*2022 IEEE 7th International conference for Convergence in Technology (I2CT)*
*Pune, India. Apr 07-09, 2022*

# Healthy Harvest: Crop Prediction And Disease Detection System

Sambhav Bhansali
*Department of Computer Engineering*
*Universal College of Engineering*
Mumbai, India
sambhavbhansali789@gmail.com

Punit Shah
*Department of Computer Engineering*
*Universal College of Engineering*
Mumbai, India
punitshah1111@gmail.com

Jinay Shah
*Department of Computer Engineering*
*Universal College of Engineering*
Mumbai, India
jinays85@gmail.com

Priyal Vyas
*Department of Computer Engineering*
*Universal College of Engineering*
Mumbai, India
priyal84vyas@gmail.com

Prof. Poonam Thakre
*Department of Computer Engineering*
*Universal College of Engineering*
Mumbai, India
poonam.thakre@universal.edu.in

2022 IEEE 7th International conference for Convergence in Technology (I2CT) | 978-1-6654-2168-3/22/$31.00 ©2022 IEEE | DOI: 10.1109/I2CT54291.2022.9825446

***Abstract*— Economy of India highly depends on agriculture. Still traditional ways of recommendations are used for agriculture. Currently, farmers use traditional ways of approximations for amount of fertilizer used and the type of crop to be sown. Agriculture extremely depends on the type of soil and climatic condition of the region. Therefore, it becomes vital to create advancement in this field. With the help of Machine Learning and Deep Learning Techniques we will create a Web-App which will be one-stop solutions for information regarding the agriculture. Crop and fertilizer recommendation system will help the farmers in increasing their yield production. We are going to take the soil parameters along with the weathers API to figure out the most suitable crop for that region. Using the decision tree and navies bayes algorithm we will make the recommendation model which will use the N-K-P, Ph. value and rainfall as the parameters for training. Basis on the crop and region of farming we will recommend the fertilizer and its uses to boost the yield productivity for farmers. Sometimes due to unwanted excess of rainfall or the pest attack can cause disease to crops. We will use the image classification technique where the user can upload the picture of the affected plant/crop and the system will figure out the type of disease which will be done using Support Vector Machine (SVM) or using the neural network techniques. And this disease detection will suggest that how that plant/crop can be cure or prevent. The aim is to make a common system for all the features and provide the results with the best accuracy for all the crops over most of the regions all over the India. Also, the price and news section will keep the farmers updated with daily market prices and government schemes and policies related to the agriculture and farming.**

***Keywords*— *Agriculture, Crop, Fertilizer, Machine Learning algorithms, Farming, Ensemble Modelling.***

## I. INTRODUCTION

Agriculture is one of the major sectors that influences a country's economic growth. Due to lack of knowledge about their soil data, Farmers don't know what type of crop to sow, which type of fertilizer is needed in order to effectively increase their crop production(yield) and gain profit. A deeply disturbing fact according to the National Crime Records Bureau (NCRB) data, 10,281 farmers committed suicide in 2019 for reasons such as debt, crop failure, environment, low produce prices, poor irrigation, increased cost of cultivation and use of chemical fertilizers. With many advancements in Machine learning and Deep learning field, we can effectively implement a Recommendation system to help farmers of India to successfully maximize their yield and turn farming into a profitable business.

Till date many farmers use the old methods of farming which leads to the less production of the yield and makes thing time consuming. Since the technology is growing at great speed it is necessary to implement the new methods in this sector. Using the Machine learning and Deep Learning techniques we can make the models and prediction system which can suggest the types of crops which can be profitable to grow in that particular region of the country. Since many small farmers are already under the debt and because of that they cannot afford the modernized machine. So, using Machine learning Algorithm they can at least increase their yield production to some extend and sometimes even the lack of communication between the government and farmers, the farmers remain unaware about the new farming technique. Moreover, the farmers can feed the system with their soil parameters and select the location on basis of that system will suggest them the best crop to grown for good production rate and profitability. Also, sometimes excess use of fertilizer can cause the soil to loosen its nutrients values. So, system will give the information on how much quantity and which quality fertilizer to use for their soil. So, the idea is that for the different crop system will give the accuracy by taking the inputs of soil parameter and location, accordingly the crop which has greater accuracy can be grown. Furthermore, Plant disease is also affecting the farmers so using the image classification to detect the type of disease and its prevention or cure can be suggested to the farmer. By keeping the Price section, the daily market price can be fetched using the API available at the data.gov.in, so that the farmers will be aware of the prices and will not sell them at lower price which can cause the losses and the news section will update the farmers with all agriculture schemes and benefit policies announced by government so they can be aware and act accordingly for the betterment.

