Crop Recommendation Analyzing Numerous Factors Through Machine Learning

by

Examination Roll:

A project report submitted to the Institute of Information Technology
in partial fulfilment of the requirements for the degree of
Professional Masters in Information Technology

Supervisor:

Institute of Information Technology
Jahangirnagar University
Savar, Dhaka-1342
August 2025

DECLARATION

I affirm that this project work is based on the results found by me. All sources of Information utilized in the project have been properly credited by reference. This thesis has never been submitted for any degree or certificate at any institution or institute before.

Roll:

ACKNOWLEDGEMENTS

I would like to take this opportunity to express my sincere appreciation and gratitude to everyone who supported and contributed to the completion of this thesis.

This work was conducted under the supervision of Dr. Md. Fazlul Karim Patwary, Professor at the Institute of Information Technology (IIT), Jahangirnagar University, Savar, Dhaka. I am truly thankful for his constant guidance, insightful suggestions, and invaluable support throughout the entire research process. His willingness to provide resources, including relevant books and research materials, played a vital role in the successful completion of this study.

I am also thankful to all those who have assisted me in various ways—whether directly or indirectly. Their help and encouragement were instrumental during this journey.

Lastly, I would like to extend my heartfelt thanks to the faculty and staff of IIT, Jahangirnagar University, as well as my friends, for their continuous support and motivation throughout this endeavor.

ABSTRACT

Agricultural crop recommendation systems rely on several input characteristics. This presents a hybrid model for crop recommendation, taking into account numerous factors like soil type, rainfall, groundwater level, temperature, fertilizers, pesticides, and season. The recommender model is constructed as a hybrid using a classifier machine learning method. The system will propose the crop based on the relevant characteristics. A technology-driven crop recommendation system for agriculture assists farmers in enhancing crop production by suggesting appropriate crops for their land based on geographic and meteorological criteria. The suggested hybrid recommender model is successful in suggesting an appropriate crop. The update of crop yield production values has significant practical importance for directing agricultural output and informing farmers of fluctuations in crop market rates. This study aims to utilize the crop selection approach to address various agricultural and farmer-related issues. This enhances the Indian economy by optimizing the yield rate of agricultural produce. Diverse categories of land conditions. The quality of the crops is assessed by a rating mechanism. This technique also communicates the rates of low-quality and high-quality crops. The use of an ensemble of classifiers facilitates improved decision-making in predictions via the application of numerous classifiers. A ranking procedure is used for decision-making to pick the outcomes of the classifiers. This approach is used to forecast the cost of crops produced for subsequent usage.

Keywords:Machine Learning, Crop Prediction, Random Forrest, Hyperparameter Tuning

LIST OF ABBREVIATIONS

ML	Machine Learning
IoT	Internet of Things
SDN	Software Defined Network
IDS	Intrusion Detection System
FIS	Fuzzy Inference System
SVM	Support Vector Machine
ANN	Artificial Neural Network
KNN	K-Nearest Neighbor

LIST OF FIGURES

Figure

LIST OF TABLES

Table

TABLE OF CONTENTS

CHAPTER I

Introduction

1.1 Overview

Society, technology, and global pandemics like COVID-19 have all contributed to rapid transformations in the agricultural, crop management, and farming industries in recent years (Pushparaj Naik). The rise of self-managed farming is only one example of how these innovations have drastically altered the way farmers run their businesses. The proliferation of digital resources, such as cellphones and the internet, has empowered farmers to assume more responsibility for their farming operations. This is an example of the trend towards data-driven precision agriculture, which is replacing more conventional methods. Modern agricultural success relies on the careful use of fertilisers according to crop needs and soil conditions. There is an immediate need for reliable, objective systems that can classify, evaluate, and alter agricultural methods as more and more farmers take charge of their own operations. Issues including poor planning, lower output, and a greater likelihood of crop damage might arise in self-managed farming due to the absence of expert advice. In light of these challenges, this research endeavour aims to develop a dependable and versatile motion correction model. This concept provides a fresh take on autonomous farming by functioning as a standalone system that can assess and modify farming operations in real-time. We want to enhance the overall quality, safety, and effectiveness of self-managed farming, which has the potential to revolutionise the agricultural business.

1.2 Problem Statement

The first issue is that conventional farming practices often make poor decisions without enough data, leading to ineffective fertiliser usage and poor crop yields. Not having specific crop and fertiliser suggestions for each farm's requirements is a major obstacle to sustainable agricultural approaches. It is difficult for farmers to determine the optimal fertiliser mix and which crops will thrive on their individual fields. Reduced production and worsened environmental damage result from farmers' inability to make educated choices in the absence of real-time data and advanced analytics. Without precise and regional recommendations for crops and fertilisers, it is difficult to maximise agricultural productivity. Unfortunately, most of the current recommendation algorithms are based on broad statistics that fail to take into account the fact that different farms have somewhat different soil types and climates.

1.3 Motivation

The world's population relies on agriculture, the oldest sector, to provide them with food. It has developed to maximise efficiency, draw in more players, and improve overall quality standards via modernisation and technology integration. A huge problem for farmers, meanwhile, is the impending loss of farmland as a result of urbanisation. When you consider that we must increase food production by more than 70% by the year 2050 to meet the demands of an ever-increasing population, you can see why creative solutions are crucial. If we want agriculture to be sustainable in the face of changing global demands, we need an automated system that can maximise harvest production with minimal inputs of resources.

1.4 Aim and Objective

I Specific goals of this thesis that should be mentioned:

- comprehensively document critical soil parameters, weather patterns, and historical crop yields from a range of agricultural sites.
- Create and refine machine learning algorithms, such as supervised learning techniques like Decision Trees, Support Vector Machines, and Neural Networks, in order to build prediction models.
- Develop algorithms to improve fertiliser composition and provide precise administration methods, ensuring focused crop development via optimal nutrient absorption.
- rigorous testing on the accuracy and performance of the developed models using a range of agricultural datasets.

1.5 Research Outline

Rest of the report is structured as follows: In **Chapter II** a literature study on related work is given including explanations for the most important terms used in this project concept and different models, reason behind choosing Random forest Model. **Chapter III** introduces system model including system architecture, algorithm and flowchart of working procedure of entire system model. **Chapter IV** explains the Performance Analysis and evaluation Random Forest Classifier Model, Classification Report of Random Forest Classifier Model, Classification Report after Hyperparameter Tuning **Chapter V** future work and conclusion is mentioned.

CHAPTER II

Literature Review

2.1 Overview

Several widely utilised methodologies, such as SVM, Decision Trees, and Random Forest algorithms, are examined in the literature on crop and fertiliser recommendation using machine learning. Soil type, climatic conditions, and nutrient levels are just a few of the factors that researchers have considered across several datasets. Research has highlighted the smart crop management potential of artificial intelligence and the internet of things (IoT) in precision agriculture. Agricultural practices, yields, and farmers' access to data-driven decision-making tools are the targets of these endeavours.

