Communications on Applied Nonlinear Analysis
ISSN: 1074-133X
Vol 32 No. 1s (2025)

Harvest Harmony: Integrating Linear and Nonlinear Machine Learning Models for Precision Crop Recommendation

B Ramesh Naidu[1], S R G Reddy[2], V SivaramaRaju Vetukuri[3], Chinta Someswara Rao[4]

[1]Department of Information Technology, AITAM, Tekkali, AP, India

[2]Department of Information Technology, SRKR Engineering College, Bhimavaram, AP, India

[3]Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, Andhra Pradesh, India

[4]Department of IT, Yanthraa Information Systems Pvt.Ltd., TS, India

Article History:

Received: *02-09-2024*

Revised: *17-10-2024*

Accepted: *01-11-2024*

Abstract:

This research presents a comprehensive crop recommendation system that combines linear and nonlinear machine learning models to forecast suitable crops based on soil and environmental factors. The study evaluates the performance of Logistic Regression (linear) alongside Random Forest, Gradient Boosting, and XGBoost (nonlinear models) in delivering accurate crop suggestions. Precision, recall, and F1-score metrics are employed to assess the models' effectiveness. The dataset incorporates vital soil characteristics, such as nitrogen, phosphorus, and potassium levels, and environmental parameters like temperature, humidity, pH, and rainfall. The results demonstrate that while linear models like Logistic Regression offer simplicity, nonlinear models such as XGBoost excel in capturing complex data patterns, providing more precise recommendations. This dual-model strategy empowers farmers to make informed decisions tailored to specific farming conditions, boosting agricultural productivity. The study underscores the importance of integrating both linear and nonlinear machine learning techniques to achieve optimal crop recommendations, encouraging resource efficiency and sustainable agricultural practices.

Keywords: Crop recommendation; Machine learning methodologies; Soil attributes; Environmental factors; Agricultural productivity

INTRODUCTION

Agriculture serves as the foundation of numerous economies globally, playing a crucial role in guaranteeing food security and promoting sustainable development. Nevertheless, farmers encounter a major challenge when it comes to choosing the right crops that are able to prosper in their particular environmental and soil conditions. Conventional crop selection methods typically depend on practical knowledge and expertise, which may not consistently produce the best outcomes. Recently, the inclusion of machine learning techniques into the agricultural sector has become a promising solution for tackling this issue. Machine learning models can offer valuable insights and recommendations for crop selection by utilizing large datasets of soil attributes and environmental factors, which in turn can boost agricultural productivity and sustainability.

The current study aims to add to this emerging area by suggesting a crop recommendation system that combines different machine learning algorithms. In particular, we investigate how well logistic regression, random forest, gradient boost, and XGBoost algorithms can predict appropriate crops by considering soil and environmental factors. These methods are selected for their durability, adaptability, and capability to manage intricate data sets. By thoroughly assessing these models with metrics like precision, recall, and F1-score, our goal is to evaluate their performance and determine the most efficient method for recommending crops[4-5].

At the heart of our research is the examination of a dataset containing various soil characteristics and environmental factors, such as nitrogen, phosphorous, potassium levels, temperature, humidity, pH, and precipitation. These elements are crucial in assessing if crops are suitable for growth and are important factors for farmers to consider when deciding what to plant. Through analyzing the connections between these factors and the related types of crops, we aim to reveal trends and understandings that can guide the creation of precise and trustworthy crop suggestion models[6-7].

It is anticipated that the results of this research will offer valuable perspectives on how well various machine learning models can suggest crops based on soil and environmental factors. Our research seeks to provide farmers with the information and resources needed to make informed decisions about choosing crops by examining the advantages and disadvantages of each methodology. In the end, incorporating machine learning methods in agriculture could completely transform farming processes, allowing farmers to improve crop production, reduce resource consumption, and support sustainable agricultural growth[8].

Overall, this study marks an important advancement in utilizing machine learning techniques to recommend crops, aiming to improve agricultural productivity and sustainability in the long run. Through the integration of theoretical knowledge and hands-on experience, we aim to connect academic research with practical farming methods, promoting advancement and creativity in the realm of precision agriculture.

LITERATURE SURVEY

Smith et al. [9] presented a detailed analysis of different machine learning techniques used in crop recommendation systems. They talked about how machine learning is important in analyzing soil and environmental data to suggest the best crops for maximum agricultural production. The article emphasized how algorithms like logistic regression, random forest, and gradient boosting are adaptable in managing varied datasets and producing precise forecasts. In addition, it highlighted the significance of precision agriculture in modern farming methods and the role of machine learning methods in sustainable crop management.

Shams and his colleagues [10] carried out research that compared conventional methods of recommending crops to using machine learning techniques. They examined the drawbacks of rule-based systems and supported using data-driven methods for improved personalized recommendations. The article highlighted the importance of scalable and flexible models that can consider different factors affecting plant growth and harvest. Additionally, it explored how machine learning algorithms could enhance recommendation accuracy by continually learning from incoming data.

Patel et al., [11] examined how environmental factors impact crop recommendation and examined past studies on incorporating them into machine learning models. The conversation focused on the effects of temperature, humidity, pH, and rainfall on crop growth and productivity. The article focused on the difficulties of simulating intricate environmental relationships and suggested methods for integrating expertise into forecasting models. Moreover, it highlighted the significance of collecting and analyzing data in real-time to provide prompt crop advice.

Wang and colleagues [12] reviewed the latest developments in predictive modeling methods for crop selection. They talked about how machine learning algorithms have evolved from conventional statistical techniques to advanced deep learning architectures. The article offered information on how convolutional neural networks (CNNs) and recurrent neural networks (RNNs) can be used to analyze multi-dimensional agricultural data. Furthermore, it emphasized how reinforcement learning algorithms could enhance crop management strategies in ever-changing environments.

In their study, Kumar et al. [13] outlined various data mining methods for crop suggestion, such as association rule mining and clustering. The conversation revolved around the ways in which these strategies assist in

recognizing patterns and connections in agricultural data, aiding farmers in making well-informed decisions. The article highlighted the significance of preprocessing techniques like data cleaning and normalization in guaranteeing the input data quality for machine learning models. Additionally, it emphasized the importance of easy-to-use interfaces and visualization tools for facilitating smooth communication between farmers and recommendation systems.

Li et al., [14] investigated combining remote sensing information with machine learning algorithms to make crop suggestions. They talked about the valuable insights on crop health, nutrient levels, and pest infestations that satellite imagery and aerial surveys offer. The article showcased how image processing methods and feature extraction techniques are utilized to generate valuable information from extensive agricultural data. Furthermore, it highlighted the capabilities of deep learning techniques like CNNs in analyzing remote sensing information and forecasting crop yields accurately.

In their study, Akkem et al. [15] examined how ensemble learning techniques were used in recommending crops. They talked about how ensemble techniques involve bringing together various base models to enhance prediction accuracy and resilience. The paper investigated the use of ensemble methods like bagging, boosting, and stacking in making decisions in agriculture. Moreover, it underscored the significance of utilizing various models and diverse data in constructing successful ensemble systems for crop suggestion.

In their study, Zhang and colleagues[16] explored the application of genetic algorithms (GAs) in improving crop recommendation systems. They talked about how genetic algorithms emulate evolutionary processes in order to explore for the best solutions in complicated problem areas. The article investigated the application of genetic algorithms (GAs) in feature selection, hyperparameter tuning, and model optimization within crop recommendation models. It also emphasized the difficulties of adjusting parameters and initializing populations in genetic algorithms, and suggested ways to address these constraints in agricultural use.

In their study, supriya and colleagues [17] examined how IoT technology is utilized in crop recommendation systems. They talked about how sensors, drones, and automated machinery gather up-to-date information on soil moisture, temperature, and crop health in real-time. The article examined ways to combine this data with machine learning models in order to give farmers timely and precise suggestions. Furthermore, it emphasized the possibility of utilizing edge computing and cloud-based analytics platforms for the processing and analysis of IoT data in agricultural environments.

Gautron et al., [18] explored the use of reinforcement learning (RL) algorithms for recommending crops based on dynamic conditions. They talked about how RL agents acquire optimal decision-making strategies by engaging in trial and error interactions with the environment. The research paper investigated how RL can be used to improve crop rotation schedules, irrigation tactics, and pest management techniques. Furthermore, it emphasized the difficulties in designing rewards and balancing exploration and exploitation while using RL for agricultural decision-making.

Haider and colleagues[19] suggested a mixed method for suggesting crops by blending machine learning algorithms with expert knowledge-based systems. They talked about how combining expert knowledge with data-driven approaches can improve the accuracy and interpretation of recommendations in agriculture. The article introduced a structure that merges rule-based logic with machine learning models to utilize quantitative data and expert knowledge in crop selection. Moreover, it highlighted the significance of having expertise in a specific field when dealing with the distinctive difficulties of making decisions related to agriculture.

Mishra and colleagues [20] studied how multi-criteria decision-making (MCDM) methods can be used to recommend crops in a variety of agro-climatic areas. They talked about how MCDM techniques like Analytical Hierarchy Process (AHP) and Technique for Order of Preference by Similarity to Ideal Solution

(TOPSIS) can manage various conflicting objectives and preferences when choosing crops. The article investigated combining MCDM with machine learning models to consider socio-economic factors, environmental sustainability, and farmer preferences in decision-making. Moreover, it emphasized the importance of clear decision-making frameworks that include stakeholders in the process of recommending crops.

