**Research paper**

# Crop Disease Detection using Convolutional Neural Network Model and Image Processing

# Komal Gupta 1, Hina Kanaujiya 2, Divika Agrawal 3, Deeksha Priya 4

# 1,2,3,4 Dept. of Computer Science and Engineering, Ajay Kumar Garg Engineering College, Ghaziabad, Uttar Pradesh, India

**ABSTRACT**

Plant and crop cultivation rates are rapidly increasing as human and animal demands increase all over the world. But, plant disease is a persistent problem for smallholder farmers, threatening their income and food security. We can use technologies such as image processing and deep learning to successfully detect plant diseases in their early stages. All necessary steps for implementing this disease recognition model are fully described throughout the paper, beginning with image collection to create a database. Deep learning frameworks, specifically Convolutional Neural Networks (CNNs), have led to significant advances in image processing and CNN training and fine-tuned to precisely match the database of a plant's leaves gathered independently for various plant diseases The developed model, which can recognise plant diseases, is deployed as a web application. For training and validating the model, a dataset of leaf images captured in a controlled environment is created. Validation results show that the proposed method has a 97.3 percent accuracy.

Key Words: Plant Disease Detection, Deep Learning, Convolutional Neural Networks Image Processing

**INTRODUCTION**

# Plant disease is a disturbance in a plant's normal state that disrupts or modifies its vital functions. Plant diseases are classified broadly based on the nature of their primary causal agent, infectious or non-infectious. A pathogenic organism, such as a fungus, bacterium, mycoplasma, virus, viroid, nematode, or parasitic flowering plant, causes infectious plant diseases. An infectious agent has the ability to reproduce within or on its host and spread from one susceptible host to another. Non-infectious plant diseases are caused by unfavourable growing conditions, such as temperature extremes, unfavourable moisture-oxygen relationships, toxic substances in the soil or atmosphere, and an excess or deficiency of an essential mineral. Non-infectious causal agents are not transmissible because they are not organisms capable of reproducing within a host.[1]

Pesticides are used to protect crops from these infestations and thus maintain yields. But use of these chemicals has a negative impact on biodiversity, including insect, bird, and fish populations, as well as soil, air, and water quality. On large farms, checking the condition of each plant several times during the season is impractical. Prospection can also be complicated by the difficulty of accessing certain crops. **To address these shortcomings, we created a plant detection system based on Convolutional Neural Networks that predicts plant disease based on the image uploaded by the farmer. This proposed work is less time-consuming, free of charge, and accurate, making it suitable for use in real-time. In recent years, the use of CNNs in plant disease classification has yielded excellent results. The multi-layered supervised network has gained favour among researchers due to the continuous emergence of superior results. Such advancements have aided in reducing training time and error rate. Above all, the evolution of architecture has been a necessary demand of large and complex datasets in the twenty-first century. Essentially, we will first train our CNN models using a large number of images of potato, pepper, and tomato. CNN reads a very large image in a straightforward manner. CNN is most commonly used to analyse visual imagery and is frequently used in image classification behind the scenes.[2]**

**The remaining paper is organised into four sections. Section 2 discusses several state-of-the-art systems for automatic plant disease detection. Section 3 describes the materials and methods used to design the proposed model. Section 4 of the paper presents the model's results for detecting diseases in tomato, potato, and pepper plants. Section 5 brings the paper to a close.**

**LITERATURE REVIEW**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **S.no.** | **Year** | **Author** | **Title(s)** | **Source(s)** | **Finding(s)** |
| 1. | 2021 | Kelman, Arthur, Rita M. , Shurtleff, Malcolm C. and Pelczar, Michael J.[1] | Plant disease plant pathology | Massachusetts Institute of Technology (A Britannica Publishing Partner) | Plant disease occurrence and prevalence vary seasonally, depending on the presence of the pathogen, environmental conditions, and the crops and varieties grown. |
| 2. | 2021 | Sk Mahmudul Hassan, Arnab Kumar Maji, Michal Jasi ´nski ,Zbigniew Leonowicz and Elzbieta Jasi nska | Identification of Plant-Leaf Diseases Using CNN and Transfer-Learning Approach | Department of Information Technology, North Eastern Hill University, Shillong, Meghalaya, Department of Electrical Engineering Fundamentals, Wroclaw University of Science and Technology | Plant disease detection using DL models (InceptionV3, InceptionResnetV2, MobileNetV2, EfficientNetB0) using healthy- and diseased-leaf images of plants. The time required to train the model was significantly less than that required by other machine-learning approaches. |
| 3. | 2020 | Mohit Agarwal, Abhishek Singh, Siddhartha Arjaria, Amit Sinha, Suneet Gupta | ToLeD: Tomato Leaf Disease Detection using Convolution Neural Network | Procedia Computer Science | As the images inside the class are not balanced, data augmentation techniques have been applied to balance the images inside the class, thus showing the benefit of the proposed model over pre-trained models. The model's testing accuracy ranges from 76% to 100% for the classes, with an average testing accuracy of 91.2%. The proposed model requires approximately 1.5 MB of storage space, whereas pretrained models require approximately 100 MB of storage space. |
| 4. | 2019 | Mercelin Francis | Disease Detection and Classification in Agricultural Plants Using Convolutional Neural Networks: A Visual Understanding | 6th international Conference on Signal Processing and Integrated Networks(SPIN) | When compared to using existing deep learning models for various applications to achieve maximum accuracy, creating and training a CNN model from scratch is a time-consuming process. When compared to other existing models, the accuracy is 88.7 with the fewest parameters, i.e., 45K. |
| 5. | 2019 | Sammy V. Militante | Plant Leaf Detection and Disease Recognition using Deep Learning | IEEE Eurasia Conference on IOT, Communication and Engineering. | CNN architecture, During the model's training, 75 epochs were used to achieve a 96.5 percent accuracy rate. When tested on random images of plant varieties and diseases, the model achieved a maximum accuracy rate of 100 percent. |
| 6. | 2016 | Sharada P. Mohanty David P. Hughes and Marcel Salathé | Using deep learning for image-based plant disease detection. Frontiers | Frontiers in Plant Science | On test set data, the model achieved an accuracy of 99.35 percent. When used on images obtained from trusted online sources, the model achieves an accuracy of 31.4 percent. While this is better than a simple random selection model, a more diverse set of training data can help to increase the accuracy. |
| 7. | 2013 | S. Arivazhagan, R. Newlin Shebiah, S. Ananthi, S. Vishnu Varthini. | Detection of unhealthy region of plant leaves and classification of plant leaf diseases using texture features | Agric Eng Int: CIGR Journal | First, a colour transformation structure is used for the input RGB image, and then, using a specific threshold value, the green pixels are detected and uninvolved, which is followed by a segmentation process, and texture statistics are computed to obtain beneficial segments. Finally, classifier is used for the extracted features that are used to classify the disease. |

