Received: 29 April 2025 | Revised: 13 August 2025 | Accepted: 19 August 2025 | Published online: 26 September 2025

RESEARCH ARTICLE

Journal of Computational and Cognitive Engineering
2025, Vol. 00(00) 1-15
DOI: 10.47852/bonviewJCCE52026049

BON VIEW PUBLISHING

Revolutionizing Agriculture Through Smart Farming by Employing Advanced Machine Learning Techniques for Optimal Crop Selection

Rajesh Natarajan[1,*], Sujatha Krishna[1], Pradeepa Ganesan[1], Amna Al Kaabi[1], Rajesh Kala[1] and Anthony Mendoza Madlambayan[1]

[1] *Department of Computing and Information Sciences, University of Technology and Applied Sciences, Oman*

Abstract: Precision agriculture relies heavily on crop selection to improve sustainability and increase productivity. The increasing problems of soil erosion, climate change, and water shortages have made it crucial to optimize crop selection through cutting-edge methods to increase farm yields and enhance resource efficiency. This research attempts to develop machine learning (ML) models, such as logistic regression (LR), Gaussian Naive Bayes (GNB), support vector classifier (SVC), K-nearest neighbors (KNN) classifier, decision tree (DT) classifier, extra tree (ET) classifier, random forest (RF) classifier, and bagging classifier, to optimize crop selection for precision agricultural systems. A large dataset comprising information on crop recommendations, weather patterns, and soil characteristics was used in this research. The data is preprocessed using the interquartile range (IQR) method to remove outliers and ensure that all features contribute equally to the model. Linear discriminant analysis (LDA) is used to extract the important features for feature extraction. The RF classifier was determined to be the most effective method for raising the precision of crop selection forecasts. This framework is designed to provide actionable insights for selecting optimal crops based on environmental conditions and resource availability. The suggested model is a valuable tool for crop selection optimization in precision agriculture, as preliminary results show that it outperforms traditional ML models in terms of F1-score, precision, accuracy, and recall at 99.88% (all measures). According to the research, modern ML algorithms can revolutionize agricultural practices and provide a sustainable solution to increase crop output while reducing resource wastage.

Keywords: crop selection, machine learning (ML) models, prediction, random forest (RF) classifier, precision agriculture

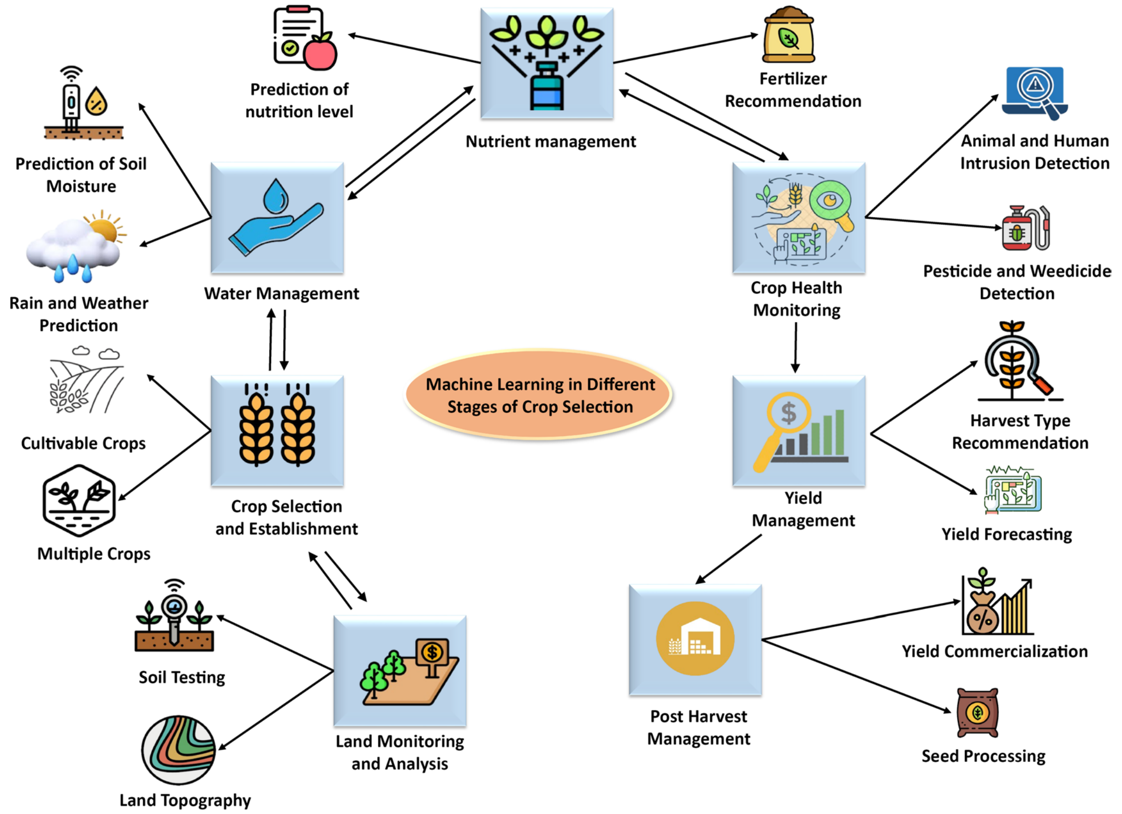
1. Introduction

Several agricultural techniques were used to cultivate crops. Around the world, traditional farming is mainly practiced. In this, methods suggested by experienced farmers are used. These methods are labor intensive and time consuming because they are not precise enough. The use of new methods and technologies to improve agricultural activities is referred to as agricultural revolution [1]. Agricultural activities can help farmers better utilize resources, increase crop yields, and improve soil quality. Agriculture is the most important sector and the foundation of the economies for most countries in the world [2]. To enhance productivity and efficiency in resource utilization, the agricultural industry is increasingly making use of crop simulations and decision-making technologies. Artificial intelligence (AI) can transform the agricultural industry by leveraging modern technologies to predict agricultural productivity [3]. Farmers can improve yields by utilizing AI technologies to monitor commodity prices, manage soil and minerals, identify plant diseases, control pests and weeds, estimate crop yields, select crop varieties, and provide real-time information about agro-product marketing [4]. Precision agriculture can increase productivity, reduce soil deterioration, improve water efficiency, minimize the use of chemicals in cultivation, and promote the use of contemporary farming techniques to reduce costs, improve quality, and increase quantity of crop production [5]. By providing farmers with accurate agricultural yield forecasts for the following year through proper planning for optimal crop growth, issues in crop production may be mitigated. Accurate forecasting of crop yields is now essential for farmers to make wise decisions. Figure 1 shows the challenges of crop selection in ML techniques.

To determine how much crops can be grown in a given area, several factors need to be considered, including crop management techniques, weather, and soil type. A reliable forecast helps estimate crops, which the government can use to formulate short-term and long-term policies aimed at reducing food shortages and import-export strategies centered on the agricultural industry [6]. Farmers can use crop yield estimates to increase production under favorable conditions and reduce production losses under unfavorable ones. Optimistic crop production forecasts are influenced by many different factors, including market pricing, weather, pesticides, fertilizers, along with farmers' actions and procedures [7]. Agricultural yield can be estimated by combining climate, area-wise production, rainfall, and statistical data on yields from previous years. Evaluation of the effects of agro-climatic conditions on the production of winter plant varieties, mainly grains, is the main challenge in the moderate temperature zone [8]. Figure 1 shows the most important challenges for using machine learning (ML) in the selection of crops. It shows elements such as diverse soil conditions, unreliable weather

***Corresponding author:** Rajesh Natarajan, Department of Computing and Information Sciences, University of Technology and Applied Sciences, Oman. Email: rajesh.natarajan@utas.edu.om

Figure 1
Challenges of crop selection in ML techniques



patterns, access to water, and compatibility of crops. These elements pose difficulties in making accurate predictions and require strong preprocessing, feature selection, and model tuning to ensure reasonable and sustainable agricultural decision-making.

Crop growth is influenced by three important factors: season, soil, and water. Soil with high nutrient content and high water-holding capacity helps crops thrive. However, the season-related weather can significantly alter crop output expectations at any time [9]. Farmers find it challenging to determine how to achieve greater adaptability and sustainability due to large climate variability. Agriculture needs to produce more with fewer inputs, according to cutting-edge farming methods and contemporary technologies. Thus, estimating crop yield is essential to determine food security problems [10]. This research aims to optimize crop selection for precision agricultural systems and develop ML models, which include Gaussian Naive Bayes (GNB), logistic regression (LR), K-nearest neighbors (KNN), decision tree (DT) classifier, extra tree (ET) classifier, random forest (RF) classifier, support vector classifier (SVC), and bagging classifier.

1.1. Contribution of research

1) To create and use a large dataset that contains vital details on crop recommendations, weather trends, and soil properties, allowing for accurate analysis to maximize crop selection in precision agriculture.
2) The interquartile range (IQR) approach can be used to preprocess data, eliminate outliers, ensure that all characteristics contribute effectively to ML models, and enhance model performance.
3) The model may focus on the most crucial factors influencing crop selection by using linear discriminant analysis (LDA) to identify the most pertinent features.
4) A range of ML models, including LR, DT, bagging, and RF classifiers, are used to implement sophisticated classification approaches. When it comes to forecasting the crop with the greatest selection, the RF classifier is the most accurate and effective.

This research is organized as follows: Section 2 presents related works on revolutionizing agriculture through smart farming using ML techniques. Section 3 outlines the current methods and clearly explains the ML techniques using the optimal crop selection process. In Section 4, we demonstrate that the proposed model performs better than conventional ML models in terms of prediction accuracy. The conclusion is given in Section 5.

2. Literature Review

In the state of Tamil Nadu, rainfall patterns were the primary determinant of crop output. These patterns were used to determine various crops according to soil pH, temperature, and levels. The research by Lad et al. [11] used an artificial intelligence (AI) approach to assess crop yield. The dependence of the model on regional rainfall patterns, which cannot translate well to other regions with distinct climates, is a major drawback. Choudhary et al. [12] recommended a methodology that allows for crop selection based on environmental and economic considerations to increase agricultural yields and meet the country's growing food needs. To forecast crop yield, the proposed model examined variables such as temperature, humidity, rainfall, soil nutrients, and soil pH. With the help of the model, farmers can maintain soil nutrient levels. The model has limited practical applicability as it ignores dynamic market dynamics and real-time field variability.

The investigation offered health computing using persistent plants based on ML techniques . Without proper ecological care, plant diseases, slower growth rate, and reduced yields can lead to premature harvesting Although several approaches have been created for crop analysis, including at an earlier stage, advanced methods must be implemented. Advanced techniques for early-stage crop analysis and ecological care were not used in the study by Kailasam et al. [13].

