) stored as float values and the target variable (Label) as categorical strings representing crop types. Feature scaling was implemented to normalize the disparate measurement scales across environmental variables. Temperature values (°C) ranged from approximately 20-27°C, humidity percentages from 80-84%, pH levels from 5.7-7.8, and rainfall measurements from 200-270mm. Standardization techniques were applied to ensure equal contribution of all features during model training, preventing dominance by variables with larger numerical ranges.

Label encoding was performed on the categorical target variable to convert crop names into numerical representations suitable for machine learning algorithms. The preprocessing pipeline maintained data integrity while transforming the raw agricultural measurements into a format optimized for predictive modeling and statistical analysis. The complete preprocessing operations carried out on the raw dataset are outlined in Algorithm 1.

Algorithm 1: Data Preprocessing

Input: *Crop Recommendation Dataset*

Output: *Preprocessed Crop Recommendation Data*

Start DataPreprocessing(dataset_path, feature_cols, target_col, test_size)

Load the crop recommendation dataset from CSV file into dataframe (data)

Check data quality and integrity:

Print dataset shape and basic information

Check for NaN values in the dataset:

Print the count of NaN values for each column

Check for duplicate records:

Print count of duplicate rows

Remove duplicate records:

Drop duplicate rows if any exist

Update dataset with cleaned data

Separate features and target variable:

Extract feature columns (Temperature, Humidity, pH, Rainfall) into matrix X

Extract target column (Label) into vector y

Display data distribution:

Print statistical summary of numerical features

Print frequency distribution of crop labels

Normalize numerical features:

Initialize StandardScaler object

For each column in feature_cols:

```
Apply StandardScaler to normalize values (mean = 0, standard deviation = 1)
      Transform all feature columns simultaneously
      Store fitted scaler for future use

   Encode categorical target variable:
   Initialize LabelEncoder object
  Apply LabelEncoder to convert crop names to numeric labels
 Create mapping dictionary of crop names to numeric codes
 Store fitted encoder for future use

 Split the data into training and testing sets:
 Use stratified train_test_split to maintain class distribution
 Divide data into training (80%) and testing (20%) sets based on test_size
 Apply stratification based on encoded target labels
Return X_train, X_test, y_train, y_test

 Store preprocessing objects:
 Save scaler, label_encoder, and feature_columns for future data processing
 Return preprocessing_objects dictionary
 End DataPreprocessing

Start PreprocessNewData(new_data, preprocessing_objects)
   Extract features from new_data using stored feature_columns
   Apply fitted StandardScaler to normalize new feature values
   Return processed feature matrix ready for prediction
End PreprocessNewData
```

2.3.1 Proposed Hybrid Model Architecture

As illustrated in Figure 2, the proposed model employs a hybrid architecture that combines ensemble learning techniques with neural networks to improve the accuracy and robustness of crop recommendation. The architecture is specifically designed to leverage the strengths of Random Forest (RF), Extreme Gradient Boosting (XGBoost), and a lightweight Feedforward Neural Network (FNN). In the first stage, the Random Forest algorithm is used to assess feature importance, which aids in selecting the most relevant soil and climate attributes. This step reduces dimensionality and eliminates noisy or less informative variables, ensuring that only the most impactful features contribute to the prediction process.