Omen Priyadharshini, Aayush Kumar, Swapneel Chakraborty, and Priyadharshini A Sowing season, soil, and geographical location are just a few of the variables that Rajendra Pooniwala included in his approach to aid farmers in crop choices. Precision agriculture, which emphasizes site-specific crop management, is also expanding in underdeveloped nations and is being used with current agricultural technology.[1]

An intelligent system called AgroConsultant was proposed by Zeel Doshi, Subhash Nadkarni, Rashi Agrawal, and Neepa Shah. Its purpose is to help Indian farmers choose the best crop to grow based on factors like sowing season, farm location, soil characteristics, and environmental conditions like temperature and rainfall.[2]

The suggested approach by Suresh Rathod and Avinash Devare is defined by a farm-based soil database, crop data supplied by agricultural specialists, and the attainment of factors like soil via data acquired in soil testing labs. In order to provide an efficient and accurate crop recommendation based on site-specific parameters, the recommendation system will leverage data collected from soil testing labs, run an ensemble model using the majority voting approach, and use support vector machines (SVMs) and artificial neural networks (ANNs) as learners.[3]

One solution offered by D. Anantha Reddy, Bhagyashri Dadore, and Aarti Watekar is a recommendation system based on soil parameters. These methods provide a categorized picture based on ground truth statistical data, which includes things like weather, crop yield, and crops broken down by state and district so that one can anticipate how much a certain crop would produce given a given set of meteorological conditions.[4]

Pabasara M.G.P. proposed we provide a crop recommendation system that calculates, based on the user's input, which crop varieties will thrive in a given region by collecting data on relevant environmental elements affecting plant development and integrating it with the system's trained sub-models.[5]

In this study, M. Kalimuthu, P. Vaishnavi, and M. Kishore show how novice farmers might benefit from using machine learning—a cutting-edge technology for crop prediction—to seed suitable crops. A supervised learning technique called Naive Bayes proposes a method to do this. Here, we gather information on the seeds of the crops, including the right conditions (such as temperature, humidity, and moisture content), so that the crops may develop to maturity. Along with the software, an Android app is also in the works. Entering factors like temperature and location will immediately start the prediction process in this program. Users are urged to do so.[6]

E. Keertanaa, P. S. Vijayabaskar, and R. Sreemathi developed a model to measure soil fertility. Based on the reading from the sensor, it also recommends what crop needs to be planted. It also gives you graphs with crop data broken down by area. It goes on to say which fertilizer is needed for the soil to boost crop yields. The farmer is able to assess the yard's fertility and plant more productive crops as a result. Additionally, it details the fertilizer that should be used with the soil.[7]

In their discussion of the potential future of machine learning in palm oil yield prediction, Nuzhat Khan covered a lot of ground, including the use of remote sensing, the identification of plant growth and diseases, the use of mapping and tree counting, and the selection of optimal features and algorithms. Last but not least, after reviewing the relevant literature, we have developed a model for the future of palm oil yield prediction using machine learning. By taking on new research problems in crop yield prediction analysis and creating a very good model for palm oil yield prediction with minimum computing effort, this technology will live up to its promise.[8]

A research on deep learning and its hybrid approaches, including artificial neural networks, recurrent neural networks, and P. Natesan, R. Thamilselvan, and S. Santhoshkumar were presented. It was useful in determining how AI technology contributes to increased harvest yields. In agriculture, the study elucidates the concept and need of recurrent neural networks and hybrid networks. Additionally, it demonstrates how it surpasses competing networks, including convolutional neural networks and artificial neural networks. We examined the data and used them to form conclusions about the future.[9]

The input and hidden layer weights and biases are initially set at random in a traditional ANN model. In this MLR-ANN hybrid model, the bias and weights of the input layers are initially set using the MLR coefficients and bias rather than random values. Utilizing performance criteria, the hybrid model's prediction accuracy is contrasted with that models. The amount of time needed to process data using hybrid MLR-ANN and traditional ANN was determined. The results demonstrate that, in comparison to the traditional models, the suggested hybrid MLR-ANN model provides superior accuracy.[10]

At its heart, Smart Crop Recommendation System is based on precision agriculture, a management strategy that employs a variety of ML models to gather, analyze, and process individual, geographical, and temporal data. Hence, DT, SVM, KNN, LGBM, and RF were used as ML models for crop prediction. Out of all the models, RF achieved the highest accuracy rate of 99.24%. Thus, the suggested method may advise on the optimal crop for given soil resources based on characteristics . Therefore, this approach may aid the government, farmers, and many others involved in agriculture in making important choices.[11]

G. Mariammal suggested comparing different wrapper feature selection methods. According to the findings of the experiments, the method that combines the Adaptive Bagging classifier with Recursive Feature Elimination is the most effective.[12]

Using a wide range of soil and environmental factors Ch. Teja Naidu explore many approaches to forecasting agricultural output. The goal of this project is to train a machine learning model to predict future outcomes. In addition to enhancing production, Reduced crop loss and optimal selling prices are two ways in which farmers may reap the rewards of yield estimate. The goals and nature of the research determine whether a machine learning model should be predictive or descriptive.[13]

Using machine learning, Himanshu Pant were able to forecast four widely grown crops in India. It is possible to apply inputs like fertilizers in a variety of ways based on estimated crop and soil demands when site-specific crop production predictions are made. In this research, we create a model to spot trends in the data and apply it to crop forecasting using Machine Learning techniques. The use of machine learning for the forecast of the four most widely grown crops in India is examined in this research. Wheat, potatoes, rice (paddy), and maize are all examples of such crops.[14]

Using a variety of machine learning approaches, Aruvansh Nigam; Saksham Garg; Archit Agrawal; and Parul Agrawal aim to forecast the crop's production. Mean absolute error is used to compare the results of various methods. In order to maximize productivity, farmers may use the predictions provided by machine learning algorithms to choose which crops to cultivate. These algorithms take into account variables such as area, rainfall, temperature, and more.[15]

In this supplementary study, Cagatay Catal found that among these deep learning algorithms, Convolutional Neural Networks (CNN) were the most popular, followed by Deep Neural Networks (DNN) and Long Short-Term Memory (LSTM).[16]

In their discussion of recent advances in machine learning-based methods for precise nitrogen status measurement and crop yield prediction, Anna Chlingaryan, Salah Sukkarieh, and Brett Whelan covered work done in the last fifteen years. The research comes to the conclusion that improved crop and environment status estimates and decision-making would be made possible by cost-effective and comprehensive solutions provided by the fast improvements in sensing technologies and ML approaches. In the not-too-distant future, precision agriculture (PA) may incorporate more customized use of sensor platforms and ML methods, the integration of various sensor modalities with expert knowledge, and the creation of hybrid systems integrating various ML and signal processing technologies.[17]

A. Parvathi, S. Bangaru Kamatchi innovative hybrid system that has been suggested combines case-based reasoning with the collaborative filtering approach. This model is unique because it uses a hybrid recommender system to analyze agricultural data on a district-by-district basis in order to forecast future weather patterns and provide crop recommendations based on those predictions, all while taking into account the district's unique agricultural pattern.[18]