To overcome these obstacles, Abdel-salam et al. [14] provided a unique paradigm for forecasting agricultural yields. To effectively improve the prediction accuracy, the framework combined the optimized SVR model with a novel hybrid feature selection technique. The effectiveness of the proposed framework was tested through several experiments. Its drawback was the research focused on a single crop and lacked generalization across multiple crop varieties and climatic circumstances.

El-Kenawy et al. [15] stated that the recent advances in deep learning (DL) and ML have provided fresh approaches for efficiently and accurately generating predictive models. These findings outline how the latest predictive models can encourage sustainable agricultural practices and informed decisions about potato production. The ML-based crop selection model by El-Kenawy et al. was mainly based on meteorological and soil factors. However, due to its focus on a single crop (potatoes) , the model has limited applicability to many agricultural situations. In Rani et al. [16], the RF classifier completed the crop selection process, while DL methods were used for weather analysis. Using the specified data size, the RF classifier generated the model in 5.34 seconds.

The research problem was resolved with the help of a decision support system. According to the research by Apat et al. [17], the AI system aids precision agriculture in increasing the general accuracy and quality of agricultural harvests. A recommendation system was selected as one alternative in this research because Industry 4.0 recommends the use of artificial intelligence (AI) and a family of ML algorithms. The drawback was that in real-world situations, where there are wildly fluctuating or invisible environmental factors, the effectiveness of the system can be reduced.

ML was essential in crop prediction, according to the research by Suruliandi et al. [18]. Climate, geography, and soil characteristics all affect crop predictions. The prediction process carried out by feature selection techniques involved the intrinsic selection of suitable traits for the appropriate crop or crops. The major drawback is that crop prediction can be affected by the complex relationships between climate and soil.

To assess the effects of various feature selection techniques on ML models trained to forecast alfalfa production, Whitmire et al. [19] analyzed the crop yield for several alfalfa types from several years of yield data in Kentucky and Georgia. Even on small datasets with few features, it has been shown that ML with feature selection may be able to forecast crop yields and that the accuracy of R and R^2 reports provides an easy way to contrast results across different crops. The research focused on a single crop (alfalfa) and its use of regional data limits its generalizability to other crops and geographical areas. The study by Tufail et al. [20] demonstrated an ML-based crop identification system for a tractor-mounted boom sprayer that can spray tobacco crops in fields at specific locations. Texture, shape, and color were carefully selected from among the distinctive features of tobacco plants, and an SVM classifier with a 96% classification accuracy was developed. Because the technique is limited to tobacco crops, it may not translate well to other crop types or different field environments.

A comprehensive strategy that combines drones, ML, and the Internet of things (IoT) to monitor crop health was suggested by Shafi et al. [21]. When multiple sensory modalities are combined , heterogeneous data with varying temporal fidelity and feature variation are generated. Rice production data were essential for crop management and food policy. Management and integration of heterogeneous, multisource data with uneven temporal resolution is a weakness of the strategy. An ML technique using time-series data from MODIS (Moderate Resolution Imaging Spectroradiometer) to predict rice crop yield in Taiwan was developed by Son et al. [22]. Despite the possible impact of boundary effects on modeling findings, the research demonstrated how ML systems successfully forecast rice yields in different locales [23].

Renju and Brunda [24] discussed the use of ML using environmental factors to enhance accuracy over traditional methods in crop yield prediction. They identified that Naïve Bayes (NB) has the highest accuracy of 99.39% compared with other conventional methods. In the study cited in Phuoc et al. [25] and covered in this review, the "SIMPLE" model stands out as an adaptable, open-source substitute with few parameters that can be modified for other crop species. Even with its benefits, such as its extensibility and ease of use, the model still has some drawbacks.

3. Research Methodology

For crop recommendations, this section uses two datasets: the crop recommendation dataset and the crop soil dataset. Data are preprocessed using the IQR technique to remove outliers, ensuring reliable and accurate inputs for the model. To identify and select the most relevant features, the LDA approach is utilized for feature extraction. Several advanced ML models, such as LR, GNB, SVC, KNN, DT, ET, RF, and bagging classifier, are developed and assessed in this study. Figure 2 shows an overview of the proposed process.

3.1. Dataset

This research focuses on crop selection optimization using two significant datasets collected from open source in Kaggle: the crop soil dataset and the crop recommendation dataset. These datasets provide vital information on crop recommendations, soil properties, and environmental conditions, which are used to develop ML models to improve crop production estimates and resource efficiency. The dataset was split using an 80/20 ratio, where 80% was used for training (with stratified 10-fold cross-validation) and 20% kept for final testing. The split was stratified according to crop class to preserve class balance in both sets. This approach ensures fair evaluation and avoids class imbalance bias.

Crop recommendation dataset: Precision farming has recently gained popularity. It assists farmers in making well-informed decisions on their agricultural approach. Based on a variety of characteristics, the data is a dataset that enables users to build a prediction model that recommends the ideal crops to cultivate in a particular farm. This dataset is created based on information on India's climate, rainfall, and fertilizer. Table 1 below shows the data field attributes.

Crop_soil.csv: Data on different crop and soil types can be found in the soil.csv collection. The crop column includes a variety of crops, such as rice, maize, chickpeas, and kidney beans, and various fruits such as bananas, mangoes, and pomegranates. Other crops include cotton and coffee. In the soil column, loamy, clayey, sandy, black, and red soils are among the soil types listed appropriate for each crop. By pairing the right crops with the appropriate soil types, the information helps determine which crops grow well in particular soil types, allowing for precision farming to implement optimal crop selection.

3.2. Data preprocessing using IQR

The preprocessing method to identify and eliminate outliers in datasets in crop selection is the IQR method. The group between the first and third quartiles (Q1 and Q3, respectively), which represent the central 50% of the data, serves as the foundation. The *IQR* is computed in Equation (1) as follows:

$$IQR = Q3 - Q1 \tag{1}$$

Figure 2
Recommended framework

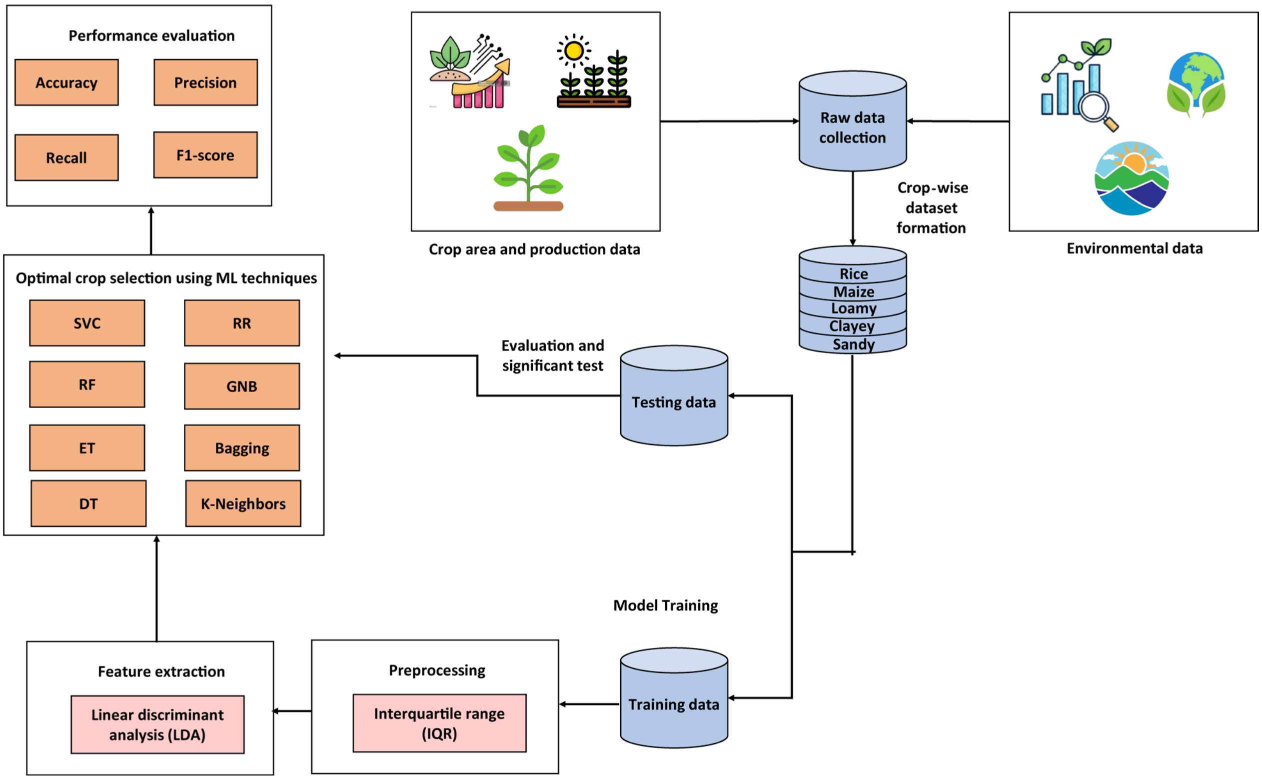


Table 1
Interpretation of the mean scale for belief, concern, and practice

Data Field Attributes	Description
N	Nitrogen content in soil ratio
P	Phosphorus content in soil ratio
K	The ratio of potassium in the soil
Temperature	degree Celsius
Rainfall	in mm
Humidity	in %
pH	soil value

Crop selection determines the lower and upper boundaries for outliers; see Equations (2) and (3).

$$Lower\ Bound = Q1 - 1.5 \times IQR \tag{2}$$

$$Upper\ Bound = Q3 + 1.5 \times IQR \tag{3}$$

Outliers are defined as data points that are either above or below the established lower bound and are typically eliminated from the dataset. Because the IQR method works regardless of the data distribution, it is reliable for both skewed and normal datasets. Outliers can significantly impact performance and are often used in data preprocessing, especially in exploratory data analysis as well as in preparing data for ML models. Figure 3 shows the outcomes of boxplots in the current classifier. The boxplots show the different environmental properties such as temperature, humidity, pH, rainfall, N, P, and K.

3.3. Extract feature using LDA

Raw data is transformed into numerical patterns through feature extraction approach, which is useful for data evaluation. LDA uses class label information to project data onto a lower-dimensional space, maximizing class separability. This not only reduces computational complexity but also significantly enhances classification accuracy by preserving discriminative features relevant to crop identification. It is a component of feature development, a procedure for better training predictive models for crop selection. In the process of feature regeneration and the continuous evaluation of performance of accounting data processing based on identification approaches and efficiency enhancements, the predictive system determines the data processing based on agricultural features such as crop recommendation and soil type. Equation (4) is used to find the linear function that defines the vector of motion to minimize the between-class dispersion structure and the within-class scattering structures in feature portions.

