

**Plant Disease Prediction Using Machine Learning**

**Capstone Project**

A Project Submitted by

M.Harshith Satya Krishna-*19BEV7022*

P.Rakesh Teja-*19BEV7023*

**Under the guidance of**

**Dr. Agam Das Goswami**

**Assistant Professor Sr, ECE, VIT-AP**

**Index**

|  |  |  |
| --- | --- | --- |
| **Chapter No.** | **Title** | **Page.NO** |
| **1** | Abstract | 1 |
| **2** | Introduction | 2-3 |
| **3** | Literature Survey | 4-5 |
| **4** | Proposed Methodology | 5-6 |
| **5** | Block Diagram | 7 |
| **6** | Result | 8-17 |
| **7** | Proposed code and Final Output | 18-21 |
| **8** | Conclusion and Future scope | 21-22 |
| **9** | References | 22-23 |

**Abstract**

Agriculture productivity is a key factor in economic growth. This is one of the reasons that plant disease detection is crucial in the sector of agriculture, as the presence of illness in plants is extremely common. If necessary precautions are not followed in this region, plants suffer major consequences, which have an impact on the quality, quantity, or productivity of the corresponding products. The use of an automatic method for plant disease detection is advantageous because it reduces the amount of time it takes to monitor large crop farms and may identify disease signs at their earliest stage, when they first emerge on plant leaves. A method for picture enhancement is presented in this work. It also includes an overview of several disease categorization methods that may be applied to the identification of plant leaf diseases. Then we build a model using CNN by using algorithm ‘VGG19’ to train the data. We also use a validation data Set. We used a large data set of 87,900 Files of the leafs to analyze which contains 88000 images .The results show that the accuracy of the method is 82.50%, which is better than the traditional method Therefore, the deep learning algorithm proposed in the paper is of great significance in intelligent agriculture, ecological protection, and agricultural production.

***Keywords: Plant disease, data preprocessing, machine learning, CNN model, VGG-19 model.***

**Introduction**

Agriculture is considered as the primary source of food production and it is the basic foundation in developing countries which creates opportunities to raise the country’s economy. Around 70 percent of its rural households in India still depend primarily on agriculture for their livelihood, with 82 percent of farmers being small and marginal. Agricultural productivity has become the foundation of the Indian economy. Pest analysis demands the statistical study of vast amounts of data to determine the association of multiple components in order to acquire the guideline for protection. Hand identification techniques have a excess of issues such as being only applicable to limited size plantations. Because of this reason, detection of disease in the plants has become a major role. Initially it is necessary to detect the disease in the plant. In conclusion, certain plant diseases may be quite similar to one another, making classification considerably more difficult and time consuming. Therefore, automated plant identification is a challenging and rising issue that has received increasing attention in recent years, especially for identification based on leaf image analysis. The ultimate goal of optimization in this scenario is to reduce categorization time and do away with the need for human specialists to handle enormous estimated lists of plant species.

Deep learning algorithms are increasingly being used in place of more traditional methods in many excellent machine learning equipment these days. Deep learning techniques done deal with the requirement for manual feature extraction by having the models automatically extract features. Furthermore, classification results are substantially better than those obtained using traditional approaches. The most often utilized characteristic for building an automated plant identification system is leaf shape. The main idea behind creating the dataset is that the leaf may provide information beyond its shape, such as its vein system. Many Deep Learning concepts have been applied to the agricultural area in recent years to solve problems, Including fruit detection, plant leaf categorization. To design a plant disease detection system, pictures of the other parts of the plant can also be taken. But the most common and easiest portion of a plant to detect sickness of a particular plant is its leaves. As a result we have used the leaves as a sample in this study to identify diseased crops.

This paper proposes a system based on Deep Learning for detecting and classification of plant disease. To evaluate performance using a minimum memory-efficient interface, we employed the model VGG19.Depending on their level of health and sickness, images of plant leaves from 33 different crops were divided into 38 distinct groups. The following point discuss the contribution of this work:

* The suggested technique uses a transfer learning principle by fitting the data into a CNN model and classifying the leaves according to the illness.
* We used Stream lit a local host to open a website where we can upload the image and predict the image disease

