Impact factor 6.2

Geoscience Journal

ISSN:1000-8527

Indexing:

- Scopus
- Google Scholar
- DOI, Zenodo
- Open Access

www.geoscience.ac

GEOSCIENCE JOURNAL
ESTD. 2020
ISSN:1000-8527

Registered

Optimization of Agricultural Parameters in Smart Farming Using Machine Learning-Based Decision Support Systems for Crop Productivity Enhancement

Amod Kumar Sahwal[1] , Dr. Sanjay Kumar[2]

1. Amod Kumar Sahwal, Research Scholar, Department of Computer Science and Engineering, N.I.T Jamshedpur, Jharkhand
2. Dr. Sanjay Kumar, Associate Professor, Department of Computer Science and Engineering, N.I.T Jamshedpur, Jharkhand

Abstract

Purpose: This study aims to develop and assess a crop recommendation system driven by soil and climatic data utilising machine learning techniques. The objective is to optimise resource utilisation and enhance crop yields by aiding farmers in picking the most appropriate crops for their respective geographical regions.

Method: The study employed a dataset that included rainfall, temperature, humidity, pH, phosphorus (P), potassium (K), and nitrogen (N). A number of machine learning models were trained and evaluated, including Random Forest, K-Nearest Neighbours, Naive Bayes, Logistic Regression, Decision Trees (Gini and Entropy), Support Vector Machines (with RBF and Linear Kernels), and Neural Networks (with ReLU and Sigmoid activations). It assessed each model's performance using criteria such as recall, precision, accuracy, and validation accuracy.

Results: The Random Forest classifier had the best accuracy (98.90%) and strong precision/recall values (0.98/0.97) of all the models. It was closely followed by SVM with a linear kernel and Decision Tree (Entropy). The results showed that ensemble and kernel-based methods are quite good at predicting what kind of crop a plant is. The study also showed what the best environmental conditions were for 22 different crops based on average feature values.

Conclusion: Machine learning techniques, particularly Random Forest and SVM, have the potential to produce highly accurate crop recommendations when trained on well-structured agro-climatic datasets. The study demonstrates how data-driven decision-making can significantly enhance precision agriculture. The technique might be incorporated into smartphone apps to assist farmers in making real-time crop selections, promoting agriculture's long-term growth.

Keywords: Crop Recommendation System, Machine Learning, Random Forest, Support Vector Machine (SVM), Agro-Climatic Parameters**.**

1. INTRODUCTION

Agriculture is very important in stabilizing the economy and ensuring that human beings do not starve to death. This is more so to countries which rely on agricultural practices such as India. However, the conventional farming does not work very efficiently since it always relies on human intelligence without easy access to scientific tools to make decisions. The world is getting more people, less land can be cultivated, and weather is less predictable. This increases the need to use intelligent systems that utilize resources the most and enlarge crop production.

The oldest industry is agriculture and it is required to feed the population of the world. It has evolved through the years in order to enhance efficiency, increase the number of people participated as well as increasing the quality standards [1]. Nonetheless, farmers have a big problem because there is a possibility of reduced arable land due to urbanisation. The necessity of new ideas is even greater with the thought of the fact that food production must grow more than 70% by 2050 to meet the growth demands of the population [2]. It is vital to introduce an automatic system that will deliver the most value out of a harvest and utilize the minimum number of resources. This will enhance more sustainable farming in response to the changing world demands [3].

Agricultural industry can be changed using machine learning that has now provided its solutions to centuries-old problems such as evaluating crop yields, [4] identifying plant species [5], and detecting diseases [6]. This brings in a new dawn in precision agriculture. Using big data sets that contain such information as soil composition, weather patterns, crop health, and yield estimates, ML algorithms provide valuable information to farmers, agronomists and academics. Over the past years, a lot of scientists deployed machine learning to develop agricultural driver-based crop recommendation algorithms.

Smart farming founded on the new technologies has proved to be one of the potential solutions to the current farmers problems. Among these emerging technologies is machine learning (ML) which has potential in the assessment of high volumes of complex information and developing valuable insight. Soil nutrient levels (Nitrogen, Phosphorus, Potassium), temperature, humidity, pH, and rainfall are just a few of the environmental factors that machine learning (ML) models take into account, enabling farmers to make informed decisions about which crops to grow in a given area.

A machine learning-based decision support system (DSS) that can recommend crops must be developed and validated as part of this assignment. Real-world agricultural data is used to teach the system how to accurately estimate the type of crop that would flourish in a certain Dung area and weather conditions. The project aims to examine several machine learning models, including Decision Tree, Random Forest, SVM, and Neural Networks, in order to determine the optimal algorithm that balances accuracy, speed, and practical application. By using data-driven smart farming practices, the study aims to support the ongoing efforts towards more sustainable farming and increased farm revenues.

1.1.Statement of the problem

Farmers in many areas still don't have the tools they need to make timely, data-driven decisions, which means that agricultural output is still not as high as it could be. When choosing crops and planning how to grow them, farmers sometimes use old methods or general rules that don't take into consideration how the environment and soil change over time. As the world's population grows and the climate changes, the agricultural sector is under a lot of pressure to improve resource use, sustainability, and yield efficiency. Machine learning and the Internet of Things (IoT) are two examples of modern technologies that could greatly help smart farming. However, there is still a big gap in how well they can be used in decision support systems that are easy for farmers to use. So, The study need to quickly create and improve smart models that can look at a lot of different agricultural factors, such soil nutrients, temperature, humidity, pH, and rainfall, and give us the best crop choices to get the most out of our work. By creating a machine learning-based decision support system that improves the accuracy and impact of crop selection in smart farming settings, this work fills in the gaps.

1.2.Research Questions

1. How well can machine learning algorithms forecast which crop would be best to grow given important agricultural factors like rainfall, pH, temperature, humidity, and soil nutrients?
2. How does are the accuracy, precision, recall, and validation scores of different machine learning models used in crop recommendation systems? Examples include Decision Trees, Random Forests, SVMs, and Neural Networks?
3. To what extent can a machine learning-based decision support system improve crop productivity and support data-driven decisions in smart farming practices?

2. LITERATURE REVIEW

The integration of machine learning (ML) with agriculture has garnered increasing attention in recent years, especially in the context of smart farming. As global food demands rise and agricultural resources remain limited, optimizing crop productivity through data-driven techniques has become critical. The potential of ML-based decision support systems (DSS) to improve yield forecasts, automate agricultural procedures, and improve crop recommendations has been the subject of numerous studies. These algorithms recommend the best crops for a location based on environmental and soil factors, including temperature, humidity, pH, rainfall, and levels of nitrogen, phosphorus, and potassium.

2.1.Overview of technology in modern agriculture

Taha et al. (2025) gave a full evaluation of these new ideas, showing how Ag5.0 changed traditional farming into a process that is highly automated and based on data. The study stressed how AI and ML may help with crop monitoring, decision-making, and predictive analysis. Their research included real-world case studies that showed how deep learning and sensor-based robots greatly improved yield forecasts and disease detection, which in turn reduced the impact on the environment [7].

Singh and Sharma (2025) looked at 188 scholarly papers over the course of six years to do a systematic assessment of IoT applications in precision agriculture. Their studies revealed that IoT is gaining significance in tracking and monitoring agricultural activities, such as smart irrigation, crop health monitoring, livestock tracking, crop health monitoring and monitoring agriculture, as well as soil nutriment analysis. The project also examined how IoT working together with ML and fuzzy logic systems assisted farmers to make a data-driven decision. Such decisions utilized resources more efficiently, minimised waste and influenced the environment less. They have also discussed the current issues about the connection, interoperability and real-time analytics and they have provided strategic ways of making efficient the use of IoT in agriculture [8].