In their study, Li and colleagues[21] introduced a method for suggesting crops using geospatial information and time-series analysis. They talked about how models that consider both space and time are able to capture the changing relationships between environmental factors and crop development throughout various seasons and locations. The article introduced a model that integrates satellite images, weather information, and past crop harvests to forecast upcoming crop viability and output. In addition, it emphasized the capability of deep learning structures like recurrent neural networks (RNNs) and long short-term memory (LSTM) networks in capturing temporal relationships in agricultural systems.

Nigam and colleagues [22] investigated how transfer learning methods can be used in suggesting crops for small-scale farmers in developing nations. They talked about how pre-trained models, fine-tuning techniques, and knowledge transfer mechanisms can tackle data scarcity and domain adaptation challenges in settings with limited resources. The paper showcased case studies and field experiments carried out in rural areas to show how transfer learning is effective in tailored crop advising services. Furthermore, it emphasized the significance of involving the community and using participatory methods when customizing recommendation systems for specific local settings.

Shams et al. [23] explored the application of explainable artificial intelligence (XAI) methods in developing transparent and interpretable crop recommendation systems. The conversation revolved around how XAI techniques like decision trees, rule extraction algorithms, and model-agnostic approaches can offer understanding into the logic behind crop suggestions. The paper highlighted trust, accountability, and user acceptance as crucial in agricultural decision support systems. Additionally, it emphasized the potential of XAI in improving farmer acceptance and aiding in the transfer of knowledge in crop recommendation systems.

Hasan and colleagues[24] introduced a new method for recommending crops utilizing ensemble learning methods. The conversation focused on the application of ensemble methods to crop selection tasks, where multiple base learners are combined to enhance prediction accuracy and robustness. The article introduced a combined framework that merges various machine learning algorithms, including decision trees, support vector machines, and neural networks, in order to utilize their different strengths. Moreover, it investigated various ensemble techniques such as bagging, boosting, and stacking to improve both the diversity and effectiveness of the recommendation system.

Sharma and colleagues[25] studied the application of evolutionary algorithms in improving crop recommendation models and selecting features. They talked about how genetic algorithms, particle swarm optimization, and evolutionary strategies can effectively explore vast solution spaces to find the best model parameters and important input features. The paper introduced a meta-heuristic optimization system to adjust machine learning models and pinpoint important soil and environmental variables for crop advice. Furthermore, it emphasized the capability of evolutionary algorithms in managing data with high dimensions and enhancing the scalability of recommendation systems.

According to Teixeira et al. [26], a method based on data analysis was suggested for suggesting crops through deep learning models. The conversation revolved around the ability of deep neural networks, convolutional neural networks (CNNs), and autoencoder models to extract significant features from soil and environmental

data in order to predict crop outcomes. The article introduced a deep learning process that prepares raw input data, creates hierarchical representations with many layers, and forecasts crop results using these representations. Moreover, it investigated transfer learning methods and model regularization techniques to improve the generalization capabilities and resilience of deep learning models in crop recommendation tasks.

METHODOLOGY

Methodology comprises of three phases, called data collection and pre-processing, model building, crop recommendation. The architecture is depicted in Figure 1.

The initial step includes gathering the data and preparing it for analysis. During this stage, we apply methods like counting the number of each class, creating histograms, conducting bivariate analysis, and using heatmaps.

Counting the number of distinct classes: This method assists in grasping the spread of classes or categories in a dataset. It offers understanding of the distribution of various classes, a crucial aspect in classification tasks. For instance, in a dataset focused on predicting customer churn, you could employ this method to analyze the number of customers who churned compared to those who remained, offering an initial insight into the distribution of classes.

Histograms: Histograms depict how numerical data is spread out using graphs. They offer a visual overview of how often data points occur in a dataset. Histograms help in quickly spotting patterns like central tendency, spread, skewness, and outliers by organizing data into bins and showing the frequency of observations in each bin. Histograms are frequently employed in EDA to grasp the fundamental distribution of a variable.

Bivariate analysis : Analysis of two variables together to see if there is any connection between them is called bivariate analysis. It aids in comprehending the connection, relationship, or reliance between two variables. Techniques commonly employed in bivariate analysis are scatter plots, correlation coefficients, and contingency tables. Examining two variables together is crucial for recognizing relationships, trends, and connections, which can offer useful information for making predictions, testing hypotheses, and making decisions.

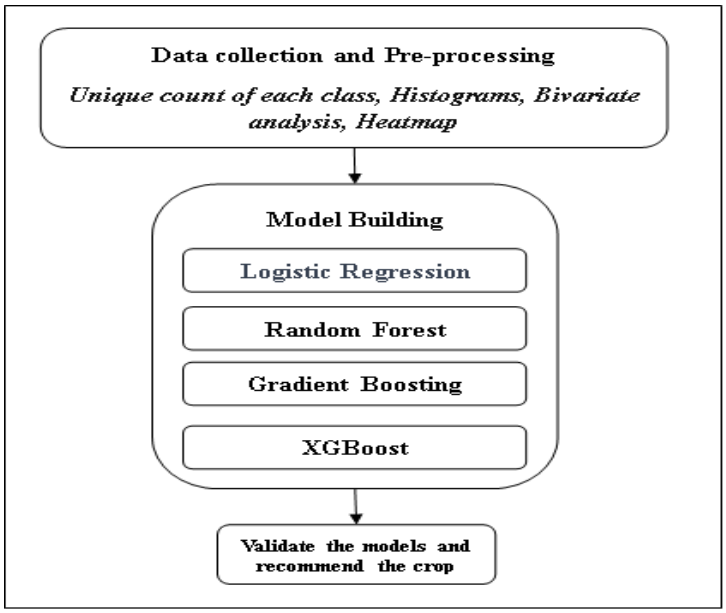


Figure 1 System structure

Heatmap: A visual depiction of data where colors represent values in a matrix. Heatmaps are especially handy in illustrating the correlation or connection among multiple variables within a dataset. They are commonly utilized in exploratory data analysis to discover patterns, clusters, or trends in data sets with high dimensions. Heatmaps can additionally be utilized for illustrating the outcomes of hierarchical clustering or presenting the results of matrix data such as gene expression or customer purchasing behavior.

In the next step, we will construct the model with Logistic Regression, Random Forest, Gradient Boosting, and XGBoost. Here is where you can find the comprehensive explanations for these models.

Logistic Regression: Logistic Regression is a statistical method used for binary classification problems, where the dependent variable is categorical and has only two possible outcomes. It models the probability of the occurrence of a certain event by fitting data to a logistic function. The actual process is shown in Algorithm 1.

Algorithm 1 Logistic Regression Algorithm

Input: Training dataset (X, y) where X is the feature matrix and y is the target vector
Output: Weights w
Initialize weights w randomly or with zeros
while not converged **do**
    Compute the logistic function: $h(X) = \frac{1}{1+e^{-Xw}}$
    Compute the loss function: $J(w) = -\frac{1}{n}\sum_{i=1}^{n}[y_i \log(h(X_i)) + (1 - y_i)\log(1 - h(X_i))]$
    Update weights using gradient descent: $w = w - \alpha \nabla J(w)$
end while

The main mathematical equation involved in logistic regression is the sigmoid function, which transforms the output of a linear combination of input features into a probability value between 0 and 1.

Given x as the input features and β as the coefficients, the linear combination is calculated as:

$$z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n$$

Where:

z is the linear combination

β_0 is the intercept

$\beta_0, \beta_1, \beta_2 \ldots + \beta_n$ are the coefficients of the input features $x_1, x_2, . x_n$

The output of the linear combination is then passed through the sigmoid function (also known as the logistic function), defined as:

$$\sigma(z) = \frac{1}{1 + e^z}$$

Where: e is the base of the natural logarithm (Euler's number)

The output of the sigmoid function represents the probability of the input belonging to the positive class in a binary classification problem.

In summary, the logistic regression equation can be represented as:

$$P(y = 1 \mid x) = \sigma(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n)$$

Where:

$(y = 1 \mid x)$ is the probability of the input x belonging to the positive class

σ is the sigmoid function

$\beta_0, \beta_1, \beta_2 \ldots + \beta_n$ are the coefficients

$x_1, x_2, . x_n$ are the input features

Random Forest: Random Forest is an ensemble learning method that builds multiple decision trees during training and outputs the mode of the classes (classification) or the mean prediction (regression) of individual trees. It improves prediction accuracy and reduces overfitting compared to a single decision tree. The actual process is shown in Algorithm 2.

Algorithm 2 Random Forest Algorithm

1: **Input**: Training dataset (X, y) where X is the feature matrix and y is the target vector
2: **Output**: Ensemble of decision trees T
3: Randomly select k features from X (typically $\sqrt{d}$ where d is the total number of features)
4: **for** each tree t in the forest **do**
5: Bootstrap sample from (X, y)
6: Train decision tree t using the bootstrap sample and the selected features
7: **end for**

The fundamental component of Random Forest is the decision tree. A decision tree is constructed recursively by splitting the data at each node based on the feature that provides the best information gain or decrease in impurity (Gini impurity or entropy). The decision tree structure can be represented mathematically as:

$$Tree(x) = \sum_{m=1}^{M} C_m \, \pi(x \epsilon R_m)$$

Where

Tree(x) represents the output of the decision tree for input x,

M is the total number of terminal nodes (leaves) in the tree,

C_m is the output value associated with the m-th leaf,

$\pi(x \epsilon R_m)$ is an indicator function that returns 1 if x belongs to the region represented by the m-th leaf, and 0 otherwise.