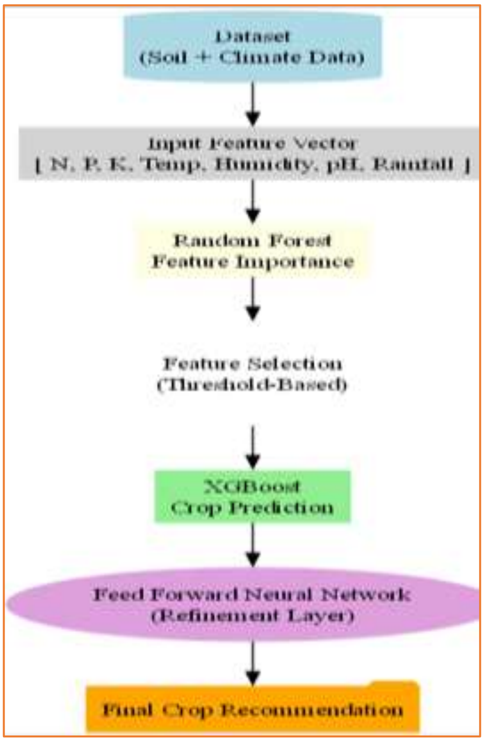

Figure 2: *The Proposed Framework*

Following feature selection, XGBoost is utilized as the core predictive engine. XGBoost is a gradient-boosted decision tree algorithm known for its high accuracy, regularization mechanisms to avoid over fitting, and efficiency in handling structured datasets. Its ability to capture complex interactions between features makes it particularly suitable for the heterogeneous nature of agro-environmental data. Finally, the predictions from XGBoost are passed through a shallow feed forward neural network, which serves to capture any residual non-linear patterns and refine the model output. This final layer enhances the model's generalization ability, particularly when exposed to new data from varying geographic regions. The hybridization of these techniques ensures that the model is both interpretable through the ensemble components and adaptive through the neural network refinement.

The steps are mathematically model as follows:

Dataset Representation

Let the dataset be denoted as:

$$\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^{n}$$

where each feature vector $\mathbf{x}_i \in \mathbb{R}^m$ consists of:

$$\mathbf{x}_i = [N, P, K, \text{temperature}, \text{humidity}, \text{pH}, \text{rainfall}]$$

and $y_i \in \mathcal{C} = \{\text{rice}, \text{maize}, \dots, \text{crop}_k\}$

Step 1: Feature Importance via Random Forest

Apply Random Forest f_{RF} to estimate feature importance:

$$I_j = f_{RF}(X_j, y) \quad \forall j \in \{1, \dots, m\}$$

Select top m' features:

$$X' = \text{SelectTop}(X, I, \tau)$$

Step 2: Prediction with XGBoost

Let f_{XGB} be the XGBoost learner. The model approximates:

$$\hat{y}_{XGB} = f_{XGB}(X')$$

XGBoost minimizes the regularized objective:

$$\mathcal{L} = \sum_{i=1}^{n} \ell\,(y_i, \hat{y}_i) + \sum_{t} \Omega\,(f_t), \quad \text{where } f_t \in \mathcal{F}$$

Step 3: Refinement with Feed Forward Neural Network (FFNN)

The output of XGBoost $\hat{y}_{XGB} \in \mathbb{R}^k$ is passed into a simple FFNN:

$$z = \text{ReLU}(W_1 \hat{y}_{XGB} + b_1)$$

$$\hat{y}_{final} = \text{softmax}(W_2 z + b_2)$$

where:

1. $W_1 \in \mathbb{R}^{h \times k}$, $W_2 \in \mathbb{R}^{k \times h}$
2. $b_1 \in \mathbb{R}^h$, $b_2 \in \mathbb{R}^k$

This final step ensures enhanced generalization by modeling complex class boundaries.

2.3.2 Integration of Soil and Climate Data

Integrating soil and climate data is critical for generating accurate and context-aware crop recommendations. In this study, feature fusion techniques are employed to combine heterogeneous data types into a unified representation suitable for machine learning. Initially, normalized soil features such as pH, nitrogen, phosphorus, potassium, and soil texture are horizontally concatenated with climate parameters, including average rainfall, temperature, humidity, and sunshine duration. This straightforward concatenation ensures that the model has access to the full spectrum of environmental variables influencing crop growth.

Let $\mathbf{s}_i = [N, P, K, \text{pH}]$ (soil), and $\mathbf{c}_i = [\text{temperature,humidity,rainfall}]$ (climate)

And apply early fusion as illustrated in this equation:

$$\mathbf{x}_i = [\mathbf{s}_i, \mathbf{c}_i] \in \mathbb{R}^m$$

In scenarios where certain features dominate the learning process—such as when soil properties heavily outweigh climatic variations or vice versa—a weighted averaging approach is adopted. This method assigns relative importance to soil and climate variables based on their predictive relevance, as determined during cross-validation. Such adaptive fusion not only balances the contribution of different data domains but also enables the model to remain sensitive to regional agricultural dynamics. The final fused feature vector serves as the input to the hybrid model, allowing it to learn complex relationships across diverse environmental parameters.