Raj et al. (2024) contributed another major role by coming up with a comprehensive report on the future of farming as a result of automation and AI. Their study explored the range of available robotics including autonomous tractors, intelligent grippers, and AI-driven drones that would allow them to perform other significant farming tasks, including planting, weeding, spraying, and harvesting without their direct involvement. These technologies enabled a constant monitoring of the plants and receiving feedback in the real-time mode that allowed increasing the quality of crops and efficiency of the operation. The researchers did note some of the issues thought however that it was difficult for the people to accept the technology greatly despite its ability to increase productivity and reduce expenditure. Such issues were high initial capital investment, data gaps in the digital infrastructure and concerns with cybersecurity [9].

Hajiyeva et al. (2024) reflected upon the significance of technical innovation in terms of bridging the agriculture sector of Azerbaijan into the contemporary from the national perspective. The paper adopted systematic processes such as abstraction, deduction, and system analysis to discuss how production was made efficient with the help of digital and environmental technology. They have found that ecosystem based management linked with agro-landscape design could preserve the environment and stabilize the economy. Another aspect emphasised by the researchers is that policies and innovation projects, which would fit the needs of the rural agricultural economy, are necessary to promote the adoption of new technologies [10].

Méndez-Zambrano et al. (2023) reviewed 280 publications in a systematic manner and selected 40 studies that examined the manner in which ICT can be implemented in digital agriculture. On the basis of their research, it was revealed that individuals are increasingly using AI, machine learning, IoT, drones and a mobile application to make soil, water, fertilisers and agrochemicals easier to manage. These implements made everything a lot efficient when compared to the impact they produced on the environment. The researchers did mention however a very large technological gap in Latin America countries. As compared to Asia, Europe, and North America, these countries could not utilize advanced agri-tech infrastructure. This contrast was sufficient proof that additional investment and government support is applicable to bridge the digital rift in agriculture on a global level [11].

2.2.Previous works on machine learning in crop recommendation

Gajbhiye et al. (2025) applied several machine learning instances, including Decision Trees, Random Forest, and SVM, to create a powerful model of crop recommendation. Their findings revealed that Random Forest classifier had the best prediction accuracy, which goes to confirm the usefulness of the model in aiding the farmer in their decision making process in the real world [12].

Arun et al. (2024) explored the feasibility of applying machine learning to precision agriculture in order to support the increasing food demand whilst consuming less resources. They examined other studies on crop recommendation systems that involved IoT-based sensing technologies, geospatial, and remote sensing solutions. The aims of such systems were to realize the maximum out of the crops and environmental protection as well. The authors arrived at a conclusion that one of the opportunities to make farming more effective in resource-intense and resource-scarce regions without facing scalability concerns was precision farming using machine learning [13].

Upadhyay (2024) proposed a system of machine learning-based crop recommendation that is to assist individuals to select the most appropriate crops to grow within a specific area which involves the application of Random Forest (RF) algorithm which considers some factors, such as the quality of soil, the climate, etc. This investigation used a methodological pipeline that included data collection, preprocessing, feature selection, training, testing, and model evaluation. The RF model proved to be dependable in recommending the best crop for a given circumstance, as seen by its 99 percent accuracy record. Also, it included an easy-to-use interface that made the advice understandable to farmers with different levels of technical expertise. This increased the system's transparency and functionality in the actual world [14].

Hasan et al. (2023) applied a machine learning method known as ensemble called K-nearest Neighbour Random Forest Ridge Regression (KRR) to predict the yield of major crops in Bangladesh where agriculture continues to play a major role in the economy. A combination of historical crop production records and national environmental data at national institutions were applied in the study. Compared to many classical and ensemble models, which were used on the same four performance measures (MAE, MSE, RMSE, and R 2), the KRR model acted superior to many classical and ensemble techniques, including Support Vector Regression, Naive Bayes, Ridge Regression, Random Forest, and CatBoost. As an illustration, the model had a R2 of 99% on Aus and Boro rice, wheat and potatoes. Using the DieboldMariano test, the model proved to be powerful to be statistically significant at the 1 percent level and at the 5 percent level of most of the comparisons. Another development, which was made using the study, was dynamic recommender system, which was to assist farmers in their selection of best crops to plant during the new growing season. This will streamline the farming to be more economically beneficial [15].

2.3.Review of machine learning models used in agriculture

Badshah et al. (2024) proposed an attractive method to categorize crops and predict harvests through state-of-art machine learning algorithms to make the process of agriculture more sustainable. In their study, they employed a Kaggle dataset involving inputs such as pH, temperature, humidity, nutrient levels in order to develop two architectures. One was on crop recommendation based on classification methods such as Extra Trees Classifier, Logistic Regression, Random Forest and among others. The other was the forecasting of wheat yields in Pakistan in 1992-2013 based on the FAO and the World Bank data. The highest accuracy in the classification was experienced by Random Forest (99.7%), whereas the highest R 2 score was experienced by Support Vector Regressor (999.9%). Feature Importance and LIME are two examples of Explainable AI (XAI) methods used in this study; K-fold cross-validation helped make the model's real-world applications more understandable [16].

Sharma et al. (2023) compared machine learning and deep learning systems to look at estimating crop yields. They applied such techniques as Decision Tree, Deep Learning, Random Forest, Random Forest, and XGBoost to data concerning the environment and production. The Random Forest model was the most accurate with the mean absolute error (MAE): 1.97, whereas CNN has the least loss (0.00060). These findings provided an indication of the usefulness of ensemble and deep learning models in yield prediction. Indicators such as RMSE and standard deviation indicated that the model was even superior [17].

Kumar et al. (2022) looked at the bigger picture of how AI and ML are being used in Indian agriculture. Their study showed how important it is to use technology to change farming because there is less land that can be farmed and more people. They said that India, where more than half of the jobs are in agriculture, had begun to use smart farming methods with the help of national programs like INDIAai and the NITI Aayog plan. The report said that the AI market in agriculture would expand from $1 billion to $4 billion by 2026. This shows how important the industry will be for preparing for the future of the economy and food security [18].

2.4.Gap in existing literature

Many research has looked at how machine learning can be used in farming, but most of them only look at how well individual models work instead of how to make a whole decision support system that meets the demands of local farmers. Most of the research that has been done so far doesn't take into account the unique environmental conditions of each region, which makes it less useful for a wide range of agricultural settings.There isn't much research that compares different machine learning models using the same input parameters, which makes it hard to find the best method for recommending crops. Also, even though many studies have shown that these models are very accurate, not much attention has been paid to how they help farmers really increase their production and make better use of their resources.

This study aims to fill in these gaps by (i) creating a decision support system based on key environmental inputs, (ii) comparing different machine learning models to find the one that works best, and (iii) using historical data to find ways to improve crop selection and output in agriculture.

2.5.Contribution and uniqueness of the current study

This study adds to the growing field of agricultural informatics by creating a decision support system based on machine learning that combines important environmental and soil factors to make reliable crop recommendations. This study is different from previous ones that just used one algorithm. It compares several models, such as Decision Tree, Random Forest, Support Vector Machine, and others, to find the best one for predicting crops in different situations.

The study is one of a kind since it uses explainable AI methods to make the recommendations clear and easy for farmers to understand. It also looks at how machine learning may help farmers use their inputs more efficiently and boost their productivity, which is good for the environment. This research is particularly unique since it has a user-accessible interface that connects advanced data science with real-world farming applications.

3. MATERIAL AND METHODS

3.1.Data Description

To prepare it for the model of the study, the study is preprocessing a dataset that already exists on Kaggle [19]. Table 1 displays the key aspects of the data.