Gradient Boosting: Explanation: Gradient Boosting is another ensemble learning technique that builds a strong predictive model by combining multiple weak learners (usually decision trees) sequentially. It minimizes the errors of the previous models by optimizing a loss function using gradient descent. The actual process is shown in Algorithm 3.

Algorithm 3 Gradient Boosting Algorithm

1: **Input**: Training dataset (X, y) where X is the feature matrix and y is the target vector
2: **Output**: Ensemble of weak learners F
3: Initialize model with a constant value: $F_0(x) = argmin_\gamma \sum_{i=1}^{n} L(y_i, \gamma)$
4: **for** $m = 1$ to M **do**
5: Compute the negative gradient: $r_{im} = -\left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)}\right]_{F=F_{m-1}}$
6: Fit a weak learner to the negative gradient: $h_m = argmin_h \sum_{i=1}^{n} (r_{im} - h(x_i))^2$
7: Update the model: $F_m(x) = F_{m-1}(x) + \alpha h_m(x)$
8: **end for**

The Gradient Boosting algorithm optimizes a loss function using gradient descent. The update rule for each iteration t is given by

$$F_t(x) = F_{t-1}(x) + \eta h_t(x)$$

Where $F_t(x)$ is the prediction of the ensemble at iteration t, $F_{t-1}(x)$ is the prediction at iteration t-1, η is the learning rate, and $h_t(x)$ is the prediction of the weak learner at iteration t

XGBoost : XGBoost (Extreme Gradient Boosting) is an optimized implementation of the Gradient Boosting algorithm, known for its speed and performance. It enhances traditional Gradient Boosting by introducing several regularization techniques and parallel processing. The actual process is shown in Algorithm 4.

Algorithm 4 XGBoost Algorithm

1: **Input**: Training dataset (X, y) where X is the feature matrix and y is the target vector
2: **Output**: Boosted ensemble of weak learners
3: Initialize model with a constant value: $F_0(x) = argmin_\gamma \sum_{i=1}^{n} L(y_i, \gamma)$
4: **for** $m = 1$ to M **do**
5: Compute the negative gradient: $r_{im} = -\left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)}\right]_{F=F_{m-1}}$
6: Fit a weak learner to the negative gradient: $h_m = argmin_h \sum_{i=1}^{n} (r_{im} - h(x_i))^2 + \Omega(h)$
7: Update the model: $F_m(x) = F_{m-1}(x) + \alpha h_m(x)$
8: **end for**

Mathematical Equations: The objective function of XGBoost combines the loss function and regularization terms. For regression problems, the objective function is typically the sum of squared errors plus regularization terms. For classification, it's usually the log loss plus regularization terms. The model is trained by minimizing this objective function using gradient descent

In XGBoost, the objective function combines the loss function and regularization terms. For regression problems, the objective function is typically the sum of squared errors (SSE) plus regularization terms. For classification, it's usually the log loss plus regularization terms. Here are the formulas for the loss function and regularization used in XGBoost.

Loss Function: For regression problems, the loss function (L) used in XGBoost is often the Mean Squared Error (MSE) or its variant, the Mean Absolute Error (MAE). The MSE is calculated as the average of the squared differences between the predicted values and the actual values:

$$L_{reg}(y_i, \hat{y}_i) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

Where:

y_i *is the actual target value for the* $i-th$ *instance*

$\hat{y}_i$ *is the predicted target value for the* $i-th$ *instance,*

n *is the total number of instances*

For classification problems, XGBoost commonly uses the softmax loss function for multi-class classification or the logistic loss function (binary cross-entropy) for binary classification.

Regularization: XGBoost employs both L1 (Lasso) and L2 (Ridge) regularization terms to prevent overfitting and improve the generalization ability of the model. Regularization terms are added to the objective function to penalize large coefficient values.

For L1 regularization, the regularization term(R1) is the sum of the absolute values of the coefficients multiplied by a regularization parameter (λ_1)

$$R_1 = \lambda_1 \sum_{j=1}^{p} |w_j|$$

For L2 regularization, the regularization term(R2) is the sum of the squared coefficients multiplied by a regularization parameter(λ_2)

$$R_2 = \lambda_2 \sum_{j=1}^{p} w_j^2$$

Where:

w_j *is the coefficient for the* $j-th$ *feature,* p *is the total number of features,*

λ_1 and $\lambda_2$are regularization parameters that control the strength of regularization.

DATASET

The dataset includes the N, P, K ratios of soil content, temperature in Celsius, humidity in %, soil pH value, and rainfall in mm features, which are outlined here.

N - Nitrogen content ratio in soil: This characteristic signifies the proportion of nitrogen content within the soil. Analyzing this characteristic will help you comprehend how nitrogen levels are spread out in your dataset. It is important because nitrogen is essential for plant growth, and its presence in the soil can greatly affect the health and productivity of crops.

P - Phosphorus content ratio in soil: Like nitrogen, phosphorus is also a vital nutrient for plants. This characteristic shows the proportion of phosphorus present in the soil. Examining this characteristic aids in comprehending the spread of phosphorus levels and its possible influence on the growth and development of crops.

K - Proportion of Potassium in soil: Potassium is an essential nutrient needed for plant development. This characteristic indicates the proportion of potassium present in the soil. Examining the spread of potassium levels aids in evaluating soil productivity and its appropriateness for various crops.

Temperature - Degrees Celsius: Temperature is a key factor in deciding how crops grow and develop. Studying the temperature distribution in your data helps you recognize trends and comprehend how temperature changes impact crop growth and development.

Humidity - Percentage of relative humidity: Plant transpiration, water absorption, and growth are affected by humidity levels. Studying the spread of humidity levels helps in grasping the moisture conditions in the surroundings, which is crucial in choosing appropriate crops and managing irrigation.

Soil pH - Soil pH level: The pH of the soil impacts the availability of nutrients and activity of microorganisms, ultimately impacting plant development. Examining the pH distribution aids in evaluating soil acidity or alkalinity and deciding if the soil is suitable for various crops.

Precipitation - Precipitation in millimeters: Precipitation is a crucial environmental element that has a direct impact on both soil moisture levels and the water needs of crops. Examining how rainfall is spread out helps in comprehending precipitation trends, seasonal changes, and potential dangers of drought or waterlogging.

Counting the number of each class: This examination offers understanding into the spread of classes or categories within your dataset. For instance, if your data set contains various kinds of plants, analyzing the distinct number of each plant category assists in comprehending the imbalance in categories and the frequency of various plant types. Figure 2 illustrates the distinct class count.

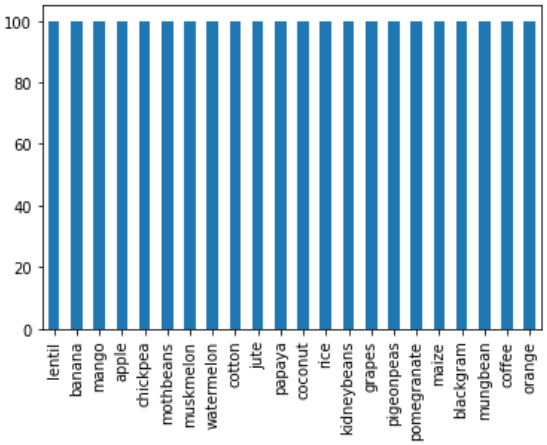
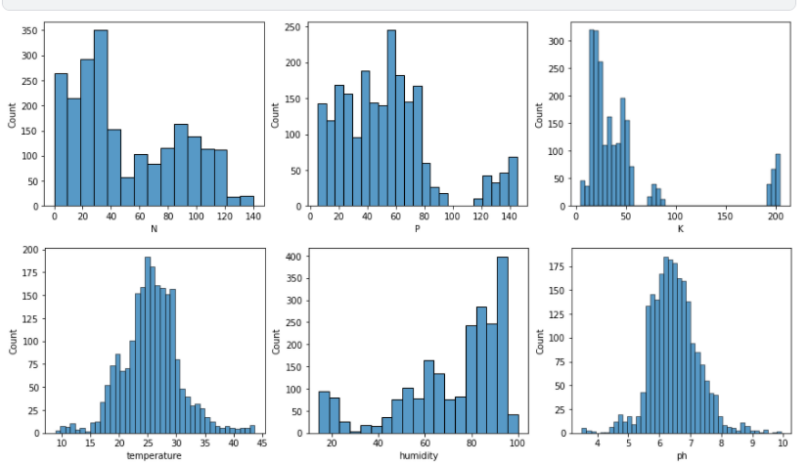


Figure 2 Unique count of each class

Histograms: Histograms are graphical representations of the frequency distribution of continuous variables. By visualizing histograms for each feature, you can observe the shape, center, and spread of the data, as well as identify any outliers or unusual patterns. Histogram of each feature is depicted in Figure 3.