Their use has been one of the driving forces behind the rise in food production since the 1950s, allowing it to meet the demands of a growing population (Cooper and Dobson, 2007). The use of these chemicals has a negative impact on biodiversity, including insect, bird, and fish populations, as well as soil, air, and water quality (Risebrough, 1986; Gill and Garg, 2014; Goulson, 2014; Sanchez-Bayo and Goka, 2014; Knillmann and Liess, 2019). Their use also poses a threat to human health, with both acute and chronic consequences (Weisenburger, 1993; Bassil et al., 2007; Kim et al., 2016). However, the amount of pesticides used worldwide is increasing, with +78 percent of tonnes of active ingredients used between 1990 and 2016. (Food and Agriculture Organization of the United Nation, 2018). Assessing the health of fields is a difficult task that necessitates a high level of expertise. A disease can manifest itself differently from one plant species to the next, or even from one variety to the next. A symptom can be caused by a variety of issues, and these issues can coexist on the same plant. Even nutritional deficiencies and pests can cause symptoms that are similar to those of certain diseases (Barbedo, 2016). Assessing the health of plots takes time as well. On large farms, checking the condition of each plant several times during the season is impractical. Prospection can also be complicated by the difficulty of accessing certain crops. By utilising automatic prospection or expert assistance tools, the automatic identification of diseases by imagery has the potential to solve all of these issues. [3]

**METHODOLOGY**

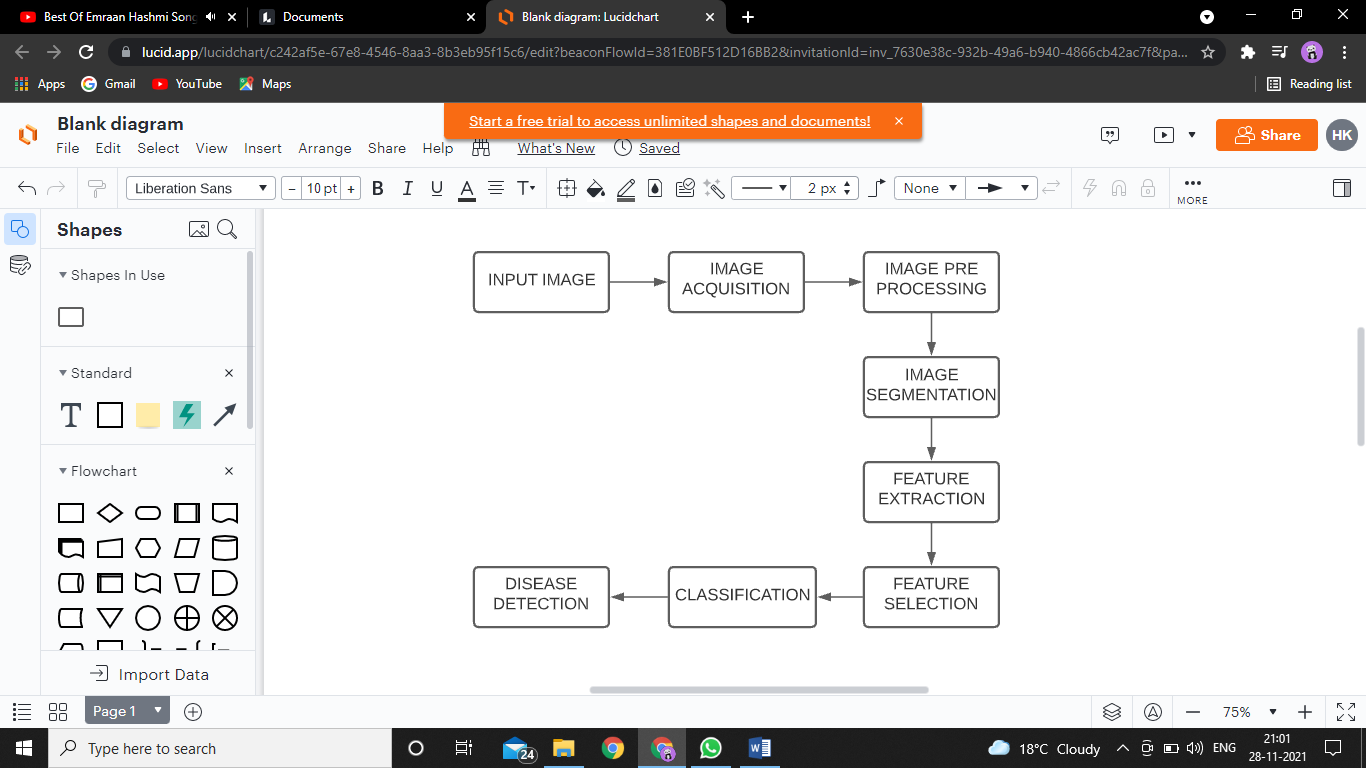
****

Fig. 1 Methodology

This section of the report illustrates the method used to categorise the leaves as diseased or healthy, and if the leaf is diseased, the name of the disease is mentioned, as well as the remedies. Our methodology is primarily based on the steps listed below.

**Dataset (Input Image) and Image Acquisition**

For this project, we will use a public dataset to detect plant leaf disease. The dataset will include images of healthy and unhealthy plant leaves, as well as some classes for experimenting with our algorithm. It accepts the image as input and processes it further. We chose the most popular image domains so that we can accept any format as input to our process, such as.bmp,.jpg,.gif. We used the Plant Village dataset to train our CNN model. This dataset contains the following plant leaf directories:

• Apple\_\_\_Apple\_scrab

• Apple\_\_healthy

• Blueberry\_\_healthy

• Cherry\_(including\_sour)\_\_\_Powdery\_mildew

• Cherry\_(including\_sour)\_\_\_healthy

• Corn\_(maize)\_\_Common\_rust

• Corn\_(maize)\_\_\_Northern\_Leaf\_Blight

• Corn\_(maize)\_\_\_healthy

• Grape\_\_\_Black\_rot

• Grape\_\_Leaf\_blight(Isariopsis\_Leaf\_Spot)

• Grape\_\_\_healthy

• Orange\_\_Haunglongbing(Citrus\_greening)