## II. LITERATURE REVIEW

In the paper [1] presented by Mythili K, Rangaraj R they have proposed the system which uses the Ant Colony Optimization (ACO) for optimizing the Deep convolution Neural Networks (DCNN) and Long Short-Term Memory (LSTM) network for crop prediction. On comparison with other algorithm, it gives greater accuracy of 95% also its can be improved in future by implementing auto encoder.

In the paper [2] presented by Pradeepa Bandara et al. they have successfully implemented a crop system in a selected small land. Data (Temp, Humidity, pH, Sunlight, Moisture) are collected at one hour interval from four sensors across land. Naive Bayes and SVM algorithms are used, with overall accuracy of 92% which can be increased up to 95% overtime with self-training. Feedback Model is also used to increase accuracy.

In the paper [3] presented by Aishwarya Bhosale et al. they have made the system using the Iot which will detect the Ph value and moisture of soil and on that basis, it will predict the best crop to grow and by using the Iot farmers don't need to visit lab for their soil testing and everything can be just control using mobile phone.

In the paper [4] presented by Devdatta A. Bondre et al. they have developed a system for prediction of higher crop yield and Fertilizer recommendation by proper implementation of machine Learning algorithms. They have used only Two Machine learning algorithms SVM and Random Forest. SVM has 86.35% accuracy compared with Random Forest algorithm which has 99.47% accuracy against SVM. Further Addition such as crop disease detection using image processing, implementation of smart irrigation is proposed.

In the paper [5] presented by D. Anantha Reddy et al. they have purposed the Ensemble model which is the combine of two or more model for greater accuracy. It will take the soil specific attributes and will recommend the crop. It uses four machine learners namely Naive bayes, K-Nearest neighbour and CHAID and Random tree. But the model is limited to Ramtek region only and the future scope is to also implement yield prediction.

In the paper [6] presented by Shima Ramesh et al. they have proposed the system of Plant Disease Detection which mainly uses the three-feature description namely Hu moments, Haralick texture and Colour Histogram. The algorithm used is random forest (70.14) which more accurate than the other machine learning algorithm. Model convert the RGB image into HSV and grayscale before computing values. Model accuracy can be increased by using global features along with local features and large number of datasets.

In the paper [7] presented by Zeel Doshi et al. they have designed the system which takes the input from the user and will recommend the most suitable crop. They have divided it into two sub-system, which are Crop Suitability Predictor and Rainfall Predictor. They used four machine learning algorithm Decision Tree (90.20), KNN (89.78), Random Forest (90.43), Neural Network (91.00) out of which clearly neural network has the highest accuracy.

In the paper [8] presented by Nidhi H Kulkarni et al. they have developed a crop recommendation system using ensemble technique with respect to only four crops. Individual learners used in this model are Random Forest, Naïve Bayes and Linear SVM. Average accuracy of crop classification into Rabi and Kharif crops is 99.91% using Majority Voting technique.

In the paper [9] presented by H D Aparna et al. they have attempted to predict the consumption of fertilizer to increase crop yield using machine learning models. They have applied three machine learning algorithm Multiple Regression (12.1461), Multi-Layer Perceptron (7.6281) and SMOreg (12.8027) on existing data out of which Multi-Layer Perceptron has lowest RMSE making it closer to actual consumption. Future Scope is to add more data variables such as temperature, rainfall, etc. and see how model reacts.

