DETECTION OF PLANT DISEASES USING MACHINE LEARNING

Suyash Srivastava[*1] ,Prakhar Pathak [*2]Mr.Kuwar Pratap Singh [*3.]

1.Department of CSE, SRM Institute of Science and Technology, Modinagar, Ghaziabad

2.Department of CSE, SRM Institute of Science and Technology, Modinagar, Ghaziabad

3.Assistant Professor, Department of CSE, SRM Institute of Science and Technology, Ghaziabad

---

ABSTRACT-

Plant leaf diseases and pests are a major cause a challenge in the agricultural sector. Quick and accurate predictions of leaf diseases in plants can help improve the treatment process early while greatly reducing economic losses. Modern development Advances in in-depth learning have allowed researchers to go to extremes improve the efficiency and accuracy of object discovery too awareness programs. In this paper, we suggest reading in depth how to get leaf diseases in many different plants using pictures of leaves of plants. Our goal is to find and develop the most effective ways to learn about our work. So, we look at three main ones beneficiary families:

 # Faster Region Based Convolutional Neural Network(Faster R-CNN)

 # Single Shot MultuBox Detector which was used for this purpose of this work. The proposed system can be identified separately types of diseases that have the ability to deal with complex conditions arising from plant area.

Keywords:Region-based fully convolutional network(R-FCN), Single Shot Multibox Detector(SSD), Machine Learning.

1.INTRODUCTION-

In early times ,People study about all the plant leaves and their respective diseases. But as the time passes, People came to know that different plants have different diseases .So it became tough for humans to learn about all the diseases existing. Then after sometime , people came to know that we can do the mentioned problem with the help of machine .So, People started doing their research on disease detection with the help of machine. In the early times of machine testing, calculations were done manually. Now as the technology improved , calculations are done by machine itself. In this research paper, we are trying to test the new algorithms of machine learning to increase the efficiency of already existing research on plant disease detection in which the photo of plant leaf is taken as input and it will predict that by which disease that plant or plant leaf is spoiled. In this Research Paper , we will diagnose on the following diseases.

1. Bacterial Spot
2. Yellow Leaf Curl Virus
3. Septorial Leaf Spot
4. Leaf Mold
5. Early Blight
6. Later Blight

ANACONDA

Anaconda is a software tool which is generally made for the Python and R languages.This tool is basically the birthplace of Data Science.This tool consists of various Python interpreters , R Studio and Visualisation tools asshown in Figure 1.

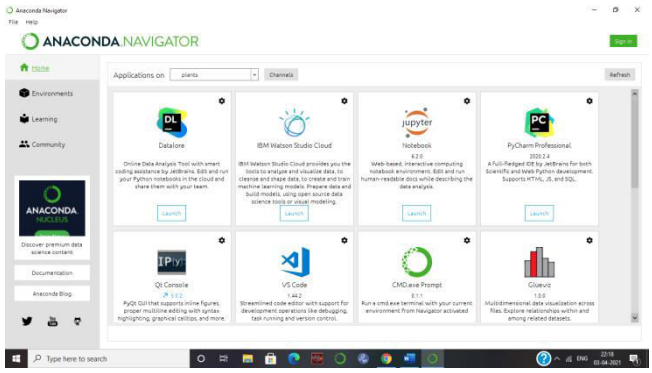

Figure 1: Anaconda interface

JUPYTER NOTEBOOK

Jupyter Notebook is one of the best Python interpreter provided by the Anaconda which is used for performing machine learning and data science processes. It is available free and comfortable to use.

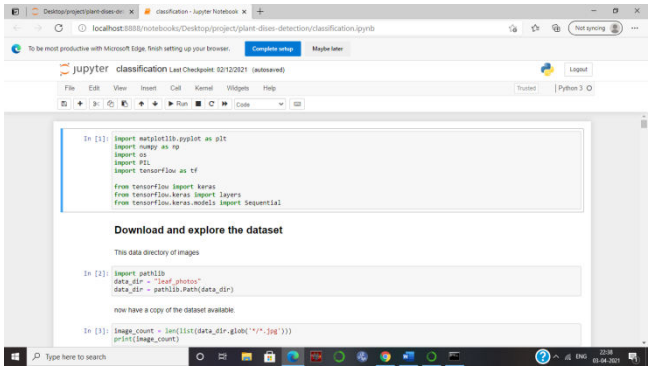

Figure 2: Jupyter Notebook interface.

