JCBI
Journal of Computing & Biomedical Informatics
ISSN: 2710 - 1606
ICASET 2024

Research Article
Collection: Intelligent Computing of Applied Sciences and Emerging Trends

Unveiling Soil Fertility Patterns via Image Analysis and Machine Learning for Accurate Crop Recommendations

Alishba Shahzad[1], Muhammad Ibrahim[1], Rana Muhammad Saleem[2*], Sibgha Zia[1], Sidra Habib[2], and Marwa Mahmood[2]

[1]The Islamia University Bahawalpur, Bahawalpur, 63100, Pakistan.
[2]University of Agriculture Fasialabad Sub Campus Burewala, Burewala, 61010, Pakistan.
[*]Corresponding Author: Rana Muhammad Saleem. Email: rana.m.saleem@uaf.edu.pk

Academic Editor: Salman Qadri **Published:** April 01, 2024

Abstract: Accurate fertilizer use is essential for precision agriculture. Crop recommendation systems should consider real-time data on soil fertility, crop type, soil nutrients, and environment to ensure sustainability and profitability. Due to its intricacy, real-time fertility of soil mapping is expensive, time-consuming, and costly. Using the Image's real-time soil fertility mapping data, we offer a crop selection approach based on artificial intelligence. It is suggested that an image architecture aid soil fertilizer mapping. Real agricultural fields have been employed to assess the precision of IMAGE-based fertility mapping employing the provided method. When evaluating IMAGE base fertility mapping compared to the process of soil chemical analysis, we look at its ability to accurately track nitrogen (N), phosphorus (P), potassium (K), and other environmental variables including humidity, precipitation, temperatures, and pH. Based on the soil fertilizer information, the following machine learning techniques are used to suggest crops: Logistic Regression (LR), Support Vector Machine (SVM), K-Nearest Neighbour (KNN), Random Forest (RF), Decision Tree (DT), and XGBoost (XGB). All machine learning algorithm performance is well good but the GNB and RF is performed most accurate performance as compared to other machine learning modules.

Keywords: Machine learning; Soil fertility; Crop recommendation; Decision tree (DT); Random forest (RF); Logistic regression (LR); Support vector machine (SVM); GNB, KNN.

1. Introduction

Agriculture is the backbone of every economy. Pakistan is mostly an agricultural nation; 47% of the nation's land area is dedicated to farming, and 61% of the population living in the countryside relies on farming as their source of income [1].In Pakistan, agriculture is a significant component of the economy, but new technologies and methods are causing it to slowly deteriorate. Food insecurity is a result of modern people's lack of knowledge about when and where crops should be grown, as well as seasonal climate changes that go against basic resources. There is no suitable technology or solution to deal with the issue.[2]. Farmers frequently choose the wrong crops, which reduces yields and revenue and makes it difficult for families to exist.[3].

Soil fertility is a key factor in determining how productive a farming system is. According to Swift & Palm (2000), it is more advantageous to think of soil fertility as an ecosystem concept that unifies the different activities of the soil rather than the most widely used definition of soil fertility, which is its ability to provide crops with nutrients [4]. Correctly nurturing the soil is crucial for the development of crops, as it helps retain fertility, improve yield, as well as safeguard the natural world. To find out which crops work best in what kinds of soil, it's essential to analyze the minerals in the soil. Finding out which nutrients are lacking right now and keeping an eye on how soil fertility fluctuates will help with good management of soil fertility. The key to obtaining excellent and profitable yields is managing the fertility of the soil at its optimum level [5].

To improve the chemical, biological, and physical properties of the soil, organic farming techniques focus on managing soil organic matter. They also control the amount of nutrients available to crops, livestock, and people, as well as diseases, pests, and weeds [6][7]. Fertilizers are used to improve productivity, leading to inefficient use of resources and soil deterioration. Soil degradation and environmental problems have become more complicated because of this [8]. The advent of intelligent and precise farming has boosted output without compromising quality. The key to long-term prosperity is improving production and making the most of scarce assets [9].

In this research, we offer an IMAGE-based soil fertility mapping architecture that uses real-time information on nutrients in the soil and environmental variables including rainfall, moisture content, temperature, and pH to produce suggestions for crops based on machine learning. What makes soil appropriate for cultivating crops is the quantity of water, air, and other nutrients it provides. Below, you can find the study's elementary achievements.

1. We propose architecture for IMAGE base fertility mapping to overcome the problems with the traditional method of evaluating soil fertility.
2. The study evaluates the accuracy of the proposed IMAGE base fertility mapping approach with respect to the established soil chemical analysis technique.

A precise fertilizer recommendation method based on machine learning models that employ in-the-moment identified soil fertility level and soil type is suggested by the study.

The study compares the efficacy of the crop recommendation system's use of the SVM, LR, DT, RF, KNN and GNB models [10]

Figure 1 shows every component/factor that is suggested for use in recommending crops and improving soil fertility. Instead of utilizing outdated techniques like mapping and tracking, modern sensor capacities, IMAGE, and ML technologies are ideal for regulating the fertility of soil effectively [11].

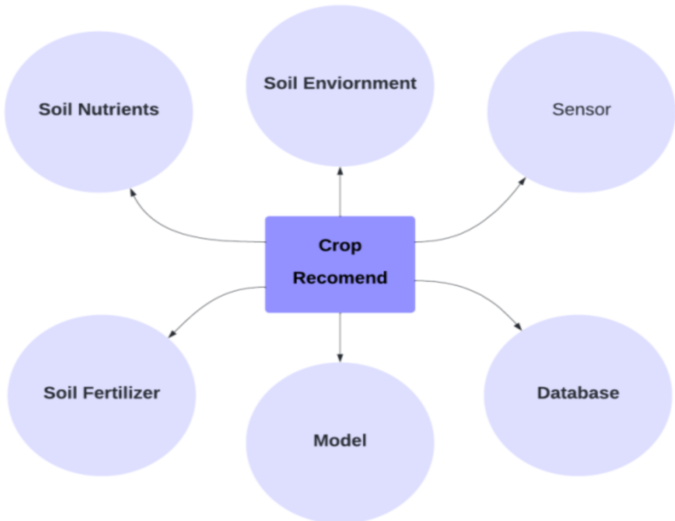


Figure 1. Component for soil fertility and crop recommend

IMAGE technology integrates the broader world into PC-based frameworks by enabling chosen things to be understood or controlled remotely. Development becomes a period for all major categories of electronic structures when IMAGE is expanded with sensors and activators. With the suspension of the envisioned make-up that can coexist inside the existing Internet domain, everything is unique [12]. From supervised to unsupervised to reinforcement learning, AI learns the algorithm. Unlike learning that is unsupervised, which relies just on inputs and does not specify desired output labels, supervised learning includes both inputs and outputs in its statistical model-building process. Mathematical frameworks are constructed using semi-supervised instruction using insufficient training data [13].

In the present piece, the following parts are presented: The literature review is in Section 2, the methodology is in Section 3, the outcomes and implementation are in the 4th section, and the conclusion and future application are in the 5th section.