The alkalinity of the soil is calculated by S. Bhanumathi, M. Vineeth, and N. Rohit after they examined several relevant factors, such as location and pH value. The data will undergo analysis of all these qualities before being trained using a variety of appropriate machine learning methods to build a model. The system is equipped with a model that can accurately estimate crop production and provide users with advice about the appropriate fertilizer ratio. These recommendations are based on the land's atmospheric and soil factors, which in turn boost crop yield and farmer income.[19]

With the help of Devdatta A. Bondre and Mr. Santosh Mahagaonkar, a method is developed and put into action to forecast agricultural productivity using historical data. This is accomplished by analyzing agricultural data using machine learning algorithms such as Random Forest and Support Vector Machine, which then prescribe fertilizer that is appropriate for each crop. Future crop production predictions are the primary emphasis of this article, which centers on the development of a prediction model. It provides a concise review of the literature on agricultural production prediction using ML methods.[20]

A data mining strategy is proposed by Rushika Ghadge and Priya R. L. to assist farmers in evaluating soil quality. In order to optimize crop output, the system recommends the right fertilizer based on the soil type and checks the soil quality to determine which crops are fit for cultivation. [21]

Traditional methods of crop prediction may have resulted in too optimistic assessments of both crops and available land, according to research by Shubham Prabhu, Prem Revandekar, Swami Shirdhankar, and Sandip Paygude. in addition, these methods often rely on labor- and expense-intensive agricultural data collection in the field. So, we're aiming to develop a time- and labor-saving automated soil testing system that can evaluate soil samples and deliver useful agricultural information. The amount of precipitation within a certain location and the fertility of the soil are both taken into account to complete this crop forecast.[22]

Rainfall has changed, with an increasing tendency during the monsoon and essentially little change during other seasons, according to Md Abiar Rahmais and Robert Macnee. Droughts are becoming more common and diminishing DTR is a threat to rice production, particularly rainfed rice. It is recommended to plant heat-and stress-tolerant rice types that need less water for irrigation. In order to lessen the impact of the unfavorable weather, studies on shifting the crop schedule and cropping pattern are required.[23]

CHAPTER III

System Model

3.1 Proposed crop recommendation system

The Figure depicts a standard machine learning pipeline, starting with raw data acquisition and advancing through several critical phases. The data is first subjected to preprocessing to cleanse and ready it for analysis. This is followed by analysis and transformation, during which the data is further refined and characteristics may be engineered. The processed data is then divided into training and testing subsets, guaranteeing an impartial model assessment. The training data is used to construct and train a prediction model, whilst the test data is allocated for evaluating the model's generalisation capability. Subsequent to training, the model's efficacy is assessed using the test data, and the outcomes are scrutinised to ascertain the overall success of the pipeline. This sequential process guarantees that each phase enhances the preceding one, resulting in a solid and dependable machine learning solution.

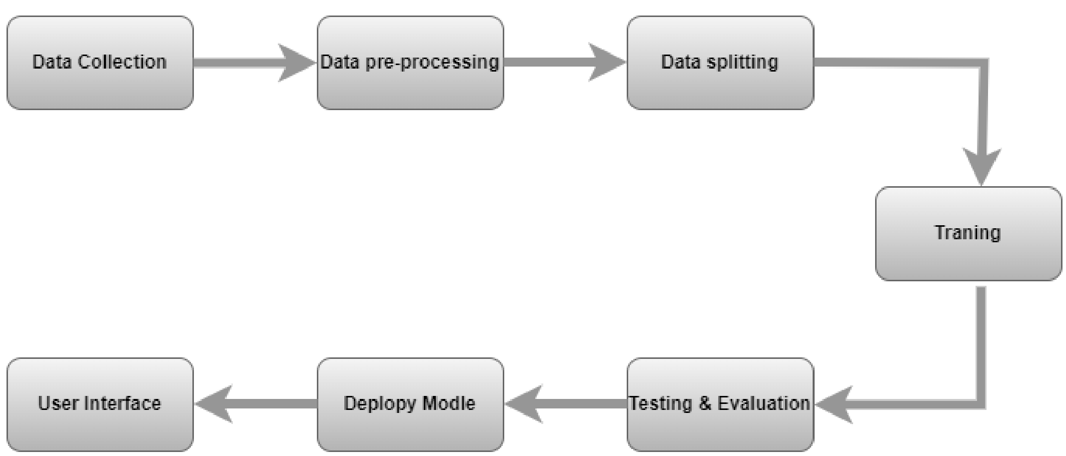

Figure 3.1: Proposed crop recommendation system

3.2 EDA Analysis

3.2.1 Distribution of Crop Labels

The figure is a horizontal bar chart showing the sample count for each crop label. Each crop type in the dataset has an identical sample size of 100. This equitable distribution guarantees that machine learning models trained on this dataset will remain impartial to any specific crop, facilitating fair and precise predictions across all crop varieties. Consistency in sample counts is essential for effective model training and assessment.

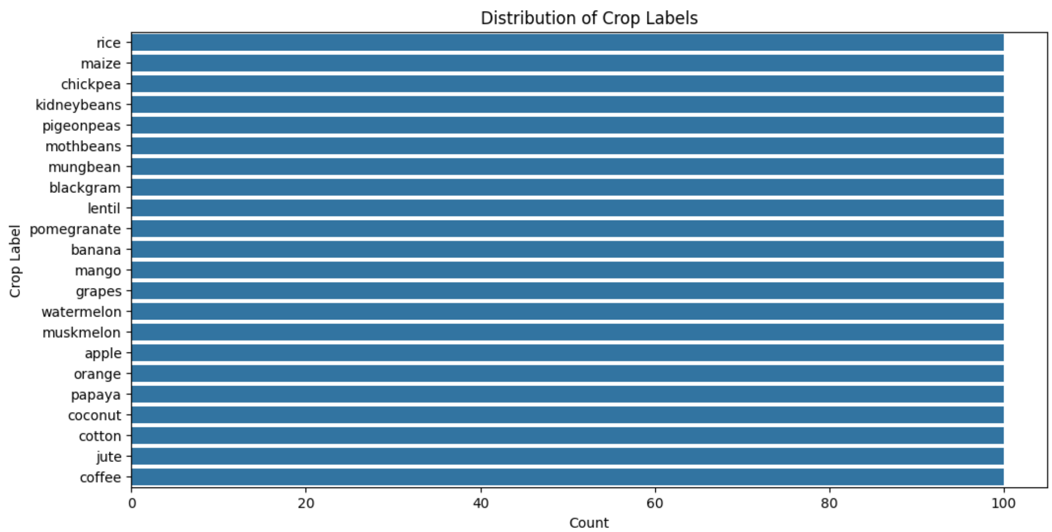

Figure 3.2: Distribution of Crop Labels

3.2.2 Distribution of numerical features

The graphic comprises histograms illustrating the distribution of each numerical characteristic over the whole dataset. Nutrient characteristics (N, P, K) have multimodal distributions, indicating the existence of crops with varying nutrient needs. Temperature and pH are somewhat regularly distributed, centred around 25–30°C and 6–7, respectively, which are ideal for optimum plant development. Humidity has a bias towards elevated values, with several samples above 70%, whilst rainfall displays a right-skewed distribution, mostly ranging from 50 to 200 mm. These distributions elucidate the diversity and spectrum of situations shown in the data.