2.LITERATURE SURVEY-

[1]Vijai Singh, Varsha, and A.K. Misra,"Detection of unhealthy region of plant leaves using Image Processing",2015 International Conference on Advances in Computer Engineering and Applications (ICACEA),2015. In this research paper , the researchers, did the image processing using Random Population and calculate the fitness of problem manually.[2] Sachin D. Khirade and A. B. Patil, "Plant Disease Detection Using Image Processing ",International Conference on Computing Communication Control and Automation,2015. Image Processing has been used by using Artificial Neural Network and Support Vector Machines ,A comparative study has been done.[3] Jitesh P. Shah, Harshadkumar B. Prajapati and Vipul K. Dabhi ,"A Survey on Detection and Classification of Rice Plant Diseases ",IEEE International Conference. A detailed study has been done on Rice plant leaves using Image Processing. The process is good working but only applicable to Rice Plants. [4] Jobin Francis,AntoSahaya, Dhas D, and Anoop B. K.,"IDENTIFICATION OF LEAF DISEASES IN PEPPER PLANTS USING SOFT COMPUTING TECHNIQUES",2016 Conference on Emerging Devices and Smart Systems. In this paper ,The experiment is done on pepper plant leaves using image acquisition and back propagation neural network for calculation and classification. Back propagation is a time taking process. [5] Monzurul , Anh Dinh, Khan Wahid, and Pankaj Bhowmik,"Detection of Potato Diseases Using Image Segmentation and Multiclass Support Vector Machine",2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE). Here SVM has been used for classification and the research work has been done on potato plant leaves . The pixel information has been used to predict the output. [6] MelikeSardogan, AdemTuncer, and YunusOzen,"Plant Leaf Disease Detection and Classification based on CNN",2018 3rd International Conference on Computer Science and Engineering. In this research work CNN Algorithm has been used . New techniques already exist which are better than CNN. [7] Korkut, and OktayYildiz, "Detection of Plant Diseases by Machine Learning",2018 26th Signal Processing and Communication Applications Conference. In this research work KNN algorithm has been used for classification and SVM to increase dimensionality. [8] Sharath D M, Akhilesh, S Arun Kumar, Rohan M G and PrathapC ,"Image based Plant Disease Detection

in Pomegranate Plant for Bacterial Blight",International Conference on Communication and Signal Processing. In this research work has been done on Pomegranate plant leaf for detecting Bacterial Blight. The image is first converted into RGB format. For calculation Gaussian filter has been used. [9] Md. Arifur Rahman, Md. Mukitul, G M Shahir Mahdee, and Md. Wasi Ul Kabir,"Improved Segmentation Approach for Plant Disease Detection", 1st International Conference on Advances in Science, Engineering and Robotics Technology 2019. In this paper, Python language has been used . The image is converted into RGB format. Deep Neural Network has been used for classification and calculation. [10] Pushkara Sharma, Pankaj Hans, and Subhash Chand Gupta, "CLASSIFICATION OF PLANT LEAF DISEASES USING MACHINE LEARNING AND IMAGE PREPROCESSING TECHNIQUES ",10th International Conference on Cloud Computing, Data Science and Engineering. In this research work , a comparative study has been done between SVM Classifier, KNN and CNN for the classification of dataset in which CNN stands the best.

3. METHODOLOGY-

The purpose of the methodology is defined in this section. Our methodology is depicted in the following steps. First of all, we used the annotated dataset.

We will use the latest version of CNN as our algorithm which is Faster Region Based Convolutional Neural Network.CNN has three types

1. Convolutional Neural Network.
2. Fast Region Based Convolutional Neural Network (25 times efficient than CNN).
3. Faster Region Based Convolutional Neural Network (250 times efficient than CNN).

