



IJIRCCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 9, Issue 10, October 2021

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 7.542



9940 572 462



6381 907 438



ijircce@gmail.com



www.ijircce.com

AI-based Farmer's on Field Plant/Crop Consultant

Sourabh Kokitkar, Parth Kajbaje, Prerana Jadhav, Pratiksha Shete, Prof. A.R.Kambale

Department of Computer Engineering, Sinhgad Institute of Technology and Science Narhe, Pune, India

ABSTRACT: As the farming fields are quite large and have very large number of crops in it, hence it becomes very difficult for the human eye to properly detect and classify each and every plant. And doing so is very important. This model will help the farmers to correctly detect and classify the plants and also detect if they're healthy. It can classify between the diseases by scanning the leaf and alert the farmers about the disease before it starts spreading. It will guide farmer by suggesting pesticides to cure the disease and also it predicts crop yield. In this model, different machine learning and deep learning algorithms like KNN, SVM, CNN, Decision tree and Random forest are trained and compared on the basis of accuracy and the algorithm that performs best in training as well as testing is taken in account.

KEYWORDS: Disease detection, CNN (Convolutional Neural Network), KNN (K-Nearest Neighbor), SVM (Support Vector Machine), Decision tree, Random forest

I. INTRODUCTION

This section illustrates brief introduction about the domain and proposed methodology of the proposed work. In recent years, Artificial Intelligence and Image processing which deals with extracting useful information from images plays a unique role in advancement of technologies. As the cultivational fields are quite large and have very large number of plants in that, hence it becomes very difficult for the human eye to properly detect and classify each and every plant. And doing so is very important as even single infected plant can spread the disease. Also, most of the farmers do not have proper knowledge of those diseases and actual cure for that disease. Hiring experts may cost them heavily and use of pesticides without knowledge will harm the land. Hence in order to solve this problem we have developed the Artificial Intelligence based solution. Our system will automatically detect disease on the plant from image of a leaf and after processing that image it will suggest preventive measures to control the spread. It will also predict crop yield. The most widely used method for plant disease detection is simply naked eye observation by experts through which identification and detection of plant diseases are done. For doing so, a large team of experts as well as continuous monitoring of experts is required, which costs very high when farms are large. At the same time, in some countries, farmers don't have proper facilities or even idea that they can contact to experts. Due to which consulting experts even cost high as well as time-consuming too. In such a condition, the suggested technique proves to be beneficial in monitoring large fields of crops. And automatic detection of the diseases by just seeing the symptoms on the plant leaves makes it easier as well as cheaper.

Plant disease identification by the visual way is a more laborious task and at the same time less accurate and can be done only in limited areas. Whereas if automatic detection technique is used it will take fewer efforts, less time and more accurately.

The purpose of this project is to obtain an image from the farmer of the diseased crop preferably the leaves through the Android Application installed on farmer's phone. The images is then processed using image-processing technique and detect type of plant/crop using SVM classifier the disease type is detected using Machine learning Approach. The diseases affected to the crop and the amount of fertilizer or the pesticide/insecticide to be used is updated to the Android Application that was previously used by the farmer to upload image. Also, in addition to this it predicts crop yield.

II. RELATED WORK

Study on Classification of plant images using Combined Classifier was carried out by ShubhamkarGhosh et al. [1] in 2015. They discussed about different classification techniques such as KNN, PNN, and SVM. The most important advantage of PNN is that training is easy and instantaneous. SVM was found competitive with the best available

machine learning algorithms in classifying high-dimensional data sets. In SVM computational complexity is reduced to quadratic optimization problem and it is easy to control complexity of decision rule and frequency of errors.

A Methodology of Plant Classification using different Machine Learning Techniques was proposed by K. B. Shobana et al. [2] in 2020. They performed two distinct plans for image classification. With an appropriate research operation, they inspect that SVM classification better than ANN classification. The performance of ANN was very low compared to SVM, this was due to the fact that the ANN classifier could not distinguish plant where as SVM was very effective in classifying plant. The accuracy percentage for SVM was found to be 98% and for ANN it was 92%.