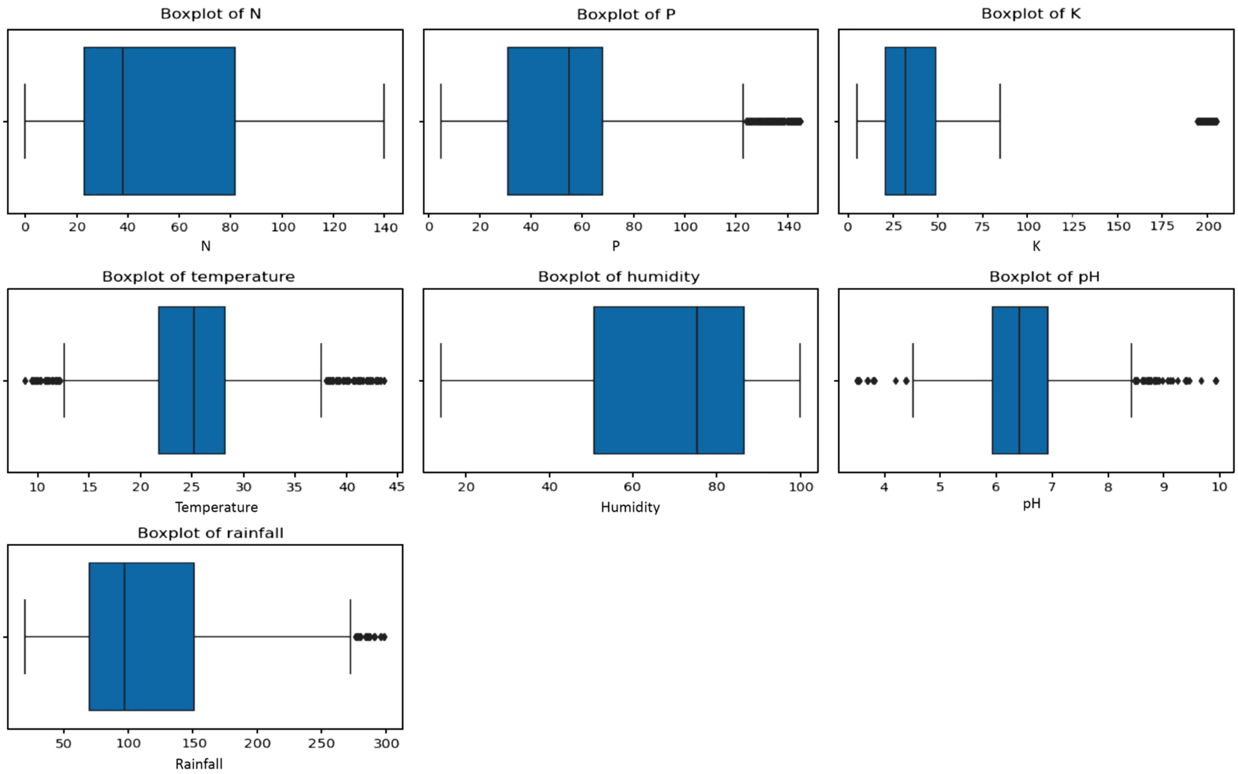
$$Z = A_1 W_{l_1} + A_2 + W_{l_2} + A_3 + W_{l_3} + + A_R + W_{l_R} \tag{4}$$

where variable $A^t = [\{A_1, A_2, \ldots, A_R\}]$ is the coefficient vector and $W_l = [W_{l_1}, W_{l_2}, \ldots, W_{l_R}]$ is the accounting data processing. Considering the linear expression $A_1, A_2, \ldots, A_R$, the LDA approach is used to determine the value coefficients. The weight of each initial feature is varied to create the updated feature space. The k^{th} finitary of the l^{th} collection is represented by each W_l.

3.4. Optimize crop selection for precision agricultural systems using ML techniques

This research attempts to create sophisticated ML models, such as LR, GNB, SVC, KNN, DT, ET, RF, and bagging classifier, to optimize

Figure 3
Outcomes for boxplots in the proposed method



crop selection in precision agriculture. By using these ML models, this research aims to improve crop selection decisions based on resource and environmental aspects, thereby ensuring increased agricultural sustainability and production. It was found that the best technique for raising the precision of crop selection forecasts was the RF classifier.

3.4.1. Logistic regression (LR)

The universal LR model in statistics is where the LR algorithm originates. When modeling the yield (upper and lower) of the agronomy-based cropping system, the logistics function is calculated in Equation (5) as follows:

$$P\left(Y = \frac{1}{X}\right) = \frac{1}{1+e^{-(\beta_0+\beta_1 X_1+\beta_2 X_2+\ldots+\beta_n X_n)}} \quad (5)$$

where $P(X) = P\ (Y = 1|X)$ is the formal expression for the potential of an input (X) (yield from the four distinct cropping systems) class ($Y = 1$ (Highlands)) that is modeled. Table 2 shows the important setup parameters for maximizing the effectiveness of LR. Figure 4 shows the categorization outcomes for each crop type using LR.

The confusion matrix demonstrates a good diagonal dominance, which means overall classification is correct. There are some slight misclassifications between crop classes under similar environmental conditions. This implies that LR performs well but may be weaker with overlapping feature spaces.

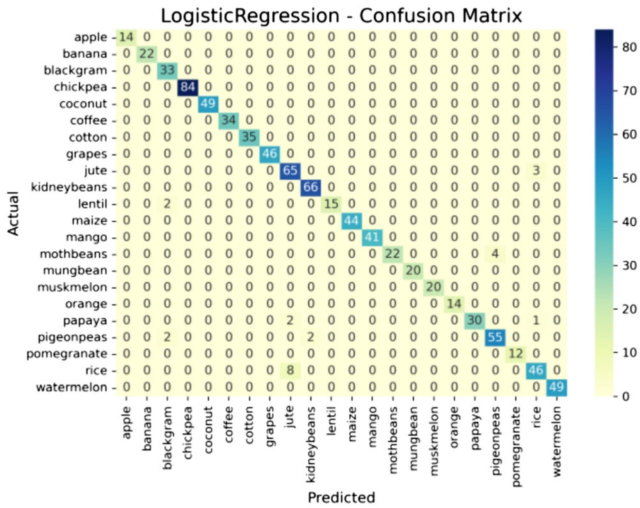
3.4.2. Gaussian Naive Bayes (GNB)

In crop selection, the probabilistic ML classification technique known as GNB assumes that each class follows a normal distribution. The premise is that each parameter can predict an agricultural outcome variable through smart farming. The foundation of GNB is the idea that an attribute or forecast exists independently of other attributes in the dataset, using Equation (6).

Table 2
Hyperparameter settings explored for LR in crop classification

Hyperparameter	Typical Values Used/Tested
Kernel	rbf, linear
Regularization C	0.1, 1.0, 10
Gamma	scale, auto
Degree (if poly kernel)	3
Probability estimates	True, False

Figure 4
Class-wise prediction performance of LR



$$O(B|A) = \frac{O(A|B)(O|B)}{O(A)} \tag{6}$$

The multinomial naïve approach was the most widely used technique for document classification problems. Meanwhile, Bernoulli Naïve often uses Boolean values in prediction, such as false and true, as well as 0 and 1. Moreover, the GNB used continuous value sets of information for prediction. The use of GNB offers benefits and drawbacks. For instance, compared to alternative techniques such as LR with less training data, the performance of GNB improves if the forecast is accurate. Although GNB is easy to set up, there is a good chance that it will lower the accuracy of the data. As a supervised ML method for classification, the NB can determine whether a particular feature is present in a class, regardless of whether other traits exist or not. Table 3 shows the typical hyperparameters for GNB applied in the classification task. Figure 5 shows the class-wise prediction breakdown using the GNB model.

This model assumes feature independence and misclassifies a number of crop types. Off-diagonal entries result from misclassification between crops with comparable nutritional profiles. It nonetheless captures the overall pattern of accurate classification.

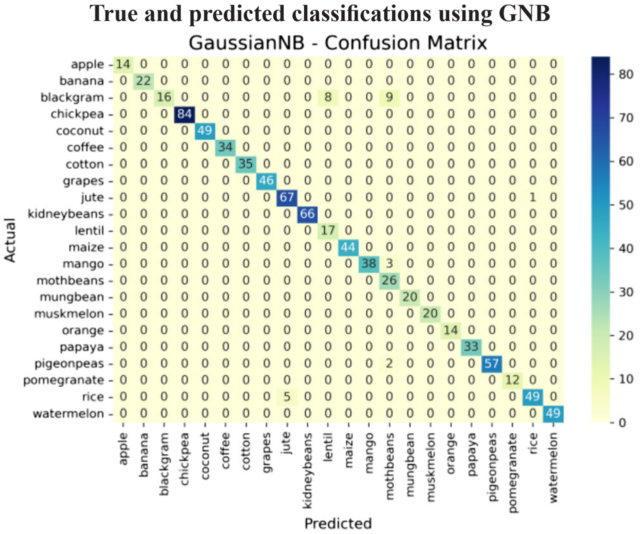
3.4.3. Support vector classifier (SVC)

The SVC is a straightforward technique for classifying crop and soil data that creates information in the test set. The model is built using data from training and testing, which are collections of data samples with multiple properties and one target value for crop selection. The identified labels indicate the dependability of the system output in smart formations by indicating the desired outcome, verifying the correctness of the system, or assisting the system in learning to operate appropriately. The goal of the SVC is to optimize the margin by determining the optimal separating hyperplane, as indicated in Equations (7–9).

Table 3
Parameters used for modeling with GNB classifier

Hyperparameter	Typical Values Used/Tested
Variance smoothing	1e-9 (default)
Prior probabilities	None
Fit prior	True, False
Distribution type	Gaussian
Assumes independence	Yes (default behavior)

Figure 5
True and predicted classifications using GNB



$$\min_{x,g,\div\xi} I\left(\vec{x}, \vec{\xi}\right) = \frac{1}{2x^s x} + D\sum_{m=1}^{M} \xi_m \tag{7}$$

$$z_m[x^s\varphi(w_m) + a] \geq 1 - \xi_{m\prime} \tag{8}$$

$$\xi_m \geq 0, m = 1, \ldots, M \tag{9}$$

where $\varphi(w_m), I\left(\vec{x}, \vec{\xi}\right)$, and g represent the transformer function, objective function, and decision function, respectively, and x^s, x, a, and ξ_m are a weight vector, input data vector, bias factor, and slack variable, respectively. Table 4 configures the hyperparameters of the SVC for multi-class crop prediction. Figure 6 shows the confusion matrix reflecting class-level accuracy of the SVC model.

