"PLANT DISEASE DETECTION"

minor project report, submitted in partial fulfillment of the requirement for the award of B.Tech. degree in computer science and engineering

by

Mahesh(2019BCS013)

विश्वजीवनामृतं ज्ञानम्

ATAL BIHARI VAJPAYEE INDIAN INSTITUTE OF INFORMATION TECHNOLOGY AND MANAGEMENT GWALIOR

GWALIOR-474 015

2021

UNDER THE SUPERVISION OF:-

Prof.Mahua Bhattacharya

Candidate Declaration

I hereby certify that the work, which is being presented in the report, entitled Plant disease detection For Online Conference , Bachelor of Technology in CSE and submitted to the institution is an authentic record of our own work carried out during the period May-2021 to August-2021 under the supervision of Prof.Mahya Bhattacharya. I also cited the reference about the text(s)/figure(s)/table(s) from where they have been taken.

Signature of Supervisor :

Date: **9.8.2021**

Candidate Declaration

I hereby certify that the work, which is being presented in the report, entitled Analysis of plants and Plant disease Detection, in partial fulfillment of the requirement for the award of the Degree of Bachelor of Technology and submitted to the institution is an authentic record of my work carried out under the supervision of Prof.Mahua Bhattacharya. I also cited the reference about the text(s)/figure(s)/table(s) from where they have been taken.

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

Signature: mahesh
Roll Number: 2019BCS-013
Date: 4/8/21

Abstract

Crop diseases are a noteworthy risk to sustenance security, however their quick distinguishing proof stays troublesome in numerous parts of the world because of the non attendance of the important foundation. Emergence of accurate techniques in the field of leaf-based image classification has shown impressive results. This paper makes use of Random Forest in identifying between healthy and diseased leaf from the data sets created. Our proposed paper includes various phases of implementation namely dataset creation, feature extraction, training the classifier and classification. The created datasets of diseased and healthy leaves are collectively trained under Random Forest to classify the diseased and healthy images. For extracting features of an image we use Histogram of an Oriented Gradient (HOG). Overall, using machine learning to train the large data sets available publicly gives us a clear way to detect the disease present in plants in a colossal scale.It described the innovative solution that provides efficient disease detection and deep learning with convolutional neural networks (CNNs) has achieved great success in the classification of various plant leaf diseases.

Keywords: Machine learning, Cotton leaves and plants, deep learning,datasets

ACKNOWLEDGEMENTS

I am highly indebted to Prof.Mahua Bhattacharya, and am obliged for giving me the autonomy of functioning and experimenting with ideas. I would like to take this opportunity to express my profound gratitude to her not only for her academic guidance but also for her personal interest in my project and constant support coupled with confidence boosting and motivating sessions which proved very fruitful and were instrumental in infusing self-assurance and trust within me. The nurturing and blossoming of the present work are mainly due to her valuable guidance, suggestions, astute judgment, constructive criticism,and an eye for perfection. My mentor always answered a myriad of my doubts with smiling graciousness and prodigious patience, never letting me feel that we are novices by always lending an ear to my views, appreciating and improving them, and by giving me a free hand in my project. It's only because of her overwhelming interest and helpful attitude, the present work has attained the stage it has.Finally, I am grateful to our Institution and colleagues whose constant encouragement served to renew my spirit, refocus my attention and energy, and helped me in carrying out this work.

(B.Mahesh)

Contents

List of Figures

List of Tables

Abbrevations

ANN Artificial Neural Network
CNN Convolutional Neural Network
FAO food and agriculture organization
HOG Histogram of an Oriented Gradient
ORM Object Relational Manager

1 INTRODUCTION

According to the Food and Agricultural Organization of the United Nations (FAO), the global population is set to reach 9.2 billion by the year 2050.To be able to feed the constantly growing population, a greater efficiency within the current farming methods is necessary. Thus, crop yield increase is one of the important targets in agriculture practices. Herbivorous insects are said to be responsible for destroying one fifth of the world's total crop production annually.

A lot of techniques were used to understand the rules and relationships from diverse data sets, to simplify the process of acquiring knowledge from empirical data. These techniques perform well on artificial test data sets, the main goal is to make sense of real-world data . Machine learning (ML) offers an alternative to the conventional engineering flow when the problem is too complex to develop a solution with guarantees.

1.1 BACKGROUND INFORMATION

Deep learning is a branch of machine learning which is completely based on neural networks. It's on hype nowadays because earlier we had a lot of data and not enough processing power. A formal definition of deep learning is- neurons Deep learning is a particular kind of machine learning that achieves great power and flexibility by learning to represent the world as a nested hierarchy of concepts.

Convolutional Neural Network is a special Neural Network used for image recognition. CNN are much better than ANN because after learning a certain pattern in a picture, a convolutional network can recognize it anywhere, CNN takes advantage of local spatial coherence of images.

