Research Article

International Journal of Advanced Technology and Engineering Exploration, Vol 11(117)
ISSN (Print): 2394-5443 ISSN (Online): 2394-7454
http://dx.doi.org/10.19101/IJATEE.2024.111100127

Soil classification and crop cultivation prediction: a comparative study of machine learning models

Fardowsi Rahman[1, 2], Md. Ashikur Rahman Khan[1*] and Zareen Tasneem[2]
Department of Information and Communication Engineering, Noakhali Science and Technology University, Noakhali 3814, Bangladesh[1]
Department of Computer Science and Engineering, Primeasia University, Dhaka, Bangladesh[2]

Received: 04-February-2024; Revised: 21-July-2024; Accepted: 27-July-2024

Abstract
Soil classification is necessary to optimize agricultural productivity, manage land use effectively, and protect the environment. The soil features, including organic matter, chemical properties, composition, and other organisms, collectively determine its type and enable its differentiation from other soil types. The problem emerges from poor soil type, directly impacting agricultural crop cultivation and presenting a fundamental issue in farming practices. In response to these concerns, this work classified soil based on chemical features such as pH levels, salinity, organic matter, nitrogen, phosphorus, sulphur, and boron, which aim to refine the soil characterization process. There are many lands in Bangladesh where the people think that cultivation is not possible at all. Due to these reasons, many lands are uncultivated in our country every year. Considering these issues, this research aims to predict crop cultivation for a particular soil so that people can select and cultivate various crops undoubtedly and correctly on their land. The soil types are determined first based on their distinct characteristics for a particular zone, then the crop cultivation is ascertained according to the soil types. Various machine learning (ML) techniques such as support vector machine (SVM), decision tree (DT), multilayer perceptron neural network, random forest (RF), logistic regression (LR), and naïve Bayes (NB) are used for prediction. The ML models are selected through a literature-informed process and intended for comparative analysis to determine the most effective techniques. A comparative analysis among different techniques is performed based on performance metrices. The results indicate that the RF algorithm is the most effective for soil classification due to its outstanding accuracy of 96.48%. For crop cultivation prediction, the SVM model outperforms other models with an accuracy of 94.95%. The outcomes of this research endeavour serve as a valuable tool for enhancing farming practices and making substantial contributions to the economy's growth.

Keywords
Soil classification, Crop cultivation prediction, Support vector machine, Random forest, Multilayer perceptron.

1.Introduction
Soil is pivotal in supporting thriving agriculture by providing the foundation for crop growth. The soil composition that supports plant growth comprises a variety of organic substances, minerals, living things, air, and water. Soil type gives information about morphological, physical, and chemical characteristics and mineral content. Determining which crops flourish in a particular soil type requires knowledge of the features and characteristics of distinct soil types. Soil classification links soil samples and natural entities on the earth's land surface [1]. There are several varieties of soil, and each has distinctive features. Based on these multiple features, several kinds of crops are grown. Understanding the properties of different soil types is crucial for selecting suitable crops and ensuring healthy and productive agriculture.

Table 1 shows an overview of the major soil types and the crops that thrive in each:

Agriculture is a cornerstone of human civilization, bearing multifaceted significance beyond mere sustenance. Agriculture ensures the fundamental necessity of food security, underpinning the nourishment of billions worldwide. Predicting and analysing crop growth is integral to contemporary farming practices, and machine learning (ML) has emerged as a potent tool in this endeavor [2, 3]. Grains (potato, sesame, wheat, pulse, rice, and other grains),

*Author for correspondence

vegetables, oil seeds, and root crops make up most of the food crops grown on the arable regions of the planet. The main cash crops include silk, rubber, cotton, tea, jute, tobacco, sugarcane, and coffee. In addition to this, the main food crops include maize, rice, wheat, pulses, oil seeds, tomato, potato, fruit, and vegetables [4–6]. Seasonal variations, biological processes, and economic considerations influence crop production in agriculture. Weather affects the growth and development of crops, resulting in considerable fluctuations in yield within a season [7–9].

Table 1 Different soil types and the crops that thrive in each

Soil type	Characteristics	Suitable crops
Loam	• Good drainage and water-holding capacity • High fertility and excellent structure • Rich in nutrients	Tomato, maize, rice, wheat, pumpkin, gourd, bean.
Clay	• Small, densely packed particles • Poor drainage and heavy when wet	Cauliflower, brinjal, broccoli, carrots.
Clay loam	• Good structure for root growth when well-managed • Excellent water-holding capacity	Pumpkin, linseed, gourd, rice, spiny gourd, bean.

In precision and intelligent farming, soil's physical and chemical properties act upon a significant rule. The cultivation of crops is based on soil that has vital nutrients. However, the increased usage of fertilizers has led to a decline in several soil nutrients. As a result of this, crop productivity is decreasing [10]. Farmers lack the necessary understanding of the soil properties and usage of the optimum quantity of inputs in their crops, which are essential for increasing productivity [11]. Rich in minerals, fertile soil is excellent for farming. Farmers' conventional methods need to be revised to fulfil the substantial demand. They injure the land by intensifying the use of toxic pesticides. It significantly impacts agricultural practices, so the land remains unproductive and barren [12]. The formation of soil is a lengthy process, and diverse soils are generated in different locations due to numerous soil-forming elements throughout the landscape [13]. Soil attributes significantly impact soil improvement, land consolidation, drainage management, soil erosion, and irrigation. At larger dimensions (such as regional and continental scales), climate, biota, and geological history are the crucial variables influencing soil's chemical and physical qualities. On the other hand, geography and human activities may be the main forces influencing soil characteristics at lower scales [14].

Agriculture is closely related to the life and livelihood of the people of Bangladesh. Numerous areas in Bangladesh are perceived as unsuitable for cultivation, leading to the notion that farming is not viable in these regions. This perception has resulted in a significant amount of land remaining unused and unexplored for agricultural purposes. The current population of Bangladesh is about 16 crores; the size of this country is petite according to population. Moreover, the population is constantly increasing all over the world. Food is the first and foremost basic need of human beings. However, many lower- and middle-income countries cannot provide enough food for their population. This problem can be significantly solved if we can produce more crops. Therefore, it is essential to increase the use of technology in agriculture.

This research addresses the challenges facing agriculture in Bangladesh, particularly the decreasing arable land due to population growth and the perception of specific areas as unsuitable for cultivation. Artificial intelligence (AI) and intelligent vertical farming methods represent a significant leap forward in agricultural technology, especially for urban areas with limited space. Vertical farming involves growing crops in vertically stacked layers, maximizing the use of urban spaces like skyscrapers and repurposed warehouses. By integrating AI, these systems benefit from automated monitoring and control, utilizing sensors and IoT to maintain optimal growing conditions, which ensures consistent crop quality and higher yields. It is necessary to increase the production of more crops on less land. For that, farmers must understand what crops can be produced according to the land. Plants require a specific kind of soil to grow correctly. For example, a crop like rice requires a very high moisture content in the soil. By contrast, a crop like wheat prefers loamy soil rich in organic material. Improper soil analysis can lead to crop decline or outright failure. Due to improper soil analysis, farmers do not get crops as expected. Therefore, determining soil type is necessary.

This research aims to categorize soil types and forecast agricultural crop cultivation based on soil type using ML techniques.

ML is employed in agriculture to enhance crop quality and productivity. It is a scientific discipline that enables machines to learn independently without human intervention [15]. Precision farming, a significant implementation of ML in agriculture, entails using data and technology to refine agricultural processes such as soil enrichment, watering systems, and pest mitigation. This approach aims to enhance both yield and quality of crops through optimized strategies [16]. Through the examination of past market trends and climatic conditions, ML algorithms are useful for forecasting various crops and recommending ideal planting schedules and locations [17].

Determining which crops flourish well on the land according to soil type is difficult. Farmers do not get the desired yield of crops because of improper analysis of the soil. In response to difficulties farmers encounter when cultivating crops, this research provides an excellent tool for enhancing farming. However, predicting crop cultivation from its parametric point of view is a very critical task. Our accomplished research differs from previous works by utilizing multiple ML techniques, such as support vector machine (SVM), decision tree (DT), multilayer perceptron neural network (MLPNN), random forest (RF), naive bayes (NB), and logistic regression (LR). This paper introduced effective methodologies for crop cultivation systems according to soil class.

The following are the contributions of our study-

- This study trained and tested distinct ML models using the collected soil and crop data. The datasets used in this study are authentic.
- This research classified soil based on soil features such as pH levels, salinity, organic matter, nitrogen, phosphorus, sulphur, and boron.
- This study carries out crop cultivation prediction based on soil type.
- This study compares the performance of six ML techniques to find an optimal model for predicting crop cultivation by classifying soil.

The findings obtained for soil classification convey pertinent insights that the RF method is the best classifier for classifying soil due to its promising results, with a fascinating accuracy rate of 96.48%. In addition, the results we obtained for predicting crop cultivation show that the SVM algorithm is the best performer compared to the others, with an accuracy rate of 94.95%.

The paper is organized as follows: Section 1 covers the research background, significance, problem analysis, and contribution. Section 2 explores the literature review. Section 3 outlines the research methodology. Section 4 explores the experimental findings, and section 5 discusses the study results and its implication. Lastly, section 6 furnishes a comprehensive conclusion.

2.Related works

In this particular section, this paper explores noteworthy studies that focus on predicting crop cultivation through soil classification. Towards the conclusion of this section, this study conducts a comparative analysis between our experimental work and existing research to underscore our study's distinctive contributions and novelty.

Palanivel and Surianarayanan stated that crop cultivation prediction is highly connected with variabilities such as weather (temperature, rainfall, and coldness), heat radiation, crop density, irrigation, fertilizers, water supply, soil class, farmhouse density, and soil chemical features [18]. The chemical measurements of the soil, such as manganese, organic matter, phosphorus, nitrogen, potassium, and soil pH, are used to recommend appropriate amounts of fertilizers and preferred crops [19, 20]. Kiran et al. take parameters such as climate, soil category, water density, chemical fertilizers, compost, and crop data collection to improve the production of crops [21].

Numerous researchers have endeavoured to forecast crop cultivation by employing soil classification methodologies. In a study by Cai et al., salt-affected soil was classified using an enhanced SVM classifier. The data for SVM classification was derived from multi-spectral characteristics and texture features. The researchers selected optimal texture features, including mean, variance, and homogeneity, to improve the accuracy of the classification process [22]. Jiang et al. (2024) present a detailed study on the changes in regional cropping structures over time and space and the factors that drive these changes [23]. Inazumi et al. used image processing to classify soil. These research projects are efficiently completed using deep learning and neural network (NN) methods. An AI model was developed for classifying clay, gravel, and sandy soils through images of practical soil textures [24]. Agarwal et al. proposed a novel and very effective automated technique for predicting soil types using soil images and soil characteristics. Moreover, the proposed classification approach has also been evaluated and tested on a dataset of soil images. The results of the experiments indicate that it has an accuracy of 80.58%, which is higher than that of other methods [25]. Kodikara et al. used soil data from mountainous areas to train and evaluate the predictive power of ML classification-based models. Their main goal was to classify mountain soils based on soil properties [26]. Mallick et al. examined the performance of ML models by using Bangladesh's Sunamganj district soil data. Furthermore, they proposed a hybrid artificial neural network (ANN) based model using complex wetland mapping data [27]. Azmin et al. (2024) describe the challenges of using ML for soil classification, including consistency issues and the availability of human specialists. Despite these difficulties, they note that ML models could replace traditional, labor-intensive methods for categorizing soil samples [28].

The essential components of soil are sand, clay, and silt. The sand, silt, and clay percentage in a particular soil can be checked by using a hydrometer test [29]. Ghadge et al. proposed a new method for testing soil quality that will be helpful for farmers. This system helps forecast suitable crops according to soil qualities and recommends the necessary fertilizers [30]. Angelaki et al. try to estimate the cumulative infiltration of soil. They used four ML algorithms to do this work and evaluated the performance through correlation coefficient (cc) and root mean square error (RMSE). According to the results, the adaptive Neuro-fuzzy inference system (ANFIS) worked best [31]. Aydın et al. are investigating novel ML techniques/algorithms that automatically classify soil to reduce process time and expense [32]. In a different study, Nguyen et al. created ML models for classifying soil using 15 input parameters related to soil characteristics, resulting in 5 soil classes. They utilize a database containing soil property data gathered from several sites in Vietnam [33].

Guo et al. (2024) discussed the factors influencing the current crop planting patterns and predicted future trends in this context using ML and the hierarchical linear model (HLM) [34]. Farmers must select suitable crops to increase their crop production and financial gain. Kumar et al. proposed a crop selection method (CSM) approach so that farmers can increase the annual production of crops and thereby simultaneously increase the country's economic growth [35]. Farmers can also choose crops based on the season, soil type, weather, location, and cost. However, these traditional methods are risky in being profitable [36]. Goswami et al. (2024) utilized ML models to categorize different soil series and recommend appropriate crops based on their geographic characteristics [37]. Waikar et al. built a user-friendly web-based system to forecast crop yields. They assert that their system helps increase the net yield rate of crops [38]. Using a weather-based crop dataset, Bhojani and Bhatt constructed an updated MLP, NN with a new activation function for agricultural production estimation [39]. Additionally, research has been done to predict agricultural cultivation utilizing various techniques, including ANN, SVM, RF, multilayer perceptron, and advanced DT [40–43]. Many researchers have tried to employ fuzzy-based techniques to classify soil and crops [44].