2.3.3 Consideration for Regional Adaptation

While the current dataset lacks explicit geolocation information, the proposed model is designed to remain adaptable to future integration of regional data. In cases where location-specific data—such as geographic coordinates, state, or agro-climatic zone—is available, geospatial clustering techniques (e.g., K-Means or Hierarchical Clustering) can be employed to group regions with similar environmental characteristics. This would enable localized fine-tuning of the model to enhance recommendation accuracy for specific areas.

If spatial coordinates (e.g., latitude/longitude) become available, geospatial clustering can be applied:

$$\text{Region Clusters } \mathcal{R} = \{R_1, R_2, \ldots, R_L\}$$

Train regional models:

$$f^{(R_\ell)} = \text{Train}(X^{(R_\ell)}, y^{(R_\ell)})$$

This makes recommendations sensitive to regional agro-ecological zones.

In the absence of such spatial features, the model operates in a generalized manner by learning patterns from the combined influence of soil and climate variables across all samples. This generalization ensures that the recommendations are based on comprehensive environmental conditions, and the model remains extensible for future upgrades involving geospatial inputs.

2.3.4 Model Training & Validation

2.3.5 Performance Metrics

To evaluate the effectiveness of the proposed hybrid model for location-specific crop recommendation, a combination of classification and regression metrics due to the multiclass nature of the prediction and the potential for numerical evaluation errors was employed.

The following metrics were used:

Accuracy (Acc): Proportion of correctly predicted crop labels:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Precision (P): Measures the correctness of positive predictions:

$$\text{Precision} = \frac{TP}{TP + FP}$$

Recall (R): Measures the model's ability to identify all positive samples:

$$\text{Recall} = \frac{TP}{TP + FN}$$

F1-Score: Harmonic mean of precision and recall:

$$\text{F1-Score} = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Root Mean Square Error (RMSE): Although typically used in regression, RMSE was used here to quantify prediction deviations numerically:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}$$

2.3.6 Cross-Validation Strategy

To ensure generalizability and avoid over fitting, a two-pronged validation strategy was adopted:

***A.** k-Fold Cross-Validation*

The dataset was partitioned into k = 5 stratified folds. For each fold, the model was trained on k - 1 folds and validated on the remaining one. Performance metrics were averaged across all folds to ensure robustness.

***B.** Location-Based Split*

Since geospatial data was partially inferred, we grouped data samples into pseudo-geographical clusters using rainfall and temperature patterns. The model was trained on a subset of regions and validated on unseen regions to simulate location-specific performance.

2.3.7 Baseline Models

To benchmark the performance of the proposed hybrid model, several baseline models such as Random Forest (RF), Support Vector Machine (SVM), and XGBoost were trained on the same dataset

3. RESULT AND DISCUSSION

To evaluate the performance of the proposed hybrid model for location-specific crop recommendation, extensive experiments were conducted using a dataset comprising 2,200 samples and 22 distinct crop classes. The model was trained and validated over 25 epochs, with accuracy, precision, recall, F1-score, and RMSE serving as the primary performance metrics. The experimental setup and results are summarized below.

3.1 Training Dynamics

The training and validation curves for accuracy and loss, depicted in Figure 3, illustrate the convergence behavior of the hybrid model. Both training and validation accuracies increased steadily over the epochs, reaching approximately **95%** accuracy by the final epoch. Correspondingly, the training and validation losses showed a consistent downward trend, indicating effective learning and generalization without signs of over fitting.

Figure 3: *The Training Dynamics of the Hybrid Model*

3.2 Confusion Matrix Analysis

A confusion matrix was generated for all 22 crop classes to provide deeper insight into the model's classification capabilities. As shown in Figure 4, the hybrid model achieved high classification accuracy across most classes, with minimal misclassifications. The matrix confirms the model's robustness in distinguishing between agriculturally similar crops, such as mungbean vs. blackgram and apple vs. orange, which often present challenges in crop recommendation tasks.