**2. Literature Survey**

**Table 1- Literature review analysis**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Sr.NO** | **Author** | **Technique** | **Advantages** | **Disadvantages** |
| **1** | Shima Ramesh,Mr. Ramachandra Hebbar[1] | HOGe (Histogram of an Oriented Gradient),CNN | The advantage of HoG feature extraction is that it operates on the cells created. | Only through color the leaf is determined. This technique may not find all the leaf’s. |
| **2** | Asma Akhtar, Aasia Khanum, Shoab A. Khan, Arslan Shaukat[2] | KNN,Decision Tree ,Naïve Bayes,RNN,SVM, | The proposed approach of combining DCT+DWT features for classification with Support Vector Machine (SVM) gives maximum accuracy of 94.45%. | It might Take long time to compare different algorithms. |
| **3** | Rehan Ullah Khan , Khalil Khan , Waleed Albattah , and Ali Mustafa Qamar [3] | Scale Invariant Feature Transform (SIFT) to find the texture information, CNN | This method achieved an accuracy of 95%. | Requires high computation time for attribute selection. |
| **4** | Saurav Roy , Ratula Ray, Satya Ranjan Dash and Mrunmay Kumar Giri[4] | SURF (Speeded Up Robust Features),K-Nearest Neighbor (KNN) | Comparison of different algorithms  Gives more information and to learn about the techniques. | It may not show possible outcomes for many other leaf”s. |
| **5** | Budi Arianto Suryo Kusumo, Ana Heryana, Oka Mahendra, and Hilman F. Pardede[5] | Support vector machines (SVM), Decision Tree (DT), Random forest (RF), and Naive Bayes (NB) | Best methods for better accuracy . | Only for corn leaf so not applicable for other leaf”s |
| **6** | Draško Radovanović, Slobodan Đukanović [6] | Support Vector Machines (SVM), k-Nearest Neighbors (k-NN), Fully connected Neural Networks (F-CNN), | We can use an appropriate model or algorithm for better accuracy in which CNN has the highest accuracy. | It may take long time to test all the models |
| **7** | Pranesh Kulkarni , Atharva Karwande , Tejas Kolhe , Soham Kamble , Akshay Joshi , Medha Wyawahare[7] | gray level co occurrence matrix (GLCM) , Support vector machine | Accuracy levels and F1 Score is approximately equal. | Many Leaf”s are not used to verify the correct accuracy |
| **8** | R. Sujatha , JyotirMoy Chatterjee , NZ Jhanji , Sarfraz Nawaz Brohi [8] | SVM,VGG16,SGD | Compared to ML The DL has the highest accuracy by using VGG16,VGG19,Inception V3 | More dataset is required to check for required accuracy |
| **9** | P.Loganathan , Dr.R.Karthikeyan  [9] | ResNet(64,32,16,8,4) | As dimension increased ResNet 64 has highest prediction rate | Many other algorithm gives more accurate value compared to ResNet |
| **10** | Archanaa.R, Shridevi.S[10] | CNN,VGG16 | High accuracy in VGG16 which is a DL algorithm | Small Data set used |

**3.Proposed Methodology**

***3.1 Data Set and Feature Extraction:***

Importing the Dataset which contains more than 70 thousand images belonging to 38 different classes from Kaggle using API (Application Programme Interface) by giving access of the programme to the application. Generating the data from the imported Dataset to make it available for further processing. Training and testing the model on the generated data. This process is done using the most common and preferable approach i.e., 70% of the data for training and 30% for testing. Feature extraction which includes two more layers in addition to the common approach.Firstly, the RGB images of leaves are acquired. Then RGB images are converted into hue saturation value (HSV) color space representation. RGB is an ideal for color generation. But the HSV model is an ideal tool for color perception. Hue is a color attribute that describes pure color as perceived by an observer. Saturation refers to the relative purity or the amount of white light added to hue and value means amplitude of light. After the transformation process, the hue component is taken for further analysis. Saturation and value are dropped since it does not give extra information.

**3.2 CNN(convolutional Neural Networks)**

We used CNN(Convolutional Neural networks) . The CNN architecture can be used to perform this disease detection system. To perform CNN the required data should be collected using Data collection, data augmentation, pre-processing, and feature extraction. Typically, training data is used to conduct the feature extraction. Based on that data, threshold values may be set, and in the testing phase, the value of the features is compared to the trained one to determine if the picture is sick or not. Initially, 88000 images have been taken which is a large Data set. After that, By using the transfer learning technique the deep learning models are trained, and to indicate the significance of the model.

**3.3 VGG 19**

We used the VGG19 model for training the data. VGG19 is a variant of VGG model which in short consists of 19 layers. There are other variants of VGG like VGG11, VGG16 and others. You can load a pre-trained version of the network trained on more than a million images from the ImageNet database. The pretrained network can classify images into 1000 object categories, such as keyboard, mouse, pencil, and many animals. As a result, the network has learned rich feature representations for a wide range of images. The network has an image input size of 224-by-224. And after training we got an accuracy of 82.5% on our dataset.