Dash et al. concentrated on linking weather factors to soil micronutrients. Then, using SVM and DT techniques, the crop type is classified based on the micronutrients [45]. Waghmare and Jhare emphasized predicting the choice of crops and enhancing the soil quality. This study is carried out by gathering many soil testing samples to determine soil fertility indices and pH values, constituting a detailed overview of the application of ML in agriculture [46]. To quickly classify rice seeds using variety purity, Ansari et al. developed an automated vision system that worked with multivariate analysis methodologies [47]. To preserve the best crop health, Priya and Yuvaraj proposed a method to increase crop productivity through data analysis and the internet of things (IoT) [48].

Patil et al. developed a method to provide information regarding the ideal timing for planting, plant growth, and plant harvesting [49]. Chakrabarty et al. created a smartphone application available for managing and accounting for agricultural crop yield on a modest scale [50]. Shastry and Sanjay developed a cloud-based farm ecosystem that enables farmers to access relevant agricultural data [51]. Using data from various agricultural research organizations and Bangladesh's meteorological department, Hasan et al. concentrated on creating a learnable dataset on agricultural crop production prediction [52]. A paper presents a statistical analysis of the features and suggests the optimal crop type based on the specified features in the context of the Indian smart town [53]. A system developed by Singh et al. [54] suggests health improvement guidelines for the selected soil sample to increase the profitability of the most suitable crop. Several ML algorithms [55], such as SVM [56], RNN [57], and ANN [58], are used by researchers to classify soil and predict crop cultivation. ML is also used in several medical fields, such as liver disease [59], heart disease, covid-19 prediction [60], and others. For promising crop yield prediction, Shetty et al. built a basic web application using Python [61].

In previous related works, researchers classified soil in various ways. Moreover, most of the existing works consider systematic categorization of soils. Few researchers focus on classifying soil based on chemical features. On the other hand, this research takes significant soil attributes like pH, salinity, organic matter, nitrogen, phosphorus-B, sulfur, and boron, which are used for soil classification. From the researchers' previous work, which crop is cultivated in a particular soil type needs to be clarified. In addition, previous researchers have either classified soils or predicted crop cultivation. However, this research has been done to predict soil classification and crop cultivation. This research aims to classify soil and predict crop cultivation depending on soil class. This paper focuses on various supervised classification algorithms for crop cultivation prediction.

3.Methodology

This research aims to ascertain the cultivation of the crops based on soil classification using the soil and crops dataset. The dataset was collected from agricultural fields of Noakhali district in Bangladesh. This research method and materials are organized into several sections: parameter selection, data collection, AI techniques, data proportion, and performance metrics. *Figure 1* depicts the research undertaking's workflow. At first, this study collects soil and crop data. For the algorithm training, it is necessary to preprocess the collected data. Afterward, the training and testing phase makes use of the processed data. Various algorithms are employed for crop cultivation prediction, including SVM, DT, MLPNN, RF, LR, and K-nearest neighbor (KNN).

3.1Attributes selection and data collection

For the proper soil classification, it is necessary to select significant soil attributes. Organic matter, pH, salinity, potassium, sulfur, zinc, boron, nitrogen, calcium, phosphorus, manganese, magnesium, copper, iron, sand, silt, clay, $CaCo_3$, and other elements are relevant to soil properties. Seven input parameters were chosen for this work from this list of features; additional information is provided in *Table 2*. These input attributes are pH, Salinity, Organic matter, Sulfur, Nitrogen, Phosphorus(B), and Boron. Data from Begumganj, Sonaimuri, Hatiya, Senbag, and Subarnachar Upazilla of Noakhali District in Bangladesh has been collected.

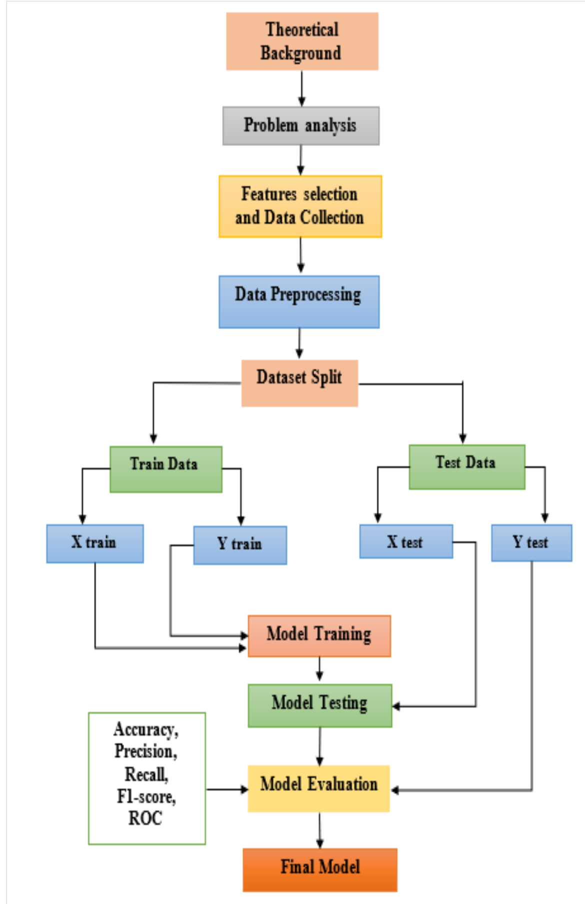

Figure 1 Workflow diagram

This research investigates chemical measurements of soil. Measuring existing chemical features is necessary to categorize soil, and determining the soil's chemical composition is essential. The total number of soil data used in this work is 1542. Soil samples were appropriately collected from agricultural land. Samples were taken six inches below the field's surface from the upper level of soil. Then, these soil samples were tested in Bangladesh's soil resource development institute (SRDI) laboratory.

For the proper prediction of crop cultivation, it is necessary to select significant attributes for crops. For crop cultivation, the soil is an essential attribute. Different types of crops grow on different types of soil. On various types of soil, many crops can flourish. There are several soil types in Bangladesh, including clay, clay loam, loam, sandy clay, sandy loam, sand, silty clay, silty clay loam, silt loam, silt, and others. Three types of soil classes—clay, clay loam, and loam are used in work from these different soil types. Four input attributes are selected for the prediction of crop cultivation, and the details are shown in *Table 3*. Here, input attributes are location, soil class, season, and pH.

Table 2 Attributes list and description of soil dataset

Attributes	Description
pH	pH value of soil
Salinity	D S\m
Organic matter	Percentage
Sulfur	Microgram/per gram soil
Nitrogen	Percentage
Phosphorus-B	Microgram/ per milliliter soil
Boron	Microgram/per milliliter soil
Soil class	Clay, clay loam, and loam

Table 3 Attributes list and description of crop dataset

Attributes	Description
Location	Different zones of Noakhali
Soil class	Clay, clay loam, and loam
Season	Rabi, kharif, and all_season
pH	pH level of soil
Crops	Rice, maize, wheat, potato, tomato, and others.

3.2Data preprocessing

Variables of different ranges are seen in the dataset, so it is necessary to preprocess the data. Data normalization is the name of a technique used to preprocess data. The data is normalized using the following Equation 1:

$$Xnew = \frac{X - Xmin}{Xmax - Xmin} \quad (1)$$

Where Xnew is the updated normalized value.

For each feature, a value is assigned between '0' and '1', with '0' being the lowest value and '1' representing the highest value. Accordingly, all other features are transformed between '0' and '1'.

This work uses the mean and median imputation methods to handle missing values. Missing values in the dataset are replaced with substituted values. Mean imputation involves replacing missing values with the mean (average) of the non-missing values in the column. Median imputation involves replacing missing values with the median (the middle value) of the non-missing values in the column.

To balance the dataset, we use the oversampling method. *Oversampling* is a technique used to address class imbalance in datasets by increasing the number of instances in the minority class. This technique helps to balance the class distribution and ensures that the ML model can learn effectively from all classes.

Table 4 shows the snapshot of the soil sample raw dataset. Here, pH, Salinity, Organic matter, Sulfur, Nitrogen, phosphorus-B, and Borer are input parameters and soil class is used as output. *Figure 2* shows the data variation of two areas through the probability distribution plot using two preprocessed soil sample data.

3.3 Distinct techniques for soil classification and crops cultivation prediction

Several techniques are used to classify soil and predict crop cultivation. Techniques such as MLPNN, DT, SVM, NB, RF, and LR were used for prediction. Those algorithms are selected based on various factors such as the problem domain, previous literature review, data characteristics, interpretability requirements, computational resources, and the need for accuracy. Understanding each algorithm's strengths and weaknesses helps to select the most appropriate model. Additionally, experimenting with multiple algorithms and comparing their performance using cross-validation will help choose the best-performing model for this crop prediction task.

Table 4 Soil sample raw dataset portion

pH	Salinity	Nitrogen	Phosphorus(B)	Sulfur	Boron	Organic matter	Soil class	Location
7.1	7.9	0.07	3	106	2	1.34	Loam	Sudaram
6.4	1.33	0.18	21	18	1.35	2.48	Loam	char panaullah
5.9	7.22	0.16	22	65	1.69	3.23	Clay loam	dakshin char majid
7.6	4.3	0.08	3	148	1.6	1.66	Loam	Sudaram
4.8	0.68	0.04	2	33	0.51	3.2	Clay loam	Begumgonj
5.4	0.73	0.11	16	44	0.62	2.98	Clay loam	Noakhali
7.8	1.5	0.08	3	96	0.91	1.35	Loam	Noakhali
6.4	6.25	0.13	20	117	1.02	2.68	Clay	Noakhali
6	0.96	0.1	105	20	0.85	2.93	Clay	uttar kachapia
7.5	0.42	0.05	0	42	0.56	0.96	Loam	Subarnochar
7	1.43	0.14	14	27	0.73	2.89	Clay loam	dakshin char majid
5.4	1.7	0.1	1	40	0.62	0.1	Loam	Sonaimuri
6.5	0.63	0.12	11	24	1.12	3.41	Clay	char vagga

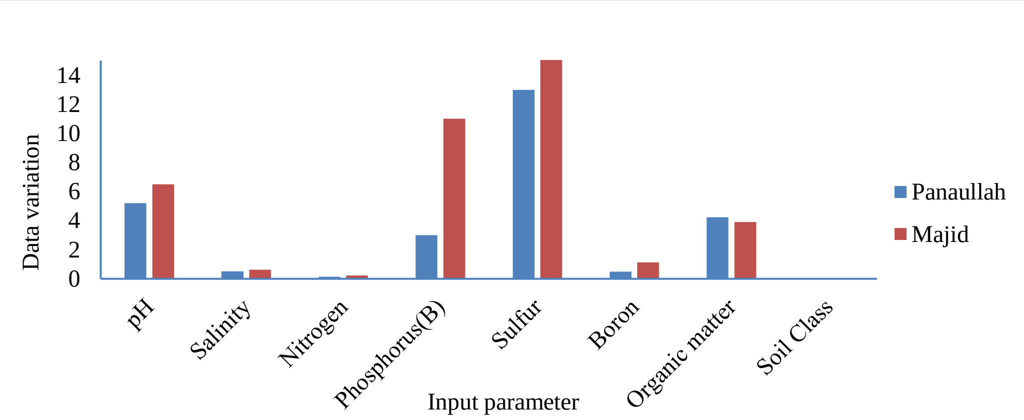

Figure 2 Probability distribution plot shows the data variation in two areas

a)Multilayer perceptron neural network

A MLP is a type of ANN model. It consists of multiple layers of neurons organized feedforward, meaning the data flows in one direction—from the input layer, through one or more hidden layers, to the output layer. Each neuron in a layer is connected to every neuron in the next layer, making MLPs fully connected networks. The input layer receives the raw input features and passes them to the first hidden layer. Each hidden layer then performs computations using weighted sums and activation functions, introducing non-linearity into the model. The output layer produces the final predictions [62]. The detailed network structures of the MLP model are shown in *Table 5*.

Table 5 MLP structure

Content	Input unit	Hidden layer 1 & 2	Activation function	Output unit
Soil Classification	7 neurons	8 to 50 neurons	Rectified Linear Unit (Relu)	Soil class
Crop Cultivation Prediction	4 neurons	8 to 50 neurons	Relu	Various crops

For soil classification, this research uses the input, hidden, and output layers of MLP. The input layer is composed of the previously mentioned seven separate soil features. Three soil types are included in the output layer: clay, clay loam, and loam. This MLP model has seven neurons in the input layer, two hidden layers, each with eight neurons, and an output layer with three neurons, as illustrated in *Figure 3*. The number of hidden layers can be adjusted as necessary, either reduced or increased. For crop prediction, the MLP model is composed of four neurons in the input layer, two hidden layers, each with eight neurons, and an output layer with 12 neurons. The network structure of MLP for crop prediction is shown in *Figure 4.*

b)Support vector machine (SVM)

SVM is a powerful supervised ML technique that works best on classification problems. It can also be used for regression problems. The SVM model tries to find a hyperplane that best distinguishes between the two groups. The best hyperplane is the one that has the farthest distance from both groups. This process involves identifying multiple hyperplanes that classify the data optimally and selecting the best hyperplane, typically the one farthest from the data points or with the most significant margin. The data points closest to the hyperplane are support vectors [63]. Kernel functions are a technique for transforming data from its input form into the format needed for processing it. SVM can employ kernels such as linear, polynomial, radial basis function (RBF), and sigmoid. Each kernel is suitable for different types of data and tasks. In this work, a linear kernel SVM is used.

