SvedbergOpen
DISSEMINATION OF KNOWLEDGE

International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: https://www.svedbergopen.com/

Research Paper Open Access

Design And Implementation Of A High-Accuracy Machine Learning Model For Multi-Crop Recommendation In Smart Farming

Rajeev Kumar[1], Sunita Dahiya[2]

[1]Research Schoolar, [2]Assistant Professor
[1,2]Department of Electronics and Communication Engineering, Deenbandhu Chhotu Ram University of Science and Technology, Haryana, India
[1]rajeevdhanda208@gmail.com, [2]sunitadahiya.ece@dcrustm.org

Abstract

Precision agriculture has proved to be a successful strategy in enhancing agricultural productivity by making decisions based on data. This study aims to develop a Machine Learning based Crop Recommendation System (CRS) which suggests appropriate crops for growers according to the nutrients in the soil and environmental condition. A well-balanced dataset with 22 crop classes, 7 input features (Nitrogen (N), Phosphorus (P), Potassium (K), temperature, humidity, pH, rainfall) and 2200 samples were used to test the performance of 26 machine learning classifiers. For the models, they were evaluated based on the Accuracy, Balanced Accuracy, Precision, Recall, F1-Score and Execution Time. Experimental results showed that the Extra Trees Classifier and Random Forest Classifier performed the best in terms of Accuracy with 99.36%, with the Extra Trees Classifier being more efficient in terms of computational time, and thus would be recommended for real-time crop selection. The statistical analysis also revealed that 15 of 26 models outperform 95% accuracy showing the framework's robustness and reliability. The developed CRS helps farmers to choose the most suitable crops based on soil fertility and climatic conditions to increase productivity, optimize use of resources and sustainable farming practices. Based on this proposed framework, an accurate, scalable and computationally efficient decision support system is built that has significant potential applications for precision agriculture and smart farming.

Keyword: Smart Farming, Machine Learning, Crop Recommendation System (CRS), Multi-Crop Classification, Precision Agriculture

Introduction

The rapid evolution of technology has made it an indispensable part of everyday life. Nowadays, smartphones, smart applications, and advanced devices are indispensable, playing a crucial role in enhancing our comfort, efficiency, and overall quality of life (1). India is one of the biggest agricultural producers in the world with over 140 million hectares of net sown area and over 50% of its population relying on agriculture for their livelihood. In agriculture, machine learning models are primarily directed at crop, soil and water management, using remote sensing technologies (2). It highlights the transformative power of Information and Communication Technology (ICT) tools such as Artificial Intelligence (AI), Internet of Things (IoT), robotics and sensors in the modernization of conventional farming. The paper also covered the challenges of implementation and directions for further research for AI in agriculture (2-3). However, Indian agriculture is still very much exposed to climate variations, reliance on monsoons, soil degradation and poor management of inputs. Indian Council of Agricultural Research (ICAR) data shows that over 70% of Indian farmers are small and marginal farmers, who lack access to scientific knowledge and tools for optimizing crop production. Traditional or classical farming in India is largely based on experience-based crop selection, failing to take into account the critical role of soil nutrients, temperature and agro climatic factors (4-5). This results in reduced productivity, over-fertilization and impractical farming practices. For instance, over cultivation of water intensive crops like rice and sugarcane in wrong places has led to depletion of the groundwater and degradation of the soils. In contrast, smart farming leverages data-driven decision-making, Internet of Things (IoT) data monitoring, and machine learning (ML) algorithms to support farmers' selection of the right crop, which is based on real-time data (6–8). Historical data include nitrogen, phosphorus, potassium concentrations, pH, temperature, humidity, and precipitation. Crop recommendation systems are essential as they propose the most appropriate crops based on a farmer's field circumstances, hence optimizing productivity and reducing environmental impact. Agricultural output is a national priority because of the increasing effects of climate change on crops and pest control. To tackle these issues, contemporary agriculture increasingly depends on Industry 4.0 technologies such as robotics, IoT, and machine learning for effective field and greenhouse management (9). The integration of ML algorithms for crop recommendation is particularly noteworthy in India, where different agricultural zones are available, unpredictable monsoon patterns, and uneven landholdings require precision and adaptation. Machine Learning Algorithms can recognize patterns and produce specific forecasts, preparing Indian farmers with appropriate action after analysing the large historic agricultural data of various domain. This article presents a multi-crop recommendation system as per soil conditions to farmers using more than 25 machine learning methods with excellent accuracy. This system helps to fill the information gap in countryside farming societies by assisting state of art technology, crop recommendation as per requirement of soil nutrients and location. The effect of this will leads to optimized resources utilization, growing yield and promote sustainable agriculture in India.