A System for Plant Classification based on its leaf features was proposed by EsraaElhariri et al. [3] in 2014. Comparison between Linear Discriminant Analysis (LDA) and Random Forests (RF) classifiers for the problem of plant classification has been presented. It has three main stages; pre-processing, feature extraction and plant classification. Classification approach was implemented by applying background removal, and extracting color components for each image. Then, feature extraction was applied to each pre-processed image, HSV color moments, shape, first order texture, GLCM and vein features are obtained as features vector.

A methodology regarding Classification of Cotton Leaf spot diseases using image processing edge detection techniques was proposed by P Revathi et al. [4] in 2012. They investigated advance computing technology to assists the farmer in plant development process. This approach is used to capture the infected cotton leaf images. RGB color feature segmentation is carried out to get disease spots. Edge detection technique is used for extraction of image features of spots to detect diseases. Neural network issued to categorize the diseases. The segmentation process is not suitable for Monocot family plant.

S. Dubey et al. [5] in 2012 proposed a system to detect infected fruit part using K-means Clustering Segmentation Technique. They explored the concept of detection and classification of apple fruit diseases. The proposed approach is composed of three steps such as segmentation, feature extraction and classification. K-means clustering technique is used for the image segmentation. The features are extracted from the segmented image and images are classified based on a Multiclass Support Vector Machine (SVM). The proposed approach is specific to apple fruit diseases and cannot be extended to other fruit diseases.

Yan, Han and Ming [6] in 2007 proposed a methodology regarding Feature selection of Cotton disease leaves images based on Fuzzy Selection Techniques. Authors select features of cotton disease leaf image by introducing fuzzy selection approach, fuzzy curves and fuzzy surfaces. The features which are extracted from fuzzy selection approach are used for diagnosing and identifying diseases. This approach removes the dependent features of image so as to reduce the number of features for classification.

An approach regarding Diagnosis and grape leaf diseases using neural networks was proposed by SannakkiRajpurohit et al. [7] in 2013. They presented an approach to diagnose the disease using image processing and artificial intelligence techniques on images of grape plant leaf. The input image of grape leaf is complex at background. The threshold is used to mask green pixels and image is processed to remove noise using anisotropic diffusion. Then segmentation is done using K-means clustering technique. The diseased portion from a segmented image is identified. The results were classified using back propagation neural network.

In 2018, Vinay Kumar et al. [8] Studied different digital image processing techniques for leaf disease detection and classification of the detected disease into its subtype. It has limited the impact of plant diseases on agricultural production using image processing techniques. The correlation between the disease's symptoms and its impacts on yield. The large number of agriculture and horticulture applications based on the detection and classification of plant leaf diseases make it difficult for someone to prospect all important concept present in the literature. This is a reason of missing potential solutions for problematic issues.

KirttiRanjanPaltasingh in 2018 [9] proposed Statistical Modelling of Crop-Weather Relationship in India i.e., a survey on evolutionary trend of methodologies. Weather factors, like other inputs such as land, labor, seeds, irrigation, fertilizer, and pesticides are also direct inputs in crop production. In a state of agriculture where the adoption and diffusion of modern technologies is very low or almost nil, weather factors count more than other inputs because of their direct and indirect effects. Thus, the link between weather and crop yield will have implications on food supply and crop forecasting and management policies. It is of immense importance to the policymakers, agricultural scientists, agricultural economists, and meteorologists to understand this relationship. The methodology for studying this relationship has undergone many improvements over time. This paper attempts to review the studies in this area done in India.

In 2018, D. S. Zingade et al. [10] proposed methodology for Machine learning based prediction of crop using Multi-Linear Regression approach. They listed out all possible crops feasible in a particular area, helping the farmer in decision making such as which crop to cultivate. A careful examination of the data related to soil, weather, pH and past year production has been done by the system and suggests which are the most profitable crops which can be cultivated in the apropos environmental condition. Also, this system examines the past production of data which will help the farmer get insight into the demand and the cost of various crops in market. As maximum types of crops will be covered under this system, farmer may get to know about the crop which may never have been cultivated.