2. Literature Review

V. by Pandith et al [5] Using an array of methods based on machine learning, we offer a way to predict mustard crop growth from soil study. Information on the mustard crop is gathered from samples of soil obtained in several districts of the Jammu region by the Department of Agriculture at Talab Tillo, Jammu. The information collected was subjected to supervised machine learning techniques, comprising K-Nearest Neighbour (KNN), Naïve Bayes, Multinomial Logistic Regression, Artificial Neural Network (ANN), and Random Forest. Soil is analysis on base soil nutrient parameter. Two machine learning algorithm KNN (94.13%) and ANN (99.44%) provide the most accurate result. Techniques for machine learning will be helpful to farmers in forecasting crop yields using soil data. One possible application case for Big Data in the years to come is food production forecasting utilizing massive soil data.

Sheeba et al. [14] Presents a machine learning algorithm for soil analysis and classification of micronutrients in IOT-enabled automated farms. The proposed system to classify soil on the base of its features and fertility indices and soil nutrients. This system used extreme learning method (ELM) to classify the soil, ELM identifies the nutrients of soil and uses different activation functions. It collects the soil data from from Tamil Nadu that have red colour, EC, N, P, K EC ets. ELM classifies the soil base on fertility indices that are useful for managing the soil and addressing nutrients deficiency issues. The result is obtained after ELM classification the soil is overall rich in potassium (35% of the samples), nitrogen (80% of the samples), and sulphur (75% of the sample) and sufficient or poor in magnesium, boron, zinc, and copper.

Khan et al. [9] Proposed the method uses machine learning and IMAGE technology to improve fertilizer recommendation accuracy in real-time soil fertility contexts according to soil fertility level, crop type, soil type. The context-aware fertilizer recommendation is implemented using SVM, KNN, LR and GNB models to access the performance of machine learning mode. Data is taken from department of agronomy Islamic university Pakistan. A practical implementation was tested in real crop fields, with mean differences of 0.34, 0.36, and -0.13 for N, P, and K observations. The Gaussian Naive Bayes (GNB) model was found to be the most accurate, achieving 96% and 94% accuracy from training and testing datasets.

N. L. Kalyani et al. [15] Proposed a novel approach using soil color measurements and deep learning techniques for estimating soil fertility. The authors used AI and image recognition for soil type classification using deep learning. They used artificial neural network model to classify soil types based on three-dimensional coordinates. This study used a CNN model for this purpose's model using five-layer convolution layer, relu layer, roiling layer, and fully connected layer to detect of red soil type with images using UI of web application. The model achieved 91% accuracy in identifying red soil and can predict significant nutrients for a specific soil.

D.Rangsung et al. [16] Propose a deep learning for crop fertilizer recommendation and disease prediction. The main aim of the paper of the project is to suggest the type of crop and the fertilizer that is suitable to be used in the soil based. The dataset collected from soil testing lab as well as general crop from online source. By using methods like DT, NB, SVM, LR, RM, and XGBoots, this approach will assist producers in making sound choices on the crops to cultivate according to certain environmental factors such as pH value, humidity, temperature, phosphorus, potassium, nitrogen, and rainfall. Based on the results of the contrast, XGBOOST is the most efficient algorithm. This system uses this suggested accurate result at right time for fertilization, real time crop analysis, select efficient parameter, making decision and improve yield.

T.Thorat et al.[17] This paper proposes a method for an intelligent insecticide and fertilizer recommendation system for smart farming. The author uses artificial intelligence and sensor technology to address the issue. It uses a soil NPK sensor for fertilizer recommendation on the base of soil analysis and uses machine vision, ANN, SVM, KNN, TPF-CNN algorithms and pest identification and insecticide recommendation. TPF-CNN is a deep learning approach that combines temporal, spectral, and spatial features of crop images. After comprising all algorithms, the TPF-CNN is performed with the best accuracy. It identifies pest images and recommends the appropriate insecticide within 10 seconds and has a 91% accurate result. This system helps the farmer to manage soil nutrients, reduce cost, reduce environmental impact, and gain maximum farm yield.

Jeedigunta et al. [18] Proposes the use of artificial neural network and supervised machine learning techniques to recommend optimal fertilizer and irrigation methods for crops. it uses machine learning algorithm such as logistic regression, support vector machine, CART, Naive Bayes, random forest classification and XGB. It tests the soil data on the base of different parameters like humidity, temperature, pressure, wind gust, wind speed and wind direction for fertilizer and irrigation of crop. After comprising the algorithm, the random forest 999.37) and Navie bayes (99.26) perform the best accurate result. The result of this study is improving crop yield and reducing water usage.

A. Gutiérrez et al. [19] this paper predicts the crop with the use of IMAGE ML. It is used to generate soil data in large remote agricultural areas and ML algorithms is used to predict the crop with help of parameters such as N, P, K, PH, temperature, humidity, and rainfall. The supervised machine learning method in weak select multilayer perception, jRip, and decision table classifier as the optimal model for classification. The accuracy of the selected classifier is 98.2273% with a maximum time of 8.05 to build a model. This paper provides a comprehensive review of IMAGE framework for precision agriculture.

T. G. Keerthan Kumar et al. [20] The paper proposes a machine learning-based system for analyzing soil properties, which grades soil and predicts suitable crops for a given land. It uses various algorithms such as Random Forest, support vector machine, and Gaussian NB. Soil nutrients like pH, EC, OC, S, K, Zn, Mn, B, and Soil Type are taken as soil feature variable. The data is collected at the ICRISAT Development Centre and Government of Andhra Pradesh. Machine learning algorithms are applied on this data set to predict soil rank and classification. After reviewing the results they produce, we are going to choose the Random Forest (72.74%) Classifier as our favorite. Farmers will have it easier with this arrangement in place.

kajla et al. [21][22] have described the image-based document recognition by using different techniques.in this paper graph-based methods have been used.it has used different evaluation metrices to evaluate the accuracy.

Table 1. Literature Review

Year	Sensor	Parameter	Objective
[5]2020	No sensor	Soil analysis (PH, EC,N,P, K,etc)	predict *mustard crop* yield from soil analysis using machine learning techniques
[14]2022	Specific sensor not mention	Fertility indices, micronutrients (soil nutri ents)	Develop an accurate and efficient system for analyzing soil samples and identifying micronutrient deficiencies to improve crop yield and quality.
[10]2022	NPK sensor	Fertilizer recommendation (fertility level, crop type, soil type, n p k)	Context aware fertilization using of IOT
[15]2022	No sensor	soil color(red clay soil, red loam soil, red laterite soiled, red yellow soil, red sandy soil, red gravel soil)	develop a method for estimating soil fertility using soil color as a parameter and deep learning techniques.
[16]2021	No sensor	Crop and Fertilizers Recommendation and Disease Prediction (N, P, K,PH, humidity, rainfall)	To develop a deep learning model for recommending crop and fertilizers, as well as predicting crop diseases.

[17]2023	Soil NPK sensor	Intelligent insecticide and fertilizer recommendation system	To develop a system for intelligent recommendation of insecticides and fertilizers in smart farming using TPF-CNN (Tri-Path Fusion Convolution Neural Network) technology.
[18]2022	No sensor	humidity, temperature, pressure, wind gust, wind speed, wind direction	To develop a machine learning technique for recommending fertilizer and irrigation to crops
[19]2022	No specific sensor	N, P, K, PH, temperature, humidity, rainfall	It develops a framework for precision agriculture that predict crop based on environmental factors
[20]2019	No sensor	Soil fertility prediction and grading	Develop ML based model for soil fertility and grading

- Previous study does not focus on the prediction of crop.
- Previous study does not focus on the rainfall effect on the soil fertility.
- Previous study just focuses on the nutrients of soil for soil fertility.
- Previous study does not focus on soil fertility accurately.
- Intelligent decision making is not used for crop recommendation.
- Previous study does not address the issue of soil moisture and its effect on soil fertility.

