2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT) | 979-8-3503-3509-5/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICCCNT56998.2023.10307726

# An Innovative Model For Soil Testing With Intelligent Crop Recommendation System

Meet Mehta*
*Dept. of Computer Engineering*
*DJSCE*
Mumbai, India
mehtameetoff2802@gmail.com

Abhijit Saha*
*Dept. of Computer Engineering*
*DJSCE*
Mumbai, India
abhijits1802@gmail.com

Siddharth Fulia*
*Dept. of Computer Engineering*
*DJSCE*
Mumbai, India
siddharthfulia18092002@gmail.com

Tushar Jain*
*Dept. of Computer Engineering*
*DJSCE*
Mumbai, India
tusharrahuljain@gmail.com

Dr. Narendra Shekokar
*Dept. of Computer Engineering*
*DJSCE*
Mumbai, India
narendra.shekokar@djsce.ac.in

***Abstract*—In India, a very small percentage of farmers do smart soil testing, with one of the major reasons being the time taken to get soil test reports from government labs and the fact that private soil testing is not affordable to the farmers. Except for this, many farmers find it difficult to interpret the soil testing report. Also, the soil testing report does not provide farmers with suggestions as to what crop they should grow. To overcome this problem, a system has been suggested in this paper that will use multiple sensors to collect soil information and send the information to the mobile app, which will suggest the best crop a farmer should grow based on the soil conditions. To provide crop suggestions, the system uses a RandomForest machine learning model, which is trained on a custom soil nutrients dataset. The dataset has many different parameters, such as nitrogen, phosphorus, rainfall, temperature, iron, sulphur, zinc, and pH value. The machine-learning model has an accuracy of about 92%.**

***Index Terms*—crop, crop recommendation, nitrogen, phosphorus, potassium, pH, zinc, sulphur, iron, internet of things**

## I. Introduction

India is a farming nation; agriculture is the primary source of livelihood for about 58% of India's population, but till today only a small percentage of farmers do soil testing. Multiple governments as well as private models have been developed to solve this problem, but none have been able to increase the percentage of farmers doing soil testing to a very great extent.

Currently, farmers have to either get the testing done at government facilities or private labs [1], [2]. Getting testing done at government labs is a very time-consuming process as many labs are not well equipped and the ones that are well equipped are overloaded. Due to this, the testing and result delivery takes time, which may cause farmers to miss the right time for sowing seeds. If farmers use private labs for testing, then the cost of testing is very high, making it unaffordable for them. Except for this, all the labs just provide a report telling the quality and amount of nutrients in the soil; none of them tells which crop should be grown based on the quality of the soil. Many farmers are unable to understand the report, which makes it difficult for them to select the right crop to grow. Recently, some startups have entered the farming industry. Their products help farmers get the soil testing done at a low cost and much faster, but they have very little coverage and also do not tell the best crop that should be grown.

As a solution, we will be making an ML model trained on a custom dataset that has multiple parameters, including nitrogen, phosphorus, rainfall, temperature, iron, sulphur, zinc, and pH value. The ML model will help farmers predict the right crop to grow based on the nutrients in the soil. First, using sensors and other techniques, all soil information such as NPK values, micronutrient content, and rainfall in that area will be collected. This information will then be entered into the ML model, which will then predict the crop that the farmer should grow. The model will be deployed using an API, and farmers will be able to access it using our mobile application. Using a mobile app will ensure that a vast number of farmers can use the crop recommendation service without having to buy or learn much about computers. Farmers will be able to track the history of soil testing and get information about other programmes that may be of use to them through the mobile application.

## II. Literature Review

Numerous research papers were consulted to gain an understanding of previous work on crop recommendation using AI and IoT. In addition, information regarding the numerous models and devices used to detect soil parameters was analysed.