Table 1: Description of Features Used for Crop Recommendation

Index	Feature Name	Description
1	Nitrogen (N)	Essential for vegetative growth and responsible for leaf development.
2	Phosphorus (P)	Promotes root development and enhances flower and fruit formation.
3	Potassium (K)	Regulates overall plant functions and strengthens plant resistance.
4	Temperature (°C)	Ambient temperature measured in degrees Celsius.
5	Humidity (%)	Relative humidity in the atmosphere, expressed as a percentage.
6	pH	Soil acidity or alkalinity, measured on the pH scale.
7	Rainfall (mm)	Total rainfall received, measured in millimeters.

Table 1 lists the main aspects of a crop recommendation system. All of them are highly essential in determining what crops should be grown on each geographical territory based on the soil, and weather conditions. It has significant soil aspects, such as; nitrogen (N), phosphorus (P) and potassium (K) that directly affect plant growth, root development and the general health of the plant. Temperature and Humidity are some of the things to learn in order to determine whether or not the climate is favorable to various crops. Soil pH will tell us the chemical composition of the soil, which will determine the effectiveness of agricultural crops to grow and the amount of nutrients that the crops can receive. Finally, Rainfall is useful in determining the required amount of water in crops and its level of tolerance. Taken in combination, these aspects form a complete image of the agro-environmental conditions that could be used to select crops that are most appropriate to a particular region.

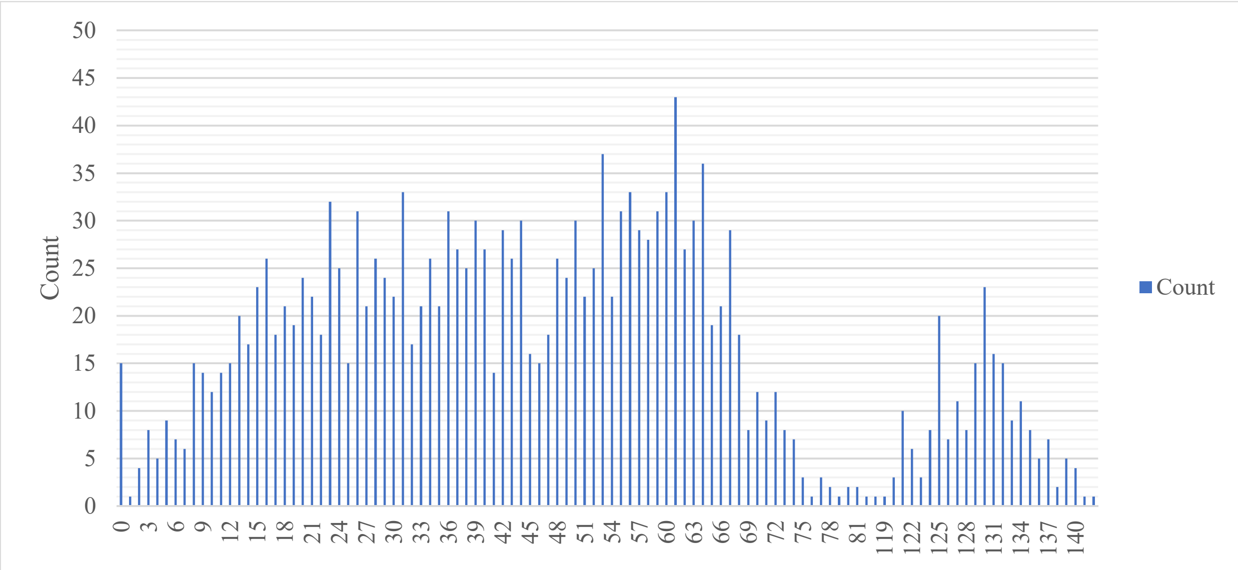


Figure1: Feature Graph for Temperature

Figure 1 displays the characteristics' number and visual representation.With each row and column denoting a distinct variable, a pairplot is a kind of statistical graph that displays the relationships between several features in a dataset as a matrix. The distribution of each variable is displayed in the diagonal plots of the matrix, and the interactions between pairs of variables are displayed in the off-diagonal plots. This dataset was produced by the addition of existing rainfall, climate, and fertiliser data for India. The dataset used in this investigation has 22 distinct crop labels. Every label represents a distinct agricultural product. Among the principal crops cultivated are rice, maize, chickpeas, kidney beans, pigeon peas, moth beans, mung beans, black gramme, and lentils. Pomegranate, banana, mango, grapes, watermelon, muskmelon, apple, orange, and papaya are among its other fruit harvests. Among the commercial and industrial crops are coffee, jute, coconut, and cotton. This wide range of choices enables the machine learning models to consider a wide variety of soil and climate preferences, resulting in more accurate and practical crop recommendations. Since there is only one good crop for a particular location, the database of about 100,000 records from which these labels are taken was trimmed to about 2.2k records.

3.2.Methodology

In this part, the studyprovides a high-level summary of the methods utilised to train the different models, as illustrated in Figure 2. It all started with a series of iterations using each of the chosen machine learning algorithms to complete the following steps.

- **Input Data:**The study made sure the data was clean and well-labeled because the quality and quantity of the data have a big effect on how accurate a model is.Figure 2 shows that the system's input is a mix of soil and ambient factors. The study trained and tested our models on some raw data, which is shown in Table 1.
- **Preprocessing:**The study cleaned the data, removed outliers, and converted it to a format your machine learning system could use. The study first removed all blanks and duplicates, separated the features from the label column, used a technique called feature engineering to generate new features from the old ones, and finally characterised and plotted the data to make sure there were no outliers.
- **Choose a machine learning algorithm:** Each time we ran the program, The study picked one of the seven algorithms we had chosen to employ. We went through the procedures of preprocessing, testing, or validating the model over and over again for each chosen algorithm to fine-tune the model.
- **Model Configurations:** We tried adjusting the activation function, epoch count, decision tree depth, and number of nearest neighbours among other parameters to see what worked best for improving test and cross validation accuracy. Additional critical characteristics are also shown in Figure 2. It is important to think about how well the model works as well.For instance, the model's performance might be negatively affected by deepening the decision tree. The performance of the neural network model would also be drastically affected if the data entering into it were to increase in frequency.
- **Training Models:** Following the data preparation in the "Preprocessing" phase, the machine learning algorithm moves on to the "Learning" phase.
- **Checking the Model's Accuracy:** We check how accurate the model we made is by comparing it to the test data. We also checked how well the model did using cross-validation. We go back to the "Model Configuration" phase if the accuracy isn't good enough. We tried out the feature engineering method in various cases. Let's say that the model's accuracy and performance are good at this point. In that situation, we go back to choose a new algorithm step and do the entire thing again with a different advanced machine learning algorithm.

3.3.Experimentation

1) **Model utilizing Multi-Class Neural Network:** A multi-class neural network can sort data into more than one class. A single-class neural network, on the other hand, can only put data into one category. We used the TensorFlow [20] framework to build a neural network with four layers. Listing 1 displays one of the examples. There are thirty neurons in the first layer (input layer), twenty neurons in the second layer, ten neurons in the third layer, and twenty-two neurons in the fourth layer (output layer). The concealed layers are the second and third layers. We also tried out different combinations of "relu," "softmax," and "sigmoid" activation functions [21], to improve the model's accuracy and performance. Lastly, we tried out different epoch [22] values on the network until we discovered the best ones. The neural network's performance goes down when the epoch value goes up. We also employed CategoricalCrossentropy as a loss function [23], which is also known as categorical

crossentropy. It is used to teach models how to classify things into more than one class. It tells you how far apart the anticipated probability and the actual labels are. The model is doing better if the categorical crossentropy is lower. Finally, we apply the Adams [24] optimizer to help us. Adam is a popular optimisation method for training DL models. A vast array of issues benefit from its enhanced performance compared to the AdaGrad [25] and RMSProp [26] methods.