3. Methodology

In this section, we present the method of Image (IOT) and ML for the fertility of soil and crop recommendation. To map soil fertility, our recommended approach employs intelligent classification of environment parameters. Unlike methods that require constant sensor rotation, our approach prioritizes accuracy in fertility mapping. We carefully monitor soil fertility to provide targeted recommendations for optimal crop growth and development.

The IMAGE sensor technology enables fertility mapping with centralized data storage and decision making for different stakeholders. This method is collecting data for proposed solution involves the use of specialized sensor that can measure various environment parameters such as N, P, K, moisture, humidity, temperature, and rainfall. These sensors are carefully selected to ensure that they can accurately measure the specified parameter for collect the data. The sensor collects data based on the specified parameter, which is used to generate a detailed profile of soil fertility level in experimental area. The sensors are deployed strategically at experimental area to ensure the data collected is representative of entire area. The data is analyzed to provide accurate recommendations for suitable crop recommendation and soil fertility improvement.

Gateway is a popular solution for enabling seamless communication between disparate systems. It is composed of multiple hardware components such as microcontroller, ardunio, and WI-FI module. The microcontroller is responsible for gathering environment parameter data and transmitting it to an Image (IOT) server. The WI-FI modules are used to connect the gateway to other device such as Smartphone or tablets, which can be used to control and monitor the system remotely.

The IMAGE server is a reliable approach for data management. When the data is generated in gateway then I transmit it to IMAGE server for processing and storage. The IMAGE server is equipped with a powerful database to handle large volume of database to handle large volume of data generated from multiple sources. The database management system utilized within the IMAGE server ensures that the data is always stored efficiently and accessible. You can export data from storage in CSV format for further analysis. When the data is generated in gateway then I transmit it to IMAGE server for processing and storage.

The solution of soil fertility is based on the IMAGE based data. After the IMAGE performance the fertilization using the machine learning algorithm that is more efficient and precise solution to the problem of soil fertilization. This model uses the machine learning algorithm like support vector machine (SVM), logistic regression (LR), K-Nearest Neighbor (KNN), Naive Bayes (NB), Random Forest (RF) and Decision Tree (DT) to predict the amount and type of fertilizer to apply a given accurate soil type for crop. The algorithms are hosted on IMAGE server and use data stored in a database to make fertilization composition and application rate. The machine learning module uses the data that is stored in database that is exported in CSV format.

Machine learning used the Collaboratory tool for analysis the data to make soil fertility information for recommend the crop that is easily accessible and understandable to the farmer. It allows farmers to make informed decisions when selecting the right crop, adjusting fertilization rate, and implementing sustainable farming practice. It also offers convenience and saves time for farmers who previously had to manually gather and analyze soil fertility data.

Show that the method of proposed solution that collect the data through the sensor and is store into the database. After storing the data in database, we implement the data applying machine learning modules. The result is predicted after training and testing the dataset. The two most important parts of this method are IMAGE sensor and machine learning algorithms. The sensor that senses the real time data for fertilize mapping of the soil and machine learning algorithm for process on dataset of soil and recommend crop. Both present a unique idea. Machine learning algorithm is:

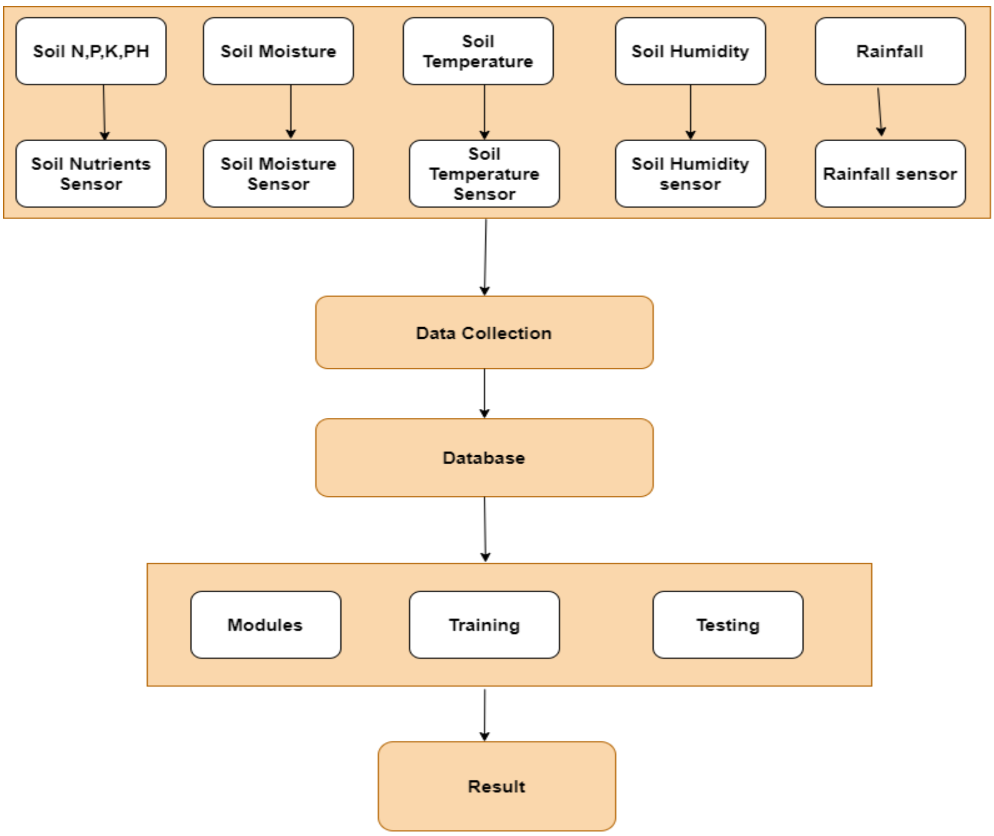


Figure 2. Method of proposed solution

3.1. Logistic Regression (LR)

The dataset has multiple crop labels, so the one of VS rest system is used to classify multiclass brackets using double classifier. The logistic regression algorithm is used to grade double markers, and the logistic regression grounded is used to grade double markers, and the logistic regression grounded double classifier is trained on each double bracket problem. Vatic nation is the class with the loftiest delicacy, but if there are fewer class markers, the model must be applied to each class, which takes time and memory.[17]

3.2. Naive Bayes

Using autonomous features and the Bayes theorem, a Naive Bayes classifier is a simple probabilistic classification technique. It works wonders with datasets that contain multiple class predictors and requires little training information to forecast labels for classes [5].

3.3. Sequential Model

Sequential models are a particular class of model architecture made up of a linear stack of layers. CNNs can be constructed in an easy and basic manner by sequentially adding each network layer.