III. MOTIVATION

Farmers are our “Annadata” and we need to revert them back through any way possible. Many a times there’s lack of knowledge about the crop/plant, their classification and about their health. Hiring expert is costly. Not dealing with it properly and using pesticides without knowledge may harm the land. Farmers’ suicide is very sensitive topic which needs to be addressed on priority basis. (More than 10 suicides daily according to figures in 2017 and 2018).

IV. PROBLEM STATEMENT AND OBJECTIVES

A. Problem Statement

To develop AI based system which will detect and classify a plant and disease on it by scanning its leaf also guide a farmer by suggesting pesticides to cure the disease and predicts crop yield.

B. Objectives

- To detect and classification of a plant from the picture of a leaf.
- To detect the diseases and classification according to its type.
- To suggest the pesticides for the detected disease
- To predict crop yield based on historical data.

V. DESIGN AND SYSTEM ARCHITECTURE

System architecture is the conceptual model that defines the structure, behavior, and more views of a system. System architecture of our project is System design defines the system architecture. It also describes the modules and interfaces.

This diagram explains the architecture of our system. The system architecture provides an insight of how the flow of process will be. Entire process of how the system will move forward that will generate the end-result is depicted.

The Fig 1 shows the system architecture of our proposed system.

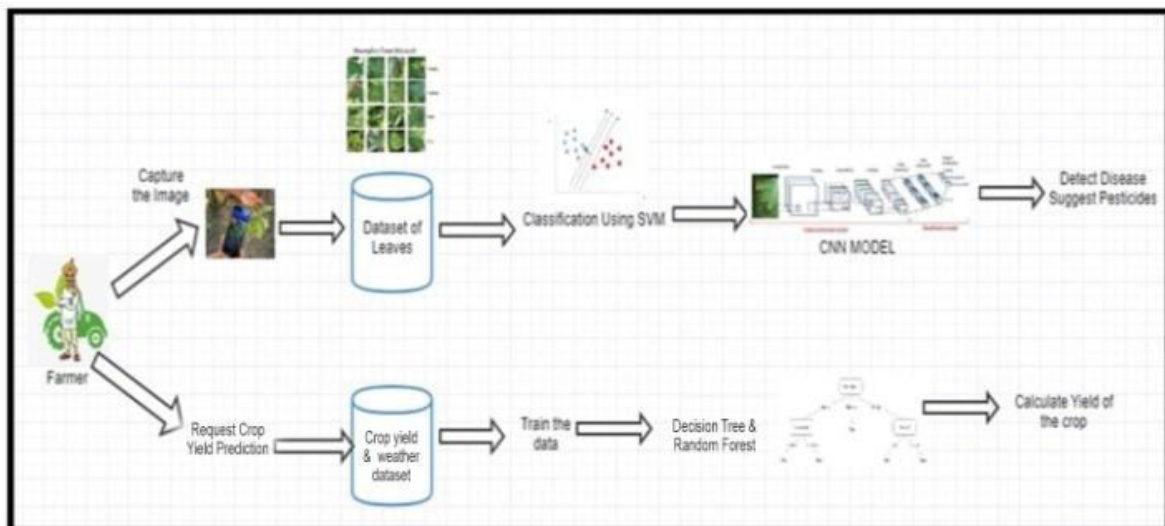


Fig 1: System Architecture

There are mainly two parts in system Architecture: In first part, the farmer will capture the image of leaf of a particular crop to detect the disease. We have a dataset of different leaves of different crops. We will carry out classification with the help of SVM (Support Vector Machine) algorithm. And then after applying CNN (Convolution neural network) model we can detect the disease only by scanning leaf of the crop. And also, we will suggest some pesticides to cure the disease. In second part, we predict current price of crop per qtl display it's prime location along with crop type. We will display the places where it has high demands and can be exported. We will predict price per qtl in every month of the upcoming year and display maximum and minimum price it can touch to With the help of Decision tree algorithm and Random Forest algorithm we will predict the yield of the crop.