In the paper [10] presented by S. Pudumalar et al. they have proposed a crop recommendation system using ensemble technique and Majority Voting Technique. Learning Model uses CHAID, Random Tree, KNN and Naive Bayes Machine Learning algorithms together to account for prediction accuracy of 88% for the whole system.

After reviewing the literature survey, we can say that more features can be added into the previous models and can provide more functionality to the system. Moreover, the accuracy can be improvised by feeding large data and using more global features. Crop and fertilizer recommendation along with disease detection can be implemented in one system for easy accessibility. Using of Neural network give the better accuracy over other algorithms and the system can be made to provide solution nationally rather than a particular region.

## III. Proposed Approach

### A. System Architecture

In Fig. 1, the farmer will have to first login/register and the details will be stored in database. Then the farmer can access the main modules, in crop prediction the farmer's input will be pre-processed and given into the model and the predicted crop will be given output. Now in fertilizer recommendation the input data will be compared with the dataset at the backend and on that basis the fertilizer that need to be added in the soil will be shown. In disease detection the farmer will upload the picture and the image will be feed into the trained CNN based image classifier MobilenetV2 which will then predict the name of disease and cure for the same will be displayed. Farmer profile section where they can check their basic information and change password. Lastly the price and news section to check latest information in agriculture sector.

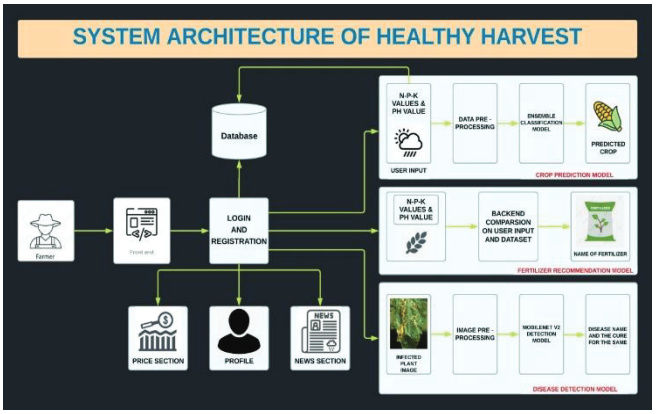

Fig. 1. System architecture of Healthy harvest.

### B. Crop Prediction

As there are multiple algorithms available for supervised learning such as Support vector machines (SVM), logistics regression, random forests, naïve bayes, decision trees and k-nearest neighbour (KNN). We decided to implement ensemble modelling technique, refer Fig. 2, in which the multiple classification algorithms are combined together

which individually predict output and then using voting classifier the majority class out of all predicted classes is selected. The predictions for each label are summed and the label with the majority vote is predicted. Algorithms used for ensemble voting are as follows:

**Logistic Regression:** Logistic regression is one of the most popular Machine Learning algorithms, which comes under the Supervised Learning technique. It is used for predicting the categorical dependent variable using a given set of independent variables. Logistic regression predicts the output of a categorical dependent variable. Therefore, the outcome must be a categorical or discrete value. It can be either Yes or No, 0 or 1, true or False, etc. but instead of giving the exact value as 0 and 1, it gives the probabilistic values which lie between 0 and 1. In Logistic regression, instead of fitting a regression line, we fit an "S" shaped logistic function, which predicts two maximum values (0 or 1).

**Random Forest:** Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML. It is based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model. Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset.

**Gaussian Naive Bayes:** Gaussian Naive Bayes is a variant of Naive Bayes that follows Gaussian normal distribution and supports continuous data. Naive Bayes are a group of supervised machine learning classification algorithms based on the Bayes theorem. It is a simple classification technique, but has high functionality. They find use when the dimensionality of the inputs is high. Complex classification problems can also be implemented by using Naive Bayes Classifier. Gaussian Naive Bayes supports continuous valued features and models each as conforming to a Gaussian (normal) distribution.

**Voting Classifier:** A Voting Classifier, refer Fig. 2, is a machine learning model that trains on an ensemble of numerous models and predicts an output (class) based on their highest probability of chosen class as the output. It simply aggregates the findings of each classifier passed into Voting Classifier and predicts the output class based on the highest majority of voting. The idea is instead of creating separate dedicated models and finding the accuracy for each them, we create a single model which trains by these models and predicts output based on their combined majority of voting for each output class.