In a sequential model, the input moves sequentially through the network, going through each layer in the prescribed order. The input data is subjected to a particular operation or transformation at each layer. Convolutional layers, pooling layers, and fully connected layers are often utilized layers in a sequential CNN model.

4. Implementation and Result

4.1. Data collection

Data is an essential component of IMAGE and machine learning systems, as it allows for the creation of models and predictions based on patterns and trends within the data. In the context of agriculture, data analysis can help farmers make informed decisions regarding planting, fertilizing, and harvesting crops We collect the data from university of agriculture Faisalabad sub campus Burewala that is used to obtain data for soil fertility and crop yields. The dataset for the crop recommendation system includes information on 22 crops commonly grown, such as rice, maize, chickpea, mango, and cotton.

4.1.1. Estimation for crop recommendation

The estimation of crop recommendation considers several parameters, such as nitrogen (N), phosphorus (P), potassium (K), and ph in the soil, as well as environmental factors such as rainfall, temperature, moisture, and humidity. By analyzing these parameters, the model can suggest which crops are best suited to growing in each area, based on the soil and environmental conditions.

4.1.2. Data collection through sensors

Sensors of N, P, K, PH, temperature, moisture, humidity, and rainfall are used to predict a suitable solution for crop management. These sensors measure the levels of nitrogen, phosphorus, and potassium in the soil, and the pH sensor measures the acidity or alkalinity of the soil. Together with temperature, humidity, moisture and rainfall sensors, these sensors provide a critical data point that allow for precise monitoring and analysis of crop growth conditions, leading to more efficient and sustainable agriculture practice.

The table. 2 shows some statistical information about the data, including the count, mean, standard deviation, minimum, 25th percentile, median (50th percentile), 75th percentile, and maximum values for each of the columns in the data. The columns are N, P, K, temperature, humidity, ph, and rainfall.

Table 2. Statically Dataset

	count	**mean**	**Std**	**Min**	**25%**	**50%**	**75%**	**max**
N	500.0	67.116	36.087	0.000	35.750	68.500	94.000	140.000
P	500.0	47.446	14.980	15.000	37.000	47.000	58.000	80.000
K	500.0	37.812	22.582	15.000	21.000	30.000	42.000	85.000
Moisture	500.0	31.552	3.742	25.730	27.862	31.585	34.380	42.250
Temperature	500.0	24.029	4.515	17.024	20.442	23.623	25.970	35.990
Humidity	500.0	58.845	24.182	14.258	47.931	65.304	80.551	84.969
Ph	500.0	6.537	0.866	4.508	5.959	6.473	7.071	8.869
Rainfall	500.0	115.222	63.357	60.652	78.325	90.966	100.782	298.560

4.2. Correlation between Different Features of Crop Data

Figure 3 shows the correlation between different features of crop data is essential for understanding the relationship between environmental factors and crop production. The table of correlation coefficients highlights the degree and direction of the correlation between different variables such as nitrogen (N), phosphorus (P), potassium (K), temperature, humidity, pH, and rainfall. The correlation coefficients of crop data show that there are both positive and negative correlations between different features of crop data, such as temperature and humidity, humidity and rainfall, nitrogen and phosphorus, and nitrogen and phosphorus. An increase in one variable. Results in a decrease in the other. Understanding the interrelationship between different variables is essential for optimizing crop production by identifying the ideal environmental conditions and making informed decisions about irrigation, fertilization, and other practices.

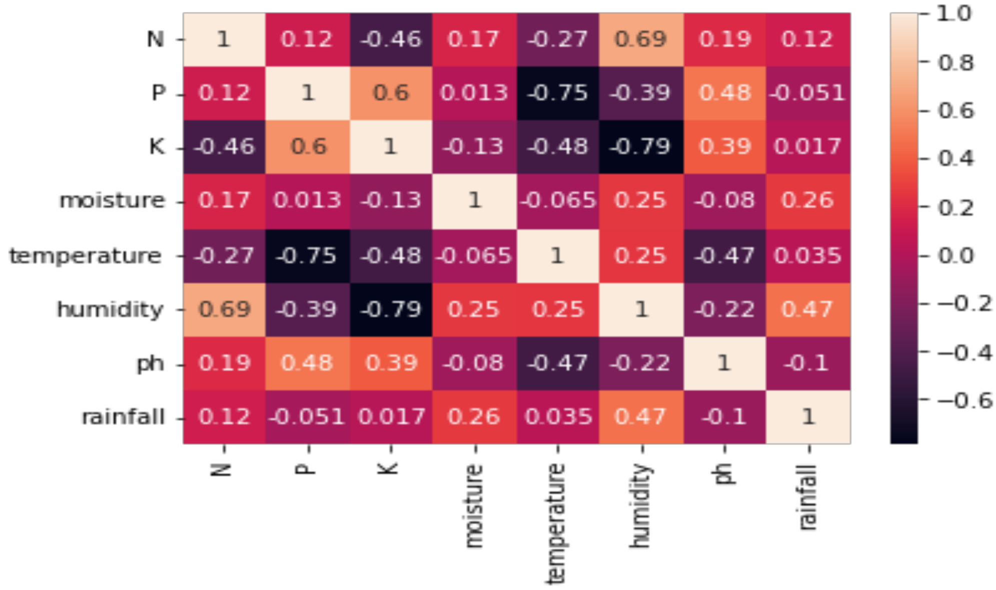


Figure 3. Correlation between different features of crop data

4.2.1. Summary of Statistics for a Data Set with 500 Data Points

The dataset consists of seven variables with 500 data points that are also shown in **table 3.**

N: The minimum value is 0, maximum value is 140, mean value is 67.116, median value is 68.5 and standard deviation is 36.087548632910284.

P: The minimum value is 15, maximum value is 80, mean value is 47.446, median value is 47.0 and standard deviation is 14.979715208757172.

K: The minimum value is 15, maximum value is 85, mean value is 37.812, median value is 30.0, and the standard deviation is 22.58212924243957.

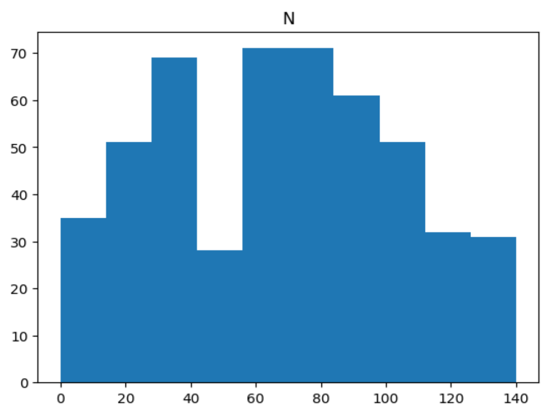
Soil moisture: The minimum value is 25.73, maximum value is 42.25, mean value is 31.5523, median value is 31.585, and standard deviation is 3.7424879079942217.