The SVC has few misclassifications and distinct class boundaries. Most of the crops are accurately identified, as can be seen from the good diagonal alignment of the confusion matrix. A small number of closely related crop types exhibit minor faults.

3.4.4. K-nearest neighbors (KNN) classifier

The supervised classification technique KNN classifier can classify non-attributes by matching them with relevant attributes in the crop selection class. Because the likelihood of errors in a straightforward classification rule is constrained by the Bayes error rate in smart farming, the Bayes decision rule for minimum probability of error is superior for creating the most significant item in pattern recognition. This model consists of a pair of formulas that calculate the distance between each variable in the dataset. Additionally, it uses a different distance metric to normalize the distance. Table 5 shows the KNN configuration variables used during model training. In Figure 7, the performance of the KNN classifier is evaluated using the class-level confusion matrix.

Manhattan Distance. Compared with other equations, Equation (10) makes it simpler to compute the distance.

Table 4
Key SVC parameter settings tested for classification efficiency

Hyperparameter	Typical Values Used/Tested
Max depth	None, 10, 20
Min samples split	2, 5, 10
Min samples	1, 2, 5
Splitter	best, random

Figure 6
SVC model classification across crops

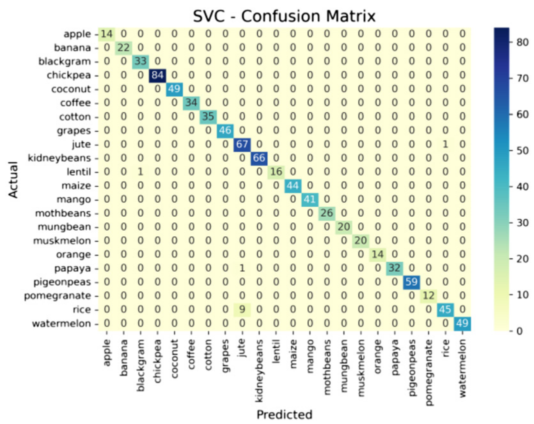
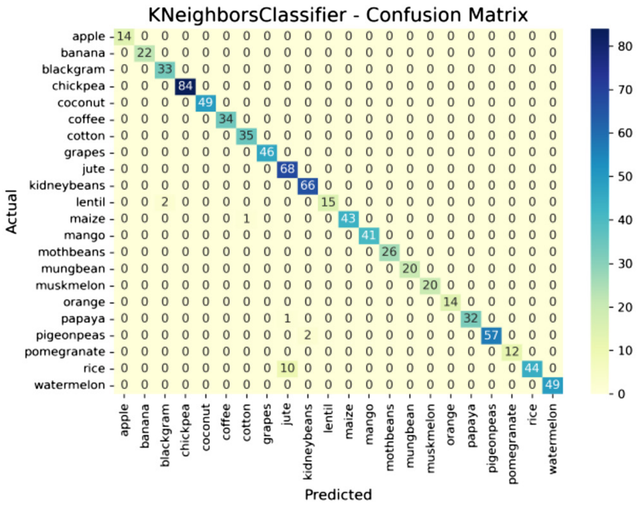


Table 5
Hyperparameters for KNN used to classify crop types based on proximity

Hyperparameter	Typical Values Used/Tested
Penalty	L2
Solver	liblinear, lbfgs
Regularization strength C	0.1, 1.0, 10
Max iterations	100, 500, 1000
Dual formulation	False, True

Figure 7
Classification output of the KNN algorithm



$$|y_2 - y_1| + |x_2 - x_1| \tag{10}$$

Euclidean distance in standard geometry is the distance between two points; see Equation (11).

$$\left((y_2 - y_1)^2 + (x_2 - x_1)^2\right)^{1/2} \tag{11}$$

Minkowski distance, as with the Euclidean distance, in this case, requires an n value $((y_2 - y_1)^p + (x_2 - x_1)^p)^{1/p}$, where x and y represent the x and y coordinates of the point on the y plane.

The KNN provides stable classification for strongly separated classes of crops but demonstrates some interference among close classes in the feature space. Its behavior is dependent on the distance metric and local data structure. Misclassifications occur mainly between crops that share overlapping environmental characteristics.

3.4.5. Decision tree (DT) classifier

In supervised ML, the DT is flowchart-like structure and is usually used for categorization and prediction. In general, DT comes in two forms, continuous and categorical, depending on the type of target feature. A perfect split can be achieved by comparing the root node of the DT with other properties or features in the dataset. If the tree is perfectly split, then outcomes of a certain class are on one side and for the opposite class are on the other side. This process splits each node until a perfect split is achieved, forming the leaf node of the tree. The choice of attributes is the true obstacle in building a DT. The Gini Index and Information Gain are two strategies that can be used to achieve splitting, as shown in Equation (12).

$$\begin{aligned} InformationGain(S, W) &= Entropy(S) - Entropy(S, W). \\ Gini\ index &= 1 - \sum(o) \wedge 2 = 1 - [(o+) \wedge 2 + (o-) \wedge 2] \end{aligned} \tag{12}$$

where S denotes the current state, W represents the selected attribute, and positive and negative outcomes and indicated $o+$ and $o-$.

Table 6 shows the important hyperparameters of the DT in crop classification. Figure 8 shows the detailed crop classification accuracy using the DT.

The confusion matrix shows almost perfect separation among classes, with high diagonal values and few off-diagonal errors. The model cleanly separates crop classes through hierarchical clustering of features, with minimal confusion in boundary cases.

3.4.6. Extra tree (ET) classifier

For challenges in crop selection classification, an ensemble ML algorithm, known as the ET classifier, makes use of the DT. While the ET classifier has important differences, it shares commonalities with RF. Unlike an RF, which builds individual trees using bootstrapped portions of the training data, the ET classifier builds each tree using the full training dataset. Although both methods incorporate randomization during the tree-building process, the ET classifier uses random thresholds at each split point. At each node, however, the RF takes into account a random selection of attributes, as shown in the following Equations (13) and (14).

$$Gini\ Impurity = \sum_{i=1}^{P} e_1(1e_i) \tag{13}$$

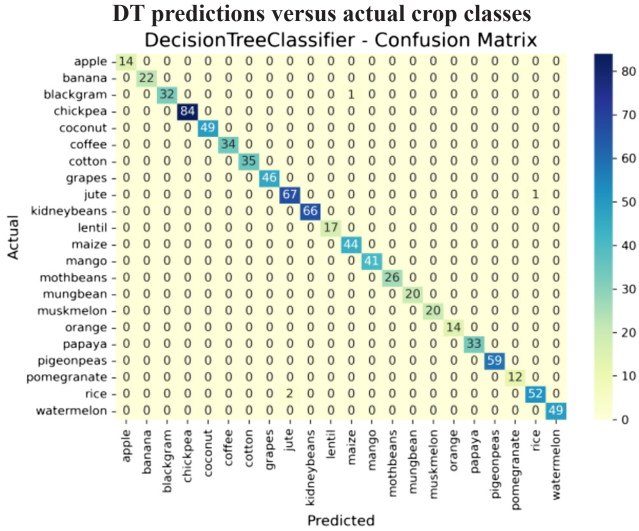
$$Entropy = \sum_{i=1}^{P} e_1 \log(e_i) \tag{14}$$

where $\log(e_i)$ is the number of distinct classes or labels and e_i represents the percentage of rows that have output label i; see Equations (6) and (7).

Table 6
DT settings tuned to improve prediction accuracy

Hyperparameter	Typical Values Used/Tested
Max depth	None, 10, 20
Minimum samples	2, 5, 10
Samples leaf	1, 2, 5
Splitter	best, random

Figure 8
DT predictions versus actual crop classes



Hyperparameter tuning was done to maximize the performance of the ET classifier model, with particular attention paid to the number of trees, minimum leaf, split sample sizes, splitting criteria, and bootstrap sampling. Table 7 and Figure 9 show the analysis of classification accuracy per class using the ET algorithm.

The ET can cause a minor instability because of the excessive randomness in the tree building process. The confusion matrix shows that there are some random misclassifications between a few classes. Although majority of predictions are correct, some crops are misclassified because of feature noise.

3.4.7. Random forest (RF) classifier

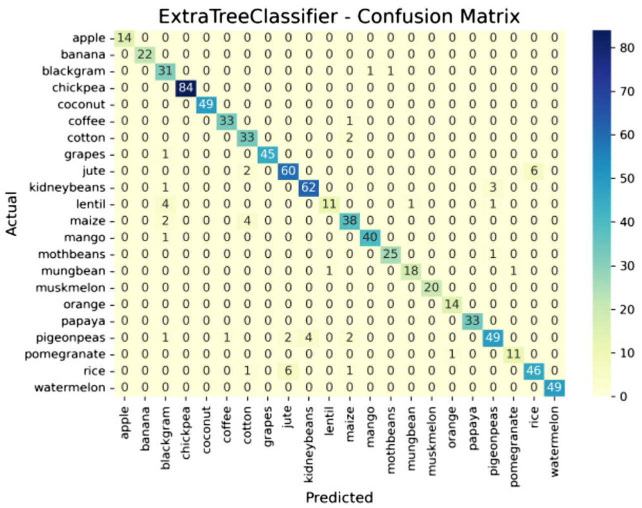
Using the RF framework, the ensemble method called RF was created for classification and regression. It predicts the average of multiple independent base models. Using multiple learning algorithms to improve the predictive performance in regression and classification is known as the ensemble method. RF uses bagging (Bootstrap Aggregation) as one of its ensemble methods. To lower the variance of a DT, bagging can be used. The RF model and the DT model have similar categorization structures. This method was implemented because a random subset of the training data is used to create each DT. The information, entropy, and data gain DTs were as expected.

$$\begin{aligned} Info_B(C) &= \sum_{i=1}^{u} \frac{|C_i|}{|C|} \times Info(C_i) \\ Info(C) &= J(o_j, m_j) = \frac{o_j}{o_j+m_j} log_2\left(\frac{o_j}{o_j+m_j}\right) \\ &\quad - \frac{m_j}{o_j+m_j} log_2\left(\frac{m_j}{o_j+m_j}\right) \end{aligned} \tag{15}$$

Table 7
Key hyperparameters tested in the ET model for crop classification

Hyperparameter	Typical Values Used/Tested
Amount of trees	100, 200
Max structures	sqrt, log2
Min models split	2, 5
Criterion	gini, entropy
Bootstrap	False (default), True

Figure 9
Decision graph of actual crop classifications versus tree projections



Entropy

$$F(B) = \frac{B1(o_j+m_j)}{o_j+m_j} J(B1(o_j+m_j)) + \cdots . + \frac{Bw(o_j+m_j)}{o_j+m_j} J(w(o_j+m_j)) \tag{16}$$

where o_j denotes a genuine and favorable value of the information, m_j is a fictitious and negative information value, $J(o_j + m_j)$ represents an amount found in the information, and $Info(C)$ is the predicted information value derived from the data o_j and m_j, as shown in Equations (15) and (16).