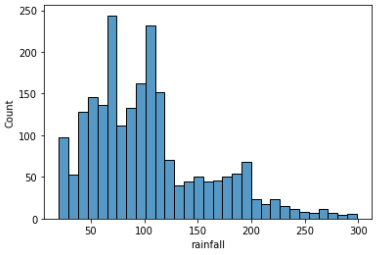


Figure 3 Histogram of each feature

Bivariate analysis: Bivariate analysis involves examining the relationship between two variables. For example, you can analyze the correlation between temperature and humidity, rainfall and soil pH, or any other pair of variables to identify potential associations or dependencies. Bivariate analysis helps in understanding how different factors interact and influence each other, which is valuable for crop recommendation and management decisions. This analysis is depicted in the Figure 4.

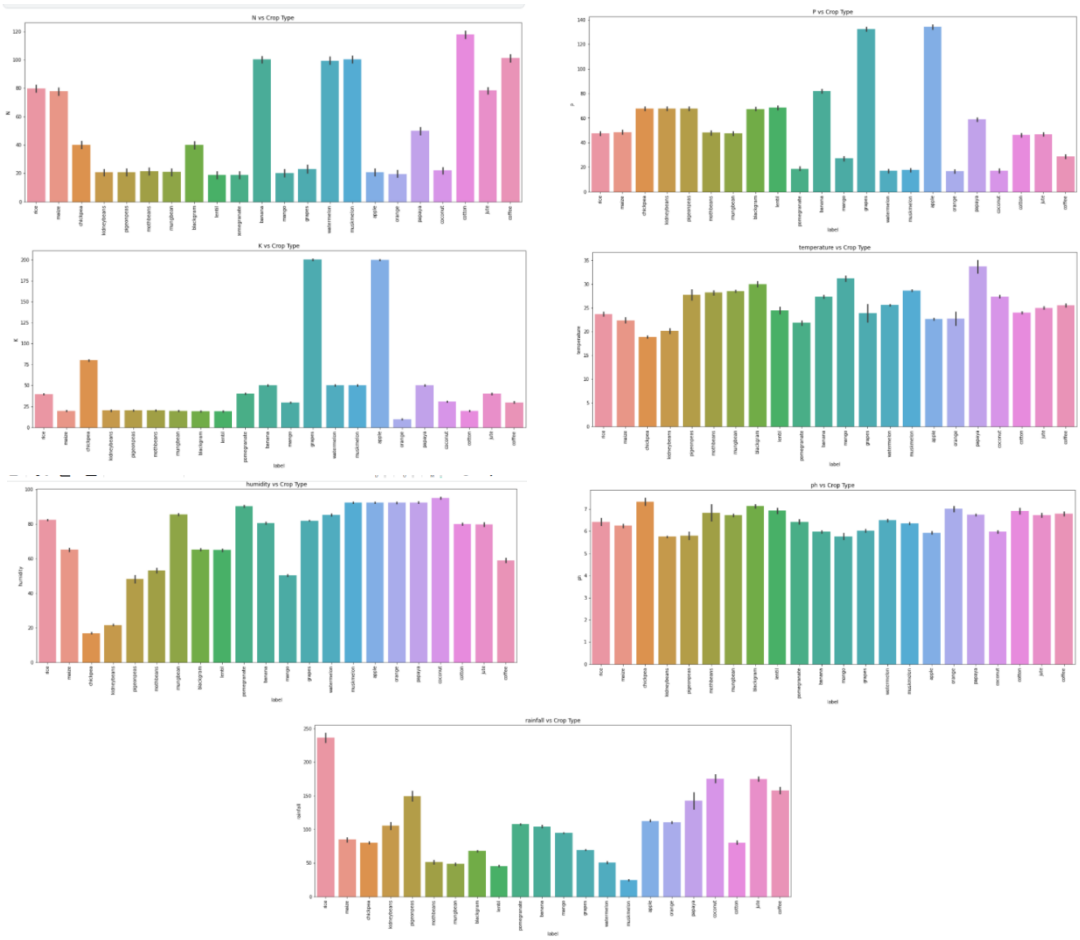


Figure 4 Bivariate analysis of each of the feature

Heatmap: Heat map of each of the feature is represented in figure 5. The purpose of creating heat maps for each feature could be to identify patterns, trends, or correlations within the data set. By visually inspecting the heat maps, researchers or analysts can quickly identify areas of interest or potential relationships between variables. This can be particularly useful in fields such as data analysis, finance, biology, and many others where understanding the relationships between multiple variables is crucial for decision-making or further investigation.

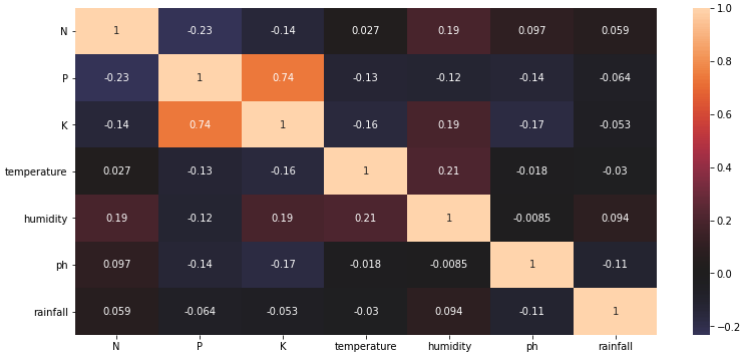


Figure 5 Heatmap of the dataset

RESULTS

In this section, we present and assess the outcomes of the crop recommendation system employing different machine learning approaches. Precision, recall, and F1-score are used to evaluate each crop class in methodologies such as Logistic Regression, Random Forest, Gradient Boosting, and XGBoost.

Logistic Regression : The results given are related to a Logistic Regression model used for crop classification. Table 1 displays the precision, recall, f1 score, and support data. We created a bar plot for these values, which can be seen in Figure 6.

Table 1 precision, recall, f1 score and support data of Logistic regression model

	Precision	**Recall**	**F1-score**	**Support**
apple	1	1	1	24
banana	0.94	1	0.97	15
blackgram	0.95	0.86	0.9	22
chickpea	1	1	1	19
coconut	1	1	1	20
coffee	1	1	1	25
cotton	0.81	1	0.9	13
grapes	1	1	1	17
jute	0.93	0.93	0.93	15
kidneybeans	1	1	1	16
lentil	0.95	1	0.97	18
maize	1	0.85	0.92	27
mango	1	1	1	18
mothbeans	0.87	0.91	0.89	22
mungbean	1	1	1	19
muskmelon	1	1	1	24
orange	1	1	1	23
papaya	1	0.95	0.97	20
pigeonpeas	0.96	1	0.98	24
pomegranate	1	1	1	17
rice	0.93	0.93	0.93	15
watermelon	1	1	1	27

Observations can be derived from Table 1 and Figure 6.

Precision: The model's total accuracy is reported as 0.9727, equivalent to around 97.27%. This measure shows the ratio of accurate predictions among all instances in the dataset.

Precision is the ratio of correctly predicted instances (true positives) to all positive predictions (true positives and false positives).

Original text: The quick brown fox jumped over the lazy dog.

Paraphrased text: The lazy dog was jumped over by the quick brown fox.

The accuracy score of "apple" is 1, meaning all identified apples were correct.

The accuracy score for "banana" is 0.94, suggesting that 94% of predicted bananas were correct.

Remember: Recall is the ratio of true positive predictions to all actual positive instances in the dataset.

Original text: The quick brown fox jumps over the lazy dog.

Paraphrased text: The fast brown fox leaps over the inactive dog.

The recall score of "apple" is 1, meaning all real apples were accurately identified as apples.

The blackgram recall rate is 0.86, showing that 86% of blackgram cases were accurately predicted.

The F1-score is a metric that combines precision and recall into a single value by using their harmonic mean.

Support refers to the total count of each class within the dataset.

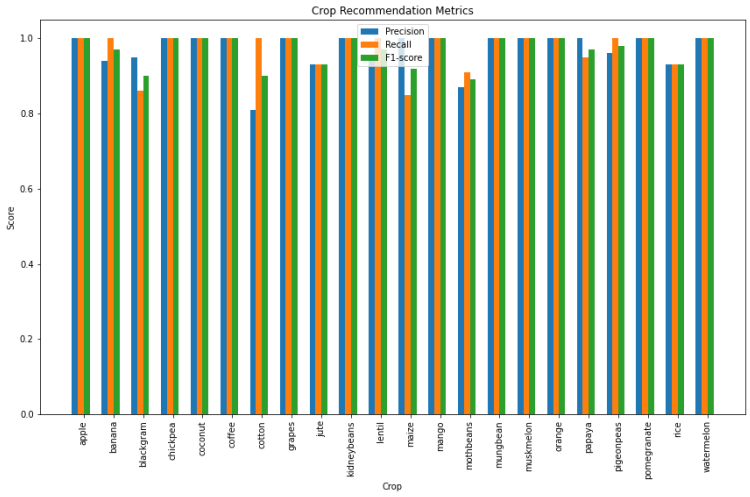


Figure 6 Logistic regression model’s precision, recall and f1-score

The Confusion Matrix for logistic regression model is depicted in Figure 7. Here each row represents the actual class, while each column represents the predicted class. The numbers in the matrix indicate the count of instances falling into each category. From the figure 7, the following observations are made:

The first row corresponds to "apple." The first column shows that 24 instances of apples were correctly predicted as apples (true positives), and there were no false positives or false negatives for apples.