Linear Kernel SVM: It is simple, efficient, and beneficial for high-dimensional data. The linear kernel's formula is expressed as follows in Equation 2:

$$k\left(x_i x_j\right) = \sum x_i \times x_j \qquad (2)$$

In the Equation 2, the product between two vectors, such as x_i and x_j, is calculated as the sum of the products of each corresponding pair of input values.

In the SVM model, the random search technique is employed for hyperparameter tuning. This technique involves randomly selecting combinations of hyperparameters to evaluate the performance.

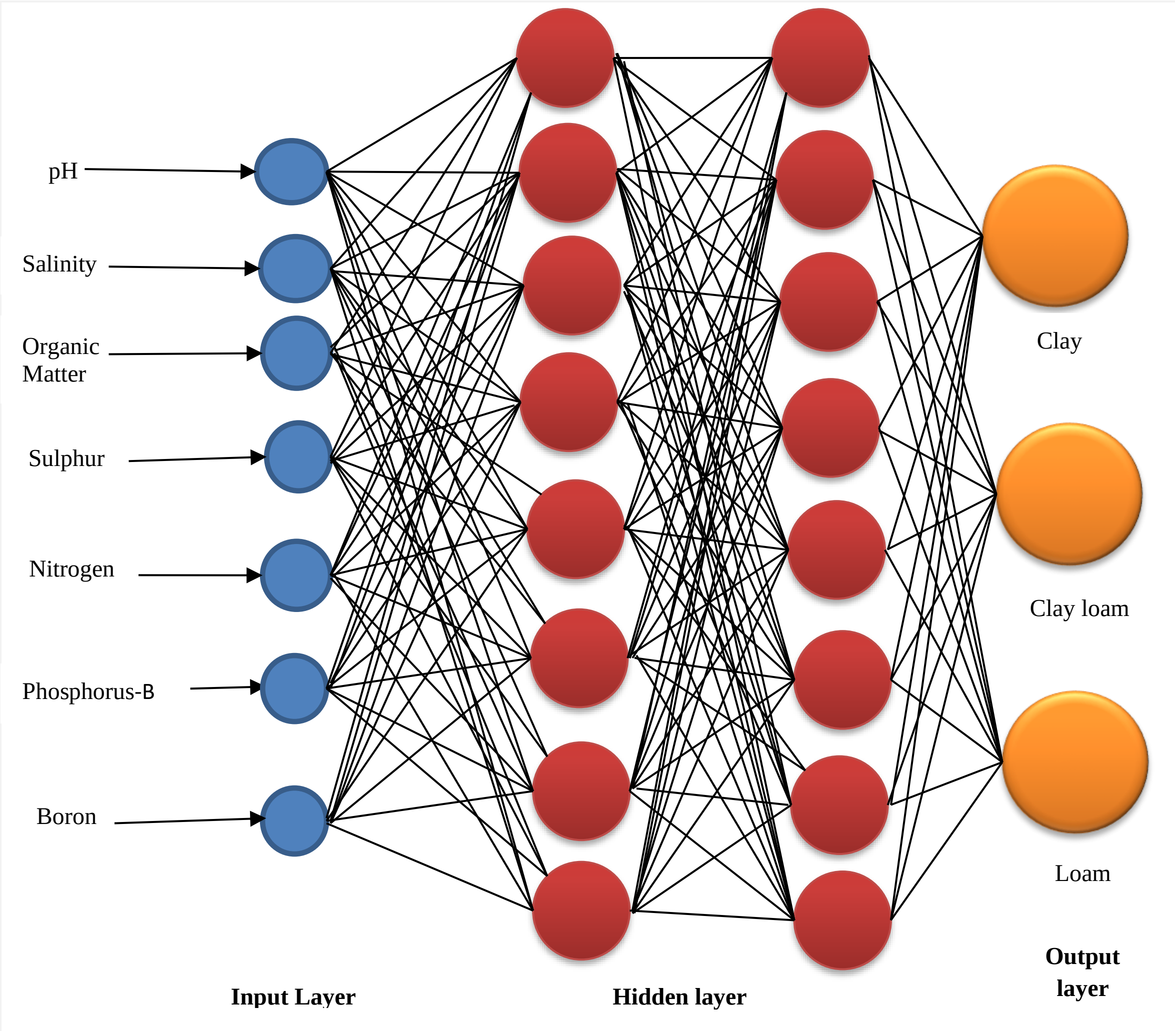

Figure 3 MLPNN structure for soil classification

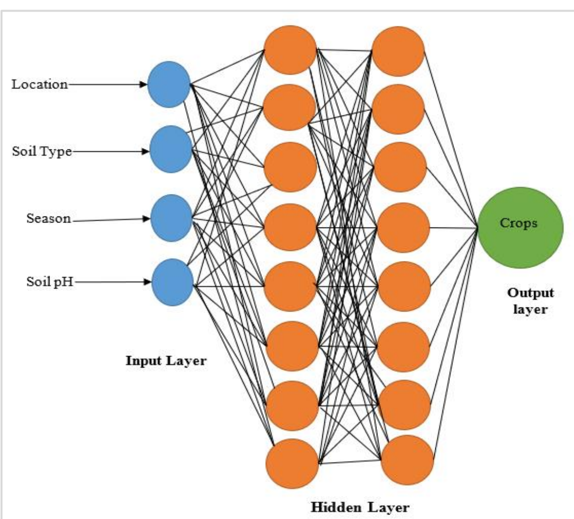

Figure 4 MLPNN structure for crop prediction

c)Random forest

The RF algorithm operates through an ensemble learning approach, combining the strengths of DTs with the principles of randomness and aggregation. At its core, RF constructs multiple DTs independently on random subsets of the training data using a process called bagging (bootstrap aggregating). During the construction of each tree, a random subset of features is selected at each node split, introducing further randomness to the learning process. This random feature subsampling ensures that each tree learns to make decisions based on different subsets of features, thereby diversifying the trees in the forest [64]. RF constructs DTs using diverse data samples and aggregates their maximum votes for soil classification and crop prediction tasks. The following steps are involved in the RF model:

Step 1: N random records are sampled from a dataset containing K total records. These random subsets, also known as bootstrap samples, are created by randomly selecting data points with replacements from the original dataset.
Step 2: Individual DTs are constructed using different subsets of the data.
Step 3: Each tree independently generates an output based on the subset of data it was trained on.
Step 4: The final output is determined based on majority voting for classifying soil and crop prediction.

d)Decision tree

The DT is a tree-structured supervised learning approach primarily employed for addressing classification and regression tasks, with a predominant application in classification problems. This hierarchical tree structure comprises a root node, branches, internal, and leaf nodes. The root node, serving as the starting point of the DT, lacks incoming branches. Branches connect to decision nodes extending from the root node to internal nodes. These internal nodes serve as decision points. The leaf nodes within the tree represent the potential outcomes of the dataset [65].

Here, the DT model is trained using the pre-processed soil data for classification. During training, the DT algorithm recursively splits the data based on the values of different attributes to create a tree-like structure. At each tree node, the algorithm selects the attribute that best separates the data into homogeneous groups (i.e., groups with similar soil types). The splitting process continues until a stopping criterion is met, such as reaching a maximum tree depth, having a minimum number of samples in each leaf node, or achieving a certain purity level (e.g., Gini impurity or entropy). The entropy or information gain is calculated through the following Equation 3:

$$E(S) = \sum_{i=1}^{n} - p_i log_2 p_i \quad (3)$$

Where, p_i is the probability of i_{th} event occurring. The goal is to find splits that maximize the reduction in entropy, leading to more homogeneous data groups at each node.

e) Logistic regression

LR is a statistical model used in ML for classification tasks, mainly for solving predictive analytics problems. When the target (output) variable is categorical, it is applied. It makes the relationship between a set of independent variables and dependent variables. The Logistic function (sigmoid function) converts anticipated values to probabilities. It maps any real value between 0 and 1 into another value. Since the LR's value must lie within the range of 0 and 1, it can never go above or below this limit, resulting in a curve shown in *Figure 5.*

LR is a valuable tool for soil classification and crop cultivation prediction. The model analyzes various soil attributes, such as pH levels, salinity, and other nutrient contents, to categorize soil class. Through training, LR estimates the probability that a soil sample belongs to a specific class, offering insight into which attributes influence soil type classification. Similarly, LR predicts the likelihood of successful crop growth based on soil type in crop cultivation prediction. Farmers can make informed choices about crop selection and land use by examining the relationship between soil class and crop suitability.

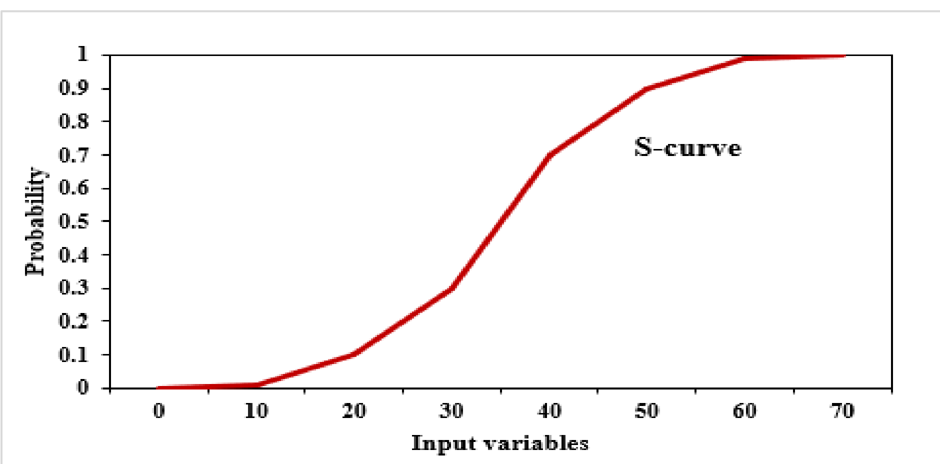

Figure 5 An example of a logistic function

f)Naïve bayes (NB)

NB stands as a probabilistic supervised ML algorithm crafted on the principles of Bayes theorem. Its primary application lies in addressing diverse classification problems, particularly prominent in text categorization scenarios with substantial training datasets. The Naive Bayes Classifier facilitates the rapid development of ML models, enabling precise predictions. At the core of this algorithm, Bayes Theorem assesses the probability of an event transpiring based on the likelihood of a prior event. The mathematical expression for Bayes' theorem is as follows in Equation 4:

$$P(A|B) = \frac{P(B|A).P(A)}{P(B)} \tag{4}$$

Here, P(A|B) denotes how often A occurs provided that B occurs, P(B|A) denotes how often B occurs given that A occurs, P(A) is how probable A is on its own, and P(B) denotes how likely B is on its own.

For soil classification, the NB model used soil samples as feature vectors. Using the training data, it estimated the likelihood of observing the feature values given each soil type. To classify a new soil sample, Naive Bayes computes the posterior probability of each soil type given the observed features using Bayes' theorem. The class with the highest posterior probability is assigned as the predicted soil type.

3.4 Training and testing

Various ML techniques—including SVM, DT, MLPNN, RF, NB, and LR—have been used to classify soil and ascertain crop cultivation. The total soil data used in this work is 1542, and the total crop data is 1162. The entire dataset is partitioned into training and testing sets, as outlined in *Table 6*. All models undergo training utilizing 70% of the data. Testing is done on 30% of the data.

Table 6 outlines the configuration and parameters used in all models training for classifying soil and predicting crops. The training process is set to run for 100 iterations, allowing the model to adjust its parameters to improve accuracy repeatedly. A random state value of 42 is used to initialize the random number generator, ensuring reproducibility of results. These random processes will yield identical outcomes every time the code is executed.

Additionally, 5-fold cross-validation is employed, dividing the dataset into five equal parts. The model is trained and tested five times, each with a different part as the test set and the remaining parts as the training set. This technique provides a more reliable assessment of the model's performance and helps prevent overfitting.

The verbose parameter is set to 2, meaning detailed information about the training process will be printed in the output logs. Finally, the n_jobs parameter is set to -1, indicating that all available CPU cores are used for computation.

Table 6 Train-test ratio and other parameters value

Contents	**Measurement**
Training part	70%
Testing part	30%
Iteration	100
Random state	42
Cross-validation(cv)	5
verbose	2
n_jobs	-1

3.4.1 Hyperparameter optimization

A random search technique is used to choose the parameters for each algorithm. This hyperparameter tuning technique is used in ML to optimize model performance by randomly sampling from a specified range of hyperparameters. Unlike grid search, which exhaustively evaluates all possible combinations, random search randomly selects a subset, making it more efficient. The process involves defining a hyperparameter space, randomly sampling combinations, and training and evaluating the model for each sample using cross-validation or a validation set. The hyperparameter space for the RF model is shown in *Table 7*. The hyperparameter space for the MLP model is shown in *Table 8.*

Table 7 Hyperparameter tunning of RF

Parameters	**Values**
'n_estimators'	randint(10,200)
'max_features'	['auto', 'sqrt', 'log2']
'max_depth'	randint(1,50)
'min_samples_split'	randint(2, 20)
'min_samples_leaf'	randint(1, 20)
'bootstrap'	[True, False]

Table 8 Hyperparameter tunning of MLPNN

Parameters	**Values**
'hidden_layer_sizes'	8 to 50 neurons
'activation'	('relu')
'solver'	('adam', 'sgd')
'alpha'	uniform(1e-5, 1e-1)
'learning_rate_init'	uniform(1e-5, 1e-1)
'batch_size'	randint(16, 256)

3.5Performance analysis

This study was carried out to assess the performance and usefulness of different classification algorithms for predicting crop cultivation through classifying soil. A model's effectiveness is evaluated based on several performance metrics. The essential performance parameters in this case are the confusion matrix, receiver operating system (ROC), precision, recall, F-measure, accuracy, and error. We will select the best method by comparing the performance metrics of different models.

a) Confusion matrix

The confusion matrix is a widely adopted and straightforward metric employed to assess the accuracy and correctness of a model. It finds particular utility in classification problems featuring two or more possible class outputs. In the case of binary classification problems, the confusion matrix consists of a two-dimensional table encompassing "Actual class" and "Predicted class" in each dimension. Rows correspond to actual classifications, while columns correspond to predicted classifications. *Table 9* is a structure of the confusion matrix for the two classes.