CNN----------→50 sec

Fast R-CNN----------→2 sec

Faster R-CNN-------→0.2 sec

ARCHITECTURAL DESIGN

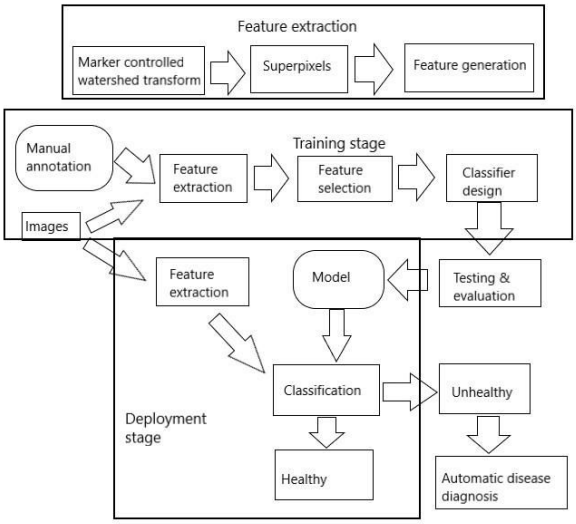

Figure 3: Architectural Design

Feature Extraction

Feature Extraction is the process of indentify the essential features of the datasets. It will will convert the image of datasets into pixels.

Training Stage

In this stage , The dataset is first converted into RGB format ,then 80% of data is trained with the algorithm used in the process.

Classification

Data is classified into various diseases with the help algorithm and then it is tested with that datasets.

Testing Stage

We will provide an input from our side as an image of plant leaf and test the dataset of plant and the confidence of the output is its efficiency.

3.1. Download and explore the dataset

The data is downloaded from the various websites like Kaggle, and Tensorflow in the form of images of plants with disease effected leaves and healthy leaves. We also used Keras from Tensorflow.

3.1.1. Tensorflow

It is a free and open-source Python library which is used to perform the Machine Learning activities.

It is specially used for performing Neural Network Machine Learning activities.

3.1.2. Keras

Keras is a Python module that is used to provide a interface between Python and Tensorflow for taking datasets and load it to the Python compiler.

3.2. Create a dataset

We created a dataset of images of plants leaves with different diseases like Bacterial Spot, Early Blight, Leaf Mold, Septorial Leaf Spot etc. Later it is converted into numpy array as RGB image.

3.2.1. Numpy

Numpy is a module in the Python language which is used to draw the arrays of the dataset. It converts the data into tabular form.

3.3. Visualize the dataset

In this part , we will check whether the dataset is properly created or not. Here we will use Matplotlib module to plot some images from the dataset.

3.3.1. Matplotlib

Matplotlib is a module in the Python language which is used to plot the graphs of the model and process.

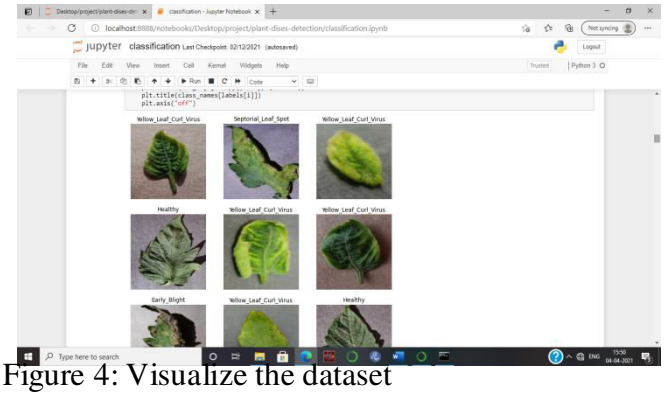

Figure 4: Visualize the dataset

3.4. Create the Model

We will create a model which will be used to predict the results. Here we will apply the algorithm as R-FCN for creating the model and classification of the data.The model is made of three convolution blocks with a max pool layer in each of them. It has a fully connected layer with 128 units on top of it that is activated by a relu activation function. This model has been tuned for high accuracy, the goal of this research is to show a standard approach.