Listing 1: Neural Network

```
import tensorflow as tf
model = tf.keras.models.Sequential([
        tf.keras.layers.Dense(30, activation = 'relu',
                input_shape = (7,)),
        tf.keras.layers.Dense(20, activation = 'relu'),
        tf.keras.layers.Dense(10, activation = 'relu'),
        tf.keras.layers.Dense(labels_count, activation
= 'softmax')
])
model.compile(
        loss = tf.keras.losses.CategoricalCrossentropy(),
        optimizer = tf.keras.optimizers.Adam(),
        metrics = ['accuracy']
)
model.fit(x_train, y_train, epochs = 60,
        validation_data = (x_test, y_test), batch_size =
        32)
```

2) **Rest of the Models:** Every model, with the exception of those that used neural networks, was created, trained, tested, and validated in a similar way. With various methods, the classifiers from Section IV were used to create the various models. This section addresses several significant changes. Two impurity metrics featured in decision trees, Gini and entropy [27], made up the decision tree method that was employed. Additionally, we made use of max depth as a parameter for decision trees. For the purpose of testing the k-nearby-neighbors technique, we employed a configuration we dubbed nneighbors. Based on these conditions, the algorithm will decide how many closest outputs to consider. We developed the SVM kernel. Of all kernel machines, the most famous one is the (SVM). Keramid machines are the algorithms that can be utilised for pattern analysis. We also experimented with cross-validating the models using listing two features.

Table2: Machine Learning Models Used

Index	Model Name
1	Logistic Regression

2	Decision Tree Classifier
3	Random Forest Classifier
4	K-Nearest Neighbors (KNN)
5	Gaussian Naive Bayes
6	Support Vector Classifier (SVC)

In the case of decision tree algorithm, we employed Gini and entropy, and these measures are impurity that occur in decision tree [27]. The maximum decision tree depth was the other parameter we employed. We tested a parameter called nneighbors in the case of k-nearest neighbours; this configuration determines the total number of nearest outputs to be taken into account. We used kernel configuration in SVM. Kernel machines are a class of methods for pattern analysis, best known by its member the support-vector machine (SVM). Additionally, we cross-checked and evaluated models using listing 2's functionality.

Listing 2: Evaluation of the Model

```
kfold = KFold(number_splits=5, shuffle=True,
        random_state=42)
predictions = model.predict(x_test_data)
accuracy = accuracy_score(predictions,
        y_test_data)
return round (accuracy*100, 3)

def perform_cross_val(my_model):
        scores = cross_val_score(model, features,
        labels, cv=kfold)
mean_score = round(scores.mean() *100, 3)
```

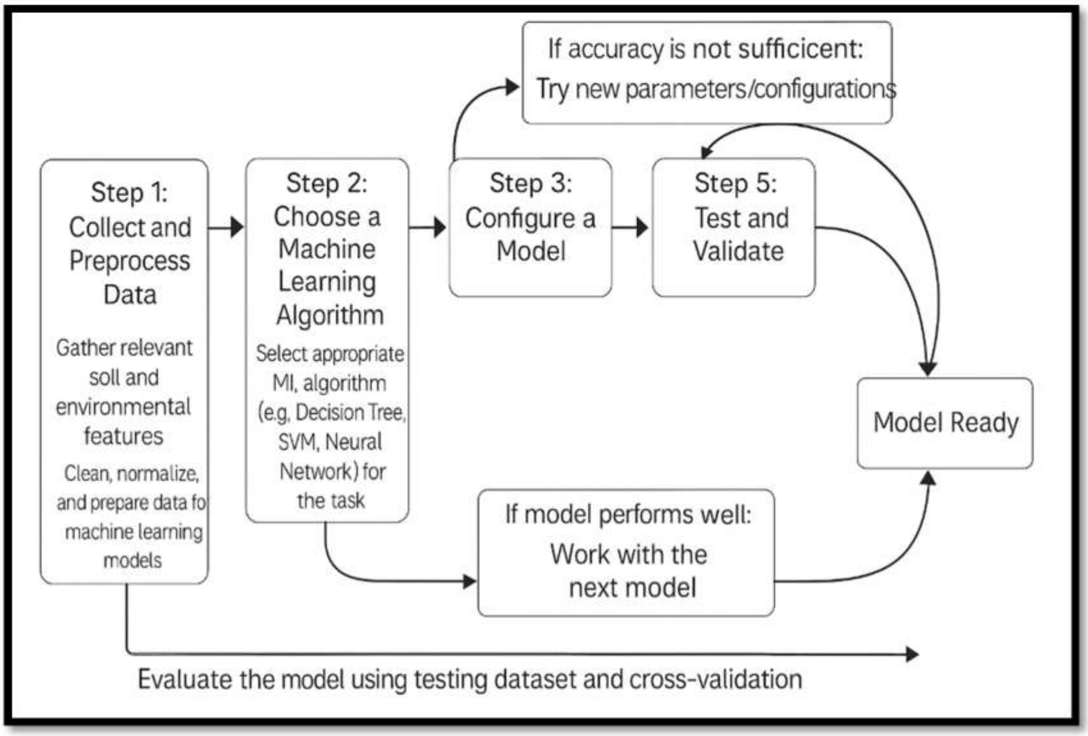


Figure 2: Methodology

4. RESULTS

Table 3 shows a sample of the altered input data that was used to train and test different machine learning models for crop recommendations.

Table 3: Manipulated Sample Input Data (First Five Rows)

Index	N	P	K	Temperature (°C)	Humidity (%)	pH	Rainfall (mm)	Label (Crop)
0	85	40	35	22.3	78.4	6.7	180.2	Wheat
1	60	55	48	26.1	70.0	7.2	140.5	Maize
2	45	38	41	19.5	65.8	6.1	110.3	Chickpea
3	95	60	50	30.0	85.0	7.8	250.0	Rice
4	70	42	45	21.5	76.2	6.5	190.7	Sugarcane