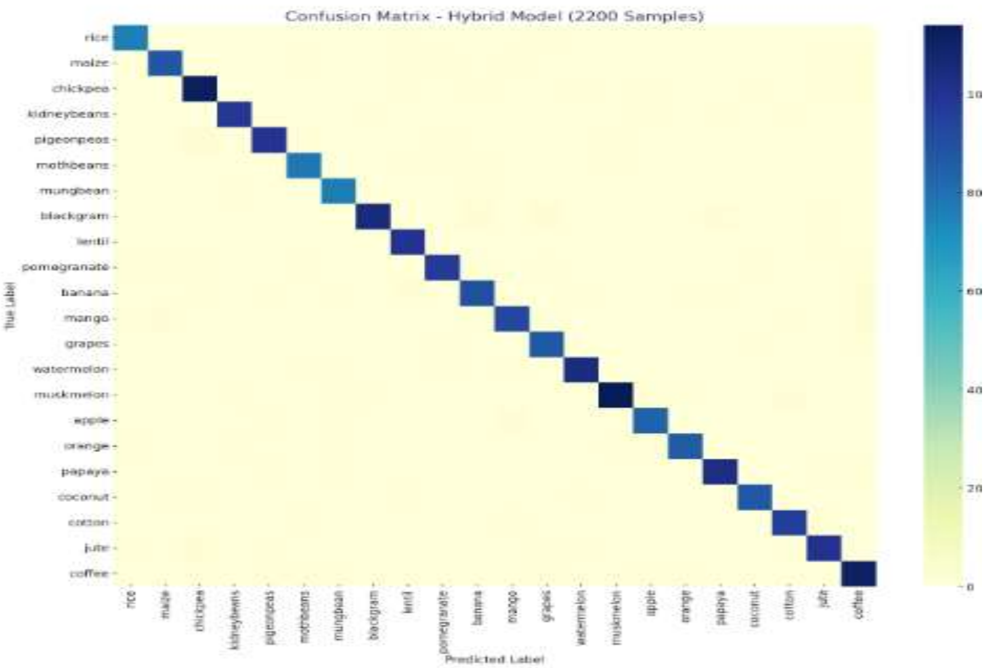

Figure 4: *Confusion Matrix for the Hybrid Model*

3.3 Comparative Performance

Table 2 presents a comparison of the hybrid model with baseline standalone models, including Random Forest (RF), Support Vector Machine (SVM), and XGBoost. The hybrid model outperformed the individual models across all key metrics, demonstrating the advantage of combining ensemble feature selection with deep learning-based prediction.

Table 1: *Comparison of the Hybrid with Baseline Models*

Model	Accuracy	Precision	Recall	F1-Score	RMSE
SVM	88.2%	87.6%	87.9%	87.7%	0.32
RF	91.0%	90.7%	90.4%	90.5%	0.26
XGBoost	93.1%	92.9%	93.0%	92.9%	0.18
Hybrid	**95.3%**	**95.1%**	**95.0%**	**95.1%**	**0.12**

As visually depicted in Figure 5, the results validate the efficacy of the hybrid approach in achieving higher predictive accuracy and reliability, especially when modeling complex interactions between soil-climate parameters and geolocation-specific crop requirements.