VI. IMPLEMENTATON

A. Data Acquisition

For implementation purpose we used Plantvillage Data set IT is public dataset of 54,305 images of diseased and healthy plant leaves collected under controlled conditions PlantVillage Dataset. The images cover 14 species of crops, including: apple, blueberry, cherry, grape, orange, peach, pepper, potato, raspberry, soy, squash, strawberry and tomato. It contains images of 17 basic diseases, 4 bacterial diseases, 2 diseases caused by mold (oomycete), 2 viral diseases and 1 disease caused by a mite. 12 crop species also have healthy leaf images that are not visibly affected by disease.

B. Plant Leaf Detection

The images are first preprocessed and then their shape, color and texture based features are extracted from the processed image.

A dataset was created using the extracted features to train and test the model. The model used was **Support Vector Machine Classifier** and was able to classify with **90.05% accuracy**.

1. Pre-processing

The following steps were followed for pre-processing the image:

- a) Conversion of RGB to Grayscale image

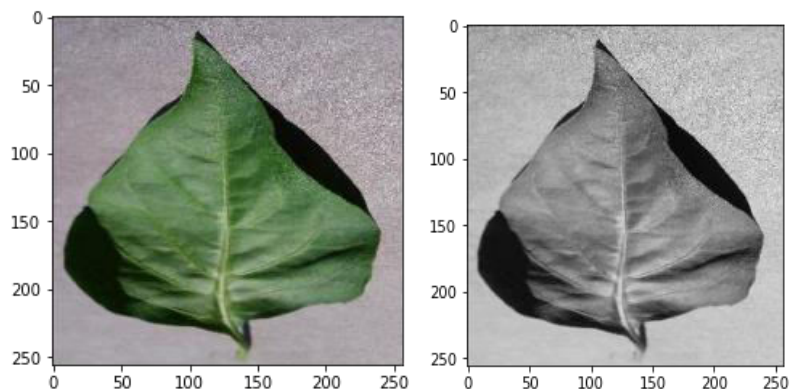


Fig 2: Conversion of RGB to Grayscale image

Fig 2 shows the conversion of RGB to Grayscale image which is very first part of preprocessing method.

- a) Smoothing image using Guassian filter
- b) Adaptive image thresholding using Otsu's thresholding method
- c) Closing of holes using Morphological Transformation
- d) Boundary extraction using contours

2. Feature extraction

Various types of leaf features were extracted from the pre-processed images which are listed as follows:

- a) *Shape based features*: In this, we are calculating physiological length, physiological width, area, perimeter, aspect ratio, rectangularity, circularity so that the design of classifier can be simplified hence resulting into more

accurate classification. Fig 3 shows the result of extraction of shape based features.

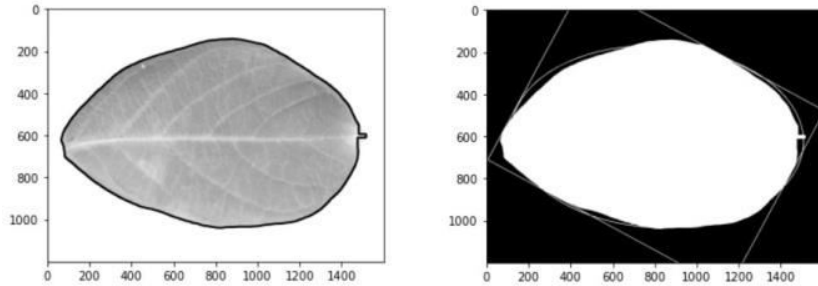


Fig 3: Extraction of Shape Based Features

b) *Color based features*: In this, we calculate mean and standard deviations of R, G and B channels. This method is used to extract the red green and blue components of input image as shown in Fig 4.