Table 9 Confusion matrix

		Predicted	
		Positive (Class 1)	Negative (Class 2)
Actual	Positive (Class 1)	TP	FN
	Negative (Class 2)	FP	TN

Four basic features (numbers) constitute the confusion matrix, determining the classifier's measurement parameters. These are the four features are:

True positives (TPs): TPs are instances when the anticipated result and the actual class of the data unit are both accurate.

True negatives (TNs): TNs are instances in which both the expected and actual classes of a data point are incorrect.

False positives (FPs): FPs occur when the anticipated outcome is accurate, but the actual class of the data point is incorrect.

False negatives (FNs): FNs occur when the projected outcome is false, but the actual class of the data point is accurate.

b) Precision

The percentage of correctly classifying positive cases from those predicted to be positive is known as the precision value. It is the accuracy of the optimistic prediction is shown in Equation 5.

$$Precision = \frac{True\ Positive(TP)}{True\ Positive(TP) + False\ Positive(FP)} \quad (5)$$

For the prediction of crop cultivation, TP indicates the number of accurately predicted crops. Moreover, FP shows that the predicted class is correct, but the actual class of the data point was false. For crop cultivation prediction, precision is as follows in Equation 6:

$$Precision = \frac{Total\ number\ of\ accurately\ predicted\ actual\ crops}{Total\ number\ of\ predicted\ crops} \quad (6)$$

c) Recall or true positive rate

The recall is the percentage of cases correctly classified as positive TP out of those actually positive as shown in Equation 7.

$$Recall = \frac{True\ Positive(TP)}{True\ Positive(TP) + False\ Negative(FN)} \quad (7)$$

d)F-measure

F-measure combines recall and precision. It is also known as the F1-score. It is expressed as in Equation 8.

$$F_1 = 2 \times \frac{Precision * recall}{Precision + recall} \quad (8)$$

e)Accuracy

The number of correct forecasts divided by the overall dataset size yields accuracy (ACC). Accuracy is compared based on how well each of the four Features (TP, TN, FP, FN) performs in Equation 9.

$$Accuracy = \frac{(TP+TN)}{(TN+FP+FN+TP)} \quad (9)$$

f)Area under the curve (AUC) and ROC curve

The ROC curve depicts the TP (recall) against the FP ratio. To evaluate the overall performance of a classification model, metrics such as the AUC are employed and calculated based on the region under the ROC curve.

g)Training and validation loss:

Training loss is the error between the model's predictions and the actual values on the training dataset. Validation loss is the error between the model's predictions and the actual values on a separate validation dataset, which is not used for training. It is computed at the end of each epoch, using the same loss function as the training loss. An example of training and validation loss is shown in *Figure 6.*

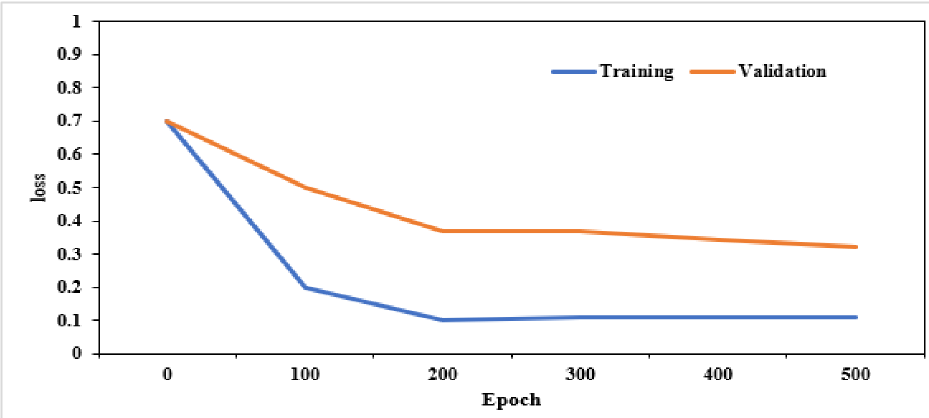

Figure 6 An example of training and validation loss plot

3.6Tools

This work utilizes Jupyter Notebook (version 6.5.4) and Google Colab to implement and evaluate ML models. Google Colab extends this functionality by offering free access to GPU and TPU acceleration, which can significantly speed up training times for deep learning models.

4.Results

The findings of the research are discussed in detail in 2 sections. The first section presents the results of soil classification. The second section shows the results of crop cultivation prediction based on soil classification. This study used ML methods for soil classification and crop cultivation prediction. Distinct Methods like SVM, DT, MLPNN, RF, NB, and LR are used. Performance parameters like precision, recall, F-measure, ROC, confusion metrics, and accuracy are used to compare the output of each chosen method. We discuss the results after comparing the performance of different methods.

4.1Soil classification

This section discusses the results of soil classification. Soil features pH, salinity, organic matter, nitrogen, phosphorus-B, sulfur, and boron, which are used as inputs. Six distinct ML techniques are used to classify soil accurately. The experimental results from the SVM, DT, MLPNN, RF, NB, and LR techniques are described below.

4.1.1Performance of distinct techniques for soil classification

Six different approaches are used to classify soil. In this part, the efficiency of each approach is determined.

a)Support vector machine (SVM)

Training and testing are done on the pre-processed soil data with a SVM algorithm. The confusion matrix produced by SVM is depicted in *Figure 7*. According to the confusion matrix figure, the output class matches the target class when the cells are green, and the opposite is true when the cells are white.

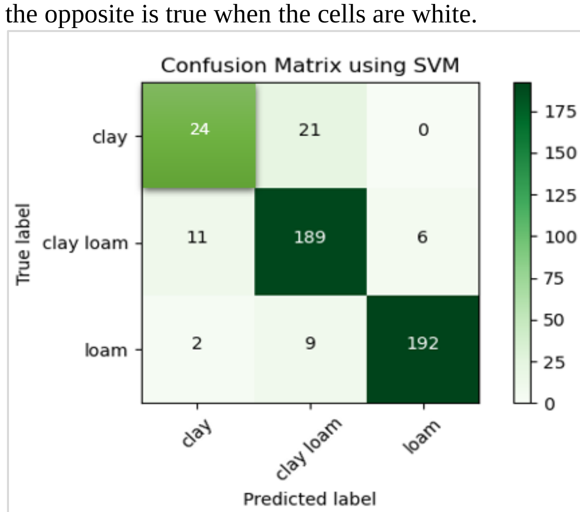

Figure 7 SVM exhibits this confusion matrix

The SVM model's confusion matrix, a reliable tool for evaluating classification accuracy, is described as follows.

- **Clay**: 24 were correctly classified as clay, 21 were classified as clay loam, and 0 were classified as loam.
- **Clay Loam**: 189 were correctly classified as clay loam, 11 were classified as clay, and 6 were classified as loam.
- **Loam**: 192 were correctly classified as loam, 2 as clay, and 9 as clay loam.

Some samples are wrongly classified. The misclassifications reasons:

- Some clay samples were classified as clay loam or loam. Because some clay loam and loam samples may have similar characteristics to clay in terms of feature composition, leading to misclassification.

- Similar reasoning applies to clay loam and loam samples being misclassified as clay or each other.

Figure 8 shows the ROC curve obtained from SVM. In the ROC curve, the actual positive rate is depicted on the y-axis, while the FP rate is represented on the x-axis. As depicted in *Figure 8*, the AUC value is 0.86 for clay, 0.92 for clay loam, and 1.00 for loam. The ROC value encompasses 92.65% of the region.

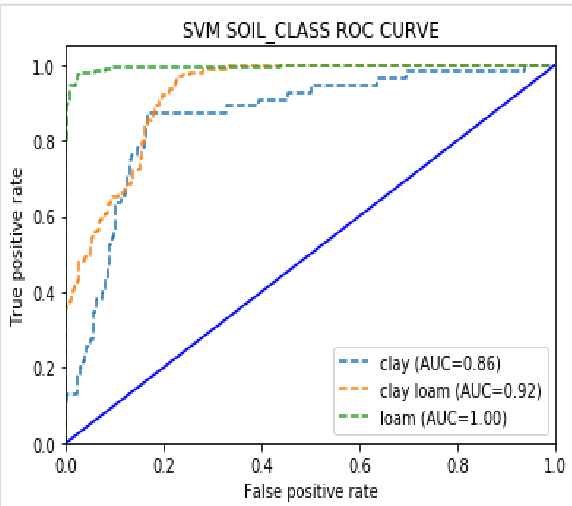

Figure 8 ROC curve obtained from SVM

Table 10 presents the results obtained from classifying soil using the SVM method. Notably, the SVM method achieves an accuracy of 89.21% in soil classification.

Table 10 Results of performance matrices obtained from SVM

Designation	Points
Precision	88.95%
F1-score	88.98%
Recall	89.21%
Accuracy	89.21%

b)Decision tree

After preprocessing, the collected soil dataset is trained and tested using the DT algorithm. Then, the confusion matrix built after running the model is shown in *Figure 9*. In the confusion matrix, the output class matches the target class when the cells are green, and the opposite is true when the cells are white. *Figure 10* shows the ROC curve obtained from DT. In the ROC curve, the y-axis represents the actual positive rate, while the x-axis indicates the false positive rate (FPR).

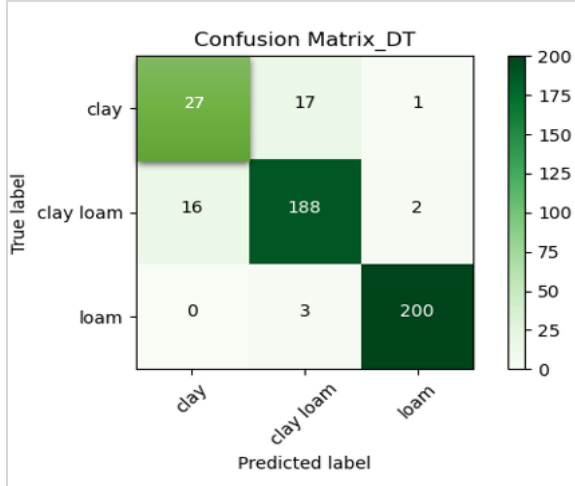

Figure 9 DT exhibits this confusion matrix

Figure 9 displays the confusion matrix for three soil classes: clay, clay loam, and loam. The DT algorithm demonstrates strong performance across three soil classes, namely clay, clay loam, and loam, as evidenced by the high number of TPs in the confusion matrix.

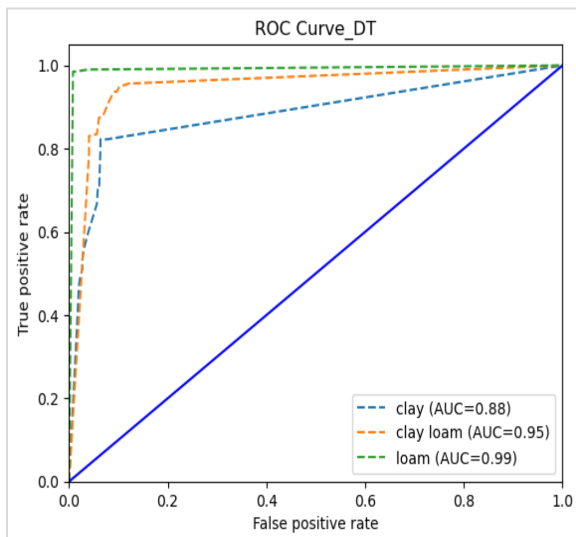

Figure 10 ROC curve obtained from DT

As shown in *Figure 10*, DT yields an AUC value of 0.88 for clay, 0.95 for clay loam, and 0.99 for the loam class. The ROC value encompasses 94.08% of the region. *Table 11* presents the soil classification results obtained using the DT algorithm, which indicates an impressive accuracy rate of 91.63%. *Figure 11* shows how decisions are made within the model for soil classification.

Table 11 Results of performance matrices obtained from DT

Designation	Points
Precision	91.46%
Recall	91.63%
F1-score	91.53%
Accuracy	91.63%

Figure 12 depicts the significant soil attributes that have a major impact on the DT model. It visually represents the prominent factors that guide the classification process, clearly understanding the variables driving the model's decisions.

Table 12 presents the outcomes regarding the feature significance impact for both the DT and RF models. Numerical values represent the significance of each feature in separate columns for DT and RF models. "Organic matter" and "Phosphorus" exhibit notably high significance values across both the DT and RF models.

Table 12 Soil feature significance impact on DT and RF model

Features	For DT	For RF
Organic matter	0.675120	0.597527
Phosphorus(B)	0.125441	0.110246
Salinity	0.063658	0.065968
Boron	0.047141	0.061432
pH	0.043273	0.061189
Nitrogen	0.034594	0.059580
Sulfur	0.010773	0.054058

c)Multilayer perceptron neural network

Training and testing are done on the preprocessed soil data with a MLPNN. Then, the confusion matrix built after running the model is shown in *Figure 13*. In the confusion matrix, the output class matches the target class when the cells are green, and the opposite is true when the cells are white. *Figure 14* shows the ROC curve obtained from MLPNN. In the ROC curve, the y-axis represents the actual positive rate, while the x-axis indicates the FP rate.