Jagdeep Yadav et al. divided their paper into three sections: "Soil Fertility Prediction," "Crop Suggestions," and "Crop Yield Prediction" [3]. Each component utilises a distinct set of data. The soil dataset with 15 properties was used for soil analysis, the crop dataset with 4 attributes (temperature, humidity, pH, and precipitation) was used for crop recommendation, and the yield dataset with 6 properties was used for

crop yield prediction. For each component, multiple kinds of models, including J48, SVM, Random Forest, and ANN, were used. ANN had the highest precision across all three datasets, followed by Random Forest.

Karan Mehta et al. in their paper suggested predicting soil fertility and did crop prediction based on the soil nutrient readings [4]. If the soil fertility is insufficient to support the growth of a specific crop that the farmer wanted to produce, they also suggested offering a list of crop fertilisers. The pH value of the soil is measured as part of the soil analysis utilising an Arduino UNO, pH metre, and Wi-fi module (ESP8266). The back-end server-side maps this data to NPK values, and the soil is categorised as Low, Medium, or High class depending on the NPK levels. This was implemented via a website that displays the mapped pH and NPK data, as well as the predicted crop and crop fertiliser list, to enhance the current nutrient value of the soil.

A crop recommender system was put forth by S. Bangaru Kamatchi et al. and trained on a 47-attribute weather dataset from the previous three years obtained from UCI Data Repository [6]. Feature selection, data preprocessing, data prediction, data recommendation, and data classification made up their technique to forecast crops depending on the selected weather characteristics. A time series function is created to get the accurate weather conditions from the dataset, and this function factors in on the recommendation of the crop later in the ANN. The hybrid recommendation system generates a list of ranked crop suggestions by their weights using ANN for crop prediction and classification. Case-Based Reasoning (CBR) is utilised to boost the success ratio of the ANN model.

For crop prediction, the majority of the papers analysed rely on fundamental NPK values, whereas the dataset proposed by us contains fundamental NPK values in addition to micronutrients such as zinc, sulphur, and iron, as well as additional attributes such as temperature, pH, and precipitation. Also, the IoT device in previous publications consisted only of a pH sensor or an NPK sensor, whereas the device proposed by us includes a temperature sensor in addition to pH and NPK sensors to obtain real-time soil values.

## III. Methodology

### A. Dataset Creation

Our model does crop recommendations using multiple parameters which include nitrogen, phosphorus, potassium, sulphur, iron, zinc, temperature, rainfall, and pH. No dataset was publicly available which covered all these parameters so a custom dataset Modified Crop Dataset (MC dataset) was created. Crop Recommendation Dataset (CR dataset), a foundation dataset that was made available on Kaggle, was used to develop CropDataset.The CR dataset includes 22 cultivars and the attributes nitrogen (N), phosphorus (P), potassium (K), temperature, humidity, pH, and rainfall. Out of the 22 crops, 12 unique crops were identified for the MC dataset, which are rice, maize, chickpea, lentil, pomegranate, banana, mango, grapes, apple, orange, cotton, and coffee. For all these crops, the attributes N, P, K, pH, rainfall, and temperature were taken from the CR dataset, and for the rest of the attributes for the MC dataset which are zinc, sulphur, and iron, various resources which included websites and published documentation were used [12]–[22]. The CR dataset has a smaller number of tuples, so a new dataset with more tuples was produced utilizing the ranges for each crop attribute from the CR dataset. The new dataset had 5000 tuples for each crop and was constructed by randomly picking values from the obtained ranges for each crop's attribute. This dataset didn't consist of any noise, thus using Condition Tabular Generative Adversarial Network (CTGAN) additional noisy data was inserted and the final MC dataset of the same size was formed [23].

CTGAN models tabular data distribution using GAN-based methodologies and samples rows from the distribution. Because it normalizes the continuous data, it was utilized to incorporate some noisy data into the MC dataset. Continuous data is challenging to represent with one-hot encoding, whereas discrete data is easily represented. Therefore, CTGAN employs mode-specific normalization to transform the continuous variable into a vector containing the continuous data's information.