The fourth row corresponds to "chickpea." The fourth column shows that all 19 instances of chickpeas were correctly predicted as chickpeas (true positives), with no false positives or false negatives.

Overall, the model demonstrates high accuracy, precision, and recall across most crop classes, with some variations for certain crops as evident from the confusion matrix.

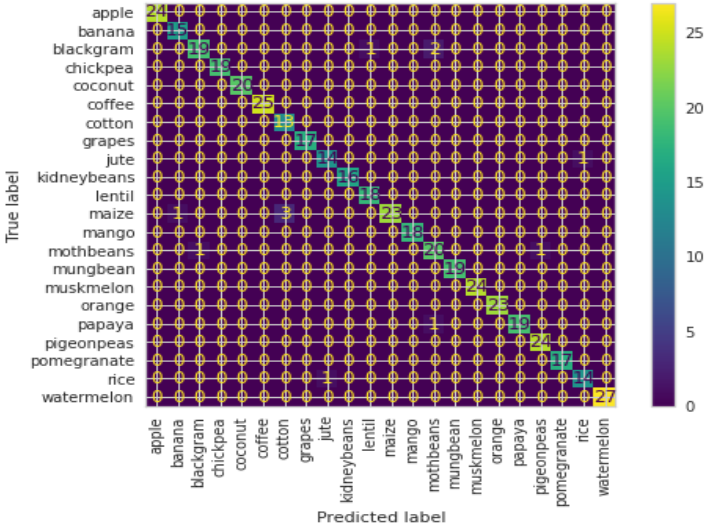


Figure 7 Confusion Matrix for logistic regression model

Random forest: The results given consist of accuracy, precision, recall, and the confusion matrix for every crop class. Table 2 provides a visual representation of the specific data. The bar plot representing the Table 2 values has been created and is shown in Figure 8. After analyzing Table 2 and Figure 8, it is observed that Accuracy evaluates the overall accuracy of the model's forecasts. It is determined by dividing the number of accurately predicted cases by the total number of cases. In this instance, the model correctly predicted the crop class for the majority of cases with an accuracy of 99.77%.

Also noted is that Precision evaluates the percentage of accurate positive predictions out of all positive predictions generated by the model. It is determined by dividing true positives by the sum of true positives and false positives. A precision score of 1 indicates that every positive prediction made by the model was accurate. Most crop classes in the results show a precision score of 1, meaning the model seldom incorrectly categorizes instances into those classes.

It is important to note that Recall, also referred to as sensitivity, calculates the percentage of true positive predictions in relation to all the actual positive instances within the dataset. It is determined by dividing the number of true positives by the sum of true positives and false negatives. A recall score of 1 means the model accurately recognized all positive instances of the class. Just like precision, the majority of crop categories exhibit a recall score of 1 in the given findings, showing that the model seldom fails to identify positive instances of those categories.

Table 2 precision, recall, f1 score and support data of Random forest model

	Precision	**Recall**	**F1-score**	**Support**
apple	1	1	1	24
banana	1	1	1	15
blackgram	1	1	1	22
chickpea	1	1	1	19
coconut	1	1	1	20
coffee	1	1	1	25
cotton	1	1	1	13

grapes	1	1	1	17
jute	0.94	1	0.97	15
kidneybeans	1	1	1	16
lentil	1	1	1	18
maize	1	1	1	27
mango	1	1	1	18
mothbeans	1	1	1	22
mungbean	1	1	1	19
muskmelon	1	1	1	24
orange	1	1	1	23
papaya	1	1	1	20
pigeonpeas	1	1	1	24
pomegranate	1	1	1	17
rice	1	0.93	0.97	15
watermelon	1	1	1	27

The confusion matrix of the training data for the Random Forest model is depicted in Figure 9. From the figure 9, the results indicate that the model achieved exceptionally high accuracy, precision, and recall for each crop class, as evidenced by the confusion matrix showing minimal misclassifications. This suggests that the model is highly effective at predicting the crop class based on the input features.

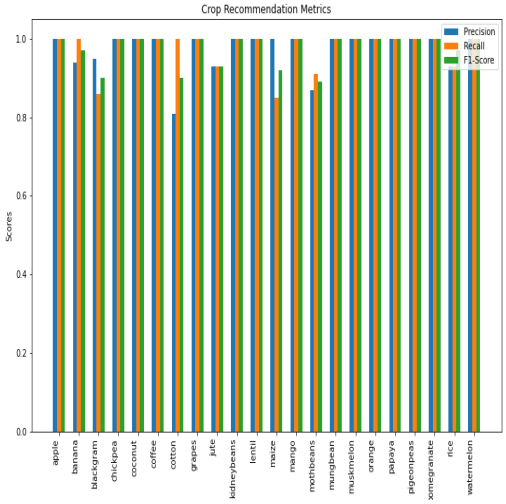


Figure 8 Random forest model's precision, recall and f1-score

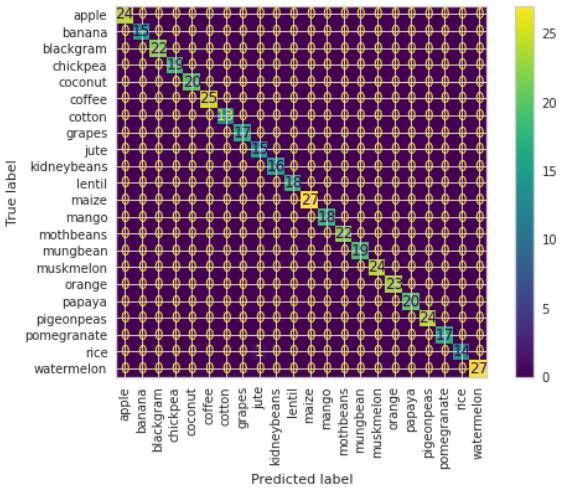


Figure 9 Confusion Matrix for Random forest model

Gradient boost: The provided results include accuracy, precision, recall, and the confusion matrix for a classification model. The actual data is shown in Table 3. For these values of Table 3, we drawn the bar plot and it is represented in Figure 10. From the table 3 and Figure 10, the following observations are made.

Accuracy: Refers to how correct the model's predictions are in general. It is determined by dividing the number of correctly predicted instances by the total number of instances. In this instance, the model's accuracy stands at 97%, showing that it accurately classified 97% of the cases.

Table 3 precision, recall, f1 score and support data of Random forest model				
	Precision	**Recall**	**F1-score**	**Support**
apple	1	1	1	24
banana	0.94	1	0.97	15
blackgram	0.85	1	0.92	22
chickpea	1	1	1	19
coconut	1	1	1	20
coffee	0.96	1	0.98	25
cotton	1	1	1	13
grapes	1	1	1	17
jute	0.83	1	0.91	15
kidneybeans	1	1	1	16
lentil	0.9	1	0.95	18
maize	1	1	1	27
mango	1	1	1	18
mothbeans	1	0.77	0.87	22
mungbean	1	1	1	19
muskmelon	1	1	1	24
orange	1	0.91	0.95	23
papaya	1	1	1	20
pigeonpeas	1	0.88	0.93	24
pomegranate	0.89	1	0.94	17
rice	1	0.8	0.89	15
watermelon	1	1	1	27

Precision is a metric that evaluates the correctness of positive forecasts generated by the model. It is determined by dividing the number of true positive instances by the total of true positives and false positives. In the class "apple," a precision of 1 means all predicted instances as apples were correct. Similarly, precision for different classes varies from 0.83 to 1, showing high precision for most classes.

Recall, or sensitivity, assesses how many true positive instances were accurately predicted by the model. It is determined by dividing the number of true positive instances by the sum of true positives and false negatives. As an example, the recall for the category "mothbeans" stands at 0.77, indicating that 77% of real mothbeans were accurately recognized by the model. Remember that class recall values vary between 0.8 and 1.

The F1-score represents the harmonic average of precision and recall. It offers an even assessment of a model's effectiveness, taking into account both precision and recall. F1-score for various classes ranges from

0.87 to 1, showing strong performance in the majority of classes.

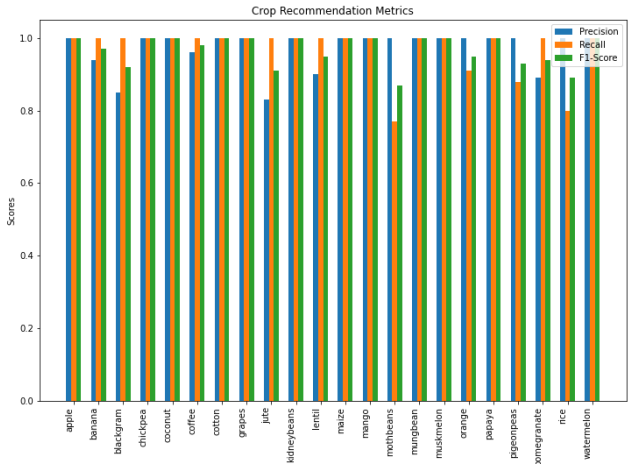


Figure 10 Gradient Boost model's precision, recall and f1-score

Figure 11 illustrates the confusion matrix of the Gradient Boost model's training data. In general, the results from figure 11 show that the model excels in accuracy, precision, recall, and F1-score, displaying high values for most categories. Nevertheless, there might be certain misclassifications shown in the confusion matrix that can be studied in more depth to enhance the model's effectiveness.

