ML Based Agriculture Crop Recommendation System by Monitoring Soil Condition

Dr.S. Gnanapriya[1], S. Chandra Bala Murugan[2]

[1]Assistant Professor, Department of Computer Applications,

Nehru college of management, Coimbatore, Tamilnadu, India

[2]Student of II MCA, Department of Computer Applications,

Nehru college of management, Coimbatore, Tamilnadu, India

Abstract— Agricultural yield generally depends on the level of soil fertility. Nitrogen (N), Phosphorus (P), Potassium (K), pH, the temperature of the soil, and moisture as soil chemical constituents are fundamental parameters for determining soil fertility. Good yield can easily be ensured by measuring their presence and applying the right amount of fertilizer in the right season. Most farmers do not produce good crops due to insufficient knowledge and the inability to use the proper amount of fertilizers. Current methods of measuring soil nutrients involve collecting soil from the field and transporting it to a laboratory for testing, which is often subjective and very expensive. This paper suggests an efficientIoT-based soil nutrient monitoring and machine learning-basedcrop recommendation system that helps farmers by offering crop- related details and recommendations for crops based on different soil and weather attributes. The proposed system deploys various types of sensors to determine soil nutrients, these sensors continuously collect the required data from the farm field and transmit it via a wireless sensor network (WSN) to a cloud database. By monitoring (N, P, K, temperature, pH, humidity, rainfall) values and analyzing the permanent and temporary behavior of the soil, the machine learning approach will recommend what types of crops have the best production potential for this land. Agriculture's use of machine-learning technology makes it easier to select the best-yielding crops by reducing the cost of unnecessary fertilizer use, which reduces manual labor in crop and crop management and increases productivity. The most appropriate crops for that cropland are suggested using machine learning algorithms in IoT-based soil nutrient monitoring, which stores data from various soil nutrients in a database. As a result, agricultural production will contribute more to national growth.

Keywords— Internet of Things(IoT), Machine Learning(ML), Wireless Sensor Networks (WSN), Potential of Hydrogen(pH), Crop Recommendation

I. INTRODUCTION

One of the reasons why many farmers, despite devoting themselves to agricultural work, are not able to improve their economic condition is that they cannot select the right crops for their cropland. They usually face financialloss due to the inability to select the right crop [1]. Only by determining the nutritional quality of the crop soil can the necessary measures be taken for profitable production. Significantly, soil fertility can be increased by applying the chemical nutrients of the soil from the outside, but the biological properties cannot be changed much. Low-nutrient soils lead to multiple plant disorders and low yields. Farmerscan be more profitable if they know which crop will grow well in which soil. Therefore, it is very important to know the information related to the nutrient quality of the soil of the farmer. The problem of choosing the right crop is more in the rural areas and for solving it is very necessary to use the techniques of the IoT approach which works amazingly [2]. In many cases, farmers without realizing it, apply excessive amounts of unwanted fertilizers, which leads to high costs and on the contrary, crop damage and production losses. Various types of sensors (such as NPK, pH,temperature, Moisture, and Humidity sensors) are employed to determine these nutrients [3], these sensors continuously collect the necessary information from the crop field via IoT devices (Arduino, Raspberry pi) and sent to the cloud platform through wireless communication (Bluetooth/ GSM Wi-Fi, etc.) module using the Wireless network [4]. What types of popular technologies are used in wireless sensor networks are shown in Table 1. The biggest advantage of data stored in a cloud database is that the data can be accessed from anywhere at any time. So there is no restriction on data access, which can be easily accessed through any smart device like a smartphone, computer, etc. Selecting a crop that does not give a high yield for a particular soil, Low yield is always the result of planting during the incorrect season and choosing the incorrect crop. Farmers often make wrong decisions regarding selectingsuch crops [5]. A number of factors that are very important to consider in making this correct decision include soil moisture, temperature, rainfall, and pH value and concentration of nutrients such as nitrogen (N), phosphorus (P), and potassium (K) alignment. All this information is considered through machine learning to recommend suitable crops. Linear Regression (LR), K Nearest Neighbor (K-NN),Decision Tree (DT), Random Forest Regression (RFR), Neural Network (NN), Support Vector Machine (SVM), and XGBoost machine learning algorithms are among the most used algorithms for machine learning [6][7]. The aim of this system is to reduce workload, increase profits and make sound decisions. Real-time Nitrogen(N), Phosphorus(P)Potassium(K), pH, Temperature, Moisture, and Humidity values from the cloud will be used as inputs inside the machine learning system to automatically recommend whichtype of crop would be best for this field as output. In this case, based on the current value from the previous learning points, it recommends the best suitable crop so that the farmer gets more profit. Through this, the farmer can select the right crop and predict the yield. Section II focuses on the literature review and related works. The suggested Methodology is presented in Section III. Section IV presents a comparative analysis of the research works under consideration, and Section V presents the conclusionTechnology