Information gain

$$Gain(B) = Info(C) - Info_B(C) \tag{17}$$

where $F(B)$ represents the entropy value that serves as the root of the DT and $B1(o_j + m_j)$ represents the integer of the information obtained from a characteristic or attribute, as shown in Equation (17).

he selection of data used for subgroups and training tasks is random. Autonomous DT is necessary for the randomization of the RF approach. The predicted information, entropy, and data gain equation, or the decision tree (CART), must first be created so to employ the RF technique. There are two types of random values in data: true and false. One metric used in RFs is the Gini Impurity. The Gini Impurity equations for the initial and subsequent layers of the RF model are shown in Equations (18) and (19).

$$J_H(m) = 1 - \sum_{j=1}^{I} (O_j)^2 \tag{18}$$

$$J_{second\ layer} = \frac{m_{left}}{m_{parent}} * J_{left\ node} + \frac{m_{right}}{m_{parent}} * J_{right\ node} \tag{19}$$

The ability of RF to handle missing information values while preserving the correctness of the missing data is an advantage. By combining random and multiple learning techniques, RF enhances the accuracy and performance of the DT method. Table 8 shows that the RF hyperparameters are tuned to achieve excellent recall and accuracy. Figure 10 shows the detailed confusion matrix, showcasing the robust classification performance of RF.

The excellent class-wise performance of RF is demonstrated by the strong concentration along the diagonal of the confusion matrix. There is very little misunderstanding in the precise distinction of all crop classifications. The ensemble method of the model guarantees reliable and consistent predictions for a range of feature inputs.

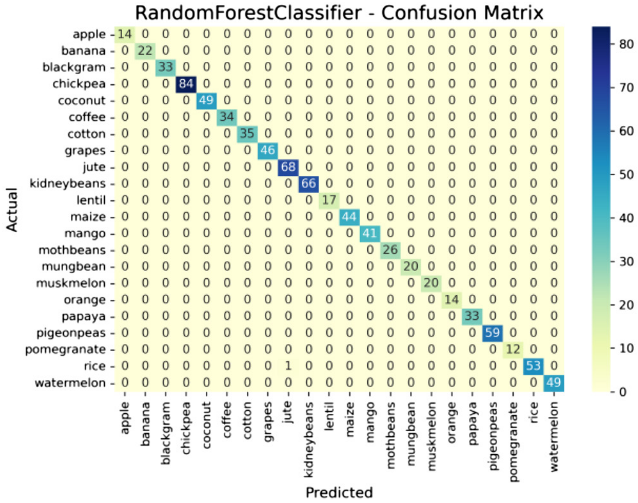
3.4.8. Bagging classifier

Bagging, sometimes referred to as bootstrap aggregation, is used with DT and greatly improves model stability by increasing accuracy and reducing variance to get rid of over-fitting. To choose the most accurate prediction for crop selection, bagging in ensemble ML aggregates multiple weak models, specializing in different regions of the feature space. A distinct model is practiced using each data group. The corresponding X is used to query each model, and all of

Table 8
Typical settings for the RF classifier in crop selection tasks

Hyperparameter	Typical Values Used/Tested
Number of estimators	100, 200, 300
Max features	sqrt, log2
Min samples	2, 5
Bootstrap	True

Figure 10
High-accuracy predictions by the RF model

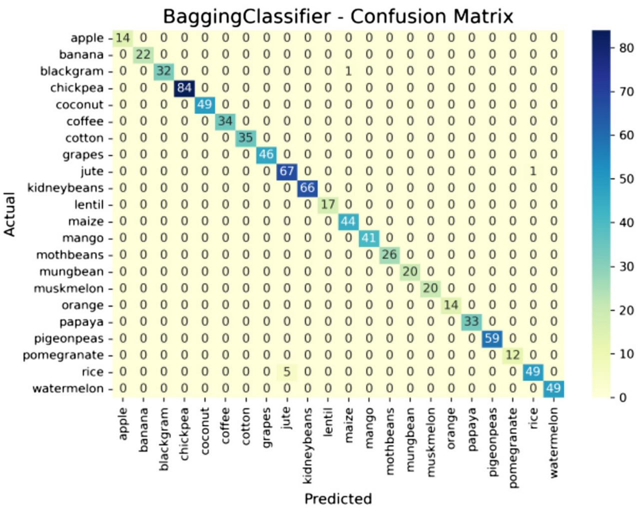


the results are gathered. To create the Y for the ensemble, the mean of the outputs from particular models is calculated. Assume K samples for bootstrap of dimension A. $\{Y_1^1, Y_2^1, \dots, Y_A^1\}$, $\{Y_1^2, Y_2^2, \dots, Y_3^2\}$, ..., $\{Y_1^K, Y_2^K, \dots, Y_A^K\}$. However, considering $Y_a^1 \equiv a$ based on the k-th bootstrapping sample, the virtually isolated K amount can be adapted $X_1(.), \dots, X_K(.)$. Table 9 displays detailed information about the hyperparameter adjustment of the DT-based bagging classifier. Figure 11 shows the results of the ensemble classification of the bagging model.

Table 9
Parameters applied in the bagging ensemble model for robust classification

Hyperparameter	Typical Values Used/Tested
Base estimator	DT
Number of estimators	50, 100
Max models	0.5, 1.0
Bootstrap	True, False
Bootstrap features	False, True

Figure 11
Outcomes visualized through the confusion matrix of the bagging model



The bagging classifier also exhibits strong diagonal dominance in the confusion matrix. Misclassifications are negligible and occur in few overlapping crop classes. The ensemble method enhances generalization and diminishes prediction variance.

4. Result and Discussion

This section explains the experimental setup, the performance of the existing and proposed methods, the comparative analysis, and, finally, the discussion.

4.1. Experimental setup

To improve the efficiency and accuracy of accounting data processing based on intelligent financial software requires a variety of hardware and software. The entire model training and testing procedure is developed using the Python ML framework, and Table 10 lists the experimental setting.

4.2. Comparative evaluation

Figure 12 shows a histogram for several crop indicators. Each histogram shows the distribution of the corresponding variable throughout the dataset. The probability density function, represented by the curves combined on the histograms, shows the general trend and distribution pattern of a variable.

Figure 12 shows the average values of several crop indicators. The average phosphorus (P) values of several crops showed a significant degree of variation. The highest P level is found in fruits such as apples and grapes, while the lowest is found in oranges and coconuts. Jute, mung bean, and rice are among the crops in the mid-range level. The following image, Figure 13, shows a preprocessed dataset with fewer outliers, indicating that the data preprocessing has improved. Figure 14 shows the outcomes of the boxplots in the proposed method

Correlation matrix: Figure 15 shows the correlation matrix of features in the proposed method. The relationship between numerical features is shown in this heatmap. Weaker correlations are shown in blue, and stronger connections are shown in darker red.

Figure 16 shows the outcomes for soil types and possible crops. This matrix shows the compatibility of various soil types with different crops. A “1” indicates the presence of a certain crop in a certain soil type. While some soil types, such as loamy and black soil, are suitable for growing a variety of crops, other soil types are less suitable. Figure 16 shows precision agricultural recommendations with the help of diverse soil-crop relationships.

The effectiveness of traditional ML methods is being compared with LR, GNB, SVC, KNN, DT classifier, ET classifier, RF classifier, and bagging classifier to optimize crop selection for precision agricultural system models. Figure 17 illustrates those that are commonly used in these described evaluations, such as accuracy, recall, precision, and F1-score.

Accuracy and precision: A popular statistic for the crop selection categorization task is accuracy. It assesses the proportion of precise

Table 10
Development tool

Environment	Frequency
GPU	NVIDIA Quadro RTX 3000
Operating system (OS)	Windows 11
Memory	16 GB
CPU	Intel Xeon Silver (4214R@2.40GHz)
Python version	3.8

Figure 12
Outcomes of the crop indicators (N, P, K, temperature, humidity, pH, and rainfall)

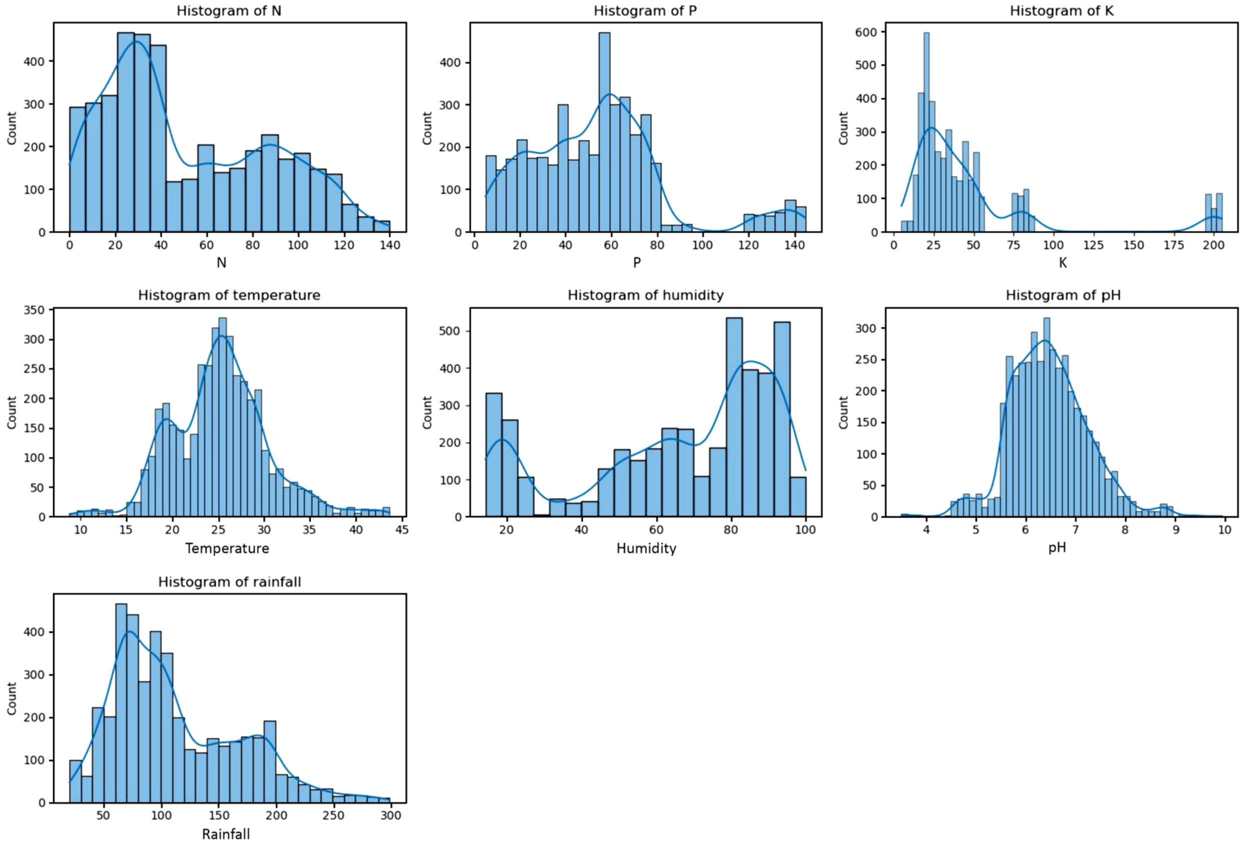


Figure 13
Comparison of the mea n values of P, K, temperature, humidity, pH, rainfall, and N for different crop plants