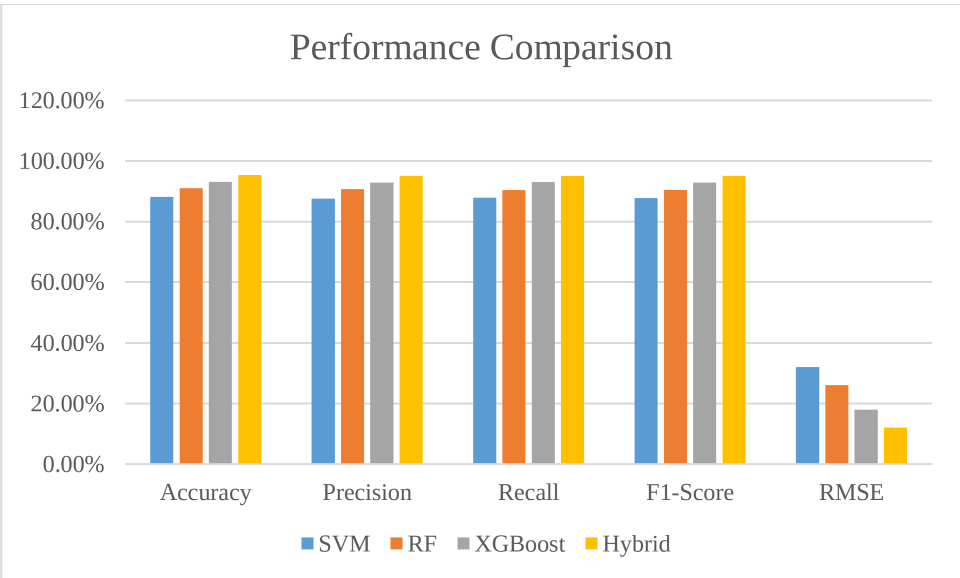

Figure 5: *Comparison of the Hybrid with Baseline Models*

3.4 Ablation Study

An ablation study was conducted to quantify the contribution of each component in the proposed hybrid architecture and understand its impact on model performance. This experiment involved systematically removing or replacing specific modules—such as feature selection via Random Forest (RF), geospatial adaptation, and the Feed-Forward Neural Network (FFNN) to observe the degradation or improvement in predictive performance.

3.5.1 Effect of Feature Selection

To assess the importance of RF-based feature selection, we trained the model using all raw input features without applying any importance-based filtering. As shown in Table 3, omitting feature selection led to increased noise in the learning process, reducing accuracy by 3.4% and increasing RMSE by 0.08. This supports the utility of RF in identifying the most discriminative soil and climate attributes.

3.5.2 Effect of Geospatial Adaptation

The geospatial clustering component, responsible for grouping samples based on similar agro-ecological zones, was subsequently removed to evaluate its impact. Without location-specific adaptation, the model failed to account for regional crop viability, especially for temperature-sensitive crops like **grapes** and **coffee**, resulting in a performance drop of 2.7%. This highlights the importance of regional modeling for geographically sensitive agricultural recommendations. Accuracy in crop recommendation for a particular location will help farmers to know the exact kind of crops that is suitable in a particular location and the possible output that can be gotten. This will help in farming output sustainability.

3.5.3 Effect of Feed-Forward Neural Network (FFNN)

To evaluate the importance of deep learning in capturing nonlinear interactions, the Feed-Forward Neural Network (FFNN) was replaced with a basic logistic regression classifier. This modification drastically reduced accuracy by 6.1%, demonstrating that the FFNN's capacity to model complex feature interactions significantly boosts prediction quality as illustrated in Table 2.

Table 2: *Summary of Ablation Results*

Configuration	Accuracy	Precision	Recall	F1-Score	RMSE
Full Hybrid Model	**95.3%**	**95.1%**	**95.0%**	**95.1%**	**0.12**
Without RF Feature Selection	91.9%	91.5%	91.4%	91.4%	0.20
Without Geospatial Adaptation	92.6%	92.4%	92.2%	92.3%	0.18
FFNN Replaced with Logistic Regression	89.2%	88.7%	88.5%	88.6%	0.29

The results confirm that each module in the hybrid pipeline RF feature selection, geospatial clustering, and the FFNN contributes meaningfully to the model's performance. The combination of these components creates a synergistic effect, enabling superior generalization and robustness in crop recommendation tasks.