/ZigBee/Radio/	DiscoveryYear	Standard	Transmission Range	Data Rate	OperatingFrequency	Cost
RFID	1987	ISO 18000-6C	1-5m	40-60Kb/s	13.56MHz	Low
Wi-Fi	1997	IEEE 802.11 a/c/b/d/g/n	20-100m	1Mb/s-7Gb/s	2.4/3.6/5/5.9/60GHz	High
GSM	1997	Many Standard	35Km	9.6Kb/s	800-900MHz	Medium
Zigbee	1998	IEEE 802.15.4	10-100m	250Kb/s	2400-2483.5MHz	Low
Bluetooth	1999	IEEE 802.15.1	10-100m	1-24Mb/s	2400-2483.5MHz	Low
WiMAX	2001	IEEE 802.16	50-90 Km	1Mb/s-1Gb/s	2-66GHz	High
6LoWPAN	2006	Open Standard	30m	250Kb/s	915MHz	Low
SigFox	2009	SigFox	30-50Km	100-600bit/s	200KHz	Low

Table 1: Wireless Technology commonly used with IoT

II. LITERATURE REVIEW AND RELATED WORKS

Agriculture's use of machine learning makes this sector incredibly efficient and easy. The three main steps in the machine learning process are data preprocessing, model building, and generalization. Cases, where human skills are insufficient, are solved through machine learning algorithms [8]. ML stands for machine learning, a field of study that involves using past experience and data to help computers make decisions. Numerous applications of machine learning (ML) exist to "feature extract" crucial features from the data and information [9]. Supervised learning and unsupervised are the two main subcategories of machine learningmethods, semi-supervised and reinforced are 2 (two) extra categories into which some researchers further subdividemachine learning algorithms. Pest control technique can be driven by ML [10]. A supervised machine model generates an output based on predetermined evidence and trainingdata, whereas a machine learning technique predicts the outcome using a defined set of input variables and labels [11]. In the paper [12], they discussed the useful classification of Machine-learning in laboratory medicine and healthcare. In [13] they established a crop production model where they collected 20 years' worth of agricultural data for their forecast. Following that, stacking analysis was employed to improve the regression model [14]. XGBoost, KNN classifier, Random Forest, and Logistic classifier were four machine learning algorithms that were tested in order topredict crop yield using the parameters rainfall and temperature. Finally, it was determined that, when comparedto other algorithms, the random forest classifier has the best accuracy [15]. The algorithm was then hybridized for more accurate yield forecasting, taking into account factors like area, rainfall, and soil type. The system was then able to determine On the basis of the aforementioned qualities, determine that what crop is appropriate for cultivation [16].

III. PROPOSED METHOD

A large part of the problems of monitoring the soil and its constituents in cropland can be solved using Internet of Things technology and crop recommendations can be made using machine learning technology to determine which type of crop will grow well on this land. This system model is divided into two modules to perform the complete task. Through Soil Fertilizer Monitoring Module, the real data of soil properties parameters Phosphorous (P), Nitrogen (N), Potassium (K), Humidity, Potential of hydrogen (pH), and temperature are collected at run time and sent to the cloud. The crop Recommendation Module implements a machine learning-based model for crop recommendation where the dataset from the Soil Fertilizer Monitoring Module will be used to decide what type of crop will grow well on this land and recommend crops accordingly as shown in Fig. 1.

A. Soil Fertilizer Monitoring Module

If the nutrient quality of the soil of cropland can be determined, it is possible to ensure profitable production by applying proper fertilizers accordingly. The amount of additional fertilizer to be applied to cropland depends on the current soil quality. Infertile soils lead to multiple plant disorders and low yields. All Soil sensor send their data to the data collector device. Camera sensors are used to provide images of crops. These images are used throughimage processing to diagnose crop diseases and determine whether pesticides are needed. The collector device (Arduino, Raspberry pi) will process the data received from the sensors and send it to the receiver End(ESP32) utilizing the transceiver and a wireless communication module, such as Bluetooth, Wi-Fi, ZigBee, Radio or GSM. The receiver end (ESP32) receives the data through the transceiver and sends it to the cloud database system using the wirelesscommunication (Wi-Fi) method. The farmer can take necessary measures by analyzing the necessary information received. It is advised to use batteries in addition to a direct power source to power all electrical components included in this proposed system. It is also advised to use renewable solar energy, as shown in Fig. 2 where the IoT-based soil monitoring system is covered. This system will have a platform where some stakeholders will be interfaced with the system according to their responsibilities. The proposed platform can be a web-based or mobile application to provide access to the system to the registered stakeholders of the system as per their needs.