• Peach\_\_\_Bacterial\_spot

• Peach\_\_\_healthy

• Pepper\_bell\_\_\_Bacterial\_spot

• Pepper\_bell\_\_\_healthy

• Potato\_\_\_Early\_blight

• Potato\_\_\_healthy

• Raspberry\_\_\_healthy

• Soybean\_\_\_healthy

• Squash\_\_\_Powdery\_mildew

• Strawberry\_\_\_Leaf\_scorch

• Strawberry\_\_\_healthy

• Tomato\_\_\_Bacterial\_spot

• Tomato\_\_\_healthy



Fig. 2 Sample image of the dataset

**Image Pre-processing**

The acquired images cannot be directly fed into the Convolutional Neural Network. The images are transformed into a machine-readable format, resized to 256\*256 pixels, then subjected to median filtering to remove noise. The digital version of the rotten leaf sample contains approximately 30% of the leaf area and the remaining 70% is the background. To achieve efficient disc storage and fast processing speed, the digital image of the leaf sample is cropped into a smaller dimension of size 16x20sq.m, saving approximately 30% of disc storage space and increasing CPU processing by 1.4 times. After pre-processing stage, the sample leaf image contains approximately 70% of the leaf area and the remaining 30% as the background. For low contrast, contrast enhancement algorithms can be used. Background removal techniques may be required in some cases if a region of interest needs to be extracted. A median filter can be used to remove noise such as salt and pepper and Wiener filter can be used to eliminate blurring.

**Image Segmentation**

The third step in our proposed method is image segmentation. Using a classifier and the k-mean clustering algorithm, the segmented images are clustered into different sectors. The RGB colour model is converted into a Lab colour model before the images are clustered. The introduction of the Lab colour model allows for the easy clustering of segmented images.

**Data pre-processing and feature extraction**

To obtain precise results, some background noise should be removed prior to feature extraction. The data was transformed into the proper format for the implementation of various classifiers. To prepare the dataset in a standard format, it is also necessary to remove any unnecessary fields, missing records, and duplicate records. After the images have been properly formatted and segmented, the next critical step is feature extraction. This step in a Convolutional Neural Network is accomplished by employing an activation function followed by max-pooling layers.

**Activation Function**

The main purpose of using an activation function in a CNN is to determine which weights and input combinations will fire the next neuron. The ReLU (Rectified Linear Unit) activation function, is used in our CNN model. This is a very simple mathematical function that returns the input if it is greater than zero and returns zero otherwise. It's written as y = max (0, x), where x is the input and y is the output. Because it activates all neurons at once, using this activation function allows the model to learn faster and perform better.

**Convolutional Neural Network (CNN) Algorithm**

1) A CNN is a type of deep neural network that is commonly used to analyse visual imagery.

2) A CNN is a Machine Learning algorithm that can take in an input image, assign importance to various aspects/objects in the image, and distinguish one from the other.

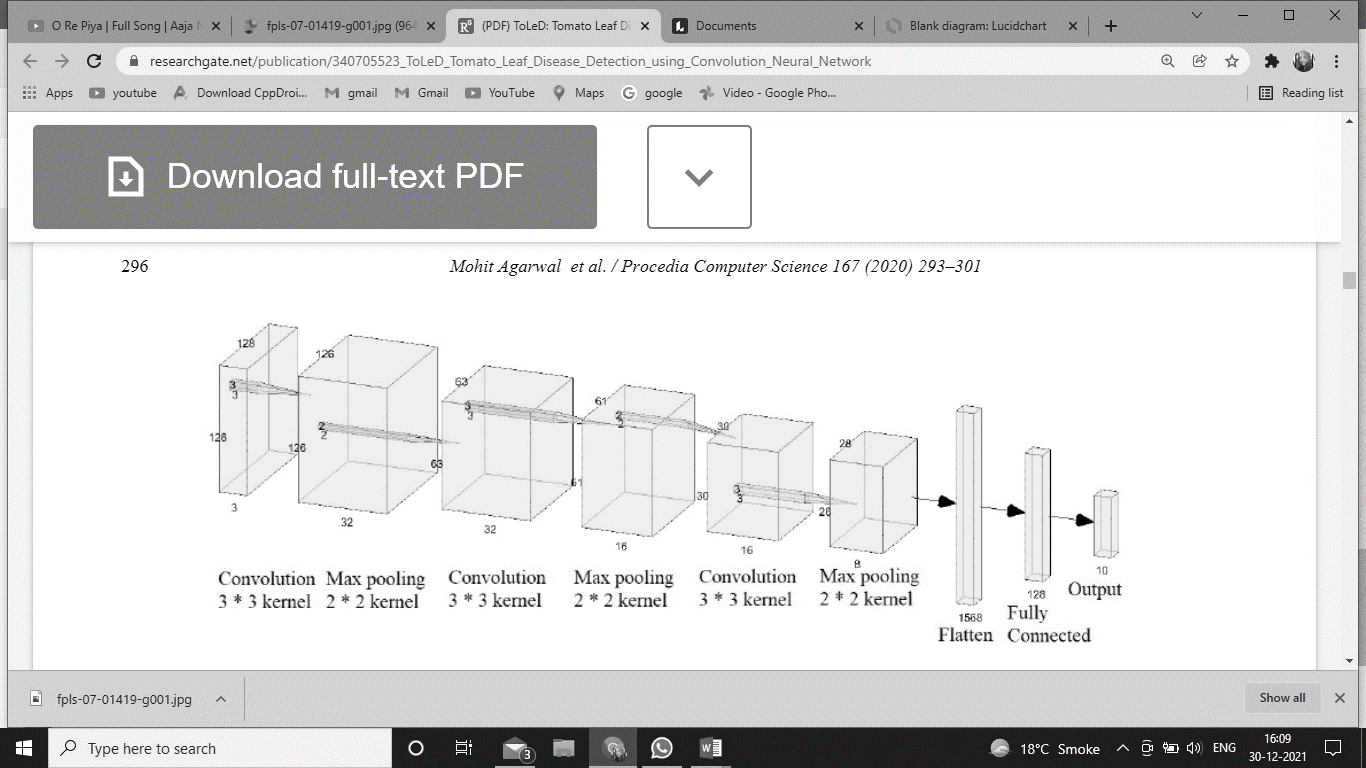


Fig.3 Pictorial representation of proposed convolutional network

**INPUT LAYER**

The image below is an RGB image that has been divided into three colour planes: red, green, and blue. Images can exist in a variety of colour spaces, including grayscale, RGB, HSV, CMYK, and others.

• The ConvEnt's role is to reduce the images into a format that is easier to process while retaining features that are critical for obtaining a good prediction.

• This is critical when designing an architecture that is not only good at learning features but also scalable to massive datasets.