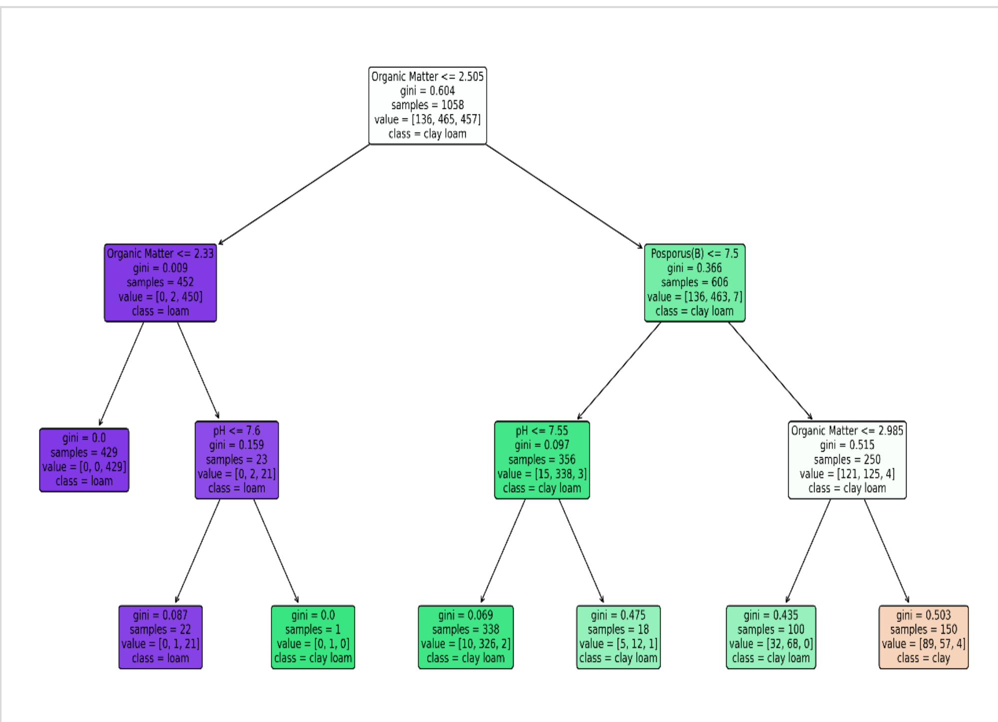

Figure 11 Decision tree for soil classification

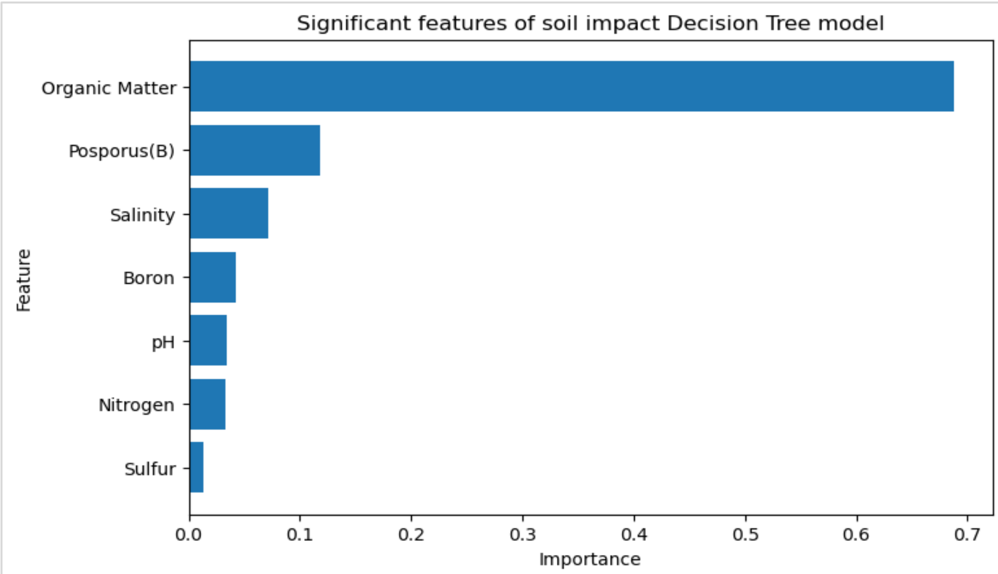

Figure 12 Impact of soil features on DT model

The MLPNN demonstrates strong performance across two soil classes, clay loam, and loam, as evidenced by the high number of TPs in the confusion matrix.

The AUC values presented in *Figure 14* further highlight this excellence, with values of 0.90 for clay, 0.95 for clay loam, and 0.98 for the loam class. The ROC value encompasses an impressive 94.59% of the region.

Table 13 provides a comprehensive performance analysis of the MLPNN, showcasing an accuracy rate of 85.68% achieved by the MLPNN algorithm.

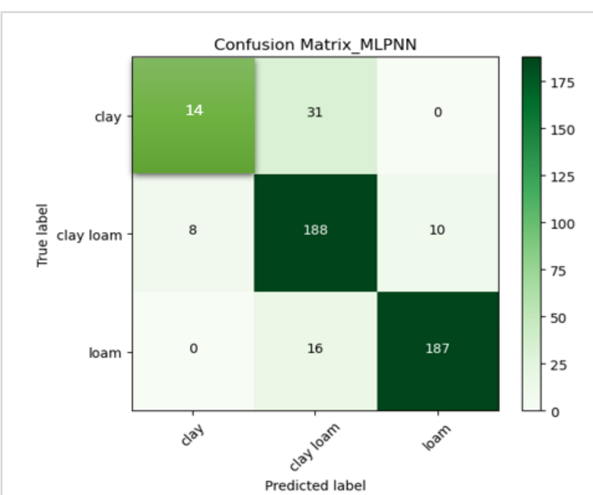

Figure 13 MLPNN exhibits this confusion matrix

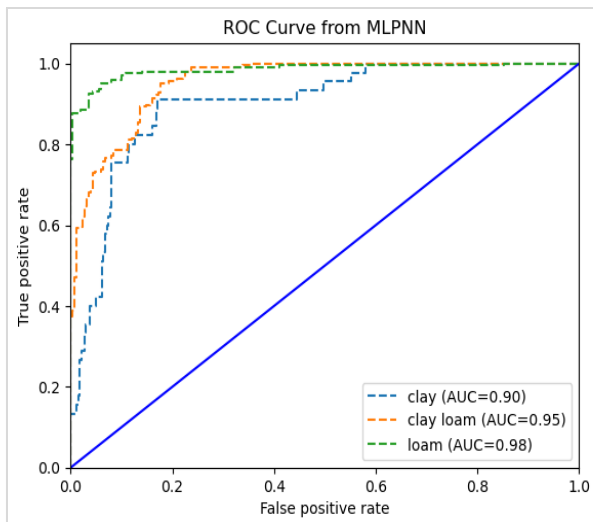

Figure 14 ROC curve obtained from MLP

Table13 Results of performance matrices obtained from MLP

Designation	**Points**
Precision	85.05%
Recall	85.68%
F1-score	84.64%
Accuracy	85.68%

d)Random forest

After preprocessing data, the collected soil dataset is trained and tested using the RF algorithm. Then, the confusion matrix built after running the model is shown in *Figure 15*. In the confusion matrix, the output class matches the target class when the cells are green, and the opposite is true when the cells are white. *Figure 16* shows the ROC curve obtained from RF. In the ROC curve, the y-axis represents the actual positive rate, while the x-axis indicates the FP rate.

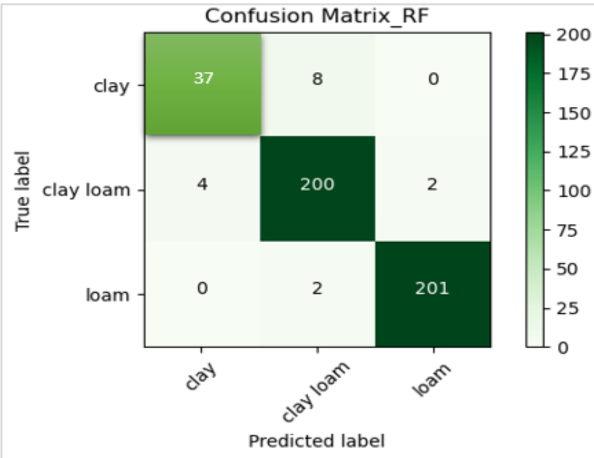

Figure 15 RF exhibits this confusion matrix

Figure 15 showcases a confusion matrix that compares the actual soil class labels with the predicted labels produced by the RF algorithm. This matrix includes three categories of soil: clay, clay loam, and loam. The high number of TP and the low number of FP and FN in the confusion matrix indicate that the RF algorithm is highly effective at accurately classifying each soil type.

As illustrated in *Figure 16*, RF achieves remarkable AUC values: 0.99 for clay, 0.99 for clay loam, and a perfect 1.00 for the loam class. Additionally, the ROC value covers an extensive 99.24% of the region. *Table 14* summarizes the results of soil classification using the RF algorithm, showcasing an impressive accuracy rate of 96.43%.

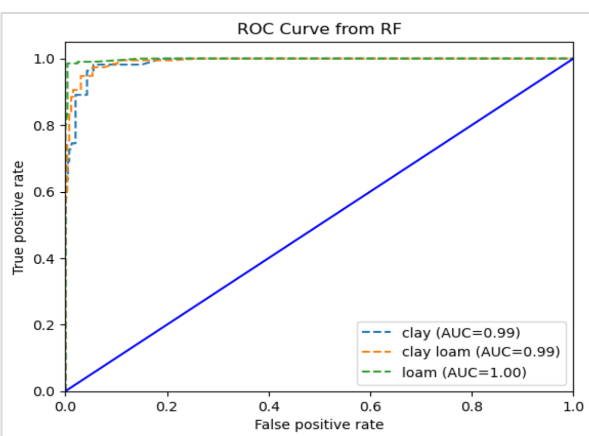

Figure 16 ROC curve obtained from RF

Table 14 Results of performance matrices obtained from RF

Designation	**Points**
Precision	96.43%
Recall	96.48%
F1-score	96.35%
Accuracy	96.48%

e)Logistic regression

Training and testing are done on the pre-processed soil data with a LR method. Then, the confusion matrix built after running the model is shown in *Figure 17*. In the confusion matrix, the output class matches the target class when the cells are green, and the opposite is true when the cells are white. *Figure 18* shows the ROC curve obtained from LR. In the ROC curve, the y-axis represents the actual positive rate, while the x-axis indicates the FPR.

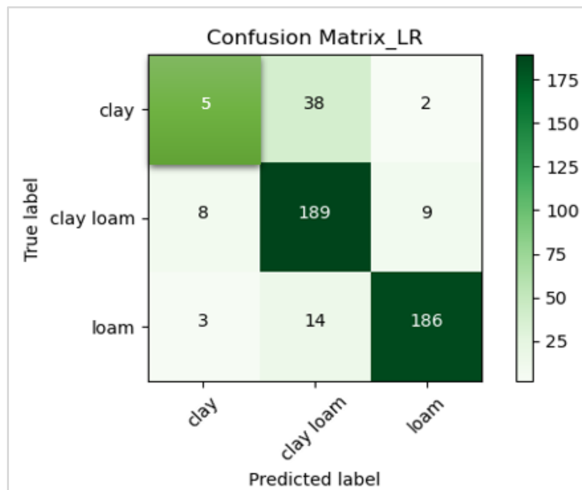

Figure 17 LR exhibits this confusion matrix

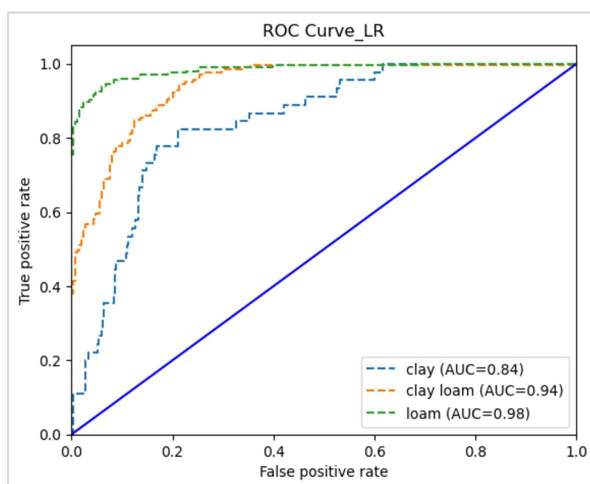

Figure 18 ROC curve obtained from LR

As depicted in *Figure 18*, the AUC values are 0.84 for clay, 0.94 for clay loam, and 0.98 for the loam class. Furthermore, the ROC value covers a substantial 92.18% of the region. The results of soil classification using the LR algorithm is presented in *Table 15*.

Table 15 Results of performance matrices obtained from LR

Designation	Points
Precision	80.9%
Recall	83.7%
F1-score	81.58%
Accuracy	83.7%

f)Naïve bayes

After preprocessing data, the collected soil dataset is trained and tested using the NB algorithm. Then, the confusion matrix built after running the model is shown in *Figure 19*. In the confusion matrix, the output class matches the target class when the cells are green, and the opposite is true when the cells are white. *Figure 20* shows the ROC curve obtained from NB. In the ROC curve, the y-axis represents the actual positive rate, while the x-axis indicates the FPR.