The conditional generator and training-by-sample techniques are utilized to address class imbalance issues. Conditional generators also referred to as conditional GANs (cGANs), are a form of generative model that can learn to generate new samples from a particular class or condition. In a conditional generator, the model of the generator is dependent on the class designation, which can help to balance the distribution of samples across classes. By conditioning the generator on a specific class identifier, the model can learn to generate samples that are representative of that class, thereby aiding in the resolution of the problem of class imbalance. Training-by-sample is a method that entails resampling the dataset to achieve a balanced class distribution. In this technique, minority class samples are duplicated or augmented to match the number of majority class samples. This technique can aid in addressing class imbalance by guaranteeing a balanced distribution of training samples for the model. However, this technique can also result in overfitting to the minority class, thereby reducing the model's generalization performance.

Based on one of the discrete columns, the conditional generator generates synthetic rows. Using training-by-sampling, the condition and training data are sampled based on the log frequency of each category, allowing CTGAN to explore all discrete values uniformly.

### B. Implemented Model

Multiple supervised learning algorithms can be used for a multi-attribute tabular MC dataset. The ML models have been selected based on their overall performance on the tabular dataset, their ability to handle large datasets, the amount of time required for training the model on large datasets, and their ability to avoid overfitting.

*1) Support Vector Machine (SVM):* SVMs are an algorithm for supervised learning that can be applied to classification and regression tasks. SVMs are based on the principle of

locating a hyperplane that defines a boundary between data categories. This hyperplane is nothing more than a line in 2-dimensional space. Each data point in the dataset is plotted in an N-dimensional space, where N is the number of features/attributes. They perform well on both high-dimensional and low-dimensional datasets, and because the MC dataset is currently of low to moderate dimensionality but will increase in the future, SVM would be beneficial in this situation. Also, different kernels can be specified for different datatypes, so if the MC dataset attributes continue to grow, the SVM model will continue to perform adequately. The SVM model's accuracy was less than 90%, so RandomForest was chosen as the next model to test.

*2) RandomForest (rf):* RandomForest is one of the widely used machine learning algorithms for classification and regression problems, it creates a decision tree based on the attributes and the importance of the attributes as they affect the final classification result. It has been used as it prevents overfitting on the dataset and ensures that the classification result is accurate. The fact that the rf operates with great accuracy on big datasets ensures that the farmers can perform soil testing reliably and quickly. As a result, over time as data is gathered and the MC dataset grows, neither the training time nor the accuracy of the model will be impacted. As the MC dataset is customized, it may contain certain values that are perhaps wrong and the noise was also added using the CTGAN model; however, utilizing rf ensures that noisy data will not affect the model's accuracy. Moreover, rf is an ML model which works effectively even if the dataset consists of multiple attributes making it one of the most appropriate models for a crop recommendation system. When used separately, the accuracy of the rf and SVM models was approximately 92% and 89%, respectively. Convolution Neural Network (CNN) was used to improve the accuracy of the deep learning model. A new model which was a combination of rf and CNN was implemented so the accuracy can be increased ensuring the farmer's best crop recommendation of the soil being tested so that they can make the highest profits.

*3) Convolutional Neural Network (CNN):* CNN is a type of algorithm for deep learning. It includes convolutional layers, pooling layers, and fully connected layers. An input layer and one or more concealed layers are connected one after the other in a fully connected Deep neural network. Each neuron receives input from the neurons of the preceding layer or the input layer. The output of one neuron becomes the input for other neurons in the next network layer, and this process continues until the output of the network is produced by the final network layer. The layers of the neural network apply a series of nonlinear transformations to the input data, enabling the network to learn complex representations of the input data, and since the MC dataset contains various attributes, it would be easy to recognize patterns among them. CNN is mostly used for images, but since it forms fully connected neuron layers it could help in increasing the accuracy of the previously applied rf model.