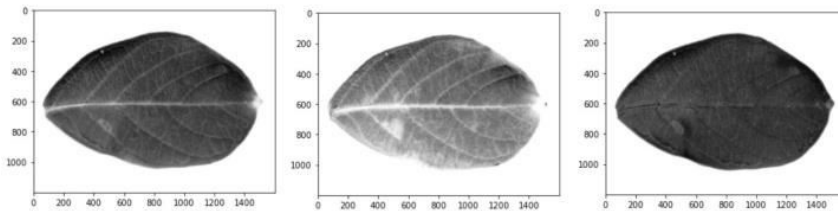


Fig 4: Extraction of Color Based Features

c) *Texture based features*: Basically, an image is a combination of pixels and texture and is defined as an entity having group of mutually related pixels within an image. In this, we are calculating contrast, correlation, inverse difference moments and entropy.

The final output of the process of extracting the texture features, shape feature and the color features is the CSV file as shown in Fig 5.

	area	perimeter	physiological_length	physiological_width	aspect_ratio	rectangularity	circularity	mean_r	mean_g	mean_b	stddev_r	stddev_g	stddev_b
0	197787.0	3479.036035	1416	759	1.865613	5.433846	61.195588	6.395667	13.643413	4.388007	24.025329	40.20091	40.20091
1	101297.0	2491.210238	1191	130	9.161538	1.528476	61.266656	7.049316	9.232018	10.876096	33.816205	37.38221	37.38221
2	88626.5	2291.511754	1096	119	9.210084	1.506590	60.616856	3.434303	6.371511	2.644757	19.975699	29.05731	29.05731
3	190481.0	2858.479352	1319	254	5.192913	1.758842	42.896164	7.670415	13.303599	6.049157	28.822885	40.22181	40.22181
4	228035.0	2920.420478	1325	286	4.632867	1.661806	37.401521	8.992028	16.671173	6.294281	30.967158	45.04011	45.04011

Fig 5 : Final Output after Preprocessing

This includes all the extracted features that are later used as a knowledge base for training the classifier to detect plant species.

3. Model building and testing

Support Vector Machine Classifier is used as the model to classify the plant species and Features are then scaled using StandardScaler Also parameter tuning is done to find the appropriate hyper parameters of the model using GridSearchCV.

Here we use CSV file which was created previously after carrying out various pre-processing techniques for feature extraction.

In the SVM algorithm we plot each data item as a point in n dimensional space (where n is no. of features you have) with the value of each feature being the value of particular co-ordinate. Then we perform classification using the hyper-plane that differentiates the two classes very well.

From the total dataset 70% data is used for training the SVM classifier and 30% is used for testing. We carry out tuning process on the model till we reach to the optimal accuracy.

After the training is completed, the classifier will use 30% of the total samples to examine the accuracy of the system depend on confusion matrix.

Then we use PCA (Principal Component Analysis) for dimensionality reduction to the overfitting model.

Using Mobile capture image.

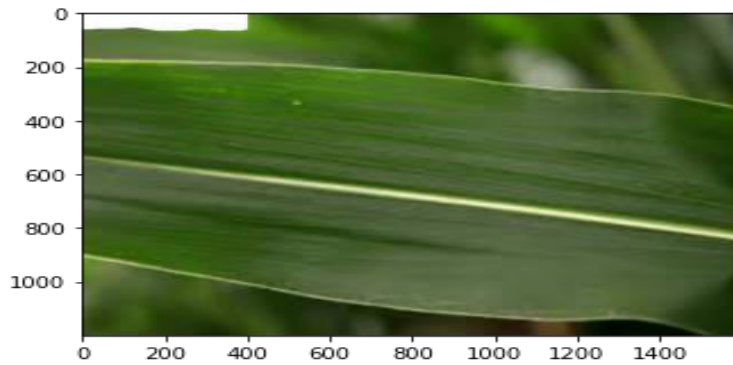


Fig 6: Any random leaf image from dataset

Fig 6 shows the random image of any plant leaf from specified dataset which is nothing but the input on which SVM model can be trained.

```

y_pred_mobile = svm_clf.predict(scaled_features)

[[-2.56384014 -2.77085584 -3.43845503 -2.55299225 -1.50209703 -0.79631745
  0.40523307 -2.71396525 -2.05786019 -3.99695032 -2.13683171 -1.35783867
 -2.33165403 -1.34585814  1.34030546  9.88659057 -5.11797209]]

In [84]: common_names = ['Pepper_Bell', 'Potato', 'Tomato', 'Apple', 'BlueBerry', 'Cherry', 'Corn', 'Grape', 'Peach', 'R
common_names[y_pred_mobile[0]]

Out[84]: 'corn'
    
```