**CONVOLUTION LAYER-The Kernel**

The green section in the following demonstration resembles our 5x5x1 input image. The element involved in performing the convolution operation in the first part of a Convolutional Layer is known as the Kernel/Filter, K, and is represented in yellow. K was chosen as a 3x3x1 matrix ((1,0,1), (0,1,0), (1,0,1)). When performing a matrix multiplication operation between K and the portion P of the image, the Kernel shifts 9 times due to Stride Length - 1.

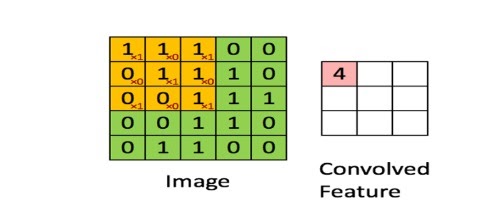


Fig. 4 Convolution Layer

With a given Stride Value, the filter moves to the right until it has parsed the entire width. Moving on, it hops down to the image's beginning (left) with the same Stride Value and repeats the process until the entire image is traversed.

**POOLING LAYER**

The Pooling layer is in charge of shrinking the spatial size of the Convolved Feature. It is also useful for extracting dominant features that are rotational and positional invariant, allowing the model to be effectively trained.

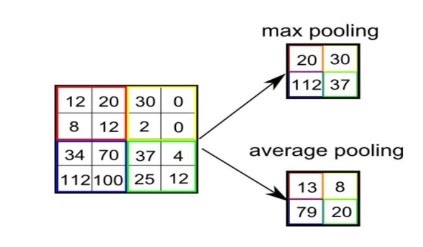


Fig. 5 Pooling

Pooling is classified into two types:

1. Max Pooling - It returns the maximum value from the Kernel-covered portion of the image.

2. Average Pooling- It returns the average of all the values from the Kernel-covered portion of the image.

This step prevents over-fitting and reduces noise distortion and reduce model over-fitting by randomly dropping neurons. The flatten layer is used to convert the output into a single long feature vector that is then fed into the next layer.

**Feature selection**

In this project, we are selecting features based on the correlation of variables with the target variable. The correlation between feature green part of leaf and green part of leaf is very high, indicating that both variables are interdependent. As a result, we've dropped one of them. Less correlated features such as green channel mean, red channel standard deviation, blue channel standard deviation, dissimilarity, and correlation will now play a minor role in model development for apple disease prediction. As a result, we removed these variables as well. Following feature selection, the data is parsed and fed into machine learning classifiers to find patterns in the data.

**Classification Algorithm**

The Classification algorithm is a Learning technique that uses training data to identify the category of new observations. We compile the model after deciding on the layers of the Convolutional Neural Network. The dataset is divided into two parts: training and testing. We have a value of 20 for the epoch, which is the number of passes the model makes over the dataset. To capture the various image features, real-time image processing functions are used. After training, the model is tested on the validation dataset.

**PROPOSED SYSTEM**

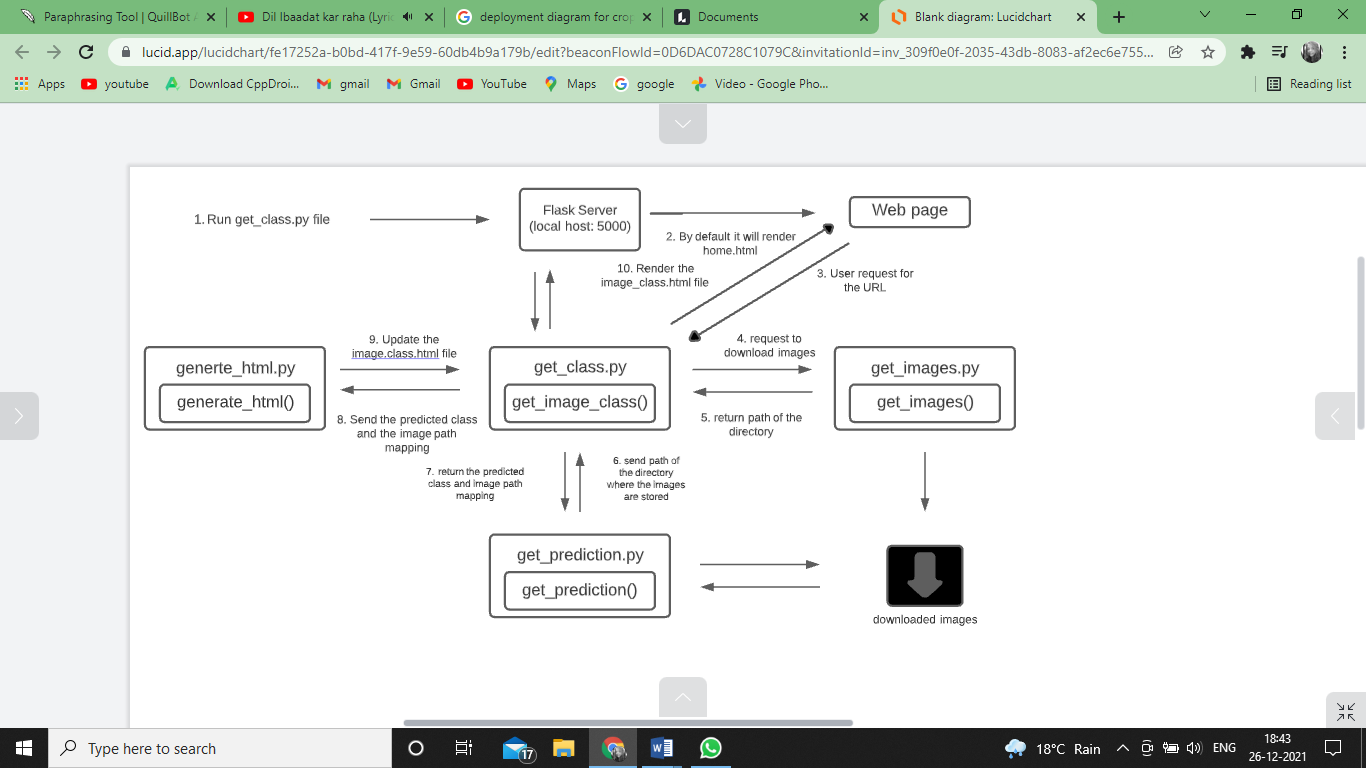


Fig.6 Proposed System

The Flask application will first render the home.html file, and whenever someone sends an image classification request, Flask will detect a post method and call the get image class function.

This function will function as follows:

1. First, it will send a request to download and save the images.

2. It will then send the directory path to the get prediction.py file, which will calculate and return the results as a dictionary.

3. Finally, it will send this dictionary to the generate html.py file, which will generate the output file and send it back to the user.

