IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 7, July 2025

DOI: 10.17148/IJIREEICE.2025.13709

Development of a Machine Learning-Based Crop Recommendation System for Sustainable Agriculture

Aarya Dhaygude[1], Aryan Karve[2], Prof. Dipali Pawar[3]

Dept. of Electronics and Computer Engineering, PES'S Modern College of Engineering, Pune, India[1]

Dept. of Electronics and Telecommunication Engineering, PES'S Modern College of Engineering, Pune, India[2]

Dept. of Electronics and Computer Engineering, PES'S Modern College of Engineering, Pune, India[3]

Abstract: The agricultural sector has become data-driven in recent years to increase agricultural yields, improve resource use, and promote sustainable agriculture. This essay describes an overview of how a Crop Recommendation System using Machine Learning was developed to assist farmers in making the optimal crop choice in relation to climatic and soil conditions. The model draws upon an open-source database of soil nutrient concentrations (nitrogen, phosphorus, potassium), temperature, humidity, pH, and rain to predict the best crop for the input conditions. Preprocessing methods were used to clean and normalize the data so that the model could be trained to be reliable. Various machine learning algorithms, such as Decision Trees, Random Forests, Support Vector Machines, and K-nearest neighbors, were trained and cross-validated to identify the most appropriate model. With heavy training, testing, and tuning for hyperparameters, the top-performing model was a Random Forest classifier, with impressive performance in accuracy, precision, recall, and F1-score measures. The system suggests crops in real-time, enabling farmers to make decisions based on weather and soil type. This increases its efficiency and decreases the overuse of fertilizers and water, thereby saving the environment. The study illustrates the potential of machine learning in transforming traditional agriculture. In order to develop a holistic system, the future can witness additional variables such as market forces, pest presence, and satellite imagery. The method discussed here is essential to realizing artificial intelligence-based sustainable agriculture.

Keywords: Crop Recommendation System, Machine Learning, Random Forest Classifier, Precision Agriculture, Sustainable Farming

I. INTRODUCTION

Agriculture has been the backbone of human civilization, providing food security, raw materials, and employment to most people worldwide. While technology has always spurred technological advancement in most sectors, agriculture in much of the world still develops based on traditional practices, experience, and intuition rather than data-driven decision-making. This typically leads to suboptimal crop selection, resource wastage, and environmental degradation. In an era of global warming, soil erosion, and a growing food demand due to an increasingly large population, the urgency for new and creative solutions for taking farming to contemporary times has never been stronger. Some fascinating emerging technologies include Machine Learning (ML) and Artificial Intelligence (AI), which have been proven to be revolutionary methods to revolutionize agricultural activities, with applications in some areas, such as the suggestion of crops.

Crop choice is a core decision that farmers take, with significant implications for yield, use of resources, and returns. Crop choice has traditionally been based on periodic seasons, conventional wisdom, and experience. With uncertainty from climate variability, soil erosion, water scarcity, and markets, experience alone will not be enough. Farmers are most likely to struggle to select an appropriate crop for soil and climatic factors, which results in poor yields, increased input, and losses of finances. An intelligent, rational, evidence-based process with several determinants will immensely improve decision-making and, thus, agricultural productivity and sustainability.

Machine Learning, a branch of Artificial Intelligence, provides strong algorithms that can learn patterns from past data and predict accurately on unseen data. Using different soil parameters like nitrogen (N), phosphorus (P), potassium (K) content, pH, temperature, humidity, and rainfall trends, ML models can suggest crops best suited to a given set of conditions. In contrast to conventional extension services, which could be inaccessible to all farmers, ML-based systems can be replicated and implemented extensively, making expert advice available to even farmers in far-flung areas. In addition, ML systems can learn and update themselves continuously with additional data input, improving their predictive performance.

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 7, July 2025

DOI: 10.17148/IJIREEICE.2025.13709

Past research has established that machine learning algorithms such as Decision Trees, Random Forests, Support Vector Machines, and K-nearest neighbors suit agricultural prediction tasks. They can analyze complex, non-linear relationships between more than one variable, which would be difficult for conventional statistical models to account for. Random Forests, in particular, have been identified to be strongly resistant to overfitting and possess high classification accuracy, rendering them suitable for crop recommendation systems. Moreover, the advancement of computational power and cloud-based systems has made it possible to deal with large agricultural datasets and produce real-time predictions at relatively lower costs.

Machine learning in agriculture is also part of the overall vision of "Precision Agriculture," which focuses on optimizing field-level management for crop farming. Precision agriculture guarantees crops and soil only what they require for their best health and productivity. Crop recommendation systems based on ML naturally belong to this framework by giving accurate suggestions customized according to the respective soil and environmental conditions. Aside from enhancing productivity, such systems also contribute to sustaining the environment as they minimize excess fertilizer usage, lower water utilization, and neutralize soil degradation due to agriculture runoff.

The present task includes deploying a crop suggestion system based on machine learning using a publicly released dataset of needed weather and soil conditions. The primary functions consist of data preprocessing, comparison and analysis of various machine learning models based on conventional evaluation criteria, and identifying the top-performing model for deployment in real-time. The dataset includes significant variables like nitrogen, phosphorus, potassium levels, temperature, humidity, pH, and rainfall, with every record labeled with its most appropriate crop. The richness and diversity of the dataset allow us to develop strong models that can be generalized in various agricultural conditions.