B. Crop Recommendation Module

A crop recommendation module based on machine learning was implemented. Because supervised machine-learning provides better accuracy than unsupervised learning in recommendation problems, four different well-known supervised machine learning algorithms are used in the model's deployment The crop recommendation model is designed using Linear Regression(LR), Random Forest Regression (RFR), Decision Tree, and XGBoost machine learning algorithms. N, P, K, Humidity, pH, Temperature of soil, and Average Rainfall in millimeters these seven (7) parameters have been taken into consideration for machine learning in implementing the crop recommendation model.

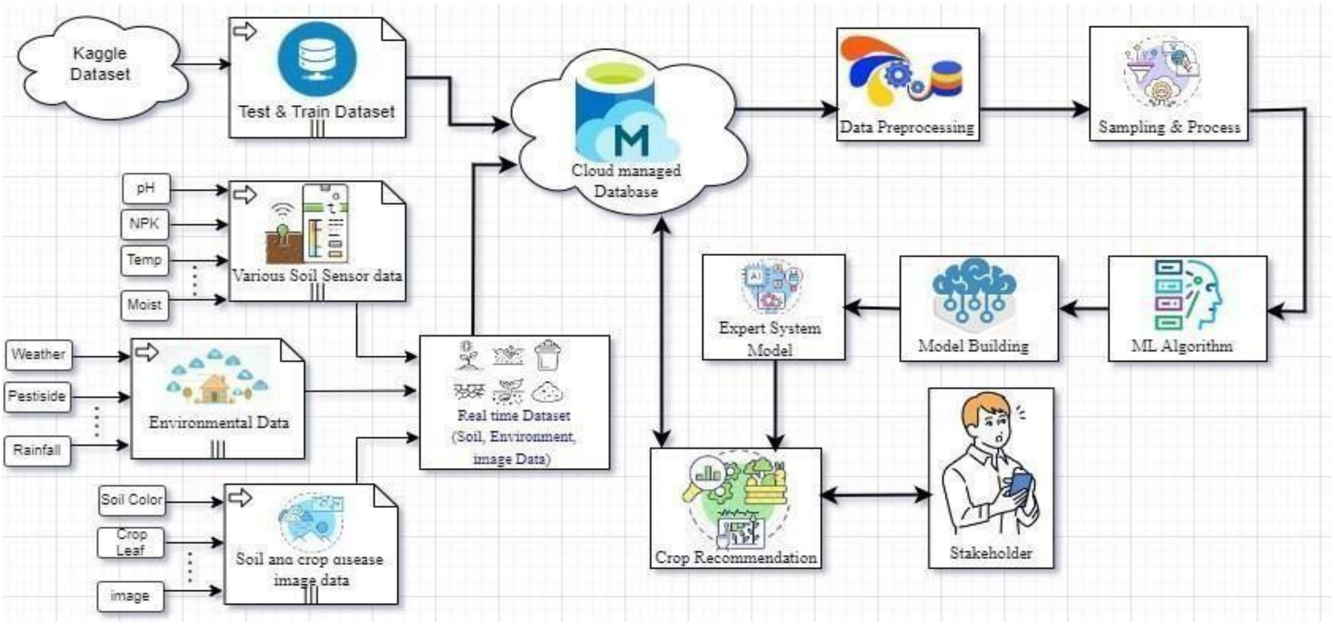


Fig. 1. Soil Fertilizer Monitoring and Crop Recommendation method

Decision Tree, and XGBoost machine learning algorithms. N, P, K, Humidity, pH, Temperature of soil, and Average Rainfall in millimeters these seven (7) parameters have been taken into consideration for machine learning in implementing the crop recommendation model. The system model will suggest what type of crop will grow best if these parameters are in which amount.