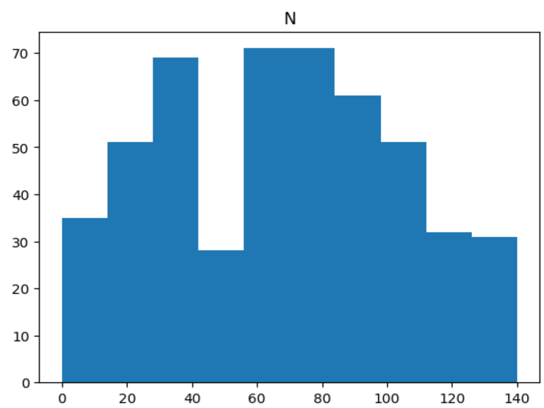
Temperature: The minimum value is 17.02498456, maximum value is 35.99009679, mean value is 24.02982218382, median value is 23.623444295, and standard deviation is 4.515330066931572.

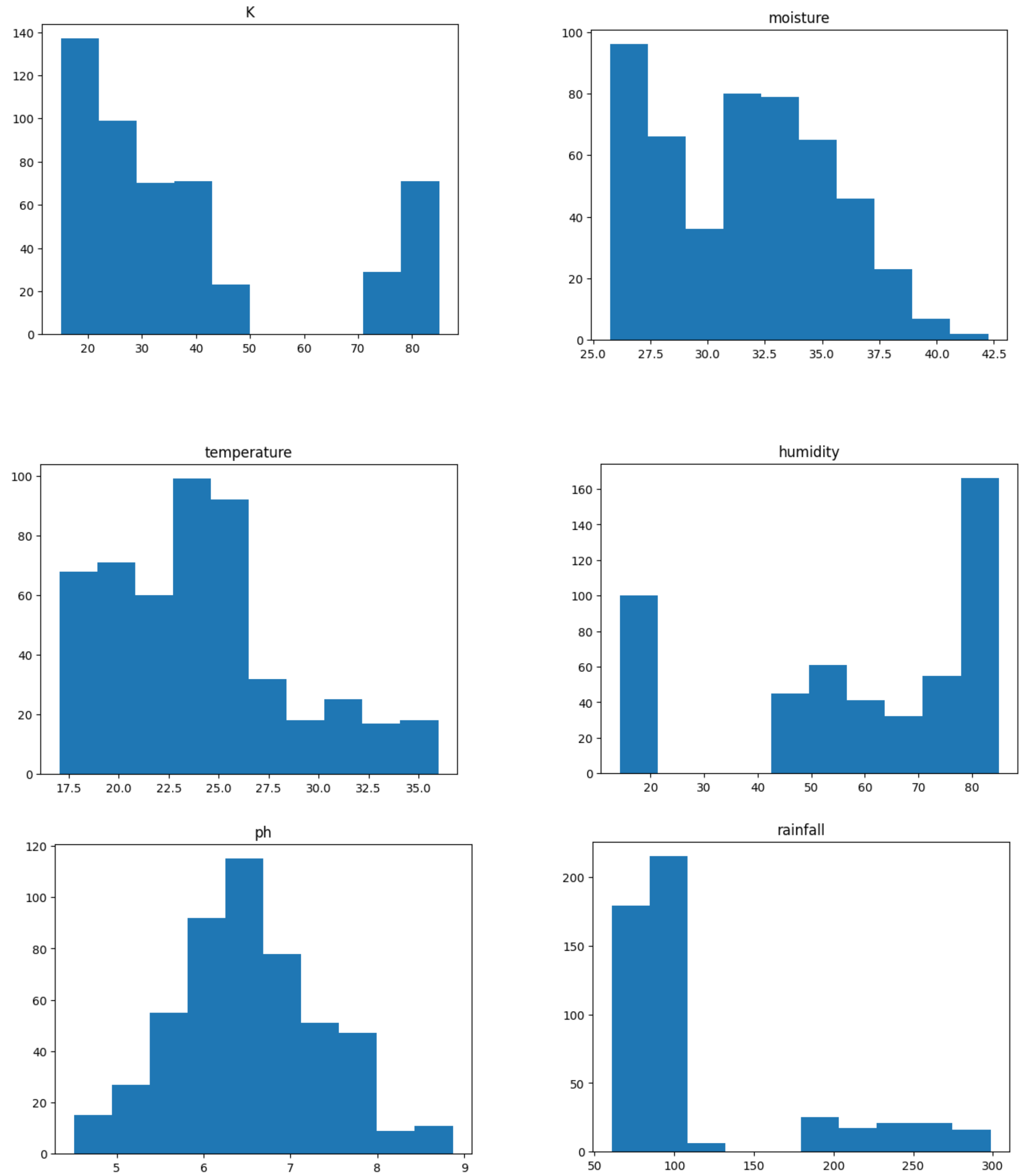
Humidity: The minimum value is 14.25803981, maximum value is 84.96907151, mean value is 58.84511147167999, median value is 65.30384518, and the standard deviation is 24.182174436631268.

pH: The minimum value is 4.507523551, maximum value is 8.868741443, mean value is 6.537333112743999, median value is 6.473399574, and standard deviation is 0.865860446879107.

Rainfall: The minimum value is 60.65171481, maximum value is 298.5601175, mean value is 115.22192733632, median value is 90.96573792000001, and standard deviation is 63.35732083287112.







4.3. Feature Selection and Data Splitting in Machine Learning

Feature selection is an important part of machine learning, that identify the most relevant variable for predict the target variable, and the crop dataset was used to illustrate how feature selection and data splitting can be done in Python using the Scikit-Learn library.

4.3.1. Features Selection

The features selected for the crop dataset are N, P, K, temperature, moisture, humidity, pH, and rainfall, which are expected to have a significant impact on predicting the label for a given crop.

4.3.2. Data Splitting

By utilizing the train_test_split method in Scikit-Learn, one can divide a dataset into an instruction set and a testing set. For sure it can be replicated we set the random state choice to 2.

4.3.3. Model Building

Once the data is split then build a machine learning model on the training dataset. The model uses Logistic Regression and Gaussian Naïve Bayes algorithms in the Scikit-Learn library, and the variables accuracy and model can be used to store the accuracy score and the trained model. On the other hand, image based data set is used by applying CNN technique on it for achieving effective accuracy.

4.3.4. Evaluation Model for Performance Measurement

The Propose a model for measuring the performance of a system. The model's performance is evaluated using three important measures: accuracy, precision, f1-score and recall. The formulae used to calculate these measures are given as follows:

- Accuracy = (TP + TN) / (TP + FP + TN + FN)
- Precision = TP / (TP + FP)
- Recall = TP / (TP + FN)
- F1 Score= 2 x (Recall x Precision) / (Recall + Precision)

TP and FP refer to cases that are predicted as positive and are positive in actual, while TN and FN refer to cases predicted as negative and are negative in actual.

4.4. Result

4.4.1. Accuracy of fertility in crops

Machine learning has revolutionized the agricultural industry by allowing farmers to accurately predict the suitability of different crops based on environmental factors. In **figure 4.** The data provides optimal environmental conditions for five different crops such as chickpea, cotton, maize, rice, and mango, which crops required most and least phosphorus.