The research methodology employed in this study consists of a set of interdependent steps, starting with data exploration and preprocessing. At the data cleaning stage, methods like missing value handling, outlier removal, and normalization are applied to prepare the dataset feature with uniformly scaled features for efficient crop recommendation. Exploratory Data Analysis (EDA) is performed to analyze relationships between varied parameters and how they affect crop selection. Several machine learning models like Decision Trees, Random Forests, Support Vector Machines, and K-nearest neighbors are implemented and compared based on their accuracy in classification. Hyperparameter optimization techniques like Grid Search and Random Search improve the model's performance. The performance of each model is assessed using metrics such as accuracy, precision, recall, F1-score, and analysis via a confusion matrix.

Out of all the models compared, the Random Forest classifier was the most accurate and provided a superb generalization to new unseen test instances. Random Forest is an ensemble machine learning model that builds numerous decision trees while training and uses them to cast a majority vote for the ultimate prediction based on the projections of trees. Its ability to process diverse input features and strong overfitting resistance make it highly suitable for application in crop recommendation systems. Furthermore, cross-validation techniques were used to stabilize the model and to achieve consistent performance on unseen data.

Apart from model development, the research is also concerned with field deployment. The system can be deployed for mobile applications, web applications, or embedded systems, making it easier for farmers to get crop advice. With the increasing smartphone and internet penetration in rural areas, digital crop advice platforms are becoming more efficient and practical. Also, the system can suggest customized fertilizers, irrigation timing, and pest control measures, which would be helpful again.

While the results are encouraging, the current study's limitations are determined. Data employed, though complete, is not likely to cover all the potential real-world conditions, like sudden weather changes, infestation, or market dynamics. Present weather, remote sensing, and socio-economic data, by being incorporated into it, can generalize the model further. Advanced deep-learning models and ensemble strategies could also be among the future research areas for improved prediction.

Optimize: In addition, joint work with farming specialists can assist in verifying and refining the advice to make it realistically feasible.

The social utility of a machine learning-based crop advisory system is high. These systems can improve income, minimize environmental degradation, and enhance food security by bringing in-time and quality information to farmers. These systems can also act as change agents in low-productivity regions of poor soil management and cropping systems. In addition, encouraging data-enabled agriculture is also in line with the United Nations Sustainable Development Goals (SDGs) of eradicating hunger, ensuring food security, promoting sustainable agriculture, and climate action.

DOI: 10.17148/IJIREEICE.2025.13709

Machine learning integration with crop recommendation is an innovation in present-day agriculture. Machine learning can generate valuable knowledge based on previous soil and climatic data and thus make recommendations for better crop selection, optimum use of resources, and optimization of farms. This study is part of upgraded research in agricultural AI applications and the development a strong, efficient, and cost-effective crop recommendation system. Since agriculture is subjected to the dual challenge of needing to feed more people and adapting to a fluctuating climate, technology innovation shines a beacon to a cleaner and brighter future for farmers everywhere.

II. LITERATURE SURVEY

Dr. V. Geetha et al. [1] suggested a study in which the Random Forest algorithm, a machine learning method, was used to predict crop growth under varied climatic conditions such as drought seasons, rising temperatures, and biophysical changes. Crop growth data were gathered from various sources to facilitate precise prediction. The data were utilized for training and testing the models, and the Random Forest algorithm was used for crop prediction. The findings revealed that the Random Forest method predicted better than other methods in terms of accuracy.

Zeel Doshi et al. [2] developed an intelligent Agro-consultant system to help Indian farmers make prompt and suitable crop cultivation decisions. The system considers various parameters such as the season for sowing, geographical conditions, soil characteristics, and climatic parameters like temperature and rain. It consists of two interacting subsystems: one is busy providing farmers with suitable crops for suggestion, and the other provides forecasts of rainfall for an area, with the forecasted information serving as an input for enhancing the prediction of crop sustainability in the first subsystem. Priyadharshini et al. [3] proposed a system to assist farmers in selecting the best possible crops by providing information otherwise difficult to access. This technique helps mitigate crop failure risk and raise overall production while preventing farmers from financial losses. The system is based on real-time monthly weather conditions for accurate yield forecasting and conducts profit analysis based on historical crop details of the previous year. A three-input and fifteen-output sequential model was employed. A Linear Regression algorithm and a Neural Network were employed as the basis of the system, developed using libraries and tools such as Pandas, NumPy, TensorFlow, Keras, and Scikit-learn, with Python as the programming framework. The accuracy in recommending crops achieved was 88.26% for the Linear Regression model and 89.88% for the Neural Network, indicating high reliability and efficiency. Similarly, Vaishnavi et al. [4] proposed a system that provides real-time accurate crop predictions to farmers depending on production and season parameters. Their system uses data mining technology to make precise predictions and offers personalized recommendations, thus enhancing crop production and farm productivity.