1.2 PROPOSED SYSTEM

We planned to design the module so that a person with no knowledge about programming can also be able to use and get the information about the plant disease. It proposed a system to predict leaf diseases. It explains the experimental analysis of our methodology. Samples of 36 images are collected that are composed of different plant diseases like Tomato, Grape, Apple and Healthy Leaves. Different number of images is collected for each disease that was classified into database images and input images. The primary attributes of the image are based upon the shape and texture oriented features.

1.3 OBJECTIVES

• The objective of the project is to do an analysis on cotton disease.

• Later develop and train a deep learning model on the leaves and plants dataset to classify the given images of affected plants vs healthy plants.

• And train some pre-trained models and compare the accuracies of the different models.

1.4 SALIENT FEATURES

• After learning a particular pattern in an image, a convnet can recognize the pattern no matter where it is in the image. AN Artificial Neural Network will have to learn the same pattern if it appears in a different location even if it's the same pattern. This makes convnets data efficient when processing images they need fewer training samples to learn representations that have generalization power.

• They can learn spatial hierarchies of patterns. The first convolution layer will learn small local patterns, a second convolution layer with maxpool and nonlinearity will learn larger patterns made of the features of the first layers,and so on. This allows convnets to efficiently learn all the features.

1.5 limitations:

• The amount of time required for the result and the cost for manual testing.

• And the limited availability of test kits.

• Producing black-box-solutions that are not interpretable, so they are only applicable to a limited set of problems.

So we need a technique that provides the result in a faster way and reduces the cost to check if the plant is affected by cotton disease or not. Still, this model can return good accuracy and can be further improved. (accuracy=0.9754)

1.6 LITERATURE SURVEY

	author	title	year	publisher	work
1.	François Cholle	Deep Learning with python	2017	Google	Understanding neural networks, CNN and Back-propagation
2.	C. Szegey et al	Going deeper with convolutions," in Proceedings of the IEEE Computer Society Conference on Computer and Pattern Recognition	2015	IEEE	Inception-V2 is a combined architecture proposed by that uses the idea of inception blocks and residual layers together
3.	K. Simonyan and A. Zisserman.	Very Deep Convolutional Networks for Large Scale Image Recognition	2015		Brief knowledge about vgg16.
4.	A. Chowdhury, Dhruba K. Bhattacharyya, Jugal K. Kalita	Co-Expression Analysis of Gene Expression	2014	google	coexpression networks, differential networking, and differential connectivity
5.	Xiyaon Guo,MingZh Yongqia dia	Image of plant disease segmentation model	2016	doi.org	image segmentation model for plant diseases based on hybrid frog-hopping algorithms
6.	Chit Su Hlaing, SaiMaung MaungZaw	Plant Diseases Recognition for Smart Farming	2017	arxiv.org	the advantages of GP distribution model and model and applied in plant disease classification

Table 1: Related work

2 METHODOLOGY

The Methodology includes 4 steps:
• Convolution operation
• Pooling
• Flattening
• Full Connection

2.1 SYSTEM ARCHITECTURE

2.1.1 Convolution

As we are dealing with images in computer terms, an image is just a 2d matrix with pixel values between 0 and 255 including them where 0 represents brightness, and 255 represents black, so a black and white image is nothing but a matrix with pixel values 0's and 255's

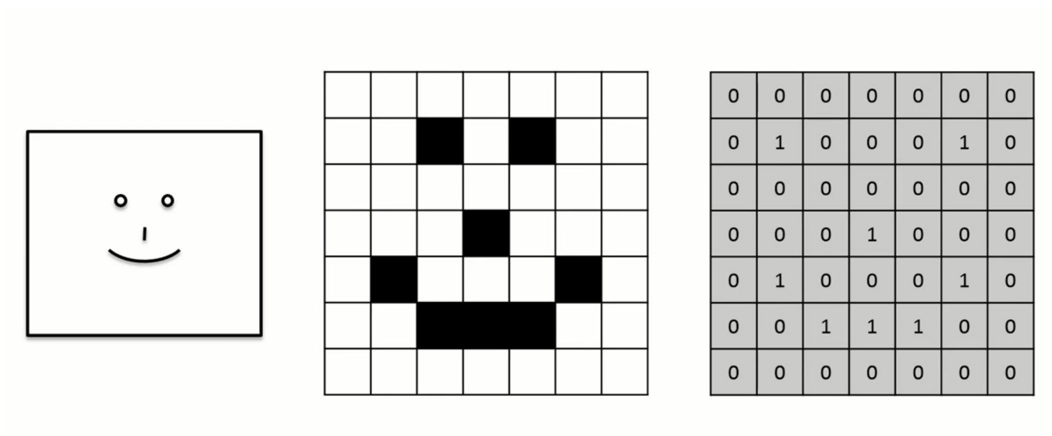

Figure 1: Convolution 2d

As we can see the image above the rightmost matrix represents the The leftmost image where the pixel value 1 represents black pixel and white as 0. We've got an input image as we discussed that's how we're going to look at images just ones and zeros as we can see the input image it's the smiling image we looked before.