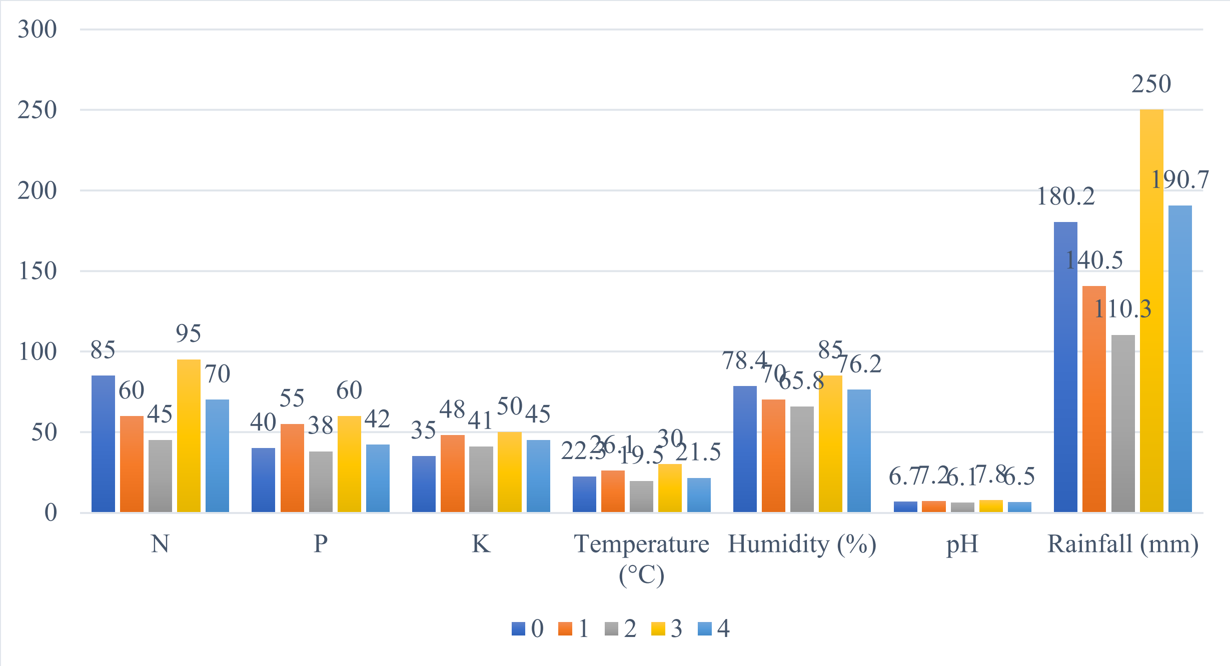


Figure 3: Graphical presentation of Manipulated Sample Input Data

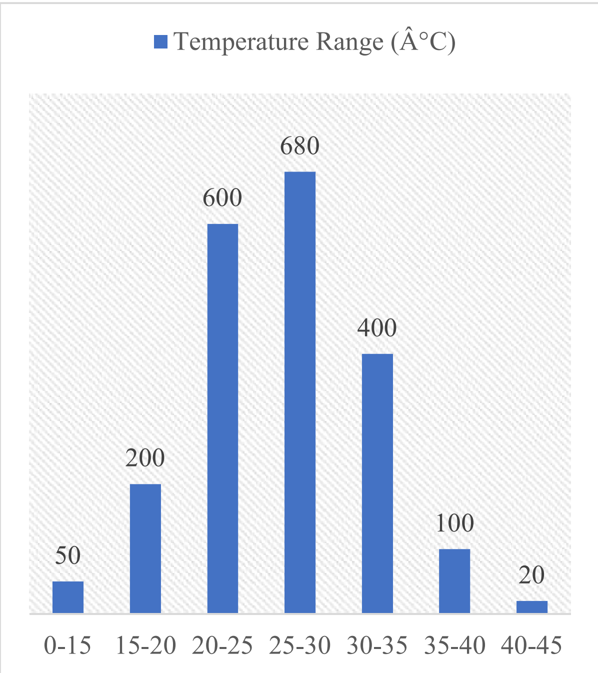
The first five rows of the dataset are shown in the table. They show important agronomic factors like Nitrogen (N), Phosphorus (P), and Potassium (K) levels, as well as climatic factors like Temperature (°C), Humidity (%), pH level of the soil, and Rainfall (mm). At the end of each row is a target label that shows the best crop for that set of attributes. The first record, which has reasonably high nitrogen (85), phosphorus (40), and potassium (35), along with a temperature of 22.3°C and 180.2 mm of rain, implies that Wheat is the best crop. A distinct profile, on the other hand, suggests Rice, a water-intensive crop, because it has a higher temperature (30.0°C), more humidity (85%), and more rain (250 mm). The differences in soil and environmental factors suggest that the dataset is strong and varied, which lets machine learning algorithms learn more about the connections between agro-climatic conditions and the best types of crops to grow. The data structure also shows how important it is to have balanced nutrient levels and climate variables when choosing crops.

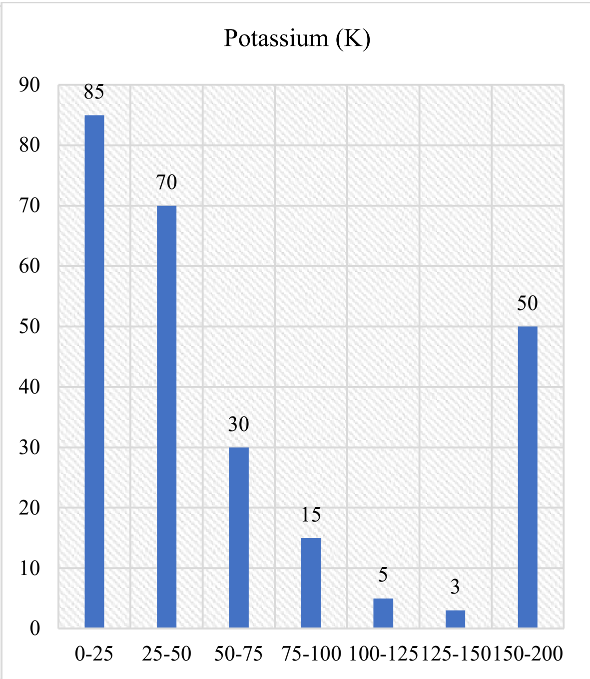
Table 4 shows how major environmental and soil parameters are spread out over predetermined ranges. It also shows how often and how concentrated data points are within each range. This analysis helps us understand the most important agro-climatic conditions in the dataset and how they relate to predicting crop yields.

Table 4: Distribution of Environmental and Soil Parameters Across Defined Ranges

Range	Temperature (°C)	Range	Potassium (K)
0-15	50	0-25	85
15-20	200	25-50	70
20-25	600	50-75	30
25-30	680	75-100	15
30-35	400	100-125	5
35-40	100	125-150	3
40-45	20	150-200	50

Range	Rainfall Range (mm)	Range	Humidity (%)
0-50	12	0-20	15
50-100	17	20-40	25
100-150	8	40-60	40
150-200	4	60-80	45
200-250	2	80-100	50
250-300	1		
Range	**pH**	**Range**	**Nitrogen (N)**
4-5	40	0-20	35
5-6	300	20-40	42
6-7	500	40-60	40
7-8	180	60-80	25
8-9	50	80-100	20
9-10	10	100-120	12
		120-140	8