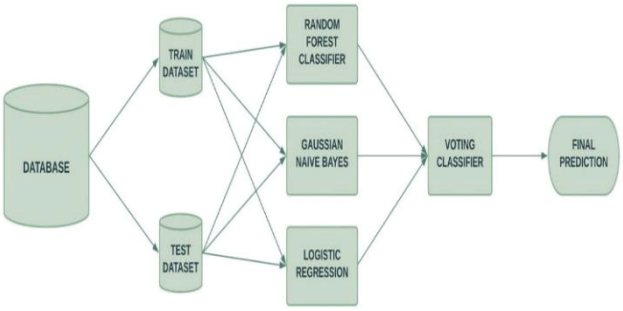


Fig. 2. Ensemble modelling technique for Crop prediction

In Fig. 2, for using Ensemble modelling technique, we first have to split crop dataset into 80:20 ratio for training and testing respectively, which is feed into Random Forest, Gaussian Naïve Bayes and Logistic Regression models for training. 20% test data is then feed into model for getting metric scores like accuracy, precision, recall, f1-score and support. In Fig. 3, we can see metric scores for Ensemble modelling technique, with 94% precision, 100% recall, 97% f1-score and 12% support. Accuracy of 98% is achieved using this modelling technique as its combines multiple algorithms and votes for the best probable result.

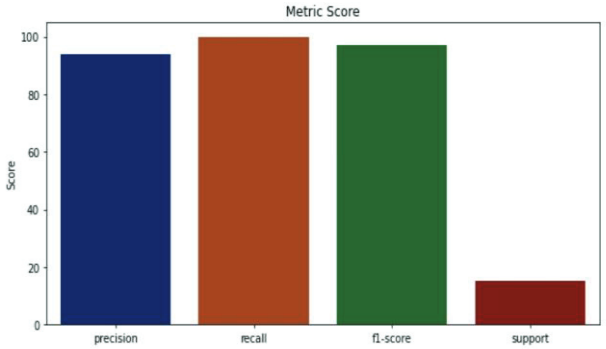


Fig. 3. Metric Scores for Ensemble modelling technique

## C. *Disease Detection*

In deep learning, a convolutional neural network (CNN, or ConvNet) is a class of artificial neural network, most commonly applied to analyse visual imagery. They have applications in image and video recognition, recommender systems, image classification, image segmentation, medical image analysis, natural language processing, brain-computer interfaces, and financial time series. CNNs are regularized versions of multilayer perceptron. Multilayer perceptron usually means fully connected networks, that is, each neuron in one layer is connected to all neurons in the next layer.

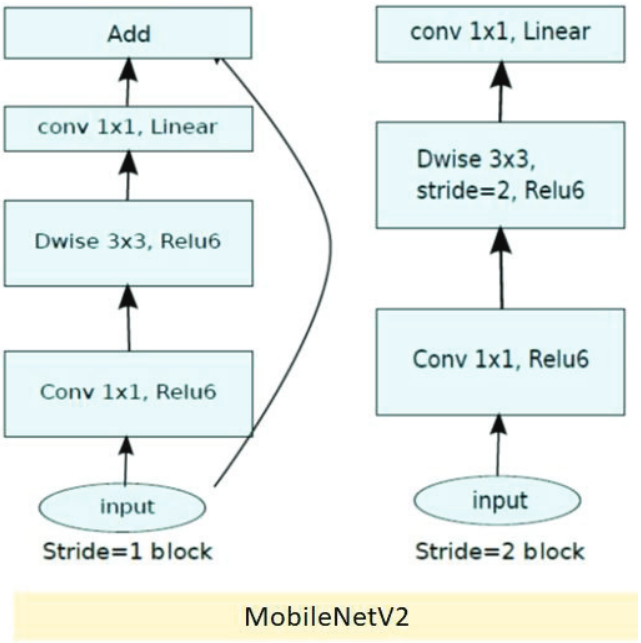


Fig. 4. MobileNetV2 architecture used for disease detection

In Fig. 4, MobileNetV2, a type of convolution neural network used for mobile computing is a family of general-purpose computer vision neural networks designed with

mobile devices in mind to support classification, detection and more. The ability to run deep networks on personal mobile devices improves user experience, offering anytime, anywhere access, with additional benefits for security, privacy, and energy consumption. MobileNetV2 provides a very efficient mobile-oriented model that can be used as a base for many visual recognition tasks. For the recognition of the image, the network takes account four main operations:

- Convolution
- Non-Linearity (ReLU)
- Pooling or Sub Sampling (For Dimensionality Reduction)
- Classification (Fully Connected Layer)

After splitting image dataset in 80:20 ratio for training and testing respectively, we then augment and shuffle train data and feed the same to base MobileNetV2 model for training. After training, we use test data to check for accuracy and loss given by base model, refer Fig. 5, which is 92% and 22% respectively. For fine-tuning the base model, we unlock the top layers of model to be trainable which allocates more trainable parameters to the model improving overall accuracy to 96% followed by some increase in loss.

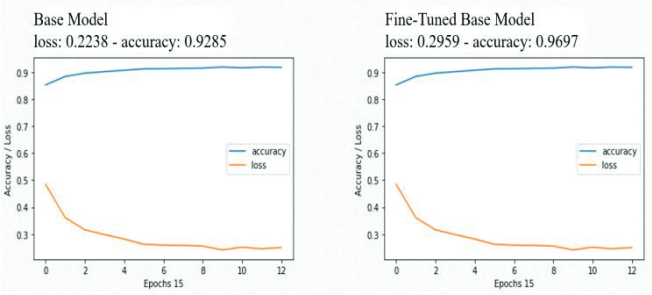


Fig. 5. Accuracy and Loss for MobileNetV2

## IV. Dataset Details

### A. Crop dataset

We custom-built crop dataset consisting 2500 rows of data for 25 types of crops that are majorly grown in India containing of information of 7 attributes such as N-P-K values, pH, rainfall, temperature and humidity of various crops across India, we used open-source data for the dataset from agricultural websites such as:

- http://www.indiaagronet.com
- https://vikaspedia.in
- https://www.apnikheti.com

### B. Disease detection dataset

Disease Detection dataset consists of around 72000 diseased and healthy crop leaf images which is classified into 49 different types of diseased and healthy categories. This dataset is been created by combining various plant and crop diseased leaf datasets on Kaggle for which links are as followed:

- https://www.kaggle.com/seroshkarim/cotton-leaf-disease-dataset
- https://www.kaggle.com/atharvaingle/crop-recommendation-dataset
- https://www.kaggle.com/amarajaved/ricediseasesimagedataset
- https://www.kaggle.com/olyadgetch/wheat-leaf-dataset
- https://www.kaggle.com/atharvaingle/crop-recommendation-dataset

## V. Results

After splitting dataset in 80:20 ratio for training and testing, fitting data into respective machine and deep learning models. Ensemble modelling technique for crop prediction has achieved an overall accuracy of 98% (This accuracy was achieved by using theoretical custom crop dataset, real world data may vary the accuracy). Fine-tuned MobileNetV2 CNN for disease detection has achieved an accuracy of 96%.

| SUBSYSTEM | HEALTHY HARVEST | | EXISITING SYSTEM | |
|---|---|---|---|---|
| | ACCURACY | MODEL NAME | ACCURACY | MODEL NAME |
| CROP PREDICITON | 98% | ENSEMBLE MODEL | 97% | RANDOM FOREST |
| DISEASE DETECTION | 96% | MOBILENETV2 | 70% | SUPPORT VECTOR MACHINE |

Fig. 6. Accuracy comparison with existing system

Referring to Fig. 6, we can see a significant increase in accuracy from SVM in disease detection model (MobileNetV2) from 70% to 96% which is achieved using latest deep learning CNN's and larger dataset. Furthermore, accuracy improvement of 1% is seen on crop prediction model using Ensemble modelling technique over existing systems.