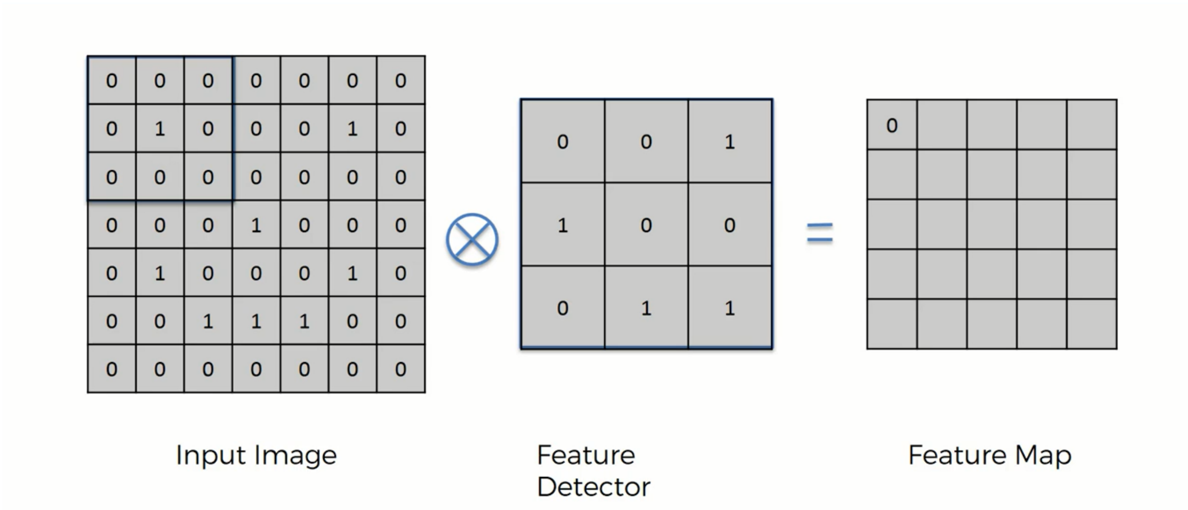

Figure 2: Feature map

This is called Feature map and sometimes called convolved feature
And that's a very important function of the feature detector of this whole convolution step is to make the image smaller because it'll be easier to process and it'll be just faster

we are losing some part of the information but at the same time, we are storing the features of the image. That's what a feature detector does for us; it detects the important features for us.We get the max value when the pattern matches in the image.

2.1.2 Max pooling

let we want to identify a cat in an image so in an image, the cheetah can look in one direction in another image it can look in another direction it can be in one part in an image it can be in another position so at the end we need to identify it like a cheetah no matter what direction it's looking or where it is in the image.

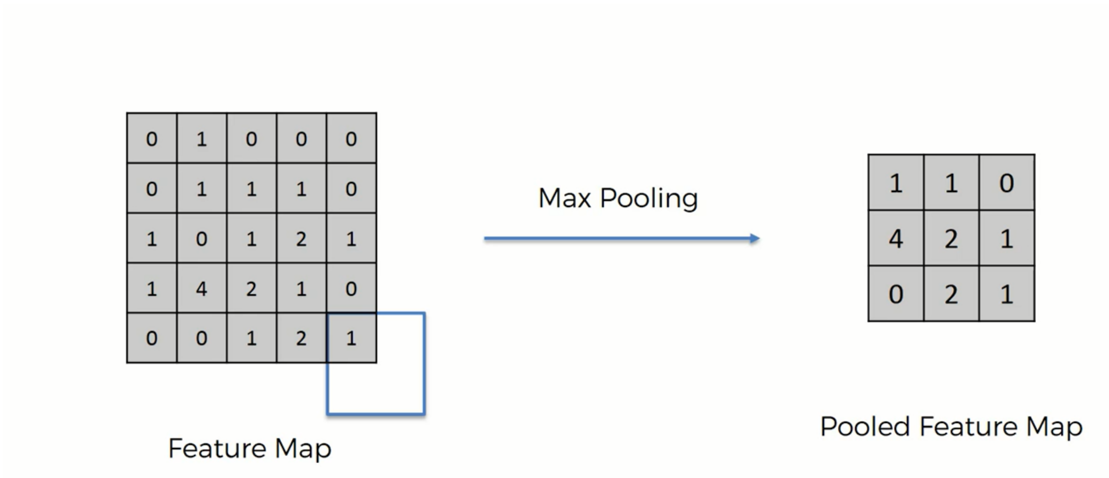

Figure 3: max pool

There are several different types of pooling methods like mean pooling Max pooling But for now, we're just applying Max pooling so we take a box of two by two pixels.

2.1.3 Flattening

We just flatten the max-pooling matrix into a vector so that we can use an input for the neural network later on for classification.