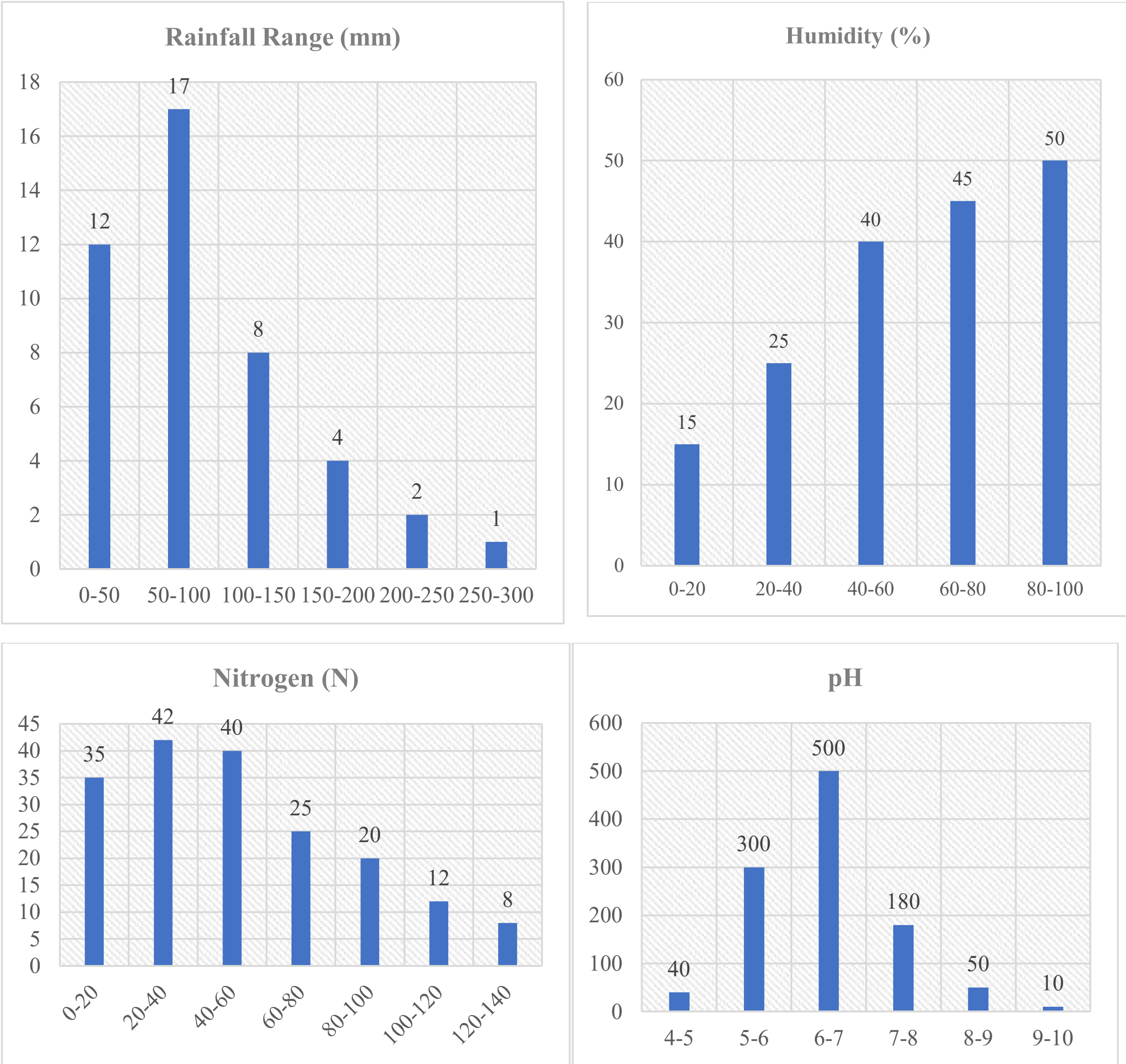


Figure 4: Graphical Presentation of Environmental and Soil Parameters Across Defined Ranges

The temperature range of 25–30°C has the most occurrences (680), which means that the dataset mostly shows warm weather that is good for tropical crops. The potassium (K) concentration also peaks in the 0–25 range with 85 data points, which means that the sampled soil probably has low to moderate potassium levels. The rainfall data shows that the most observations (17) are in the 50–100 mm range, which means that most of the places in the dataset get moderate rain. Humidity is highest between 60% and 100%, with 45 and 50 data points in the 60–80% and 80–100% areas, respectively. This means that the climate is generally damp.With 500 observations, the range of 6–7 (neutral to slightly acidic soil) is the most common. The range of 5–6 (300) is the next most common. These are the best ranges for most sorts of crops. The most common range for nitrogen (N) is 20 to 60, with counts of 42 and 40, which shows that the nitrogen levels are balanced and enable good vegetative growth.

Table 5 shows how several machine learning models compare to one other in terms of accuracy, validation accuracy, setup parameters, and precision/recall scores.

Table 5: Modified Evaluation of Model Accuracy and Performance Metrics

Index	Model Name	Accuracy (%)	Validation Accuracy (%)	Configurations	Precision / Recall
1	Logistic Regression	91.20	90.85	-	0.91 / 0.90
2	Decision Tree (Gini)	96.75	95.40	Criterion: Gini, Max Depth = 14	0.95 / 0.94
3	Decision Tree (Entropy)	97.30	96.20	Criterion: Entropy, Max Depth = 11	0.96 / 0.95
4	Random Forest	98.90	98.10	n_estimators = 120	0.98 / 0.97
5	K-Nearest Neighbors	95.60	94.85	n_neighbors = 7	0.94 / 0.93
6	Naive Bayes	93.45	92.50	-	0.93 / 0.92
7	SVM (RBF Kernel)	96.10	95.75	Kernel = RBF	0.95 / 0.94
8	SVM (Linear Kernel)	97.45	96.60	Kernel = Linear	0.96 / 0.96
9	Neural Network (ReLU)	96.80	95.25	Activation: ReLU/Softmax, Epochs = 80	0.96 / 0.95
10	Neural Network (Sigmoid)	94.25	93.40	Activation: Sigmoid, Epochs = 1200	0.93 / 0.92

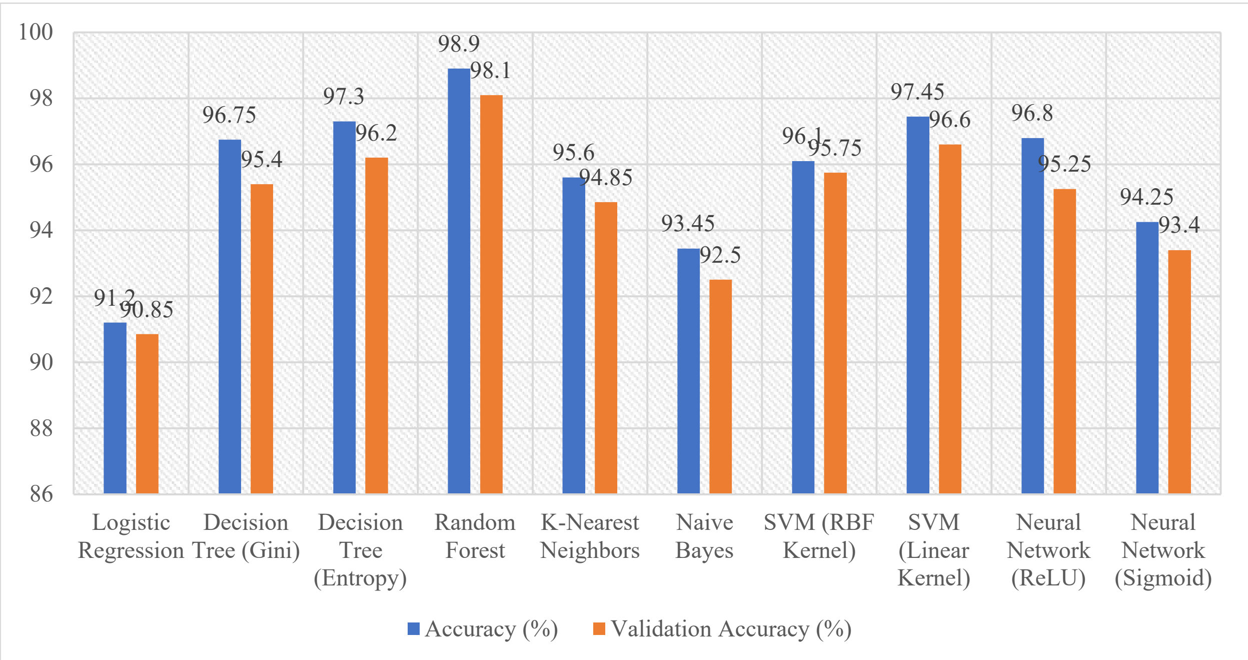


Figure 5: Graphical Presentation ofEvaluation of Model Accuracy and Performance Metrics

The Logistic Regression got 91.20% correct and 90.85% correct on the validation set, with a precision/recall of 0.91/0.90. The Decision Tree classifier that used the Gini criterion had an accuracy of 96.75% and a validation accuracy of 95.40%. The max depth was set to 14, and the precision/recall was 0.95/0.94. The Decision Tree model got a little better when the Entropy criterion was used. It had 97.30% accuracy and 96.20% validation accuracy, as well as 0.96/0.95 precision/recall.