**5. Proposed Block Diagram**

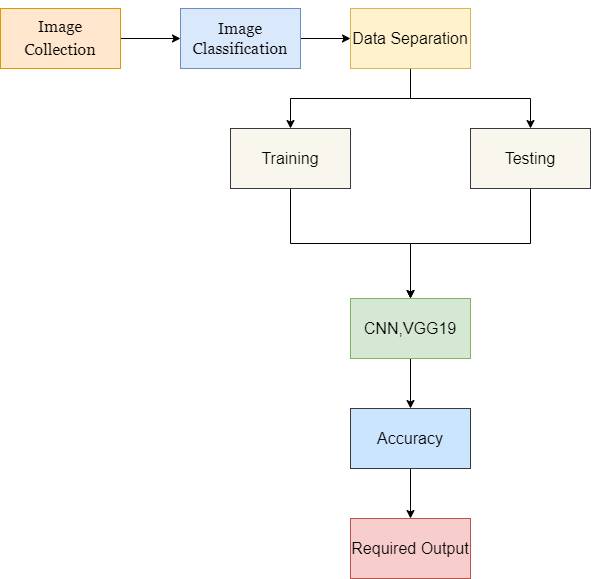
****

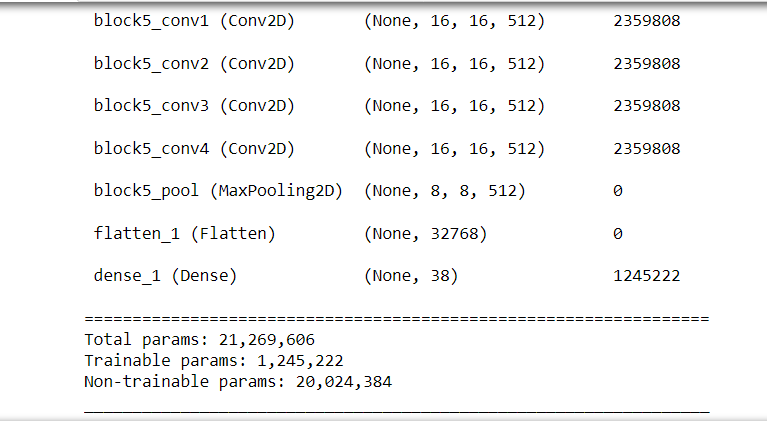
Fig. 1. Proposed methodology design

**4.RESULT**

**Model performance with other approaches**

|  |  |  |
| --- | --- | --- |
| **S.NO** | **Author** | **Accuracy** |
| **1** | Amanda Ramcharan, Peter McCloskey , Kelsee Baranowski, Neema Mbilinyi [11] | 79% |
| **2** | Navneet Kaur, V. Devendran band Sahil Verma [12] | 82.2% |
| **3** | P.Loganathan , Dr.R.Karthikeyan[9] | 81.08% |
| **4** | Hailay Beyene (PhD scholar), Dr. Narayan A. Joshi, Dr. Ketan Kotecha.[13] | 79.04% |
| **5** | Kshyanaprava Panda Panigrahi, Himansu Das, Abhaya Kumar Sahoo and Suresh Chandra Moharana[14] | 79.23% |
| **6** | **Proposed model in this work** | 82.50% |

**-Schema of Architecture**

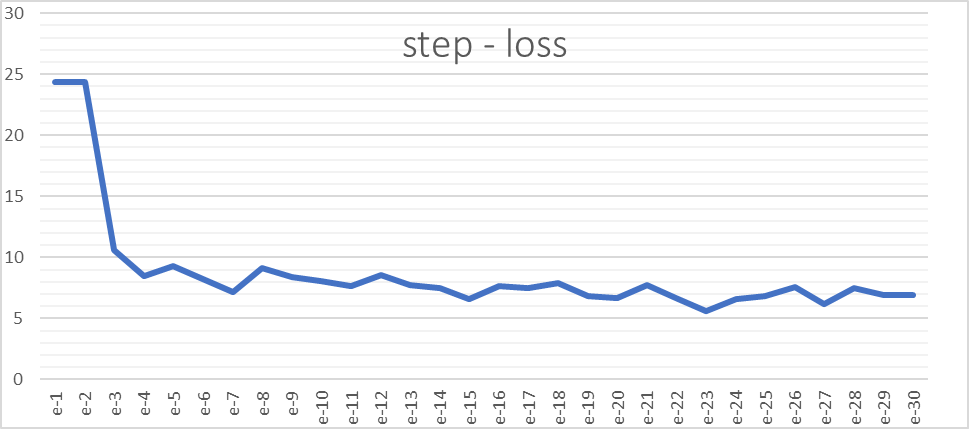
****

From the output shown in the above figure we can see that the total parameters are 20,024,384 trainable parameters are 20,024,384. This indicates that the model is trained over all the available parameters to reach certain efficiency. This process requires more time and memory which is one of the considerable factors while creating a model.

