UNIVERSITY OF MUMBAI

PROJECT REPORT ON

**PLANT DISEASE PREDICTION**

PROJECT

SUBMITTED BY

ASHISH RAMJANAM MALLAH

UNDER THE GUIDANCE OF

PROF. AQUILA SHAIKH



LATE BHAUSAHEB HIRAY SMARNIKA SAMITI TRUST  
**HIRAY GROUP OF INSTITUTES**

MUMBAI - 400051

MAHARASHTRA

MCA SEM III [ 2021-2022]



LATE BHAUSAHEB HIRAY S.S. TRUST’S INSTITUTE OF COMPUTER APPLICATION

ISO 90012008 CERTIFIED

**S.N. 341, Next to New English School, Govt. Colony,**

**Bandra (East), Mumbai – 400051,**

**Tel: 91-22-26570892/3181**

# Date:

**CERTIFICATE**

**This is to certify that Mr. ASHISH RAMJANAM MALLAH**

**--------------------------------------------------------------Roll No. 202114**

**is a student of MCA of 2th year Semester-III has completed successfully full-semester Mini-Project of subject “PLANT DISEASE PREDICTION” for the academic year 2020 – 21.**

----------------- ---------------

**Subject In-Charge Director**

--------------------

**External Examiner**

# PROFORMA FOR THE APPROVAL PROJECT PROPOSAL

**PNR No.:-** 2017016400250142 **SEAT No.:-**202114

**Name of the Student:-** ASHISH RAMJANAM MALLAH

**Title of the Project:**- PLANT DISEASE PREDICTION

**Name of the Guide:-** Prof. AQUILA SHAIKH

**Teaching experience of the Guide:**

Is this your first submission? Yes No

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Signature of the Student** **Signature of the Guide**

Date: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Date:\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Signature of the Coordinator:** \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Date: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**ACKNOWLEDGEMENT**

I extend my deepest appreciation to my esteemed guide,

**Prof. AQUILA SHAIKH & RASHMITA PRADHAN** for providing me with the possibility to complete this project with the right guidance and advice.

Special gratitude I give to my respected head of the division

**PROF. VIKRAM PATALBANSI**, for allowing me to use the facilities available and also help me to coordinate my project Furthermore, I would also like to acknowledge with much appreciation the crucial role of faculty members on this occasion.

Last but not least, I would like to thank friends who help me to assemble the parts and gave a suggestion about the project.

# ----------------------

**Abstract**

This is end to end deep learning project in agriculture domain. Farmers every year face economics loss and crop waste due to various diseases in plants. We will use image classification using CNN and built a application using which a farmer can take a picture and application will tell you if the plant has a disease or not.

## **TABLE OF CONTENTS**

|  |  |  |
| --- | --- | --- |
| **CONTENTS** |  | **PAGE NO.** |
| **INTRODUCTION**  1)Introduction and Importance  1.1) Topic Intro  1.2) Need and Motivation |  |  |
| **DATA SET**  2)Steps in Preparing Data for Model  2.1) Data Collection  2.2) Data Cleaning  2.3) Feature Engineering  2.4) Training a model |  |  |
| **LANGUAGE AND MODELS USED**  3.1) Python  3.2) Numpy and Pandas for data cleaning  3.3) Matplotlib for data visualization  3.4) Sklearn for model building  3.5) Jupyter notebook, google colab & pycharm as IDE  3.6) create a app using Streamlit  **3B) Models Used**  3B.1) CNN |  |  |
| **4) RESULTS AND DISCUSSIONS**  4.1) BEST SUITED CNN MODEL SCORE  4.2) DEPLOYMENT APP |  |  |
| 5) ABBREVIATIONS |  |  |
| 6) CONCLUSION |  |  |
| 7) REFERENCE |  |  |

LIST OF TABLES

|  |  |  |
| --- | --- | --- |
| CONTENT |  | PAGE N0. |
| 4.1)Plants Disease image Dataset |  |  |
| 4.2) Transformed Dataset images |  |  |

**LIST OF FIGURES**

|  |  |  |
| --- | --- | --- |
| **Contents** |  | **Page No**. |
| 1. Real Vs Predicted Plots |  |  |
| 2. Comparison Bar Plot |  |  |
| 3. Website View |  |  |

Plants Disease Prediction Project

INTRODUCTION

INTRO & STEP:\_

we are beginning end to end machine learning project or data science project for Plants Disease image classification. During this project, I will implement how projects are executed in big companies in a typical corporate environment. We will try to classify an image of my 3 plants Diseases. I am just giving an introduction and going over business requirements etc. In next one we will talk about data collection. While working on this project you will learn,

Steps:-

1) In this project we will create a Convolutional Neural Network which will be able to predict whether a plant is suffering from a disease. We will use different layers and other hyperparameters for building, training and testing this classifictaion model.We will be using tensorflow and keras for this project.

2) First we will mount our google drive on colab so that we can use the dataset directly from our drive. For this you first need to upload the data on your drive and then mount the drive on colab.

3) After mounting our drive we will locate the folder where our data is stored to use it in our colab notebook. Here you can see that I have 2 folders in my drive and 'Plant images' contains the images that we will work on.

4) Next we will import all the required libraries. As we are making a CNN model we will import all the required layers, activations, optimizers, etc.

5) Now we will observe some of the images that are their in our dataset. We will plot 12 images here using the matplotlib library.

6) After visualizing the images let us move forward and create a function which will convert the images into a numpy array. It is required because we will normalize our dataset after this.

7) Now we will convert all the images into numpy array.

8) We will also observe the number of images under different classes to see if the dataset is balanced or not

9) Next we will observe the shape of the image.

10) Checking the total number of the images which is the length of the labels list.

11) Next we will use sklearn train\_test\_split to split the dataset into testing and training data. Here I have taken test size as 0.2 so my data will be divided into 80% training and 20% testing data.

12) Now we will normalize the dataset of our images. As pixel values ranges from 0 to 255 so we will divide each image pixel with 255 to normalize the dataset.

13) Next we will create a network architecture for the model. We have used different types of layers according to their features namely Conv\_2d (It is used to create a convolutional kernel that is convolved with the input layer to produce the output tensor), max\_pooling2d (It is a downsampling technique which takes out the maximum value over the window defined by poolsize), flatten (It flattens the input and creates a 1D output), Dense (Dense layer produce the output as the dot product of input and kernel).

14) While compiling the model we need to set the type of loss which will be Binary Crossentropy for our model alongwith this we also need to set the optimizer and the metrics respectively.

15) Next we will split the dataset into validation and training data.

16) Fitting the model with the data and finding out the accuracy at each epoch to see how our model is learning. Now we will train our model on 10 epochs and a batch size of 128. You can try using more number of epochs to increase accuracy but here we can see that the model has already raeched a very high accuracy so we don't need to run it for more. During each epochs we can see how the model is performing by viewing the training and validation accuracy.

17) Saving the model using different techniques.

18) Next we will plot the accuracy of the model for the trainig history.

19) Evaluating the model to know the accuracy of the model.

20) Next we will use our model to predict predicting the testing dataset label.

21) Printing out the original and the predicted label.

**REQUIREMENTS**

* **Hardware Requirement:**
* Processor –Core i3
* Hard Disk – 160 GB
* Memory – 1GB RAM
* **Software Requirement:**
* Windows 7 or higher
* Python
* python Streamlit server

**3.LANGUAGE & MODELS USED:-**

**Technology and tools wise this project covers:-**

1. Python
2. Numpy and Pandas for data cleaning.
3. Matplotlib for data visualization.
4. Tensorflow and keras for model building using CNN.
5. Jupyter notebook, Google Colab and pycharm as IDE.
6. Using Python Streamlit server.

* **Advantages**
* Saves time
* Easy to access the system anywhere and anytime.
* **Limitation**
* Requires an active internet connection.
* **Application**
* This system can be used by the multiple peoples to get the counselling sessions online.
* **Modules:**

The system comprises of 3 major modules with their sub-modules as follows:

1. **Admin:**
   * **Add Image:** Admin can dray n drop image.
   * **View Image:** Admin can View the added image.
   * **Check Label with Class:** Admin can see Class and Label of the classify image predicted.
2. **User:**
   * **View Image:** User can view the image.
   * **View Classification Image:** User can view the predicted image with Disease classification.

DATA COLLECTION:-

There are 3 different ways of collecting data for our project,

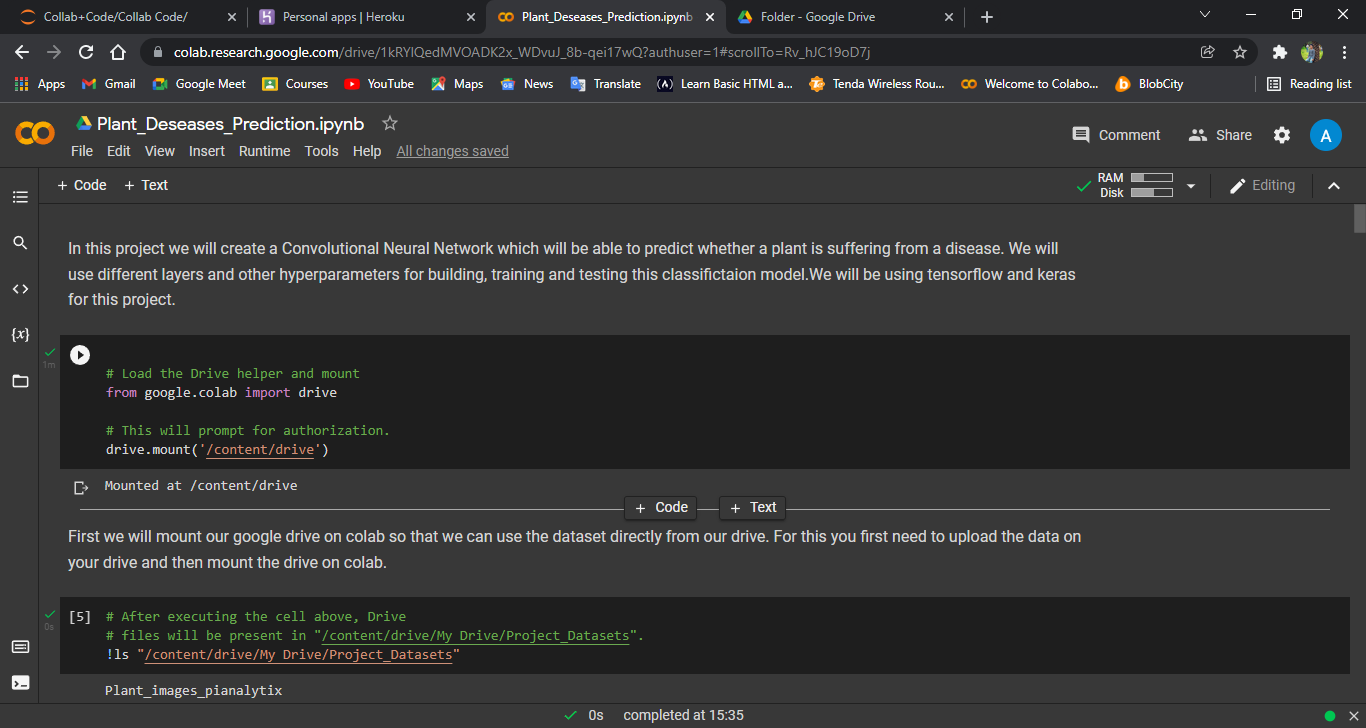
(1) Manually download images from google images