2. RELATED WORK

Agronomy is poised for a radical change led by Machine Learning (ML) in the modern era of digital data explosion. Emphasizing the transforming power of machine learning on modern farming approaches, this study investigates the dynamic merging of Information and Communications Technology (ICT) with conventional agricultural techniques (10). On two levels, the proposed approach offers a novel ensemble-based machine learning categorization system. First level (level-0) basic classifiers include logistic regression, CART, SVM, and KNN are used to provide first predictions. The outputs provide input data for the Random Forest level-1 meta-classifier, which runs the last classification. With a phenomenal accuracy rate of 99% (11). This stacked ensemble approach shows great prediction ability in separating 22 different crop categories. Three main criteria—MSE, R^2, and MAE, evaluate success in this work. With a rather low MSE of 0.9412 and exceptional accuracy in projecting crop yields (12). The empirical results clearly highlight the outstanding efficacy of the XAI-CROP model. When compared to other ML models used for crop advisory uses, the suggested approach proved to be remarkably dependable, accurate, and generally efficient. It confirmed its rank as the most exact solution in the research by beating all other evaluated techniques with an amazing accuracy rate of 99.31% (13). In order to improve land cover classification accuracy using LiDAR-derived DSMs, particularly with limited training data, this work presents a unique feature space using local kernel descriptors obtained from morphological profiles (MPs.). Tested on Houston and Trento datasets, the suggested SVM-based methodology attained a 93.75% accuracy—much above conventional techniques by up to 57% (14). Using supervised learning, the optimal model was developed through designated ML algorithms in WEKA. The chosen classifiers—Multilayer Perceptron, JRip, and Decision Table—accomplished a performance accuracy of 98.23%, with a weighted average ROC of 1.0 and a maximum model-building time of 8.05 seconds (15). This article proposes a context for sensing and categorizing invasions in IoT networks utilized in agriculture. With privacy and security being key issues in the Internet of Things (IoT) world, the study highlights the significance of these issues in an agricultural context. It reviews different ML algorithms with respect to different parameters such as Recall, accuracy, and precision (16). Real-time context-appropriate fertilizer recommendations are essential for precision agriculture to be both sustainable and efficient. A number of current systems neglect some of the important considerations such as the type of crop, soil type, and soil fertility. In this solution, the proposed model is Gaussian Naive Bayes (GNB) which outperforms all other machine learning models with 96% accuracy in training set and 94% accuracy in testing set (17). The study aims to fill this gap in the literature by proposing a new architecture combining IoT and the use of Machine Learning for continuous crop quality assessment. The architecture consists of three integrated layers that gather, process and analyze data from various sources. The results of the experiments demonstrate that the error in the prediction can be reduced by aggregating data from various sources as shown in figure 6, which results in a percentage error of 6.59 which is less than the percentage error of 6.71 in case of sensor data only (18).

Energy use becomes a key component in several fields of smart farming such as crop cultivation, plant growth, device mechanism and energy management. Comparative evaluation of the ML algorithms, like ANN, SVR, Random forest, KNN, XGBoost and ARIMA were used to find the most optimized prediction model. The best prediction accuracy of 92% (19) was achieved by the Random Forest model with seven performance criteria which is better than the other models. In this work, a real-time, field curated database of grape leaf images from India is provided, which is then used to use deep learning to identify grape diseases. The highest accuracy in the classification of healthy and disease leaves was obtained by the transfer learning approach with CNN models; particularly the ResNet50V2 model which is better than the current approaches and provides a worthwhile benchmark for future research (20). In the agriculture field, yield prediction is a crucial but difficult problem which is the primary goal of making better predictions of higher yield using machine learning algorithms. Due to its emerging significance, this article provides a detailed overview of ML techniques used for crop yield prediction, with a specific emphasis on palm oil yield prediction (21). A review of various research work is carried out to provide an overview of the automation practice in agriculture, especially with the help of robots and drones in the field of weeding. It covers various soil water sensing methods and two automated weeding methods, and also the use of drone for spraying and crop monitoring (22). Agriculture contributes 6.4% of the world economic output and is the main economic activity in at least nine countries. The main objectives of this paper are to review various applications like meticulousness farming, disease identification in the farm, crop phenotyping etc. that are using advanced technologies like AI, ML, Internet of things (IoT), WSN, CNN and other technologies (23). The main object of this work is to focus on fertilization recommendation for crop for rural farmers of Hua County. It attacks the wheat, maize and peanut crops and uses Object-Oriented approach to design a model of the balance of fertilizing crops of vegetables. The development of a village-scale decision making system with ArcGIS Server showed that the system would be more expansible than the previous systems (24). For optimizing network performance and energy economy in smart agriculture, data clustering and selection of cluster head is optimized in an IoT-based WSN framework using the two methods: Trilevel K-means clustering, and Arithmetic optimization method. The proposed method has higher throughput (96 kbps), better packet delivery (37%), shorter latency (0.28%) and less energy consumption (0.52 J) (24) compared to the existing methods. This paper discusses the applications of WSNs in Agriculture and compares wireless protocols (ZigBee, LoRa, Wi-Fi, 5G) and routing techniques for the effective monitoring and control. It focuses on energy efficient technologies, primary barriers and smart farming solutions based on IoT (24) from 2018 to 2023. The results of these studies all show significant advances in the application of ML to create multi-crop recommendation systems, advancing the development of smart farming practices.

3. METHODOLOGY

3.1 Dataset Description

This study is based on a structured and balanced dataset of agriculture, especially created for the development of a high accuracy multi-crop recommendation system. The data set included 2200 cases, and 100 samples for each of the 22 crop kinds. The data chunks are created from publicly available information from various sources such as Kaggle, IEEE DataPort, UC Irvine Machine Learning Repository etc. Each data instance contains 7 important features for input to show the environment and soil conditions for a particular land plot. These qualities were selected because they play an important role in the health and growth stages of plants and general production potential. Nitrogen (N), Phosphorus (P) and Potassium (K) measured in ppm (parts per million)

- **Temperature** – measured in degrees Celsius (°C)
- **Humidity** – measured as a percentage (%)
- **pH** – representing the soil acidity or alkalinity
- **Rainfall** – measured in millimeters (mm)

Target Variable: Crop Label: The dataset used in this research article consists of 22 different crop classes which we have to recommend to the farmers. Dataset having seven input features and target (output) variable is multiclass which represents the suitable crop selection under the given circumstances.

Figure 1. Crop classes used for Crop Recommendation

Covering a wide range of grains, pulses, fruits, cash crops, and legumes, these crops were chosen for their agricultural importance both in India and outside. This multiplicity enables the model to view the crop broadly that is from different crops and areas.