3.4.1. Region-based Fully Convolutional Network

Faster R-CNN is a deep convolutional network used for object detection, that appears to the user as a single, end-to-end, unified inter-connection. The network can accurately and quickly predict the locations of different objects.

3.5. Train the Model

We train the 80% of the data model and test the remaining 20% of the data for predicting the results. But due to increase in number of data ,the possibility for Overfitting increases.

Overfitting

It is an error in machine learning that arises when the model is trained with too much data which contains some noisy or irrelevant data that will disturb accuracy of the prediction of model.

Here we will train the model first with Overfitting and then after removing the Overfitting.

3.5.1. Train the Model with Overfitting

After training the model , the Overfitting occurs which reduces the accuracy of the model.

To reduce the Overfitting , We can use the following methods which is mentioned below.

Data Augmentation

Data Augmentation methods which is used to expand the amount of data by adding modified copies of already existing data or newly created synthetic data from existing dataset. It helps to recognize various possibilities of data to generalize better.

Dropout

Dropout means ignoring units i.e. neurons during the training phase of certain set of neurons which is chosen at random.

3.5.2. Train the Model without Overfitting

After removing the Overfitting , We will again train the data set and then visualize it.

3.6.Predict on new data

After the training of the data set , we will introduce another algorithm named Single Shot Multibox Detector which will converge all the classified diseases into single unit which will increase the efficiency of the process and then we will provide some input in the form of plant leaf image and it will predict that by which disease that plant is affected and it will mention how much sure it is which will be shown as it's efficiency.

4.RESULT-

4.1. Result in Training Stage

4.1.1 Result of Training with Overfitting

In this stage , the data is trained and it is showing 100% training efficiency but some noise or irrelevant data is also added which is removed by Data Augmentation and Dropout.

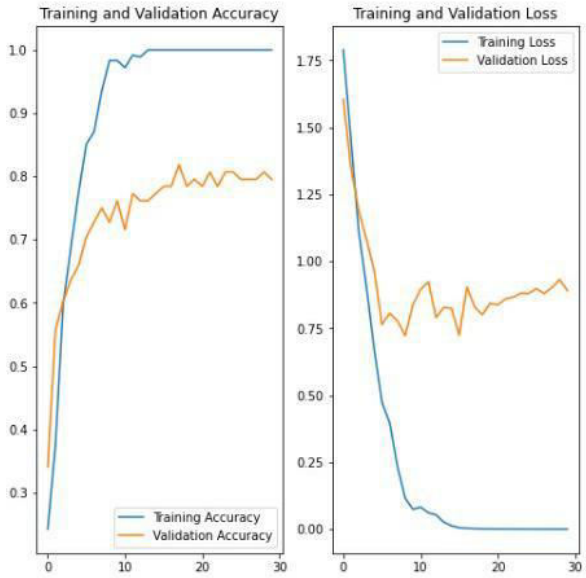

Figure 5:Outputof the Training with Overfitting.

4.1.2 Result of Training without Overfitting

When we remove the Overfitting in the process the machine is getting trained with 92.1% efficiency as in the Overfitting part some irrelevant data was added it was removed then it is showing the given efficiency.

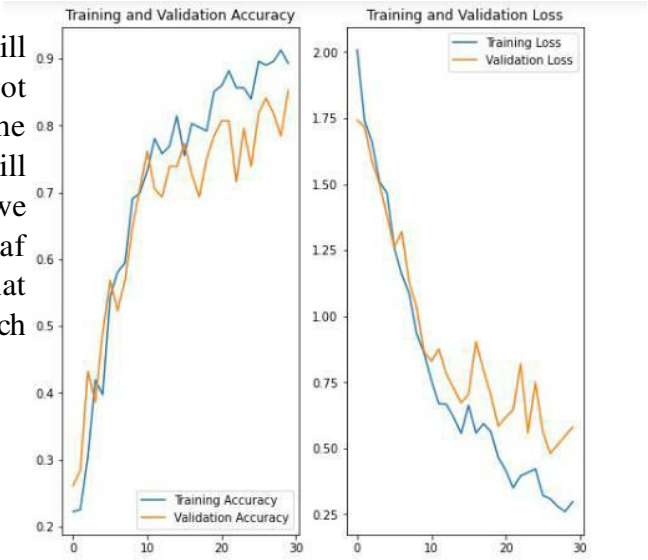

Figure 6: Output of the Training without Overfitting.