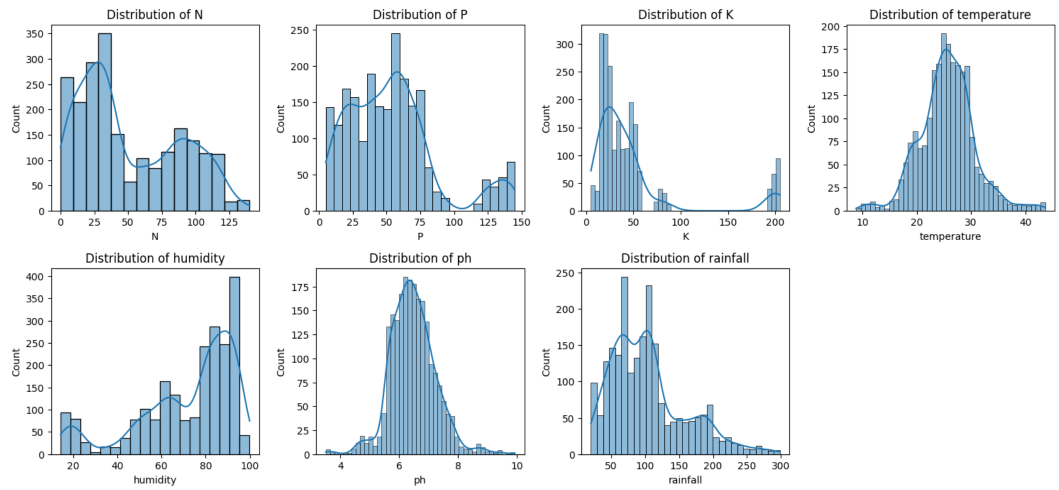

Figure 3.3: Distribution of numerical features

3.2.3 Correlation matrix of numerical features

The graphic displays a correlation matrix for the numerical attributes of the dataset: Nitrogen (N), Phosphorus (P), Potassium (K), temperature, humidity, pH, and rainfall. Each column displays the correlation coefficient between two variables, spanning from -1 (indicating perfect negative correlation) to 1 (indicating perfect positive correlation). The main association exists between P and K, exhibiting a robust positive correlation of 0.74, signifying that these two nutrients tend to rise together. The majority of other feature pairings have weak or insignificant correlations, with coefficients approaching 0, indicating that these variables are mostly independent within this dataset.

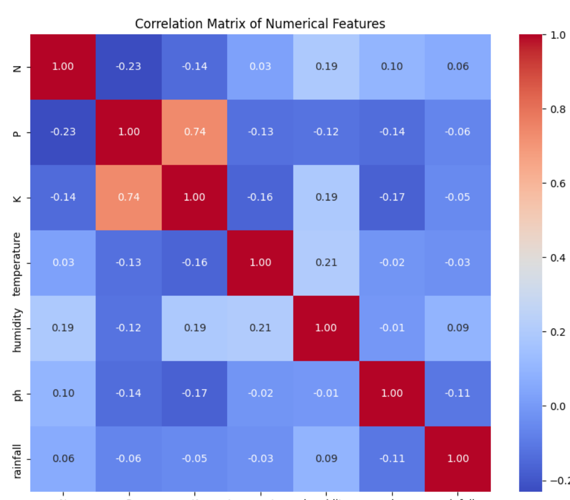

Figure 3.4: Correlation matrix of numerical features

3.2.4 Relationships between numerical features and the target variable using box plots

The figure presents boxplots for each numerical characteristic categorised by crop label. Each subplot demonstrates the variation in the distribution of a variable (such as nitrogen, phosphorus, potassium, temperature, humidity, pH, or precipitation) among several crops. Crops such as rice and maize have elevated median values for rainfall and humidity, whilst others like grapes and watermelon display unique distributions for temperature and nutritional needs. These visualisations emphasise the distinct climatic and nutritional requirements of each crop, which is beneficial for categorisation and recommendation jobs in agriculture.

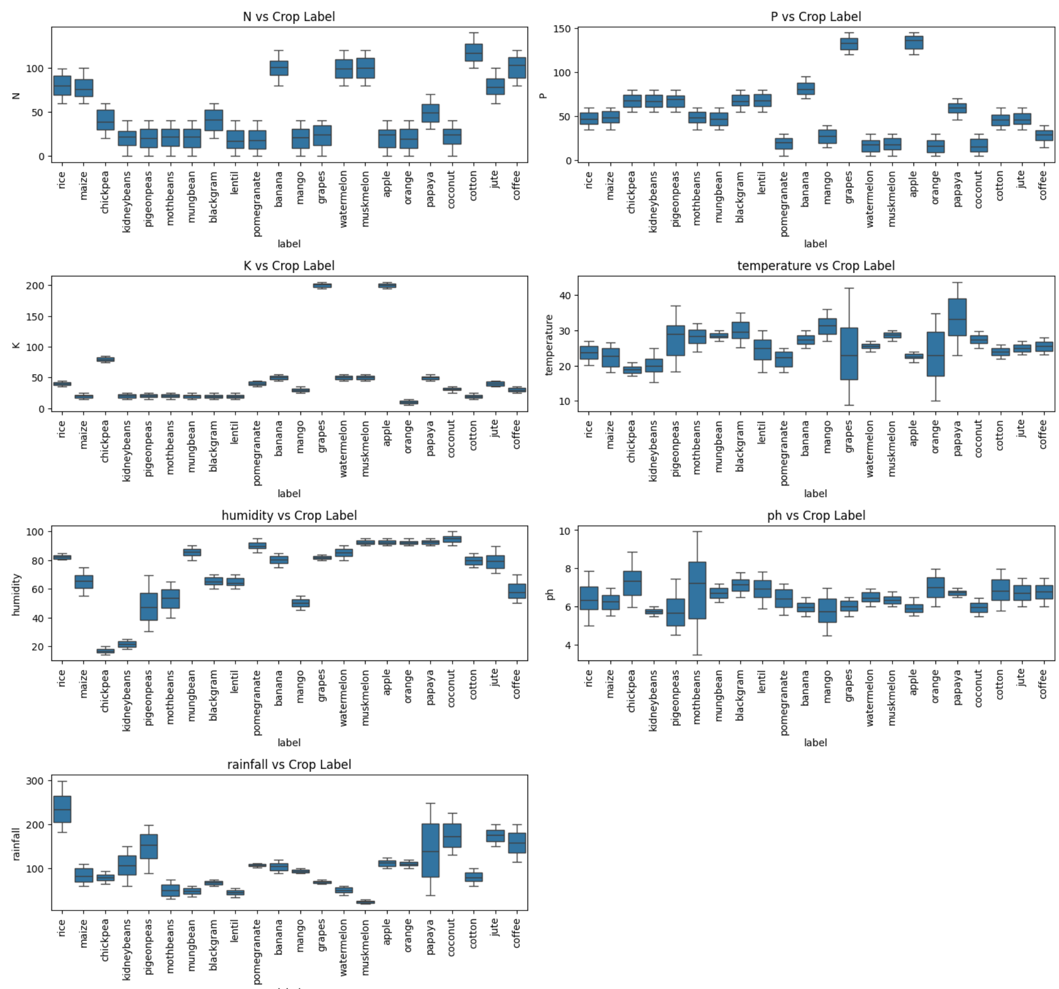

Figure 3.5: Relationships between numerical features and the target variable using box plots