3.2 PREPROCESING

A uniform sample size of 100 samples per crop ensures a class balance required to train unbiased ML classification models. In addition, pre-processing ensures:

- No missing values
- Proper normalization using Min-Max Scaling
- Randomized distribution to avoid sequence bias
- 80:20 split for training and testing sets

This dataset provides a robust foundation for building an effective ML-driven recommendation system that can accurately match environmental parameters to suitable crops, enabling informed and sustainable decisions in smart farming.

3.3 MACHINE LEARNING MODELS

A wide range of over 26 machine learning methods were used in this work to build a comprehensive multi-crop recommendation model. These models were selected to cover a wide spectrum of algorithmic paradigms, including ensemble learning, probabilistic classifiers, kernel-based models, linear models, instance-based learning, and graph-based semi-supervised learning. Out of these ML algorithms random forest classifier and extra trees classifier are illustrated below.

Extra Trees Classifier:

The Extra Trees Classifier (ETC) is an collaborative algorithm that builds numerous decision trees, introducing maximum randomness in two key areas:

- Selecting the splitting thresholds randomly for each feature.
- Using the entire training dataset (no bootstrapping).

Each tree T_m in the ensemble is trained on the original dataset, and predictions are made by majority voting across all M trees. The implemented models are listed below by category:

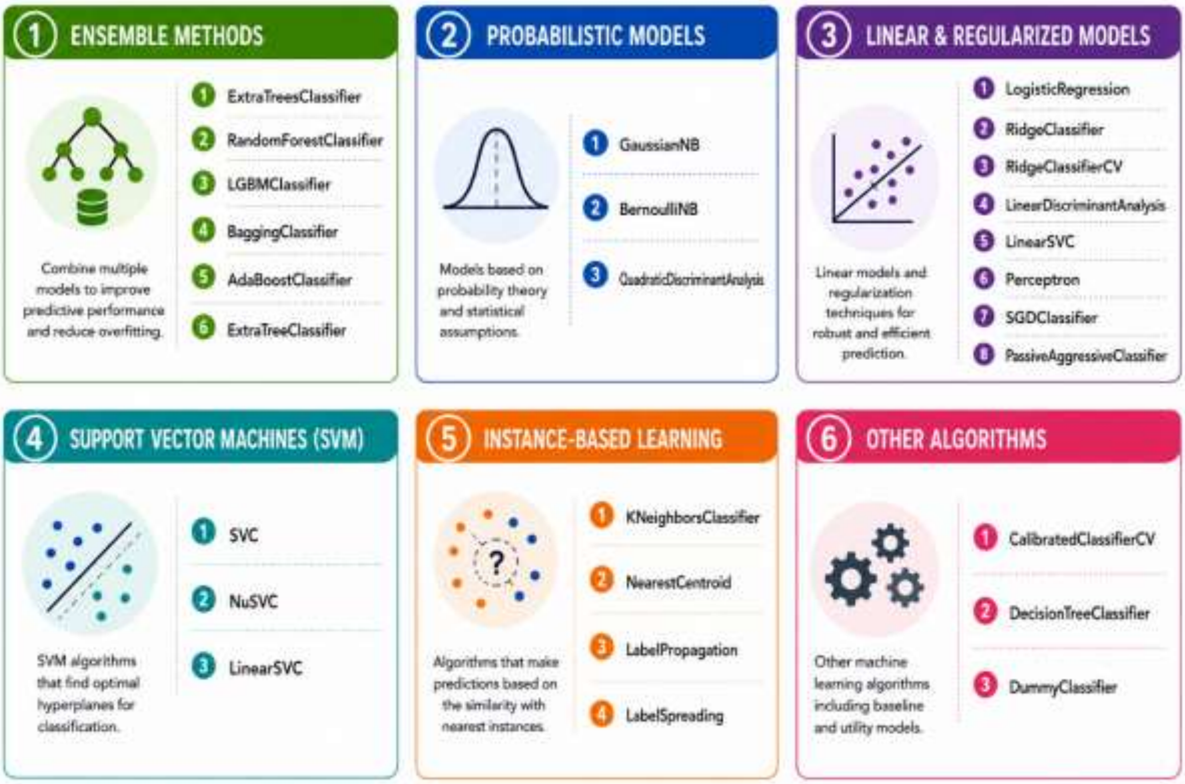


Figure 2. Categorization of Machine Learning Algorithms Evaluated for Crop Recommendation

The predicted class label $\hat{y}$ for an input vector x is given by:

$$\hat{y} = arg\,\max_{c \in C} \sum_{m=1}^{M} I\,(Tm(\dot{x}) = C) \ldots \ldots \ldots \ldots (1)$$

Where:

- C is the set of all crop classes (22 in this case),
- $T_m(x)$ is the prediction of the m^{th} tree,
- I is the indicator function that returns 1 if the condition is true, 0 otherwise.

Random Forest Classifier

The Random Forest Classifier (RFC) also builds an ensemble of decision trees, but differs in that:

- It uses bootstrap samples (random sampling with replacement) for training each tree.
- It searches for the optimal split among a random subset of features at each node.

Prediction:

$$\hat{y} = arg\,\max_{c \in C} \sum_{m=1}^{M} I\,(hm(\dot{x}) = C) \ldots \ldots \ldots \ldots (2)$$

Where:

- $h_m(x)$ is the class prediction of the mmm-th tree trained on bootstrap sample $D_m \subset D$.