As demonstrated in *Figure 20*, NB provides AUC values of 0.91 for clay, 0.91 for clay loam, and 0.96 for the loam class. Additionally, the ROC value covers a substantial 92.51% of the region.

The results of soil classification using the NB algorithm are detailed in *Table 16*.

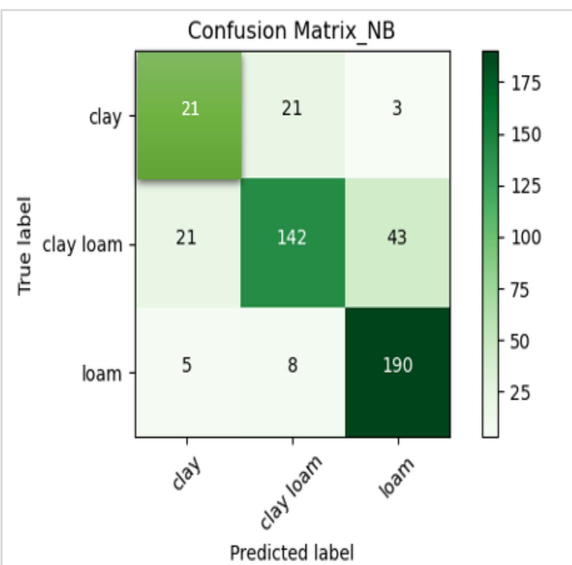

Figure 19 NB exhibits this confusion matrix

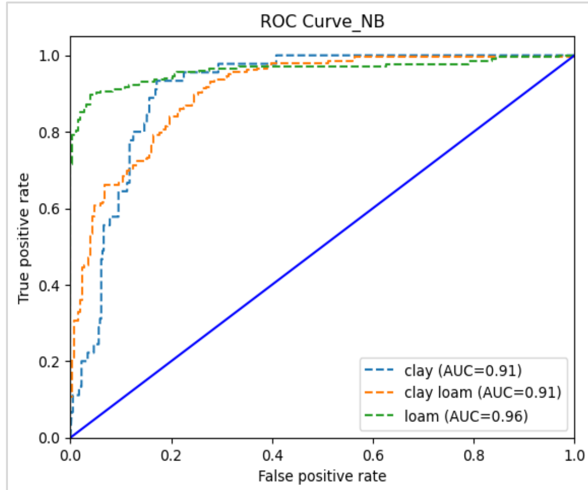

Figure 20 ROC curve obtained from NB

Table 16 Results of performance matrices obtained from NB

Designation	Points
Precision	78.11%
Recall	77.75%
F1-score	77.41%
Accuracy	77.75%

4.1.2Comparison of distinct techniques for soil classification

Comparisons among the performances of different ML algorithms- SVM, DT, MLPNN, RF, LR, and NB- are carried out in this section. Accordingly, *Table 17* compares different methods' prediction accuracy, error, and ROC. *Table 17* shows that the RF algorithm provides more accuracy than other algorithms in soil classification, at 96.48%. The ROC score for RF is also higher than that of other classifiers. *Table 18* compares the performance of the SVM, DT, MLPNN, RF, LG, and NB algorithms regarding prediction accuracy, precision, recall, and F1-score. *Table 18* demonstrates that the RF algorithm outperforms the other algorithms regarding maximum accuracy, precision, recall, and F1-score (performance metrics). Therefore, based on the discussion in this section, the RF algorithm is selected for soil classification.

Table 17 Comparison of accuracy, error, and roc between different algorithms

Name of algorithm	Accuracy (%)	Error (%)	ROC
SVM	89.21	10.79	92.65
DT	91.63	8.37	94.08
MLPNN	85.68	14.32	94.59
RF	96.48	3.52	99.24
LG	83.7	16.2	92.18
NB	77.75	22.25	92.51

Table 18 Comparison of performance between different algorithms

Name of algorithm	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-score (%)**
SVM	89.21	88.95	89.21	88.98
DT	91.63	91.46	91.63	91.53
Multilayer Perceptron NN	85.68	85.05	85.68	84.64
RF	96.48	96.43	96.48	96.35
LR	83.7	80.90	83.7	81.58
NB	77.75	78.11	77.75	77.41

Multi-class SVM [29]	Accuracy: 91.37%
Hybrid ANN [27]	Kappa:89.4% and RMSE:0.13
SVM [11]	Accuracy: 82.55%

4.2Crop cultivation prediction

Several ML techniques are applied to ascertain agricultural crop cultivation. SVM, DT, MLP, RF, LR, and NB are used for prediction. In this part, the effectiveness of each technique is assessed. The visualization of different types of soil is shown in *Figure 21*.

a) Clay b) Clay loam c) Loam

Figure 21 Visualization of different types of soil

Clay soil is sticky and smooth when wet. Hard and compact when dry. Clay loam is a mix of fine clay, coarser silt, and sand particles. To identify soil, moisten a soil sample and try to form a ball.

Clay: Forms a hard, durable ball.

Clay Loam: Forms a firm ball but is slightly crumbly.

Loam: Forms a loose, crumbly ball.

4.2.1Performance of distinct techniques for crop cultivation prediction

Six distinct techniques are applied for forecasting crop cultivation, and their efficiencies are assessed in this segment.

a)Support vector machine (SVM)

After preprocessing data, the crop dataset is trained and tested using the SVM algorithm. The predicted crop list includes rice, maize, beans, brinjal, potato, tomato, cauliflower, gourd, pumpkin, and others. *Figure 22* shows the confusion matrix generated by the SVM model for predicting crops. In this matrix, green cells indicate that the output class matches the target class, while white cells indicate a mismatch between the output and target classes. *Figure 23* exhibits the ROC curve derived from SVM model. On the ROC curve, the y-axis denotes the true positive rate, while the x-axis represents the FPR.

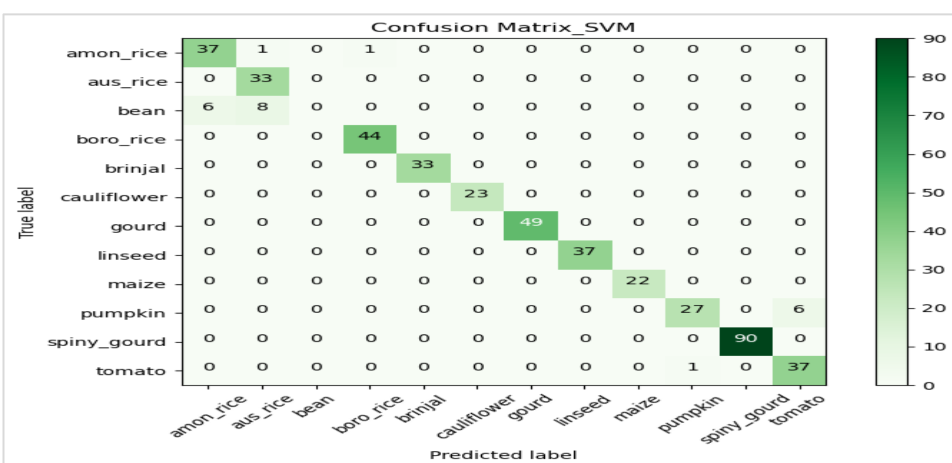

Figure 22 Confusion matrix using SVM

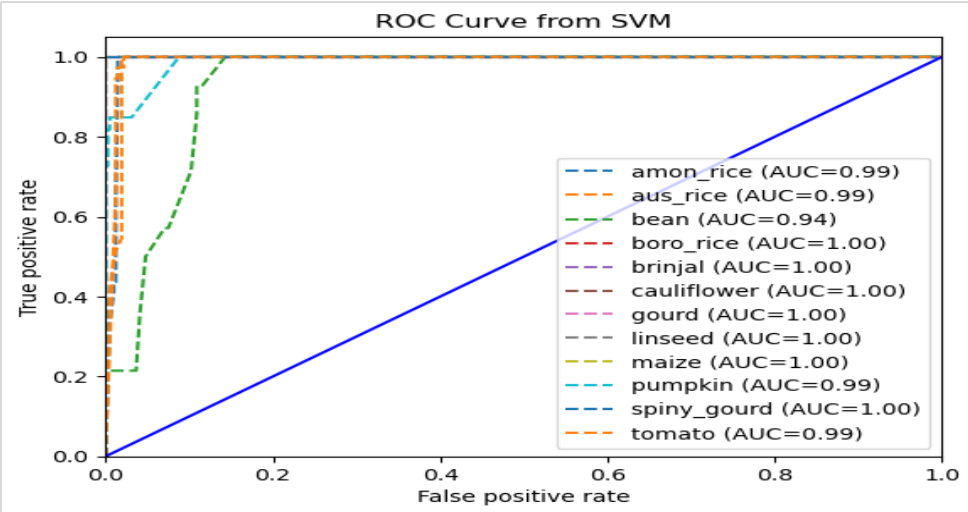

Figure 23 ROC Curve using the SVM model

Table 19 shows the results after crop cultivation prediction using the SVM algorithm. The SVM algorithm achieves a promising accuracy rate of 94.95%.

b)Decision tree

After preprocessing data, the crop dataset is trained and tested using the DT algorithm. *Figure 24* shows the confusion matrix generated by the DT. In this matrix, green cells indicate that the output class matches the target class, while white cells indicate a mismatch between the output and target classes.

Figure 25 exhibits the ROC curve derived from the SVM model. The y-axis denotes the true positive rate, while the x-axis represents the FPR.

Table 19 Results of performance matrices obtained from SVM

Designation	Points
Precision	93.53%
Recall	94.95%
F1-score	93.55%
Accuracy	94.95%
ROC	99.19%

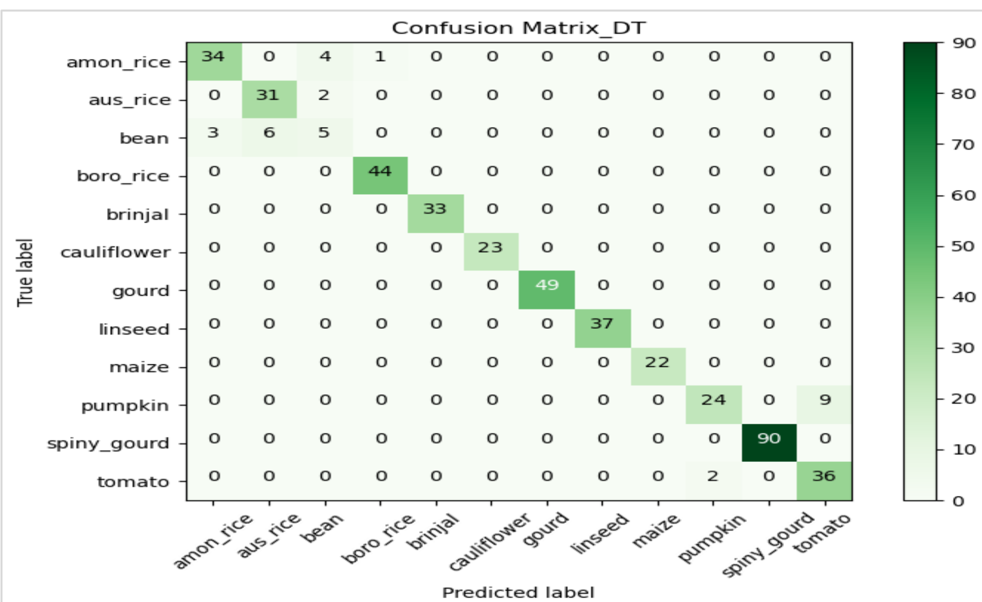

Figure 24 Confusion matrix using DT

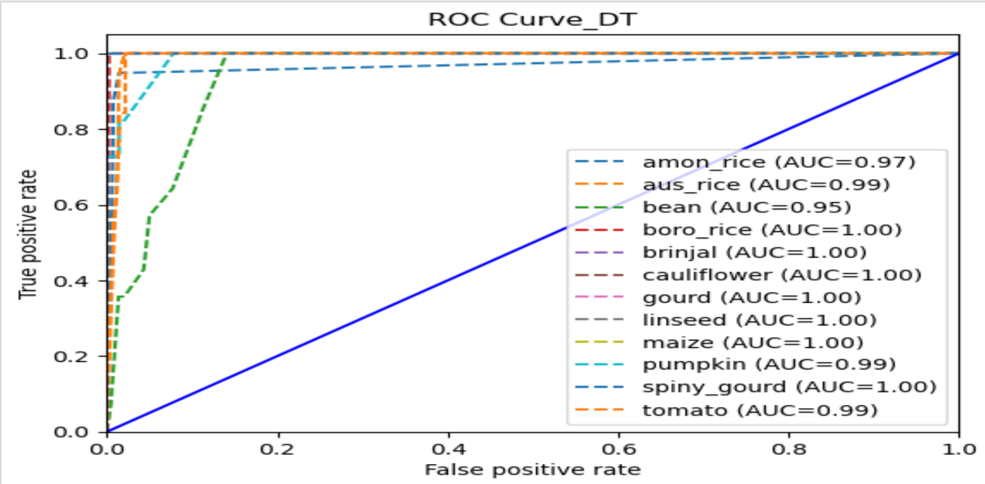

Figure 25 ROC Curve using the DT model

The results obtained after crop cultivation prediction using the DT algorithm are given in Table 20.