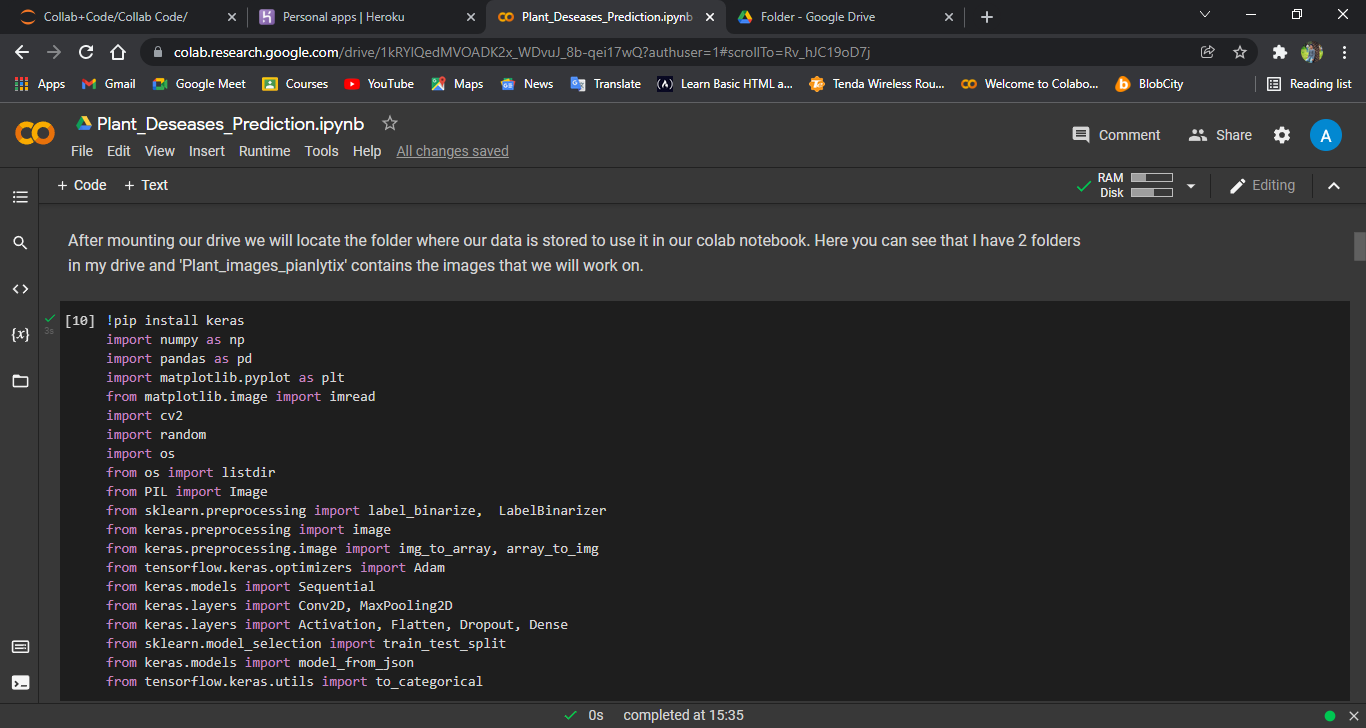
(2) Use python and web scrapping to automate downloading images from google

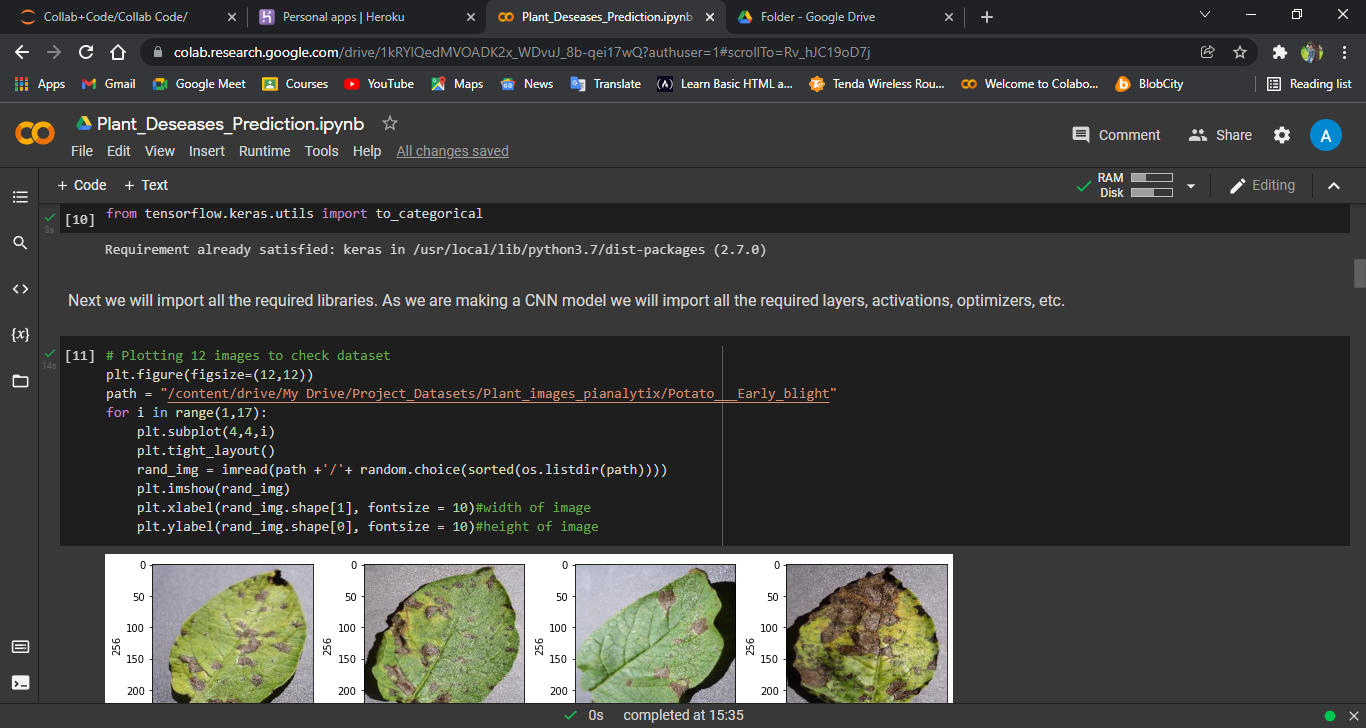
(3) Use a chrome extention called fatkun. You can download that from here: https://chrome.google.com/webstore/de...

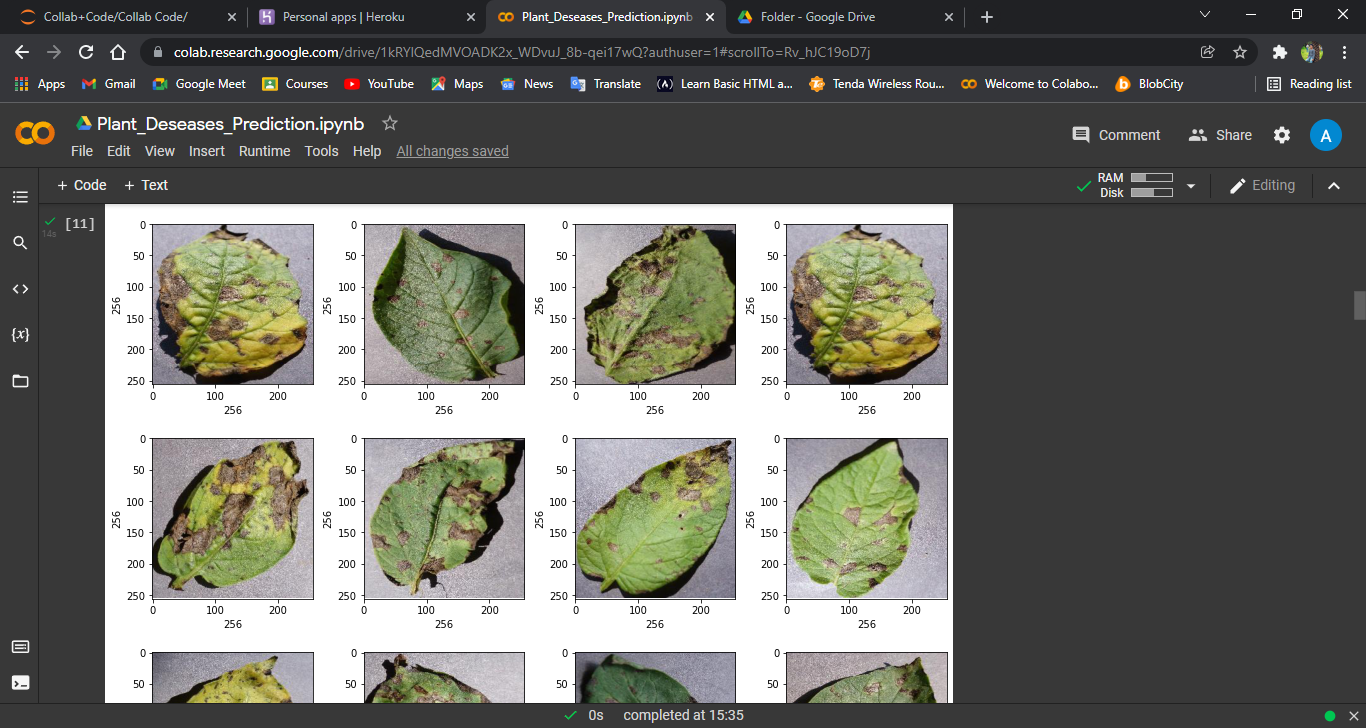
4) using kaggle datasets.