Tree Construction:

For a given node:

- Let Ƒ⊂F be a randomly selected subset of features.
- Find the best split based on impurity minimization (e.g., using **Gini impurity** or **entropy**):

Entropy = $-\sum_{i=1}^{M} p_i \, log_2 \, p_i$................(3)

The tree grows recursively until stopping conditions are met (e.g., max depth or min samples per leaf). The randomness in sampling and feature selection leads to decorrelated trees, reducing variance and improving generalization.

4. RESULT AND DISCUSSION

4.1 Software Used and Programming Used: Python is chosen as the programming language because it is widely used for data science and machine learning applications and is versatile. There are several integrated development environments (IDEs) available for developing Python programs, each with its own set of features designed to improve the coding experience and efficiency. Python programming has several integrated development environments (IDEs) available, each with its own unique features that can make programming more efficient and user-friendly.

4.2 Results Analysis

The dataset consists of 2200 samples of 22 crops and has seven features that are of importance in the agriculture field: NPK, Temperature, Humidity, pH, and Rainfall.

Table 1. Descriptive Statistical Analysis of Agricultural Features Dataset

Statistic	N	P	K	Temperature	Humidity	pH	Rainfall
count	2200	2200	2200	2200	2200	2200	2200
mean	50.55	53.36	48.15	25.62	71.48	6.47	103.46
std	36.92	32.99	50.65	5.06	22.26	0.77	54.96
min	0	5	5	8.83	14.26	3.50	20.21
25%	21	28	20	22.77	60.26	5.97	64.55
50%	37	51	32	25.60	80.47	6.43	94.87
75%	84.25	68	49	28.56	89.95	6.92	124.27
max	140	145	205	43.68	99.98	9.94	298.56

Summary of the Descriptive Statistics

Macronutrients (N, P, K):

- The mean value of nitrogen (N) was 50.55 with a standard deviation of 36.92 and the range was from 0 to 140, showing large variation in the amount of nitrogen in soils.
- The range of phosphorus (P) was 5-145, with an average of 53.36, and a standard deviation of 32.99, indicating that there is a wide distribution of phosphorus levels.
- Potassium (K) has a range of 5 to 205, a mean of 48.15 and highest standard deviation of 50.65 of the three nutrients, indicating that there is a large variation in potassium concentration.
- The high standard deviations for N, P and K in the data overall suggest that a diverse range of soil nutrient contents are present, thereby enabling the development of strong crop recommendation models.
- **Environmental Conditions:**
- Temperature ranges from 8.83°C to 43.68°C with a mean of 25.62°C and a standard deviation of 5.06°C, indicating climatic conditions from relatively cool to hot agricultural regions.
- The humidity ranges from 14.26% to 99.98% with a mean of 71.48% suggests a wide spectrum of humidity conditions.
- Soils are very acidic to alkaline (pH 3.50 to 9.94) with a mean of 6.47, considered good for most crops.
- The range of the rainfall is wide from 20.21 mm to 298.56 mm and the mean is 103.46 mm with a standard deviation of 54.96 mm.

Distribution Characteristics:

- The data set is composed of 2200 observations for each feature, which is more than is needed for machine learning analysis.
- Internally, the interquartile ranges (25th–75th percentile) show that the middle 50% of observations are well spread, and therefore do not have a great impact.
- Overall, the richness of the nutrient concentrations and environmental conditions contributes to the diversity of the data, which is well-suited for the creation and assessment of machine learning models designed to recommend crops and for precision agriculture.

Table 2: Training classification report for Random Forest Classifier across various crops

Class	Precision	Recall	F1-Score	Support
Apple	1	1	1	82
Banana	1	1	1	82
Blackgram	0.97	1	0.99	78
Chickpea	1	1	1	77
Coconut	1	1	1	85
Coffee	1	1	1	83
Cotton	1	1	1	84
Grapes	1	1	1	82
Jute	0.86	0.95	0.9	79
Kidney beans	1	1	1	80
Lentil	0.96	0.98	0.97	83
Maize	1	1	1	82
Mango	1	1	1	79
Moth beans	1	0.96	0.98	75
Mung bean	1	1	1	83
Muskmelon	1	1	1	77
Orange	1	1	1	77
Papaya	0.99	1	0.99	79
Pigeonpeas	1	1	1	78
Pomegranate	1	1	1	77
Rice	0.95	0.84	0.89	75
Watermelon	1	1	1	83
accuracy		0.99	0.99	1760
macro avg	0.99	0.99	0.99	1760
weighted avg	0.99	0.99	0.99	1760

Table 3. Testing classification report for Random Forest Classifier across various crops

Class	Precision	Recall	F1-Score	Support
Apple	1	1	1	18
Banana	1	1	1	18
Blackgram	1	1	1	22
Chickpea	1	1	1	23
Coconut	1	1	1	15
Coffee	1	1	1	17
Cotton	1	1	1	16
Grapes	1	1	1	18
Jute	0.91	0.95	0.93	21
Kidney beans	0.95	1	0.98	20
Lentil	0.94	0.94	0.94	17
Maize	1	1	1	18
Mango	1	1	1	21
Moth beans	0.96	0.96	0.96	25
Mung bean	1	1	1	17
Muskmelon	1	1	1	23
Orange	1	1	1	23
Papaya	1	1	1	21
Pigeonpeas	1	0.95	0.98	22
Pomegranate	1	1	1	23
Rice	0.96	0.92	0.94	25
Watermelon	1	1	1	17
accuracy			0.99	440
macro avg	0.99	0.99	0.99	440
weighted avg	0.99	0.99	0.99	440