Dr. Y. Jeevan Nagendra Kumar et al. [5] applied a supervised learning technique to predict crop yield in agriculture. Past data was analyzed in their work to forecast the future yields of crops, with a specific focus on weather and the application of pesticides as key control factors. The data set, which was collected from Kaggle, included parameters such as rainfall, precipitation, temperature, and production, amounting to 3101 historical samples. Two-thirds of such information was utilized for model training. The Random Forest algorithm was also tested using its most significant parameters, i.e., number of trees (n tree), features to examine at every split (m try), and node size. Random Forest proved more accurate than other algorithms and produced the most precise crop yield forecasts. Being more accurate translates to increased profits for farmers through better yield predictions. The system also encompasses a broad range of crops under consideration; hence, it benefits farmers in India. Similarly, Ms. Kavita et al. [6] built a model that predicted the crop yield in India from 1950 to 2018 for five principal crops: Rice, Wheat, Jowar, Bajra, Tobacco, and Maize.

The dataset contains 745 instances, which are divided in such a manner that 70% is to be used in training and the remaining 30% for testing. The dataset contains attributes like rainfall, cropland, irrigated area, crop type, production season, production stage, and yield records from 1950 to 2018. The study uses various models, such as Decision Tree, Linear Regression, Lasso Regression, and Ridge Regression. Among them, the Decision Tree model yielded an % accuracy rate of 98.62%. This system is helpful for small farmers because accurate crop yield prediction assists in better planning and agricultural output.

.Meeradevi et al.[7] has submitted a paper in which they have developed an application through which the farmer can forecast the yield of particular crops in a specific area. The crop production is forecasted based on physical parameters like rainfall and temperature. This model predicts the crop based on different datasets of different crops in various parts of India and rainfall and temperature datasets in the same regions. A modified ARIMA model identifies the predicted rainfall and temperature values. These datasets are used with the linear regression model, which is accountable for further prediction. The suggested system is primarily an application that operates based on the individuality of the crops. The system provides farmers with a precise list of recommendations to improve their crop choice according to different parameters such as location, farm size, temperature, rainfall, and other crop databases. Sonal Agarwal et al. [8] have given a paper in which a new model has been developed using Machine Learning (SVM) and Deep Learning (LSTM, RNN) methods.

DOI: 10.17148/IJIREEICE.2025.13709

Following the analysis of prediction parameters, the model was constructed to accurately forecast crops that need to be cultivated on land with fewer costs from several available crops. It uses soil and temperature parameters to predict crops. The accuracy attained by the model proposed is 97%. It performs an analysis of the provided data and assists the farmer in forecasting the crop, which will help them enhance their profit in return.

N. Manjunathan et al. [9] have introduced a paper where they develop a machine-learning model that can precisely forecast the yield of rice crops. Since the production of crops relies on numerous such factors, tracking them for fruitful crop yield is a must. So, we suggest a model based on a Support Vector Machine (SVM) algorithm to predict crop yield precisely. SVM is employed to classify the crops according to the area and season factors. They have also developed a web application that allows users to engage with the trained ML model and make predictions using the input they provide. The system employs the Weka tool to develop machine learning algorithms and HTML, CSS, and JavaScript to create the web application. The accuracy it achieved was 96%. Shubham Prabhu et al. [10] have submitted a paper introducing the soil analysis and crop forecasting model.

This paper aims to build a prediction tool for determining the best crop suitable for the input-given soil. As a preliminary task, the authors endeavored to forecast the proper crop yield by estimating the soil fertility and rain of the place mentioned as the user's input. The author took into consideration five samples of soil from various locations, and then the analysis was conducted based on temperature, humidity, and moisture, which is performed at a fixed interval of 24 hours, and the data is uploaded, displayed, and updated at 2 hours. All the data analyzed is monitored continuously, displayed, and uploaded to the IoT cloud. They have used three algorithms: Naïve Bayes, Logistics, and C4.5, in which C4.7 has a maximum accuracy of 85%.

Kusum Lata et al. [11] suggested using four algorithms for classification via the WEKA tool to enhance agricultural data analysis efficiency and effectiveness. It brings to focus the urgency for upgrading vast volumes of agricultural datasets and putting them in the hands of farmers to assist them in more effective decision-making and increase yields. The study aims to compare and determine the best among the Random Tree, J48, Bayes Net, and K-Star algorithms. Findings indicated that the Random Tree classifier was good, with a lower error rate, and performed better than the other three algorithms—K-Star, Bayes Net, and J48. The paper's primary aim is to use Machine Learning methods for valuable insights generation from agricultural information to improve crop yield predictions for main crops grown in Maharashtra's Nashik district.
D. Jayanarayana Reddy et al. [12] proposed a crop yield forecasting system using Machine Learning methodologies, stressing the valuable contribution of ML towards increasing agricultural yield. The final goal of crop yield prediction (CYP) is to maximize crop yield. The process has several steps: initially, farm data is collected and processed using preprocessing. Following this, vital attributes regarding field, soil management, nutrients, and moisture are intensified and prepared for analysis. Post-processing data is used to train and evaluate Machine Learning algorithms, including CNN, LSTM, ANN, KNN, and DNN algorithms. Trained algorithms heavily reduced relative prediction error. The paper also delivers a comparative evaluation of different Machine Learning techniques through intensive prediction accuracy evaluation.

Saeed Khaki et al. [13] employed two Deep Neural Networks, one for yield and the other for checking yield, and used their difference in outputs as the output for yield difference. The utilized models are DNN (G), DNN(S), and DNN (W). This model framework was superior to using a single neural network to produce differences since the environment and genotype impacts are more naturally linked with the yield, and their difference is confirmed. Hyperparameters were employed in training to increase the model's computation. To achieve optimal prediction accuracy while minimizing the possibility of overfitting, researchers used neural networks with 21 hidden layers and 50 neurons per layer, and ultimately, this configuration provided the optimal balance between these two variables. Even after experimenting with deeper network architectures, the researchers found that this particular layer and neuron size combination always produced the optimal result.