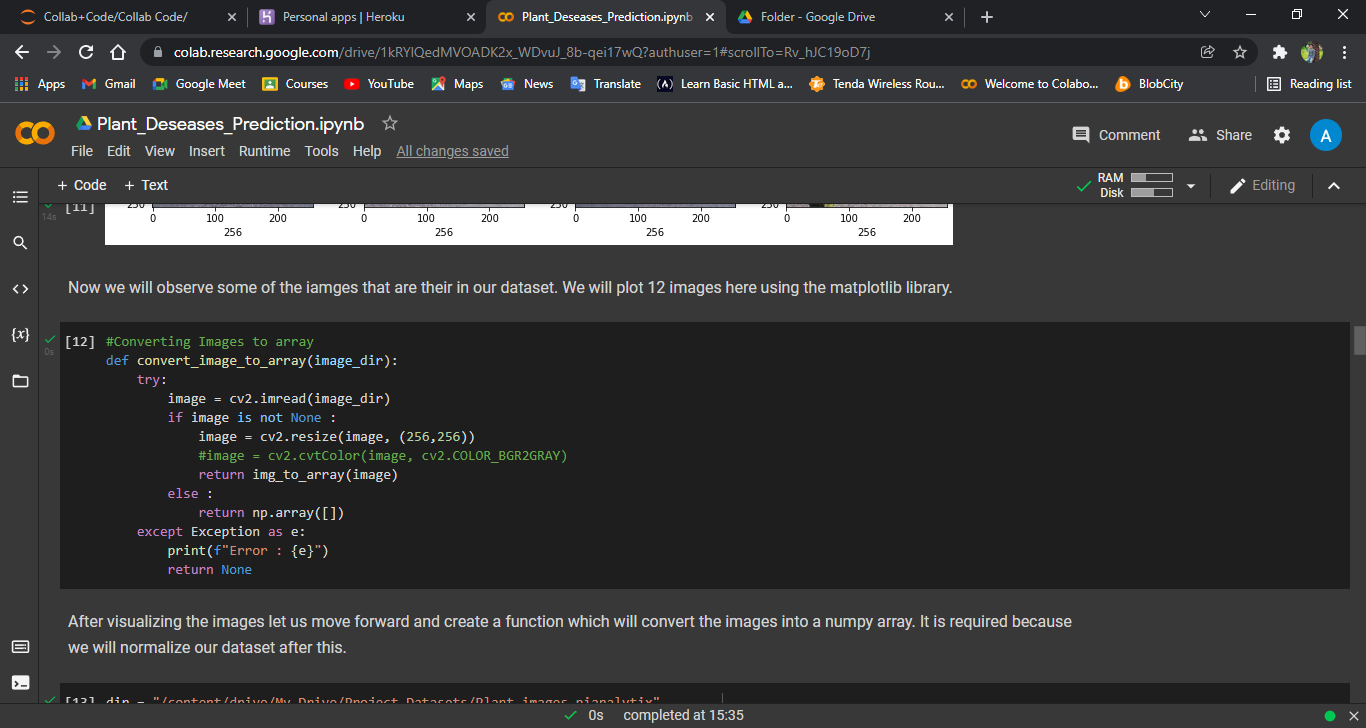
**Google Colab ScreenShot:-**

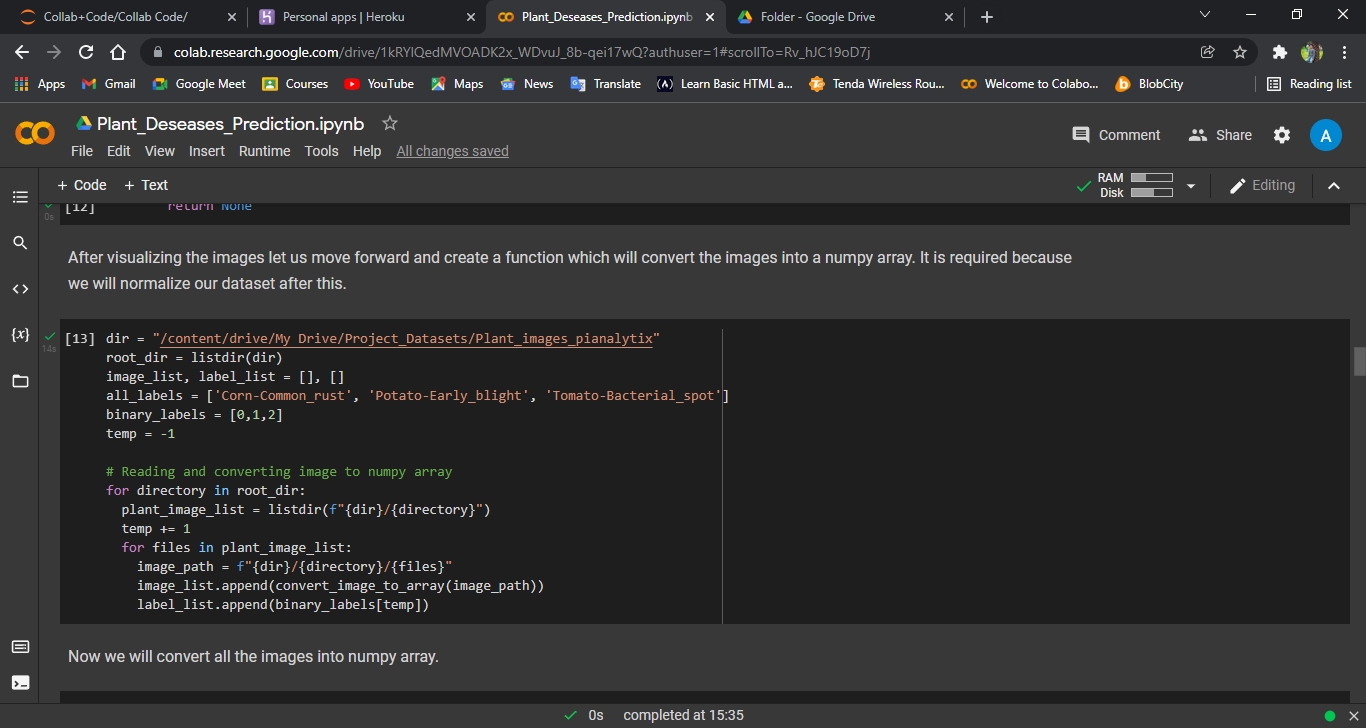


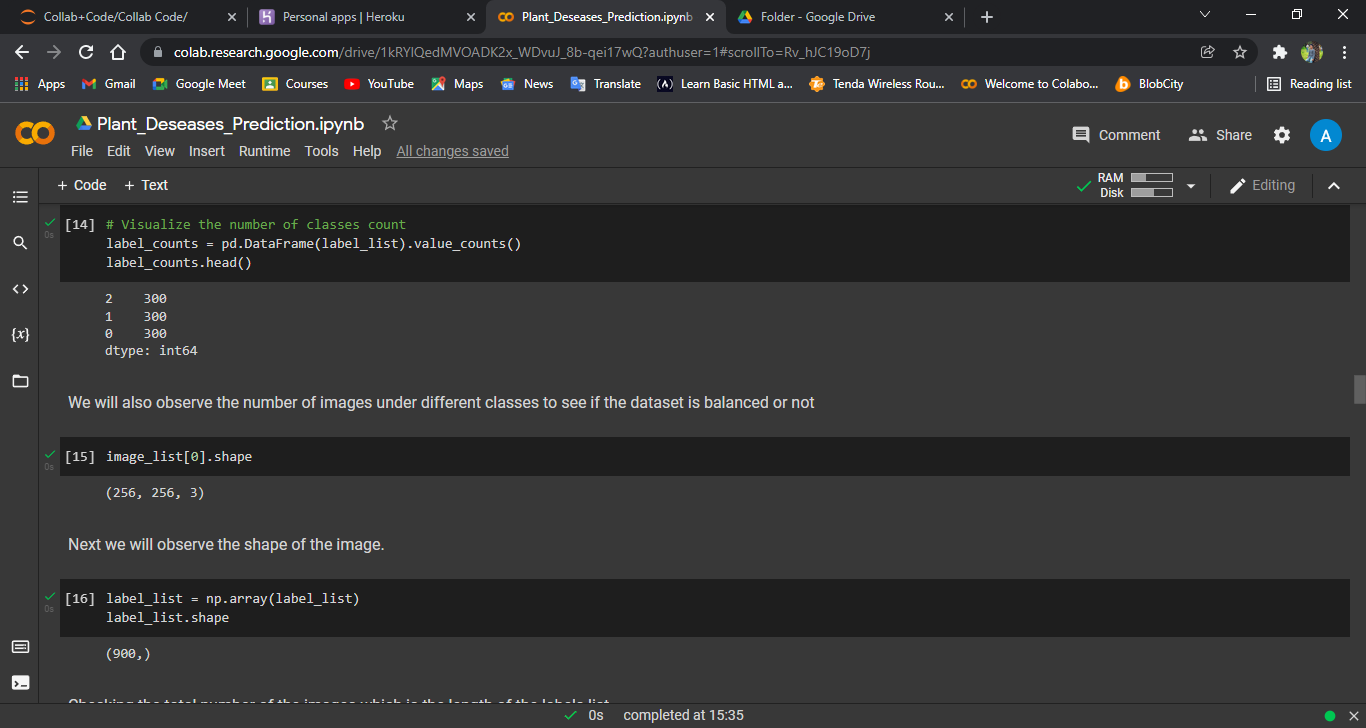


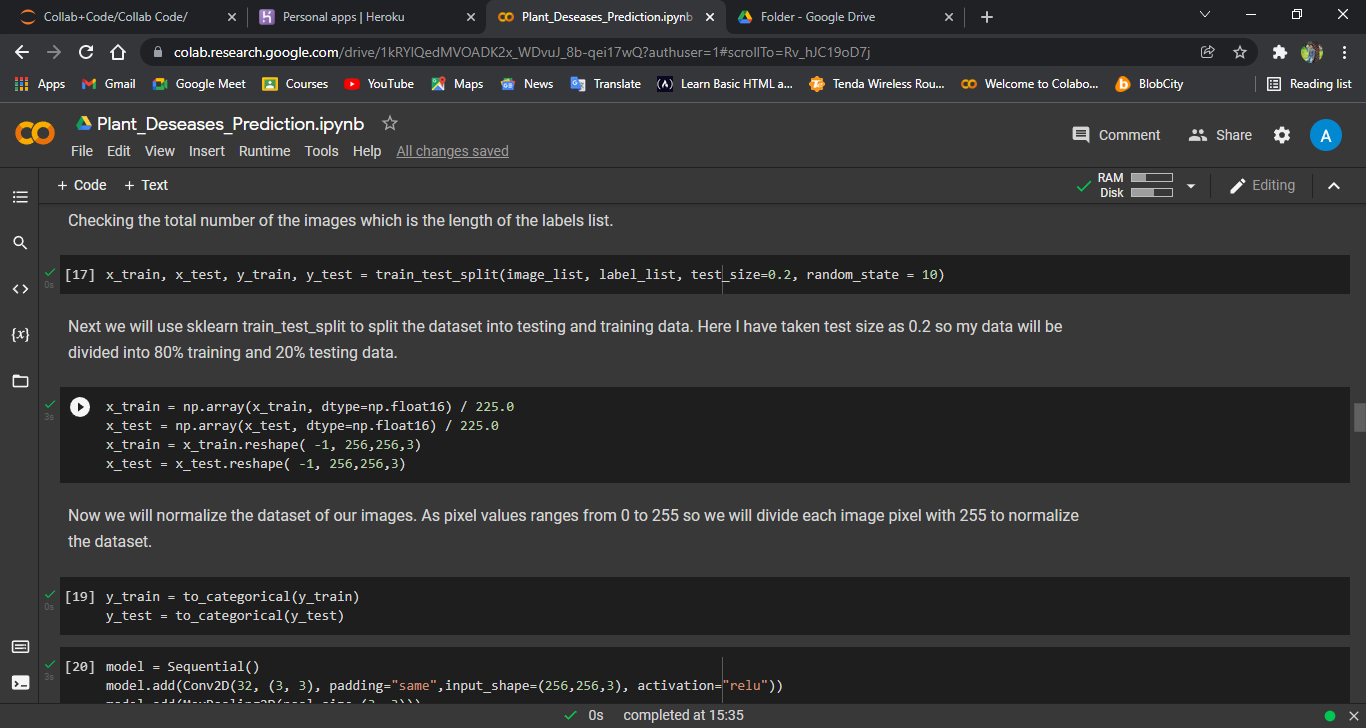