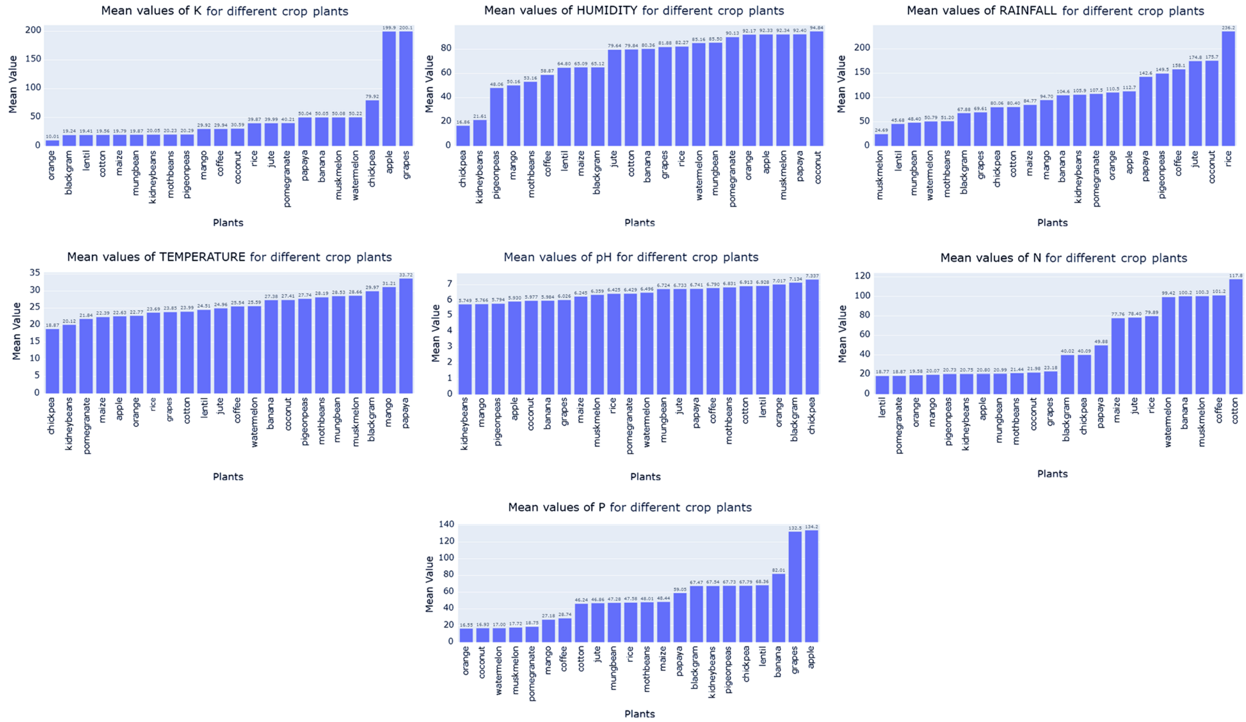


Figure 14
Outcomes for boxplots in the proposed method

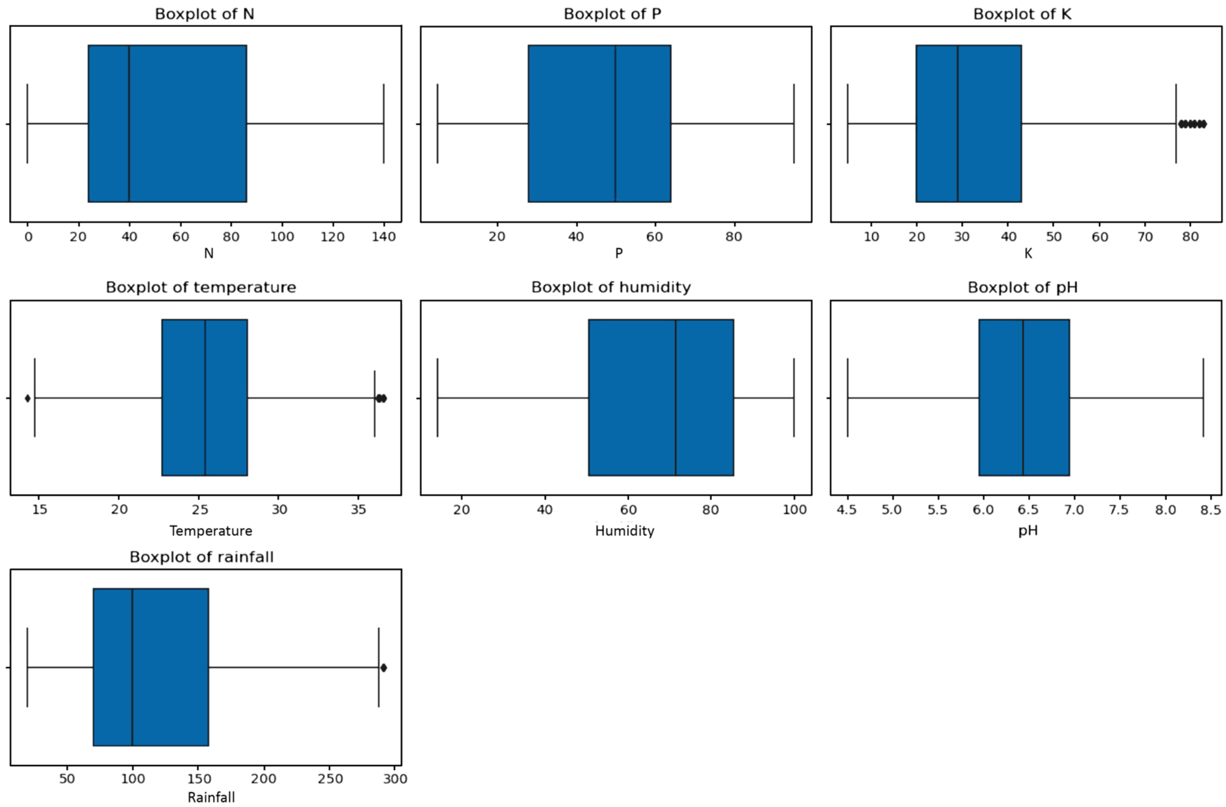


Figure 15
Correlation matrix of features in the proposed method

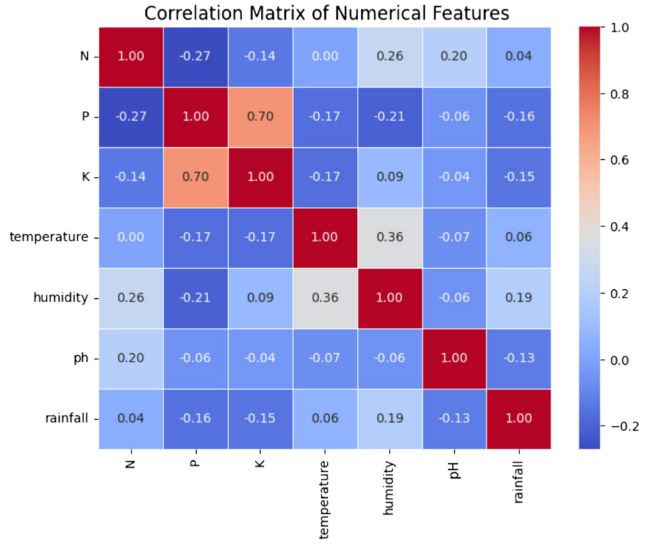
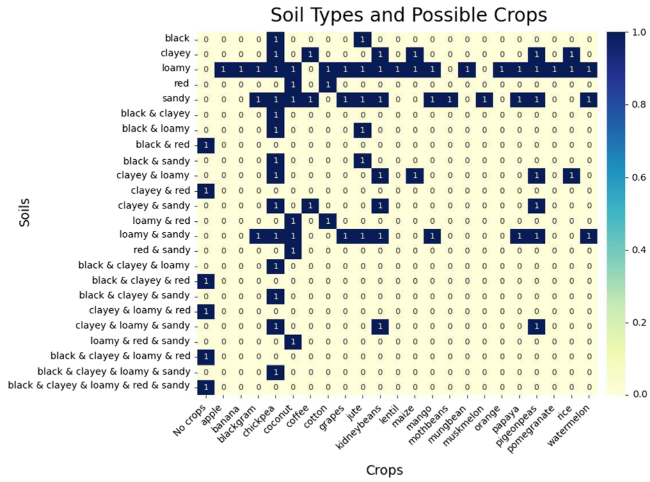


Figure 16
Outcomes of the soil types and possible crops



forecasts in each prediction category of the model. A performance indicator called precision determines the ratio of accurate positive predictions to every negative prediction made by the model. It assesses the accuracy of positive predictions by computing Equations (20) and (21), which determine the percentage of expected positive values. Table 11 shows the results of accuracy and precision of the classifier.

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \tag{20}$$

$$Precision = \frac{TP}{TP+FP} \tag{21}$$

Note: TN means true negative, FN means false negative, TP means true positive, and FP means false positive.

Figure 18 shows a comparison of the performance of different ML classifiers in terms of accuracy for crop selection. With the highest scores (both at 99.88%), RF had the best predictive power. Bagging and DT both performed well. GNB and ET showed somewhat lower precision, indicating somewhat less dependable predictions, but SVC and KNN had strong results.