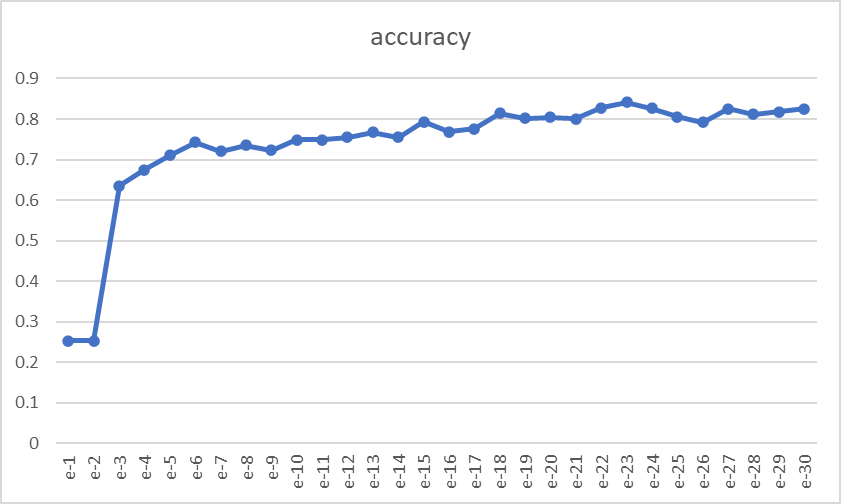
**Training Epoch**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Epoches** | **step – loss** | **accuracy** | **val\_loss** | **val\_accuracy** |
| e-1 | 24.3861 | 0.2537 | 12.0251 | 0.5013 |
| e-2 | 24.3861 | 0.2537 | 12.0251 | 0.5013 |
| e-3 | 10.6198 | 0.635 | 8.2369 | 0.6988 |
| e-4 | 8.4799 | 0.675 | 6.2271 | 0.7738 |
| e-5 | 9.2578 | 0.71 | 6.4477 | 0.7638 |
| e-6 | 8.2275 | 0.7425 | 6.889 | 0.7613 |
| e-7 | 7.1302 | 0.72 | 5.7158 | 0.77 |
| e-8 | 9.0917 | 0.735 | 6.4417 | 0.7987 |
| e-9 | 8.3501 | 0.7225 | 7.576 | 0.7675 |
| e-10 | 8.0419 | 0.7487 | 5.8454 | 0.7975 |
| e-11 | 7.6769 | 0.7487 | 5.2354 | 0.8313 |
| e-12 | 8.5634 | 0.755 | 5.71 | 0.825 |
| e-13 | 7.7512 | 0.7675 | 5.2813 | 0.8388 |
| e-14 | 7.4964 | 0.755 | 5.0708 | 0.8275 |
| e-15 | 6.5953 | 0.7937 | 4.4732 | 0.86 |
| e-16 | 7.6114 | 0.7688 | 4.5157 | 0.8537 |
| e-17 | 7.4557 | 0.7763 | 6.2981 | 0.8487 |
| e-18 | 7.9149 | 0.8138 | 5.3819 | 0.85 |
| e-19 | 6.7928 | 0.8012 | 6.7991 | 0.845 |
| e-20 | 6.7055 | 0.8037 | 5.2628 | 0.85 |
| e-21 | 7.7365 | 0.8 | 4.3952 | 0.8687 |
| e-22 | 6.6399 | 0.8275 | 4.712 | 0.88 |
| e-23 | 5.6056 | 0.8413 | 4.2648 | 0.8813 |
| e-24 | 6.5545 | 0.8263 | 6.8932 | 0.8213 |
| e-25 | 6.8368 | 0.805 | 4.7461 | 0.8763 |
| e-26 | 7.6007 | 0.7925 | 6.1902 | 0.8512 |
| e-27 | 6.1465 | 0.825 | 4.8073 | 0.88 |
| e-28 | 7.4893 | 0.8112 | 6.789 | 0.8413 |
| e-29 | 6.9495 | 0.8175 | 4.7485 | 0.885 |
| e-30 | 6.8821 | 0.825 | 6.8182 | 0.8537 |