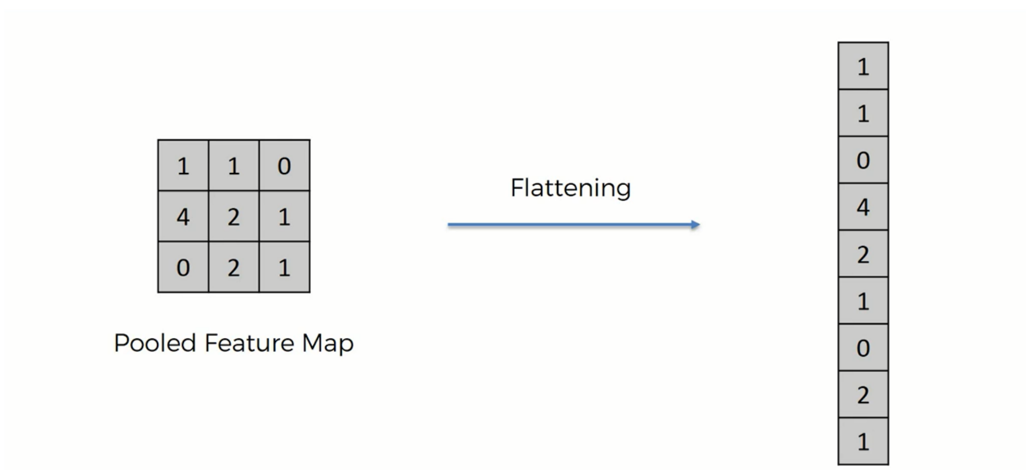

Figure 4: Flat

2.1.4 Fullconnection

The flatten max pooled matrix is the input for artificial neural networks.Initially the neurons in the neural network have random weights and the weights are changed accordingly to maximize the output accuracy.

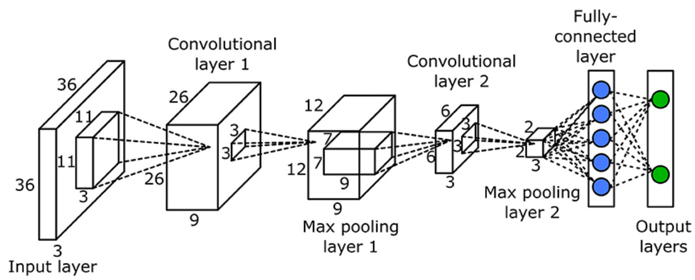

Figure 5: System architecture

2.2 DATA PREPROCESSING

These image paths are stored in a dataset pathimages in a column and labels of the image and then we iterate a loop over the dataset and classify these images into their respective directories.

2.2.1 BGR-RGB

As we are making use of CNN. CNN takes an RGB image. which are of size 224*224 as most of the images are BGR we convert them to RGB using the cv2 module.

2.2.2 Scaling

Scaling the pixel values are between 0 and 255 we divide each pixel value by255 so that we scale them down to 0 and 1

2.2.3 Splitting of Data into Train, Test, validation

Total data in the datasets are divided into Test, Train data accordingly.
Train Data: 80(precentage)
Validation Data: 10(precentage)
Test Data: 1(precentage)

Data	Train	Test
diseased cotton leaf	288	3
diseased cotton plant	815	5
fresh cotton leaf	427	5
fresh cotton plant	421	5

Table 2: Train and test set

2.3 IMPLEMENTATION DETAILS

2.3.1 Dataset

• A dataset containing information regarding plants and leaves.
• It consists of three parts Train Test Validation
• The dataset is a collection of diseased cotton leaf,diseased cotton plant,fresh cotton leaf,fresh cotton plants.

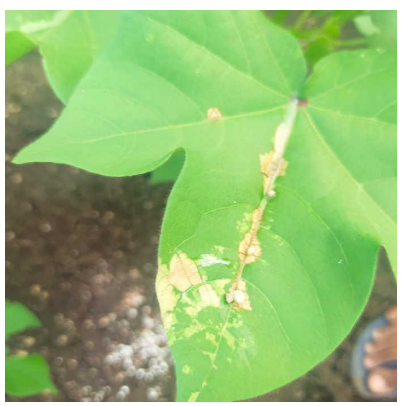

Figure 6: Diseased cotton leaf

Figure 7: Diseased cotton plant

Figure 8: Fresh cotton leaf

Figure 9: Fresh cotton leaf

2.3.2 Tools and Hardware

• Python
• Numpy
• Pandas
• Keras or Tensorflow
• Seaborn
• sklearn
• CPU or GPU(recommended)

2.3.3 Hardware Setup

• If your computer does not have GPU(GPU is recommended for faster Computation) we can use GPUs that are provided by Google or Kaggle.

• Kaggle is recommended as the datasets are available on kaggle. Datasets can be loaded with ease if you are using Kaggle kernel. If you want to do it on Google Colab you need to upload the datasets to your Google Drive first.