1) Importing Libraries and dataset

Data preprocessing tools, Machine Learning Algorithms as well required Libraries need to be imported. The system is made productive and efficient to predict the best crop by these tools and libraries. Popular libraries were imported, including label encoder, train test split, pickle, pandas, seaborn, matplotlib, and numpy. The system also includes models like XGBoost, Decision Trees, Linear Regression, and Random Forest Regression. To train the system, the dataset collected from Kaggle is combined with some manual datasets collected from real fields to form a new reliable dataset according to BARC guideline in Bangladesh.Most of the data in this dataset are used for machine learning and the remaining data is used for testing the system model. The dataset comprises 2,420 rows and eight columns (8 attributes). The last column enlisted the crop list.7 columns of attributes describe below:

S.L	Attribute	Description
1	N	Nitrogen(s)
2	P	Phosphorous(s)
3	K	Potassium(s)
4	H	Humidity (%)
5	T	Temperature in Celsius
6	PH	Soil pH
7	RF	Rainfall in mm
8	Crop	Crop Grown

Table 2: Details meaning of the Dataset

2) *Data Analysis and Visualization*

To obtain the best model, data analysis and visualization should be performed prior. It is very important to check for missing values per attribute as well as distinguish unique values of dependent and independent variables. Visualization is easy after getting the analyzed dataset.

N	P	K	H	T	PH	RF	Crop
79	51	16	68.4	25.3	6.5	96.4	Maize
90	42	43	82.0	20.8	6.5	202.9	Rice
43	79	79	18.9	19.4	7.8	80.2	Chickpea
17	77	23	20.8	24.5	5.6	64.1	KidneyBeans
28	59	22	52.7	30.9	7.0	170.9	Pigeon peas
99	15	27	56.6	27.4	6.0	127.9	Coffee

Table 3: Sample Instances of Dataset

3) *Label encoding*

The label values for the dependent label variables must be converted into numeric values before being nourished into the predictive model for future prediction because they have categorical and non-numeric values. Because processing non-numeric values are quite complex and time-consuming, label encoding plays a major role.

4) *Separating the dataset into train and test sets*

Data sets are separated into training and testing sets formodel building after data analysis and visualization. Initial data is required to train the machine. Machine learning algorithms are taught how to make precise predictions or carry out the desired action using training datasets. 60% of the data from this data set was used for machine learning and the remaining 40% of the data was used for testing the system model shown in Fig. 4.

5) *Model building*

The best crop recommendation model is found using four machine-learning algorithms in the proposed work. Steps to build a model should be performed

- To build the model, the library has to be imported.
- Being formed is a model.
- The model is fitted with both training and testing data, and after training, the testing data set is used to test the model.
- Metrics for evaluation and confusion are computed.