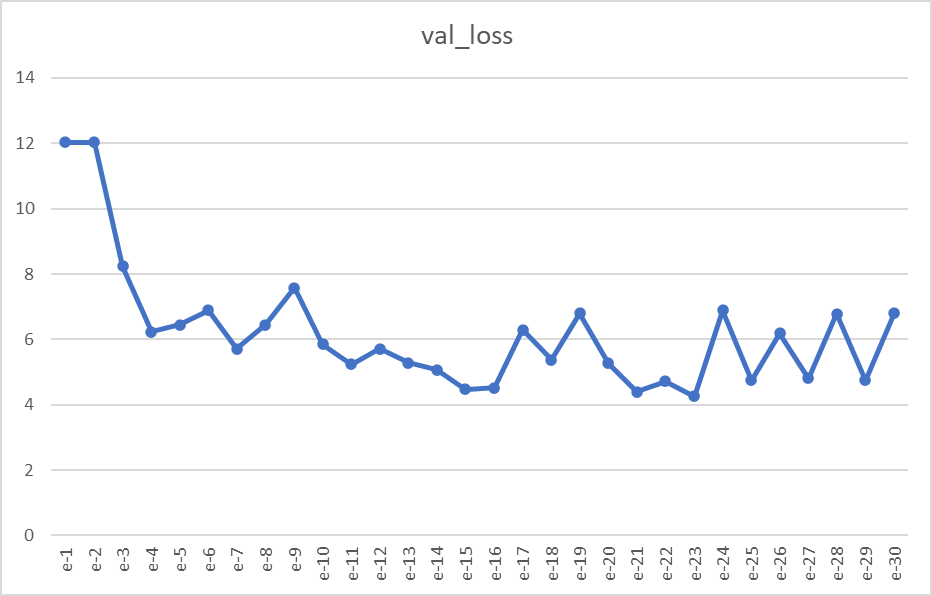
**-Training Data**

****

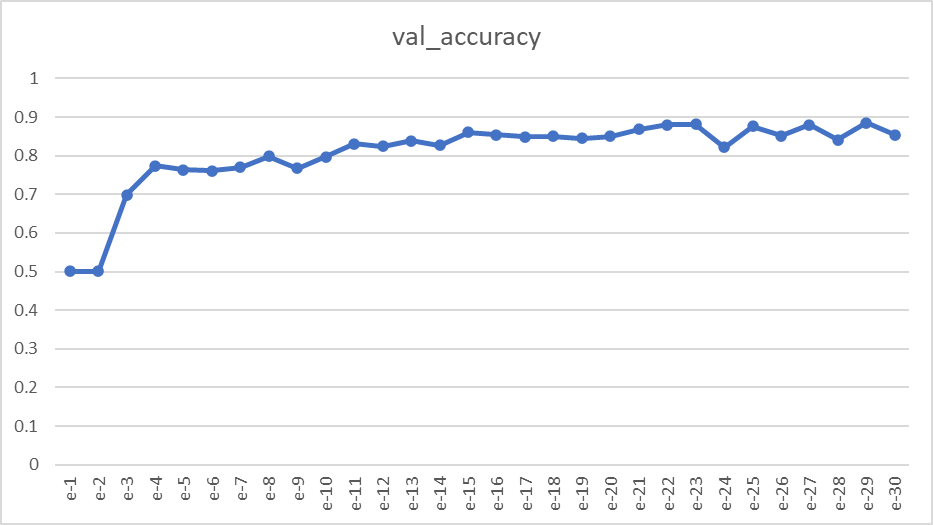
**Fig 2. Step loss**

****

**Fig 3. Accuracy**

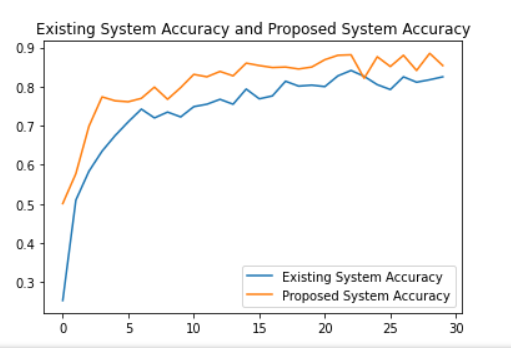
****

**Fig 4. Val Loss**

****

**Fig 5. Val\_Accuracy**

**ROC(receiver operating characteristic curve) Graph**

****

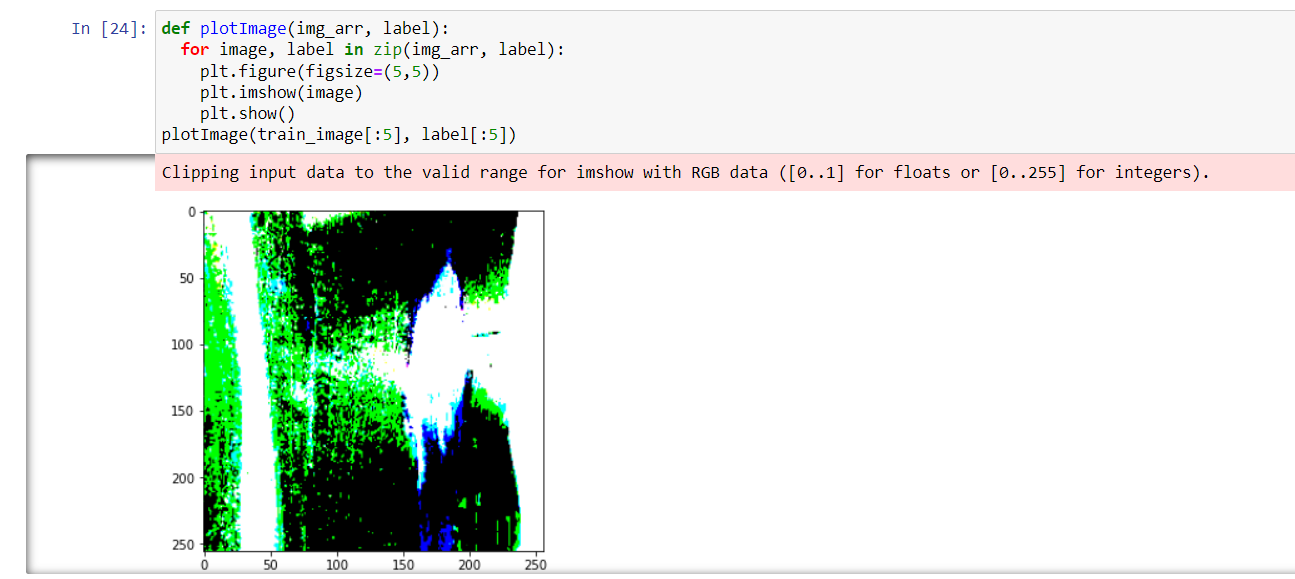
**Fig 6. ROC curve**

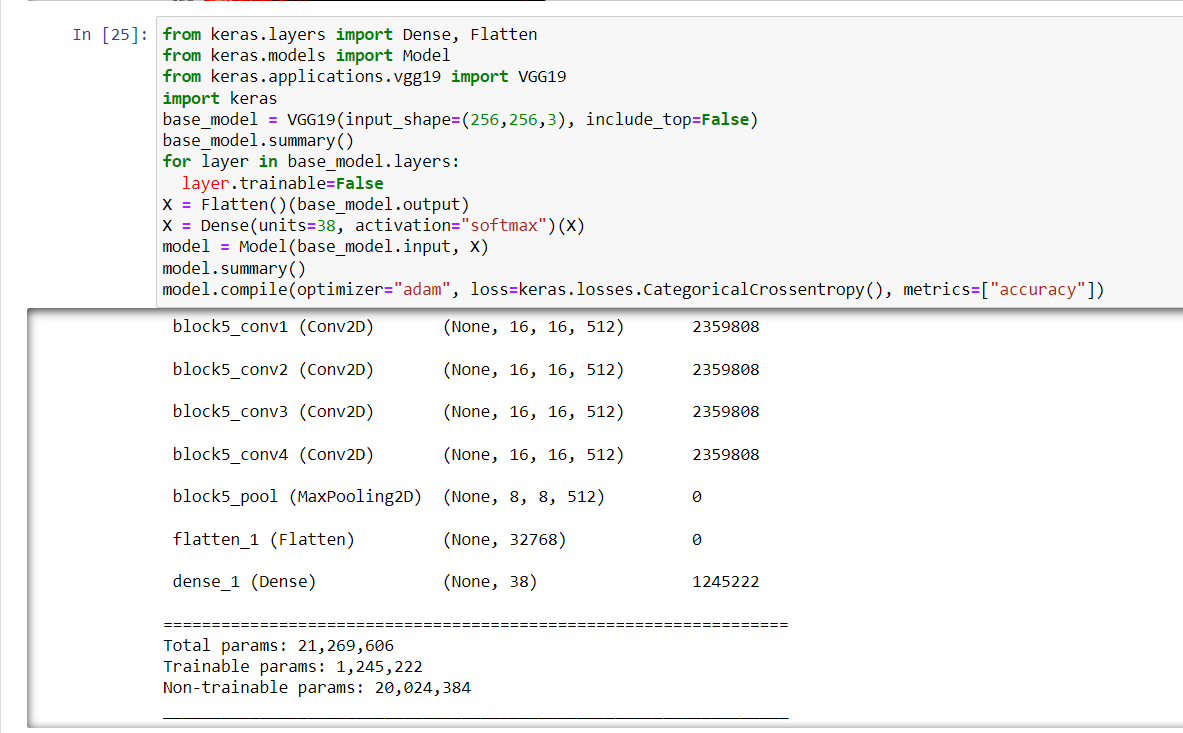
**Few Accuracy for Images**

|  |  |  |  |
| --- | --- | --- | --- |
| **S.No** | **Leaf Image** | **Disease Name** | **Accuracy** |
| 1 |  | Apple\_\_\_Apple\_scab | 82.5% |
| 2 |  | Blueberry\_healthy | 82.5% |
| 3 |  | Cherry\_(including\_sour)\_\_\_Powdery\_mildew | 82.5% |
| 4 |  | Corn\_(maize)\_\_\_Cercospora\_leaf\_spot Gray\_leaf\_spot | 82.5% |
| 5 |  | Grape\_\_\_Black\_rot | 82.5% |
| 6 |  | Orange\_\_\_Haunglongbing\_(Citrus\_greening) | 82.5% |
| 7 |  | Peach\_\_\_Bacterial\_spot | 82.5% |
| 8 |  | Pepper,\_bell\_\_\_Bacterial\_spot | 82.5% |
| 9 |  | Pepper,\_bell\_\_\_healthy | 82.5% |
| 10 |  | Potato\_\_\_Early\_blight | 82.5% |
| 11 |  | Potato\_\_\_Late\_blight | 82.5% |
| 12 |  | Squash\_\_\_Powdery\_mildew | 82.5% |
| 13 |  | Strawberry\_\_\_Leaf\_scorch | 82.5% |
| 14 |  | Tomato\_\_\_Bacterial\_spot | 82.5% |
| 15 |  | Tomato\_\_\_Spider\_mites Two-spotted\_spider\_mite | 82.5% |
|  |  |  |  |