Figure 17
Outcomes of the accuracy of the proposed method

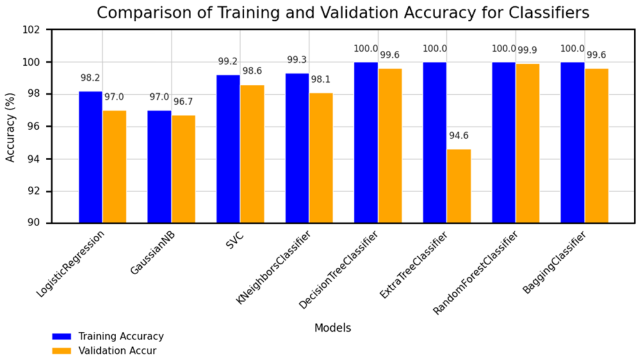


Figure 19
Precision of the different ML classifiers

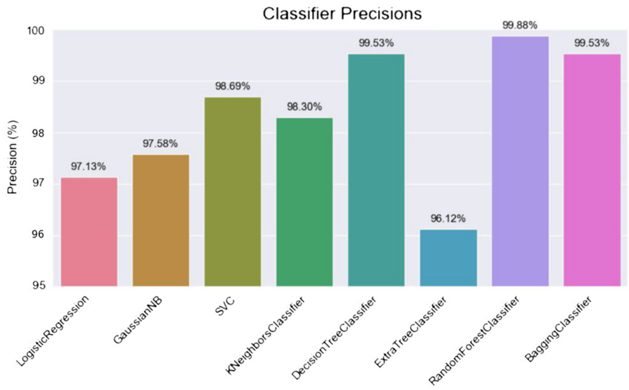


Table 11
Outcomes of accuracy and precision of classifiers

Methods	Accuracy (%)	Precision (%)
LR	97.02	97.13
GNB	96.67	97.58
SVC	98.57	98.69
KNN	98.10	98.30
DT	99.64	99.53
ET	97.50	96.12
RF	**99.88**	**99.88**
Bagging	99.64	99.53

Table 12
Numerical outcomes of the recall and F1-score of the classifiers

Methods	*Recall* (%)	*F1-score* (%)
LR	97.02	97.01
GNB	96.67	96.57
SVC	98.57	98.69
KNN	98.10	98.08
DT	99.64	99.52
ET	96.55	95.32
RF	**99.88**	**99.88**
Bagging	99.76	99.40

Figure 18
Accuracy performance of different ML classifiers

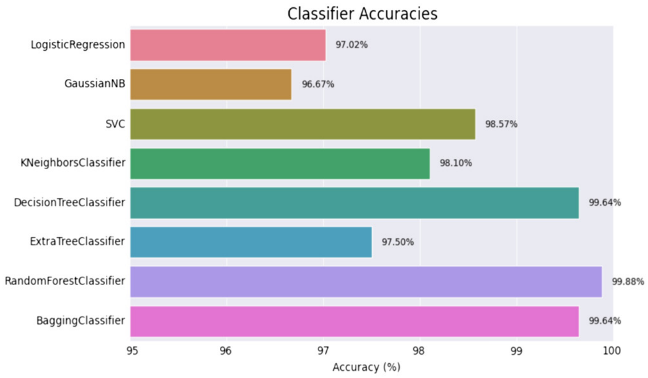


Figure 19 shows a comparison of the precision of each classifier. With 99.88% precision, the RF classifier outperforms other methods. These results show the exceptional performance of RF.

The bar charts show the performance of different classification algorithms. The RF classifier had the highest performance (99.88%), closely followed by the bagging classifier and DT Classifier with 99.64%. GNB had the worst performance at 96.67%. In general, ensemble models such as RF and bagging performed much better than basic models such as LR and NB, indicating their suitability for predictive classification tasks.

The above graph representation shows the precision levels of the classifiers. RF leads with 99.88%, closely followed by bagging (99.53%) and DT (99.53%). SVC (98.69%) and KNN (98.30%) exhibit good performance. LR (97.13%) and GNB (97.58%) perform well, although ET has the lowest precision level (96.12%), indicating low confidence in its positive predictions (Table 12).

Recall and F1-score: This metric measures the proportion of instances in a given positive class that were accurately categorized by crop selection prediction. Equations (22) and (23) demonstrate how the F1-score can be used to evaluate the predictive ability for crop selection by calculating the harmonic mean for accuracy and recall. The numerical results of the recall and F1-score of the classifiers are shown in Table 12.

$$Recall = \frac{TP}{TP+FN} \tag{22}$$

$$F1 - score = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{23}$$

All models perform well, as shown in the table, but RF has the greatest recall and F1-score (99.88%), showing exceptional accuracy in crop class identification. Although ET and GNB exhibit comparatively lower results, both DT and bagging show good performance. This demonstrates the reliability and success rate of RF for crop classification in precision farming. Figure 20 shows the performance evaluation of recall.

RF had the greatest F1-score and recall of 99.88% among the evaluated models. With respective recall rates of 99.76% and 99.40% and F1-scores of 99.52% (Figure 21). These best-performing models exhibit outstanding classification capabilities. Although SVC and KNN also have competitive scores, ET and GNB fall slightly behind, reflecting reduced resilience in coping with intricate crop classification.

10-fold cross-validation: This technique divides the data into ten equal sets (folds) and is a powerful model testing technique. Each run uses nine folds for training and one fold for testing. This process is

Figure 20
Performance evaluation of recall of the different ML classifiers

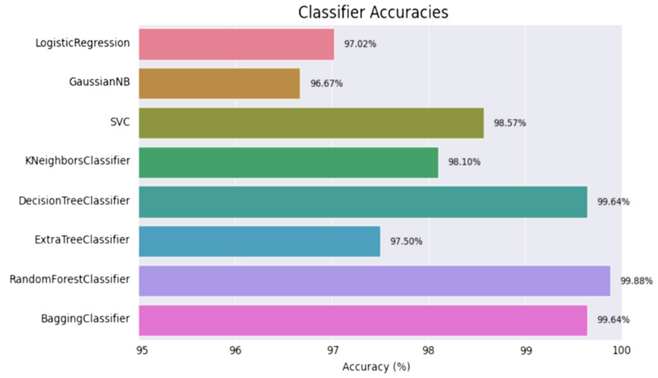
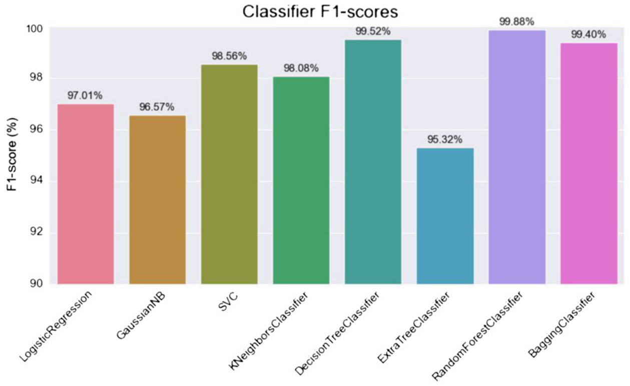


Figure 21
Outcomes of the F1-score



repeated ten times to ensure that each dataset is used once for testing and once for training. Table 13 provides a more accurate measure of model performance, where the results are averaged. This study confirms the consistency of the RF classifier across all assessment metrics.

The table shows the 10-fold cross-validation performance of the RF classifier, which consistently shows high performance in all folds. Accuracy, precision, recall, and F1-score values are all above 99.7% for each fold, indicating the robustness and stability of the model. The average value of all four measures is 99.88%, implying outstanding classification reliability and generalizability. The high consistency indicates that the model is suitable for reliable and efficient crop selection under precision agriculture operations.

Table 13
10-fold cross validation across the metrics

Fold	*Accuracy* (%)	*Precision* (%)	*Recall* (%)	*F1-Score* (%)
Fold 1	99.76	99.70	99.78	99.74
Fold 2	99.88	99.85	99.87	99.86
Fold 3	99.92	99.90	99.93	99.91
Fold 4	99.84	99.80	99.85	99.82
Fold 5	99.88	99.89	99.88	99.88
Fold 6	99.80	99.76	99.79	99.77
Fold 7	99.90	99.91	99.89	99.90
Fold 8	99.87	99.88	99.86	99.87
Fold 9	99.85	99.83	99.84	99.83
Fold 10	99.89	99.90	99.88	99.89
Average	99.88	99.88	99.88	99.88

4.3. Comparison phase

In the comparison phase, the performance of several ML models is evaluated to ascertain whether they are appropriate for crop prediction. The RF model consistently exhibits higher accuracy and dependability across assessment measures compared to all other methods.

Accuracy: The term measures the ratio of correct crop predictions by the model to the total predictions. The aim of the study is to maximize crop selection through ML optimization, with high accuracy showing consistent classification based on environmental and soil characteristics. It shows the efficiency of the model in precision agriculture decision-making. Table 14 shows the comparison of prediction accuracy and fit of the crop yield estimation model.

The table shows a comparison of the crop forecast accuracy of various ML classifiers. The suggested RF model achieves the highest accuracy (98.80%), indicating superior performance. Bagging, DT, and SVC perform well, while existing research techniques, such as KNN and LR, reflect relatively low accuracy, indicating achieved improvements by the proposed model.

Mean absolute error (MAE): This metric calculates the mean magnitude differences between model-predicted and actual crop yield values. It helps determine how well the model forecasts approximate actual crop conditions and aids in precise crop selection.

R^2: The term determines the extent to which the input variables predict crop yield variability. The greater the value of R^2, the better the predictive power of the model and the objective of data-based crop optimization. Table 15 shows that RF is the most accurate model compared to other models based on MAE and R^2 assessments.

The MAE and R^2 for various regression models are used in crop forecasting. The new RF model has the best performance with the lowest MAE and highest R^2, indicating high prediction accuracy and excellent model fit. Traditional models and the existing research report more errors, reflecting the superiority of the new approach in yield prediction .

4.4. Discussion

This research demonstrates how well-suited sophisticated ML models are for precision agricultural crop selection optimization.

Table 14
Evaluation of classifier accuracy across models used in crop selection

Methods	Accuracy (%)
KNN [23]	97.72
LR [23]	94.69
GNB	97.20
SVC	97.90
KNN	97.60
DT	97.95
ET	97.10
RF	**98.80**
Bagging	97.98

Table 15
Comparison of prediction error and fit by model for estimating crop yield

Methods	MAE	R^2
Linear Regression [23]	1.08	0.92
RF Regressor [23]	0.64	0.96
GNB	0.61	0.89
SVC	0.60	0.93
KNN	0.59	0.94
DT	0.58	0.95
ET	1.57	0.92
RF	0.52	0.97
Bagging	0.55	0.96

The research by Tufail et al. [20] is dedicated solely to predicting alfalfa yield based on data from Georgia and Kentucky, restricting the generalizability of the findings to other crops and locations. Similarly, the study by Shafi et al. [21] introduces a crop/weed detection system based entirely on ML tailored for tobacco crops, relying on visual characteristics such as texture, shape, and color. The research by Patil et al. [23] deals with one type of crop and is based on regional data, which will not work well in different climates and different soils, thus limiting its generalizability to other types of agricultural settings. Moreover, the lack of multi-crop testing and restricted feature variability diminish its suitability in wider crop selection applications. The current research comparison results clearly show that the RF model proposed here is better than any classifier with respect to accuracy, MAE, and R^2. It has the maximum classification accuracy and minimum prediction error, proving it is robust. It gives more accurate and consistent results for crop selection and yield prediction compared to current methods.