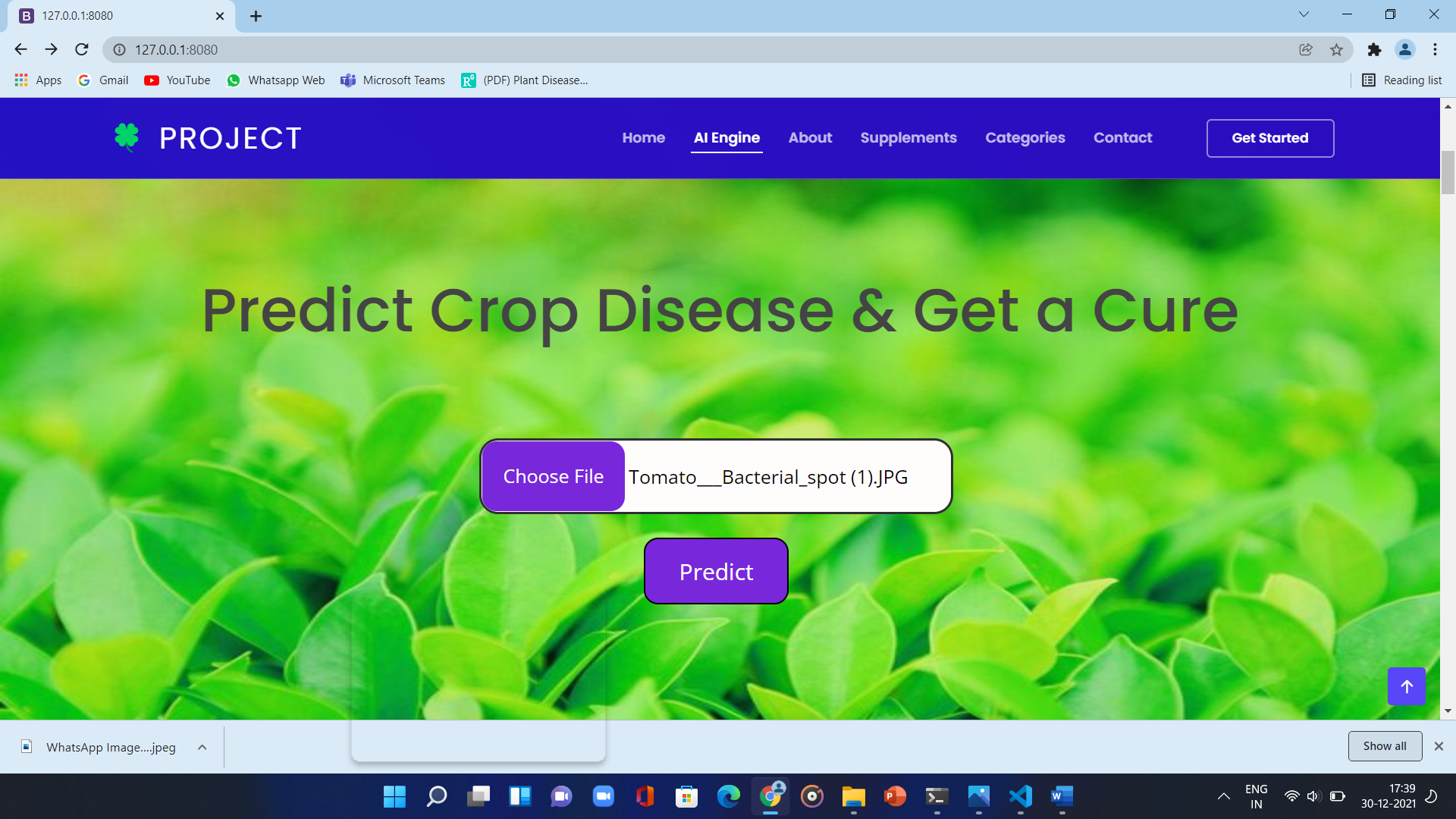


Fig. 7 AI Engine Input Page

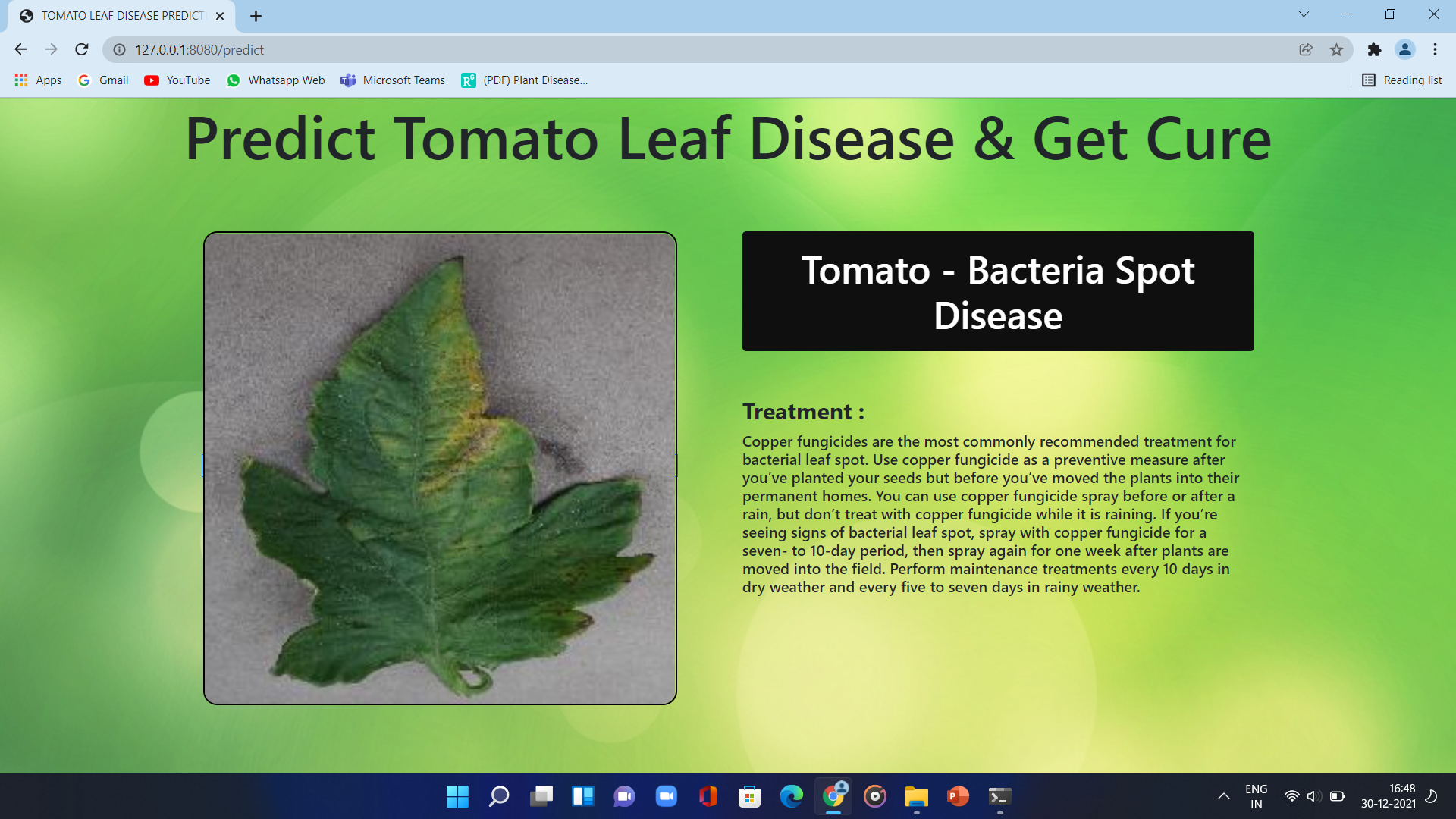


Fig. 8 AI Engine Output Page

First the images of various leaves are acquired using a high resolution camera so as to get better results & efficiency. Then image processing techniques are applied to these images to extract useful features which will be required for further analysis.

The main purpose of the proposed system is to detect the diseases of plant leaves by using feature extraction methods where features such as shape, color, and texture are taken into consideration. Convolutional neural network (CNN), a machine learning technique is used in classifying the plant leaves into healthy or diseased and if it is a diseased plant leaf, CNN will give the name of that particular disease. Suggesting remedies for particular diseases is made which will help in growing healthy plants and improve productivity.

The following are the benefits of the proposed algorithm:

• It also offers eco friendly recovery measures for the identified species.

• Using estimators for automatic cluster centre initialization eliminates the need for user input during segmentation.

• The proposed algorithm improves detection accuracy.

• The proposed method is completely automatic, whereas existing methods require the user to choose the best segmentation of the input image disease.

**OBJECTIVES OF THE PROPOSED SYSTEM**

• Image acquisition and image pre-processing are used to improve the given input image.

• Using texture analysis and segmentation, identify the affected part.

• Using feature extraction and classification, classify the healthy and affected leaf parts.

• For accurate results, train the model with testing data.