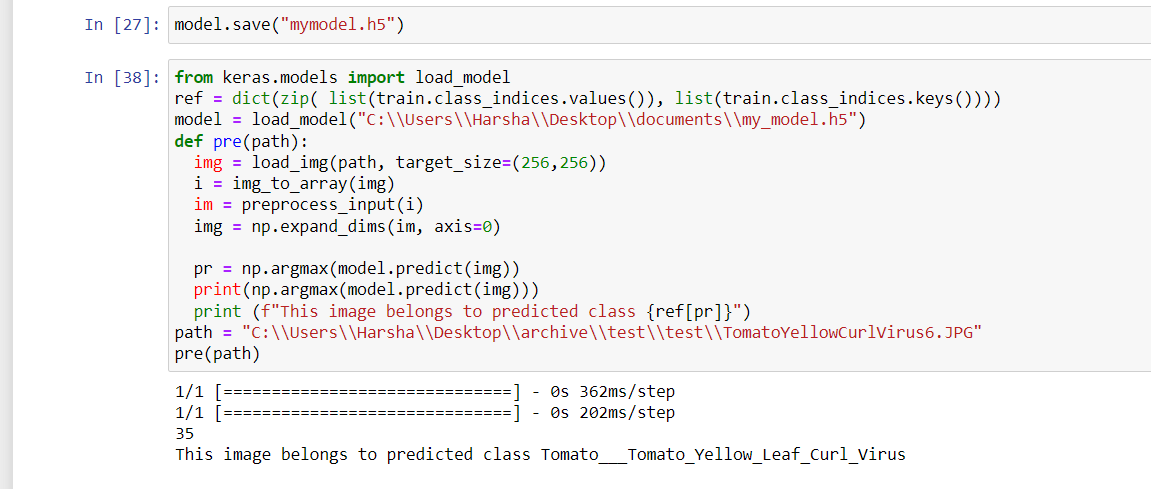
**5. Proposed Code-**





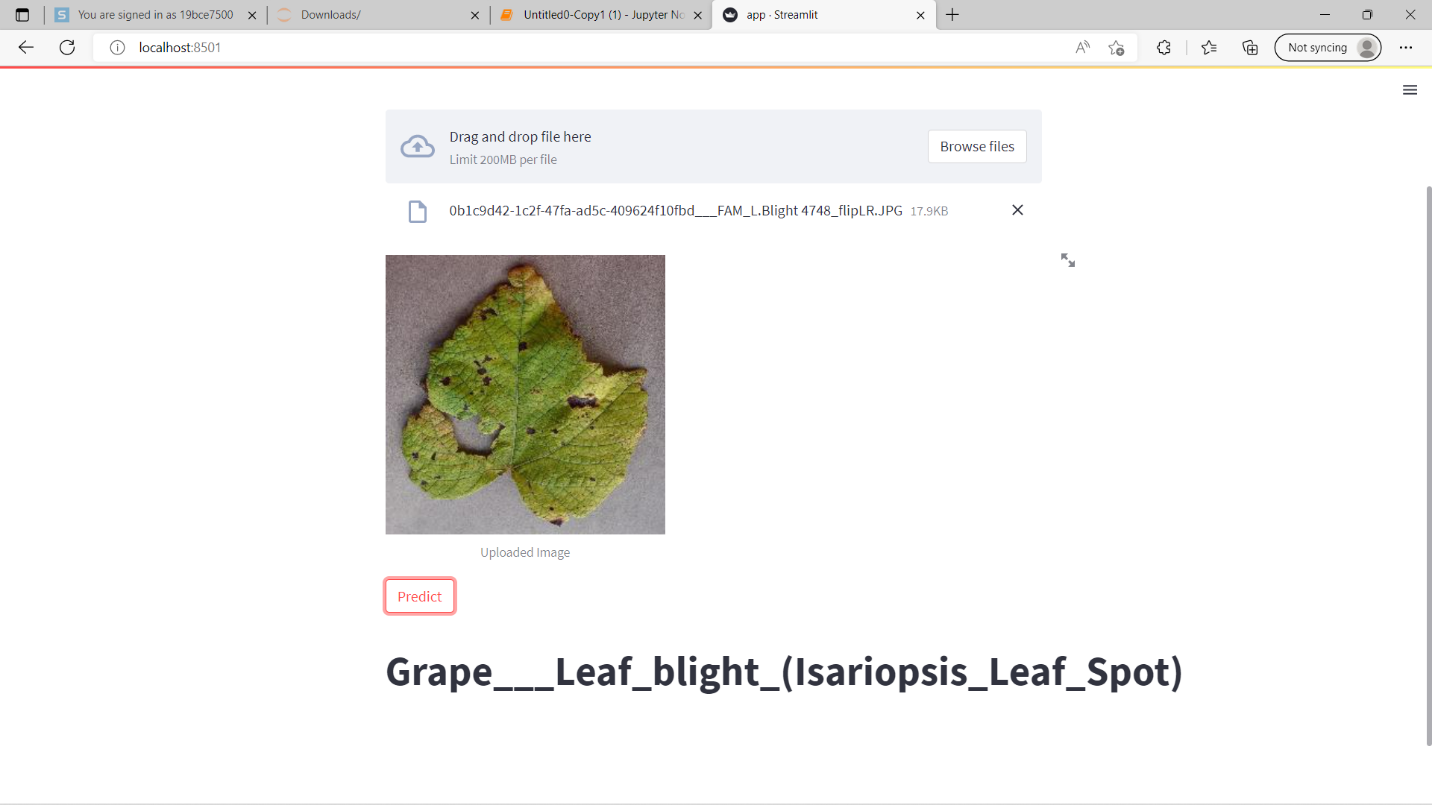








**Final output in Local Host Website**

****

**6. Conclusion and future Scope**

Agriculture is not an easy endeavor. However, that which matters most in our world. There are a number of factors that may be used to define the impacts on plants and diseases caused by environmental factors, including temperature, humidity, excess or deficiency in nutrition, light, and the most prevalent illnesses, which include bacterial, and fungus-related illnesses. Specialized deep learning models were created for our system based on the disease detection in plants using photographs of healthy or sick leaves. Our detector used pictures different cameras on-site, as well as data gathered from a variety of sources. Results of our experiment. Further comparisons between other deep-architectures with feature extractors showed how our deep-learning-based detector is able to accurately identify numerous plant illnesses in different species also provide a cure for the aforementioned disorders. Healthy plants are surviving on the nutrient-contaminated soil effectively ward against insect attacks. We anticipate that the suggested system will be a useful addition to the agricultural study. In future, the progress of identifying the plant diseases has been further modified to analyze various parts of plants such as stem, flower, leaf etc., and improve the performance speed.

**References**

[1] S. R. Maniyath *et al.*, “Plant disease detection using machine learning,” in *Proceedings - 2018 International Conference on Design Innovations for 3Cs Compute Communicate Control, ICDI3C 2018*, Aug. 2018, pp. 41–45. doi: 10.1109/ICDI3C.2018.00017.

[2] A. Akhtar, A. Khanum, S. A. Khan, and A. Shaukat, “Automated plant disease analysis (APDA): Performance comparison of machine learning techniques,” in *Proceedings - 11th International Conference on Frontiers of Information Technology, FIT 2013*, 2013, pp. 60–65. doi: 10.1109/FIT.2013.19.

[3] R. U. Khan, K. Khan, W. Albattah, and A. M. Qamar, “Image-Based Detection of Plant Diseases: From Classical Machine Learning to Deep Learning Journey,” *Wireless Communications and Mobile Computing*, vol. 2021. Hindawi Limited, 2021. doi: 10.1155/2021/5541859.