• In a kaggle kernel we can use the GPU thats provided for us by Select the Settings tab. Then select the checkbox for Enable GPU. Verify the GPU is attached to your kernel in the console bar, where it should show GPU ON next to your resource usage metrics. And turn on the internet to download the relvent libraries.

3 Results and Discussions

3.1 CNN

I have done some research on websites after choosing the topic.Since I am familiar with Tensor flow and keras applications. so I have trained the dataset model using CNN using model.fitgenrator and giving epochs 20 Training set gives accuracy 0.9754 and loss 0.1898.

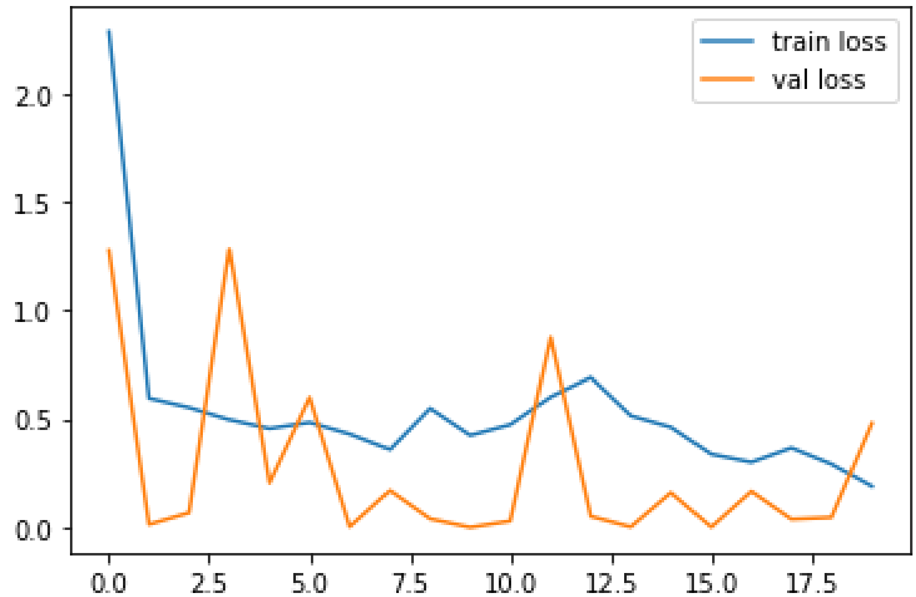

Figure 10: Loss metrics

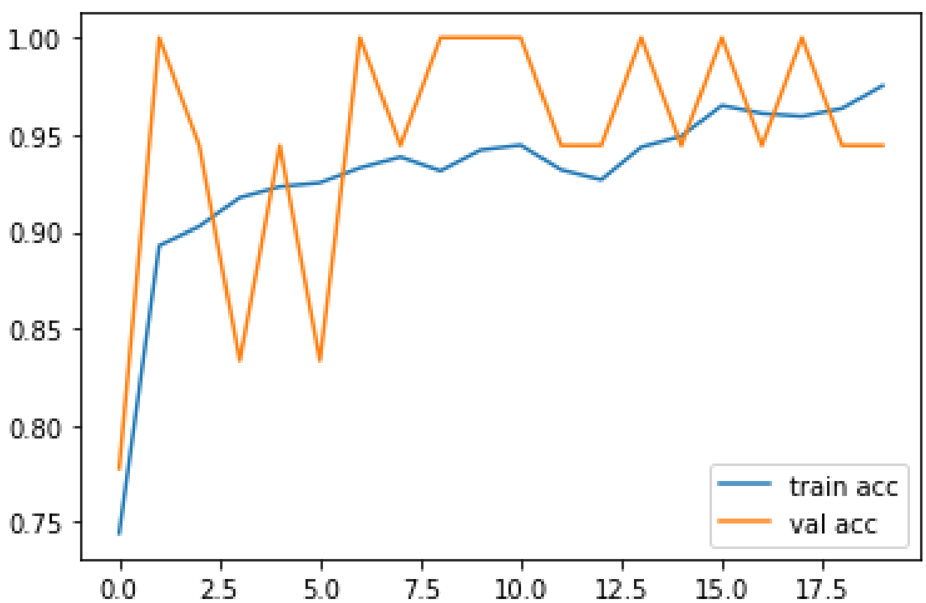

Figure 11: accuracy gain metrics

The performance metrics used to evaluate the model are the accuracy and the loss. our main aim was to increase the accuracy at the same time decrease the loss.

• The figure shows the comparison between accuracy and epochs, and also between loss and epochs. From the figure the accuracies i.e., both the training and validation accuracies are increasing from the first epoch itself which tells that the model is learning.

• TensorFlow Framework was used. The dataset is taken from the Kaggle.

