MACHINE LEARNING IN CROP DISEASE DETECTION AND YIELD PREDICTION

Preethi P[1], Vimarsha CR[2]

[1,2]B.E Students, Dept of ISE, Sir MVIT, Bangalore, Karnataka
[3]Guided by ***Raghav S****, Associate Professor, Dept. of ISE, Sir MVIT, Bangalore, Karnataka*

---***---

Abstract: Agriculture plays a major role in the Indian Economy. The Indian agriculture division provides eighteen percent of India's gross domestic product (GDP) approximately. Over a few years there is a drastic reduction in agriculture in India. The main source for these radical changes are climatic change and utilization of low level technology. Predicting the crop yield well ahead of its harvest and monitoring the crop would help the farmers. Machine learning is an essential approach for achieving practical and effective solutions for many agricultural problems. In this paper, we discuss about agriculture problems such as crop disease detection and its yield prediction using machine learning algorithms. Convolution neural network (CNN) algorithm along with the inbuilt image processing function is used to build the model. This model is used in the process of disease detection using TensorFlow. Another model is built using the Random forest Algorithm in order to predict the crop yield.

Keywords: Disease Detection, Yield Prediction, Machine Learning, Convolution neural network(CNN), Random forest Regression, Image processing, User Interface Web Application

I. Introduction

Agriculture is the backbone of India were almost 50 to 60 per cent of the workforce in India is directly dependent on agriculture which influences their livelihood. Agriculture is the significant source of food supply which provides a constant supply of food for a huge amount of the population of our country where more or less 60 percent of household consumption is alone satisfied with agricultural outcome. Farmers can select suitable crops either vegetable or fruit from a wide range of diversity. But there's always a considerable risk factor for the farmers when deciding to grow a particular crop during a particular season, on a particular piece of land. However, the cultivation of these crops for optimum yield and quality produce is highly technical irrespective of the capital put in terms of soil nutrients, water and seed quality, the crop may fail to bring disastrous losses to the farmer and his family, eventually leading to more serious problems like debt and suicide. The main root of this problem is lack of knowledge of the new technologies since they still use the conventional way of agriculture. If the farmer would be able to know the yield of the crop in advance, it will help him to choose which crop he should cultivate. After the selection of the proper crop to be cultivated it is the responsibility of the farmer to monitor the crop that is to control the plant diseases which leads to significant influence on the yield production. The application developed in this research will provide farmers a user friendly interface to predict crop yield and its disease with suitable remedies. All such problems related to agriculture can be best answered by the Machine Learning approach. Machine Learning uses algorithms to analyze the data and learn from it. After this, it is used to determine or predict about something in the world. Our application is developed by building the model using supervised machine learning where this model is trained using datasets of previous crop production and various image datasets of disease present in the particular crop. This trained model is used to predict the yield and the disease for the given input.

II. Methodology

A. Random Forest Regression

Random forest is an algorithm that uses ensemble learning method for regression and classification.It is a Supervised Learning algorithm. A technique that combines the predictions from multiple machine learning algorithms together is an Ensemble method. Ensemble models make more accurate predictions than any individual model. It has Bootstrap and Boosting Aggregation learning types. Random forest is not a boosting technique but a bagging technique. In bagging each model runs independently and then at the end output is aggregated without preference to any model. Bagging model constructs a multitude of decision trees on which it operates on training time. Decision tree classifiers combine their input which can be aggregated into a random forest ensemble. Through model votes or averaging the above results are aggregated into a single ensemble model.

Some modification which helps in this process are:

1. The number of features that can be split on at each node is limited to some percentage of the total. This ensures that the ensemble model does not rely too heavily on any individual feature, and makes fair use of all potentially predictive features.
2. Each tree draws a random sample from the original data set when generating its splits, adding a further element of randomness that prevents overfitting.

The main advantages and features of Random Forest are:

1. It is the most accurate learning algorithm available.
2. It produces a highly accurate classifier.
3. It runs efficiently on large databases.
4. Without variable deletion it can handle thousands of input variables.
5. Very effective method for estimating missing data.
6. When a large proportion of the data is missing it maintains accuracy.