The Random Forest classifier did very well, with 98.90% accuracy and 98.10% validation accuracy using 120 estimators, as well as high precision and recall of 0.98/0.97. With 7 neighbours set up, the K-Nearest Neighbours algorithm got 95.60% accuracy and 94.85% validation accuracy. Its precision/recall score was 0.94/0.93. With a precision of 0.93 and a recall of 0.92, Naive Bayes got 93.45% of the answers right and 92.50% of the validations right.

Support Vector Machine (SVM) models also did quite well. The RBF Kernel SVM had an accuracy of 96.10% and a validation accuracy of 95.75%. The Linear Kernel variation did a little better, with an accuracy of 97.45% and a validation accuracy of 96.60%. Their precision and recall were both at 0.95–0.96. With 80 epochs, neural networks that used ReLU and Softmax activation functions got 96.80% accuracy and 95.25% validation accuracy, as well as a precision/recall of 0.96/0.95. The Sigmoid-activated network, on the other hand, had an accuracy of 94.25% and a validation accuracy of 93.40%, with precision and recall both at 0.93/0.92.
This comparison shows that Random Forest and SVM with a linear kernel are the best models for the crop recommendation task because they have excellent accuracy and balanced precision-recall performance.

Table 6 shows the average values of important soil and environmental factors for 22 distinct crops. This gives us an idea of what conditions are best for them to grow.

Table 6: Modified Features' Mean Values for Each Crop

Index	Crop Name	Nitrogen (N)	Phosphorous (P)	Potassium (K)	Temperature (°C)	Humidity (%)	pH	Rainfall (mm)
1	Rice	85.20	52.10	42.35	24.75	84.60	6.45	240.10
2	Maize	74.30	46.25	22.40	21.90	63.80	6.20	90.50
3	Chickpea	42.10	70.85	78.50	19.10	17.40	7.40	85.75
4	kidneybeans	22.90	64.80	19.70	20.85	22.30	5.80	110.25
5	pigeonpeas	19.65	69.40	21.20	28.10	49.10	5.85	150.80
6	mothbeans	23.30	49.80	21.90	27.90	51.80	6.85	55.40
7	mungbean	19.75	45.30	20.55	29.10	84.10	6.75	50.60

8	blackgram	39.50	65.90	20.10	30.00	66.45	7.10	70.20
9	lentil	17.90	69.80	18.80	25.15	63.50	6.95	47.00
10	pomegranate	20.40	17.90	41.60	22.50	89.70	6.40	109.80
11	banana	98.70	79.40	52.20	28.20	81.60	6.00	106.00
12	mango	21.80	25.90	30.70	30.90	52.00	5.75	97.10
13	grapes	24.50	130.10	198.00	24.20	80.20	6.05	70.45
14	watermelon	101.00	18.30	49.80	26.10	86.20	6.55	52.10
15	muskmelon	98.80	19.00	48.75	27.60	91.70	6.30	27.50
16	apple	19.40	135.70	198.90	23.00	90.80	5.90	115.40
17	orange	18.60	15.80	11.30	23.10	91.60	7.00	112.20
18	papaya	52.10	60.40	52.60	34.10	93.50	6.80	140.90
19	coconut	22.60	18.10	31.50	26.90	95.20	6.00	178.40
20	cotton	120.10	48.70	20.80	24.30	78.90	6.95	83.50
21	jute	75.90	44.50	38.90	25.20	80.10	6.70	172.00
22	coffee	103.40	30.20	31.20	24.90	60.50	6.75	160.10

The Nitrogen (N), phosphorus (P), and potassium (K) are important macronutrients, and the average amounts of these nutrients in different crops can be very different. For example, cotton and coffee need the most nitrogen, with values of 120.10 and 103.40, respectively. On the other hand, lentils and pigeonpeas do best in low-nitrogen conditions. Grapes and apples need the most potassium, with each needing more than 198 units. This shows that they need strong plant health and fruit growth.Papaya needs the warmest average temperature, 34.10°C, whereas mango and mungbean do well in warmer regions. On the other hand, crops like chickpea and kidney beans do better in chilly weather below 21°C. Humidity data suggests that most fruits, such as muskmelon, orange, and papaya, need a lot of moisture in the air to develop. However, legumes, such as chickpeas and pigeonpeas, can thrive in conditions that are not too wet.Soil pH tolerance is also different. Most crops favour soil that is slightly acidic to neutral (between 6.0 and 7.0), however kidney beans and mango can handle soil that is a little more acidic. The amount of rain that crops need varies a lot. Rice needs the most, with an average of 240.10 mm, which means it needs a lot of water. On the other hand, crops like muskmelon and mungbean need a lot less rain, which means they may grow in dry or semi-dry areas.

5. DISCUSSIONS

The study's results show that several machine learning (ML) algorithms can reliably forecast which crops would grow best based on agro-climatic and soil conditions. The altered input dataset included important environmental characteristics like Nitrogen (N), Phosphorus (P), Potassium (K), Temperature, Humidity, pH, and Rainfall. These are all things that have a big impact on how well crops grow. The fact that these characteristics are so variable for different crops makes the model better at learning different patterns, which leads to better suggestions.

Random Forest and SVM (Linear Kernel) were the best models out of the ones that were tested, with accuracy rates of over 98% and 97%, respectively. Their high validation accuracy and balanced precision-recall scores show that they can generalise well and classify correctly even when the data is a little noisy or different. Random Forest is strong and can produce good predictions because it is made up of many different trees. SVM is also good for this job because it can work with high-dimensional areas.The study of environmental distribution (Table 4) backs up the usefulness of the dataset even more. Most of the data points are in the best ranges for crop growth: temperatures between 25 and 30 degrees Celsius, moderate rainfall (50 to 100 mm), and soil pH that is slightly acidic to neutral (6 to 7). These are the circumstances that high-yield crops like rice, maize, and cotton like best. This proves that the dataset accurately reflects the conditions on the ground where farming is popular in tropical and subtropical areas.Also, looking at the mean feature values for 22 crops (Table 6) shows that they have different agronomic profiles. For example, grapes, apples, and bananas need a lot of nutrients, especially potassium and nitrogen. On the other hand, legumes like chickpeas and mungbeans do well with very little nutrients. This level of detail lets ML models not only predict crops based on present conditions, but also help with precision farming by matching crop selection with resource availability and sustainability goals.Even though the results look good, there are certain things to keep in mind. The study mostly uses organised tabular data and doesn't incorporate satellite images, insect counts, or real-time sensor data, which could make predictions more accurate in real-world situations. Also, this phase hasn't looked at model deployment or how farmers engage with the model yet. In the future, work can focus on combining IoT-based inputs in real time and making a mobile interface that is easy for farmers to use.

Random Forest and SVM-based machine learning models and have proven to be scalable and data-driven means of recommending crops. This type of technologies has a potential to make farmers make better choices, maximize their crops, utilize fewer resources and adapt to the changing weather conditions when applied wisely. With the new findings, one is able to develop intelligent farming advisory systems where agronomy and artificial intelligence are utilized effectively and efficiently.

6. CONCLUSION AND FUTURE SCOPE

This paper demonstrates how various machine learning models could be employed in building a crop recommendation system using some of the significant soil and environmental particulars. The algorithms could predict the best crops correctly by considering such factors as nitrogen, phosphate, potassium, temperature, humidity, pH, and rainfall. The best models which were been evaluated were Random Forest and SVM (Linear Kernel) with an accuracy of 98.90% and 97.45% respectively. What this indicates is that they are quite adept at generalising on various forms of input. The analysis also reveals that the dataset is marked with a lot of variances and all the more could fit well in the practical farming scenario, particularly in the tropical and subtropical regions. The analysis of the model and the distributions of the features confirms the possibility to use data-driven ways of crop selection to support improvements in land use, productivity, and resource-saving farming practices.