Using Flask, a micro web python framework, to deploy Healthy Harvest where Farmer can input crop data as given in Fig. 7 and in Fig. 8, Result page shows prediction given by model is displayed to farmer along with its climatic conditions.

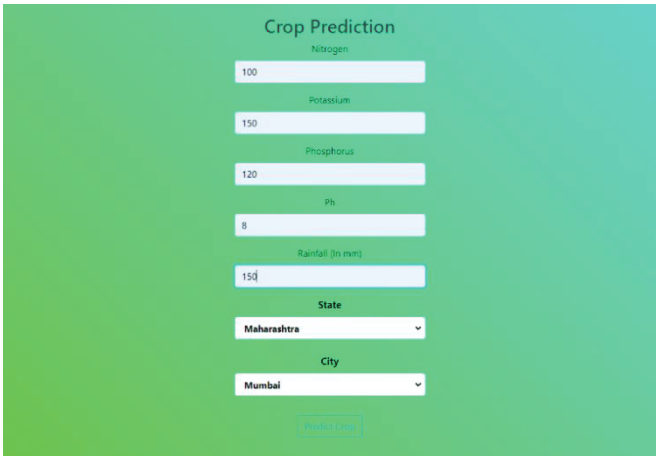


Fig. 7. Soil data input page

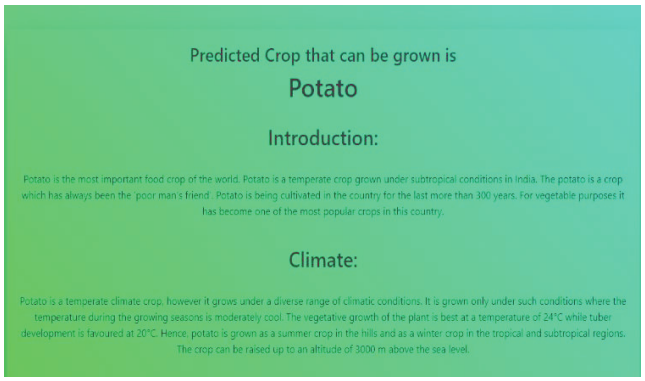


Fig. 8. Crop prediction result

In Fig. 9, Fertilizer recommendation result page displays fertilizer suggestion for given crop data and also its usage.

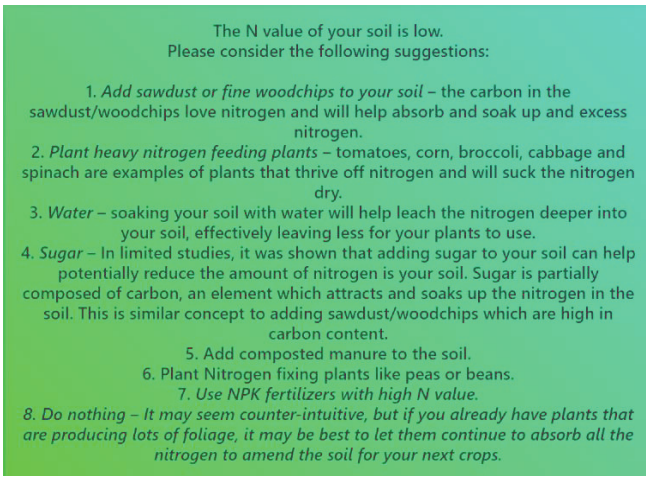


Fig. 9. Fertilizer recommendation result

In Fig. 10, shows disease detection output page where it displays disease of infected crop uploaded by user and also its prevention and cure.