B. Convolution Neural Network

A Convolutional Neural Network(ConvNet) is a Deep Learning algorithm which takes an input image, assigns learnable weights and biases to various objects in the image and is able to differentiate one from the other. The ConvNet's architecture is analogous to the connectivity pattern of Neurons in the Human Brain. It is inspired by the organization of the Visual Cortex. Usually CNN works by pushing an image to the network which is inputting the image. Next infinite number of steps is applied on the input image which is a convolutional part of the network. Finally, we can predict the output of the image using a neural network.

Four main components of a CNN are:

1. Convolution
2. Non Linearity(ReLU)
3. Pooling or Sub Sampling
4. Classification (Fully Connected Layer)

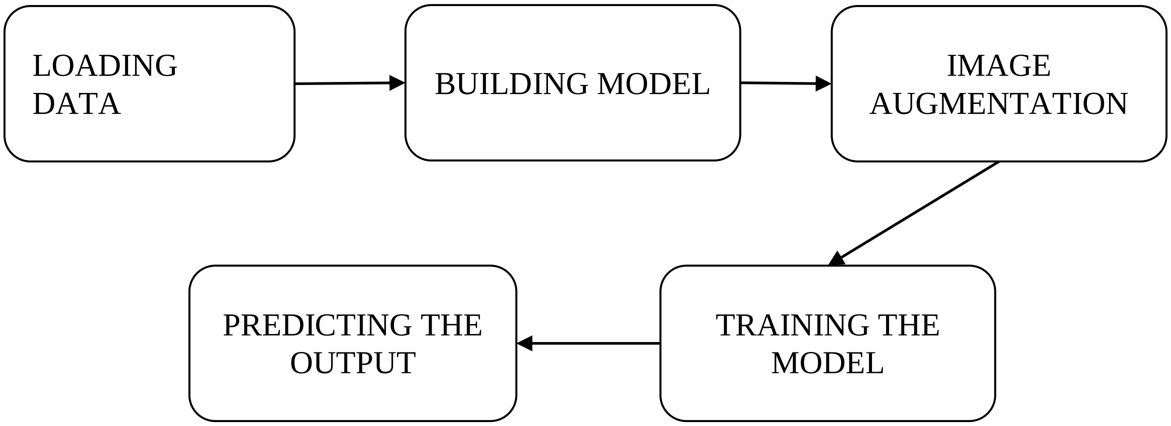
The input image is passed through a series of all the above layers and then the output is generated.

The first layer, Convolution layer, is used to draw out features from an input image. Convolution conserves the relationship between pixels by learning image features using small squares of input data which is called filter. The filter is used to multiply its values by the original pixel values. Later all these multiplications are summed up. At the end one number is obtained. After passing the filter across all positions, a matrix is obtained, but smaller than an input matrix. Convolution of an image with different filters can perform functions such as edge detection, blur and sharpen by applying filters. The pooling layer accompanies the nonlinear layer. It works with width and height of the image and performs a downsampling operation on them. As a result the image volume is reduced. The fully connected layer takes the output information from convolutional networks. Attaching a fully connected layer to the end of the network results in an N dimensional vector, where N is the amount of classes from which the model selects the desired class.