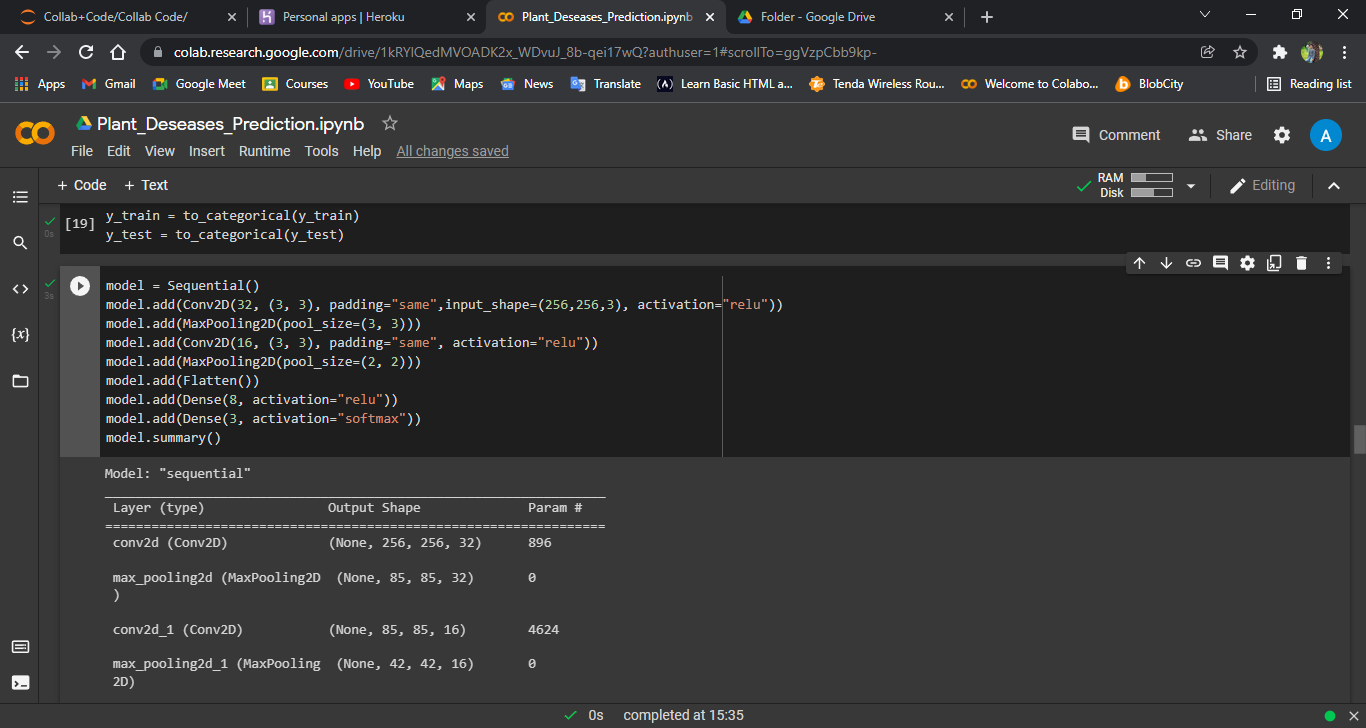


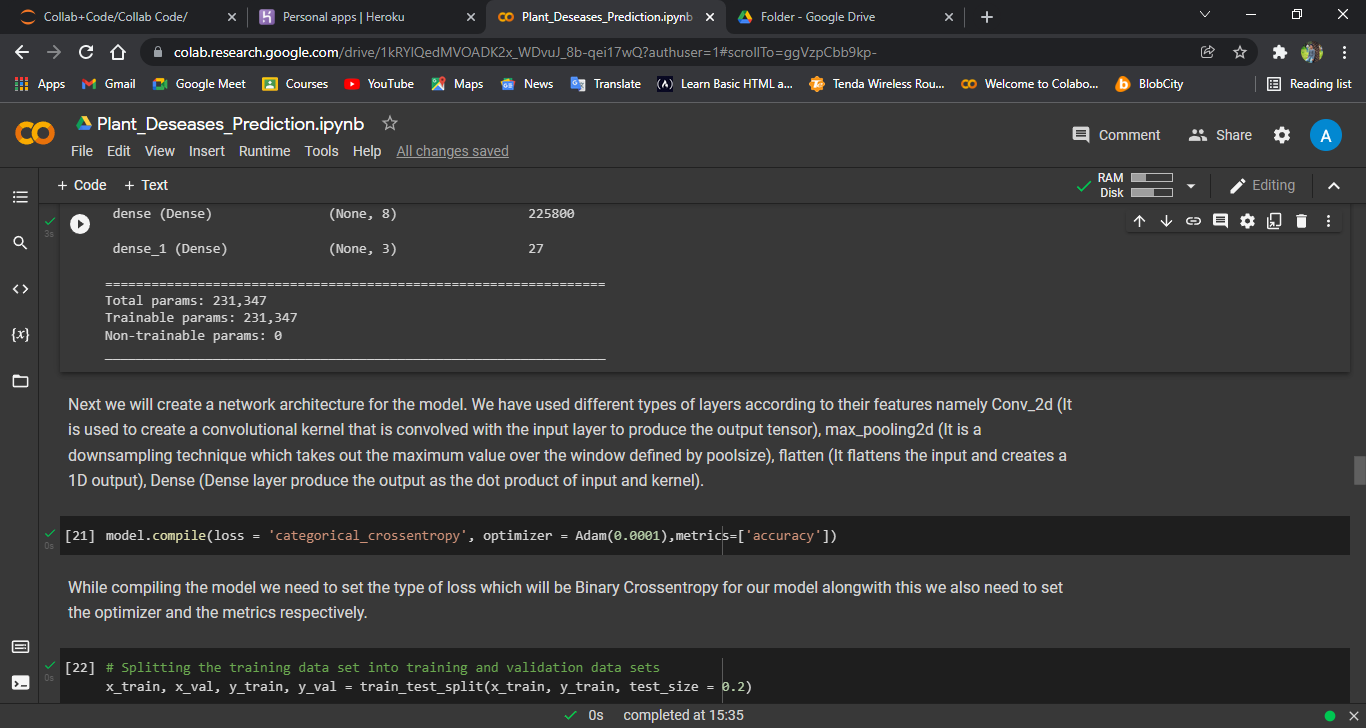


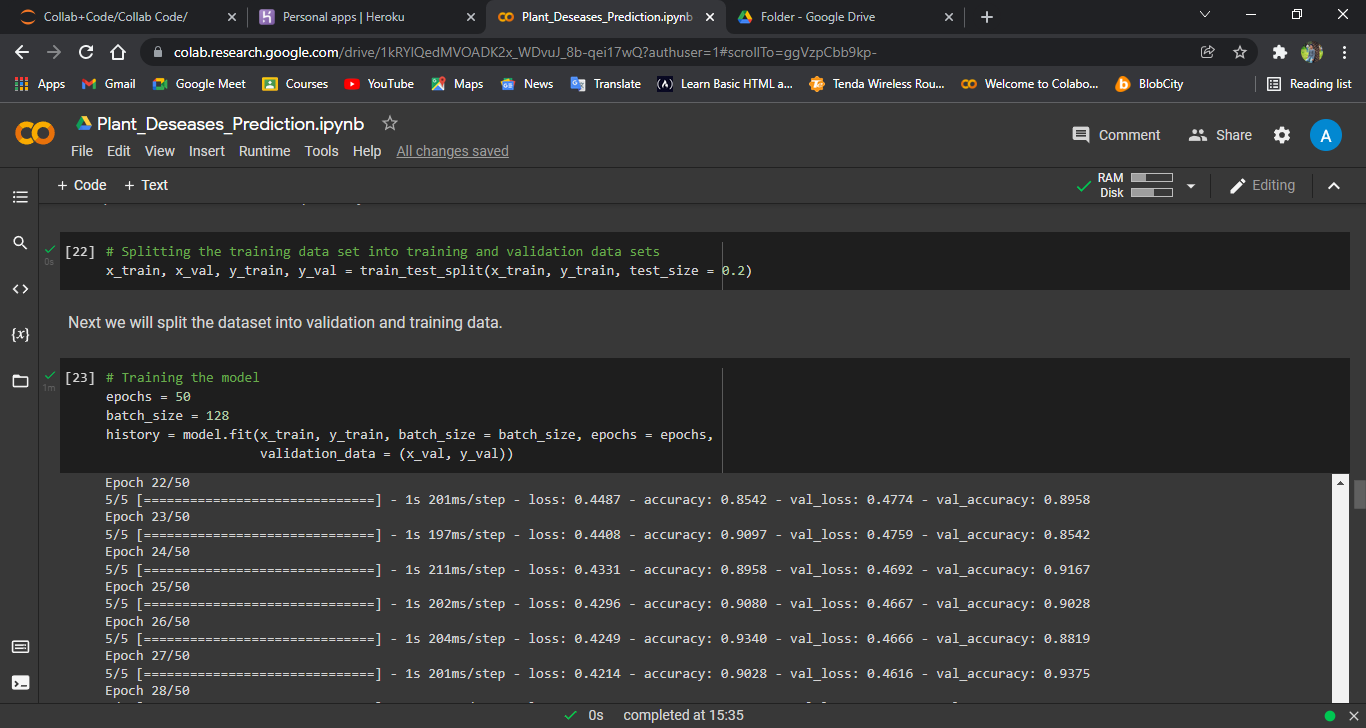


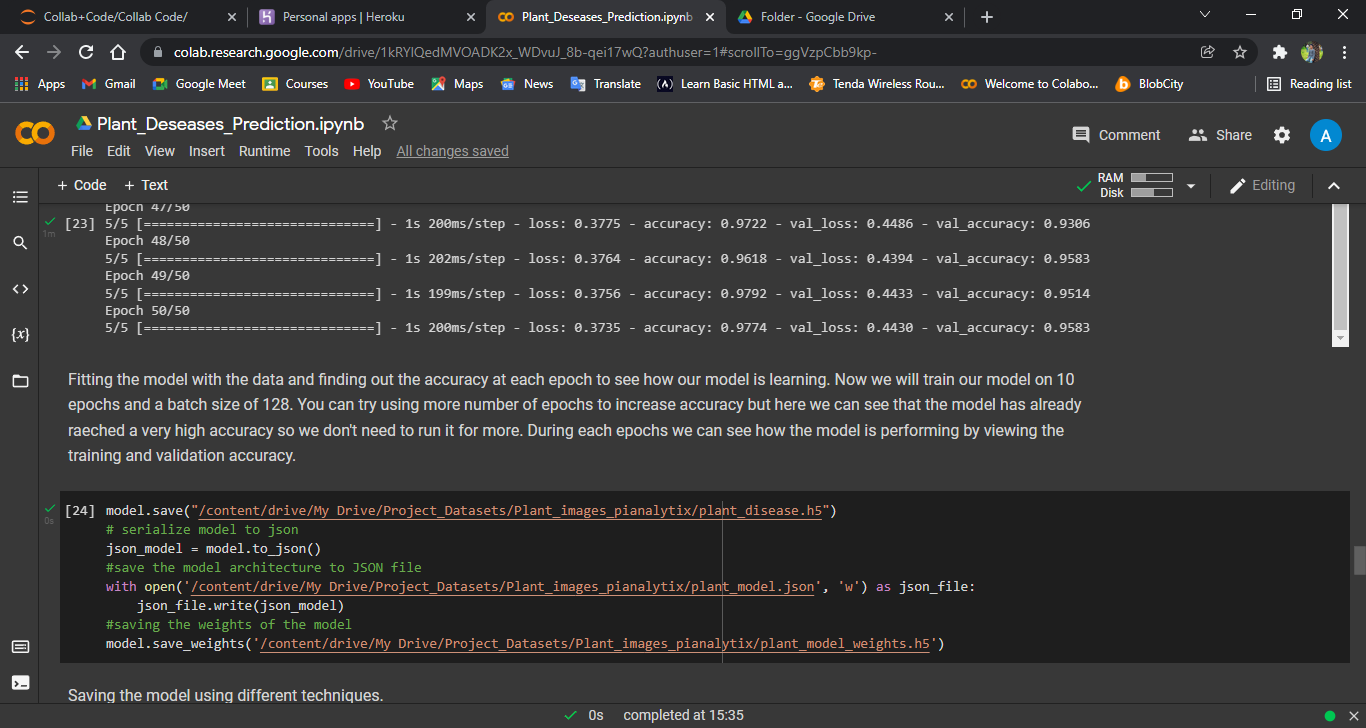


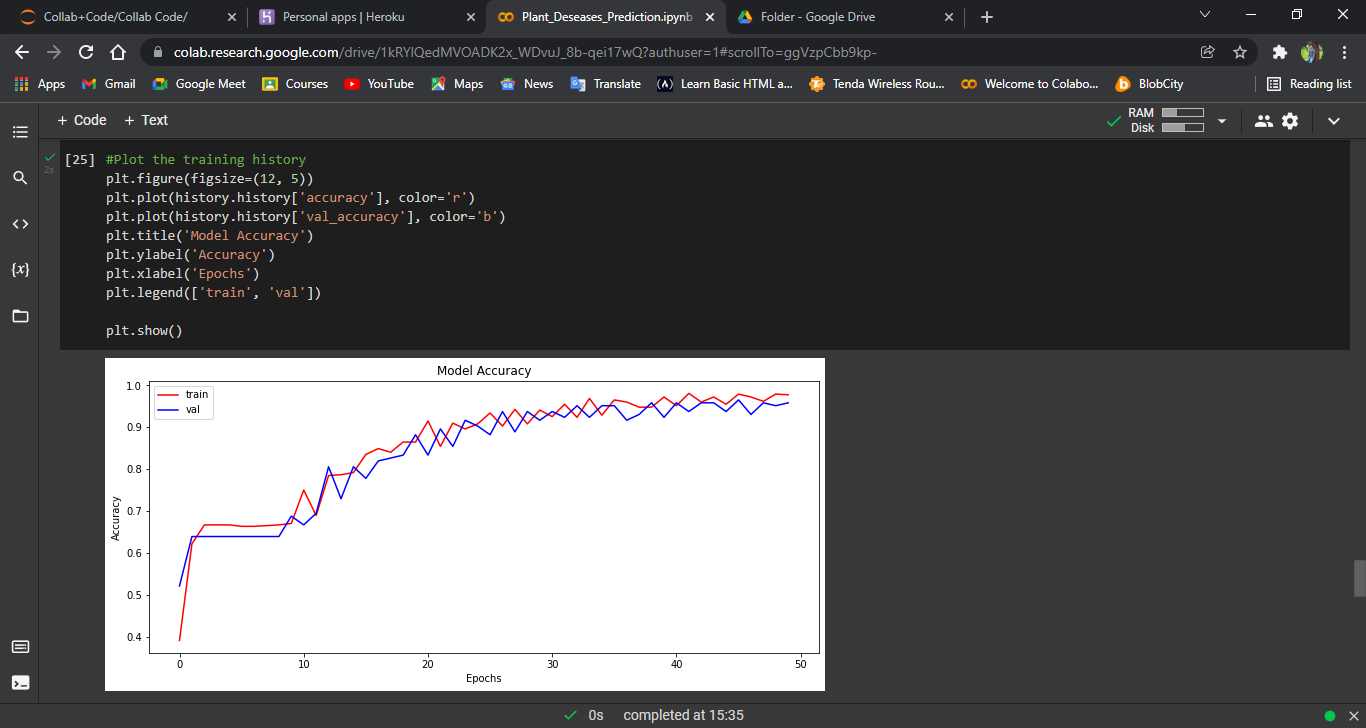