Data	Train	Test
diseased cotton leaf	288	3
diseased cotton plant	815	5
fresh cotton leaf	427	5
fresh cotton plant	421	5

Table 3: Train and test set

Note: These results are for 291 diseased cotton leaf,820 diseased cotton plant,432 fresh cotton leaf,426 fresh cotton plant these metrics may change if the no.of input images change.

3.2 Symptoms

• SEEDLING BLIGHT: Small, water-soaked, circular or irregular lesions develop on the cotyledons.Later, the infection spreads to stem through petiole and cause withering and death of seedlings.

• ANGULAR LEAF SPOT: Small, dark green, water soaked areas develop on lower surface of leaves,enlarge gradually and become angular when restricted by veins and veinlets and spots are visible on both the surface of leaves.They turn to reddish brown colour and infection spreads to veins and veinlets.

• BLACK VEIN: The infection of veins causes blackening of the veins and veinlets, gives a typical ‘blighting’ appearance. On the lower surface of the leaf,bacterial oozes are formed as crusts or scales. The affected leaves become crinkled and twisted inward and show withering. The infection also spreads from veins to petiole and cause blighting leading to defoliation.

• BLACK ARM: On the stem and fruiting branches, dark brown to black lesions are formed, which may girdle the stem and branches to cause premature drooping off of the leaves, cracking of stem and gummosis, resulting in breaking of the stem which hang typically as dry black twig to give a characteristic “black arm” symptom.

• BOLL ROT: On the bolls, water soaked lesions appear and turn into dark black and sunken irregular spots. The infection slowly spreads to entire boll and shedding occurs. The bacterium spreads inside the boll and lint gets stained yellow because of bacterial ooze and looses its appearance and market value. The pathogen also infects the seed and causes reduction in size and viability of the seeds

3.3 Flask

Flask is a web framework, it’s a Python module that lets you develop web applications easily. It’s has a small and easy-to-extend core: it’s a microframework that doesn’t include an ORM (Object Relational Manager) or such features.

It does have many cool features like url routing, template engine. It is a WSGI web app framework.

• Save the training model in ’model.h5’ with keras and pass it into the flask app.

• Predict the model using if-else constraints.

Data	preds
diseased cotton leaf	0
diseased cotton plant	1
fresh cotton leaf	2
fresh cotton plant	3

Table 4: Prediction of data set

• Using app route it will create web application(through html) and deploy the image of plant it will give you the result of plant or leaf whether it has disease or not

