2022 Third International Conference on Intelligent Computing Instrumentation and Control Technologies (ICICICT) | 978-1-6654-1005-2/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICICICT54557.2022.9917584

Soil Mineral Prediction of Crops Using Machine Learning

Prema T. Akkasaligar
Department of Computer Science and Engineering
BLDEA's V. P. Dr. P. G. Halakatti College of Engineering and Technology
Vijayapur-586103, Karnataka, India
premasb@rediffmail.com

Sunanda Biradar
Department of Computer Science and Engineering(Aritifical Intelligence and Machine Learning)
BLDEA's V. P. Dr. P. G. Halakatti College of Engineering and Technology
Vijayapur-586103, Karnataka, India
sunanda_biradar@rediffmail.com

Manjula C. Gudgeri
Department of Engg science and Humanities, Mathematics
KLE Dr M S Sheshgiri College of Engg and Tech Belagavi, Karnataka, India
manjulagudgeri11@gmail.com

Sana Mohammadi A Mulla
Department of Computer Science and Engineering
BLDEA's V. P. Dr. P. G. Halakatti College of Engineering and Technology
Vijayapur-586103, Karnataka, India
sanamohammadiaus@gmail.com

***Abstract*—In India, agriculture is the primary source of income. Agriculture employs the majority of the people. Because India's financial system is so reliant on agriculture, there is a pressing need to boost agricultural output generally. Soil is the most important natural resource in agriculture. The most important requirement is good soil quality. The pH of the soil is crucial in this regard. The acidity of the soil is measured by its pH. The total yield is influenced by the pH level. Nitrogen, Potassium, and Phosphorus all influence pH of soil. Farmers can then begin cultivating crops as a result of this. pH sensors are commonly used to determine the pH of soil. However, because they increase the cost of manufacture, they are not always practicable. To address this problem, the proposed method determines the pH value while simultaneously recommending a crop. The crop recommendation method takes into account factors such as soil pH, season, and temperature.**

Keywords—Nitrogen-N, Potassium-P, Phosphorus-K, pH level and Soil prediction.

I. INTRODUCTION

India is a country with a wide range of physical and cultural characteristics. Our country's primary occupation is agriculture. The agriculture industry is the backbone of the Indian economy. With the advancement of technology, it is more important than ever to educate farmers on diverse agricultural approaches. This will make farming easier and provide the farmer with a variety of benefits. As a result, it is critical to educate farmers in a variety of ways, such as anticipating disease and taking precautions to avoid loss, anticipating soil pH values without the ranchers visiting research centres. The crop suggestion is likewise conceivable and made accessible through current innovations, climate data, and advertising data for the field. In [1], system that can forecast soil pH, which is one of the most important factors in improving crop maturity and quality. They use a soil picture to measure the pH value, and by extracting the soil characteristics, they have proposed a crop by predicting the PH value with the assistance of a neural network. This technique would be extremely beneficial to Indian farmers in learning more about their field soil and maximising crop productivity. In [2], there are several types of soil, each with its own set of characteristics for different crops. Different strategies and models are now being employed in agriculture sector to boost crop quantity. The major goal of this system is to develop a model that will assist farmers in determining which crops should be planted in specific soil types. Farmers may also use contemporary technology to assist them export agricultural produce and preserve their fields. With the use of image processing, technology may assist in determining the pH of soil. Processed soil elements may be recognised, and precise results can be given for farmers.

The proposed method deals with the prediction of the crop suitable to the particular field depending the soil contents (like NPK and pH). The main objectives are,

•Prediction of crop using Random Forest Algorithm.

•Prediction of crop based on season.

•Prediction of pH value using N, P, K of a soil for particular region.

II. LITERATURE SURVEY

Agriculture is a significant part of India's economy. As a result of industrialization and pesticide overuse, soil resilience has been decreased in recent years. In many circumstances, agriculture's methods for raising output are insufficient. The most common issue that Indian farmers encounter is a lack of understanding about the best crop to grow based on their soil requirements, which affects productivity.