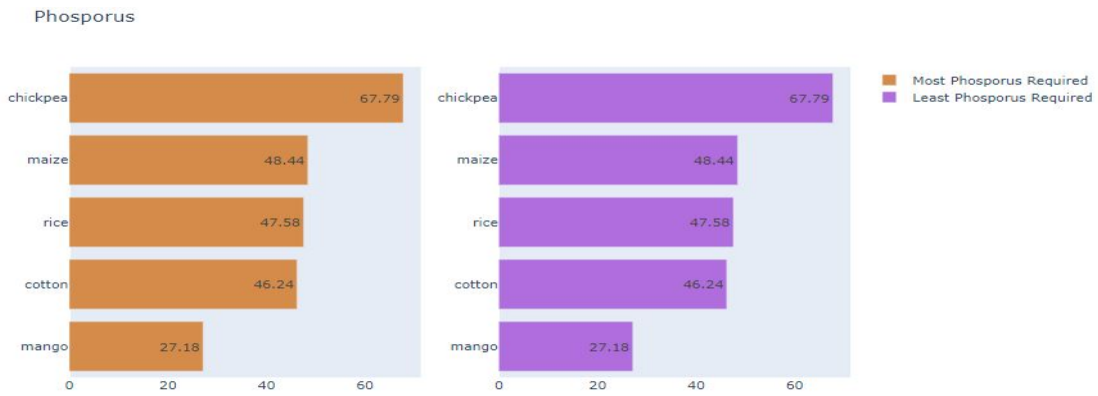


Figure 4. Phosphorus detail

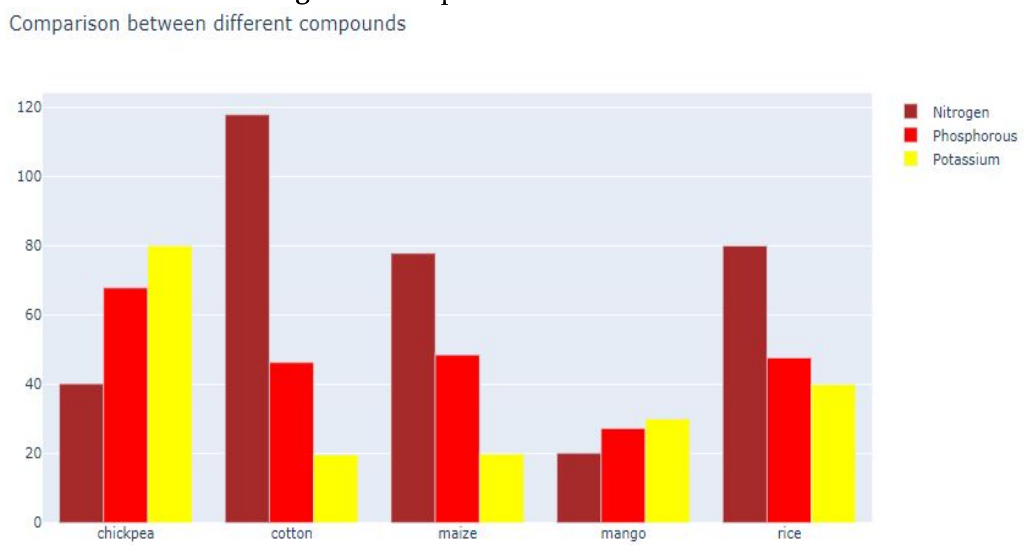


Figure 5. Comparison between different compounds

In **figure 5.** Five different crops such as chickpea, cotton, maize, rice and mango compared with each with respect to the presence of nitrogen, phosphorus and potassium. Different kinds of clustering is applied on the feature-based data set. Machine learning uses clustering to find underlying patterns and assemble related data points in feature-based datasets. By displaying the dataset's internal structure and relationships, it aids with data exploration. By locating and treating outliers or noisy data, clustering also helps with data preprocessing. In figure 6, 7 and 8 clustering is applied.

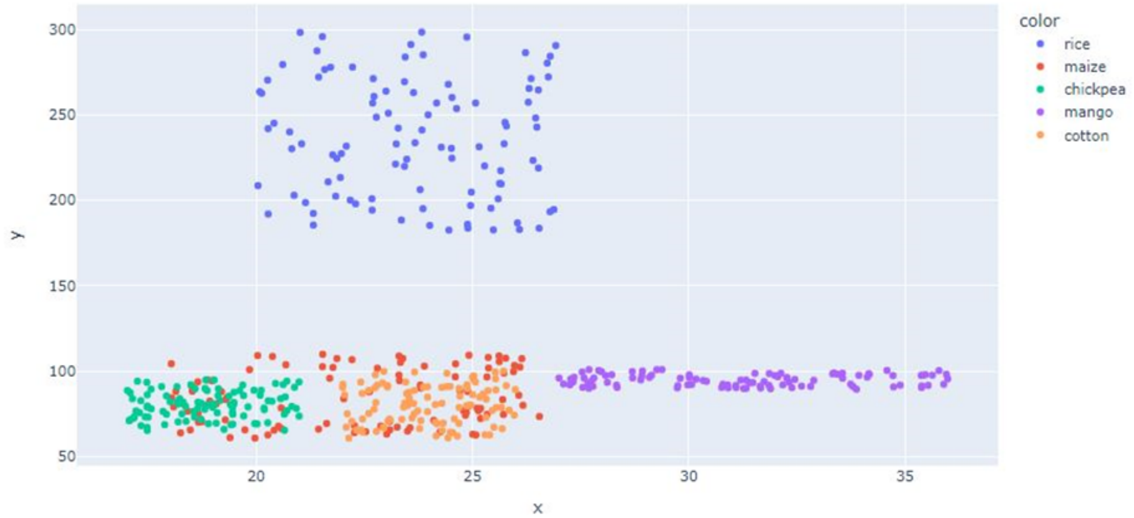


Figure 6. Cluster scatter plot of available crops.

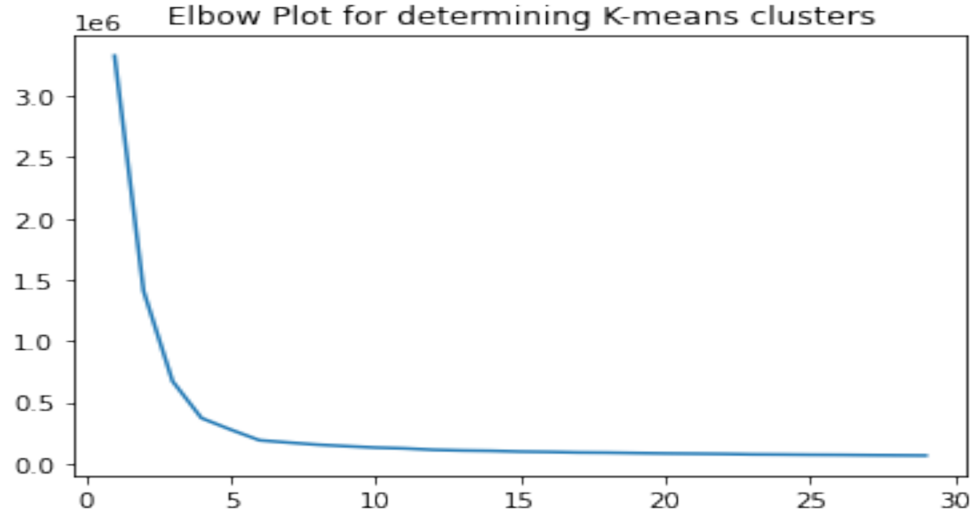


Figure 7. K-mean clustering

4.4.2. Accuracy of Machine learning modules

The purpose of developing models for ML is to make it possible to use data for decision-making and forecasting. Data quality and value, algorithm selection, and models all have a role in how precise they are. Accuracy, precision, recall, and F1 score are some of the metrics used to assess an ML model's effectiveness. The precision score for each class is calculated as the ratio of true positives and false positives. The precision score is the proportion of correctly predicted positive samples among all samples that the model classified as positive. A precision score of 1.00 indicates that all samples classified as positive are correct.