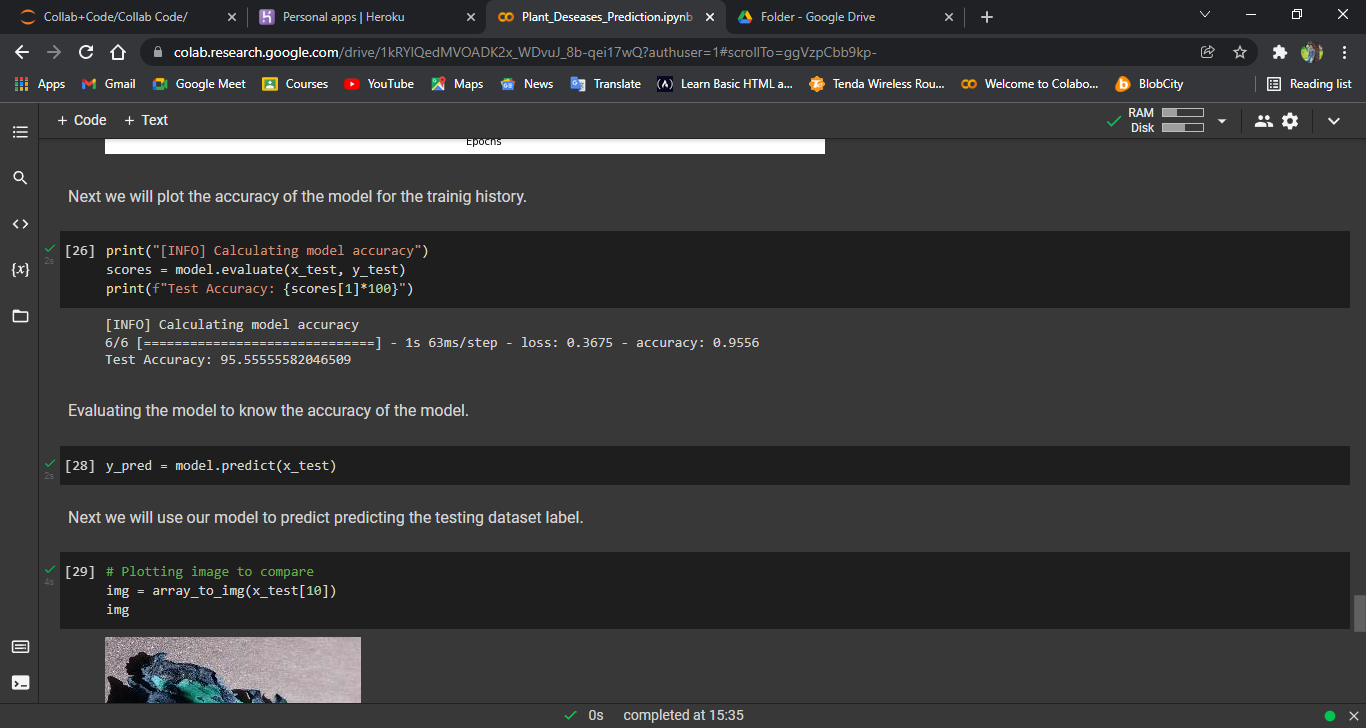


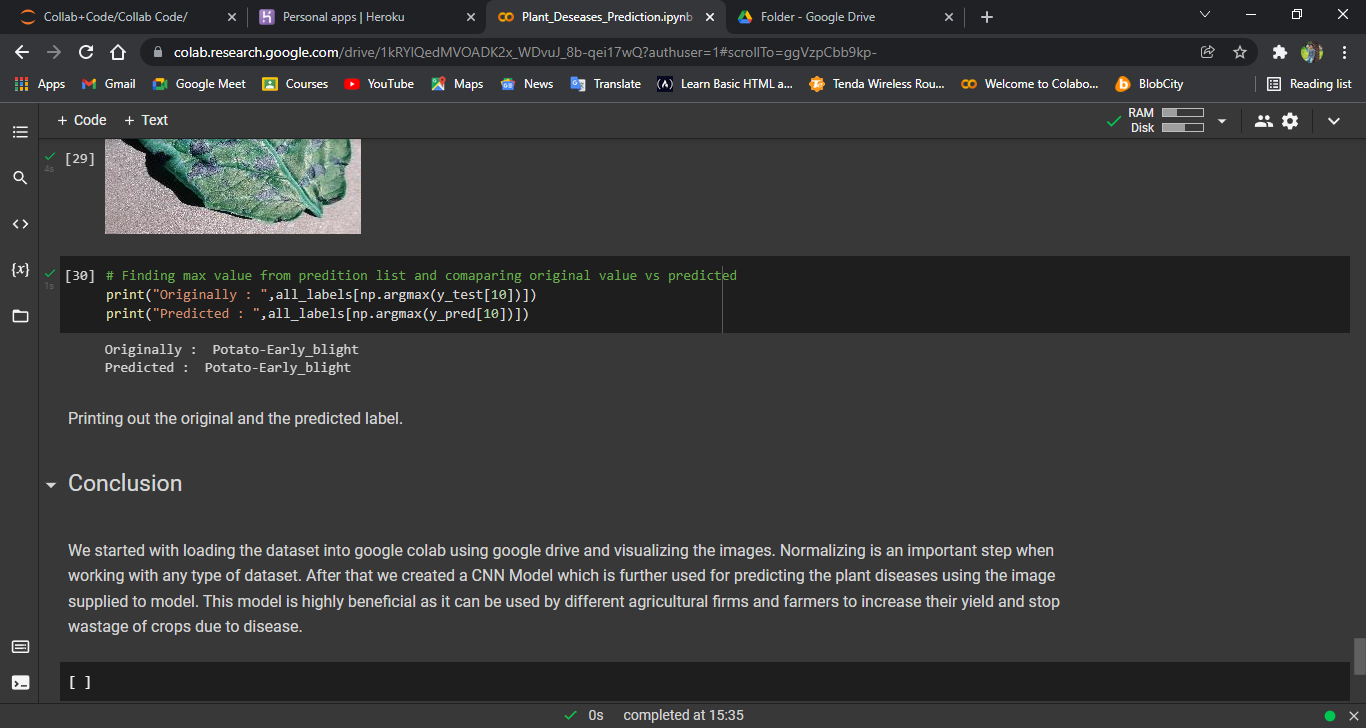






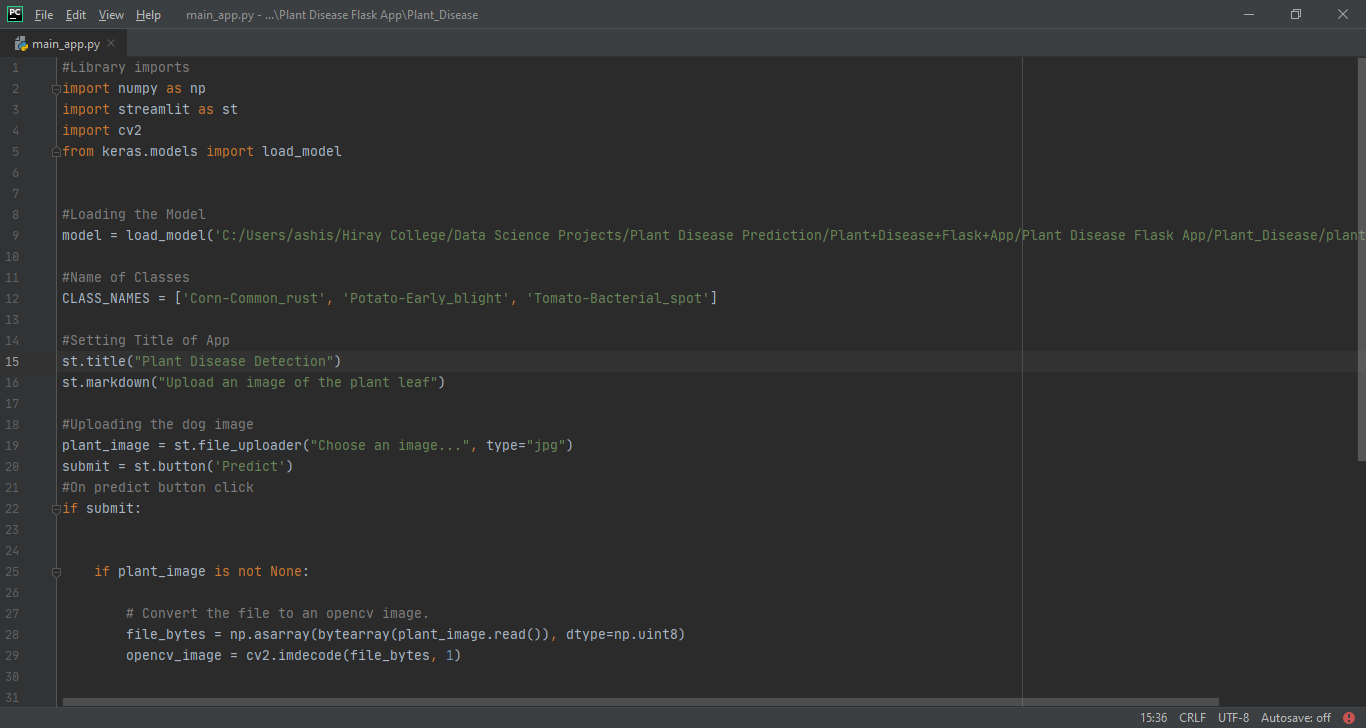


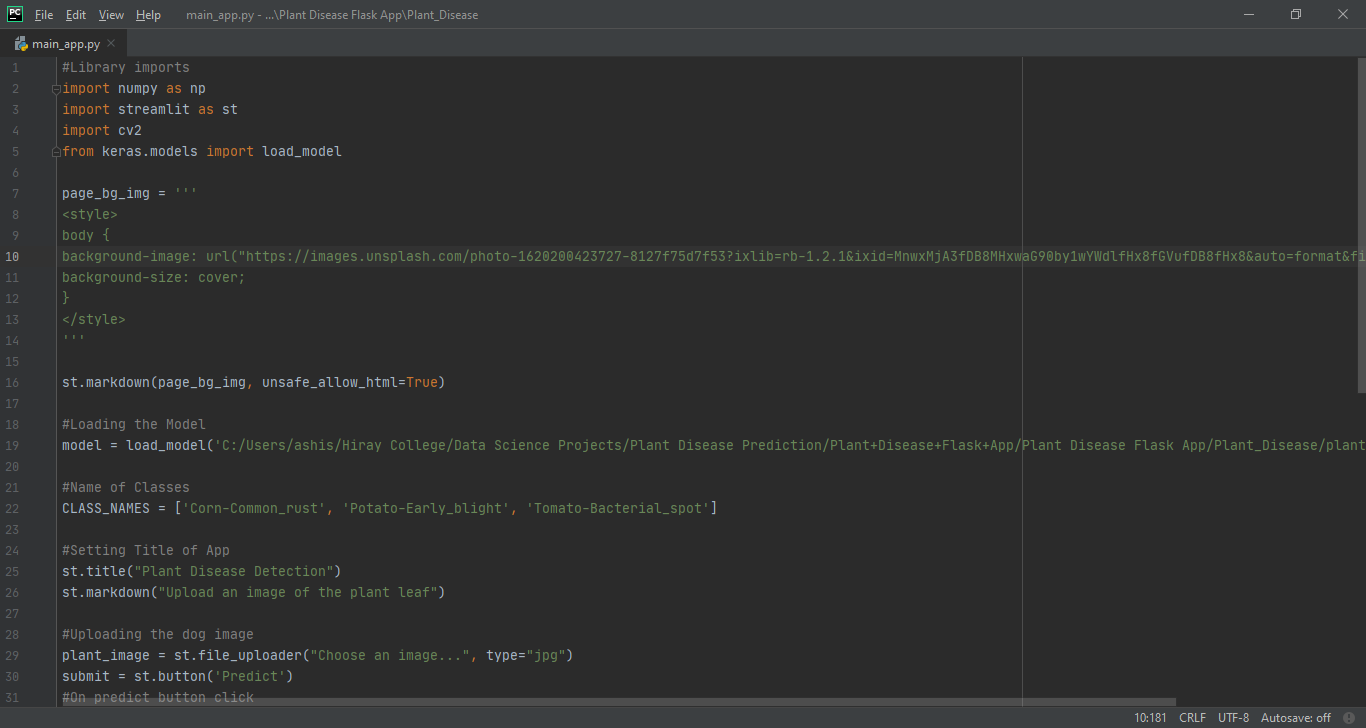


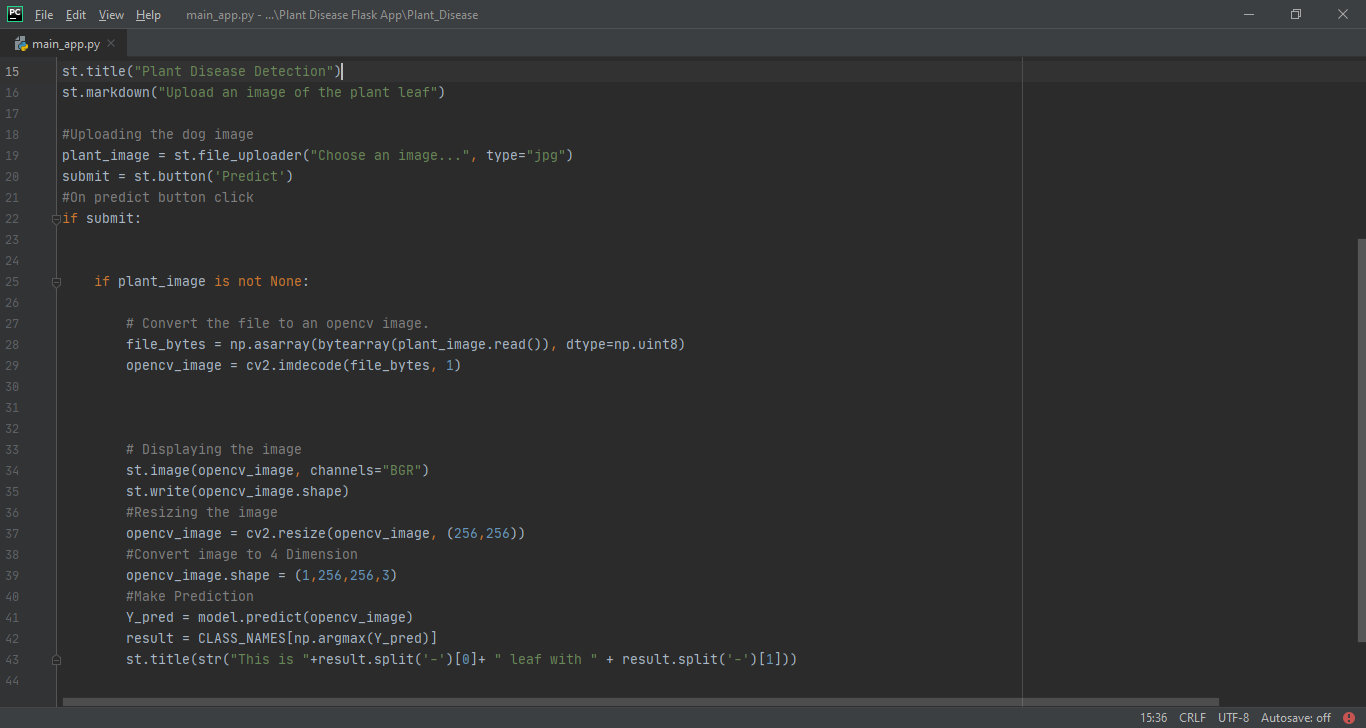