In [3], author created a Matlab-based programme to determine the soil colour using digital image processing. The dirt part is separated from the backdrop using HSV segmentation. The KNN classifier labels the photos with Munsell soil notation and classifies them based on their RGB values.

In [4], author have implemented the support vector machine approach for categorization of soil types, image capture, image pre-processing, feature extraction, and classification are all methods in the soil classification process. The low pass filter, Gabor filter, and colour quantization approach are used to extract texture information from soil photographs. The statistical metrics used include mean amplitude, HSV histogram, and standard deviation

In [5], authors have implemented, They examine the effectiveness of Artificial Neural Network and Support Vector

Machine classifiers in identifying and classifying four pairs of similar-looking food grains: Finger Millet, Split Greengram, Soyabean, Pigeon Pea, Mustard, Aniseed, Cumin-seeds, and Split Blackgram. To capture and process colour photographs of these grains' samples, algorithms are being developed. The algorithms that have been created include 18 colors-Hue Saturation Value and 42 wavelet-based texture characteristics were extracted using this method. Neural Back Propagation. A BPNN-based classifier is created utilising three feature sets: colour – HSV, wavelet-texture, and their combinations. A hybrid models works well for the classification.

In [6], author have proposed the extensible crop yield prediction framework which allows for crop, dependent and independent factors, and dataset selection for crop yield prediction in precision agriculture.

In [7], the author has proposed the Crop Selection Method (CSM) to solve the crop selection problem and optimise crop net productivity rate over season, resulting in the country's maximum financial growth.

In [8], the author has shown forecast crop production and recommend the optimum crop for enhancing the agriculture sector's quality and profitability. Further, soil type, temperature, humidity, water level, spacing, depth, soil pH, season, fertiliser, and months are all factors considered by the system. This prediction may be made using the Bayesian process, which results in excellent prediction accuracy.

In [9], authors have implemented, the acquired data is analysed and trained in order to create a model using machine learning methods. The trained model is extremely valuable for more correctly and effectively estimating crop production. The needed fertiliser ratio is also predicted, allowing the former to achieve a higher yield. With enhanced new technology of machine learning algorithms, this notion primarily assists the former in increasing his earnings and making appropriate judgments when spreading the seeds into the field. The study focuses on the comparison of two machine learning algorithms, agricultural yield prediction, and fertiliser amount prediction. The results of the random forest and backpropagation algorithms are compared. When compared to the Backpropagation algorithm, the random forest approach is more efficient. Machine learning algorithms are successfully used to estimate crop production and fertiliser application amounts. The web applications might be improved to assist the formers in obtaining a higher yield and more income.

In [10], author have discussed as leak detection, fire detection, room temperature-humidity detection, and avoiding overflow of water from the above tank are all included in the proposed home security system. They suggested system intends to create a safe home system by making accessibility simpler for disabled people, especially youngsters and the elderly. Temperature-humidity sensor, water level monitoring sensor, fire-gas detection sensor, and motion detection sensor are all used in the proposed work. DTH11 is used in the suggested model to capture the room's temperature and humidity. It also uses a conductor to monitor the water level in a home's overhead tank. The MQ3 sensor detects fire and gas, while the PIR sensor detects human presence. In the event of unexpected events, alarm messages are delivered to the owner. There project's effort has resulted in a hardware device that can be used to monitor the house or office for any potentially dangerous actions. This item improves the comfort and security of your house.

In [11], author have suggested and tests, a technique for predicting agricultural productivity based on historical data. This is accomplished through the use of association rule mining on agricultural data. The goal of this study is to develop a crop production, forecast model that may be utilised in the future. This work gives a brief investigation of agricultural production prediction utilising data mining techniques based on association rules for the selected region, namely the Tamilnadu district in India. The results of the experiments reveal that the suggested work accurately predicts agricultural yield output. The goal of this study is to develop and deploy a rule-based system for predicting agricultural yield output based on historical data.