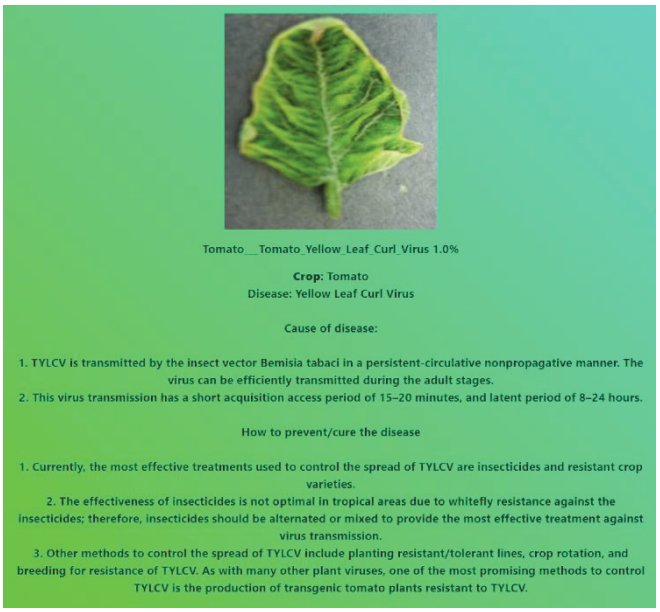


Fig. 10. Disease detection result

## VI. Conclusion and Future Work

Using this Crop and fertilizer recommendation system, the work of farmers will be simplified they will no longer need to depend on other sources, just input the require parameters into the system and a suitable crop/fertilizer would be recommended for the user. Farmers will get better yield production which will in return improve their financial condition. Increased production can also solve the hunger problem of some countries. Sometime due to unforeseen changes in nature causes the crop and the plants to get infected, to mitigate that disease detection can detect the disease uploaded by user and also recommend it's cure.

There is more work that can be done using the Internet of things (IOT) by building the hardware which contains the Ph., rainfall, moistures sensors. This can be installed in the farmlands so the farmers can get the real-time data about the soil right on their mobile and that data can be given to recommendation system. Overtime models can be re-trained using real-time data and improve overall accuracy of prediction system.

## References

[1] Mythili K, Rangaraj R (2021), "Crop Recommendation for Better Crop Yield for Precision Agriculture Using Ant Colony Optimization with Deep Learning Method", SCOPUS, April 2021.

[2] Pradeepa Bandara, Thilini Weerasooriya, Ruchirawya T.H., W.J.M. Nanayakkara, Dimantha M.A.C, Pabasara M.G.P (2020), "Crop Recommendation System", IJAC, October 2020.

[3] Aishwarya Bhosale, Nikita Asode, Mayur Ahuja, Rutank Thanekar, Indira Joshi (2020), "Soil Based Fertilizer Recommendation System using IoT", IRJET, April 2020.

[4] Devdatta A. Bondre, Mr. Santosh Mahagaonkar (2019), "Prediction of Crop yield and Fertilizer Recommendation using Machine Learning Algorithms.", IJEAST, September 2019.

[5] D. Anantha Reddy, Bhagyashri Dadore, Aarti Watekar (2019), "Crop Recommendation System to Maximize Crop Yield in Ramtek region using Machine Learning", Research-Gate, February 2019.

[6] Shima Ramesh, Mr. Ramachandra Hebbar, Niveditha M, Pooja R, Prasad Bhat N, Shashank N, Mr. P V Vinod (2018), "Plant Disease Detection Using Machine Learning", IEEE, August 2018.

[7] Zeel Doshi, Rashi Agrawal, Prof. Neepa Shah, Subhash Nadkarni (2019), "AgroConsultant: Intelligent Crop Recommendation System Using Machine Learning Algorithms", IEEE, April 2019.

[8] Nidhi H Kulkarni, Dr. G N Srinivasan, Dr. B M Sagar, Dr. N K Cauvery (2018), "Improving Crop Productivity Through A Crop Recommendation System Using Ensembling Technique ", IEEE, July 2018.

[9] H D Aparna, Dr. Kavitha K S, Dr. Kavitha C (2017), "Use of Machine Learning Techniques to Help in Predicting Fertilizer Usage in Agriculture Production", IRJET, May 2017.

[10] S. Pudumalar, E. Ramanujam, R.Harine Rajashree, C.Kavya, T.Kiruthika, J.Nisha (2016), "Crop Recommendation System for Precision Agriculture", IEEE, June 2016.