PyCharm Screenshot:-

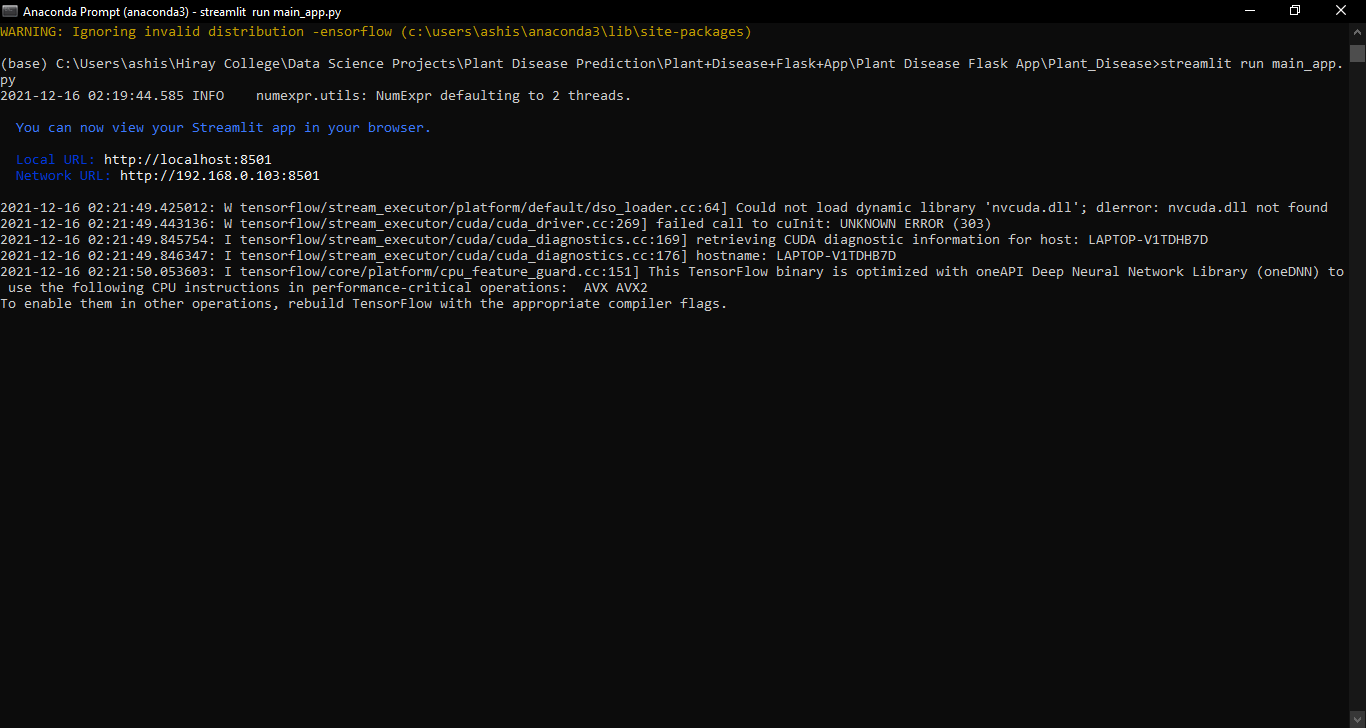
StreamlitServer:-







Anaconda Command Prompt to run streamlit:-

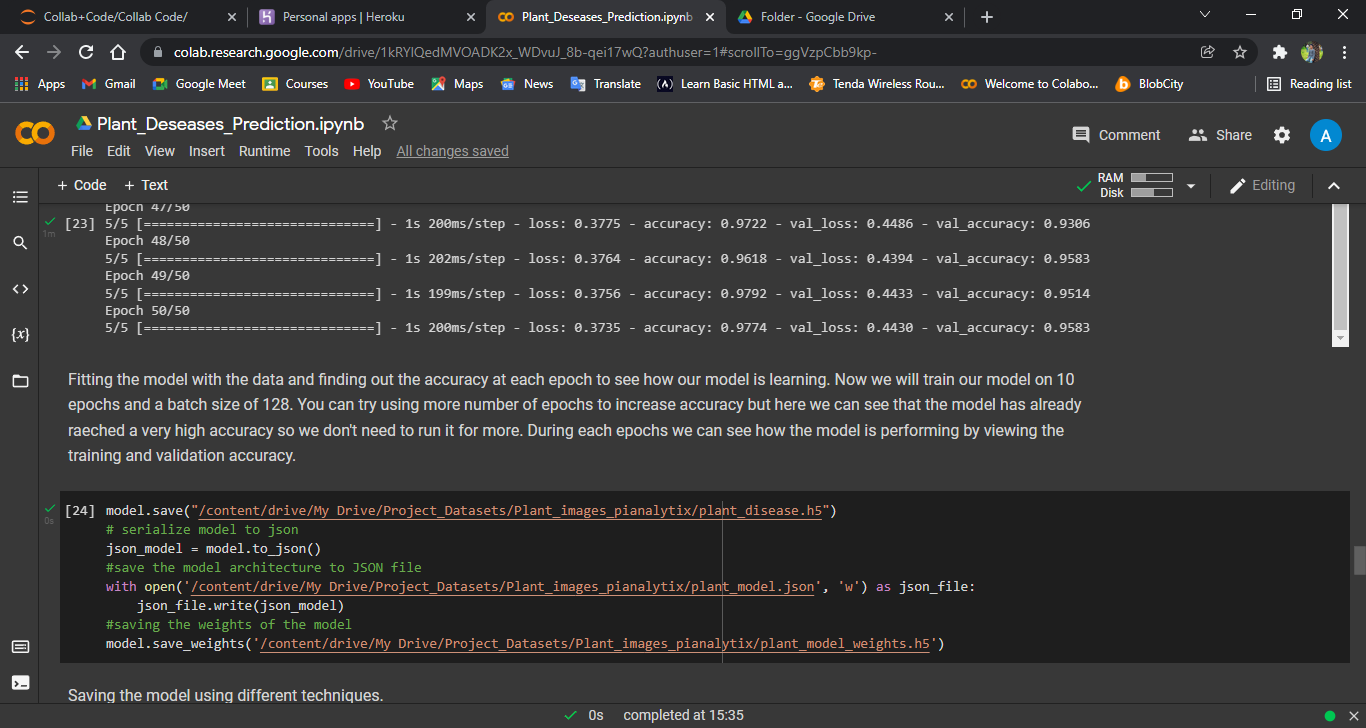


**RESULTS AND DISCUSSIONS**