Fig 7: Output after training of SVM model

After applied SVM model on Fig 6 we get the output as shown in Fig 7 which is the plant name of that input image. Here, we successfully detect the plant species from image of that particular leaf.

b) Plant Leaf Disease Detection

Data Preprocessing

First of all we divide dataset into 80% training dataset and 20% testing dataset. The database is pre-processed such as image re-shaping, resizing and conversion to an array form is carried out. Set height, width and color of input image to get the images which are dimensionally similar.

Build the Model

CNN Model Steps

The objectives of the convolutional operations are to extract the high-level features like edges, colours, shape from the input image.

1.Conv2D: Conv2D is a 2D Convolution Layer, used to determines the number of output filters in the convolution. This layer creates a convolution kernel that's wind with layers input which helps produce a tensor of outputs.

2.max_pooling: Pooling layer is used for reducing the spatial size of the Convolved Feature.It is useful for extracting dominant features, thus the output after max-pooling level would be a feature map.

3. Flatten: It is present in between convolutional layer and fully connected layer. Flatten is used for converting the data into a 1-dimensional array. We flatten the output of the convolutional layers to create a single long feature vector.

4. Dense: Dense layer is deeply connected neural network layer. Dense layer receives input from all the neurons present in its previous layer. This layer perform operations like rotation, scaling ,matrix-vector multiplication.

5. Dropout: Dropout may be implemented on all hidden layers, input layer in the network. Dropout is a simple Way to Prevent Neural Networks from Overfitting.

Depending upon the complexities present in input images, the number of layers may be increased to capturing the details from input image.

Train the Model

We train model using validation dataset to validate each step. After 1 epochs, we get 79% for accuracy, we can improve this accuracy more than 98% by using fine-tuning.

After the CNN model is trained successfully, the system can identify the disease if the plant species is contained in the database. After successful training and preprocessing, comparison of the test image and trained model takes place to predict the disease.

Checking Performance

Fig 8 consists of two graphs, one is showing training and validation accuracy and other is showing training and validation loss.

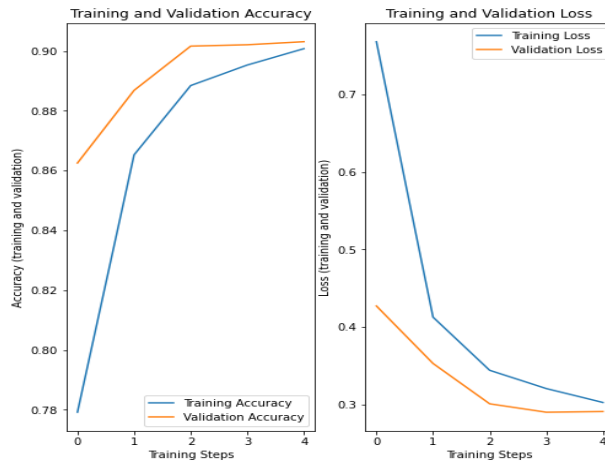


Fig 8: Graphical Representation of Validation Accuracy and Validation Loss

Random Test



Fig 9: Output after training CNN model

In Fig 9 we taken a random image of leaf and trained CNN model successfully so that we can detect the disease of that plant.