Rainfall is highly variable, varying from 20.2 mm to 298.6 mm, the average being ~103.47 mm. The data set was split into training data set and testing data set using 80:20 ratio. This means that 80% of the data is assigned to train the machine learning (ML) models and 20% is reserved to assess the model's accuracy on unseen data, ensuring the model is consistently evaluated against unseen data for a consistent assessment of model generalization performance.

```
# Manually input the 7 input features for testing
N = float(input("Enter N: "))
P = float(input("Enter P: "))
K = float(input("Enter K: "))
Temperature = float(input("Enter Temperature: "))
Humidity = float(input("Enter Humidity: "))
pH = float(input("Enter pH: "))
Rainfall = float(input("Enter Rainfall: "))
```

```
Enter N: 45
Enter P: 34
Enter K: 40
Enter Temperature: 35
Enter Humidity: 60
Enter pH: 7
Enter Rainfall: 80
```

```
new_data = sc.transform([[N, P, K, Temperature, Humidity, pH, Rainfall]])
crop_label = classifier.predict(new_data)
print("Predicted crop label: ", crop_label)
```

```
Predicted crop label:  ['mango']
```

Figure 3. Crop recommended by Random Forest Classifier ML model

The recommendation of the crop by the Random Forest Classifier with all the seven soil parameters and environmental parameters N, P, K, pH, Rainfall, Temperature, and Humidity are shown in the figure 3.

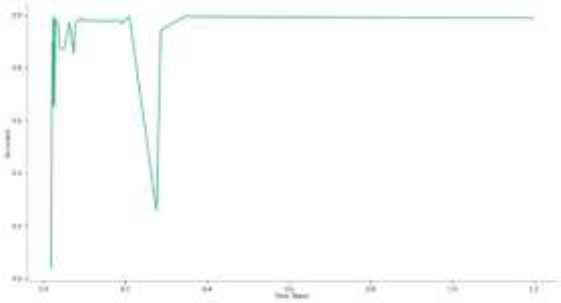


Figure 4. Time series graph for accuracy and time taken

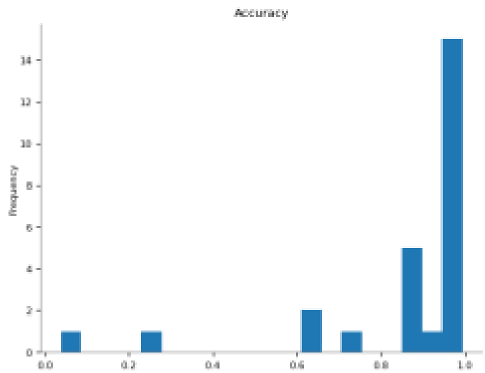


Figure 5. Distribution of Model Accuracy Scores

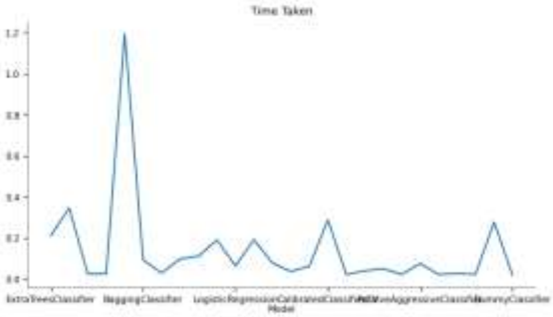


Figure 6. Values: Time taken (in sec) by implemented ML models taken

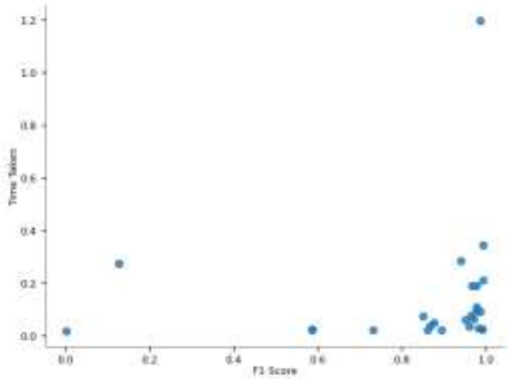


Figure 7. 2D Distribution of F1 score vs time

4.2 Statistical Analysis of Model Performance

A descriptive statistical analysis was conducted using the following evaluation metrics: Accuracy, Balanced Accuracy, F1-Score, Precision, Recall and Execution Time to comprehensively assess the effectiveness of the machine learning algorithms implemented. The statistical summary contains important statistics such as the mean, minimum, maximum, quartiles, median and distribution of execution time, which give an idea of the predictive ability and the computational efficiency of the classifiers tested.

Table 4. Performances of various implemented ML modles

Model	Accuracy (%)	Balanced Accuracy (%)	F1 Score (%)	Precision (%)	Recall (%)	Time Taken
Extra Trees Classifier	99.36	99.31	99.36	99.41	99.36	0.3904
Random Forest Classifier	99.36	99.29	99.36	99.44	99.36	0.4230
Gaussian NB	99.18	99.14	99.18	99.21	99.18	0.0418
Quadratic Discriminant Analysis	99.09	99.04	99.09	99.11	99.09	0.0651
LGBM Classifier	98.73	98.66	98.72	98.79	98.73	1.8398
Bagging Classifier	98.73	98.64	98.72	98.85	98.73	0.1957
Decision Tree Classifier	98.27	98.27	98.27	98.3	98.27	0.0734
SVC	97.82	97.69	97.8	97.93	97.82	0.2951
Label Spreading	97.73	97.66	97.72	97.88	97.73	0.2913
Label Propagation	97.73	97.66	97.72	97.88	97.73	0.2059
Logistic Regression	97.09	96.98	97.09	97.13	97.09	0.1128
Nu SVC	96.73	96.59	96.7	96.93	96.73	0.3181
K Neighbors Classifier	96.45	96.36	96.43	96.71	96.45	0.0595