III. Implementation

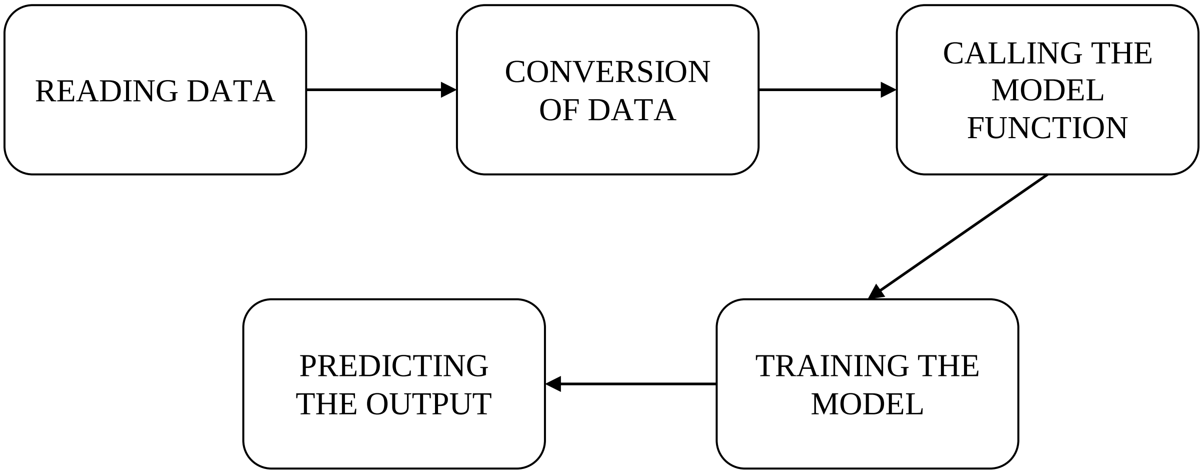
In this research, we have developed a user friendly web application, the farmer will interact with the user interface of this system. The web application is further divided into two main divisions, those are disease detection and crop yield prediction. In the disease detection division the user is asked to select the particular crop and then the user is directed to the next web page where he/she has to upload the image of the diseased part of the selected crop. This uploaded image is then sent to the model and then the disease is detected and the remedies for that disease are displayed on the web page. In the yield prediction division the user is asked to select the district, season and crop, then he is asked to enter the area. Once the user has given all this information, this info is then passed on to the model where it predicts the yield for the given inputs. Then this predicted value is displayed on the web page.

Disease Detection:



The workflow of the disease detection part is as shown in the above diagram. The image datasets of certain plants are loaded into a particular list, which is then converted into a numpy array and saved as a '.npy' file. The CNN model is built using the TensorFlow and Keras inbuilt functions. This model consists of convolution, activation, pooling and fully connected layers. After this the ImageDataGenerator function is called to perform the image augmentation process where the saved .npy files are passed to fit this function after performing certain mathematical operations on it. The output obtained from this function is used to train the CNN model, which is already built and then this model is saved. The saved model is loaded in order to predict the disease for the given input.

Yield Prediction:

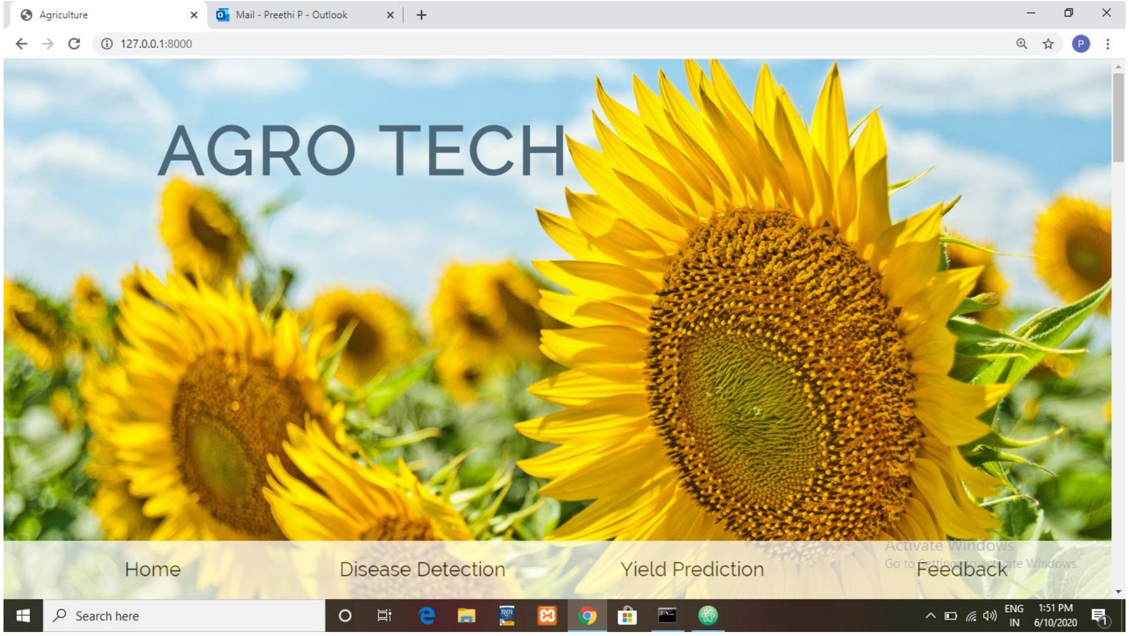


The workflow of the yield prediction part is as shown in the above diagram. The datasets which are stored in the form of a '.csv' file are loaded to a particular variable. This loaded data is then converted to NumPy array and saved. The sklearn package is imported from which Random Forest Regressor model function is called. The variables in which NumPy array values are stored, are used to train the model by calling the fit function. After the model is trained, it is saved as a pickle file. This saved model is later loaded in order to predict the output for the given input.

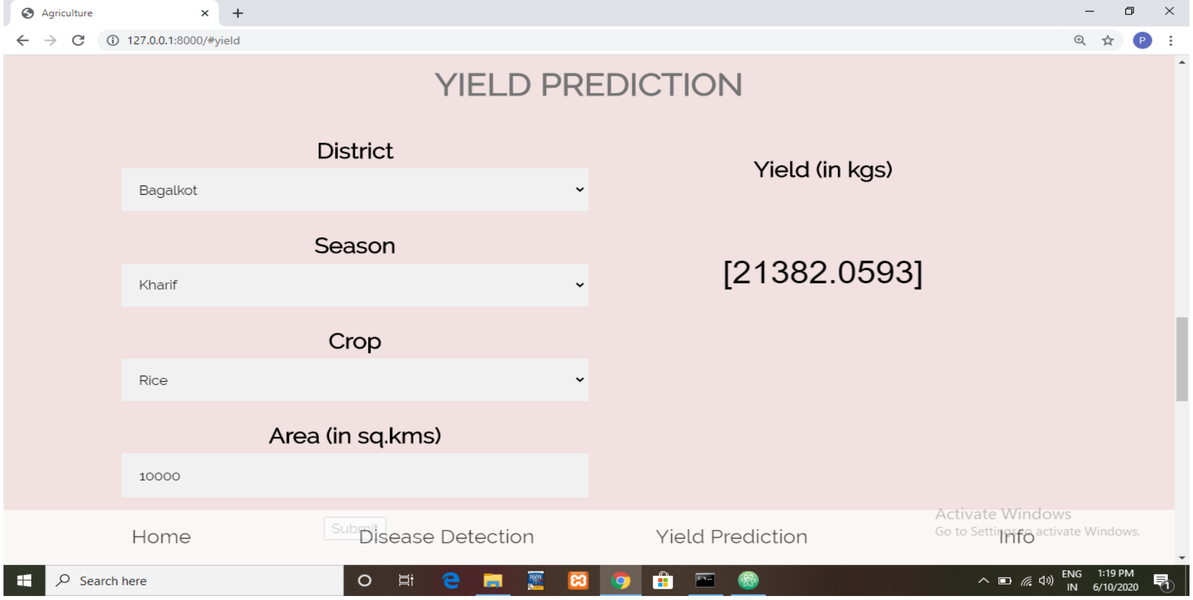
IV. Result

The aim of the research was to construct a user friendly web application that will help the farmers to predict the crop yield well ahead of its harvest and to detect any disease in a particular crop.

The backend of the system was made using python and the results of this system were rendered into a web page by using Django.



The above figure shows the home page of the web application where the user is given the option to predict the yield or to detect the disease and then to give feedback.