4.2. Layers of the Neural Network

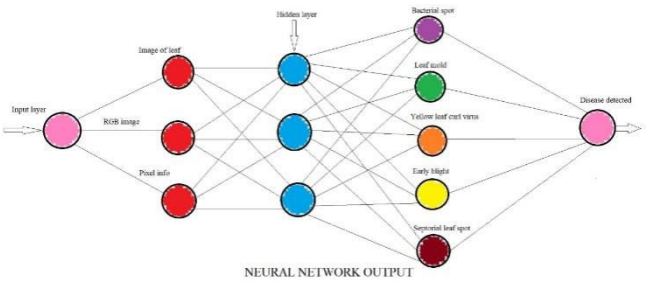

Figure 7: Neural Network formed for the training and testing.

4.3. Resultin Testing Stage

After the training of the dataset , We provided the image of the leaf and we come to know that the image is efficient to detect that by which disease that particular leaf is contaminated.

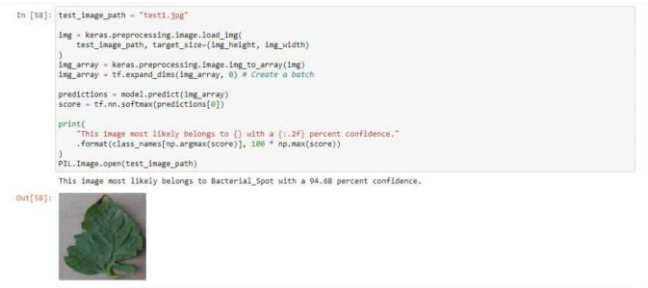

Figure 8: Output ofTesting Stage.

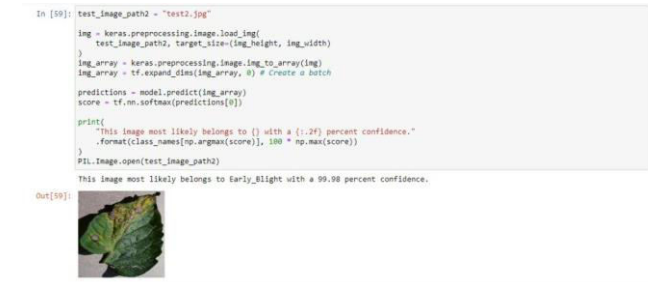

Figure 9: Output of Testing Stage.

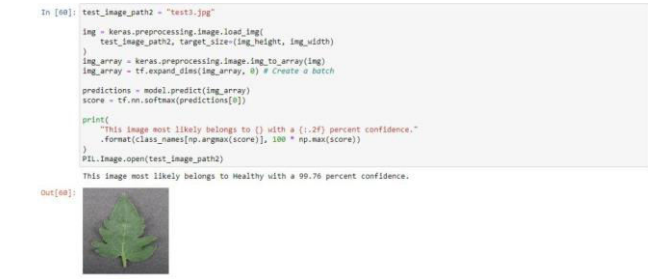

Figure 10: Output of Testing Stage.

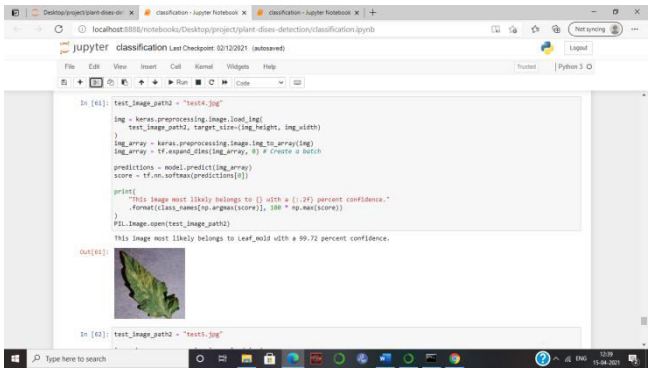

Figure 11: Output of Testing Stage.