Mrs. R. Usha Devi et al. [14] proposed a Machine Learning-based recommendation system to suggest suitable crops based on soil qualities to solve agriculture problems and improve crop yields. By monitoring agricultural fields and suggesting optimal crops for farmers to cultivate based on input soil conditions, this study aims to improve production and decrease losses by a considerable percentage. The classifier models utilized in the study combine Logistic Regression, Naïve Bayes, and Random Forest data, with Random Forest results being the best. The study concludes that agricultural production is predicted using machine-learning algorithms to help farmers choose what crops to harvest based on rainfall, temperature, and geolocation.

Nischitha K et al. [15] put forward a system meant to overcome many of the problems encountered by farmers. With an easy-to-use graphical user interface, the system forecasts the best crop to be planted on a given land using soil

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 7, July 2025

DOI: 10.17148/IJIREEICE.2025.13709

characteristics and weather conditions. The system also offers information on needed nutrients, seeds to be used in cultivation, crop yield, and market prices. The system uses a Support Vector Machine (SVM) algorithm for rainfall prediction and a Decision Tree algorithm for crop recommendation. Farmers can use this system to grow new crop types, thus expanding their profit margins while at the same time decreasing soil pollution.

III. PROPOSED SYSTEM

The block diagram of the proposed Machine Learning-based Crop Recommendation System is shown in Fig.1. It outlines a structured flow from data acquisition to crop recommendation. Each block of the system is explained in detail below.

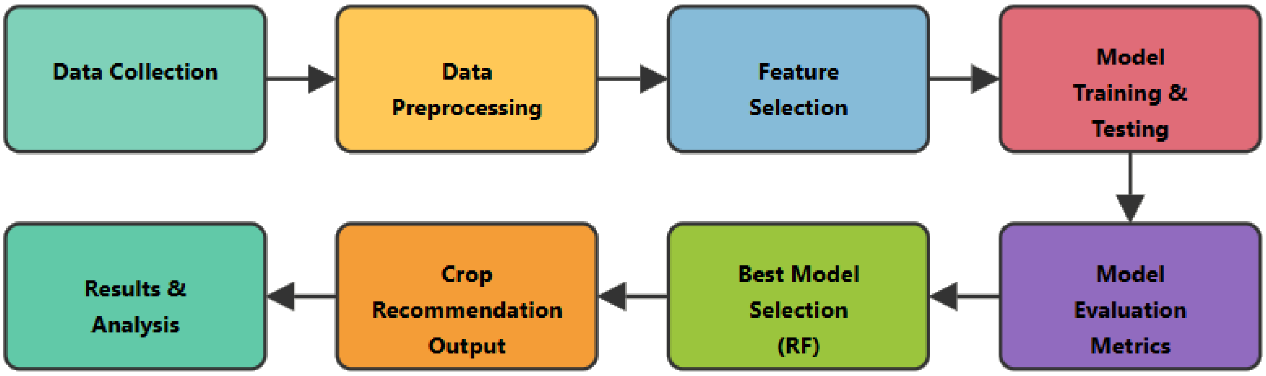

Fig. 1. Block diagram of proposed Crop prediction system

A. Input Dataset

The design of the previously mentioned crop recommendation system is based on data collection. The machine-learning models' effectiveness and reliability were ensured using a strong dataset containing comparative agriculture parameters. Such features that drive crop growth, such as soil nutrients, nitrogen (N), phosphorus (P), and potassium (K), as well as such critical environmental parameters as temperature, humidity, pH level, and rainfall, were present in the dataset. These factors were chosen since they are applicable in determining whether the crop is appropriate and the likely yield. They are based on data collected from open-source stores, different soils, climatic conditions, and geographical cultivation practices. The system has been made capable of effective generalization across geographies and seasons through a varied dataset. The precaution was taken to ensure the integrity and completeness of the data collected before going to the preprocessing phase since any inconsistency or loss of values in the raw data may negatively affect the learning capacity of the model. The complete data collection process guarantees the system is adequately prepared to offer reliable and meaningful crop recommendations to farmers following actual agricultural conditions. The dataset samples are shown in Fig.2.

	N	P	K	temperature	humidity	ph	rainfall	label
0	0.642857	0.264286	0.190	0.345886	0.790267	0.466264	0.656458	20
1	0.607143	0.378571	0.180	0.371445	0.770633	0.549480	0.741675	20
2	0.428571	0.357143	0.195	0.406854	0.793977	0.674219	0.875710	20
3	0.528571	0.214286	0.175	0.506901	0.768751	0.540508	0.799905	20
4	0.557143	0.264286	0.185	0.324378	0.785626	0.641291	0.871231	20

Fig. 2. Samples of dataset

B. Preprocessing

Data preprocessing is an initial process that keeps the gathered dataset from being used optimally to train and evaluate the model. Specific critical methods enhance data quality, accuracy, and dependability. Missing values are first located and either substituted with statistically meaningful estimates or removed by deleting incomplete records to keep the dataset sound. Outlier detection techniques are used to identify and remove the abnormal records that can affect learning. Feature normalization is used to rescale the numerical features into a comparable range to avoid larger-scale features from dominating the learning process by the model. Normalization is needed explicitly with distance-based methods like K-Nearest Neighbors (KNN) and Support Vector Machine (SVM).