Describing dynamic aspects of the system in following diagram. This diagram is essentially a flowchart that depicts the flow from one activity to another. This can be described as system operation. The control flow is directed from one operation to the next.

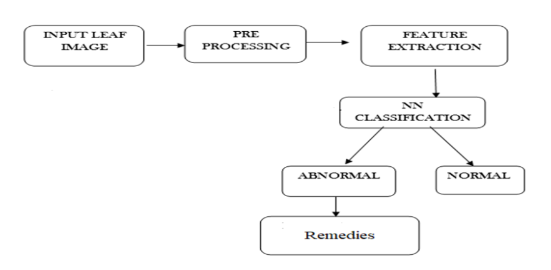


Figure. 9 Activity Diagram

The above flow represents the flow from one activity to another; the activity begins with an input leaf image captured with a digital camera, and then the input leaf is pre-processed to extract features such as colour, shape, texture, and so on. The processed image is now classified as Normal or Abnormal; if leaf is Abnormal, remedies are recommended.

**RESULTS AND DISCUSSION**

In our research, the proposed model is given expected output after training, testing, and validation using various datasets; validation results show that the proposed method can achieve an accuracy of 97.3 %.

CNNs surpass other technologies in tasks involving crop disease classification and detection. They are capable of dealing with complex issues under difficult imaging conditions. Because of their robustness, they may now be able to leave the research environment and become part of operational tools. However, before tools for expert assistance and automatic screening can be implemented, a few steps must be tested and integrated. In this section, we will first discuss the best practises to implement throughout the development chain so that trained models can handle the real-world complexities of agricultural and phytosanitary problems. We then identify the elements that need to be addressed further in order for such tools to be fully operational, including potential research directions.