Pesticides Suggestions

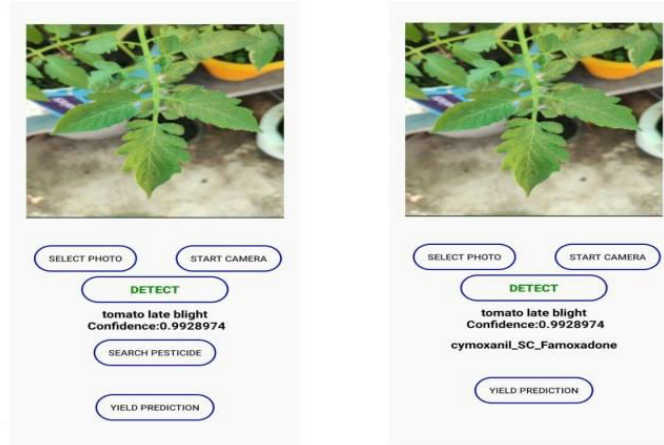


Fig 10: Output of Pesticide Suggestion

After knowing the disease of that particular plant our proposed system will suggest the pesticides to cure that disease as shown in Fig 10.

At the end of this module, we will save this model and we convert it to TensorFlow lite to deploy this model into android app.

c) Crop Yield Prediction

The dataset for this module was taken from the website <https://data.gov.in/>. It contains Annual Rainfall and Wholesale Price Index (WPI) from April 2012 to December 2018 for 23 different crops namely – Arhar, Bajra, Barley, Copra, Cotton, Gram, Groundnut, Jowar, Jute, Maize, Masoor, Moong, Niger, Paddy, Ragi, Rape, Safflower, Sesamum, Soyabean, Sugarcane, Sunflower, Urad and Wheat. Each one contains attributes like, ‘Month’, ‘Year’, ‘rain’, ‘WPI’.

The Dataset is preprocessed. In this module there are two machine learning algorithms used. One of them is Decision tree and Random forest.

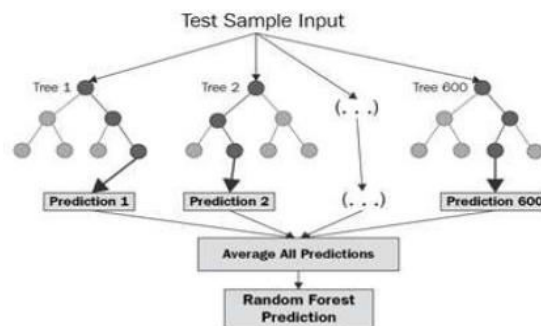


Fig 11: Test Sample Input

The decision tree’s role in model building is to build regression and classification models in the form of a tree structure. It is used to break down a dataset into a smaller subset and simultaneously an associated decision tree is developed. The tree consists of nodes and leaves which represents values for attributes tested. This tree handles categorical and numerical data. ID3 is the core algorithm. It uses a greedy top-down search through the space of possible branches with no backtracking. The other algorithm is Random forest which is a Supervised Learning Algorithm and is a bagging technique which uses Ensemble learning. Ensemble combines the predictions obtained together to make accurate predictions. It will be used as a meta-estimator which aggregates many decision trees, with some helpful modifications. This generated 600 decision trees where we got in total of 600 predictions, one from each tree. Then all the predictions are gathered and an average is calculated as shown in Fig 11. It is then fed into the random forest predictor which eventually gives us an output. This is done by the ensemble so that we will get an exact accuracy.



Fig 12: UI of Crop Prediction Module

A python package index (PPI) is used here called Flask. It is used to deploy the project. An instance for the flask is being created. It acts as a Central configuration object which sets up the pieces of the application required. Flask Framework does not have built in database facilities. It does have a package which helps in connecting the UI interface to a SQL database. The User interface is made with HTML/CSS to display the details of each crop as shown in Fig 12. The increase and decrease in the price of each crop according to the months and a graph depicting the same. The UI is connected to the model with the help of Flask framework.