**Best Suited Model**

So, our study showed that……..

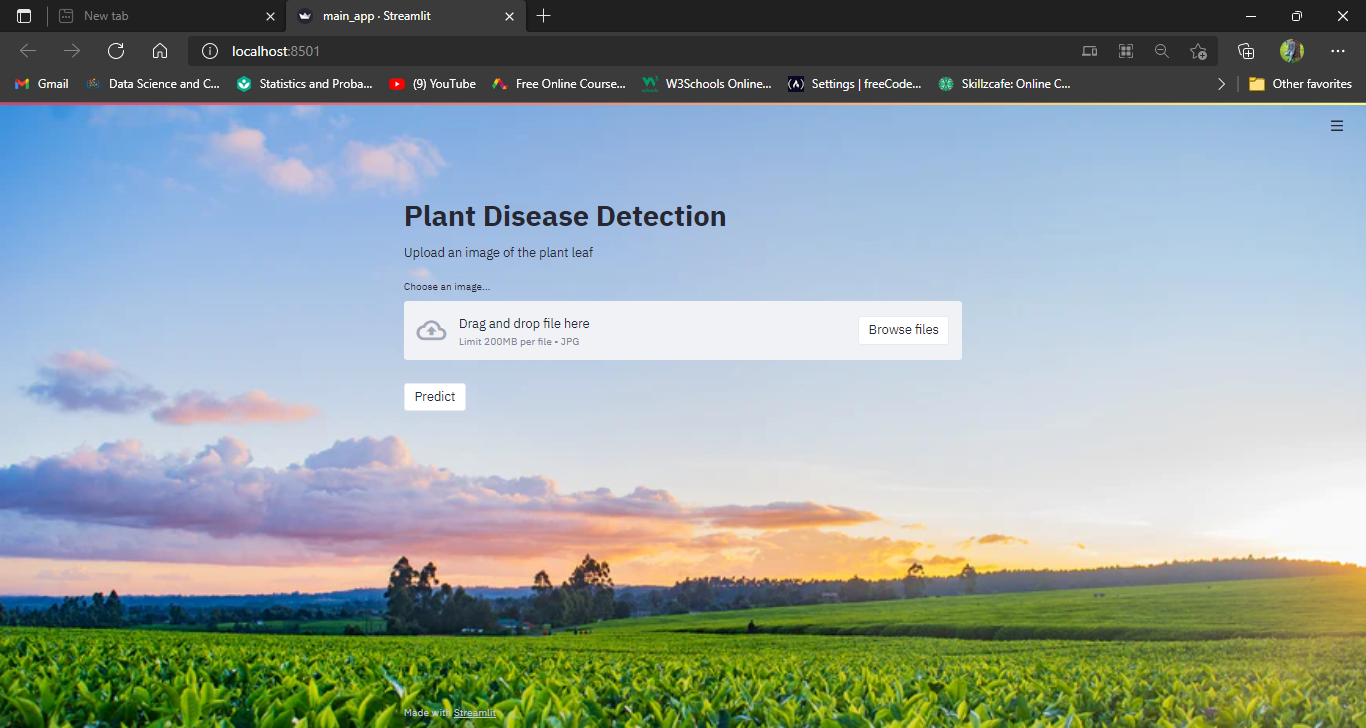
Convolutional Neural Networks displayed the best performance for this Dataset and can be used for deploying purposes.