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⅀ Peer-reviewed & Refereed journal ⅀ Vol. 13, Issue 7, July 2025

DOI: 10.17148/IJIREEICE.2025.13709

Additionally, exploratory data analysis (EDA) is carried out to identify feature distributions, identify patterns, and reveal latent interdependencies between crop types and soil parameters. By EDA, preliminary impressions of feature importance and interdependencies are derived and direct feature selection and model building. Preprocessing enables data input into machine learning models to be adequately cleaned, in scale, and representative, significantly increasing the ability of the model to learn reasonable and generalizable patterns.

C. Feature Selection

Feature selection is vital in enhancing the performance and efficacy of the developed system for recommending crops. Feature selection implies picking the best possible features among the data, holding onto them, and figuring out which contributes the most value towards predicting the crops and dropping redundant features or features contributing the least usefulness. Nutrient nutrients like nitrogen (N), phosphorus (P), and potassium (K), as well as temperature, humidity, pH level, and rainfalls, were initially utilized as input features in the current study. Their relevance was verified by correlation analysis and feature ranking methods, with the assurance that every planned feature would improve the model's ability to predict. While pinpointing the most significant variables, feature selection keeps computation light, enhances model understanding, and reduces the risk of overfitting. Even though all seven features were kept due to their agricultural relevance and statistical significance, this stringent evaluation ensures that the machine learning models learn from high-quality, non-redundant data. Good feature selection, therefore, improves the model's generalizability over large ranges of agricultural conditions and the overall robustness of crop recommendations.

D. Model Training and Testing

Model training and testing are the crux of building the crop recommendation system because they train the machine learning models on the past data and accurately predict the new unseen data. It achieves this by dividing the preprocessed data into a testing dataset and a training dataset in the proportion of 80:20 or 70:30. The training data is trained on machine learning models such as the Decision Tree Classifier, Random Forest Classifier, Support Vector Machine (SVM), and K-Nearest Neighbors (KNN). All these models are trained to identify the complex non-linear relationship between the soil parameters and the crop varieties that must be grown. Various techniques, such as cross-validation, grid search, and hyperparameter tuning, are performed during training to maximize the performance. The approaches allow the best model parameter estimation and avoidance of overfitting to achieve better generalization of new data. The models are then evaluated on the test set, and the quality of their predictions is evaluated using standard quality measures. This step ensures that the models are proficient at memorizing training data and resilient at predicting the right crop for varied soil and environmental conditions. The best and most accurate model will be deployed in the final system through relative comparison based on model performance.

E. Model Evaluation

Performing performance testing is of utmost necessity to make the trained machine learning models effective and reliable for the crop recommendation system. In this research work, specific performance measures were considered to evaluate the predictive ability of the models comprehensively. A straightforward measure of accuracy was used as a ratio of correct predictions to total predictions. However, precision alone may not be good enough, particularly in class imbalance or multiclass classification. Thus, other values, such as recall and F1-score, were considered. Precision tracks the proportion of correctly predicted positive instances to all the predicted positive cases. At the same time, recall verifies how well the model performs in correctly labeling all the correct positive instances. The F1-score, a harmonic mean of precision and recall, gives the composite measure of the model's performance. Along with this, a confusion matrix was also used to check classification vs. actual classification across the different types of crops so that the crop types, in general, most likely to get misclassified, were determined. By ensuring the execution of a combination of these steps, the evaluation process provides a solid, unprejudiced, and holistic choice of the best-performing model for deployment in the end-crop recommendation system.

F. Best Model Selection

Among the comparison of results of different machine learning models, the Random Forest Classifier was the best-performing model under crop recommendations. As an ensemble model, multiple decision trees are created during training, making it a better, more consistent, and generalizing model compared to Decision Trees, Support Vector Machines, and K-nearest neighbors. Its power lies in its ability to process data of high dimensions and learn complex, non-linear patterns without overfitting. Random Forest aggregates predictions from many shallow decision trees to eliminate variance and increase prediction confidence. It posted the highest score on all the evaluation measures, such as precision, recall, accuracy, and F1-score in the test, which is a testimony to its high performance. Due to its high precision and interpretability, the Random Forest model was deployed in the ultimate crop recommendation system to provide farmers with reliable and confident recommendations in different agricultural contexts.

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 7, July 2025

DOI: 10.17148/IJIREEICE.2025.13709

G. Best Model Selection

The final step of the system suggested is producing the output of crop recommendations based on the inputs provided by users. Once soil and environmental parameters—such as nitrogen, phosphorus, potassium, temperature, humidity, pH, and rain—enter the system, the trained Random Forest model evaluates these inputs, and output is produced on the best crop to grow. This suggestion is arrived at through an analysis of the patterns acquired from the training data so that the suggested crop would be appropriate for the area's specific soil characteristics and climatic patterns. The system will give outputs in real time, making it highly practical and helpful for farmers and agricultural officers. Through this recommendation system, users can make informed, knowledge-based decisions for crop selection that lead to maximum yield, minimum waste of resources, and minimum environmental impact. Furthermore, the system can be developed at affordable platforms such as mobile applications or web interfaces, thus ensuring maximum reach and utilization by the farmer population. Generally, the crop recommendation output allows users to make an informed decision with actionable knowledge, thereby supporting sustainable agriculture and improving economic benefits.