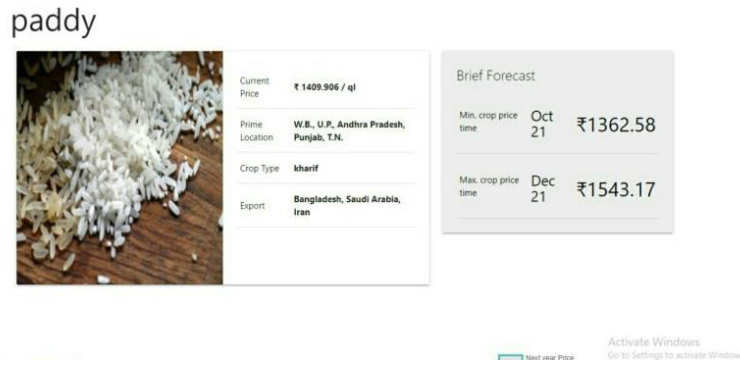


Fig 13: Output of Crop Yield Prediction

After carrying out all the processes we predict current price of crop per qtl display it's prime location along with crop type. We will display the places where it has high demands and can be exported. We will predict price per qtl in every month of the upcoming year and display maximum and minimum price it can touch to as shown in Fig 13 .



Fig 14: Graphical Representation of Forecast Trends

We will display graph where Y axis will have price in rupees per qtl and X axis will have months in respective year as

parameters. There are graphs for previous year and upcoming year showing the price changes as shown in Fig 14.

VII. CONCLUSION

It is important to analyze crops health, identify diseases on them and to use appropriate fertilizers and pesticides for cure. It is bit difficult to keep track of each crop on the large farmlands, hence in order to do that efficiently, an android application is to be developed which includes very accurate artificial intelligence solution for detecting and classifying different plant leaf disease it also considers weather conditions around the crop and then it will suggest line of action which includes appropriate pesticides/fertilizers to be used prediction of crop yield. We are using SVM algorithm for classification of crops/plants, CNN for accurately detecting diseases and their types, decision tree algorithm and random forests for predicting crop yield.

VIII. FUTURE WORK

In the future, work should be focused on considering more number of plants and diseases on them. Also, instead of just suggesting pesticide work should be done to suggest other remedies to cure the disease.

REFERENCES

- [1]. H. Al-Hiary, S. Bani-Ahmad, M. Reyalat, M. Braikand Z. AL Rahamneh, "Fast and Accurate Detection and Classification of Plant Diseases", International Journal of Computer Applications(0975 – 8887) March 2011.
- [2]. Mr. AshishNage, Mr. AshishNage, "Detection and Identification of Plant Leaf Diseases Based onPython", International Journal of Engineering Research & Technology (IJERT), May 2019.
- [3]. ShivanshiShokeen, Archana Singh, "Color Transform Based Approach for Disease Spot Detection on Plant Leaf", International Journal of Computer Science and Telecommunications, Jun 2012.
- [4]. P. Revathi and M. Hemalatha, "Classification of cotton leaf spot diseases using image processingedge detection techniques", International conference on emerging trends in science ,2012
- [5]. Shiv Ram Dubey, PushkarDixit,Nishant Singh, Jay Prakash Gupta," Infected Fruit Part Detectionusing K-Means Clustering Segmentation Technique ", International Journal of Artificial Intelligenceand Interactive Multimedia, Vol. 2, No 2.,2013
- [6]. KirttiRanjanPaltasingh, PhanindraGoyari, "Statistical Modeling of Crop-Weather Relationship in India: A Survey on Evolutionary Trend of Methodologies", Asian Journal of Agriculture andDevelopment, Vol. 15 No. 1, June 2018
- [7]. Yan, Han and Ming, "Features selection of cotton disease leaves image based on fuzzy feature selection techniques",2007
- [8]. SanjeevSannakki, Vijay Rajpurohi, Vijay Nargund and PallaviKulkarni "Diagnosis and classification of grape leaf diseases using neural networks", Conference: 2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT) July 2013.
- [9]. Gittaly, "Study of digital image processing techniques for leaf disease detection andclassification",2018
- [10]. KirttiRanjanPaltasingh, "Statistical Modeling of Crop-Weather Relationship in India: A Survey onEvolutionary Trend of Methodologies",2018.



INNO  **SPACE**
SJIF Scientific Journal Impact Factor
Impact Factor: 7.542



ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

 **9940 572 462**  **6381 907 438**  **ijircce@gmail.com**



www.ijircce.com

Scan to save the contact details