Table 20 Results of performance matrices obtained from DT

Designation	Points
Precision	93.01%
Recall	93.07%
F1-score	92.85%
Accuracy	93.07%
ROC	99.06%

c)Multilayer perceptron neural network

After preprocessing data, the crop dataset is trained and tested using the MLPNN algorithm. The results we get after crop cultivation prediction using the MLP algorithm are given in *Table 21*.

Table 21 Results of performance matrices obtained from MLP

Designation	Points
Precision	89.62%
Recall	91.87%
F1-score	90.11%
Accuracy	91.87%
ROC	95.13%

d)Random forest

After preprocessing data, the crop dataset is trained and tested using the RF algorithm. The results we get after crop cultivation prediction using the RF algorithm are given in *Table 22*.

Table 22 Results of performance matrices obtained from RF

Designation	Points
Precision	93.60%
Recall	93.85%
F1-score	93.68%
Accuracy	93.85%
ROC	99.03%

e)Logistic regression

After preprocessing data, the crop dataset is trained and tested using the LR algorithm. The results we get after crop cultivation prediction using the LR algorithm are given in *Table 23.*

Table 23 Results of performance matrices obtained from LR

Designation	Points
Precision	69.21%
Recall	70.35%
F1-score	69.22%
Accuracy	70.35%
ROC	97.60%

f)Naïve Bayes

After preprocessing data, the crop dataset is trained and tested using the NB algorithm. The results we get after crop cultivation prediction using the NB algorithm are given in *Table 24.*

Table 24 Results of performance matrices obtained from NB

Designation	Points
Precision	88.12%
Recall	88.34%
F1-score	88.03%
Accuracy	88.34%
ROC	98.56%

4.2.2Comparative analysis among distinct techniques

Table 25 compares the performance of SVM, DT, MLPNN, RF, NB, and LR regarding prediction accuracy, error, and ROC.

Table 25 demonstrates that the SVM algorithm's prediction accuracy, which is 94.95%, is higher than that of other algorithms. *Table 26* compares the performance of the SVM, DT, MLPNN, RF, LG, and NB algorithms regarding prediction accuracy, precision, recall, and F1-score. *It* demonstrates that the SVM algorithm outperforms the other algorithms regarding maximum accuracy, precision, recall, and F1-score (performance metrics). So, based on the discussion in this section, the SVM model is selected for crop cultivation prediction.

Table 25 Comparison of several algorithms' accuracy, error, and roc

Name of algorithm	Accuracy (%)	Error (%)	ROC
SVM	94.95	5.05	99.19
DT	93.07	6.93	99.06
Multilayer Perceptron NN	91.87	8.13	98.91
RF	93.85	6.15	99.09
LR	70.34	29.65	99.60
NB	88.34	11.66	98.56

Table 26 Comparison of performance between different algorithms

Name of algorithm	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
SVM	94.95	92.53	94.95	93.55
DT	93.07	93.01	93.07	92.85
Multilayer Perceptron NN	91.87	89.62	91.87	90.11
RF	93.85	93.60	93.85	93.68
LR	70.34	69.21	70.34	69.22
NB	88.34	88.12	88.34	88.03

5.Discussion

This study implemented the research using several ML models, such as SVM, DT, MLPNN, RF, LR, and NB. These methods played a leading role in completing our research. Peoples can now easily understand how to determine the soil class and ascertain crop cultivation according to soil type.

5.1Summary of key findings

The performance, such as precision, recall, F1-score, ROC, accuracy, error, roc curve, confusion matrix, and other metrics of distinct techniques, is determined, analyzed, and discussed here. For soil classification, it is found that the SVM, DT, MLPNN, RF, NB, and LR methods results are 89.21%, 91.63%, 85.68%, 96.48%, 77.75%, and 83.7% accuracy, respectively. *Figure 21* shows performance metrics representation graphically. The RF method has the highest prediction accuracy of 96.48% and is the best algorithm for classifying soil, according to all algorithms' experimental data analysis. On the other hand, the DT also performed well. MLPNN, SVM, and LR models provide adequate, promising accuracy. In contrast, NB are not treated well for classifying soil. Meanwhile, the RF method gives better accuracy so that it can be selected for soil classification prediction.

For crop cultivation prediction, it is found that SVM, DT, MLPNN, RF, NB, and LR methods provide perspective results of 94.95%, 93.07%, 91.87%, 93.85%, 88.34%, and 70.35% accuracy, respectively. The SVM method is the most effective technique for crop cultivation prediction because of its maximum prediction accuracy, 94.95%, based on the comparative analysis of experimental data of all algorithms. On the other hand, RF also performed well. DT, MLP, and NB provide promising accuracy. In contrast, LR is not treated well in this study. Ultimately, we can conclude that SVM gives better accuracy so that it can be selected for crop cultivation prediction.

5.2Interpretations

The high accuracy rates achieved by the RF and SVM algorithms indicate their robustness and efficiency in handling soil and crop data complexities. The RF model superior performance in soil classification can be attributed to its ability to handle large datasets with higher dimensionality and its effectiveness in reducing overfitting through ensemble learning. Similarly, the SVM model effectiveness in crop cultivation prediction is due to its capability to construct hyperplanes in a high-dimensional space, effectively separating different classes.

5.3Implications and limitations

These findings have significant implications for precision agriculture. Accurate soil classification allows for more tailored and efficient use of fertilizers and water resources, potentially leading to increased crop yields and sustainability. Similarly, precise crop cultivation predictions can aid farmers in selecting the most suitable crops for their soil type, ultimately improving agricultural productivity and profitability.

Despite the promising results, the study acknowledges that it has limitations. The dataset size and diversity might limit the generalizability of the findings across different regions and soil types.

5.4Recommendations and comparative analysis

A comparative analysis of the different algorithms based on their performance metrics reveals:

- **RF** stands out in soil classification due to its highest accuracy (96.48%) and lowest error rate (3.52%). It also performed well in crop cultivation prediction with a high accuracy (93.85%).
- **SVM** excels in crop cultivation prediction with the highest accuracy (94.95%) and performs well in soil classification (89.21%).
- **The DT performs strongly** in both tasks, making it a reliable alternative with high accuracy in soil classification (91.63%) and crop prediction (93.07%).
- **MLPNN** performs moderately in both tasks, indicating room for improvement, especially in handling complex soil-crop relationships.
- **LR** and **NB** showed lower accuracies, suggesting they are less effective for these specific applications.

Based on the comparative analysis, the study recommends using the RF algorithm for soil classification and the SVM for crop cultivation prediction.

5.5Result impact and overall analysis

The results of this study have significant implications for precision agriculture. By leveraging the power of advanced ML techniques, farmers and agronomists can make more informed decisions that enhance productivity and sustainability. The comparative study highlights the strengths and weaknesses of each algorithm, providing valuable insights into their applicability in different agricultural contexts. Overall, the findings indicate the potential of ML to revolutionize agricultural practices, leading to more efficient and sustainable food production systems. *Table 27* illustrates a comparative analysis between our experimental research and several existing works.

This study also concentrated on the various constraints faced by farmers during cultivation. Farmers don't get the desired crop production due to the prior traditional methods. In response to difficulties faced by farmers, this study classifies the soil and predicts crop cultivation based on soil class.

The findings from this study can found to be helpful in enhancing cultivation and making a substantial contribution to the economy. The complete list of abbreviations is shown in *Appendix I.*

Table 27 Comparison with existing works

Author, year	Dataset	Methods	Results
Mallick et al. [27], 2021	Wet-land mapping dataset	RF, SVM, KNN, ANN.	The proposed fused ANN-based hybrid model outperformed well than others in terms of accuracy (89.4% kappa and RMSE: 0.13)
Barman and Choudhury [29], 2020	Soil texture dataset	SVM	91.37% accuracy
Dash et al. [45], 2021	Crops data set	ML techniques.	SVM provides the highest accuracy in terms of F1-score 91%.
Waiker et al. [38], 2020	Crops dataset	ANN, NB	ANN gives 92% accuracy, and NB gives 76% accuracy for crop yield prediction.
Harlianto et al. [11], 2017	Soil related data	SVM, NN	The result shows that SVM, using a linear function kernel, outperforms the other algorithms. The SVM provides an accuracy of 82.35%.
Our Research Experiments	Soil data	SVM, DT, MLPNN, RF, LR, NB	RF outperforms others with 96.04% accuracy for soil classification.
	Crops data	SVM, DT, MLPNN, RF, LR, NB	SVM demonstrates superior performance, achieving an accuracy of 94.95% in predicting crop cultivation compared to other methods.

6.Conclusion and future work

This research has been done to classify soil and predict crop cultivation based on soil type. This research concentrated on various supervised classification methods to classify soil and forecast crop cultivation. Soil is an essential ingredient of agriculture. Here is the real-time dataset of soil and crops collected from different zones of Bangladesh. The algorithm used for prediction is SVM, DT, MLPNN, RF, LR, and NB. For soil classification, the prediction accuracy of RF is higher (96.48%) than the remaining five algorithms and a lower error rate (3.52%) than the other algorithms. All the models are trained and tested using the same input data (dimensions, features, and samples are all the same). Among these six techniques, DT yields the second-best performance for soil classification.

On the other hand, the SVM algorithm has a greater prediction accuracy (94.95%) than the remaining five algorithms for crop cultivation prediction and a lower error rate (5.05%) than the other algorithms. Among these six techniques, NB yields the second-best performance for crop cultivation prediction.

Future work will focus on expanding the dataset to include more diverse soil types and crops from various regions to improve model generalizability. Additional soil features and advanced ML techniques like deep learning will enhance prediction accuracy. Developing real-time soil and crop monitoring systems using IoT devices will facilitate continuous data collection and dynamic updates. Integrating climate and weather data into prediction models will provide more context-specific recommendations. Collaborating with agricultural experts to create user-friendly decision-support systems will ensure practical and beneficial applications for farmers.

The results of this research, focused on predicting crop cultivation based on soil type, will be disseminated to farmers through multiple channels to ensure practical usage of the information. Local workshops and training sessions will provide hands-on training and demonstrations to educate farmers on interpreting and applying the predictions. A user-friendly mobile application will also be developed to deliver crop cultivation predictions directly to farmers' smartphones, offering real-time updates, personalized recommendations, and alerts.

Acknowledgment

This work is funded and supported by the ICT Division, Ministry of Posts, Telecommunications, and Information Technology, Bangladesh, under the ICT fellowship scheme (tracking number 22FS25873). We express our gratitude to the Soil Resource Development Institute (SRDI) of the Government of the People's Republic of Bangladesh for testing our soil samples and providing essential soil data that contributed to the execution of our experiment.

Conflicts of interest

The authors have no conflicts of interest to declare.

Data availability

The datasets used in this study were gathered from various regions of Noakhali District, Bangladesh. The data are not publicly accessible but can be requested from the first author upon reasonable request.

Author's contribution statement

Fardowsi Rahman: Data collection, Conceptualization, Work design, Modeling, Coding, Simulation, Paper writing-Review, Editing, and Revision. **Md. Ashikur Rahman Khan:** Conceptualization, Supervision, Investigation, Manuscript revision, Paper writing, Review, Editing, and Revision. All authors reviewed the results and approved the final manuscript of this work.

References

[1] Eswaran H, Ahrens R, Rice TJ, Stewart BA. Soil classification: a global desk reference. CRC press; 2002.

[2] Van KT, Kassahun A, Catal C. Crop yield prediction using machine learning: a systematic literature review. Computers and Electronics in Agriculture. 2020; 177:105709.

[3] Elbasi E, Zaki C, Topcu AE, Abdelbaki W, Zreikat AI, Cina E, et al. Crop prediction model using machine learning algorithms. Applied Sciences. 2023; 13(16):9288.

[4] Hirel B, Le GJ, Ney B, Gallais A. The challenge of improving nitrogen use efficiency in crop plants: towards a more central role for genetic variability and quantitative genetics within integrated approaches. Journal of Experimental Botany. 2007; 58(9):2369-87.

[5] Rahman S, Hasan MK. Energy productivity and efficiency of wheat farming in Bangladesh. Energy. 2014; 66:107-14.

[6] Elsoragaby S, Yahya A, Mahadi MR, Nawi NM, Mairghany M. Energy utilization in major crop cultivation. Energy. 2019; 173:1285-303.

[7] Mishra S, Mishra D, Santra GH. Applications of machine learning techniques in agricultural crop production: a review paper. Indian Journal of Science and Technology. 2016; 9(38):1-4.

[8] PS MG. Performance evaluation of best feature subsets for crop yield prediction using machine learning algorithms. Applied Artificial Intelligence. 2019; 33(7):621-42.

[9] Elavarasan D, Vincent DR, Sharma V, Zomaya AY, Srinivasan K. Forecasting yield by integrating agrarian factors and machine learning models: a survey. Computers and Electronics in Agriculture. 2018; 155:257-82.

[10] Rajeswari AM, Anushiya AS, Fathima KS, Priya SS, Mathumithaa N. Fuzzy decision support system for recommendation of crop cultivation based on soil type. In 4th international conference on trends in electronics and informatics 2020 (pp. 768-73). IEEE.

[11] Harlianto PA, Adji TB, Setiawan NA. Comparison of machine learning algorithms for soil type classification. In international conference on science and technology-computer 2017 (pp. 7-10). IEEE.

[12] Yadav J, Chopra S, Vijayalakshmi M. Soil analysis and crop fertility prediction using machine learning. Machine Learning. 2021; 8(3):41-9.