In [12], authors have proposed a model for crop prediction. It assists farmers to determine the form of crop to be developed based on the climatic situations and vitamins present within the soil. This paper compares k-Nearest Neighbor, Decision Tree, and Random Forest Classifier for the usage of different criterions Gini and Entropy. The result shows that Random Forest offers the highest accuracy. The comparative analysis of three unique supervised system models shows that Random Forest predicts the fine-suitable crop for the unique land that can assist farmers to develop vegetation more successfully.

In [13], authors have implemented method which was created by the author utilising machine learning algorithms for the benefit of farmers. Based on content and climate, their technology will recommend the ideal crop for a certain piece of land. weather conditions in addition, the system offers data. regarding the needed fertiliser content and amount seeds for planting. As a result, using their method, farmers may develop a new crop type to increase profit margins and soil contamination can be avoided. they designed a farmer-friendly technology with the goal of reducing such losses. A graphical user interface that predicts the appropriate crop for each situation.

This system would also provide for specific land and information about necessary nutrients to add up, and necessary seeds Expected yield and market price for cultivation. As a result, farmers are more likely to make the best crop selection decisions, resulting in the development of the agricultural industry original concept.

Literature review shows, the scope for crop predication using supervised learning exists.

III. BASIC DECISION TREE CONCEPT

The notion of a decision tree is more related to a rule-based system. The decision tree algorithm will generate a set of rules based on the training dataset's targets and characteristics. On the test dataset, the same set of rules may be utilized to make the prediction as shown in Figure 3.1
The information gain and Gini index computations will be used in the decision tree method to calculate these nodes and build the rules. Instead of utilizing information gain or the Gini index to calculate the root node in the random forest algorithm, the process of locating the root node and separating the feature nodes will be done at random.

A. Machine Learning using Random Forest Algorithm

Let's start with the algorithm for creating a random forest. Consider ‘a’ as N/P/K value of soil and ‘b’ be the Total number of (N+P+K) features, then ‘c’ be the resultant crop predication.

Step 1. Start

Step 2. Choose "a" feature at random from the total number of "b" characteristics.

where a<< b

Step 3. Among the “a” feature, compute the node “c” using best split point.

Step 4. Using the best split, split the node into daughter nodes.

Step 5. Repeat steps 1–3 until single node is attained.

Step 6. Build a forest by repeating steps 1 through 4 "n" times to create "n" trees.

Step 7. Stop

The random forest method begins by picking "a" feature at random from a total of "b" features. For prediction of crops, We applied the Random Forest algorithm, which is the most widely used classification system in machine learning. Both classification and regression problems may be solved using the random forest methodology. The random forest method is a supervised classification technique. It creates a dense forest with many trees. The greater the number of trees in a forest, the more powerful it looks to be. Similarly, with the random forest classifier, the bigger the number of trees in the forest, the more accurate the results. So, in our project we use Random Forest algorithm as a classifier algorithm.

B. How Random Forest algorithm works

To create accurate and reliable prediction, random forest creates several decision trees and merges them together. Randomness can be added to the model by performing this approach. While splitting any node, random forest searches for the important parameters, and then seeks for the best among the subset of random characteristics. As a result, a model with increased precision in a wide range of situations emerges.

The working of algorithm has two stages.

•Random forest creation.

• Prediction from the created random forest classifier.

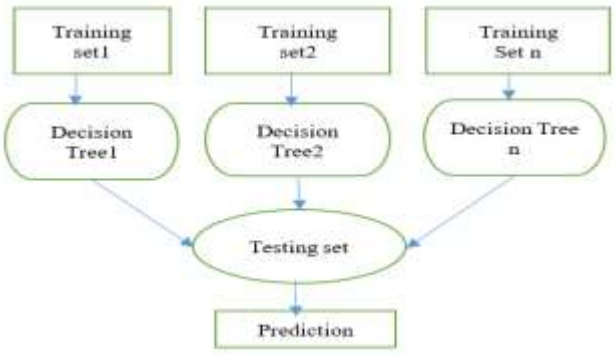

Figure 3.1. Tree diagram of Random Forest Algorithm

IV. METHODOLOGY

The flowchart for the python script for predicting the crops using random forest algorithm is show in Figure 4.1 below. At very first step when input is received from farmer through smart kit and GUI, the data (headers & values) from the csv file is read and treat them as test case values. At the same time the data (headers & values) from the training dataset are also read.