### C. Hardware

As we can see in image 1 hardware used in the system includes a DS18B20 temperature sensor, a pH sensor, an NPK sensor, an Arduino UNO, an RS-485 to TTL converter, and an ESP32 WiFi module. The image 1 illustrates that DS18B20 temperature sensor is connected to pin 5 (light yellow wire) of the Arduino UNO, while the NPK and pH sensors transfer data to the Arduino UNO using the RS-485 protocol.Ultimately, all data is sent to the ESP32 WiFi Module, which relays the information to the server. The WiFi module is connected to the Arduino board using its 5V and ground ports. For the purposes of communication, the Wifi module is attached to the Arduino nano through the TX and RX pins, which correspond to pins 1(as shown in green wire in image 1) and 0(as shown in orange wire in image 1) on the Arduino UNO,respectively. Both the pH and NPK sensors' input voltages are connected to an external source, and the RS-485 and external adapter serve as grounding points. Both the NPK sensor and the pH sensor have a modbus A pin and a modbus B pin, which are connected to the RS-485 module. The modbus A pin is denoted by the colour yellow, whereas the modbus B pin is denoted by the colour blue. The temperature sensor has an input voltage of 5V, and the data wire is attached to a resistor whose value is 4.7K ohms.

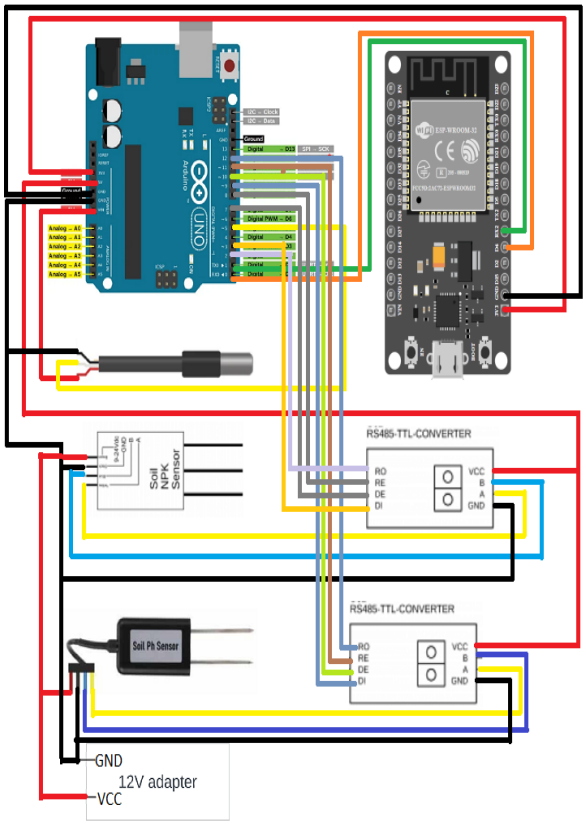



Fig. 1: IOT Diagram

## IV. Implemented Architecture

A soil sample will be collected and placed within the container. The temperature, pH, and NPK sensors will then be submerged in the soil. Using the ESP32 Wi-Fi Module, the collected data from the sensors will be transmitted to the backend. This information will then be displayed on customizable fields on the app (Figure 5(a)) that a farmer could also use to obtain produce recommendations from an existing soil testing report. The rainfall values would need to be entered. For this purpose, a link to the IMD website is provided below the input field. The farmer must then determine the rainfall amount in their region and input the value. Once all values have been input, the app would display the suggested crop by calling the custom API where the model has been hosted. The model performs all calculations regarding the inputted values and returns the result to the application. The results are stored in the app so that farmers can access them at any time (Figure 5(a)) and can view the details for each recommendation result (Figure 5(b)).