Linear Discriminant Analysis	96	95.78	95.96	96.33	96	0.0673
Linear SVC	95.09	95.25	95.01	95.56	95.09	0.0660
Calibrated Classifier CV	94.18	94.37	94.09	95.07	94.18	1.3636
Extra Tree Classifier	89.45	89.28	89.47	89.67	89.45	0.0617
Perceptron	87.09	87.54	86.67	89.51	87.09	0.0540
SGD Classifier	87	87.13	87.73	91.5	87	0.0671
Nearest Centroid	86.64	86.44	86.21	87.29	86.64	0.0692
Passive Aggressive Classifier	85.27	85.58	84.92	87.65	85.27	0.0585
Bernoulli NB	75.09	75.73	73.17	77.74	75.09	0.0671
Ridge Classifier	65.27	66.7	58.71	69.64	65.27	0.0413
Ridge Classifier CV	65.27	66.7	58.67	69.47	65.27	0.0380
AdaBoost Classifier	25.64	26.69	12.76	9.04	25.64	0.6592
Dummy Classifier	3.82	4.55	0.28	0.15	3.82	0.0505

Table 5. Statistical Summary Model Performance

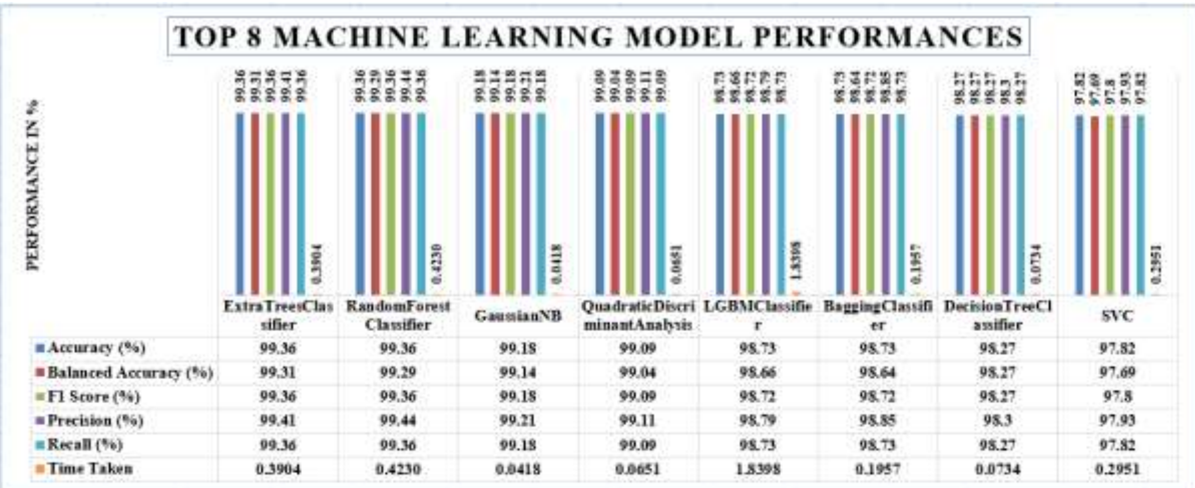
Statistic	Accuracy (%)	Balanced Accuracy (%)	F1 Score (%)	Precision (%)	Recall (%)	Time (seconds)
Mean	85.85	85.81	84.61	86.64	85.85	0.15
Minimum	3.82	4.55	0.28	0.15	3.82	0.016
Maximum	99.36	99.31	99.36	99.44	99.36	1.48
25th Percentile	86.73	86.51	86.33	87.47	86.73	0.024
Median	96.23	96.09	96.2	96.52	96.23	0.057
75th Percentile	98.16	98.13	98.15	98.47	98.16	0.162

The following statistical observations can be drawn from the results:

- The average values of Accuracy (across 26 classifiers), Balanced Accuracy, F1-Score, Precision and Recall were 85.85%, 85.81%, 84.61%, 86.64% and 85.85%, respectively. These average values are high showing that most of the models implemented are performing well and are consistent in classification performance.
- The highest Accuracy, Balanced accuracy, F1-Score and Recall were obtained as 99.36%, 99.31%, 99.36% and 99.36% respectively, whereas the highest Precision was 99.44% obtained by the Random Forest Classifier. The results show the outstanding prediction performance of ensemble based learning algorithms.
- The baseline model Dummy Classifier yielded the minimum values for Accuracy (3.82%), Balanced Accuracy (4.55%), F1-Score (0.28%), Precision (0.15%) and Recall (3.82%). This proves that any algorithm that is worth considering for machine learning has a significant advantage over random prediction.
- The first quartile (25th percentile) values are 86.73% (Accuracy), 86.51% (Balanced Accuracy), 86.33% (F1-Score), 87.47% (Precision) and 86.73% (Recall) which show that about 75% of studied classifiers had performance above 86% on all evaluation metrics.
- At least half of the evaluated classifiers showed high levels of predictive performance in both the median values of Accuracy (96.23%), Balanced Accuracy (96.09%), F1-Score (96.20%), Precision (96.52%) and Recall (96.23%).
- The third quartile (75th percentile) values were 98.16% (Accuracy), 98.13% (Balanced Accuracy), 98.15% (F1-Score), 98.47% (Precision), and 98.16% (Recall). The above results show that the most successful 25% of the classifiers were able to achieve a classification performance greater than 98% in every instance, further demonstrating the success of ensemble and tree based learning methods.
- Around 15 out of 26 classifiers (≈58%) performed better than 95% in terms of Accuracy, followed by comparable classifiers that exceeded 95% in terms of Balanced Accuracy, Precision, Recall and F1-Score,

which means that the predictions of more than half of the evaluated classifiers are highly reliable for crop recommendation.