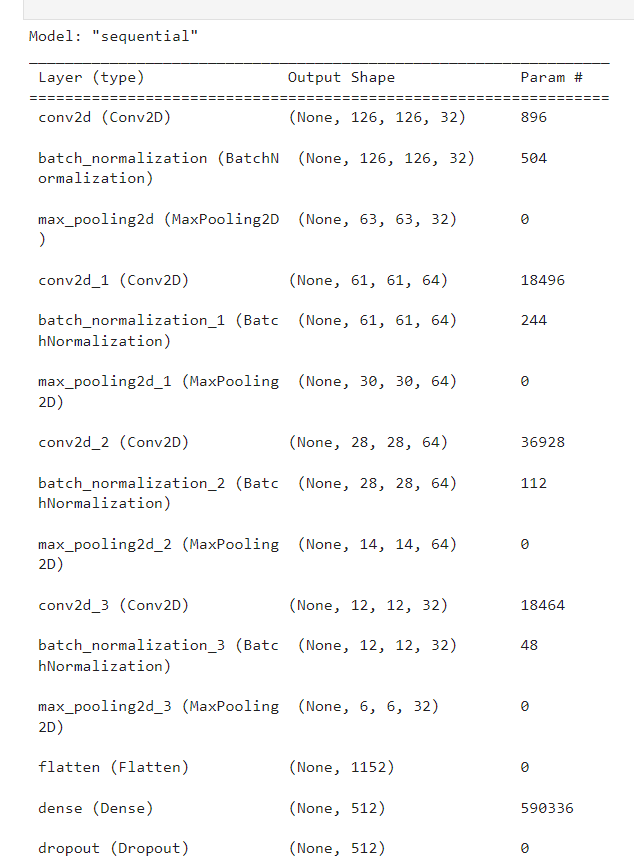


Fig. 10

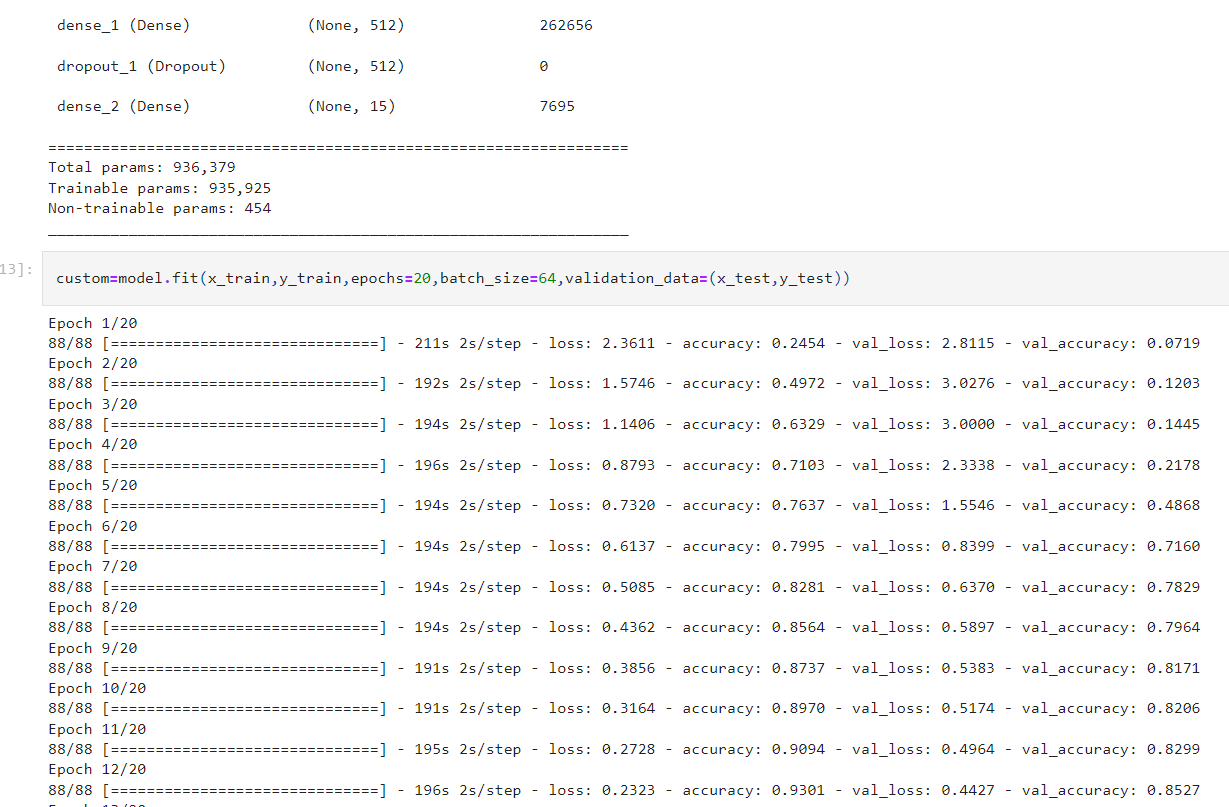


Fig. 11

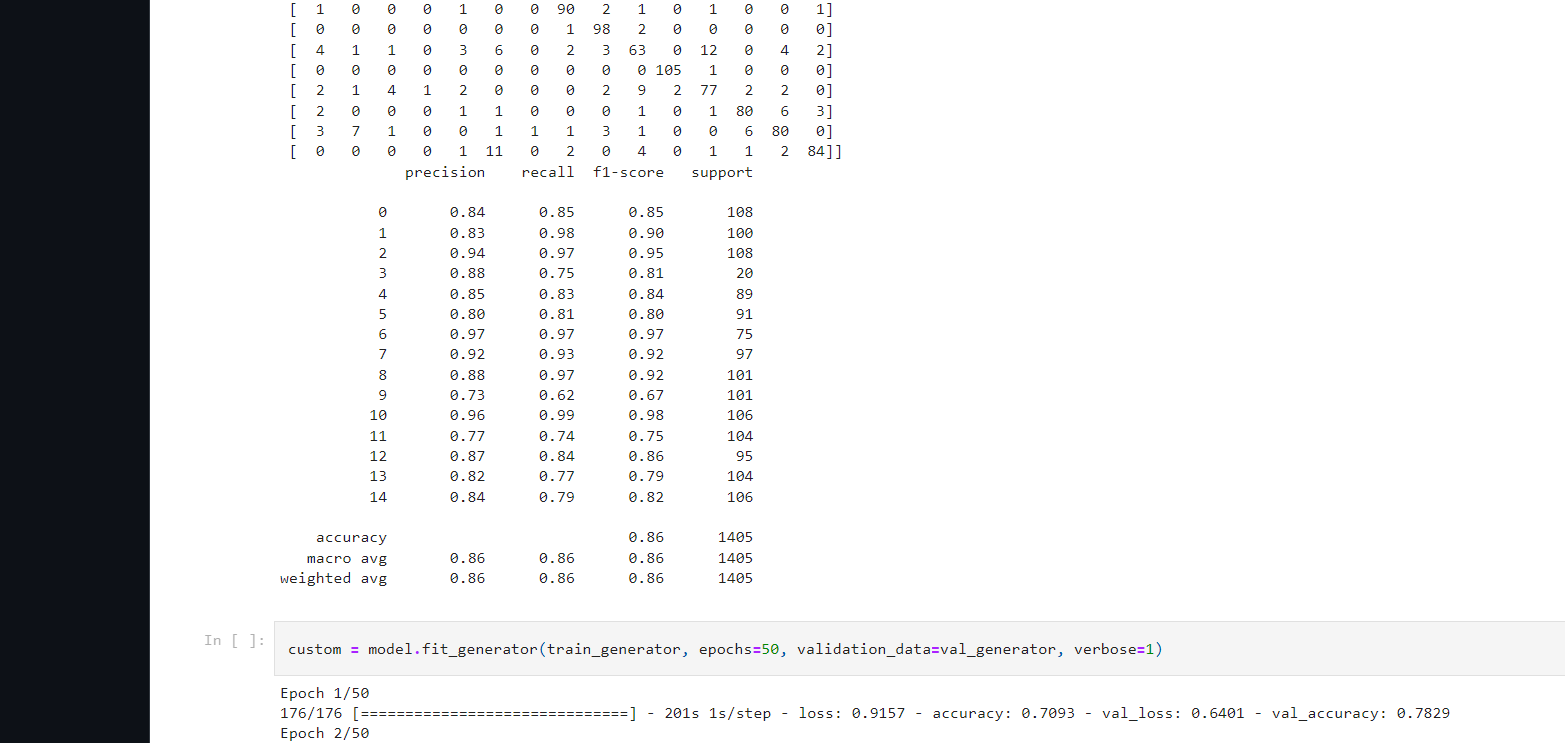


Fig. 12

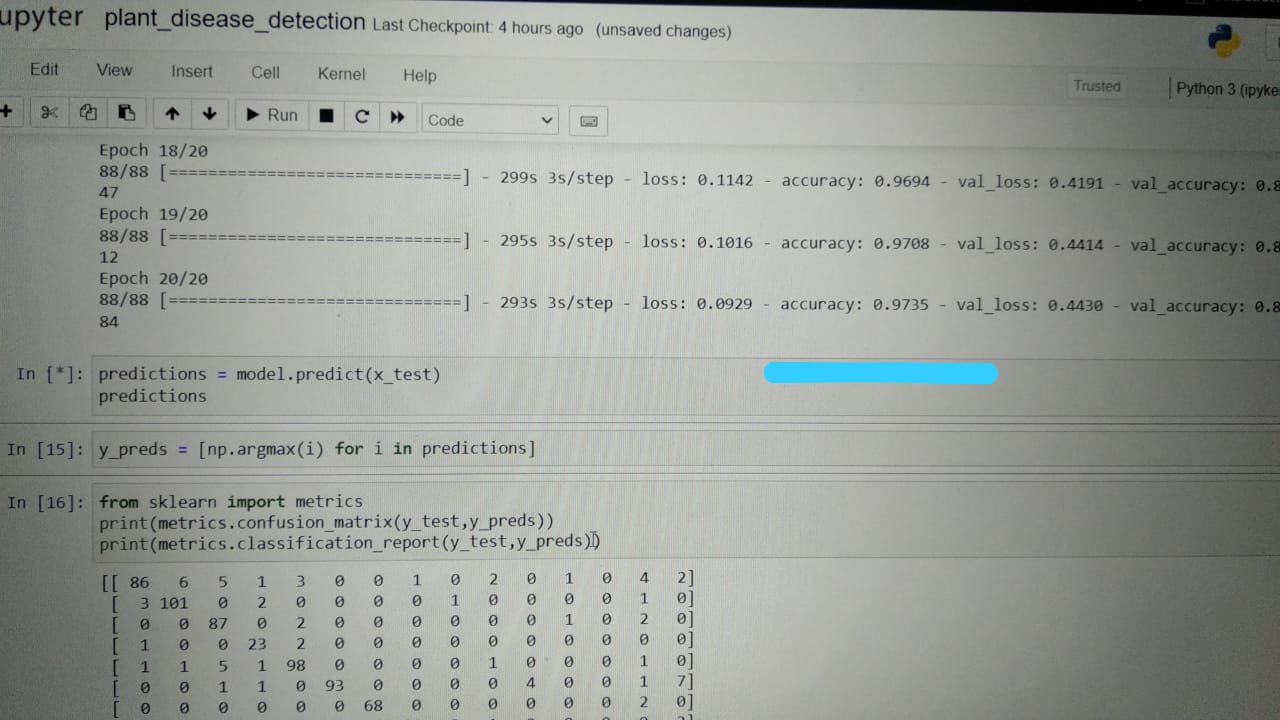


Fig. 13

**CONCLUSIONS AND FUTURE DIRECTIONS**

• The Plant Village dataset was used to evaluate the accuracy and performance of the respective DL models/architectures in the majority of the. Despite the fact that this dataset contains a large number of images of various plant species and their diseases, it has a simple/plain background. However, for a realistic scenario, the real environment should be taken into account.

• Hyperspectral/multispectral imaging is a new technology that has been used in a variety of fields of study. As a result, it should be combined with efficient DL architectures to detect plant diseases even before symptoms appear.

• A more efficient method of visualising disease spots in plants should be implemented to save money by avoiding the unnecessary application of fungicide/pesticide/herbicide. Because the severity of plant diseases varies over time, DL models should be improved so that they can detect and classify diseases throughout their entire life cycle.

• The DL model should be efficient for a wide range of lighting conditions, so the datasets should include images taken in a variety of field scenarios as well as images taken in the real world.

• A comprehensive study is required to understand the factors influencing plant disease detection, such as dataset classes and size, learning rate, illumination and so on.

**ACKNOWLEDGEMENT**

The authors would like to thank Mrs. Inderjeet Kaur for her continuous motivation and guidance for the development of this paper.

**REFERENCES**

[1] Identification of Plant-Leaf Diseases Using CNN and Transfer-Learning Approach, Sk Mahmudul Hassan, Arnab Kumar Maji, Michal Jasi ´nski , Zbigniew Leonowicz and Elzbieta Jasi nska , Electronics 2021, 10, 1388. , https://doi.org/10.3390/ electronics10121388