3.2.5 Pairwise relationships for numerical features

The Figure displays a pairplot of all numerical characteristics, illustrating their pairwise correlations and distributions. Each scatter plot illustrates the interaction between two variables, whilst the diagonal plots depict the distribution of each characteristic. The distributions of nitrogen (N), phosphorus (P), and potassium (K) are dispersed among specific ranges, whereas rainfall exhibits a right-skewed distribution, with the majority of data points concentrated below 200 mm. The scatter plots do not demonstrate robust linear connections among the majority of feature pairs, corroborating the previous conclusion that these variables are mostly independent, with the exception of a few mild correlations. This visualisation aids in identifying patterns, clusters, or possible outliers within the data.

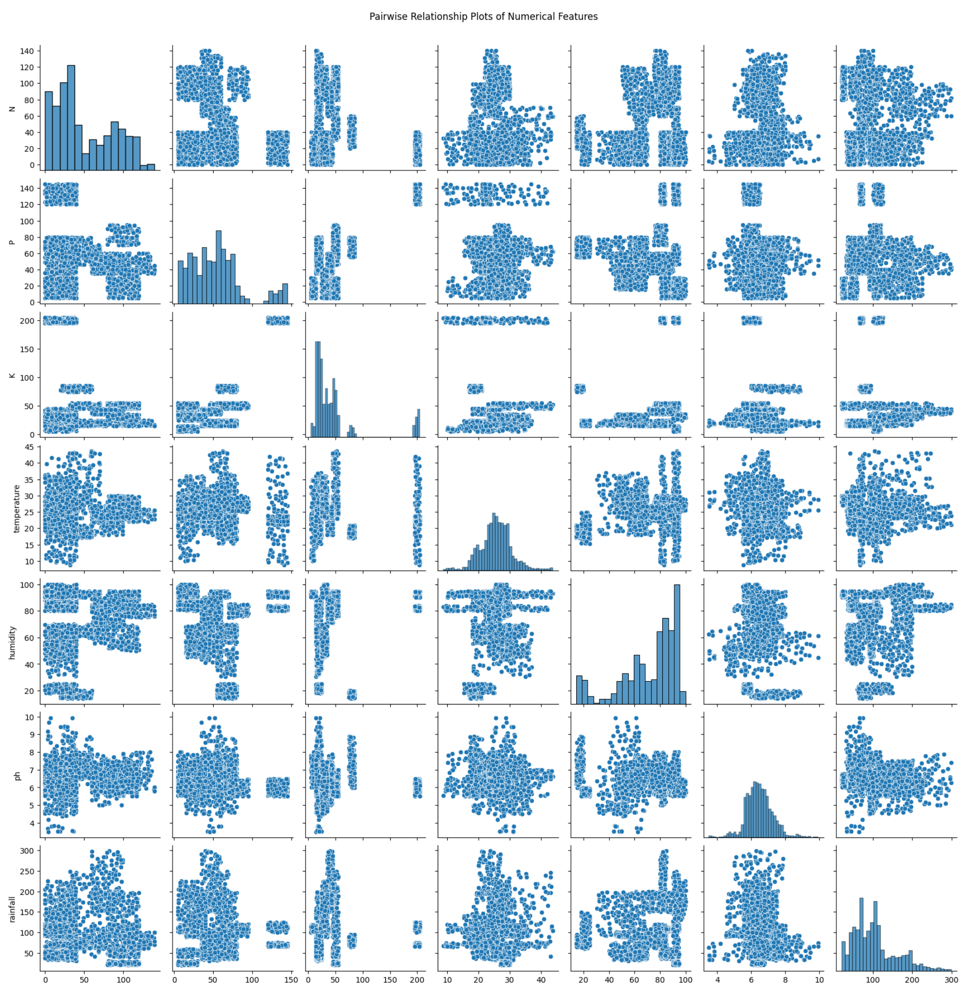

Figure 3.6: Pairwise relationships for numerical features

3.2.6 Feature importances

The graphic presents the feature importances derived from a machine learning model evaluating agricultural data. Rainfall and humidity are the predominant contributors, with significance values of roughly 0.23 and 0.20, respectively. The soil nutrients potassium (K) and phosphorus (P) are prioritised, whilst nitrogen (N), temperature, and pH have lower scores, signifying their diminished contribution to the model's predictions. This ranking indicates that environmental elements such as rainfall and humidity are more significant in influencing crop results than certain soil nutrients or pH in this dataset.