- The high closeness of all the values of Accuracy, Balanced Accuracy, Precision, Recall, and F1-Score indicates that the classifiers have a balanced performance, so that they are not biased to any of the crop classes. Moreover, due to the minimal difference between Accuracy and Balanced Accuracy, it is concluded that the dataset is well balanced and does not have significant class imbalance.
- The results showed that the average time taken for the classifiers was around 0.15 seconds, suggesting that most of the classifiers are capable of making quick predictions as required for precision agriculture in real-time.
- The different classifiers exhibited the following times for prediction: the fastest took about 0.016 seconds; the slowest took 1.48 seconds, showing a computational trade-off between the different learning algorithms. Even the slowest model, however, was still computationally tractable and practical for application in practice.
- The median execution time was 0.057 seconds, meaning that half of all the classifiers tested in the paper had execution times under 60 milliseconds. Likewise, the 25th percentile (0.024 s) and 75th percentile (0.162 s) show that most classifiers have a very low computation overhead.



	ExtraTreesClassifier	RandomForestClassifier	GaussianNB	QuadraticDiscriminantAnalysis	LGBMClassifier	BaggingClassifier	DecisionTreeClassifier	SVC
Accuracy (%)	99.36	99.36	99.18	99.09	98.73	98.73	98.27	97.82
Balanced Accuracy (%)	99.31	99.29	99.14	99.04	98.66	98.64	98.27	97.69
F1 Score (%)	99.36	99.36	99.18	99.09	98.72	98.72	98.27	97.8
Precision (%)	99.41	99.44	99.21	99.11	98.79	98.85	98.3	97.93
Recall (%)	99.36	99.36	99.18	99.09	98.73	98.73	98.27	97.82
Time Taken	0.3904	0.4230	0.0418	0.0651	1.8398	0.1957	0.0734	0.2951

Figure 8. Comparative Performance of the Top 8 Machine Learning Models

5. CONCLUSION

This study proposed a machine learning based Crop Recommendation System (CRS) for precision agriculture for a balanced dataset of 2,200 samples, 22 crop classes and 7 Agronomic features (Nitrogen, Phosphorus, Potassium, temperature, humidity, pH, and rainfall). 26 machine learning classifiers were tested to determine which is the most suitable for crop recommendation. The results showed that ensemble learning algorithms performed best over all the other classifiers. The Extra Trees Classifier and Random Forest Classifier gave the highest accuracy of 99.36%, which resulted in a high value for Balanced Accuracy, F1-Score, Precision and Recall of 99.31%, 99.36%, 99.41% and 99.44% respectively, showing that the predictions were highly reliable and balanced. The Extra Trees Classifier also needed a lesser computing time, which made it the most appropriate model for real time agricultural application. Based on statistical analysis, 15 out of 26 models had an accuracy above 95%, and the good generalization for all crop classes was confirmed by the close fitting of Accuracy, Balanced Accuracy, Precision, Recall and F1-Score results. Furthermore, the execution times are very small, thus showing the practical viability of the proposed framework in real smart farming settings. Implementing the proposed CRS can help farmers choose the right crop considering soil nutrients and environmental conditions, which helps them to boost the crop productivity and use resources optimally and accordingly in sustainable agriculture. The system is a bit helpful for the Indian agriculture system, in which agro-climatic conditions are diverse and land holdings are fragmented which require intelligent decision support system.

Future work will focus on:

- Integration with real-time IoT-based sensors for dynamic data collection,
- Exploration of deep learning architectures (e.g., LSTM, CNN, and attention-based models) for improved adaptability,
- Development of mobile or web-based deployment platforms for farmer accessibility,

- Region-specific model tuning using local agricultural datasets and weather APIs.
- By bridging the gap between machine learning research and practical agricultural challenges, this study contributes a scalable, intelligent solution for enhancing crop planning and resource optimization in smart farming.