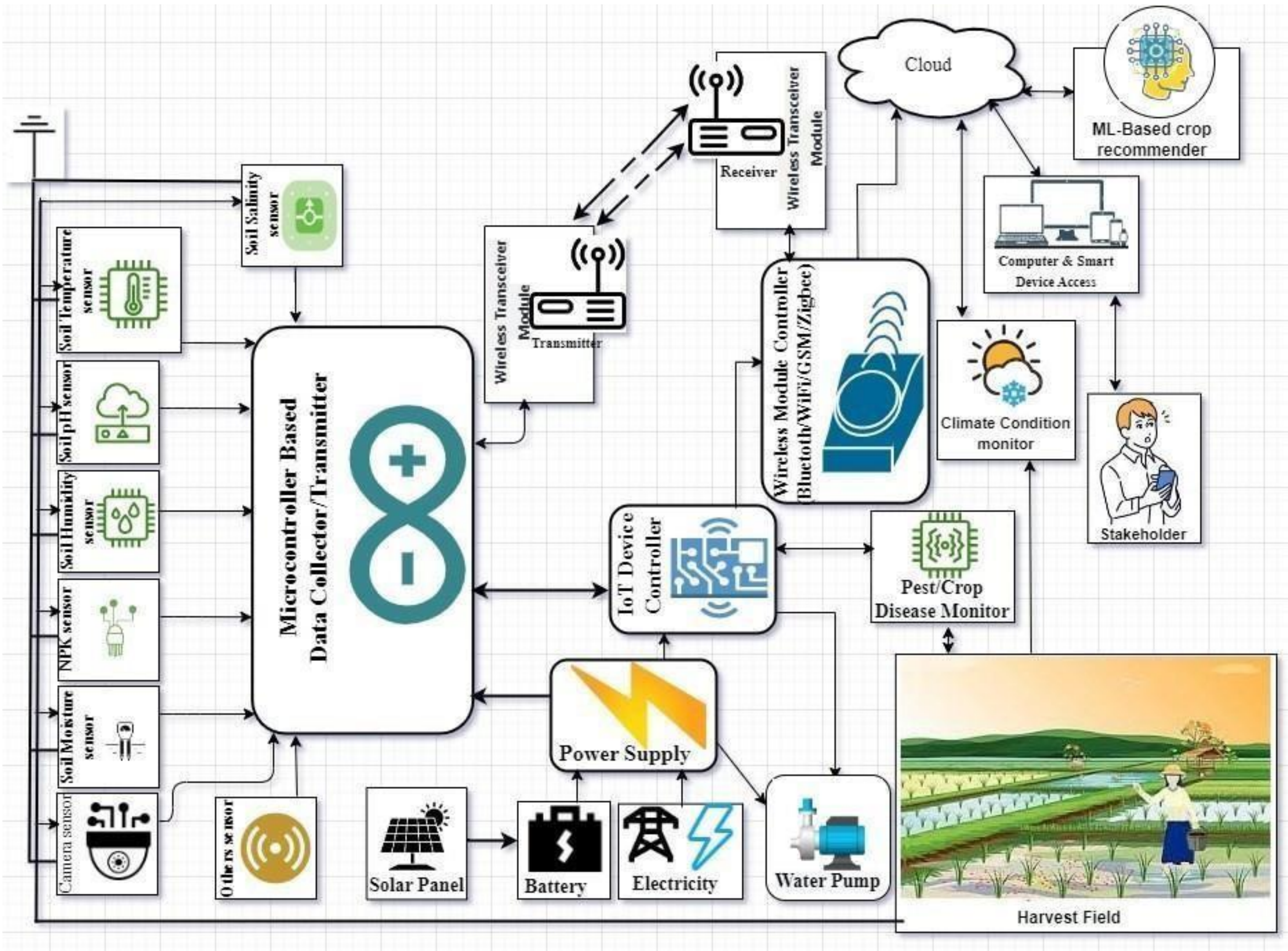


Fig. 2. Overall IoT base Soil monitoring System

Among these four machine-learning algorithms Linear Regression (LR), Random Forest Regression (RFR), Decision Tree (DT) and XGBoost, the one with the best performance are used for model building. Random Forest Regression and XGBoost algorithm performs better than the two. Model building is done using the Random Forest Regression algorithm.

6) Predicting the crop

Choosing the best method for crop assumption based oncomparison and performance assessment. The model needs inputs for N, P, K, temperature, moisture, pH, and rainfall in order to predict the best crops. According to the parameter values, the best-predicted crop name would be recommended.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

The IoT-based smart soil fertilizer monitoring module was tested in garden tubs as well as agricultural crop fields and found satisfactory results in the Indian subcontinent specially in Bangladesh. Over time various cycles of testing, multiple factors were considered and improvements weremade according to satisfactory results. Fig.4 shows the clouddatabase table content collected from the IoT-based smart soil fertilizer monitoring module. The proposed supervised machine learning-based crop recommendation models are implemented using the four algorithms mentioned above. The model was implemented by experimenting with 2420 samples from the field and 22 different crops. The algorithmic model will suggest crops for the targeted land based on the match between the targeted land and the required crop property. The dataset used to train the model consists of 7 soil properties. Dataset has been collected from IoT based fertilizer monitoring module using Mysql database shown on Fig.3.

+ Options

S.L	Nitrogen	Phosphorous	Potassium	Humidity	Temperature in Celsius	PH	Timestamp
1	28	56	22	52	30	7	2022-11-04 13:16:16.519138
2	23	48	23	58	33	7.5	2022-11-04 13:36:16.519168
3	34	67	43	58	43	6.5	2022-11-04 14:36:16.519158
4	54	57	65	63	54	6.5	2022-11-04 15:36:16.519128
5	43	79	80	18.9	19.4	7.8	2022-11-05 23:19:54.411323
6	45	34	76	34	24	6.4	2022-11-05 23:32:16.519188
7	45	67	54	47	28	6.7	2022-11-05 23:33:09.555397
8	28	34	54	45	24	6.4	2022-11-05 23:34:16.240124
9	32	49	59	43	29	5.9	2022-11-05 23:35:06.005177
10	54	67	59	47	31	6.2	2022-11-05 23:35:54.172716
11	43	79	59	34	29	6.4	2022-11-05 23:36:27.168105

Fig. 3. Tabular data from the fertilizer monitoring module

Since every parameter in the dataset contributes to the assignment of prediction, the similarity graph that was obtained indicates that there is no particular attribute that might be removed. Four supervised machine-learning Algorithms are used to implement machine learning-based crop recommendation, and their performance has been compared in order to reap the benefits. Therefore, the Random Forest Regression-based machine learning model works better than other supervised machine learning models. Also XGBoost-based machine learning model gained satisfaction shown in Fig.5. proposed system module given better accuracy output than crop yield prediction using machine learning algorithms [15].