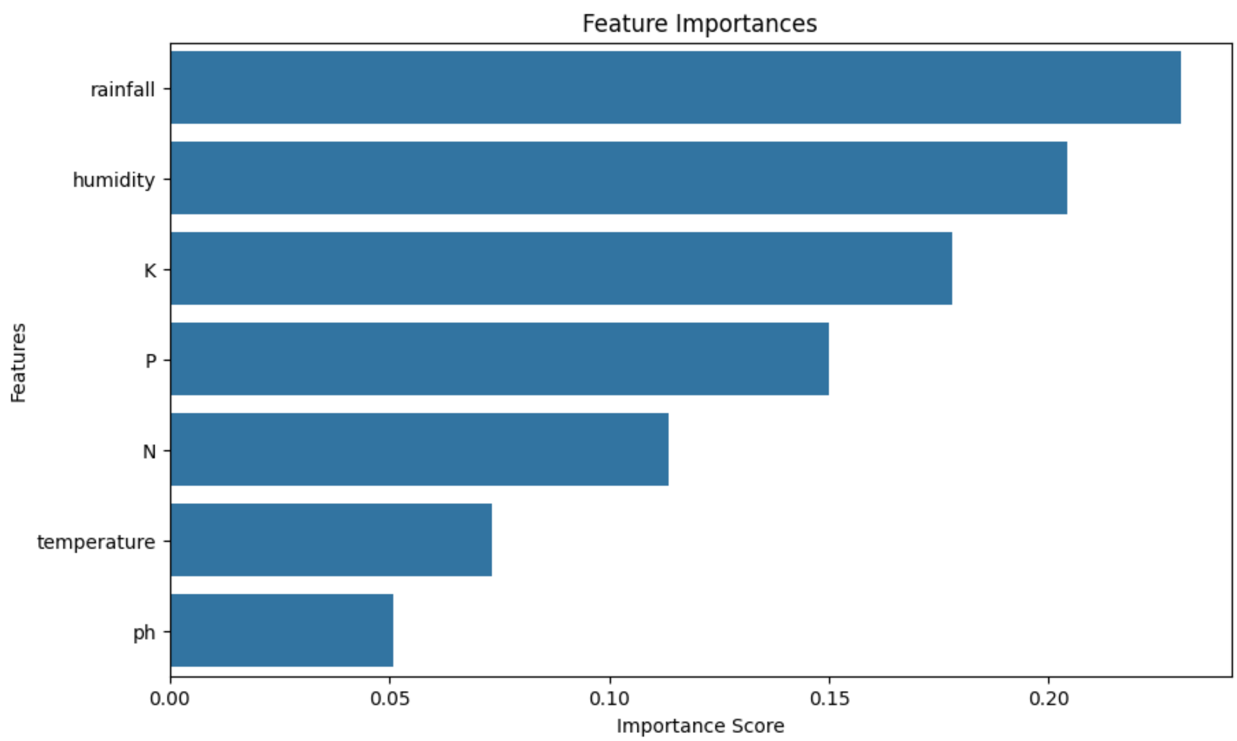

Figure 3.7: Feature importances

CHAPTER IV

Performance Analysis

4.1 Evaluation of Random Forest Classifier Model

4.1.1 Confusion Matrix of Random Forest Classifier Model

A multi-class classification model trained to recognise 21 distinct crop types (e.g., apple, banana, blackgram, etc.) is evaluated in depth in the image's confusion matrix. In this matrix, the actual label is shown by each row, while the anticipated label is shown by each column. The number of right predictions for each class is shown by the diagonal components. The most common value is 25, which means that all 25 samples were properly identified for almost every kind of crop. Except for a few cases, the model got 24 predictions right and one wrong; for example, it misclassified one sample of blackgram, lentil, and rice. With the exception of these three instances, all off-diagonal components are 0, suggesting that the model correctly classified the crops almost every time.

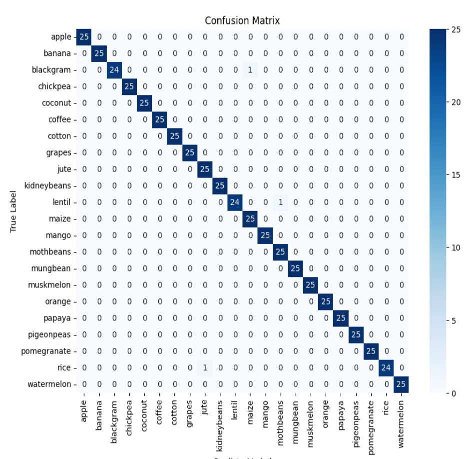

Figure 4.1: Confusion Matrix of Random Forest Classifier Model

The consistency of the confusion matrix, where most predictions lie exactly along the diagonal, lends more credence to this excellent degree of accuracy. The model's outstanding performance is shown by the fact that out of 525 predictions, there were only three misclassifications, indicating an overall accuracy rate more than 99%. This was achieved by dividing the classes into 25 groups and using 25 samples from each group. These findings demonstrate that the model learns the unique characteristics of each crop class quite well, which makes it a trustworthy tool for agricultural decision-making and practical crop identification applications.

4.1.2 Classification Report of Random Forest Classifier Model

The classification report in the image provides a detailed performance summary for a multi-class crop classification model evaluated on 21 different crop types, including apple, banana, blackgram, chickpea, coconut, and others. Each crop class has a "support" of 25, meaning there are 25 test samples for each crop. The model achieved perfect precision, recall, and F1-score (all 1.00) for the majority of classes, such as apple, banana, chickpea, coconut, coffee, cotton, grapes, lentil, mango, mungbean, muskmelon, orange, papaya, pigeonpeas, pomegranate, and watermelon. Only a few classes—blackgram, jute, kidneybeans, maize, mothbeans, and rice—show slightly lower recall and F1-scores (0.96 and 0.98), indicating that for these classes, the model misclassified one sample out of 25, resulting in 24 correct predictions and one error for each.

The overall performance metrics are outstanding, with an accuracy of 0.99 across all 550 samples (21 classes $\times$ 25 samples each plus a possible rounding adjustment), and both macro and weighted averages for precision, recall, and F1-score at 0.99. This means that, out of 550 predictions, only a handful were incorrect, and the model was able to distinguish nearly all crop types with extremely high reliability. Such results demonstrate the model's robustness and suitability for real-world agricultural applications, where accurate crop identification is crucial for decision-making and resource optimization.

4.2 Evaluation after Hyperparameter Tuning

4.2.1 Confusion Matrix after Hyperparameter Tuning

A multi-class classification model trained to differentiate among 21 distinct crop varieties, such as chickpeas, bananas, apples, and blackgrams, is evaluated using the

Table 4.1: Classification Report of Random Forest Classifier Model

Class	**Precision**	**Recall**	**F1-score**	**Support**
apple	1.00	1.00	1.00	25
banana	1.00	1.00	1.00	25
blackgram	1.00	0.96	0.98	25
chickpea	1.00	1.00	1.00	25
coconut	1.00	1.00	1.00	25
coffee	1.00	1.00	1.00	25
cotton	1.00	1.00	1.00	25
grapes	1.00	1.00	1.00	25
jute	0.96	1.00	0.98	25
kidneybeans	1.00	0.96	0.98	25
lentil	1.00	0.96	0.98	25
maize	0.96	1.00	0.98	25
mango	1.00	1.00	1.00	25
mothbeans	0.96	1.00	0.98	25
mungbean	1.00	1.00	1.00	25
muskmelon	1.00	1.00	1.00	25
orange	1.00	1.00	1.00	25
papaya	1.00	1.00	1.00	25
pigeonpeas	1.00	1.00	1.00	25
pomegranate	1.00	1.00	1.00	25
rice	1.00	0.96	0.98	25
watermelon	1.00	1.00	1.00	25
accuracy			0.99	550
macro avg	0.99	0.99	0.99	550
weighted avg	0.99	0.99	0.99	550

exhibited confusion matrix. There is a row for the actual class and a column for the anticipated class. All 25 test samples for those crops were properly categorised, as shown by the diagonal elements, which in this matrix mostly had values of 25, which imply valid predictions for each crop type. The algorithm accurately identified 24 out of 25 samples with one misclassification each for blackgram, lentil, and rice, making them the only three outliers. There seems to be very little misunderstanding between the classes, since all other crops have flawless classification and all off-diagonal elements are zero with the exception of these three little mistakes.

The statistics shown here prove that the model is quite accurate and dependable. With just three cases of misclassification out of 525 predictions (21 classes $\times$ 25 samples per class), the overall accuracy was about 99.4%. The alignment along the diagonal of the matrix is almost flawless, which shows that the model learnt and distinguished between the various types of crops quite well. Due to its excellent performance, the model is well-suited for use in practical agricultural settings, where accurate crop identification is essential for managing resources and making informed decisions.