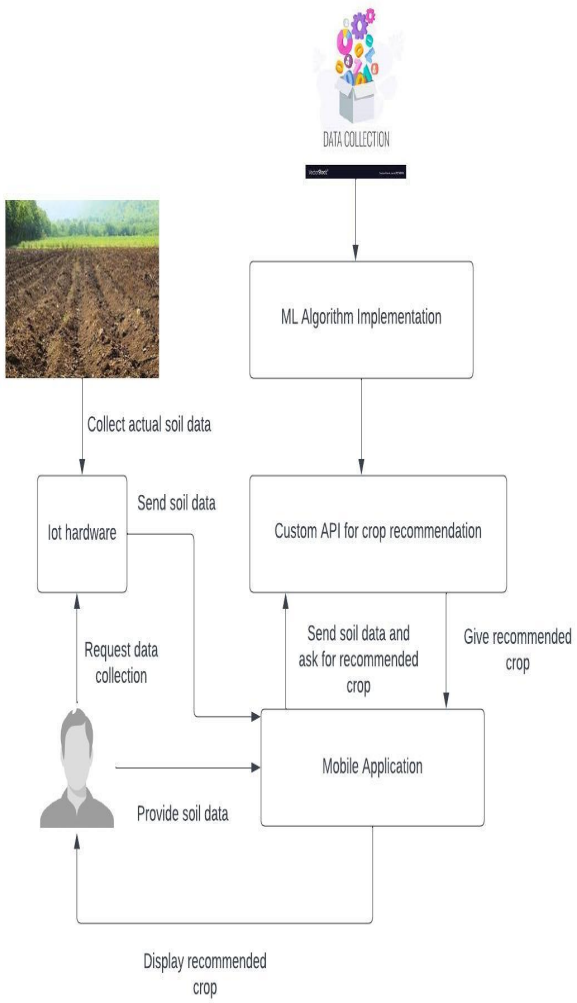
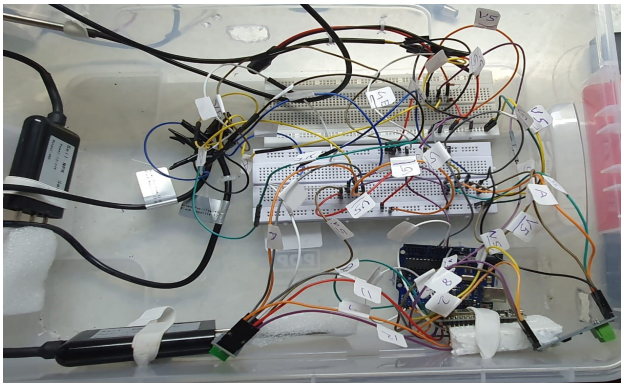
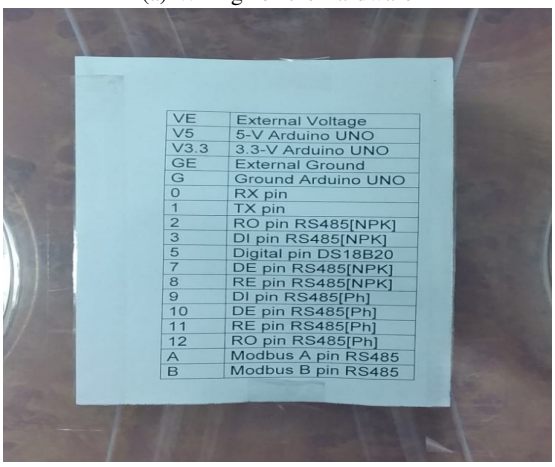