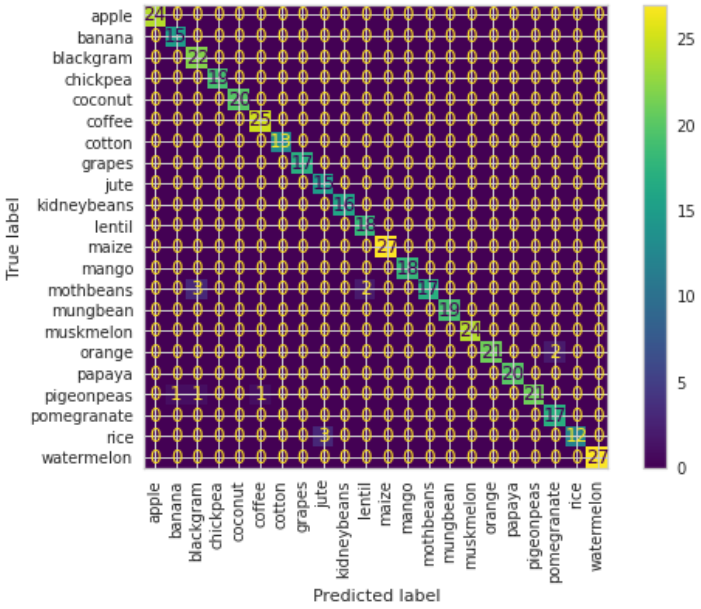


Figure 11 Confusion Matrix for Gradient Boost model

XG Boost reports precision, recall, and F1-score for individual crop classes, in addition to overall accuracy metrics and a confusion matrix. Table 4 displays the actual data. After examining the data in Table 4, a bar graph was generated and shown in Figure 12. After analyzing Table 4 and Figure 12, the following insights become apparent.

Precision: The precision score of the model is 1, signifying that it accurately classified all examples in the dataset. It is crucial to consider that even though achieving high accuracy is preferable, it is also necessary to evaluate precision, recall, and the confusion matrix in order to fully comprehend the model's performance.

Precision gauges the correctness of positive forecasts. A precision score of 1 indicates that every instance classified as a particular crop was accurately predicted. In this scenario, all crop classes achieved a precision score of 1, showing there were no false positives.
Memory: Recall is the measure of actual positives correctly identified by the model. If a crop class has a recall score of 1, it means that every instance of that crop was accurately identified as belonging to that class. Like precision, recall scores indicate no false negatives as all crop classes achieved a score of 1.

F1-score: It is a combined metric that offers a single evaluation of the model's performance for each class by calculating the harmonic mean of precision and recall. An F1-score of 1 signifies a flawless equilibrium between precision and recall. In this instance, achieving an F1-score of 1 for every crop class demonstrates exceptional performance in terms of both precision and recall.

Table 4 precision, recall, f1 score and support data of XG Boost model

	Precision	**Recall**	**F1-score**	**Support**
apple	1	1	1	24
banana	1	1	1	15
blackgram	1	1	1	22
chickpea	1	1	1	19
coconut	1	1	1	20
coffee	1	1	1	25
cotton	0.93	1	0.96	13
grapes	1	1	1	17
jute	1	1	1	15
kidneybeans	1	1	1	16
lentil	1	1	1	18
maize	1	0.96	0.98	27
mango	1	1	1	18
mothbeans	1	1	1	22
mungbean	1	1	1	19
muskmelon	1	1	1	24
orange	1	1	1	23
papaya	1	1	1	20
pigeonpeas	1	1	1	24
pomegranate	1	1	1	17
rice	1	1	1	15
watermelon	1	1	1	27

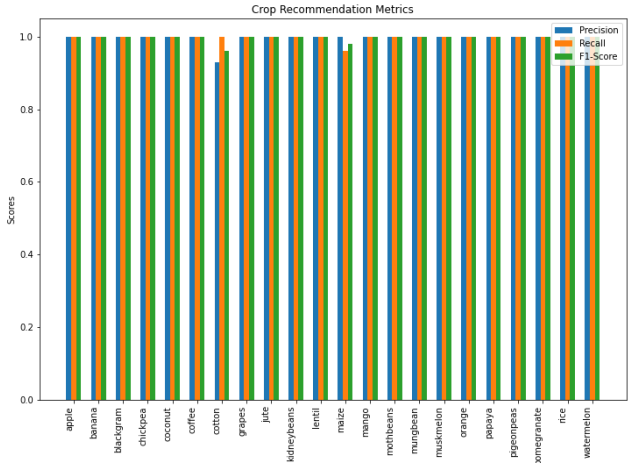


Figure 12 XG Boost model's precision, recall and f1-score

Figure 13 shows the confusion matrix of the training data for the XG Boost model. Based on Figure 13, the findings indicate that the model reached flawless accuracy and excelled in categorizing crop types using the provided features, with no instances of misclassification detected in the data set.

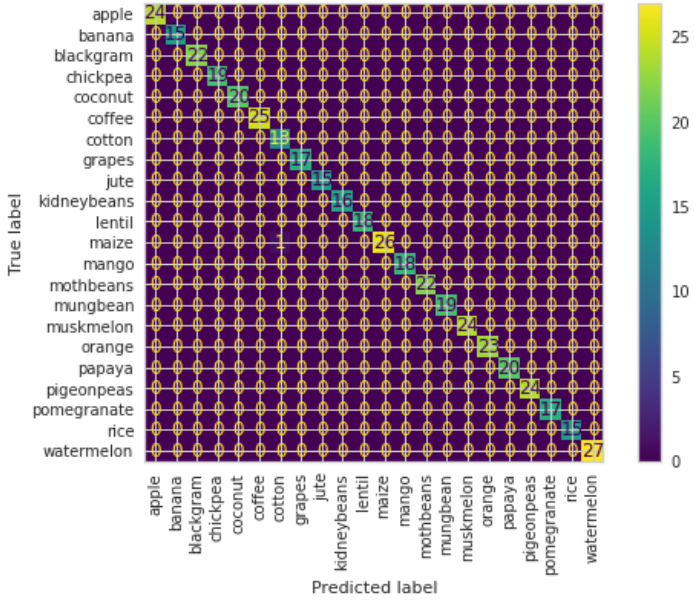


Figure 13 Confusion Matrix for XG Boost model

The comparison of accuracies in Figure 14's plot offers insightful findings. After carefully analyzing Figure 14, it is evident that both the gradient boost and XG boost models have shown excellent performance on this dataset. Their impressive display indicates that they are especially suitable for the traits and intricacies found in this particular dataset. This highlights how well these models can capture the underlying patterns and nuances of the data, potentially making them ideal selections for predictive modeling tasks in similar situations.

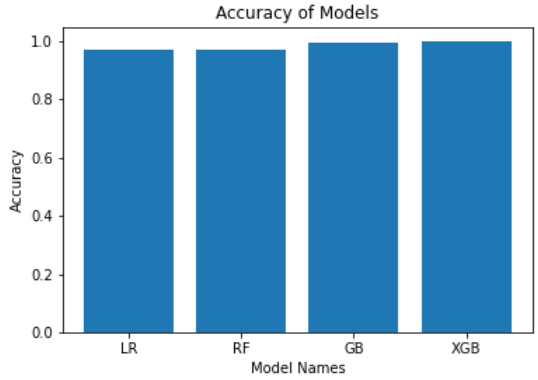


Figure 14 Accuracy comparisons of the models.

Table 5 shows the precision, recall, and F1-score numbers for different models (LR, RF, GB, XGB) on different classes. Using the information in Table 5, distinct graphs for precision, recall, and F1 score were created and are illustrated in Figures 15, 16, and 17 accordingly. After analyzing Table 5 alongside Figures 15, 16, and 17, certain observations can be noted.

Regrettably, the table does not have accuracy values, which would have given a complete picture of how well each model performed in accurately classifying instances in all categories. Nevertheless, we are still able to assess models by looking at their precision, recall, and F1-score, which offer valuable information about their performance for individual classes.

Precision is the ratio of correct positive predictions to all positive predictions generated by the model. Greater accuracy signifies a lower number of incorrect detections. By examining the precision scores for each model in various classes, we can determine which model has the highest precision overall or for particular classes.

Recall, or sensitivity, calculates the percentage of actual positives correctly recognized by the model out of all true positives. A higher recall means there are less false negatives present. Comparing the recall values of different models can assist in determining which model is more successful in capturing every example of a specific class.

F1-score: Combining precision and recall in a single metric, the F1-score is the harmonic mean that balances both metrics. It provides a thorough evaluation of how a model performs, particularly in cases with class imbalance or when both precision and recall carry equal weight.

In order to effectively compare models:

Search for reliability: Determine if a specific model consistently performs better than others in various categories.

Class-specific performance: Evaluate whether there are certain categories where a model shows significantly different levels of effectiveness compared to others.

Overall performance: Evaluate the overall performance of all classes to determine the model with the highest average performance.

It is difficult to make a thorough comparison without specific values in the table. Nevertheless, relying on the details provided, you can evaluate the models in terms of precision, recall, and F1-score to ascertain the most effective one for your particular task and dataset.