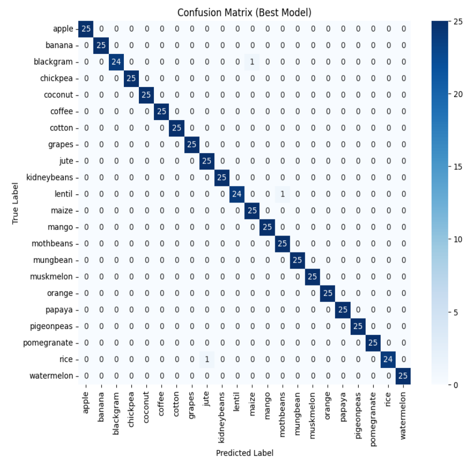

Figure 4.2: Confusion Matrix after Hyperparameter Tuning

4.2.2 Classification Report after Hyperparameter Tuning

The classification report presents a comprehensive evaluation of a multi-class crop classification model tested on 21 crop types, each with 25 samples, totaling 550 predictions. For most crops, including apple, banana, chickpea, coconut, coffee, cotton, grapes, lentil, mango, mungbean, muskmelon, orange, papaya, pigeonpeas, pomegranate, rice, and watermelon, the model achieved perfect precision, recall, and F1-score of 1.00, indicating all test samples were classified correctly. However, for a few crops—blackgram, jute, maize, kidneybeans, and mothbeans—the recall and F1-score dropped slightly to 0.96 and 0.98, respectively, meaning that for each of these classes, one out of 25 samples was misclassified, resulting in 24 correct predictions and one error per crop.

Overall, the model demonstrates exceptional performance, with an accuracy of 0.99 across all 550 samples and macro and weighted averages for precision, recall, and F1-score also at 0.99. This signifies that only a handful of predictions were incorrect out of the entire test set, and the model reliably distinguishes between nearly all crop types. Such high accuracy and consistency across classes highlight the model's robustness and suitability for practical agricultural applications, where precise crop identification is essential for effective decision-making and resource management.

Table 4.2: Classification Report after Hyperparameter Tuning

Class	Precision	Recall	F1-score	Support
apple	1.00	1.00	1.00	25
banana	1.00	1.00	1.00	25
blackgram	0.96	1.00	0.98	25
chickpea	1.00	1.00	1.00	25
coconut	1.00	1.00	1.00	25
coffee	1.00	1.00	1.00	25
cotton	1.00	1.00	1.00	25
grapes	1.00	1.00	1.00	25
jute	0.96	1.00	0.98	25
kidneybeans	1.00	0.96	0.98	25
lentil	1.00	0.96	0.98	25
maize	0.96	1.00	0.98	25
mango	1.00	1.00	1.00	25
mothbeans	0.96	1.00	0.98	25
mungbean	1.00	1.00	1.00	25
muskmelon	1.00	1.00	1.00	25
orange	1.00	1.00	1.00	25
papaya	1.00	1.00	1.00	25
pigeonpeas	1.00	1.00	1.00	25
pomegranate	1.00	1.00	1.00	25
rice	1.00	0.96	0.98	25
watermelon	1.00	1.00	1.00	25
accuracy			0.99	550
macro avg	0.99	0.99	0.99	550
weighted avg	0.99	0.99	0.99	550

4.3 User Interface

The user interface is named ”Crop Recommendation App” and is intended to suggest an appropriate crop based on specified environmental parameters.

The components and their respective roles are shown in the illustration.

The title ”Crop Recommendation App” explicitly indicates the function of this interface.

The directive ”Input the environmental conditions to receive a crop recommendation” instructs users on the necessary action.

Input Parameters (Environmental Conditions) Every environmental component is shown as a named slider, accompanied by a range and a value box:

- Nitrogen (N): Slider spans from 0 to 140, now positioned at 41.

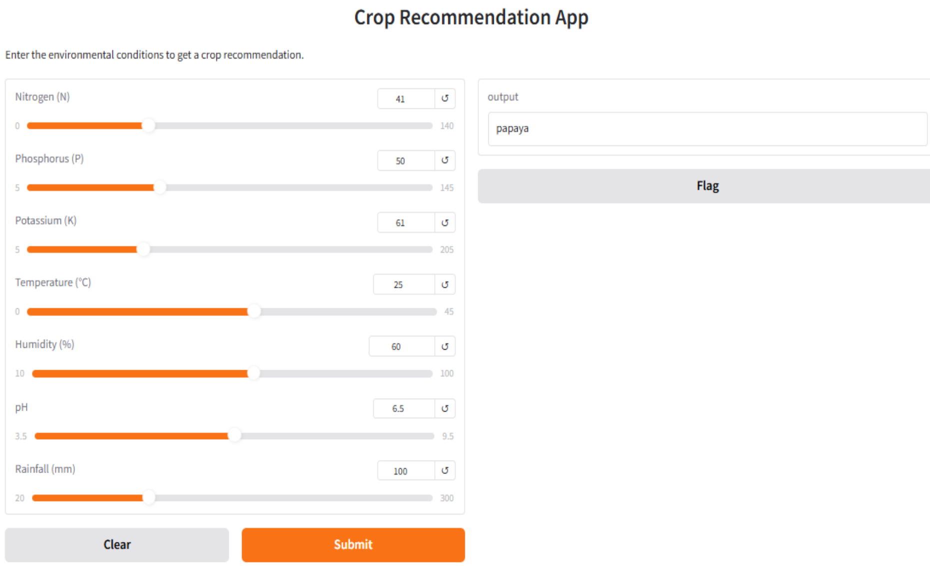

Figure 4.3: User Interface

- Phosphorus (P): A slider with a range of 5 to 145, now positioned at 50.
- Potassium (K): Slider spans from 5 to 205, now positioned at 61.
- Temperature (°C): A slider spanning from 0 to 45, now positioned at 25.
- Humidity (%): A slider with a range of 10 to 100, currently positioned at 60.
- pH: A slider with a range of 3.5 to 9.5, now positioned at 6.5.
- Precipitation (mm): Slider spanning from 20 to 300, now positioned at 100.

Each input comprises a slider for value adjustment and a numeric input field that displays the chosen value.

Buttons Clear: Positioned at the bottom left, it resets all input fields.

Submit: Positioned in the bottom center, it transmits the inputted data for a crop suggestion.

Result Section Label "output": Designates the location where the suggested crop will be shown.

The text box presents the outcome; in this instance, it indicates "papaya" as the recommended crop, signifying that papaya is advised for the specified environmental circumstances.

CHAPTER V

Future Work & Conclusion

5.1 Conclusion

Employing a large number of people and making a significant dent in national gross domestic product, agriculture is the backbone of any global economy. But there are a lot of obstacles to agricultural output, such as different soil types, unpredictable weather, and a lack of access to current farming techniques. Overcoming these challenges and boosting agricultural output requires the development of intelligent agricultural systems. A crop recommendation system based on machine learning and customised to the agricultural environment is presented in this research to satisfy this demand. It is the goal of the proposed system to use past data on weather, soil, crop yields, and farmer preferences to provide tailored crop recommendations. This work aims to provide practical suggestions for crop selection by evaluating nine various ML models: LR, SVM, KNN, DT, RF, BG, AB, GB, and ET. To make sure the data is suitable for training models, it is cleaned and normalised using several preprocessing approaches. In order to establish a correlation between environmental and agronomic parameters and crop yields, machine learning models are trained on historical datasets that include variables such as soil pH, nutrient levels, rainfall, humidity, and temperature. Optimising the models' performance and ensuring their resilience is achieved by fine-tuning using approaches like cross-validation. With an accuracy percentage of 99.31%, Random Forest stands out as the best performer among these models. One potential solution to the problems farmers have is the Crop Recommendation system that is based on machine learning. Improved agricultural production, food security, and economic prosperity are the end results of the system's use of modern data analytics and artificial intelligence tools, which enable it to provide farmers with timely and personalised advice. Although our present research makes good use of the accessible information to build a crop recommendation system, we are cognisant of the fact that it has certain constraints. In particular, important factors for precise recommendations but under-represented in our dataset include changes in land quality, temperature, and historical data on crop planting. Improving the accuracy of soil and climatic data, together with planting records for different crops across time, will be the primary goals of future studies. To further tackle the intricacies of agricultural decision-making, we want to include farmers' expert knowledge and investigate cutting-edge Machine Learning methods and hybrid models. The system's accuracy and flexibility will be enhanced with these upgrades, leading to more personalised and effective suggestions for farmers.