6. REFERENCES

1. Zaborowicz M, Frankowski J. Big Data Analytics and Machine Learning for Smart Agriculture. Agriculture. 2025 Mar 31;15(7):757. 10.3390/agriculture15070757
2. Menaga A, Vasantha S. Smart Sustainable Agriculture Using Machine Learning and AI: A Review. In: Lecture Notes in Networks and Systems. Springer Science and Business Media Deutschland GmbH; 2022. p. 447–58. 10.1007/978-981-16-7952-0_42
3. Shaikh TA, Mir WA, Rasool T, Sofi S. Machine Learning for Smart Agriculture and Precision Farming: Towards Making the Fields Talk. Vol. 29, Archives of Computational Methods in Engineering. Springer Science and Business Media B.V.; 2022. p. 4557–97. 10.1007/s11831-022-09761-4
4. Durai SKS, Shamili MD. Smart farming using Machine Learning and Deep Learning techniques. Decision Analytics Journal. 2022 Jun 1;3. 10.1016/j.dajour.2022.100041
5. Sharma A, Jain A, Gupta P, Chowdary V. Machine Learning Applications for Precision Agriculture: A Comprehensive Review. Vol. 9, IEEE Access. Institute of Electrical and Electronics Engineers Inc.; 2021. p. 4843–73. 10.1109/ACCESS.2020.3048415
6. Ayaz M, Ammad-Uddin M, Sharif Z, Mansour A, Aggoune EHM. Internet-of-Things (IoT)-based smart agriculture: Toward making the fields talk. IEEE Access. 2019; 7:129551–83. 10.1109/ACCESS.2019.2932609
7. Akhter R, Sofi SA. Precision agriculture using IoT data analytics and machine learning. Vol. 34, Journal of King Saud University - Computer and Information Sciences. King Saud bin Abdulaziz University; 2022. p. 5602–18. 10.1016/j.jksuci.2021.05.013
8. Qazi S, Khawaja BA, Farooq QU. IoT-Equipped and AI-Enabled Next Generation Smart Agriculture: A Critical Review, Current Challenges and Future Trends. Vol. 10, IEEE Access. Institute of Electrical and Electronics Engineers Inc.; 2022. p. 21219–35. 10.1109/ACCESS.2022.3152544
9. Rosero-Montalvo PD, Gordillo-Gordillo CA, Hernandez W. Smart Farming Robot for Detecting Environmental Conditions in a Greenhouse. IEEE Access. 2023; 11:57843–53. 10.1109/ACCESS.2023.3283986
10. Mohyuddin G, Khan MA, Haseeb A, Mahpara S, Waseem M, Saleh AM. Evaluation of Machine Learning Approaches for Precision Farming in Smart Agriculture System: A Comprehensive Review. IEEE Access. 2024; 12:60155–84. 10.1109/ACCESS.2024.3390581
11. Syed L. Smart Agriculture using Ensemble Machine Learning Techniques in IoT Environment. In: Procedia Computer Science. Elsevier B.V.; 2024. p. 2269–78. 10.1016/j.procs.2024.04.215
12. Shams MY, Gamel SA, Talaat FM. Enhancing crop recommendation systems with explainable artificial intelligence: a study on agricultural decision-making. Neural Comput Appl. 2024 Apr 1;36(11):5695–714. 10.1007/s00521-023-09391-2
13. Garg D, Alam M. An effective crop recommendation method using machine learning techniques. International Journal of Advanced Technology and Engineering Exploration. 2023 May 1;10(102):498–514. 10.19101/ijatee.2022.10100456
14. Asghari Beirami B, Mokhtarzade M. Land Covers Classification from LiDAR-DSM Data Based on Local Kernel Matrix Features of Morphological Profiles. International Journal of Engineering Transactions C: Aspects. 2023 Sep 1;36(9):1611–7. 10.5829/ije.2023.36.09c.04
15. Gutiérrez A, Bakthavatchalam K, Karthik B, Thiruvengadam V, Muthal S, Jose D, et al. IoT Framework for Measurement and Precision Agriculture: Predicting the Crop Using Machine Learning Algorithms. 2022. 10.3390/technologies10010013
16. Raghuvanshi A, Singh UK, Sajja GS, Pallathadka H, Asenso E, Kamal M et al. Intrusion Detection Using Machine Learning for Risk Mitigation in IoT-Enabled Smart Irrigation in Smart Farming. Journal of Food Quality. 2022;2022: 3955514. doi: 10.1155/2022/3955514
17. Khan AA, Faheem M, Bashir RN, Wechtaisong C, Abbas MZ. Internet of Things (IoT) Assisted Context Aware Fertilizer Recommendation. IEEE Access. 2022; 10:129505–19. 10.1109/ACCESS.2022.3228160
18. Perales Gómez ÁL, López-de-Teruel PE, Ruiz A, García-Mateos G, Bernabé García G, García Clemente FJ. FARMIT: continuous assessment of crop quality using machine learning and deep learning techniques for IoT-based smart farming. Cluster Comput. 2022 Jun 1;25(3):2163–78. 10.1007/s10586-021-03489-9
19. Venkatesan S, Lim J, Ko H, Cho Y. A Machine Learning Based Model for Energy Usage Peak Prediction in Smart Farms. Electronics (Switzerland). 2022 Jan 1;11(2). 10.3390/electronics11020218
20. Shah SK, Kumbhar V, Singh TP. Grape (Vitis Vinifera) Leaf Disease Detection and Classification Using Deep Learning Techniques: A Study on Real-Time Grape Leaf Image Dataset in India. International Journal of Engineering, Transactions B: Applications. 2024;37(8):1522–33 10.5829/ije.2024.37.08b.06

21. Rashid M, Bari BS, Yusup Y, Kamaruddin MA, Khan N. A Comprehensive Review of Crop Yield Prediction Using Machine Learning Approaches with Special Emphasis on Palm Oil Yield Prediction. Vol. 9, IEEE Access. Institute of Electrical and Electronics Engineers Inc.; 2021. p. 63406–39. 10.1109/ACCESS.2021.3075159
22. Talaviya T, Shah D, Patel N, Yagnik H, Shah M. Implementation of artificial intelligence in agriculture for optimisation of irrigation and application of pesticides and herbicides. Vol. 4, Artificial Intelligence in Agriculture. KeAi Communications Co.; 2020. p. 58–73. 10.1016/j.aiia.2020.04.002
23. Pathan M, Patel N, Yagnik H, Shah M. Artificial cognition for applications in smart agriculture: A comprehensive review. Vol. 4, Artificial Intelligence in Agriculture. KeAi Communications Co.; 2020. p. 81–95. 10.1016/j.aiia.2020.06.001
24. Li D, Liu Y, Chen Y, Zhang H, Zhang L, Ren Y, et al. Design and Implementation of Crop Recommendation Fertilization Decision System Based on WEBGIS at Village Scale. Vol. 345, IFIP AICT. 2011. 10.1007/978-3-642-18336-2_44
25. Nandal V, Dahiya S. An energy-efficient information aggregation protocol with optimized trilevel K-means clustering for IoT-based WSN framework: A case study on smart agriculture. *Int J Commun Syst*. 2025; 38:e70020. doi:10.1002/dac.70020.
26. Nandal V, Dahiya S. A comprehensive review of energy efficient wireless communication and routing protocols in smart agriculture. *Int J Sensor Netw*. 2023;43(2):99-115. doi:10.1504/IJSNET.2023.134308.