In next step the code extracts the headers ([“Nitrogen”, “Phosphorous”, “Potassium”, “Organic Carbon”, “pH”]) from training dataset and from test case it extracts just ([“Nitrogen”, “Phosphorous” , ”Potassium”]) as we need to compute pH for test inputs to predict crop.

Then a function checks weather both the file format is csv or not. If “yes” then it proceeds and if “no” then another function converts the file type to csv format. Further the values of training dataset are splatted into features (the values to NPK) and target (pH). These values of training dataset are used to train the random forest classifier using algorithm. Once the classifier is done with training then next it goes to the computing the pH using test input given by farmer. So, depending the pH value of soil computed, suitable crops are predicted and these predicted crops are displayed using GUI.

The main benefits are:

•An artful guide system for the beginner in the agriculture field.

•An IoT based smart kit, explores the uneducated farmers to globe.

•This system helps in predicting the correct crop to be grown depending on the soil content and season.

•Sensing the moisture of the field.

•Providing the intellectual crop selection system.

•Helps in the confusion of the various situations like, what to grow? , when to grow? and what all the things to be considered for the beginners.

IoT plays a vital role and many applications like smart home, agriculture field, security system, etc. As the hardware implementation we use the below sensors for smart agriculture field for soil testing.

Soil N, P, K Sensor:

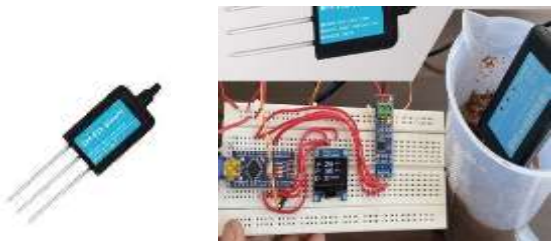

Figure 4.2: Soil N, P, K sensor and the connectivity with the Arduino uno

To determine the nitrogen, phosphorous, and potassium content of a soil. The JXCT Soil NPK sensor is a low-cost, rapid measurement speed, high-precision, measure multiple types of soil and portable Modbus RS485 sensor. As shown in Figure 4.2. It has 4 wires, brown wire is the vcc wire and should be connected to 9v to 24v dc power supply, the black wire is the ground wire and it should be connected with the

Arduino's ground pin and the remaining 2 wire are blue and yellow wire and these two wires should be connected with b and a pin of max 485 modbus module. All you have to do is stick the probe into the ground and use Arduino to read the data. So, further detail which will further give these N, P, k values as input to the farmers and hence predicts the pH and Crop Predication.