[13] Bhargavi P, Jyothi S. Soil classification using data mining techniques: a comparative study. International

Journal of Engineering Trends and Technology. 2011; 2(1):55-9.
[14] Wu W, Li AD, He XH, Ma R, Liu HB, Lv JK. A comparison of support vector machines, artificial neural network and classification tree for identifying soil texture classes in southwest China. Computers and Electronics in Agriculture. 2018; 144:86-93.
[15] Bali N, Singla A. Emerging trends in machine learning to predict crop yield and study its influential factors: a survey. Archives of Computational Methods in Engineering. 2022; 29(1):95-112.
[16] Durai SK, Shamili MD. Smart farming using machine learning and deep learning techniques. Decision Analytics Journal. 2022; 3:100041.
[17] Mathi S, Akshaya R, Sreejith K. An internet of things-based efficient solution for smart farming. Procedia Computer Science. 2023; 218:2806-19.
[18] Palanivel K, Surianarayanan C. An approach for prediction of crop yield using machine learning and big data techniques. International Journal of Computer Engineering and Technology. 2019; 10(3):110-8.
[19] Sirsat MS, Cernadas E, Fernández-delgado M, Khan R. Classification of agricultural soil parameters in India. Computers and Electronics in Agriculture. 2017; 135:269-79.
[20] Radhika K, Latha DM. Machine learning model for automation of soil texture classification. Indian Journal of Agricultural Research. 2019; 53(1):78-82.
[21] Kiran MP, Deepak NR. Crop prediction based on influencing parameters for different states in india-the data mining approach. In 5th international conference on intelligent computing and control systems 2021 (pp. 1785-91). IEEE.
[22] Cai S, Zhang R, Liu L, Zhou D. A method of salt-affected soil information extraction based on a support vector machine with texture features. Mathematical and Computer Modelling. 2010; 51(11-12):1319-25.
[23] Jiang S, Yu J, Li S, Liu J, Yang G, Wang G, et al. Evolution of crop planting structure in traditional agricultural areas and its influence factors: a case study in alar reclamation. Agronomy. 2024; 14(3):1-24.
[24] Inazumi S, Intui S, Jotisankasa A, Chaiprakaikeow S, Kojima K. Artificial intelligence system for supporting soil classification. Results in Engineering. 2020; 8:100188.
[25] Agarwal R, Shekhawat NS, Luhach AK. Automated classification of soil images using chaotic Henry's gas solubility optimization: smart agricultural system. Microprocessors and Microsystems. 2021: 103854.
[26] Kodikara GR, Mchenry LJ, Van DMFD. Application of deep learning and spectral deconvolution for estimating mineral abundances of zeolite, Mg-sulfate and montmorillonite mixtures and its implications for Mars. Planetary and Space Science. 2022; 223:105579.
[27] Mallick J, Talukdar S, Pal S, Rahman A. A novel classifier for improving wetland mapping by integrating image fusion techniques and ensemble machine learning classifiers. Ecological Informatics. 2021; 65:101426.
[28] Azmin SN, Rahman HM, Lim NN, Arbaiy N. Machine learning for soil classification: challenges and opportunities. Journal of Applied Science, Technology and Computing. 2024; 1(1):29-38.
[29] Barman U, Choudhury RD. Soil texture classification using multi class support vector machine. Information Processing in Agriculture. 2020; 7(2):318-32.
[30] Ghadge R, Kulkarni J, More P, Nene S, Priya RL. Prediction of crop yield using machine learning. International Research Journal of Engineering and Technology. 2018; 5:2237-9.
[31] Angelaki A, Singh NS, Singh V, Sihag P. Estimation of models for cumulative infiltration of soil using machine learning methods. ISH Journal of Hydraulic Engineering. 2021; 27(2):162-9.
[32] Aydın Y, Işıkdağ Ü, Bekdaş G, Nigdeli SM, Geem ZW. Use of machine learning techniques in soil classification. Sustainability. 2023; 15(3):1-18.
[33] Nguyen MD, Costache R, Sy AH, Ahmadzadeh H, Van LH, Prakash I, et al. Novel approach for soil classification using machine learning methods. Bulletin of Engineering Geology and the Environment. 2022; 81(11):468.
[34] Guo W, Huang Y, Huang Y, Li Y, Song X, Shen J, et al. Develop agricultural planting structure prediction model based on machine learning: the aging of the population has prompted a shift in the planting structure toward food crops. Computers and Electronics in Agriculture. 2024; 221:108941.
[35] Kumar R, Singh MP, Kumar P, Singh JP. Crop selection method to maximize crop yield rate using machine learning technique. In international conference on smart technologies and management for computing, communication, controls, energy and materials 2015 (pp. 138-45). IEEE.
[36] Priyadharshini A, Chakraborty S, Kumar A, Pooniwala OR. Intelligent crop recommendation system using machine learning. In 5th international conference on computing methodologies and communication 2021 (pp. 843-8). IEEE.
[37] Goswami T, Ganapathi DM, Goswami P. Soil classification and crop prediction using machine learning techniques. In intelligent robots and drones for precision agriculture 2024 (pp. 101-18). Cham: Springer Nature Switzerland.
[38] Waikar VC, Thorat SY, Ghute AA, Rajput PP, Shinde MS. Crop prediction based on soil classification using machine learning with classifier ensembling. International Research Journal of Engineering and Technology. 2020; 7(5):4857-61.
[39] Bhojani SH, Bhatt N. Wheat crop yield prediction using new activation functions in neural network. Neural Computing and Applications. 2020; 32(17):13941-51.
[40] Fegade TK, Pawar BV. Crop prediction using artificial neural network and support vector machine. In data management, analytics and innovation: proceedings of ICDMAI 2019, 2020 (pp. 311-24). Springer Singapore.
[41] Rale N, Solanki R, Bein D, Andro-vasko J, Bein W. Prediction of crop cultivation. In 9th annual computing

and communication workshop and conference 2019 (pp. 227-32). IEEE.
[42] Khan MA, Akther J, Rahman F. Comparative analysis of potato blight diseases BARI-72 and BARI-73 using a simplified convolutional neural network method. International Journal of Advance Technology and Engineering Exploration. 2024; 11(115):819-37.
[43] Rajeswari S, Suthendran K. C5. 0: advanced decision tree (ADT) classification model for agricultural data analysis on cloud. Computers and Electronics in Agriculture. 2019; 156:530-9.
[44] Murmu S, Biswas S. Application of fuzzy logic and neural network in crop classification: a review. Aquatic Procedia. 2015; 4:1203-10.
[45] Dash R, Dash DK, Biswal GC. Classification of crop based on macronutrients and weather data using machine learning techniques. Results in Engineering. 2021; 9:100203.
[46] Waghmare KA, Jhare SA. A review: improving the village wise soil parameter and predict the crop suggestion. International Journal of Scientific Research in Computer Science, Engineering and Information Technology. 2020; 6(3):120-5.
[47] Ansari N, Ratri SS, Jahan A, Ashik-e-rabbani M, Rahman A. Inspection of paddy seed varietal purity using machine vision and multivariate analysis. Journal of Agriculture and Food Research. 2021; 3:100109.
[48] Priya PK, Yuvaraj N. An IoT based gradient descent approach for precision crop suggestion using MLP. In journal of physics: conference series 2019 (pp. 1-17). IOP Publishing.
[49] Patil P, Panpatil V, Kokate S. Crop prediction system using machine learning algorithms. International Research Journal of Engineering and Technology. 2020; 7(2):748-53.
[50] Chakrabarty A, Mansoor N, Uddin MI, Al-adaileh MH, Alsharif N, Alsaade FW. Prediction approaches for smart cultivation: a comparative study. Complexity. 2021; 2021(1):5534379.
[51] Shastry KA, Sanjay HA. Cloud-based agricultural framework for soil classification and crop yield prediction as a service. Emerging research in computing, information, communication and applications. Springer, Singapore. 2019:685-96.
[52] Hasan M, Marjan MA, Uddin MP, Afjal MI, Kardy S, Ma S, et al. Ensemble machine learning-based recommendation system for effective prediction of suitable agricultural crop cultivation. Frontiers in Plant Science. 2023; 14:1234555.
[53] Bhuyan BP, Tomar R, Singh TP, Cherif AR. Crop type prediction: a statistical and machine learning approach. Sustainability. 2022; 15(1):1-17.
[54] Singh J, Pandey P, Pandey PK. Decision-making system for crop selection based on soil. In AI, edge and IoT-based smart agriculture 2022 (pp. 449-75). Academic Press.
[55] Nischitha K, Vishwakarma D, Ashwini MN, Manjuraju MR. Crop prediction using machine learning approaches. International Journal of Engineering Research & Technology. 2020; 9(8):23-6.
[56] hybrid approach for crop yield prediction using machine learning and deep learning algorithms. In journal of physics: conference series 2021 (pp. 1-10). IOP Publishing.
[57] Chandraprabha M, Dhanaraj RK. Machine learning based pedantic analysis of predictive algorithms in crop yield management. In 4th international conference on electronics, communication and aerospace technology 2020 (pp. 1340-5). IEEE.
[58] Yamaç SS, Todorovic M. Estimation of daily potato crop evapotranspiration using three different machine learning algorithms and four scenarios of available meteorological data. Agricultural Water Management. 2020; 228:105875.
[59] Khan MA, Afrin F, Prity FS, Ahammad I, Fatema S, Prosad R, et al. An effective approach for early liver disease prediction and sensitivity analysis. Iran Journal of Computer Science. 2023; 6(4):277-95.
[60] Chaity TT, Khan MA, Rahman F. Prediction of impacts and outbreak of COVID-19 on the society using distinct machine learning algorithms. Iran Journal of Computer Science. 2024; 7(2):155-75.
[61] Shetty SA, Padmashree T, Sagar BM, Cauvery NK. Performance analysis on machine learning algorithms with deep learning model for crop yield prediction. In proceedings of data intelligence and cognitive informatics 2020 (pp. 739-50). Springer Singapore.
[62] Park YS, Lek S. Artificial neural networks: multilayer perceptron for ecological modeling. In Developments in Environmental Modelling 2016 (pp. 123-40). Elsevier.
[63] Gaye B, Zhang D, Wulamu A. Improvement of support vector machine algorithm in big data background. Mathematical Problems in Engineering. 2021; 2021(1):5594899.
[64] Belgiu M, Drăguţ L. Random forest in remote sensing: a review of applications and future directions. ISPRS Journal of Photogrammetry and Remote Sensing. 2016; 114:24-31.
[65] De VB. Decision trees. Wiley Interdisciplinary Reviews: Computational Statistics. 2013; 5(6):448-55.

Fardowsi Rahman is a Lecturer in the Department of Computer Science and Engineering at Primeasia University in Dhaka, Bangladesh. She completed her undergraduate and postgraduate studies in Information and Communication Engineering at Noakhali Science and Technology University, Bangladesh. She was the University Rank Holder, securing the 1st merit position at both undergraduate and postgraduate levels. Her research interests include data Mining, Artificial Intelligence, Machine Learning, Neural Networks, Cloud Computing, Cybersecurity, and Big Data.
Email: fardowsirahman09@gmail.com

Md. Ashikur Rahman Khan is a Professor in the Department of Information and Communication Engineering at Noakhali Science and Technology University (NSTU), Bangladesh. He began his teaching career at NSTU in 2006 as a Lecturer. Throughout his professional journey at NSTU, he has served as the Dean of the Faculty of Engineering and Technology, along with numerous other roles and responsibilities at the university. Professor Khan was the Director of the ICT Cell at NSTU from 2021 to 2024. He has also contributed to the admission process for the General Science and Technology Universities, serving as a member of the technical committee for twenty-four universities in the country from 2020 to 2021. Before joining NSTU, he worked as an Engineer in the Inspection and Testing division of the Rural Electrification Board, Bangladesh, for over five years (2001-2006). He earned his Ph.D. in 2012 from the University Malaysia Pahang, Malaysia. Dr. Khan actively participates as a reviewer and editorial board member for various journals. His research interests include Advanced Machining, Artificial Intelligence, Neural Networks, Machine Learning, and Renewable Energy.
Email: ashik@nstu.edu.bd

Zareen Tasneem is an Assistant Professor in the Department of Computer Science and Engineering at Primeasia University, Dhaka, Bangladesh. She completed her M.Sc. and B.Sc. in Computer Science and Engineering from the Military Institute of Science and Technology. Her research interests include Machine Learning, Neural Networks, Bioinformatics, Big Data, Artificial Intelligence, Network Security, Human-Computer Interaction, and Theoretical Computer Science.
Email: zareen.tasneem@primeasia.edu.bd

Appendix I

S. No.	Abbreviation	Description
1	AI	Artificial Intelligence
2	ANFIS	Adaptive Neuro-fuzzy Inference System
3	ANN	Artificial Neural Network
4	AUC	Area Under the Curve
5	CSM	Crop Selection Method
6	DT	Decision Tree
7	FN	False Negative
8	FP	False Positive
9	FPR	False Positive Rate
10	HLM	Hierarchical Linear Model
11	KNN	K-Nearest Neighbor
12	LR	Logistic Regression
13	ML	Machine Learning
14	MLPNN	Multilayer Perceptron Neural Network
15	MSE	Mean Squared Error
16	NB	Naïve Bayes
17	NN	Neural Network
18	RBF	Radial Basis Function
19	RF	Random Forest
20	RMSE	Root Mean Square Error
21	ROC	Receiver Operating System
22	SRDI	Soil Resource Development Institute
23	SVM	Support Vector Machine
24	TN	True Negative
25	TP	True Positive