5.2 Future Work

To identify the most effective approaches for predicting agricultural output in Bangladesh, we have evaluated five machine learning classifiers and three deep learning models. Nevertheless, owing to restricted data availability, we were able to get good outcomes from three machine learning models and one bespoke deep learning model. Our primary future objective is to enhance our dataset to improve the usefulness and reliability of our models. Furthermore, we want to openly distribute our dataset in the future. We will moreover endeavor to evaluate new cutting-edge models using our enhanced dataset.

References

[1] A. Priyadharshini, S. Chakraborty, A. Kumar, and O. R. Pooniwala, "Intelligent crop recommendation system using machine learning," in *2021 5th international conference on computing methodologies and communication (ICCMC)*, pp. 843–848, IEEE, 2021.

[2] Z. Doshi, S. Nadkarni, R. Agrawal, and N. Shah, "Agroconsultant: intelligent crop recommendation system using machine learning algorithms," in *2018 fourth international conference on computing communication control and automation (ICCUBEA)*, pp. 1–6, IEEE, 2018.

[3] R. K. Rajak, A. Pawar, M. Pendke, P. Shinde, S. Rathod, and A. Devare, "Crop recommendation system to maximize crop yield using machine learning technique," *International Research Journal of Engineering and Technology*, vol. 4, no. 12, pp. 950–953, 2017.

[4] D. A. Reddy, B. Dadore, and A. Watekar, "Crop recommendation system to maximize crop yield in ramtek region using machine learning," *International Journal of Scientific Research in Science and Technology*, vol. 6, no. 1, pp. 485–489, 2019.

[5] P. Bandara, T. Weerasooriya, T. Ruchirawya, W. Nanayakkara, M. Dimantha, and M. Pabasara, "Crop recommendation system," *International Journal of Computer Applications*, vol. 975, p. 8887, 2020.

[6] M. Kalimuthu, P. Vaishnavi, and M. Kishore, "Crop prediction using machine learning," in *2020 third international conference on smart systems and inventive technology (ICSSIT)*, pp. 926–932, IEEE, 2020.

[7] P. Vijayabaskar, R. Sreemathi, and E. Keertanaa, "Crop prediction using predictive analytics," in *2017 international conference on computation of power, energy information and commuincation (ICCPEIC)*, pp. 370–373, IEEE, 2017.

[8] M. Rashid, B. S. Bari, Y. Yusup, M. A. Kamaruddin, and N. Khan, “A comprehensive review of crop yield prediction using machine learning approaches with special emphasis on palm oil yield prediction,” *IEEE access*, vol. 9, pp. 63406–63439, 2021.

[9] M. Dharani, R. Thamilselvan, P. Natesan, P. Kalaivaani, and S. Santhoshkumar, “Review on crop prediction using deep learning techniques,” in *Journal of physics: conference series*, vol. 1767, p. 012026, IOP Publishing, 2021.

[10] P. M. Gopal and R. Bhargavi, “A novel approach for efficient crop yield prediction,” *Computers and Electronics in Agriculture*, vol. 165, p. 104968, 2019.

[11] P. Kathiria, U. Patel, S. Madhwani, and C. Mansuri, “Smart crop recommendation system: A machine learning approach for precision agriculture,” in *Machine Intelligence Techniques for Data Analysis and Signal Processing: Proceedings of the 4th International Conference MISP 2022, Volume 1*, pp. 841–850, Springer, 2023.

[12] A. Suruliandi, G. Mariammal, and S. Raja, “Crop prediction based on soil and environmental characteristics using feature selection techniques,” *Mathematical and Computer Modelling of Dynamical Systems*, vol. 27, no. 1, pp. 117–140, 2021.

[13] S. Iniyan, V. A. Varma, and C. T. Naidu, “Crop yield prediction using machine learning techniques,” *Advances in Engineering Software*, vol. 175, p. 103326, 2023.

[14] J. Pant, R. Pant, M. K. Singh, D. P. Singh, and H. Pant, “Analysis of agricultural crop yield prediction using statistical techniques of machine learning,” *Materials Today: Proceedings*, vol. 46, pp. 10922–10926, 2021.

[15] A. Nigam, S. Garg, A. Agrawal, and P. Agrawal, “Crop yield prediction using machine learning algorithms,” in *2019 fifth international conference on image information processing (ICIIP)*, pp. 125–130, IEEE, 2019.

[16] T. Van Klompenburg, A. Kassahun, and C. Catal, “Crop yield prediction using machine learning: A systematic literature review,” *Computers and electronics in agriculture*, vol. 177, p. 105709, 2020.

[17] A. Chlingaryan, S. Sukkarieh, and B. Whelan, “Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review,” *Computers and electronics in agriculture*, vol. 151, pp. 61–69, 2018.

[18] S. B. Kamatchi and R. Parvathi, "Improvement of crop production using recommender system by weather forecasts," *Procedia Computer Science*, vol. 165, pp. 724–732, 2019.

[19] S. Bhanumathi, M. Vineeth, and N. Rohit, "Crop yield prediction and efficient use of fertilizers," in *2019 International Conference on Communication and Signal Processing (ICCSP)*, pp. 0769–0773, IEEE, 2019.

[20] D. A. Bondre and S. Mahagaonkar, "Prediction of crop yield and fertilizer recommendation using machine learning algorithms," *International Journal of Engineering Applied Sciences and Technology*, vol. 4, no. 5, pp. 371–376, 2019.

[21] R. Ghadge, J. Kulkarni, P. More, S. Nene, and R. Priya, "Prediction of crop yield using machine learning," *Int. Res. J. Eng. Technol.(IRJET)*, vol. 5, pp. 2237–2239, 2018.

[22] S. Prabhu, P. Revandekar, S. Shirdhankar, and S. Paygude, "Soil analysis and crop prediction," *International Journal of Scientific Research in Science and Technology*, vol. 7, no. 4, pp. 117–123, 2020.

[23] M. A. Rahman, S. Kang, N. Nagabhatla, and R. Macnee, "Impacts of temperature and rainfall variation on rice productivity in major ecosystems of bangladesh," *Agriculture & Food Security*, vol. 6, pp. 1–11, 2017.

Appendix