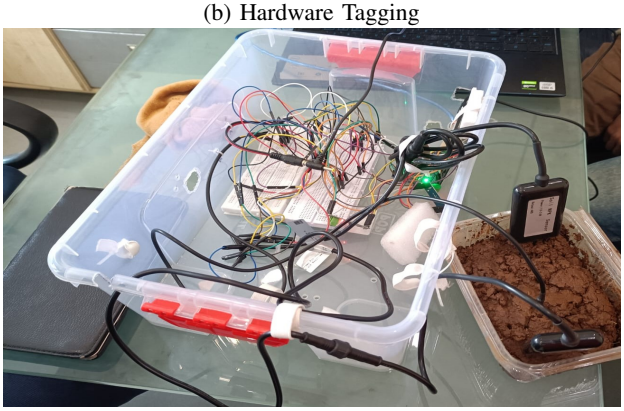

Fig. 2: Complete Architecture



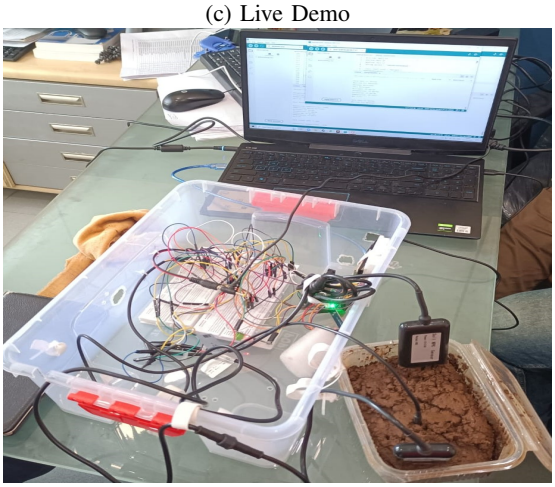
(a) Wiring for the hardware




(b) Hardware Tagging

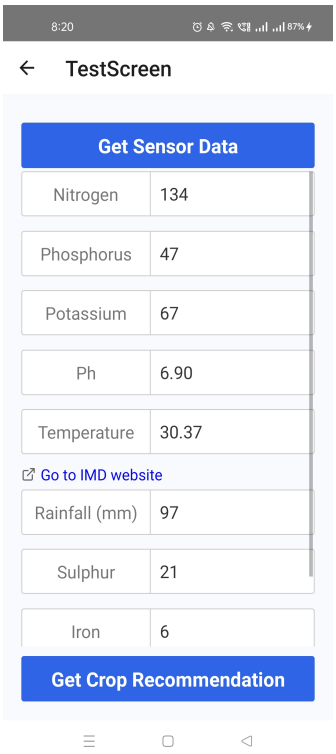


(c) Live Demo



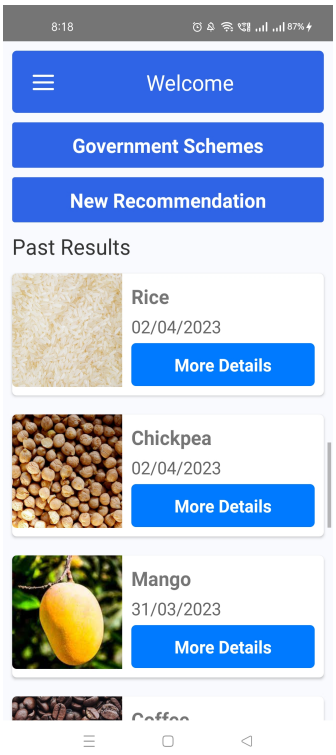
(d) Live Demo

Fig. 3: IOT Hardware Implemenation




(a) Crop Recommendation Form


(b) Predicted Crop - Rice




(a) Past Crop Recommendation


(b) List of Government Schemes

Fig. 5: App implementation

## V. Result

Models like Random Forest (rf) and other permutations of Random Forest and Convolution Neural Network (CNN) were among those tried out for the recommendation system. The initial model was a straightforward rf one, and it was only approximately 92% accurate. The second model used hyperparameter adjustment to produce an accuracy of around 89% using a Support Vector Classifier (svc) or Support Vector Machine (svm). The third model used a convolutional neural network (CNN) with two dense layers (rf+CNN2). Roughly 90% accuracy has been achieved with rf+CNN2. Similar to rf+CNN2, the third model is rf+CNN4, which added 4 dense layers for improved accuracy to around 91%. Following rf+CNN2 and rf+CNN4, the rf+CNN6 model added 6 dense and achieved around 88% accuracy. The rf model, which was the last to be created and is now being utilised in the system, proved to be the most accurate. Table 1 and the Fig 9 highlights and compares the accuracy of different models.