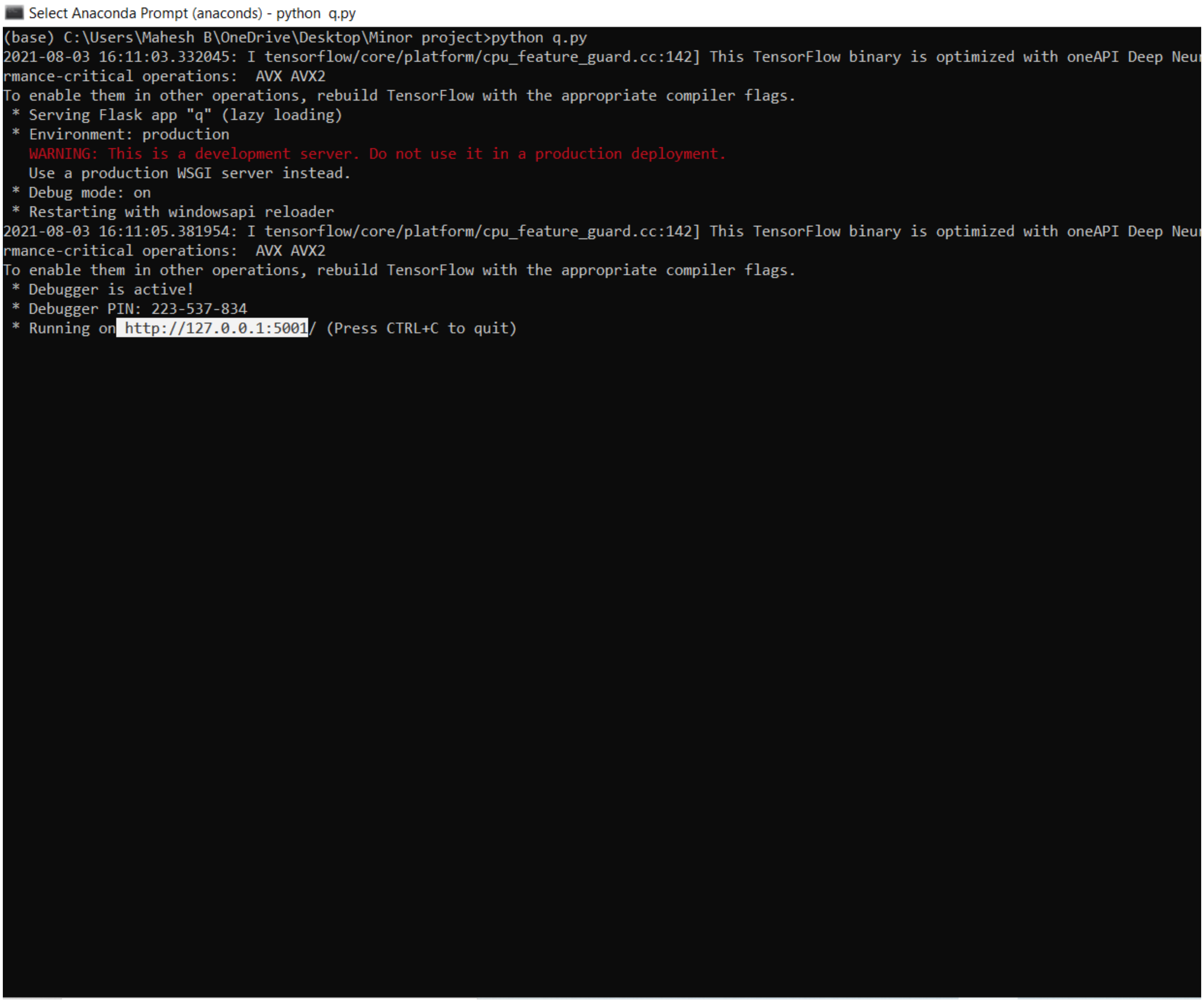

Figure 12: Compiling flask app

- Copy this url and paste it on google we will get the result of the plant.