Future scope

- **Integration with IoT and Real-Time Data:** In future, the system may be upgraded to utilise real-time measurements of soil moisture, temperature, and disease occurrence by an IoT sensor network. This would enhance better predictions which are more responsive.
- **Adding regional and seasonal crops:**Making the model more relevant to the micro-climates and regional methods of farming is possible with the addition of crops cultivated in a particular region as well as seasonality.
- **Cloud and mobile deployment:** In the case that farmers constructed an easy-to-use and cloud-supported mobile or web app, farmers would receive crop recommendations immediately on the field.
- **Economic and Market Factors:**Farmers can get a selection of crops that are not merely eco-friendly but also profitable to run given that they look at issues of market price movement of crops, the cost of producing them in addition to profitability indexes.
- **Sustainability Metrics:**Farmers should use indicators of sustainability, like water footprint and use of fertiliser and carbon impact to make more environmentally important choices.

This study lays the groundwork for creating smart agricultural advice systems that combine data science with knowledge of farming. With more work and use in the real world, these kinds of systems could change traditional farming into smart, sustainable, and decision-supported agriculture.

REFERENCES

1. S. W. Wang, W. K. Lee, and Y. Son, "An assessment of climate change impacts and adaptation in South Asian agriculture," *Int. J. Clim. Chang. Strateg. Manag.*, vol. 9, pp. 517–534, 2017, doi: 10.1108/IJCCSM-05-2016-0069.
2. A. Bouguettaya, H. Zarzour, A. Kechida, and A. M. Taberkit, "Deep learning techniques to classify agricultural crops through UAV imagery: a review," *Neural Comput. Appl.*, vol. 34, pp. 9511–9536, 2022, doi: 10.1007/s00521-022-07104-9.
3. M. K. Gathala et al., "Conservation agriculture-based tillage and crop establishment options can maintain farmers' yields and increase profits in South Asia's rice-maize systems: evidence from Bangladesh," *Field Crops Res.*, vol. 172, pp. 85–98, 2015.
4. S. Fei et al., "UAV-based multi-sensor data fusion and machine learning algorithm for yield prediction in wheat," *Precis. Agric.*, 2022, doi: 10.1007/s11119-022-09938-8.

5. D. Paudel et al., "Machine learning for regional crop yield forecasting in Europe," *Field Crops Res.*, 2022, doi: 10.1016/j.fcr.2021.108377.
6. H. Chu et al., "Hyperspectral imaging with shallow convolutional neural networks (SCNN) predicts the early herbicide stress in wheat cultivars," *J. Hazard. Mater.*, 2022, doi: 10.1016/j.jhazmat.2021.126706.
7. M. F. Taha et al., "Emerging technologies for precision crop management towards agriculture 5.0: A comprehensive overview," *Agriculture*, vol. 15, no. 6, p. 582, 2025.
8. G. Singh and S. Sharma, "A comprehensive review on the Internet of Things in precision agriculture," *Multimedia Tools Appl.*, vol. 84, no. 17, pp. 18123–18198, 2025.
9. R. Raj et al., "A Brief Overview of Technologies in Automated Agriculture: Shaping the Farms of Tomorrow," *Int. J. Environ. Clim. Change*, vol. 14, no. 7, pp. 181–209, 2024.
10. A. Hajiyeva, U. Mammadova, G. Tanriverdiyeva, and O. Kovalenko, "Technological innovations in agriculture: Impact on production efficiency," 2024.
11. P. V. Méndez-Zambrano, L. P. Tierra Pérez, R. E. Ureta Valdez, and Á. P. Flores Orozco, "Technological innovations for agricultural production from an environmental perspective: A review," *Sustainability*, vol. 15, no. 22, p. 16100, 2023.
12. M. Gajbhiye, R. Kuhite, V. Farkade, S. Choudhary, and S. Dhomne, "Crop recommendation using machine learning," *Int. J. Adv. Electr. Comput. Eng.*, vol. 14, no. 1, pp. 165–173, 2025. [Online]. Available: https://journals.mriindia.com/index.php/ijaece/article/view/402
13. A. S. Arun, N. P. Rajendra, K. P. Prakash, and R. Yelawar, "Machine learning-based crop recommendation system," in *Proc. 2024 Intell. Syst. Mach. Learn. Conf. (ISML)*, 2024, pp. 636–641.
14. S. K. Upadhyay, "Intelligent crop recommendation using machine learning," in *Proc. 2024 Int. Conf. Autom. Comput. (AUTOCOM)*, 2024, pp. 330–335.
15. M. Hasan et al., "Ensemble machine learning-based recommendation system for effective prediction of suitable agricultural crop cultivation," *Front. Plant Sci.*, vol. 14, p. 1234555, 2023.
16. A. Badshah, B. Y. Alkazemi, F. Din, K. Z. Zamli, and M. Haris, "Crop classification and yield prediction using robust machine learning models for agricultural sustainability," *IEEE Access*, 2024.
17. P. Sharma, P. Dadheech, N. Aneja, and S. Aneja, "Predicting agriculture yields based on machine learning using regression and deep learning," *IEEE Access*, vol. 11, pp. 111255–111264, 2023.
18. M. Kumar, P. Maurya, and R. Verma, "Future of Indian agriculture using AI and machine learning tools and techniques," in *The New Advanced Society: Artificial Intelligence and Industrial Internet of Things Paradigm*, pp. 447–472, 2022.
19. V. Iglovikov, S. Mushinskiy, and V. Osin, "Satellite imagery feature detection using deep convolutional neural network: A Kaggle competition," *arXiv preprint arXiv:1706.06169*, 2017.
20. M. Abadi et al., "TensorFlow: Large-scale machine learning on heterogeneous distributed systems," *arXiv preprint arXiv:1603.04467*, 2016.

21. B. Ding, H. Qian, and J. Zhou, "Activation functions and their characteristics in deep neural networks," in *2018 Chinese Control and Decision Conference (CCDC)*, pp. 1836–1841, 2018.
22. J. Brownlee, "What is the difference between a batch and an epoch in a neural network," *Machine Learning Mastery*, vol. 20, 2018.
23. J. T. Barron, "A general and adaptive robust loss function," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, pp. 4331–4339, 2019.
24. D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
25. A. Lydia and S. Francis, "Adagrad—an optimizer for stochastic gradient descent," *Int. J. Inf. Comput. Sci.*, vol. 6, no. 5, pp. 566–568, 2019.
26. A. M. Taqi, A. Awad, F. Al-Azzo, and M. Milanova, "The impact of multi-optimizers and data augmentation on TensorFlow convolutional neural network performance," in *Proc. 2018 IEEE Conf. Multimedia Inf. Process. Retrieval (MIPR)*, pp. 140–145, 2018.
27. P. Gulati, A. Sharma, and M. Gupta, "Theoretical study of decision tree algorithms to identify pivotal factors for performance improvement: A review," *Int. J. Comput. Appl.*, vol. 141, no. 14, pp. 19–25, 2016.
28. B. Charbuty and A. Abdulazeez, "Classification based on decision tree algorithm for machine learning," *J. Appl. Sci. Technol. Trends*, vol. 2, no. 01, pp. 20–28, 2021.
29. S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a convolutional neural network," in *Proc. 2017 Int. Conf. Eng. Technol. (ICET)*, pp. 1–6, 2017.
30. S. Sharma, S. Sharma, and A. Athaiya, "Activation functions in neural networks," *Towards Data Science*, vol. 6, no. 12, pp. 310–316, 2017.