5. Conclusion

The research focuses on applying sophisticated ML models to optimize crop selection for precision agriculture. Because of concerns such as soil erosion, climate change, and water scarcity, crop selection has become essential to sustainable farming. The research creates an ML model to optimize crop selection by assessing several classifiers, such as LR, GNB, SVC, KNN, DT, ET, RF, and bagging. With preprocessing techniques such as IQR to eliminate outliers and LDA for feature extraction, the dataset includes crop suggestions, meteorological information, and soil properties. The RF classifier performed better than the others, achieving 99.88% accuracy, 99.88% precision, 99.88% recall, and 99.88% F1-score. In precision agriculture, the results show how contemporary ML models can improve crop selection, optimize resource utilization, and increase output. High computing costs and the time required to train efficient ML models, especially with huge datasets, are among the drawbacks of the research. The intricacy of the model can also make it difficult to implement in environments with limited resources. Future research should focus on utilizing AI-driven systems to investigate further sustainable agricultural options, enhance model efficiency, and integrate real-time data for dynamic crop suggestions.

Funding Support

This research project was funded by the Ministry of Higher Education, Research and Innovation through the Block Funding Program-2024, grant number MoHERI/BFP/UTAS/2024.

Ethical Statement

This study does not contain any studies with human or animal subjects performed by any of the authors.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

The Crop Recommendation datasets that support the findings of this study are openly available in Kaggle at https://www.kaggle.com/datasets/atharvaingle/crop-recommendation-dataset. The data that support the findings of this study are openly available in Kaggle at https://www.kaggle.com/datasets/shekharyada/crop-soilcsv.

Author Contribution Statement

Rajesh Natarajan: Conceptualization, Methodology, Validation, Writing – original draft, Writing – review & editing, Funding acquisition. **Sujatha Krishna:** Validation, Formal analysis, Writing – original draft, Writing – review & editing. **Pradeepa Ganesan:** Software, Formal analysis, Visualization. **Amna Al Kaabi:** Writing – review & editing, Supervision. **Rajesh Kala:** Resources, Data curation. **Anthony Mendoza Madlambayan:** Resources, Data curation, Visualization.

References

[1] Naqvi, R. Z., Farooq, M., Naqvi, S. A. A., Siddiqui, H. A., Amin, I., Asif, M., & Mansoor, S. (2021). Big data analytics and advanced technologies for sustainable agriculture. In C. M. Hussain & P. Di Sia (Eds.), *Handbook of smart materials, technologies, and devices: Applications of Industry 4.0* (pp. 1-27). Springer International Publishing. https://doi.org/10.1007/978-3-030-58675-1_82-1

[2] Sethuramalingam, S. & Perumal, S. A. (2022). Revolutionizing agriculture: Precision farming through IoT, AI, and machine learning for sustainable growth. *NeuroQuantology*, *20*(11), 9603-9615.

[3] Dutta, S., Rakshit, S., & Chatterjee, D. (2020). Use of artificial intelligence in Indian agriculture. *Food and Scientific Reports*, *1*(4), 65–72.

[4] Bakthavatchalam, K., Karthik, B., Thiruvengadam, V., Muthal, S., Jose, D., Kotecha, K., & Varadarajan, V. (2022). IoT framework for measurement and precision agriculture: Predicting the crop using machine learning algorithms. *Technologies*, *10*(1), 13. https://doi.org/10.3390/technologies10010013

[5] Durai, S. K. S. & Shamili, M. D. (2022). Smart farming using machine learning and deep learning techniques. *Decision Analytics Journal*, *3*, 100041. https://doi.org/10.1016/j.dajour.2022.100041

[6] Hasan, M., Marjan, M. A., Uddin, M. P., Afjal, M. I., Kardy, S., Ma, S., & Nam, Y. (2023). Ensemble machine learning-based recommendation system for effective prediction of suitable agricultural crop cultivation. *Frontiers in Plant Science*, *14*, 1234555. https://doi.org/10.3389/fpls.2023.1234555

[7] Sharma, P., Dadheech, P., Aneja, N., & Aneja, S. (2023). Predicting agriculture yields based on machine learning using regression and deep learning. *IEEE Access*, *11*, 111255–111264. https://doi.org/10.1109/ACCESS.2023.3321861

[8] Raja, S. P., Sawicka, B., Stamenkovic, Z., & Mariammal, G. (2022). Crop prediction based on characteristics of the agricultural environment using various feature selection techniques and classifiers. *IEEE Access*, *10*, 23625–23641. https://doi.org/10.1109/ACCESS.2022.3154350

[9] Swaminathan, B., Palani, S., & Vairavasundaram, S. (2022). Meta learning-based dynamic ensemble model for crop selection. *Applied Artificial Intelligence*, *36*(1), 2145646. https://doi.org/10.1080/08839514.2022.2145646

[10] Cravero, A. & Sepúlveda, S. (2021). Use and adaptations of machine learning in big data—Applications in real cases in agriculture. *Electronics*, *10*(5), 552. https://doi.org/10.3390/electronics10050552

[11] Lad, A. M., Bharathi, K. M., Saravanan, B. A., & Karthik, R. (2022). Factors affecting agriculture and estimation of crop yield using supervised learning algorithms. *Materials Today: Proceedings*, *62*, 4629–4634. https://doi.org/10.1016/j.matpr.2022.03.080

[12] Choudhary, M., Sartandel, R., Arun, A., & Ladge, L. (2023). Crop recommendation system and plant disease classification using machine learning for precision agriculture. In S. Hiranwal & G. Mathur (Eds.), *Artificial intelligence and communication technologies* (pp. 39–49). Soft Computing Research Society. https://doi.org/10.52458/978-81-955020-5-9-4

[13] Kailasam, S., Achanta, S. D. M., Rama Koteswara Rao, P., Vatambeti, R., & Kayam, S. (2022). An IoT-based agriculture maintenance using pervasive computing with machine learning technique. *International Journal of Intelligent Computing and Cybernetics*, *15*(2), 184–197. https://doi.org/10.1108/IJICC-06-2021-0101

[14] Abdel-salam, M., Kumar, N., & Mahajan, S. (2024). A proposed framework for crop yield prediction using hybrid feature selection approach and optimized machine learning. *Neural Computing and Applications*, *36*(33), 20723–20750. https://doi.org/10.1007/s00521-024-10226-x

[15] El-Kenawy, E.-S. M., Alhussan, A. A., Khodadadi, N., Mirjalili, S., & Eid, M. M. (2025). Predicting potato crop yield with machine learning and deep learning for sustainable agriculture. *Potato Research*, *68*(1), 759–792. https://doi.org/10.1007/s11540-024-09753-w

[16] Rani, S., Mishra, A. K., Kataria, A., Mallik, S., & Qin, H. (2023). Machine learning-based optimal crop selection system in smart agriculture. *Scientific Reports*, *13*(1), 15997. https://doi.org/10.1038/s41598-023-42356-y

[17] Apat, S. K., Mishra, J., Raju, K. S., & Padhy, N. (2023). An artificial intelligence-based crop recommendation system using machine learning. *Journal of Scientific & Industrial Research*, *82*(05), 558–567. https://doi.org/10.56042/jsir.v82i05.1092

[18] Suruliandi, A., Mariammal, G., & Raja, S. P. (2021). Crop prediction based on soil and environmental characteristics using feature selection techniques. *Mathematical and Computer Modelling of Dynamical Systems*, *27*(1), 117–140. https://doi.org/10.1080/13873954.2021.1882505

[19] Whitmire, C. D., Vance, J. M., Rasheed, H. K., Missaoui, A., Rasheed, K. M., & Maier, F. W. (2021). Using machine learning and feature selection for alfalfa yield prediction. *AI*, *2*(1), 71–88. https://doi.org/10.3390/ai2010006

[20] Tufail, M., Iqbal, J., Tiwana, M. I., Alam, M. S., Khan, Z. A., & Khan, M. T. (2021). Identification of tobacco crop based on machine learning for a precision agricultural sprayer. *IEEE Access*, *9*, 23814–23825. https://doi.org/10.1109/ACCESS.2021.3056577

[21] Shafi, U., Mumtaz, R., Iqbal, N., Zaidi, S. M. H., Zaidi, S. A. R., Hussain, I., & Mahmood, Z. (2020). A multi-modal approach for crop health mapping using low altitude remote sensing, internet of things (IoT) and machine learning. *IEEE Access*, *8*, 112708–112724. https://doi.org/10.1109/ACCESS.2020.3002948

[22] Son, N.-T., Chen, C.-F., Chen, C.-R., Guo, H.-Y., Cheng, Y.-S., Chen, S.-L., ..., & Chen, S.-H. (2020). Machine learning approaches for rice crop yield predictions using time-series satellite data in Taiwan. *International Journal of Remote Sensing*, *41*(20), 7868–7888. https://doi.org/10.1080/01431161.2020.1766148

[23] Patil, P., Athavale, P., Bothara, M., Tambolkar, S., & More, A. (2023). Crop selection and yield prediction using machine learning approach. *Current Agriculture Research Journal*, *11*(3), 968–980.

[24] Renju, K., & Brunda, V. (2024). Optimizing crop yield prediction through multiple models: An ensemble stacking approach. *International Journal of Data Informatics and Intelligent Computing*, *3*(2), 52–58. https://doi.org/10.59461/ijdiic.v3i2.120

[25] Phuoc, L. H., Suliansyah, I., Arlius, F., Chaniago, I., Xuan, N. T. T., & van Quang, P. (2023). Literature review crop modeling and introduction a simple crop model. *Journal of Applied Agricultural Science and Technology*, *7*(3), 197–216. https://doi.org/10.55043/jaast.v7i3.123

How to Cite: Natarajan, R., Krishna, S., Ganesan, P., Al Kaabi, A., Kala, R., & Madlambayan, A. M. (2025). Revolutionizing Agriculture Through Smart Farming by Employing Advanced Machine Learning Techniques for Optimal Crop Selection. *Journal of Computational and Cognitive Engineering*. https://doi.org/10.47852/bonviewJCCE52026049