IV. RESULTS AND DISCUSSION

In this study, four machine learning classifiers—Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Decision Tree (DT), and Random Forest (RF)—were trained, tested, and evaluated for the crop recommendation task using the prepared dataset.

A. Results of SVM

The SVM classifier performed at 95.68%. While SVM performed well, its inability to deal with multiclass classification problems and non-linear relationships somehow constrained its performance relative to ensemble models. Examining the confusion matrix indicates that SVM had misclassified certain crop classes where data points contained similar feature values.

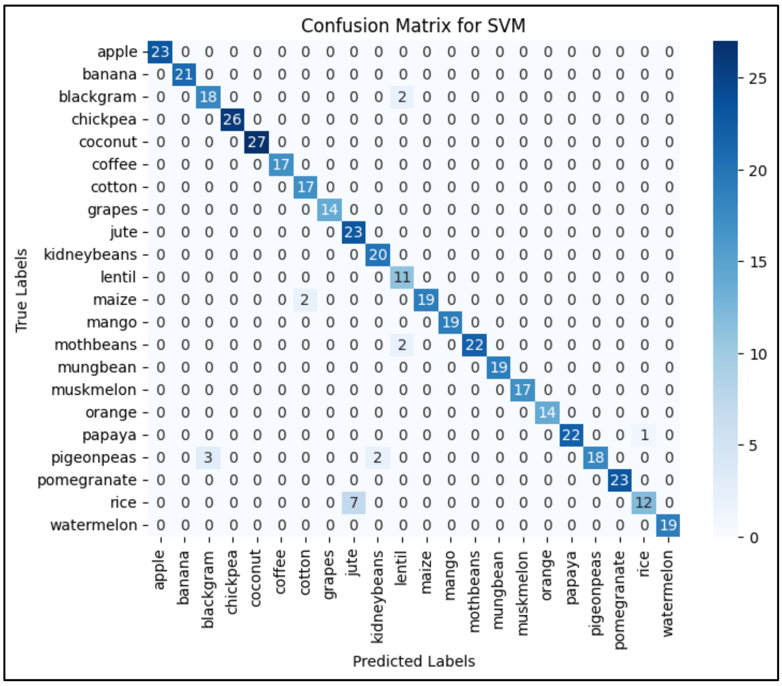

Fig. 3. Confusion matrix of SVM

B. Results of KNN

KNN classifier provided an enhanced accuracy of 96.81%. KNN improved class discrimination within the feature space as a representative instance learner. However, sensitivity to choice of k-value and computational inactivity when working on larger data were noted as limitation points.

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 7, July 2025

DOI: 10.17148/IJIREEICE.2025.13709

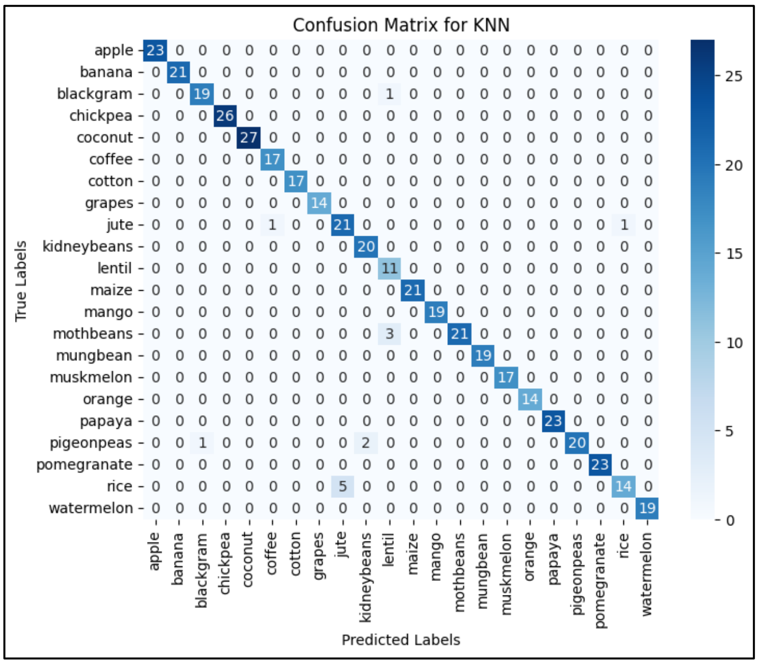

Fig. 4. Confusion matrix of KNN

C. Results of DT

The Decision Tree classifier worked surprisingly well at 98.63% accuracy. The model was able to learn the complex relationships well and was also quick to train and predict. Interpretability of the decision paths and feature importance analysis contributed to its performance further.