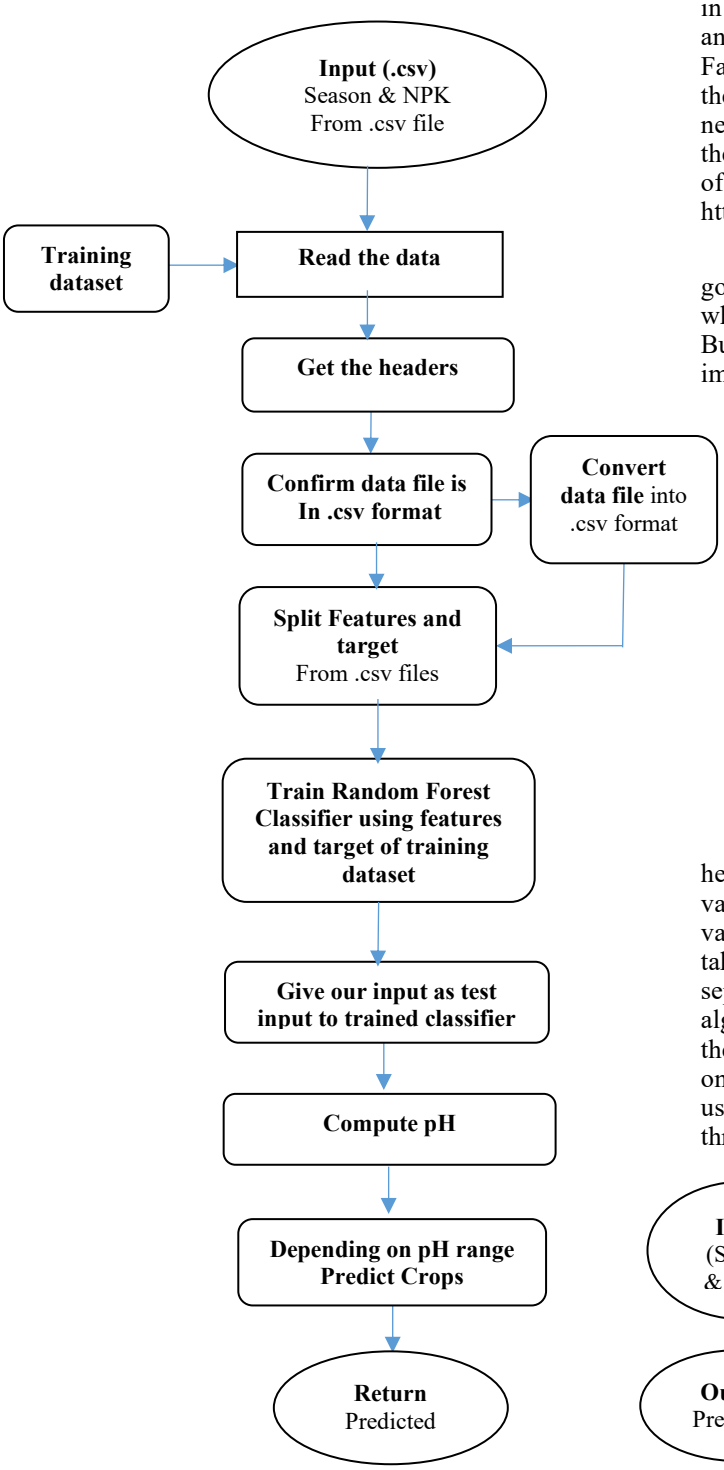

Figure 4.1: Flowchart of Script for Predicting Crops

V. Results and Discussion

The Random Forest Algorithm is implemented on 11th Gen Intel Core i5 system, x64-based processor with 16GB RAM. The software required is Python 3.9 or higher version.

A. Inputs to the Algorithm

So for in our application, we provide the input to algorithm in terms of soil contents like Season, Nitrogen, Phosphorous and Potassium. These values are given to code in csv format. Farmers can get these soil contents (NPK) value by soil test of their field and this could be done in the soil testing laboratories near them. There are many government soil testing labs in all the districts and even some private labs are present. The details of govt. soil testing labs can be seen in govt. web portal https://farmer.gov.in/STL.aspx. As shown in Figure 5.1.

The department of agriculture and farmers' welfare, government of India provide soil health card after soil testing, which contains the several details of field and soil contents. But mainly the pH, Nitrogen, Phosphorous, Potassium are important. we require these itself for our project

Figure 5.1: Soil Health card (Kannada version)

As discussed, in before the farmer needs to get his soil health card for giving the input to smart kit. Once we get the values of NPK after soil testing, farmer needs to enter the NPK values to our system through TFT LCD touch screen. So, GUI takes NPK values and insert them into a .csv (comma separated values) file. Which is then given to random forest algorithm which computes the pH of the soil and depending the pH value it predicts the suitable crops which can be grown on that field. This predicted crop list is displayed to farmer using TFT screen[14]. Every input and output events happen through the GUI only as shown in Figure 5.2.