| *Model* | *Accuracy* |
|---|---|
| rf | 0.9189 |
| svm | 0.8923 |
| rf + CNN2 | 0.9032 |
| rf + CNN4 | 0.9052 |
| rf + CNN6 | 0.8792 |

TABLE I: Model Performance

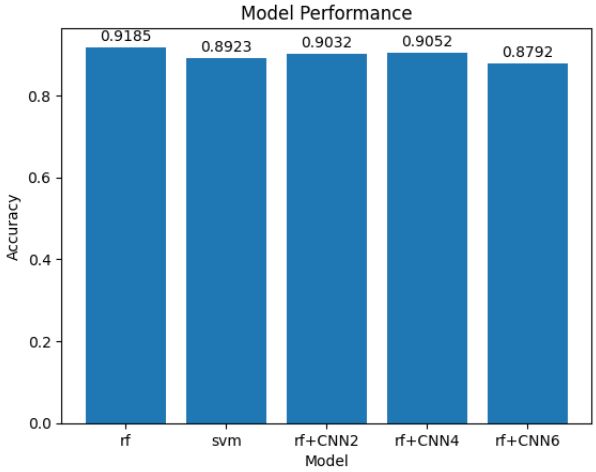



Fig. 6: Model Performance

## VI. Conclusion and Future Scope

The main objective the paper is to develop a smart integrated system that allows farmers to do soil testing and get results along with best crop recommendation in minimum amout of time. The paper suggest using rf model trained on the custom MC dataset. The model has been suggested after researching about several supervised learning, deep learning and a combination of supervised and deep learning models.

The rf model has been used at it provides high accuracy, low training time and prevents overfitting even with large datasets. The accuracy provided by the rf model trained on MC dataset is about 92%. The model has been implemented by creating a custom API and can be access using the mobile appilication.

As shown in figure 3(c), the Arduino then compiles the informational data packet into a single data packet that is received by the ESP32 WiFi module and sent to the server for uploading.Figure 3(d) illustrates how the application will subsequently retrieve the data from the server after that is finished.

Farmers can use the suggested system to do soil testing usin the iot hardware which sends the data to mobile app or input the data manually in the app. The application will then provide the best crop recommendation using the ML model. The mobile application can will also provide history of soil testing and other information like information of various schemes which can be used by farmers.

The ML model used in the proposed solution was trained on a small dataset of about 60000 rows and 12 unique crops. In order to overcome this, over time, as more data is gathered from farmers, the model could be retrained in order to increase accuracy and cover a much larger range of crops for the recommendation. The machine learning algorithm only recommends one crop at the moment, but when more data is gathered, other features like multiple crop suggestions can be offered, giving the farmer the choice of selecting which crop he wants to grow based on other parameters like return on investment. Except this farmers frequently decide which crops to produce based on which crop provided the best return on investment the previous year. As a result of multiple farmers growing the same item, the market price may rise overall. In order to make it simpler for farmers to choose a crop that would provide them with the best return on investment, a feature that displays the demand for a specific crop as well as which crops are grown in a given region can be introduced.

## REFERENCES


[1] BharatAgri App [Online:Available] https://www.bharatagri.com.

[2] Government Labs [Online:Available], https://services.india.gov.in/service/detail/locate-soil-testing-laboratory.

[3] Jagdeep Yadav, Shalu Chopra, Vijaylakshmi M, "SOIL ANALYSIS AND CROP FERTILITY PREDICTION USING MACHINE LEARNING", Article ID MRAE10082, ISSN: 2349-2163.

[4] Karan Mehta, Hiral Mehta, Alok Pai, Kaveri Sawant, "Soil Analysis and Crop Fertility Prediction", International Journal of Emerging Technologies and Innovative Research (www.jetir.org), ISSN:2349-5162.