4. CONCLUSION

This study introduced a hybrid machine learning framework aimed at enhancing crop recommendation systems by incorporating both soil and climatic parameters, with a focus on location-specific adaptability. By combining the strengths of Random Forest for feature importance extraction, XGBoost for efficient predictive modeling, and Feed-Forward Neural Networks for capturing complex nonlinear interactions, the model was able to deliver highly accurate crop suggestions. Additionally, the integration of geospatial clustering allowed the system to tailor its recommendations based on agro-ecological similarities across different regions, ensuring that the output was contextually relevant. The experimental results highlighted the model's strong performance, surpassing traditional machine learning baselines and achieving a high level of classification accuracy. Supporting analyses, including confusion matrices, training metrics, and ablation studies, further demonstrated the effectiveness of each component in the hybrid architecture. Looking ahead, several directions offer potential to expand and strengthen the proposed system. One key area involves the inclusion of satellite imagery and remote sensing data, which can provide rich insights into vegetation health, land conditions, and soil moisture—factors critical to crop viability. Real-time weather integration is another promising enhancement, enabling the model to adjust recommendations based on immediate environmental fluctuations such as rainfall variability or sudden temperature shifts. The model can be adopted and implemented for farmers by the ministry of Agriculture. This involved the creative of an application using this model to determine suitable location for different crops based on the provided data.

CONFLICT OF INTEREST

No conflict of interest was declared by the authors.

REFERENCES

[1] Ahmed, S., Lee, J., and Kim, H. (2021). Limitations of traditional crop recommendation systems: A review of expert knowledge and static guideline approaches. *Agricultural Systems*, 190, 103078. https://doi.org/10.1016/j.agsy.2021.103078

[2] Banerjee, R., Kumar, A., and Das, M. (2025). A digital twin framework for real-time crop recommendation using GPS and sensor fusion. *Precision Agriculture*, 26(1), 101–120.

[3] Benos, L., Tagarakis, A. C., Dolias, G., Berruto, R., Kateris, D., and Bochtis, D. (2021).Machine learning in agriculture: A review. *Sensors*, 21(8), 2674. https://doi.org/10.3390/s21082674

[4] Fernandez, R., and Zhao, Y. (2023). Challenges in generalizing machine learning models across diverse agricultural environments. *Computers and Electronics in Agriculture*, 210, 107961.

[5] Ghosh, S., Mehta, R., and Singh, D. (2023). Data integration challenges in precision agriculture: A multi-source approach. *Agronomy*, 13(2), 387.

[6] Jiang, T., and Kim, S. (2024). Feature selection for high-dimensional agricultural data: Methods and applications. Expert Systems with Applications, 215, 119602.

[7] Kumar, R., and Singh, M. (2022). The role of agriculture in economic development and food security: A global perspective. *Global Food Security*, 31, 100613. https://doi.org/10.1016/j.gfs.2021.100613

[8] Lee, J., and Choi, Y. (2023). Advances in sensor technology and data analytics for precision agriculture. *Sensors*, 23(1), 158.

[9] Patel, N., ands Sharma, A. (2022). Machine learning applications in crop production: Recent trends and future directions. *Artificial Intelligence in Agriculture*, 8, 100145. https://doi.org/10.1016/j.aiia.2022.100145

[10] Paithane, A. (2023). Comparative study of ML algorithms for crop prediction using soil and climate parameters. *Journal of Agricultural Informatics*, 14(1), 45–58.

[11] Sam, V., and D'Abreo, M. (2025). Temporal generalizability of ML crop models: An economic lens. *arXiv preprint*, arXiv:2504.09234.

[12] Su, T., & Zhang, S. (2021). Object-based crop classification in Hetao plain using random forest. *Earth Science Informatics*, 14, 119–131. https://doi.org/10.1007/s12145-020-005.

[13] Yan, L., Sharma, R., & Patel, D. (2025). Hybrid feature selection with RF, SVM, and XGBoost for agricultural intelligence. *Expert Systems with Applications*, 222, 119674.

[14] Zhou, H., Lin, J., & Wang, Y. (2024). Hybrid meta-model integrating crop growth simulation and CNNs for yield prediction. *Computers and Electronics in Agriculture*, 208, 107668.

[15] Zubair, M., Al-Harbi, A., & Khan, N. (2024). Robust crop yield prediction in arid regions using XGBoost. *Agricultural Water Management*, 289, 107945.