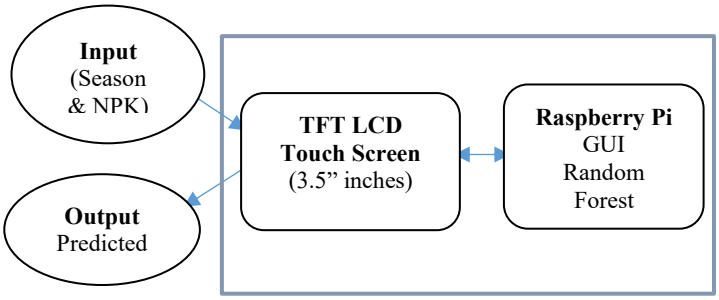

Figure 5.2: Block diagram of System

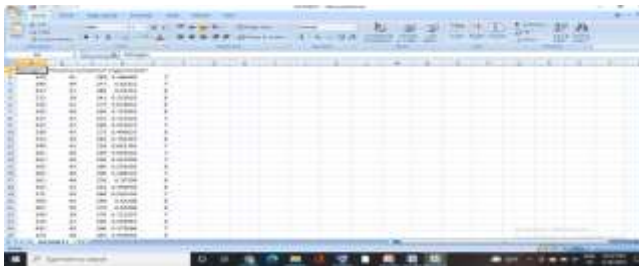

Figure 5.3: Snapshot of training dataset1

The above Figure 5.3 shows the 300-training dataset for monsoon season. The Nitrogen value for soil ranges from 240-480kg/ha, Phosphorus value ranges from 11-22 Kg/ha and Potassium ranges from 110-280Kg/ha these values are calculated by random forest algorithm and fetches the pH value which ranges from 5.5-7.5.

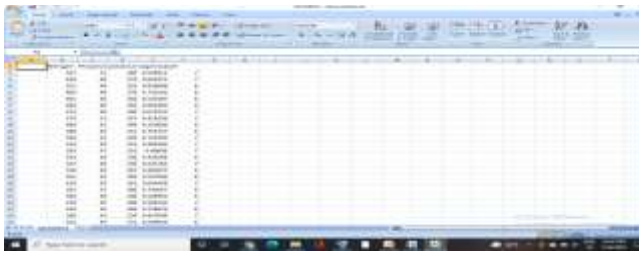

Figure 5.4.: Stimulation of the soil moisture

The above Figure 5.4 shows the stimulation of the soil moisture in Tinkercad software. It is a virtual hardware model using the potentiometer for detection of the soil pH value with the connection of Arduino uno.

Figure 5.5.: Snapshot of training dataset3

The above Figure 5.5 shows the 300-training dataset for summer season.

Figure 5.6.: Snapshot of test dataset

The Figure 5.6 shows the test dataset for test data.

Figure 5.7: Snapshot of GUI

The Figure 5.7 shows the GUI representation, prediction of crop. Detecting the suitable crop predication of the soil by using the N, P, K value and calculating the pH value of the soil. The values of training dataset are used to train the random forest classifier. Once the classifier is done with training, then next it goes to the computing the pH using test input given by farmer. So, depending the pH value of soil computed, suitable crops are predicted and these predicted crops are displayed using GUI.

Table 1. Records of the predicting crops with its pH value

Enter the nutrient values			Enter the season	pH	Predicated crops
N	P	K			
10	20	30	Monsoon	7	Sunflower
50	30	20		8	Not suitable for any crops
30	20	60		6	Groundnut and Maize
30	20	60	Rainy	5	Maize
20	30	60		7	Jowar
50	30	40		8	Not suitable for any crops
60	30	40	Summer	5	Maize
70	40	50		8	Not suitable for any crops
30	40	50		6	Groundnut and Maize

The Table 1 shows the comparative accuracy of the nutrient values which will predict the crop and pH value accordingly.

VI. CONCLUSION

This project helps the farmer to know the suitable crops for his field that can be grown based on a particular season and nutrients present in the soil of that particular land/field. This project connects the farmers to the global technology, gain the knowledge of latest technology in world and intern increases literacy rate by bringing awareness for education. This project taken up for helping the farmers of the India. At present situation technology can be helpful in making agriculture more accurate and easier to get higher yields intern higher rate of profit, so that youngsters' turn-up towards agriculture rather than doing jobs in govt. sector or IT sector. So, we conclude saying that, India is an agriculture-based country, most of the land in India is used for agriculture hence we call farmers as the backbone of our country and to keep the country developing in any sector/field, first the farmer, the backbone of the country should be happy and strong then only country is able to develop.