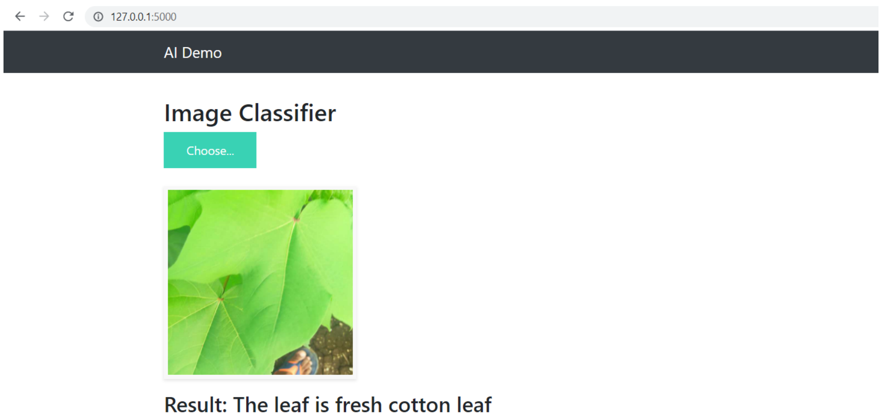

Figure 13: Fresh cotton leaf

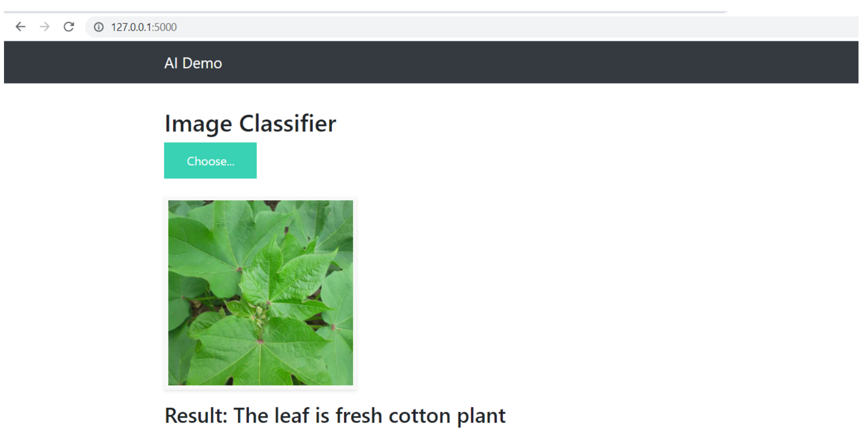

Figure 14: cotton plant

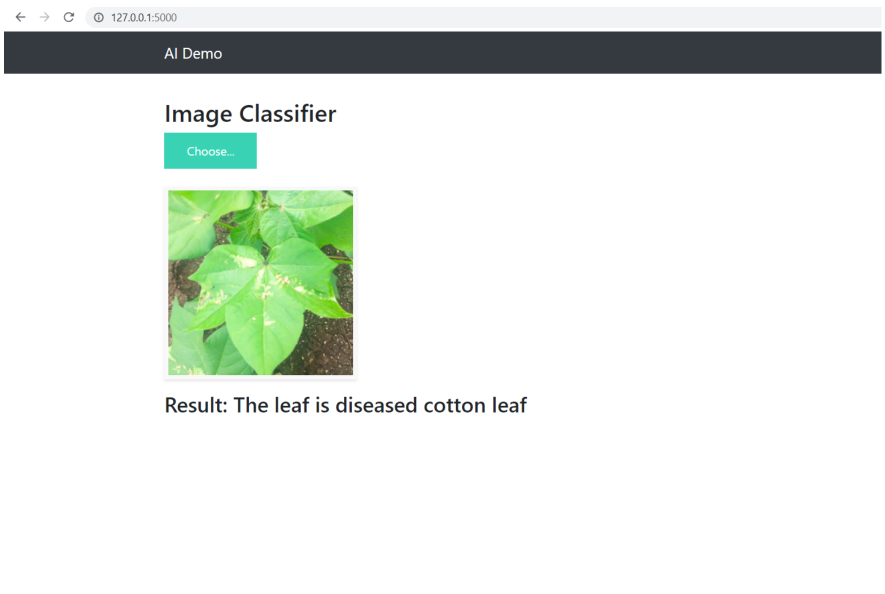

Figure 15: Diseased cotton leaf

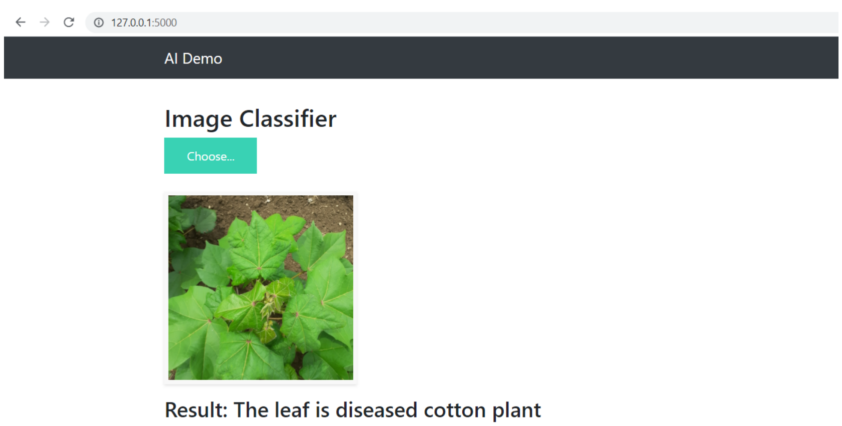

Figure 16: Diseased cotton plant

4 CONCLUSION

This report provides a critical analysis for CNN-Architecture, proposed originally for image analysis. These CNN-Architectures are used to differentiate disease vs healthy based on plant and leaf images.

We also proposed a simple CNN architecture which gave us good accuracy by giving epochs 50,The result of the CNN that we built from scratch and load in .h5 file and pass it to the flask app.

With the help from these techniques we implemented neural network based on plant leaf disease detection and it takes less amount of time to for detection.

4.1 Advantages

1. The amount of time required for the result and the cost is less when compared to manual testing.

2. And the limited availability of test kits.

3. And can be used for different plants and leaves for classification.

4. Increased layers of deep learning algorithm to get most accurate and appropriate result.

4.2 Future scope

The future work may be:

1. Removing other unnecessary noise in the image such as text writing marked on plant and leaf images for a better vision and understanding.

2. And the collection of more plant and leaf Images will improve the models accuracy

5 REFERNECS

[1] Monica Jhuria, Ashwani Kumar, and Rushikesh Borse, “Image Processing for Smart Farming: Detection of Disease and Fruit Grading”, Proceedings of the 2013 IEEE Second International Conference on Image Information Processing (ICIIP-2013.

[2] Sachin D. Khirade, and A. B. Patil, “Plant disease detection using image processing,” IEEE

[3] Shivani K. Tichkule, Prof. Dhanashri. H. Gawali, “Plant Diseases Detection Using Image Processing Techniques”, 2016 Online International Conference on Green Engineering and Technologies (IC-GET).

[4] Di Cui, Qin Zhang, Minzan Li, Glen L. Hartman, Youfu Zhao, “Image processing methods for quantitatively detecting soybean rust from multispectral images”, ELSEVIER-Science Direct, Vol.22, No.4, pp.186193, 2010.

[5] Rittika Raichaudhuri and Rashmi Sharma. ”On Analysis of Wheat Leaf Infection by Using Image Processing” Proceedings of the International Conference on Data Engineering and Communication Technology, Advances in Intelligent Systems and Computing. 2016;1; 978- 981.

[6] Savita N. Ghaiwat, ParulArora, “Detection and Classification of Plant Leaf Diseases Using Image processing Techniques: A Review”, International Journal of Recent Advances in Engineering and Technology, ISSN: 2347-2812, Volume 2, Issue 3, 2014.

[7] Prof. Sanjay B. Dhaygude and Mr. Nitin P. Kumbhar, “Agricultural plant Leaf Disease Detection Using Image Processing” International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, Volume 2, Issue , 2013.

[8] M. Egmont-Petersen, “ Image processing with neural networks”, Elsevier, Volume 35, Issue 10, October 2002, Pages 2279–2301 2002.