[4] S. Roy, R. Ray, S. R. Dash, and M. K. Giri, “Plant Disease Detection Using Machine Learning Tools With an Overview on Dimensionality Reduction,” in *Data Analytics in Bioinformatics: A Machine Learning Perspective*, wiley, 2021, pp. 109–144. doi: 10.1002/9781119785620.ch5.

[5] B. S. Kusumo, A. Heryana, O. Mahendra, and H. F. Pardede, “Machine Learning-based for Automatic Detection of Corn-Plant Diseases Using Image Processing,” in *2018 International Conference on Computer, Control, Informatics and its Applications: Recent Challenges in Machine Learning for Computing Applications, IC3INA 2018 - Proceeding*, Jan. 2019, pp. 93–97. doi: 10.1109/IC3INA.2018.8629507.

[6] M. Kumar, A. Kumar, and V. S. Palaparthy, “Soil Sensors-Based Prediction System for Plant Diseases Using Exploratory Data Analysis and Machine Learning,” *IEEE Sens. J.*, vol. 21, no. 16, pp. 17455–17468, Aug. 2021, doi: 10.1109/JSEN.2020.3046295.

[7] P. Kulkarni, A. Karwande, T. Kolhe, S. Kamble, A. Joshi, and M. Wyawahare, “Plant Disease Detection Using Image Processing and Machine Learning.”

[8] R. Sujatha, J. M. Chatterjee, N. Z. Jhanjhi, and S. N. Brohi, “Performance of deep learning vs machine learning in plant leaf disease detection,” *Microprocess. Microsyst.*, vol. 80, Feb. 2021, doi: 10.1016/j.micpro.2020.103615.

[9] P. Loganathan, R. Karthikeyan, and R. Scholar, “Residual Neural Network (ResNet) Based Plant Leaf Disease Detection and Classification,” 2021.

[10] A. R and S. S, “Plant leaf Disease Identification using Deep learning techniques,” *Int. J. Eng. Adv. Technol.*, vol. 9, no. 5, pp. 462–464, Jun. 2020, doi: 10.35940/ijeat.E9683.069520.

[11] A. Ramcharan *et al.*, “A mobile-based deep learning model for cassava disease diagnosis,” *Front. Plant Sci.*, vol. 10, no. March, pp. 1–8, 2019, doi: 10.3389/fpls.2019.00272.

[12] N. Kaur, V. Devendran, and S. Verma, “Crop leaf disease identification based on ensemble classification,” 2021.

[13] H. BEYENE, D. N. A. JOSHI, and D. K. KOTECHA, “PLANT DISEASES PREDICTION USING IMAGE PROCESSING AND MACHINE LEARNING TECHNIQUES: SURVEY,” *Int. J. Comput. Appl.*, vol. 1, no. 8, 2018, doi: 10.26808/rs.ca.i8v1.23.

[14] K. P. Panigrahi, H. Das, A. K. Sahoo, and S. C. Moharana, “Maize Leaf Disease Detection and Classification Using Machine Learning Algorithms,” in *Advances in Intelligent Systems and Computing*, 2020, vol. 1119, pp. 659–669. doi: 10.1007/978-981-15-2414-1\_66.