[5] Lakshmi Patil, Dr. K Saraswathi, "Crop Yield Prediction on the Basis of Soil Composition using Machine Learning Algorithms", ISSN:1001-1749.

[6] S. Bangru Kamatchi, R. Parvathi, "Improvement of Crop Production Using Recommender System by Weather Forecasts", Procedia Computer Science 165 (2019) 724–732.

[7] "Iot based soil nutrient monitoring with arduino esp32." https://how2electronics.com/iot-based-soil-nutrient-monitoring-with-arduino-esp32/ (accessed Aug. 10,2022).

[8] https://how2electronics.com/diy-soil-ph-meter-using-soil-ph-sensor-arduino/ (accessed Dec. 09,2022).

[9] https://how2electronics.com/iot-based-soil-nutrient-monitoring-with-arduino-esp32/#Capacitive_Soil_Moisture_Sensor (accessed Dec. 09,2022).

[10] https://how2electronics.com/iot-based-soil-nutrient-monitoring-with-arduino-esp32/#DS18B20_Waterproof_Temperature_Sensor (accessed Dec. 09,2022).

[11] https://robu.in/arduino-pin-configuration/#: :text=Analog%20Pins%3A%20The%20pins%20A0%20to%20A11%20are%20used%20as,known%20as%20a%20UART%20pin (accessed Sept. 01, 2022).

[12] "Crop Guide: Potato Nutritional Requirements." https://www.haifa-group.com/crop-guide/field-crops/crop-guide-potato/nutrients-growing-potatoes (accessed Sept. 01, 2022).

[13] "Zinc content in rice" (accessed Sept. 01, 2022) http://www.dietandfitnesstoday.com/zinc-in-rice.php.

[14] "Iron content in rice" (accessed Sept. 01, 2022) http://www.dietandfitnesstoday.com/iron-in-rice.php.

[15] https://www.researchgate.net/publication/330727473_Effect_of_Different_Sources_and_Levels_of_Sulphur_on_Growth_and_Yield_of_Maize_Zea_mays_L.

[16] Muhammad Amir Maqbool AbduRahman Beshir "Zinc biofortification of maize (Zea mays L.): Status and challenges." https://onlinelibrary.wiley.com/doi/full/10.1111/pbr.12658 (accessed Sept. 01, 2022).

[17] "Lentil – Fertility" https://albertapulse.com/lentil-seeding/lentil-fertility/ (accessed Sept. 01,2022).

[18] M. Hasani, Z. Zamani, G. Savaghebi, R. Fatahi. "EFFECTS OF ZINC AND MANGANESE AS FOLIAR SPRAY ON POMEGRANATE YIELD, FRUIT QUALITY, AND LEAF MINERALS. "https://www.uniwinchemical.com/news-posts/effects-of-zinc-and-manganese-as-foliar-spray-on-pomegranate-yield-fruit-quality-and-leaf-minerals/ (accessed Sept. 01, 2022).

[19] "Climate & Soil Requirements of Pomegranate" https://www.kalliergeia.com/en/climate-soil-requirements-of-pomegranate/ (accessed Sept. 01,2022).

[20] L.Richard. "Nutrient needs by apple trees" https://www.goodfruit.com/nutrient-needs-by-apple-trees/ (accessed Sept. 01, 2022).

[21] "Boron and Zinc Use Recommendations for Cotton." https://agrihunt.com/articles/pak-agri-outlook/boron-and-zinc-use-recommendations-for-cotton/ (accessed Sept. 01, 2022).

[22] "NUTRITIONAL REQUIREMENTS OF COTTON DURING FLOWERING AND FRUITING" https://www.cotton.org/foundation/upload/F-F-Chapter-4.pdf (accessed Sept. 01,2022).

[23] https://docs.google.com/spreadsheets/d/1DiImSMaIw4Na0w-yPqp9Cy2PXpdXlzmnxDK3HP5yKRg/edit?usp=sharing (accessed Sept. 01, 2022).