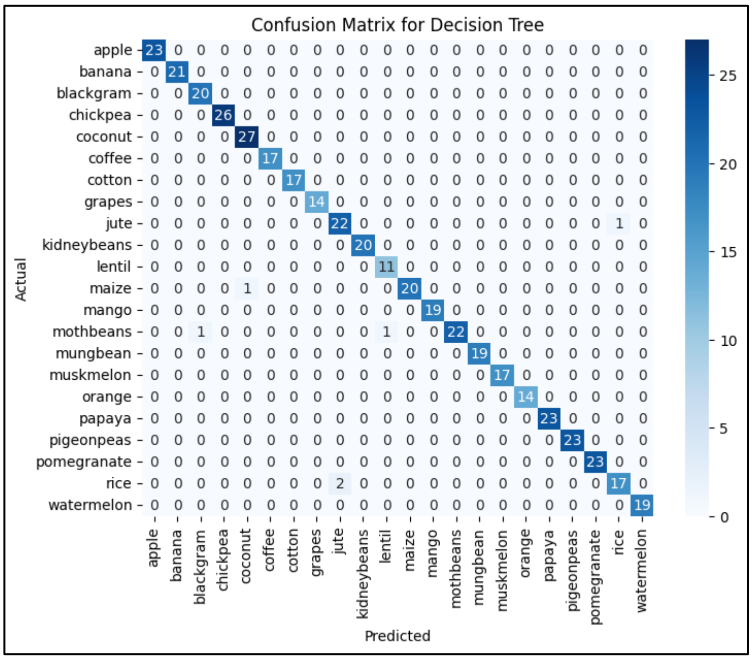

Fig. 5. Confusion matrix of DT

D. Results of RF

The Random Forest model resulted in a maximum accuracy of 99.31% from all the classifiers. RF's ensemble property of averaging over multiple decision trees gave excellent generalization on unseen samples with minimal overfitting. RF also provided feature importance data, confirming that the nitrogen (N), phosphorus (P), and potassium (K) levels were the most influential parameters that contributed to crop prediction.

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⪤ Peer-reviewed & Refereed journal ⪤ Vol. 13, Issue 7, July 2025

DOI: 10.17148/IJIREEICE.2025.13709

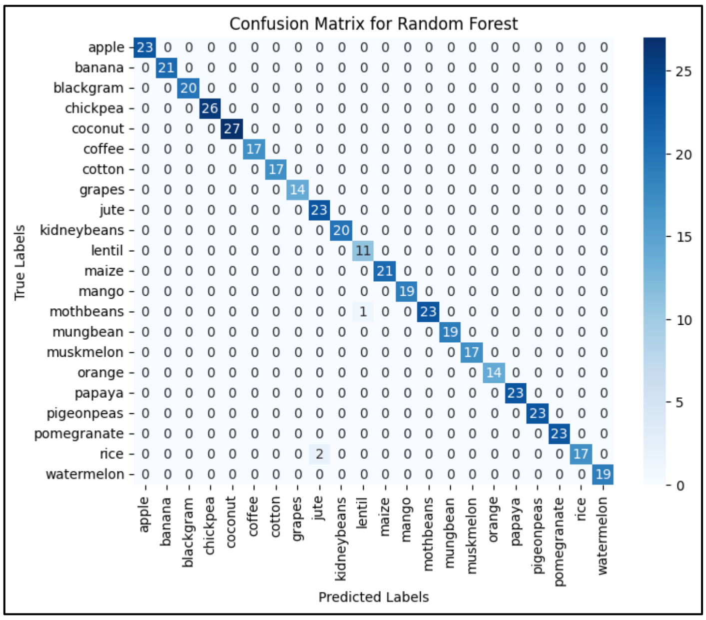

Fig. 6. Confusion matrix of RF

Fig. 7.

Table I presents a comparative analysis of the proposed system for crop recommendation.

TABLE I. PERFORMANCE OF ML CLASSIFIER FOR CROP RECOMMENDATION

	Classifier	Accuracy
0	SVM	0.956818
1	KNN	0.968182
2	Decision Tree	0.986364
3	Random Forest	0.993182

Table I captures the performance of some of the machine learning classifiers deployed in crop recommendation concerning accuracy. Of the models tested, the Support Vector Machine (SVM) had an accuracy of 95.68%, while K-Nearest Neighbors (KNN) performed better with a 96.81% accuracy. The Decision Tree (DT) classifier performed much better, with an accuracy of 98.63%, indicating the model's capability to learn complex decision boundaries. But then again, the Random Forest (RF) classifier performed better than the rest at a maximum accuracy of 99.31% because it uses the ensemble learning method, combining several decision trees to improve generalization and avoid overfitting. The results show that although all models showed excellent predictive power, Random Forest is the most stable and consistent model for reliable crop recommendation and, therefore, the most suitable to be deployed in the system proposed.

V. CONCLUSION

This paper presents the design and testing of a machine-learning-based crop advisory system for improving sustainable agriculture. Using a data set containing vital soil and environmental parameters like nitrogen, phosphorus, and potassium content, temperature, humidity, pH, and rainfall, different machine-learning models were trained and tested according to their predictability for crops. Among the models tested, the Random Forest classifier gave the best accuracy of 99.31%, outperforming SVM, KNN, and Decision Tree classifiers. The Random Forest model can handle complex non-linear relationships, has high robustness, and is less prone to overfitting, making it a suitable model to be deployed in the field. The proposed system empowers the farmers with improved decision-making abilities regarding crop choice, optimal use of yield resources, and implementation support for green agriculture practices. Through the supply of data-driven information, the proposed system empowers precision agriculture and sustainable food production activities per global initiatives for improving food security and mitigating climate change.

While the proposed system exhibits high accuracy and practical relevance, several avenues exist for future improvement and expansion. Real-time weather data, satellite imagery, and dynamic soil health monitoring can enhance the system's predictive capabilities.