YIELD PREDICTION

District: Bagalkot

Season: Kharif

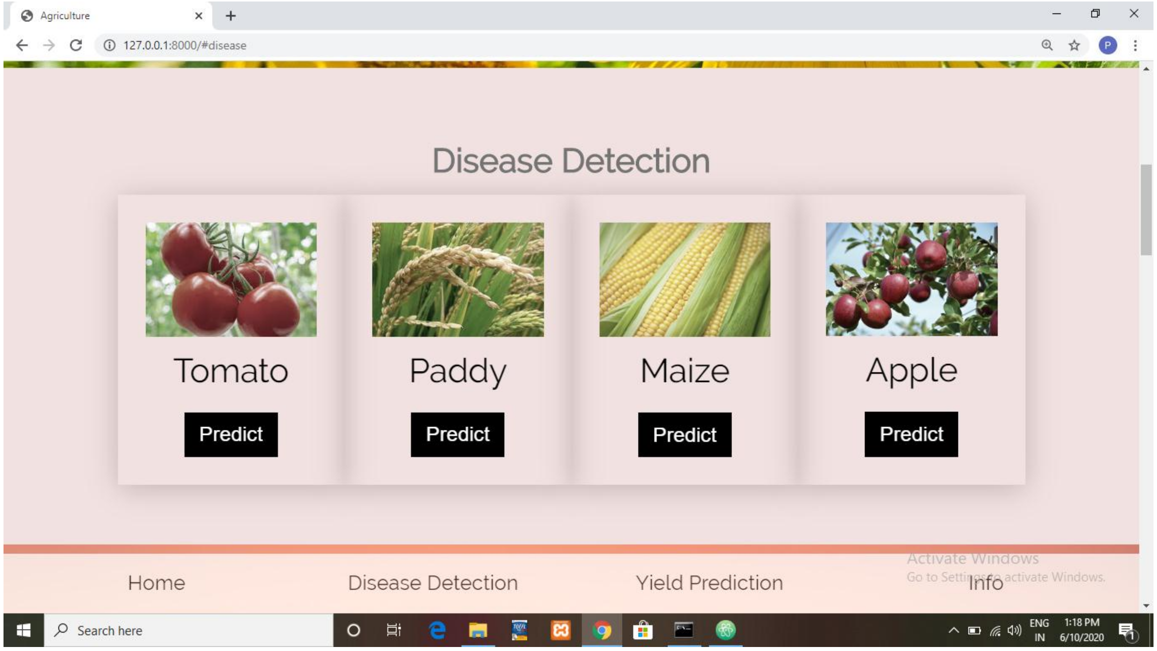
Crop: Rice

Area (in sq.kms): 10000

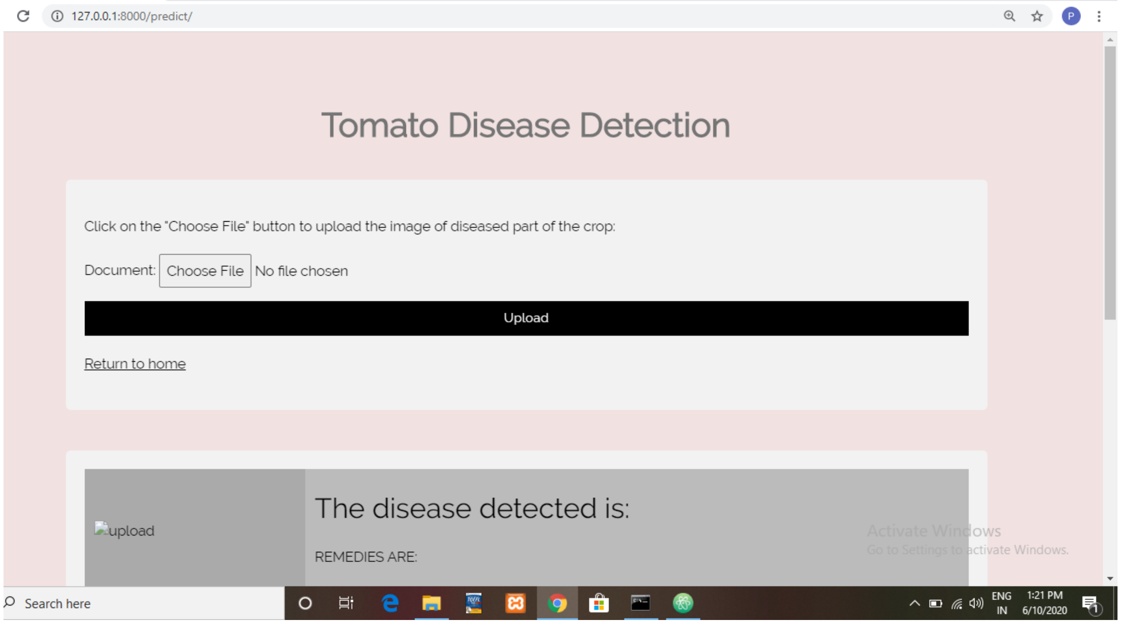
Yield (in kgs)

[21382.0593]

This figure shows the sub-web application to predict the yield. Here the user is asked to select the crop, district, season and then to enter the area. The web-application, then displays the predicted yield value for the given input values.