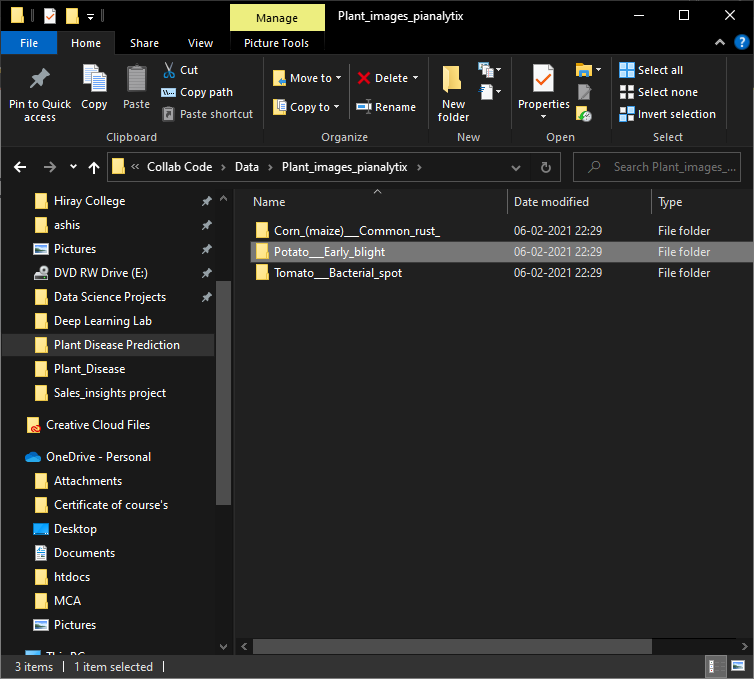


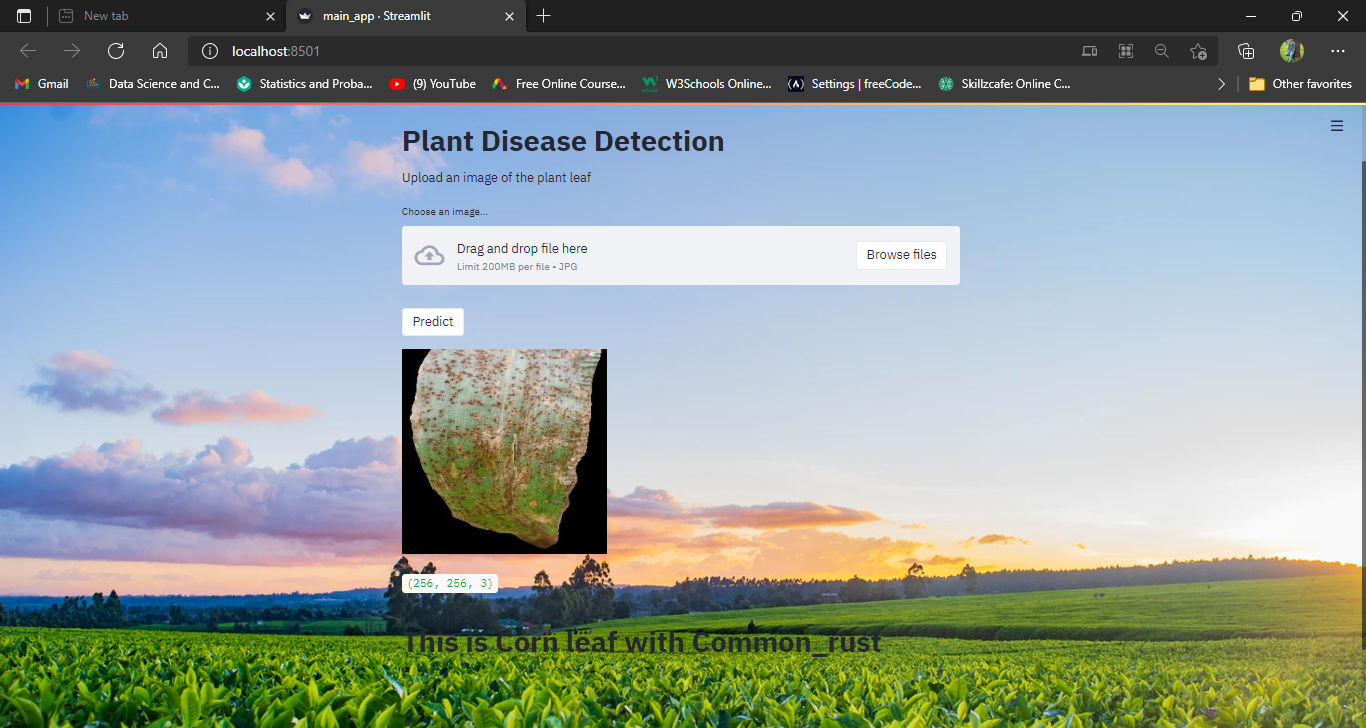
**Deployment App**

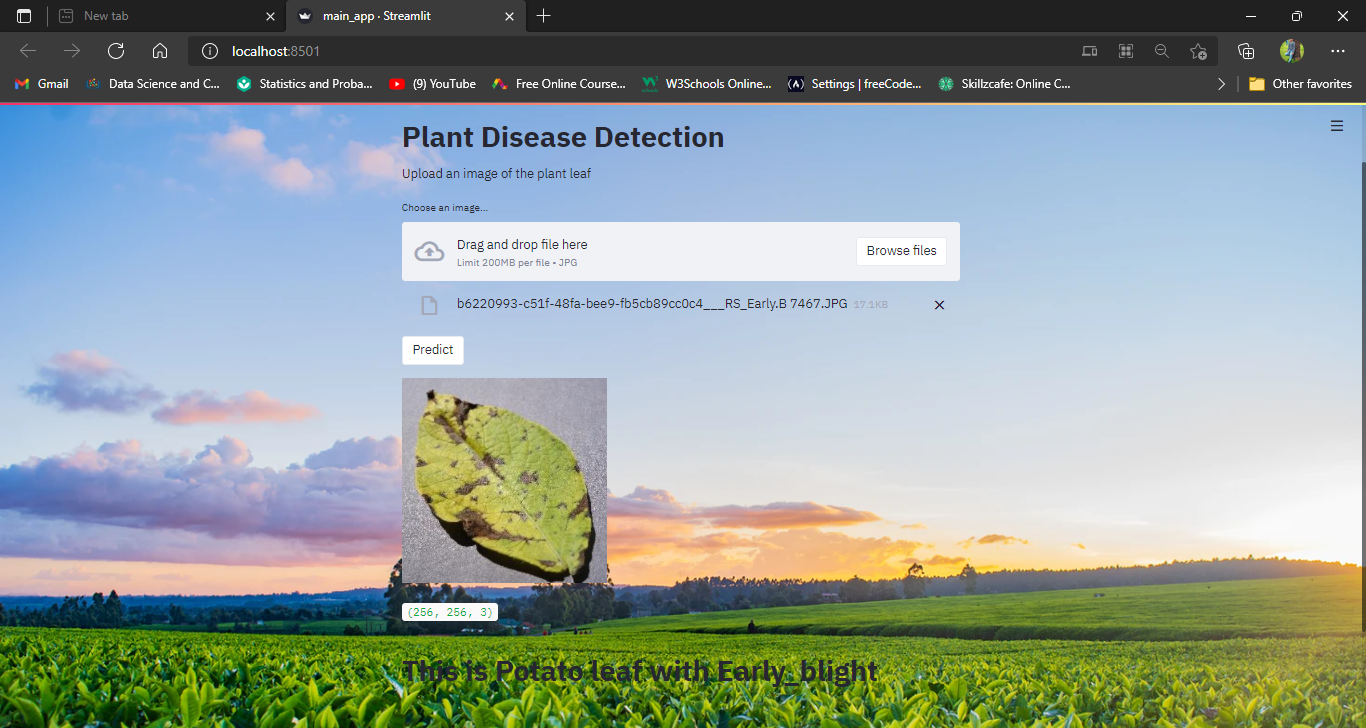
The Model is deployed through Python Web App Streamlit

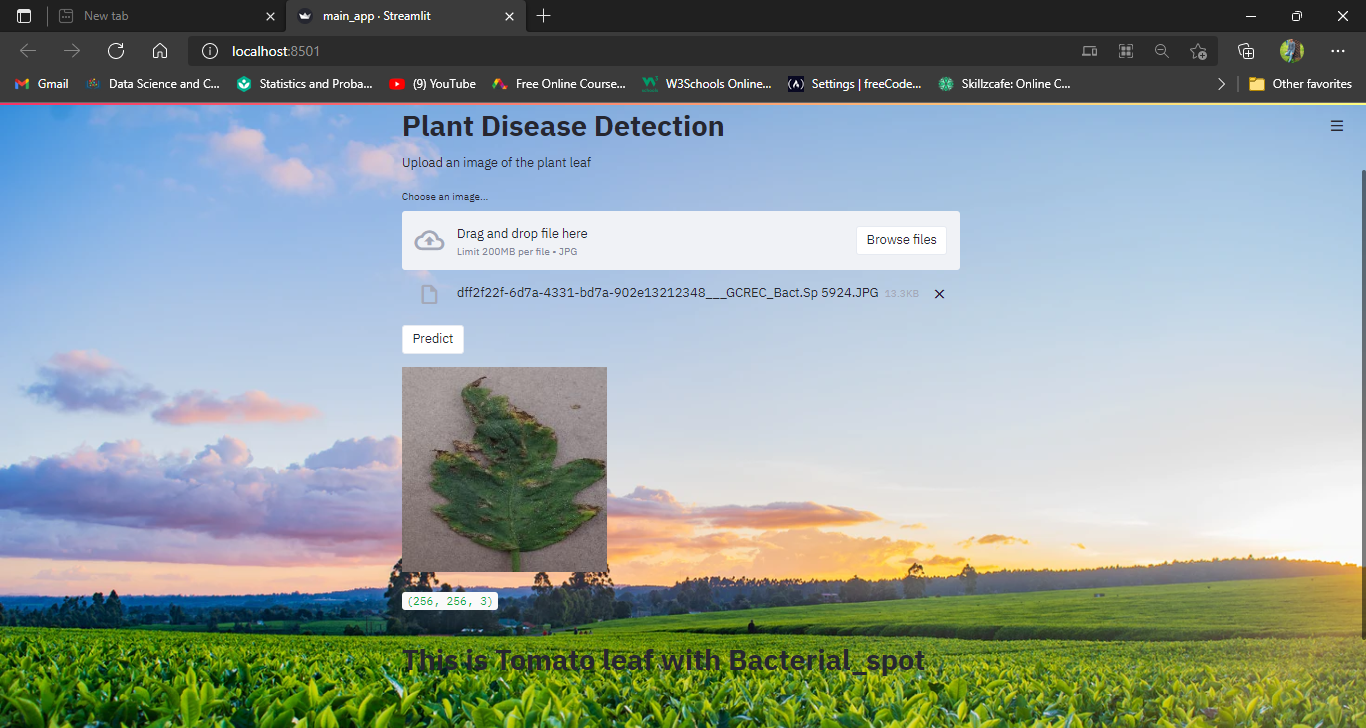
Project Output ScreenShot:-











==================================================================

CONCLUSION

We started with loading the dataset into google colab using google drive and visualizing the images. Normalizing is an important step when working with any type of dataset. After that we created a CNN Model which is further used for predicting the plant diseases using the image supplied to model. This model is highly beneficial as it can be used by different agricultural firms and farmers to increase their yield and stop wastage of crops due to disease.

* Model for our Dataset with BEST ACCURACY of 0.95.
* **Bibliography**
* [Find Open Datasets and Machine Learning Projects | Kaggle](https://www.kaggle.com/datasets)
* [Towards Data Science](https://towardsdatascience.com/)
* [RxJS, ggplot2, Python Data Persistence, Caffe2, PyBrain, Python Data Access, H2O, Colab, Theano, Flutter, KNime, Mean.js, Weka, Solidity (tutorialspoint.com)](https://www.tutorialspoint.com/index.htm)
* [Learn R, Python & Data Science Online | DataCamp](https://www.datacamp.com/)
* [YouTube](https://www.youtube.com/)
* <https://www.khanacademy.org>
* Analyticsvidhya.com
* Machinelearningplus.com
* Simplilearn.com
* [www.tutorialpoint.com](http://www.tutorialpoint.com)

=========================================================