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering
Impact Factor 8.414 ⪤ Peer-reviewed & Refereed journal ⪤ Vol. 13, Issue 7, July 2025
DOI: 10.17148/IJIREEICE.2025.13709

Integration of market trends, pest incidence forecasting, and economic viability analysis would allow the system to offer agronomic and financial guidance to farmers. Advanced deep learning models, such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), can be explored for even more accurate and adaptive crop recommendations. Additionally, deploying the system as a mobile or IoT-based application will increase accessibility for farmers, especially in rural areas. Collaboration with agricultural experts for continuous model validation and user feedback will ensure the system remains practical, accurate, and beneficial in real-world farm scenarios. These future enhancements will transform the system from a recommendation engine into a comprehensive digital farming assistant.

REFERENCES

[1] V. Geetha, A. Punitha, M. Abarna, M. Akshaya, S. Illakiya and A. P. Janani, "An Effective Crop Prediction Using Random Forest Algorithm," *2020 International Conference on System, Computation, Automation and Networking (ICSCAN)*, Pondicherry, India, 2020, pp. 1-5, doi: 10.1109/ICSCAN49426.2020.9262311.

[2] Z. Doshi, S. Nadkarni, R. Agrawal and N. Shah, "AgroConsultant: Intelligent Crop Recommendation System Using Machine Learning Algorithms," 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), Pune, India, 2018, pp. 1-6, doi: 10.1109/ICCUBEA.2018.8697349.

[3] P. A, S. Chakraborty, A. Kumar, and O. R. Pooniwala, "Intelligent Crop Recommendation System using Machine Learning," 2021 5th International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2021, pp. 843-848, doi: 10.1109/ICCMC51019.2021.9418375.

[4] Y. J. N. Kumar, V. Spandana, V. S. Vaishnavi, K. Neha, and V. G. R. R. Devi, "Supervised Machine Learning Approach for Crop Yield Prediction in Agriculture Sector," 2020 5th International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, 2020, pp. 736-741, doi: 10.1109/ICCES48766.2020.9137868.

[5] Y. J. N. Kumar, V. Spandana, V. S. Vaishnavi, K. Neha, and V. G. R. R. Devi, "Supervised Machine Learning Approach for Crop Yield Prediction in Agriculture Sector," *2020 5th International Conference on Communication and Electronics Systems (ICCES)*, Coimbatore, India, 2020, pp. 736-741, doi: 10.1109/ICCES48766.2020.9137868.

[6] M. Kavita and P. Mathur, "Crop Yield Estimation in India Using Machine Learning," *2020 IEEE 5th International Conference on Computing Communication and Automation (ICCCA)*, Greater Noida, India, 2020, pp. 220-224, doi: 10.1109/ICCCA49541.2020.9250915.

[7] Meeradevi and H. Salpekar, "Design and Implementation of Mobile Application for Crop Yield Prediction using Machine Learning," *2019 Global Conference for Advancement in Technology (GCAT)*, Bangalore, India, 2019, pp. 1-6, doi: 10.1109/GCAT47503.2019.8978315.

[8] S. Agarwal and S. Tarar, "A hybrid approach for crop yield prediction using machine learning and deep learning algorithms," *J. Phys.: Conf. Ser.*, vol. 1714, p. 012012, 2021, doi: 10.1088/1742-6596/1714/1/012012.

[9] N. Manjunathan, P. Rajesh, E. Thangadurai, and A. Suresh, "Crop yield prediction using linear support vector machine," in *Proc. Int. Conf.*, 2020. [Online]. Available: https://api.semanticscholar.org/CorpusID:246470919

[10] S. Prabhu, P. Revandekar, S. Shirdhankar, and S. Paygude, "Soil analysis and crop prediction," *Int. J. Sci. Res. Sci. Technol.*, vol. 7, no. 4, pp. 117–123, Jul. 2020, doi: 10.32628/IJSRST207433.

[11] K. Lata and S. Khan, "Experimental analysis of machine learning algorithms based on agricultural dataset for improving crop yield prediction," *Int. J. Eng. Adv. Technol.*, vol. 9, no. 1, pp. 3246–3251, Oct. 2019, doi: 10.35940/ijeat.F9308.109119.

[12] D. Jayanarayana Reddy and Dr. Kumar, "Crop yield prediction using machine learning algorithm," in *Proc. 5th Int. Conf. Intelligent Computing and Control Systems (ICICCS)*, Madurai, India, May 2021, pp. 1466–1470, doi: 10.1109/ICICCS51141.2021.9432236.

[13] S. Khaki and L. Wang, "Crop yield prediction using deep neural networks," *Frontiers in Plant Science*, vol. 10, p. 621, May 2019, doi: 10.3389/fpls.2019.00621.

[14] U. Devi and S. Selvakumari, "Crop prediction and mapping using soil features with different machine learning techniques," in *Proc. Int. Conf. Innovative Computing & Communication (ICICC)*, 2022. [Online]. Available: https://doi.org/10.2139/ssrn.4097213.

[15] N. Mahendra, D. Vishwakarma, K. Nischitha, Ashwini, and M. R. Manjuraju, "Crop prediction using machine learning approaches," *International Journal of Engineering Research & Technology (IJERT)*, vol. 9, no. 8, pp. –, Aug. 2020.