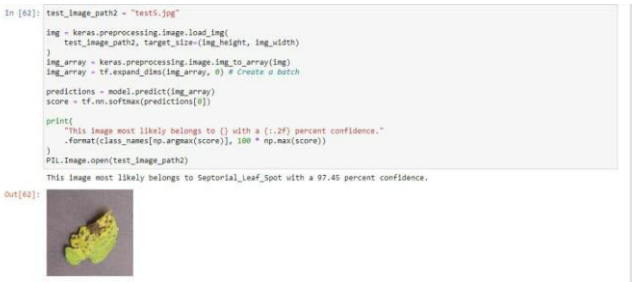

Figure 12: Output of Testing Stage.

Figure 13: Output of Testing Stage.

Efficiency

For calculating the efficiency of the following process , We will find the average of the confidence which is obtained in the output of the various testing images.

Efficiency = Sum of all the confidence / No of testing observations.

5.CONCLUSION-

This Research Paper is made for the purpose of detecting the plant diseases with the help of plant leaf. We tried to make this process more efficient as compared to previous research works. We included here two algorithms which is Faster Region Based Convlutional Neural Network and Single Shot Multibox Detector. We trained 80% of dataset and tested the remaining data. We worked on 443 images of dataset and Python language for coding. We used Jupyter Notebook as Python interpreter. We used the modules as Tensorflow, Keras, Numpy, Pathlib, Matplotlib and PIL for performing the experiment. After performing the experiment , We come to conclusion that our Research work is 98.59% efficient as per our study on previous Research Papers.

6.REFERENCES-

[1] Vijai Singh, Varsha, and A.K. Misra,"Detection of unhealthy region of plant leaves using Image Processing",2015 International Conference on Advances in Computer Engineering and Applications (ICACEA),2015.

[2] Sachin D. Khirade and A. B. Patil, "Plant Disease Detection Using Image Processing ",International Conference on Computing Communication Control and Automation,2015.

[3] Jitesh P. Shah, Harshadkumar B. Prajapati and Vipul K. Dabhi,"A Survey on Detection and Classification of Rice Plant Diseases ",IEEE International Conference,March 10 2016.

[4] Jobin Francis,AntoSahaya, Dhas D, and Anoop B. K.,"IDENTIFICATION OF LEAF DISEASES IN PEPPER PLANTS USING SOFT COMPUTING TECHNIQUES",2016 Conference on Emerging Devices and Smart Systems,2016.

[5] Monzurul , Anh Dinh, Khan Wahid, and Pankaj Bhowmik,"Detection of Potato Diseases Using Image Segmentation and Multiclass Support Vector Machine",2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE),2017.

[6] MelikeSardogan, AdemTuncer, and YunusOzen,"Plant Leaf Disease Detection and Classification based on CNN",2018 3rd International Conference on Computer Science and Engineering,2018.

[7] Korkut, and OktayYildiz, "Detection of Plant Diseases by Machine Learning",2018 26th Signal Processing and Communication Applications Conference , 2018.

[8] Sharath D M, Akhilesh, S Arun Kumar, Rohan M G and PrathapC ,"Image based Plant Disease Detection inPomegranate Plant for Bacterial Blight",International Conference on Communication and Signal Processing, April 4-6, 2019.

[9] Md. Arifur Rahman, Md. Mukitul, G M Shahir Mahdee, and Md. Wasi Ul Kabir,"Improved Segmentation

Approach for Plant Disease Detection", 1st International Conference on Advances in Science, Engineering and Robotics Technology 2019 (ICASERT 2019).

[10] Pushkara Sharma, Pankaj Hans, and Subhash Chand Gupta, "CLASSIFICATION OF PLANT LEAF DISEASES USING MACHINE LEARNING AND IMAGE PREPROCESSING TECHNIQUES ",10th International Conference on Cloud Computing, Data Science and Engineering ,2020.