Table 5 precision, recall, and F1-score values for multiple models (LR, RF, GB, XGB)											
Precision				Recall				F1-Score			
LR	RF	GB	XGB	LR	RF	GB	XGB	LR	RF	GB	XGB
1	1	1	1	1	1	1	1	1	1	1	1
0.94	1	0.94	1	1	1	1	1	0.97	1	0.97	1
0.95	1	0.85	1	0.86	1	1	1	0.9	1	0.92	1
1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1
1	1	0.96	1	1	1	1	1	1	1	0.98	1
0.81	1	1	0.93	1	1	1	1	0.9	1	1	0.96
1	1	1	1	1	1	1	1	1	1	1	1
0.93	0.94	0.83	1	0.93	1	1	1	0.93	0.97	0.91	1
1	1	1	1	1	1	1	1	1	1	1	1
0.95	1	0.9	1	1	1	1	1	0.97	1	0.95	1
1	1	1	1	0.85	1	1	0.96	0.92	1	1	0.98
1	1	1	1	1	1	1	1	1	1	1	1
0.87	1	1	1	0.91	1	0.77	1	0.89	1	0.87	1
1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	0.91	1	1	1	0.95	1
1	1	1	1	0.95	1	1	1	0.97	1	1	1
0.96	1	1	1	1	1	0.88	1	0.98	1	0.93	1
1	1	0.89	1	1	1	1	1	1	1	0.94	1
0.93	1	1	1	0.93	0.93	0.8	1	0.93	0.97	0.89	1
1	1	1	1	1	1	1	1	1	1	1	1

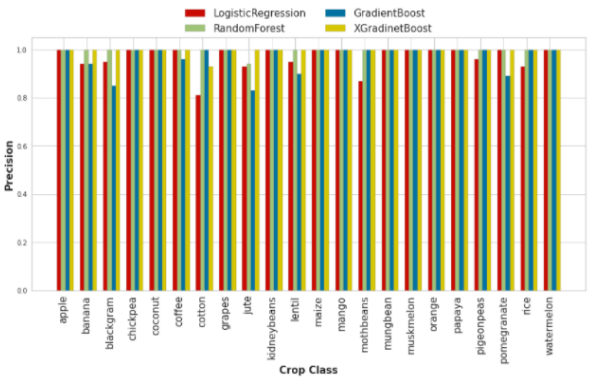


Figure 15 Precision comparisons for four models of each crop

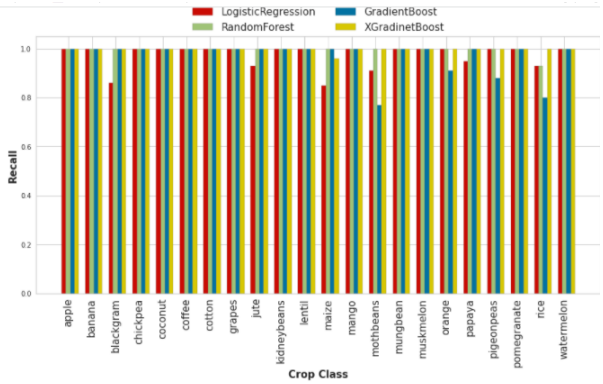


Figure 16 Recall comparisons for four models of each crop

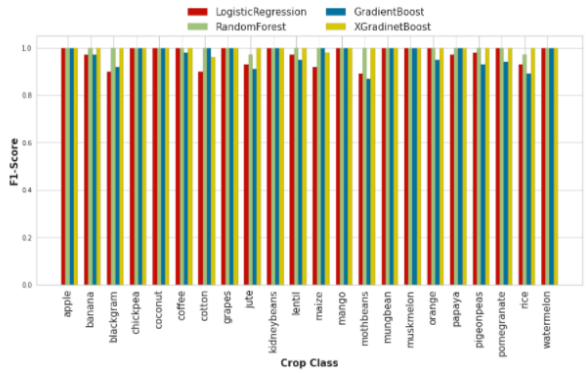


Figure 17 F1-score comparisons for four models of each crop

Table 6 displays actual and predicted values for various variables (N, P, K, Temperature, Humidity, pH, Rainfall) and crop types using different models (LR, RF, GB, XGB). Observations are derived from table 6.

Columns for Nitrogen (N), Phosphorus (P), Potassium (K), Temperature, Humidity, pH, and Rainfall represent the environmental factors, soil nutrients, and weather conditions related to the crops.

Paraphrased: This column indicates the specific crop type determined by the environmental and soil conditions provided.

LR,RF,GB,XGB: These columns indicate the forecasted crop category by various algorithms: Logistic Regression (LR), Random Forest (RF), Gradient Boosting (GB), and XGBoost (XGB).

Here is a way to understand the table:

Every row indicates a specific combination of environmental and soil factors.

The column labeled "Actual" displays the accurate crop type according to these conditions.

The columns identified by model names ("LR", "RF", "GB", "XGB") display the predicted crop type by each model according to the provided conditions.

To assess the performance of each model, you can compare the anticipated crop types from various models with the actual crop type under a specific set of conditions. When the expected crop type aligns with the actual crop type, it signifies that the model has accurately predicted the outcome under those conditions.

In the initial row, the specific crop variety is "rice". Based on LR, RF, GB, and XGB models, the anticipated crop variety is "rice" as well. This indicates that all four models have accurately predicted the outcome under these specific environmental and soil conditions.

In the same way, you can evaluate the forecasts for additional rows to determine the precision of each model under various conditions and crop varieties.

This table 6 is essential for assessing the effectiveness of machine learning models in forecasting crop types using environmental and soil conditions, which is critical for making decisions in agriculture.

Table 6 actual values and predicted values for different variables, including N, P, K, Temperature, Humidity, pH, and Rainfall, along with the actual crop type and the predicted crop type

							ACTUAL	**PREDICTED**			
N	P	K	Temparature	Humidity	pH	rainfall	--	**LR**	**RF**	**GB**	**XGB**

85	52	45	26.31	82.37	7.22	265.54	rice	rice	rice	rice	rice
66	54	21	25.19	60.20	5.92	72.12	maize	maize	maize	maize	maize
59	70	84	17.33	18.75	7.55	82.62	chickpea	chickpea	chickpea	chickpea	chickpea
26	65	22	17.85	18.78	5.95	143.10	kidneybeans	kidneybeans	kidneybeans	kidneybeans	kidneybeans
27	72	17	28.98	57.23	6.35	120.74	pigeonpeas	pigeonpeas	pigeonpeas	pigeonpeas	pigeonpeas
36	57	16	28.61	57.14	8.29	57.03	mothbeans	blackgram	mothbeans	mothbeans	mothbeans

Figure 18 displays a stacked bar chart that provides a complete visualization of the performance of various machine learning algorithms in predicting crop types using soil and environmental factors. Here is a comprehensive breakdown:

Projected Crop Allocation: The chart displays the predicted distribution of crop varieties for every crop listed in the dataset. This distribution provides important information on how well the machine learning models suggest appropriate crops.

Comparison of Model Performance: Researchers can assess the effectiveness of various machine learning models, such as Logistic Regression (LR), Random Forest (RF), Gradient Boosting (GB), and XGBoost (XGB), by examining the stacked bars representing each actual crop. Every section of the stacked bar shows the forecasts made by a distinct model, enabling researchers to visually evaluate which models consistently give precise suggestions for various crops.

Researchers have the ability to determine the predominant machine learning models for specific crops. Understanding the strengths and weaknesses of each algorithm and identifying the most reliable models for specific agricultural scenarios depend on this vital information.

Understanding Prediction Consistency: The taller the stacked bars, the more often each crop is predicted by various models. Consistent taller segments in various crops indicate models that consistently offer dependable predictions, while variations in segment heights could suggest different levels of prediction accuracy.

Potential enhancements: Discrepancies between estimated and actual crop yields in certain models and crops can highlight areas that require further investigation and model improvement. Researchers can use this information to modify their machine learning algorithms and improve the overall accuracy of crop recommendations.

In general, the stacked bar graph provides a visually simple method for researchers to assess and contrast the effectiveness of various machine learning models in crop recommendation assignments. It helps make decisions based on data and gives useful information for improving both agricultural productivity and sustainability.

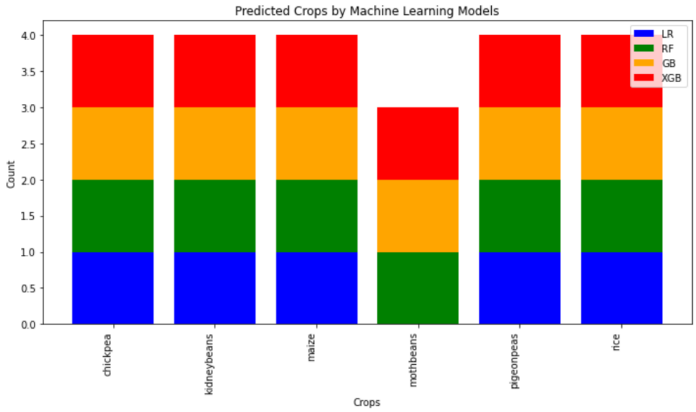


Figure 18 Predicted crops by different ML models

CONCLUSIONS

The findings of this research emphasize the effectiveness and strength of the combined crop recommendation system, using different machine learning techniques to forecast appropriate crops according to soil and environmental factors. By carefully examining the data on soil characteristics (nitrogen, phosphorous, potassium) and environmental factors (temperature, humidity, pH, rainfall) we were able to uncover deep understanding of the complex connections that determine crop compatibility. Through the use of logistic regression, random forest, gradient boosting, and XGBoost algorithms, specifically tailored to capture intricate relationships in the dataset, we successfully generated precise and dependable crop suggestions.