Precision = TP / (TP + FP)

A class's recall score is a percentage of correct answers compared to the total number of right and wrong responses. The percentage of actual samples that were positive split by the total number of positive data anticipated accurately is the recall score. A perfect recall score of 1.00 indicates that the model accurately identified every good data.

Recall = TP / (TP + FN)

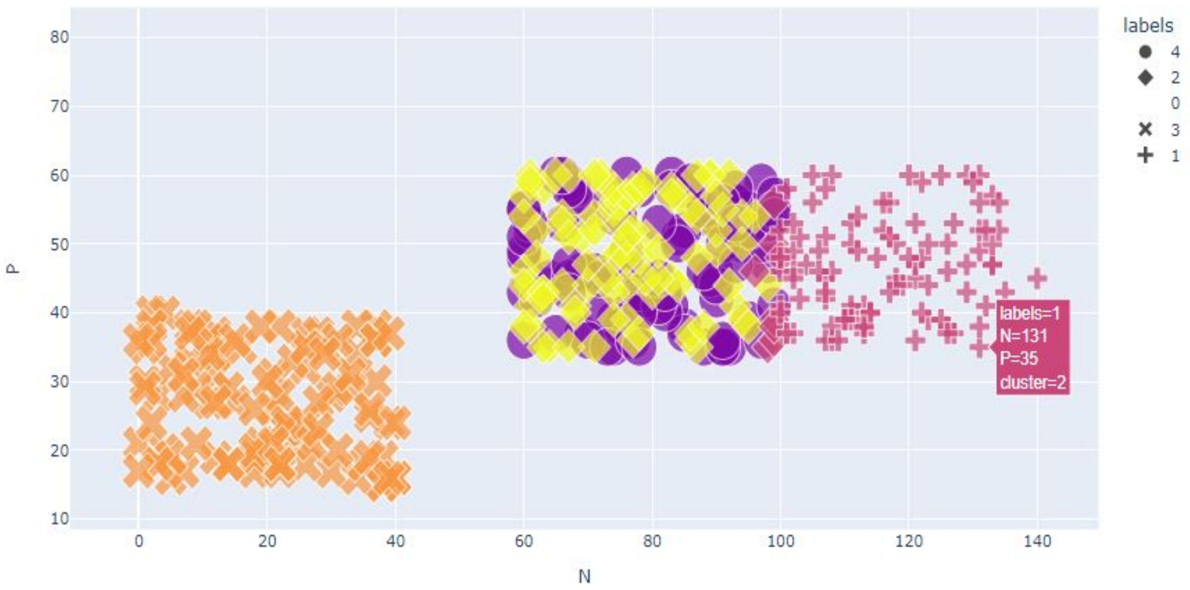


Figure 8. Cluster of all crops

The F1-score is a measure of precision and recall, with 1.00 indicating perfect accuracy.

F1-score = 2 * (precision * recall) / (precision + recall)

The support column indicates the number of samples in the testing set that belong to each class.

The macro average of precision, recall, and F1-score is calculated by taking the average of the scores for each class, giving equal weight to each class.

macro avg = (precision_chickpea + precision_cotton + precision_maize + precision_mango + precision_rice) / 5

weighted avg = (precision_chickpea * support_chickpea + precision_cotton * support_cotton + precision_maize * support_maize + precision_mango * support_mango + precision_rice * support_rice) / total_support.

We used six machine learning modules for prediction of suitable crop. The figures show the accuracy scores of different classifiers on a testing set.

4.4.3. Applied Techniques

For feature-based datasets, two methods are used Gaussian Naive Bayes (GNB) and Logistic Regression (LR). When these two algorithms are applied to the dataset, the results are over fit and do not match expectations in figure 9.

```
Model LG Confusion Matrix
[[30  0  0  0  0]
 [ 0 30  0  0  0]
 [ 0  0 28  0  0]
 [ 0  0  0 35  0]
 [ 0  0  0  0 27]]

1.0
Model GNB Confusion Matrix
[[30  0  0  0  0]
 [ 0 30  0  0  0]
 [ 0  0 28  0  0]
 [ 0  0  0 35  0]
 [ 0  0  0  0 27]]

1.0
```

Figure 9. Confusion matrix of LG and GNB

4.4.4. CNN Technique

Images from a dataset are processed using CNN algorithms. Used image-based data collection for good accuracy outcomes when ML algorithms results over fit. Nine classes are being held for both the images from the data sets and for the testing and training of classes. Its. Accuracy improves and meets expectations when a sequence-based model is used. Number classes and graphics are displayed in figure 10.

```
Found 23 images belonging to 2 classes.
Found 63 images belonging to 2 classes.
```

Figure 10. Testing and training classes

Sequential model is the model that was used to train the data set. For the training model to view, learn, and adjust itself to the given data, the actual dataset is used. In the training phase, the sequential model's parameter sets are epoch 15 as illustrated in figure 11.

```
Epoch 15/15
2/2 [==============================] - 8s 4s/step - loss: 0.2277 - accuracy: 0.9206 - val_loss: 0.2405 - val_accuracy: 0.9130
```

Figure 11. Accuracy result of training model

4.4.5. Confusion Matrix

After applying a binary matrix during the training phase, confusion matrix is applied to the images to verify our results for classification challenges. It provides projected and actual value and is used to evaluate performance. Confusion matrix is displayed in picture 12 below.

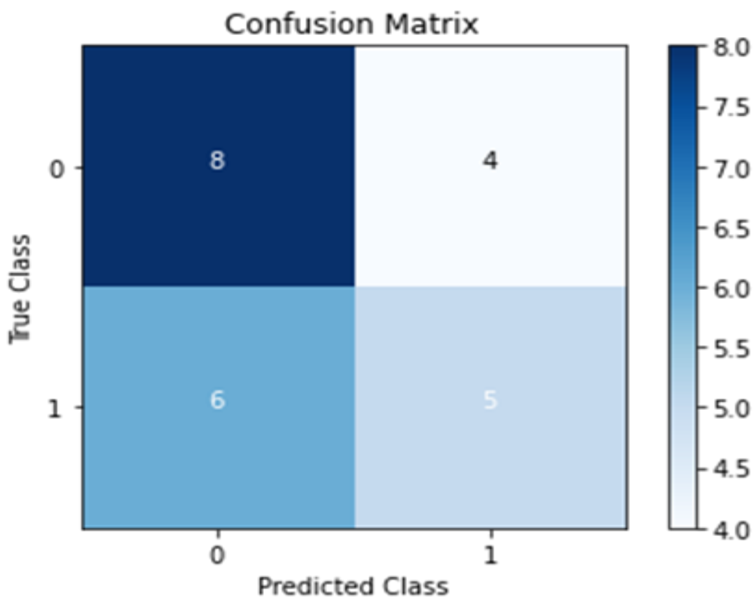


Figure 12. Confusion Matrix

4.4.6. Measuring Parameters

The measuring parameter is assessed using the confusion metrics mentioned above. Accuracy, Precision, Recall, and F1 Score are the parameters that were calculated and are listed in the tables below.