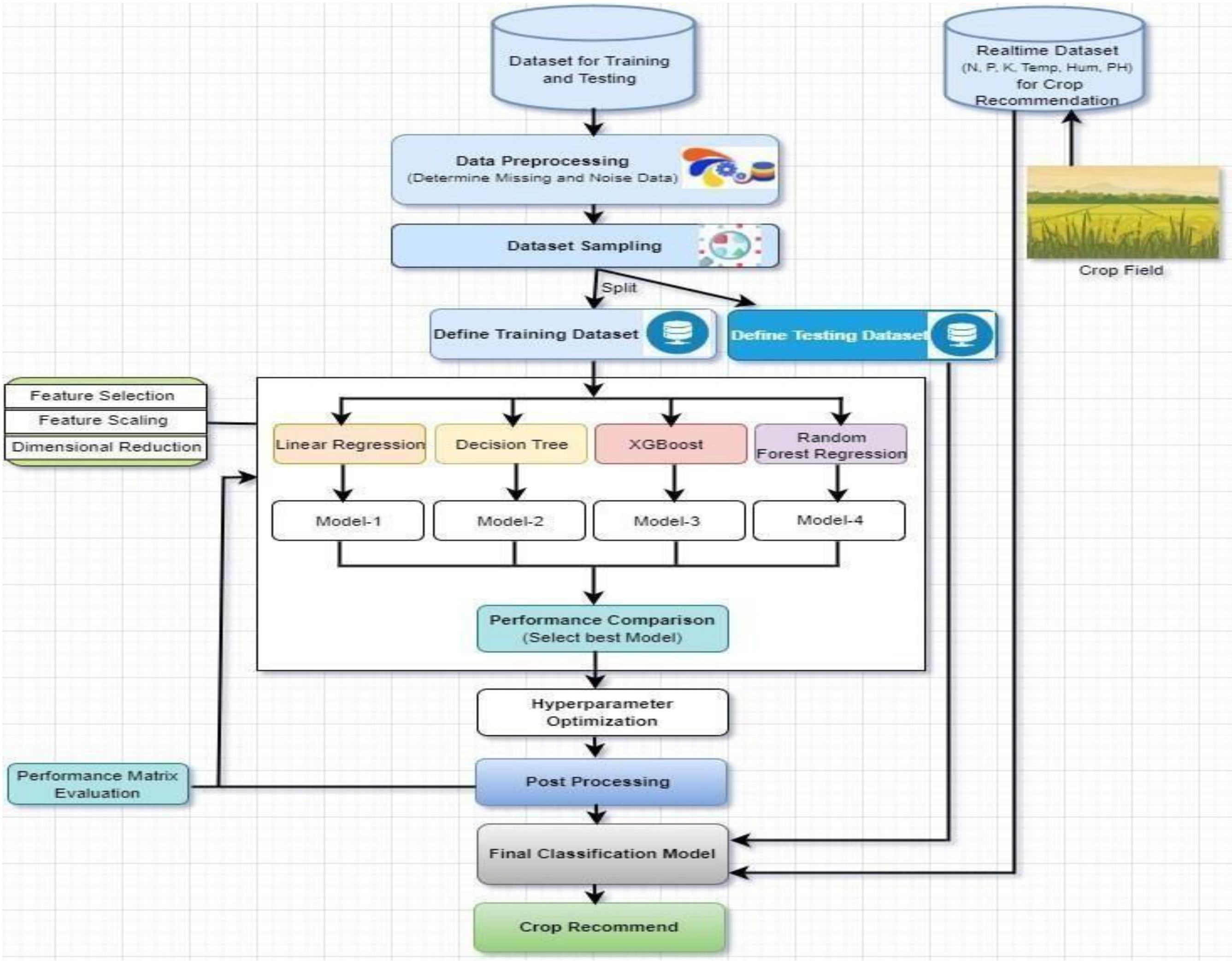


Fig. 4. Sequence Process of Crop Recommendation

The Train accuracy of Linear Regression, Random Forest Regression, Decision tree, and XGBoost are 72.03%, 100%, 94.48%, and 100% respectively. Similarly, The Test accuracy of Linear Regression, Random Forest Regression, Decision tree, and XGBoost are 70.81%, 99.09%, 93.48%, and 98.63% respectively Table 4. According to soil parameter value Output is shown in Fig.6. Fig.7 shows the fertilizer monitoring of the garden tub through the working method of the soil fertilizer monitoring module.