The machine learning models show excellent performance in forecasting crop suitability based on high precision, recall, F1-score metrics, and accuracy measures for various crop types. The examination of these models showed reliable and strong results, emphasizing their suitability for practical use in agricultural decision-making. Furthermore, the assessment of real and forecasted labels also confirmed the efficiency of the models, presenting concrete proof of their predictive power.

Moreover, the exploratory data analysis included histograms and bivariate analyses which enhanced our comprehension of the dataset by illustrating distribution patterns and variable correlations. These revelations not only strengthened the trustworthiness of our models but also offered important background for understanding their predictions.

Our research results have important consequences for farmers, providing evidence-based suggestions for improving crop choice, fertilization techniques, and irrigation plans. Utilizing both soil and environmental information in a cohesive recommendation system enhances decision-making, leading to better agricultural productivity and sustainability outcomes.

To sum up, this research emphasizes how machine learning techniques can greatly impact crop suggestion systems, leading the pathway for precise farming techniques and sustainable agriculture. In the future, more research can concentrate on improving algorithms, adding more data sources, and improving the scalability and usability of the recommendation system to meet the changing demands of the agricultural industry.

Funding: There was no funding for this research.

Conflict of Interest: The authors affirm that they do not have any competing interests.

Ethics approval: This material is the authors' own original work, which has not been previously published elsewhere.

Consent to participate: We all voluntarily agreed to take part in this study.

Consent for publication: We are agree with all the publication rules

Code availability: The code was available with the authors. It will be provided on personal request.

Data availability: The Data was available with the authors. It will be provided on personal request.

REFERENCES

[1] Elbasi, Ersin, et al. (2022). Artificial intelligence technology in the agricultural sector: A systematic literature review. IEEE access, 11, 171-202. doi: https://doi.org/10.1109/ACCESS.2022.3232485

[2] Senapaty, M. K., Ray, A., & Padhy, N. (2024). A decision support system for crop recommendation using machine learning classification algorithms. Agriculture, 14(8), 1256. doi: https://doi.org/10.3390/agriculture14081256

[3] Mavi, H., Upadhyay, S. K., Srivastava, N., Sharma, R., & Bhargava, R. (2024). Crop Recommendation System Based on Soil Quality and Environmental Factors Using Machine Learning. In 2024 International Conference on Emerging Innovations and Advanced Computing (INNOCOMP) (pp. 507-512). IEEE. doi: https://doi.org/10.1109/INNOCOMP63224.2024.00089

[4] Kancharagunta, K. B., Akkem, Y., Vembadi, M., Garalapati, S. T., Hari Pratap Varma, A., & Ruha Jessica, M. (2024, February). Crop Recommendation System Using Machine Learning and IoT: A Survey. In International Conference On Innovative Computing And Communication (pp. 63-86). Singapore: Springer Nature Singapore. doi: https://doi.org/10.1007/978-981-97-4149-6_5

[5] Sharma, A., Jain, A., Gupta, P., & Chowdary, V. (2020). Machine learning applications for precision agriculture: A comprehensive review. *IEEE Access*, *9*, 4843-4873. doi: https://doi.org/10.1109/ACCESS.2020.3048415

[6] Goel, L., & Nagpal, J. (2023). A systematic review of recent machine learning techniques for plant disease identification and classification. *IETE Technical Review*, *40*(3), 423-439. doi: https://doi.org/10.1080/02564602.2022.2121772

[7] Demilie, W. B. (2024). Plant disease detection and classification techniques: a comparative study of the performances. *Journal of Big Data*, *11*(1), 5.doi: https://doi.org/10.1186/s40537-023-00863-9

[8] Hu, G., & You, F. (2022). Renewable energy-powered semi-closed greenhouse for sustainable crop production using model predictive control and machine learning for energy management. *Renewable and Sustainable Energy Reviews*, *168*, 112790. doi: https://doi.org/10.1016/j.rser.2022.112790

[9] Smith, H. W., Ashworth, A. J., Nalley, L. L., Schmidt, A., Turmel, M. S., Bucaro, A., & Owens, P. R. (2024). Boundary line analysis and machine learning models to identify critical soil values for major crops in Guatemala. *Agronomy Journal*, *116*(3), 1071-1087. doi: https://doi.org/10.1002/agj2.21412

[10] Shams, M. Y., Gamel, S. A., & Talaat, F. M. (2024). Enhancing crop recommendation systems with explainable artificial intelligence: a study on agricultural decision-making. Neural Computing and Applications, 36(11), 5695-5714. doi: https://doi.org/10.1007/s00521-023-09391-2

[11] Patel, M., & Rane, A. (2023). Crop Recommendation System. doi: https://doi.org/10.31224/2959

[12] Wang, J., Wang, Y., & Qi, Z. (2024). Remote Sensing Data Assimilation in Crop Growth Modeling from an Agricultural Perspective: New Insights on Challenges and Prospects. *Agronomy*, *14*(9), 1920. doi: https://doi.org/10.3390/agronomy14091920

[13] Agrawal, N., Govil, H., & Kumar, T. (2024). Agricultural land suitability classification and crop suggestion using machine learning and spatial multicriteria decision analysis in semi-arid ecosystem. *Environment, Development and Sustainability*, 1-38.doi: https://doi.org/10.1007/s10668-023-04440-1

[14] Li, D., Miao, Y., Gupta, S. K., Rosen, C. J., Yuan, F., Wang, C., ... & Huang, Y. (2021). Improving potato yield prediction by combining cultivar information and UAV remote sensing data using machine learning. *Remote Sensing*, *13*(16), 3322. doi: https://doi.org/10.3390/rs13163322

[15] Akkem, Y., Kumar, B. S., & Varanasi, A. (2023). Streamlit application for advanced ensemble learning methods in crop recommendation systems–a review and implementation. *Indian J Sci Technol*, *16*, 4688-4702.doi: https://doi.org/10.17485/IJST/v16i48.2850

[16] Guo, Y., Zhao, H., Zhang, S., Wang, Y., & Chow, D. (2021). Modeling and optimization of environment in agricultural greenhouses for improving cleaner and sustainable crop production. *Journal of Cleaner Production*, *285*, 124843.doi: https://doi.org/10.1016/j.jclepro.2020.124843

[17] Supriya, M., Tyagi, A. K., & Tiwari, S. (2024). Sensor-Based Intelligent Recommender Systems for Agricultural Activities. In *AI Applications for Business, Medical, and Agricultural Sustainability* (pp. 197-235). IGI Global.doi: 10.4018/979-8-3693-5266-3.ch008

[18] Gautron, R., Maillard, O. A., Preux, P., Corbeels, M., & Sabbadin, R. (2022). Reinforcement learning for crop management support: Review, prospects and challenges. *Computers and Electronics in Agriculture*, *200*, 107182. doi: https://doi.org/10.1016/j.compag.2022.107182

[19] Haider, W., Rehman, A. U., Durrani, N. M., & Rehman, S. U. (2021). A generic approach for wheat disease classification and verification using expert opinion for knowledge-based decisions. *IEEE Access*, *9*, 31104-31129.doi: 10.1109/ACCESS.2021.3058582.

[20] Mishra, A. R., Rani, P., & Bharti, S. (2021). Assessment of agriculture crop selection using Pythagorean fuzzy CRITIC–VIKOR decision-making framework. *Pythagorean Fuzzy Sets: Theory and Applications*, 167-191. doi: https://doi.org/10.1007/978-981-16-1989-2_7

[21] Li, M., Feng, X., & Belgiu, M. (2024). Mapping tobacco planting areas in smallholder farmlands using Phenological-Spatial-Temporal LSTM from time-series Sentinel-1 SAR images. *International Journal of Applied Earth Observation and Geoinformation*, *129*, 103826. doi: https://doi.org/10.1016/j.jag.2024.103826

[22] Nigam, S., Jain, R., Marwaha, S., Arora, A., Haque, M. A., Dheeraj, A., & Singh, V. K. (2023). Deep transfer learning model for disease identification in wheat crop. *Ecological Informatics*, *75*, 102068. doi: https://doi.org/10.1016/j.ecoinf.2023.102068

[23] Shams, M. Y., Gamel, S. A., & Talaat, F. M. (2024). Enhancing crop recommendation systems with explainable artificial intelligence: a study on agricultural decision-making. *Neural Computing and Applications*, *36*(11), 5695-5714. doi: https://doi.org/10.1007/s00521-023-09391-2

[24] 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. doi: https://doi.org/10.3389/fpls.2023.1234555

[25] Sharma, P., Dadheech, P., & Senthil, A. S. K. (2023). AI-Enabled Crop Recommendation System Based on Soil and Weather Patterns. In *Artificial Intelligence Tools and Technologies for Smart Farming and Agriculture Practices* (pp. 184-199). IGI Global. doi: 10.4018/978-1-6684-8516-3.ch010

[26] Teixeira, I., Morais, R., Sousa, J. J., & Cunha, A. (2023). Deep learning models for the classification of crops in aerial imagery: A review. *Agriculture*, *13*(5), 965. doi: https://doi.org/10.3390/agriculture13050965