REFERENCES

[1] Shabnam Shaikh, Gargi Dhuri, Harshada Mhetre,Prajakta Raut,Prof. Amol Dhakne, "Soil pH Value Prediction and Crop Recommendation", International Journal of Research and Analytical Reviews (IJRAR), December 2018, Vol. 5 Issue 4 , pp. 670-673.

[2] Vrushali C. Waikar, Sheetal Y. Thorat, Ashlesha A. Ghute, Priya P. Rajput, Mahesh S. Shinde, "Crop Prediction based on Soil Classification using Machine Learning with Classifier Ensembling", International Research Journal of Engineering and Technology (IRJET), May 2020,Vol. 07 Issue: 05, pp. 4857-4861.

[3] Shima Ramesh Maniyath, Akshatha K N, Architha L, "S,Soil Color Detection Using Knn", Classifier International Conference on Design Innovations for 3Cs Compute Communicate Control,2018,pp. 52-55.

[4] K. Srunitha; S. Padmavathi, "Performance of SVM classifier for image based soil classification", International Conference on signal processing, communication, power and embedded system(SCOPES),2016, pp. 411-415.

[5] B.S.Anami , Sunanda. D. Biradar, D.G. Savakar , P.V. Kulkarni, "Identification and classification of similar looking food grains", International Conference on Communication and Electronics System Design, 201, Vol. 8760, pp.1-8.

[6] Aakunuri Manjula, G. Narsimha, "An Effective Soil Classification and Prediction of Crop Yield Using Spatial Big Data", International Journal of Innovative Technology and Exploring Engineering (IJITEE), Vol.-8 Issue-7, May, 2019, pp. 2263-227.

[7] Roushan Kumar, Rakesh Kumar Dutta, Kamlesh Dutta, "Mobile App using ASTM System of Soil Classification",CPUH-Research Journal: 2015, 1(2),pp. 2455-6076.

[8] Shakil Ahamed, Navid Tanzeem Mahmood, Nazmul Hossain, Mohammad Tanzir Kabir, Kallal Das, Faridur Rahman, Rashedur M Rahman, "Applying Data Mining Techniques to Predict Annual Yield of Major Crops and Recommend Planting Different Crops in Different Districts in Bangladesh A.T.M" IEEE, 1 June 2015, pp. 1-7.

[9] Lakshmi Patil Dr. K. Saraswathi, "Crop Yield Prediction on the Basis of Soil Composition using Machine Learning Algorithms", wutan hustan jisuen jishu journal, pp.201-205.

[10] Prema T. Akkasaligar, Sunanda Biradar, and Rohini Pujari, "Internet of Things Based Smart Secure Home System", Lecture Note on Data Engineering and Communication Technologies, 38, 2020,, pp. 348-355,.

[11] E. Manjula S. Djodiltachoumy, "A Model for Prediction of Crop Yield", International Journal of Computational Intelligence and Informatics, March 2017m Vol. 6: No. 4, pp. 298-305.

[12] Madhuri Shripathi Rao, Arushi Singh, N.V. Subba Reddy and Dinesh U, "Crop prediction using machine learning", AICECS-2021 pp. 1-11.

[13] Nischitha K, Dhanush Vishwakarma, Ashwini, Mahendra N, Manjuraju M.R, "Crop Prediction using Machine Learning Approaches",International Journal of Engineering Research & Technology (IJERT),Vol. 9 Issue 08, August-2020,pp. 23-26.

[14] Hiremath P. S., Prema T. Akkasaligar and Badiger SR, "An Optimal Wavelet Filter for Despeckling Echocardiographic Images ", 2008, Proceedings Int. Conf. Computational Intelligence and Multimedia Applications, ICCIMA – 2007,3,4426375,pp.245-249.