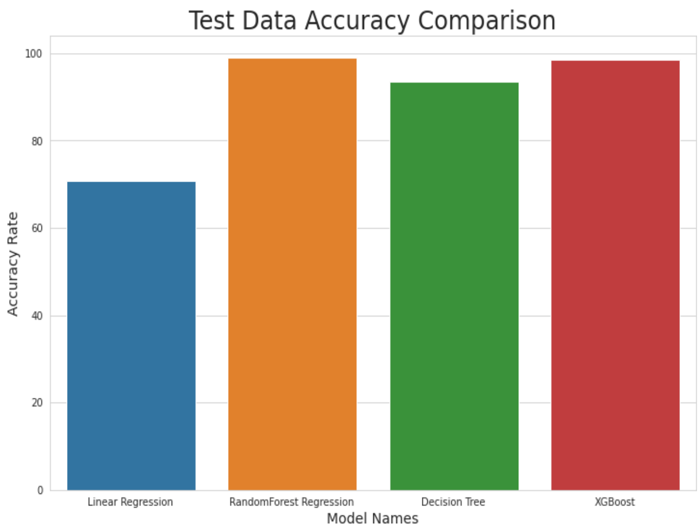


Fig. 5. Model accuracy comparison

```
N = 90
P = 42
K = 43
temperature = 20.82312
humidity = 82.00284
ph = 6.50232
rainfall = 202.93536
sample = [N, P, K, temperature, humidity, ph, rainfall]
single_sample = np.array(sample).reshape(1,-1)
pred = model.predict(single_sample)
pred.item().title()
```

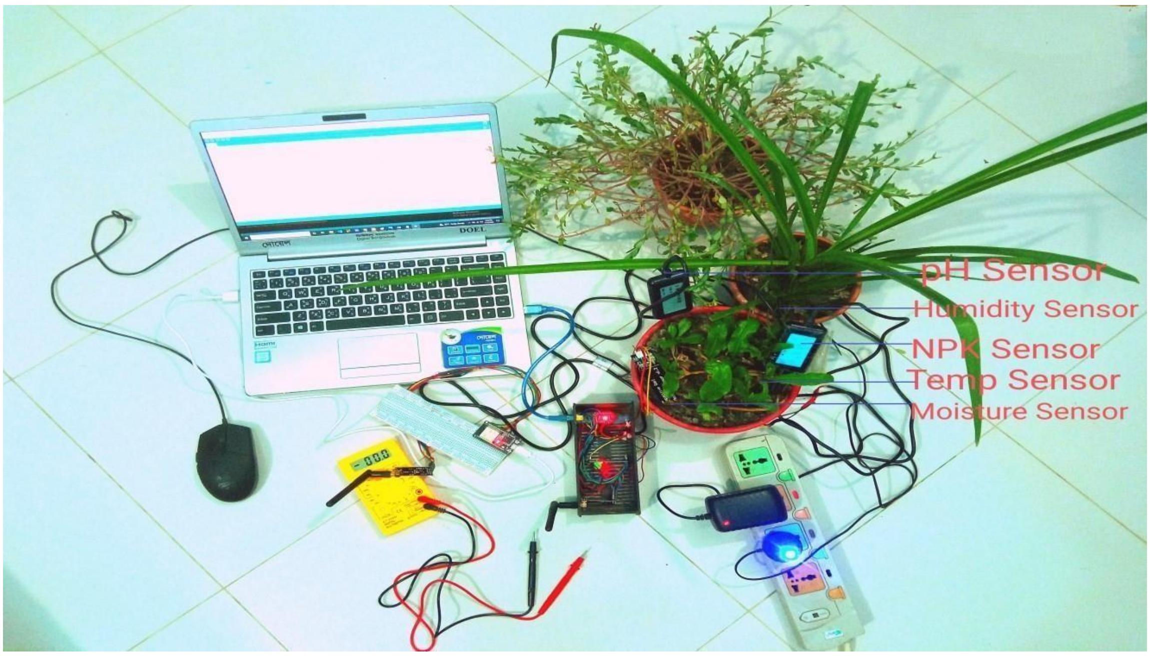
Fig. 6. Calling recommendation model method

Output: Rice

Table 4: Train and Test accuracy score

	model_name	accuracy_train	accuracy_test
1	Linear Regression	72.025319	70.808524
2	Random Forest Regression	100.000000	99.090909
3	Decision Tree	94.480519	93.484848
4	XGBoost	100.000000	98.636364

Fig. 7. Fertilizer monitoring of the garden tub



V. CONCLUSION AND FUTURE WORKS

The main focus of this work is to assist growers in making an informed decision before planting by advising them to produce the best crop depending on various factors. It is possible to grow more profitable crops if the correct information about the soil of the agricultural land is known for growing agricultural crops. The work modules will act asan effective tool for farmers to quantify the different soil properties of their land and recommend the most profitable crops suitable for that land. Farmers in rural areas in particular will be encouraged to use new technologies (IoT, Cloud Computing, AI, ML) to increase agricultural yield through which smart villages can be implemented. The proposed framework will help open up new horizons in making soil and agriculture researchers more efficient with effective solutions by using them to develop an IoT and machine learning-based integrated agricultural management system. This system can be improved to include anabundance of features depending on the season, such as weather prediction, extreme drought, flood detection, and agrarian price forecasting. To make it easier for farmers to use, a Mobile application would be developed and embedded with the suggested system.