The above figure displays the sub-web application which provides the user to select a particular crop for disease detection. Once the crop is selected, the user is then directed to another page which is shown below.



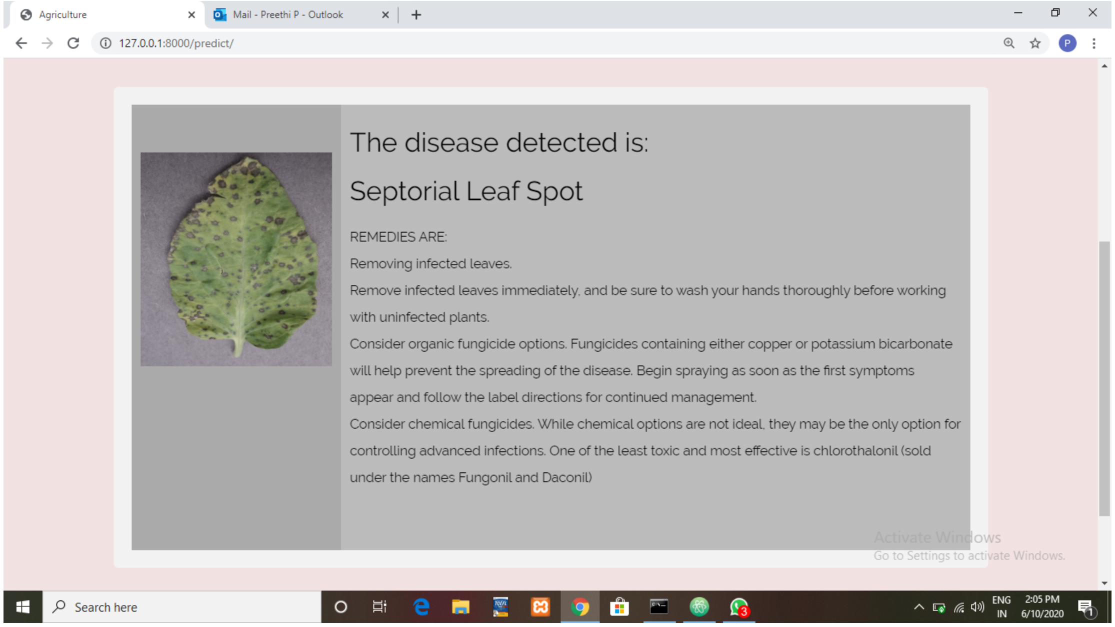
Tomato Disease Detection

Click on the "Choose File" button to upload the image of diseased part of the crop:

Document: Choose File No file chosen

The disease detected is:

REMEDIES ARE:



The disease detected is:

Septorial Leaf Spot

REMEDIES ARE:

Removing infected leaves.

Remove infected leaves immediately, and be sure to wash your hands thoroughly before working with uninfected plants.

Consider organic fungicide options. Fungicides containing either copper or potassium bicarbonate will help prevent the spreading of the disease. Begin spraying as soon as the first symptoms appear and follow the label directions for continued management.

Consider chemical fungicides. While chemical options are not ideal, they may be the only option for controlling advanced infections. One of the least toxic and most effective is chlorothalonil (sold under the names Fungonil and Daconil)

Here the user is asked to upload the image of the leaf-part which is diseased, then the disease is detected in the uploaded input and the disease along with its remedies is displayed.

V. Conclusion

The model used to build this web application provides an accuracy of approximately 90%. In the yield prediction part we are considering datasets of previous year's crop yield, which is developed by considering factors such as season, crop, rainfall at a particular place to predict the crop in coming years.

But some factors such as soil fertility and pests are not considered here, whereas these factors also can be taken into account for further improvisation. In the disease detection part we are predicting the disease only for a few particular crops like tomato, sugarcane, rice, etc which are mainly grown in Karnataka. This can also be extended to predict the disease for other crops as well.