[2] Mercelin Francis, “Disease Detection and Classification in Agricultural Plants Using Convolutional Neural Networks: A Visual Understanding”, 2019 6th international Conference on Signal Processing and Integrated Networks (SPIN). (7-8 March,2019)

[3] Mohanty, S. P., Hughes, D. P., & Salathé, M. (2016). Using deep learning for image-based plant disease detection. Frontiers in Plant Science, 7(September), [1419]. <https://doi.org/10.3389/fpls.2016.01419>.

[4] Mohit Agarwal, Abhishek Singh, Siddhartha Arjaria, Amit Sinha, Suneet Gupta, “ToLeD: Tomato Leaf Disease Detection using Convolution Neural Network” - 2020

[5] Sammy V. Militante, “Plant Leaf Detection and Disease Recognition using Deep Learning” - 2019 IEEE Eurasia Conference on IOT, Communication and Engineering. (3- 6 October,2019)

[6] S. Arivazhagan, R. Newlin Shebiah, S. Ananthi, S. Vishnu Varthini. 2013. Detection of unhealthy region of plant leaves and classification of plant leaf diseases using texture features. Agric Eng Int: CIGR Journal.

[7] Plant Disease Detection using CNN, September 2020 DOI:[10.13140/RG.2.2.36485.99048](http://dx.doi.org/10.13140/RG.2.2.36485.99048) Project: [Plant Disease Recognition using CNN](https://www.researchgate.net/project/Plant-Disease-Recognition-using-CNN), Emma Harte School of Computing National College of Ireland Mayor Street, IFSC, Dublin 1 Dublin, Ireland ; [asked](https://www.goeduhub.com/10519/project-leaf-disease-detection-and-recognition-using-cnn) Jun 17, 2020 in [AI-ML-Data Science Projects](https://www.goeduhub.com/ai-ml-data-science-projects) by [Mohit Bansal](https://www.goeduhub.com/user/Mohit+bansal) (471 points) edited Jun 18, 2020 by [Mohit Bansal](https://www.goeduhub.com/user/Mohit+bansal)

[8] Plant Disease Detection using Convolutional Neural Network Sanmati RM1, Utkarsh Srivastava2, Vaishnavi S Korlahalli3, Varshitha K4 1,2,3,4Dept. Of Computer Science and Engineering, Sir .M Visvesvaraya Institute of Technology, Karnataka, India, 08 Issue: 07 | July 2021