REFERENCES

[1] A. Chlingaryana, S. Sukkarieha and B. Whelan, "Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review," *Computers and electronics in agriculture*, 2018, vol. 151, pp. 61-69.

[2] N. Ahmed, D. De and I. Hussain, "Internet of Things (IoT) for Smart Precision Agriculture and Farming in Rural Areas," *IEEE INTERNET OF THINGS JOURNAL*, 2018, vol. 5, no. 6, pp. 4890 - 4899.

[3] N. Ananthi, J. Divya, M. Divya and V. Janani, "IoT based smart soil monitoring system for agricultural production," *2017 IEEETechnological Innovations in ICT for Agriculture and Rural Development (TIAR)*, Chennai, India, 2017, pp. 209-214.

[4] C. Bepery, M. S. S. Sozol, M. M. Rahman, M. M. Alam and M. N. Rahman, "Framework for Internet of Things in Remote Soil Monitoring," 2020 23rd International Conference on Computer and Information Technology (ICCIT), DHAKA, Bangladesh, 2020, pp. 1- 6, doi: 10.1109/ICCIT51783.2020.9392658.

[5] R. Maheswari, H. Azath, P. Sharmila and S. Sheeba Rani Gnanamalar, "Smart Village: Solar Based Smart Agriculture with IoT Enabled for Climatic Change and Fertilization of Soil," *2019 IEEE 5th International Conference on Mechatronics System and Robots (ICMSR)*, Singapore, 2019, pp. 102-105.

[6] R. Dash, D. Ku Dash and G. C. Biswal, "Classification of crop based on macronutrients and weather data using machine learning techniques," *Results in Engineering*, 2021, vol. 9, no. 100203.

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

[8] A. Sharma, A. Jain, P. Gupta and V. Chowdary, "Machine Learning Applications for Precision Agriculture: A Comprehensive Review," in *IEEE Access*, vol. 9, pp. 4843-4873, 2021, doi: 10.1109/ACCESS.2020.3048415.

[9] M. I. Jordan and T. M. Mitchell, "Machine learning: Trends, perspectives, and prospects," *SCIENCE*, 2015, vol. 349, no. 6245, pp. 255-260, doi: 10.1126/science.aaa8415.

[10] A. Kumar, S. Sarkar and C. Pradhan, "Recommendation System for Crop Identification and Pest Control Technique in Agriculture," *2019 International Conference on Communication and Signal Processing (ICCSP)*, Chennai, India, 2019, pp. 0185-0189.

[11] P. C. Sen, M. Hajra and M. Ghosh, "Supervised Classification Algorithms in Machine Learning: A Survey and Review," *Emerging Technology in Modelling and Graphics*, 2019, vol. 937, pp. 99-111.

[12] H. H. Rashidi, N. Tran, S. Albahra and L. T. Dang, "Machine learning in health care and laboratory medicine: General overview of supervised learning and Auto-ML," *Int J Lab Hematol*, 2021, vol. 43, pp. 15-22.

[13] A. Suresh, P. Ganesh Kumar and M. Ramalatha, "Prediction of major crop yields of Tamilnadu using K-means and Modified KNN," *2018 3rd International Conference on Communication and Electronics Systems (ICCES)*, Coimbatore, India, 2018, pp. 88-93.

[14] P. S. Nishant, P. Sai Venkat, B. L. Avinash and B. Jabber, "Crop Yield Prediction based on Indian Agriculture using Machine Learning," *2020 International Conference for Emerging Technology (INCET)*, Belgaum, India, 2020, pp. 1-4.

[15] A. Nigam, S. Garg, A. Agrawal and P. Agrawal, "Crop Yield Prediction Using Machine Learning Algorithms," *2019 Fifth International Conference on Image Information Processing (ICIIP)*, Shimla, India, 2019, pp. 125-130.

[16] S. V. Bhosale, R. A. Thombare, P. G. Dhemey and A. N. Chaudhari, "Crop Yield Prediction Using Data Analytics and Hybrid Approach," *2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)*, Pune, India, 2018, pp. 1-5.