Table 3. Measuring Parameters

F1 Score	Precision	Recall	Accuracy
30%	30%	29%	91.30%

5. Conclusion

The system of fertilizer and recommendation is based on the machine learning according to the real time data of soil parameters like N, P, K, PH, temperature, humidity, moisture, and rainfall that is capture through the IMAGE system for soil fertility mapping. Various machine learning algorithms are used for this purpose. Algorithms have different accuracy but the Logistic regression and Gaussian naive bayes accuracy over fits. So, for that reason image-based dataset is used and sequential model CNN technique is applied and get the accuracy of 91%. Sequential algorithm is recommended for rice crops on the base of soil fertility data. Adding a fresh dataset that facilitates the use of machine learning could prove to be an excellent over-time addition.

References

1. "Machine Learning Prediction Analysis using Image for Smart Farming," Int. J. Emerg. Trends Eng. Res., vol. 8, no. 9, pp. 6482–6487, Sep. 2020, doi: 10.30534/ijeter/2020/250892020.
2. P. Priya, U. Muthaiah, and & M. Balamurugan, "IJESRT INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY PREDICTING YIELD OF THE CROP USING MACHINE LEARNING ALGORITHM," Int. J. Eng. Sci. Res. Technol., no. 1, doi: 10.5281/zenodo.1212821.
3. M. Choudhary, R. Sartandel, A. Arun, and L. ladge, "Crop Recommendation System and Plant Disease Classification using Machine Learning for Precision Agriculture," in Artificial Intelligence and Communication Technologies, Soft Computing Research Society, 2022, pp. 39–49. doi: 10.52458/978-81-955020-5-9-4.
4. Watson C.A.*, D. Atkinson, P. Gosling, L. R. Jackson, and F. W. Rayns, "Managing soil fertility in organic farming systems," Soil Use Manag., vol. 18, no. 3, pp. 239–247, 2002, doi: 10.1079/sum2002131.
5. V. Pandith, H. Kour, S. Singh, J. Manhas, and V. Sharma, "Performance Evaluation of Machine Learning Techniques for Mustard Crop Yield Prediction from Soil Analysis," J. Sci. Res., vol. 64, no. 02, pp. 394–398, 2020, doi: 10.37398/jsr.2020.640254.
6. E. A. Stockdale et al., "Agronomic and environmental implications of organic farming systems," Adv. Agron., vol. 70, no. February 2021, 2001, doi: 10.1016/s0065-2113(01)70007-7.
7. Ullah, Saif, Nadeem Iqbal, Ali Haider Khan, Muhammad Sajid, Zohaib Ahmad, Haroon Ahmad, and Muzammil Hussain. "EMPOWERING AGRICULTURE: A GREEN REVOLUTION WITH INTERNET OF ENERGY-DRIVEN FARM ENERGY MANAGEMENT FOR SUSTAINABLE AND ECO-FRIENDLY PRACTICES." Journal of Population Therapeutics and Clinical Pharmacology 30, no. 19 (2023): 975-992.
8. H. Kolbe, "Comparative Analysis of Soil Fertility, Productivity and Sustainability of Organic Farming in Central Europe—Part 1: Effect of Medium Manifestations on Conversion, Fertilizer Types and Cropping Systems," Agronomy, vol. 12, no. 9, 2022, doi: 10.3390/agronomy12092001.
9. S. Qazi, B. A. Khawaja, and Q. U. Farooq, "IoT-Equipped and AI-Enabled Next Generation Smart Agriculture: A Critical Review, Current Challenges and Future Trends," IEEE Access, vol. 10. Institute of Electrical and Electronics Engineers Inc., pp. 21219–21235, 2022. doi: 10.1109/ACCESS.2022.3152544.
10. A. A. Khan, M. Faheem, R. N. Bashir, C. Wechtaisong, and M. Z. Abbas, "Internet of Things (IoT) Assisted Context Aware Fertilizer Recommendation," IEEE Access, vol. 10, pp. 129505–129519, 2022, doi: 10.1109/ACCESS.2022.3228160.
11. S. J. Suji Prasad et al., "An efficient LoRa-based smart agriculture management and monitoring system using wireless sensor networks," Int. J. Ambient Energy, vol. 43, no. 1, pp. 5447–5450, 2022, doi: 10.1080/01430750.2021.1953591.
12. P. Lashitha Vishnu Priya, N. Sai Harshith, and N. V. K. Ramesh, "Smart agriculture monitoring system using IoT," Int. J. Eng. Technol., vol. 7, no. July, pp. 308–311, 2018, doi: 10.35940/ijeat.d7916.049420.
13. D. A. Bondre and M. Santosh Mahagaonkar, "PREDICTION OF CROP YIELD AND FERTILIZER RECOMMENDATION USING MACHINE LEARNING ALGORITHMS," 2019. [Online]. Available: http://www.ijeast.com
14. T. Blesslin Sheeba et al., "Machine Learning Algorithm for Soil Analysis and Classification of Micronutrients in IoT-Enabled Automated Farms," J. Nanomater., vol. 2022, 2022, doi: 10.1155/2022/5343965.
15. N. L. Kalyani and K. B. Prakash, "Soil Color as a Measurement for Estimation of Fertility using Deep Learning Techniques," Int. J. Adv. Comput. Sci. Appl., vol. 13, no. 5, pp. 305–310, 2022, doi: 10.14569/IJACSA.2022.0130536.
16. D. Rangsang, "Crop and Fertilizers Recommendation and Disease Prediction using Deep Learning," no. 14, pp. 846–854, 2021.
17. T. Thorat, B. K. Patle, and S. K. Kashyap, "Intelligent insecticide and fertilizer recommendation system based on TPF-CNN for smart farming," Smart Agric. Technol., vol. 3, no. June 2022, p. 100114, 2023, doi: 10.1016/j.atech.2022.100114.
18. S. Jeedigunta, "Using Machine Learning Technique for Recommendation of Fertilizer and Irrigation in Crops," vol. 5, no. Iv, pp. 15–19, 2022.
19. A. Gutiérrez et al., "Image Framework for Measurement and Precision Agriculture: Predicting the Crop Using Machine Learning Algorithms," 2022, doi: 10.3390/technologies.
20. T. G. Keerthan Kumar, C. Shubha, and S. A. Sushma, "Random forest algorithm for soil fertility prediction and grading using machine learning," Int. J. Innov. Technol. Explor. Eng., vol. 9, no. 1, pp. 1301–1304, 2019, doi: 10.35940/ijitee.L3609.119119.
21. Kajla, N. I., Missen, M. M. S., Luqman, M. M., & Coustaty, M. (2021). Graph neural networks using local descriptions in attributed graphs: an application to symbol recognition and hand written character recognition. IEEE Access, 9, 99103-99111.
22. Ali Awan, M.D., Ali, S., Samad, A., Iqbal, N., Saad Missen, M.M. and Ullah, N., 2021. Sentence Classification Using N-Grams in Urdu Language Text. Scientific Programming, 2021, pp.1-11.